



Universidad Nacional Autónoma de México
Programa de Posgrado en Ciencias de la Administración

Modelo de *Big Data Analytics* para la toma de decisiones en las organizaciones con manejo de grandes volúmenes de datos

T e s i s

Que para optar por el grado de:

Maestra en Administración

Presenta:

Nayana María Guerrero Ramírez

Tutor:

M.A. Stephen García Garibay
Facultad de Contaduría y Administración

Ciudad de México, abril de 2017



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INDICE

Índice de figuras	4
Índice de tablas	5
Introducción	7
Capítulo 1: Big Data: una herramienta indispensable para las organizaciones	11
1.1 Concepto de Big Data	11
1.2 Surgimiento de Big Data	13
1.3 Analíticos de Grandes Volúmenes de Datos (Big Data Analytics)	16
1.4 Herramientas para el análisis de grandes volúmenes de datos	23
1.4.2 Ciencia de datos	30
1.4.3 Datos abiertos como política pública	32
1.5 Big Data en la actualidad	35
1.6 Impacto de Big Data en la toma de decisiones en las organizaciones	39
1.7 Desafíos del uso de Big Data en las organizaciones	48
1.8 Resumen de aportaciones	54
Capítulo 2: Modelos de Big Data	56
2.1 Marco de trabajo de Bill Schmarzo	56
2.1.1 Técnicas de creación de valor de Bill SCHMARZO	59
2.1.2 Factores motrices del valor de Big Data	61
2.2 Modelo de Gobierno de Datos WebGestiones de Dr. Gabriel Guerrero	68
2.3 Comparativo del marco de trabajo de Bill Schmarzo y de Modelo de gobierno de datos de Dr. Gabriel Guerrero	71
2.4 Resumen de aportaciones	74
Capítulo 3: Marco referencial	76
3.1 Documento de estrategia para instrumentar una solución Big Data para Starbucks por Bill Schmarzo	76
3.2 Aplicación de las técnicas de creación de valor de Bill Schmarzo	80
2.4.1 Cuatro factores motrices de Big Data: Ejemplo de merchandising en Foot Locker	81
3.3 Resumen de aportaciones	83
Capítulo 4: Metodología	85
4.1 Planteamiento del problema	85
4.2 Preguntas de investigación	87
4.2.1 General	87
4.2.2 Específicas	87
4.3 Objetivos de la investigación	88
4.3.1 General	88
4.3.2 Específicos	88
4.4 Hipótesis de trabajo	89
4.5 Cuadro de Congruencia	90
4.6 Taxonomía del estudio	91
4.7 Diseño de la investigación	91
4.8 Procedimiento para la construcción del modelo	92
4.8.2 Fase de inicio	95

4.8.3 Fase de Recolección, construcción y análisis	96
4.8.4 Fase de teorización y construcción del modelo	97
4.9 Modelo propuesto	99
4.9.2 Comparativo del modelo propuesto respecto a los modelos existentes	110
4.10 Validación del modelo propuesto	111
4.10.2 Instrumento de medición	111
4.10.3 Selección de los informantes	111
4.10.4 Resultados de las entrevistas	112
4.11 Ajustes al modelo	114
Capítulo 5 Modelo Final	116
Conclusiones	133
Bibliografía	138
Glosario	146
Apéndices	148
Análisis Bibliométrico	148
Resumen general enviado a los informantes	158
Guión empleado en la entrevista	184

Índice de figuras

Figura 1 Las 7 V's de Big Data	15
Figura 2 Diagrama de las relaciones existentes entre cada proceso del modelo de minería de datos y las tecnologías Microsoft SQL Server	27
Figura 3 Modelo de estrategia para Big Data por Bill Schmarzo	59
Figura 4 Big Data guía el proceso de creación de valor.	61
Figura 5 Documento de estrategia de Big Data para Starbucks	80
Figura 6 Fases del Proceso de Investigación	94
Figura 7 Diseño general de investigación	96
Figura 8 Modelos	98
Figura 9 Fuentes generadoras de grandes volúmenes de datos	100
Figura 10 Big Data Diamond	108
Figura 12 FASE I: entradas y salidas	121
Figura 13 FASE II: entradas y salidas	122
Figura 14 FASE III: entradas y salidas	123
Figura 15 FASE IV: entradas y salidas	125
Figura 16 FASE V: entradas y salidas	127
Figura 17 FASE VI: entradas y salidas	128
Figura 18 FASE VII: entradas y salidas	129
Figura 19 FASE VIII: entradas y salidas	131
Figura 20 Modelo de Big Data Analytics para la toma de decisiones	132
Figura 22 Histograma por frecuencia sobre el campo SC de Big Data	150
Figura 23 Red simétrica sobre el campo C1 de Big Data	151
Figura 24 Red simétrica sobre el campo (DE) de Big Data	152
Figura 25 Red simétrica sobre el campo (ID) de Big Data	153
Figura 26 Visualización de red con VOSviewer de Big Data	154
Figura 27 Gráfica de coocurrencia de palabras clave de Big Data	155
Figura 28 Gráfica temporal para la identificación de términos emergentes de palabras clave de Big Data	156
Figura 29 Fuentes generadoras de grandes volúmenes de datos	170
Figura 30 Big Data Diamond	176

Índice de tablas

Tabla 1 Tipos de análisis que se requieren para aplicar Big Data Analytics _____	21
Tabla 2 Sistemas de minerías de datos más utilizados en el mercado y las técnicas utilizadas _____	26
Tabla 3 Los cuatro factores motrices de Big Data _____	63
Tabla 5 Modelo de gobierno de datos WebGestiones de Dr. Gabriel Guerrero ____	71
Tabla 6 Comparativo del modelos _____	72
Tabla 7 Comparativo sobre cómo implementar un modelo de Big Data _____	74
Tabla 9 Selección de informantes _____	112
Tabla 10 Proceso Administrativo integrado al modelo final _____	119

Agradecimientos

A mis padres, Gloria y Gabriel por que han fomentado en mi el espíritu investigativo. A mi esposo, Carlos por ser un soporte fundamental durante mis estudios y en mi vida. A mis hermanos, Ianis, Rachid y Alister por motivarme en todo momento y por brindarme una vida llena de alegría y experiencias compartidas.

A mis sinodales, Dra. Gabriela Bribiesca Correa, Dr. Roberto Ernesto López Martínez, M.A. Rita Aurora Fabregat Tinajero, M. en E.A.A.D Alfredo Corona Cabrera y M.A. Stephen García Garibay por su retroalimentación y tiempo destinado a esta tesis.

A la máxima casa de estudios, la Universidad Nacional Autónoma de México, por ofrecerme un espacio de aprendizaje para ampliar mis conocimientos a través de los estudios de maestría.

Nayana María Guerrero Ramírez

Introducción

El crecimiento del universo digital y la acumulación de datos en las organizaciones puede ser alarmante para las mismas. No obstante, si se emplean las herramientas adecuadas, los datos se pueden convertir en información relevante para obtener mejores resultados y cifras que impacten de forma positiva a dichas organizaciones. Hoy en día, se requieren mejores decisiones con información relevante proveniente de todas las áreas de la organización.

El uso de Big Data impacta en toda la organización ya que su buen uso ayuda a descubrir nuevas posibilidades de monetización. Efectivamente, con esta herramienta se pueden identificar nuevos elementos clave de la empresa con una rapidez que permita tomar las medidas necesarias para la creación de valor y de esta forma, obtener conocimientos empresariales en tiempo real.

Algunas organizaciones conocen el término *Big Data Analytics* pero desconocen cómo se podría aplicar para potenciar sus principales iniciativas institucionales. Su impacto en las organizaciones tiene un gran potencial en los procesos de creación de valor empresariales.

*Big Data*¹ ha venido a revolucionar la forma en que podemos aprovechar las nuevas fuentes de datos y las nuevas capacidades analíticas para contestar preguntas clave con el objetivo de descubrir nuevos conocimientos sobre la toma de decisiones. Esta herramienta permite “tomar decisiones con un nivel más de detalle para poder descubrir nuevos conocimientos sobre nuestros clientes, productos y operaciones y aplicarlos para responder a las preguntas empresariales clave. Asimismo, perfecciona y acelera nuestra capacidad para

¹ “Big Data es de alto volumen, alta velocidad y los activos de información de gran variedad que exigen, formas innovadoras y rentables de procesamiento de la información para mejorar la comprensión y la toma de decisiones.” (LANEY, 2001)

identificar áreas específicas del negocio y procesos empresariales específicos en lo que pueden devolver un valor empresarial inmediato.” (SCHMARZO, 2014)

Esta investigación tiene como objetivo el diseño de un modelo de *Big Data Analytics* que se pueda aplicar a las organizaciones con manejo de grandes volúmenes de datos para mejorar la toma de decisiones a nivel directivo.

Mediante el estudio en la literatura de algunos modelos de *Big Data Analytics* en las organizaciones y la investigación de su definición, se pretende dar a conocer la importancia de esta herramienta en la toma de decisiones así como considerar si sirve como un proceso de creación de valor para las organizaciones e identificar las etapas que debe incluir el desarrollo de un modelo de *Big Data Analytics*.

Para esta investigación se retoman los modelos *Big Data Analytics* de los científicos de datos Bill Schmarzo y Gabriel Guerrero. Estos modelos permiten identificar dónde y cómo pueden repercutir los Big Data en la toma de decisiones en una organización.

En el primer capítulo se analiza el surgimiento, las definiciones y la importancia de los Analíticos de Grandes Volúmenes de Datos, mejor conocidos como *Big Data Analytics*, en la toma de decisiones a través de la información y la bibliografía de dos variables; la primera, el desarrollo de esta herramienta que se acompaña de una aceleración del volumen de datos que manejan las organizaciones; la segunda, mediante el estudio de autores como LANEY Doug (2001), LOSHIN David (2013) y SCHMARZO Bill (2013) quienes subrayan que los métodos y tecnologías tradicionales no son suficientes para la toma de decisiones. *Big Data Analytics*, de acuerdo a estos autores es una herramienta insoslayable en la época actual para las organizaciones.

En el segundo capítulo, se estudian y comparan dos modelos de *Big Data*: un modelo de gobierno de datos mexicano creado por el experto científico de datos Dr. Gabriel Guerrero y otro modelo para el análisis de grandes volúmenes de datos diseñado por Bill Schmarzo autor del libro “*Big Data: el poder de los datos*”. Se examina la utilización de tecnologías para la administración de la información así como las técnicas de creación de valor para enfrentar el desafío del crecimiento de los datos en las organizaciones. Asimismo, se estudian las buenas prácticas de ambos modelos para el desarrollo del modelo propuesto en esta investigación.

En el tercer capítulo se estudia el impacto empresarial de *Big Data Analytics* que como menciona Bill Schmarzo (2013) *Big Data* “está más relacionados con la transformación de las empresas que con la transformación de las tecnologías de información” (SCHMARZO, 2014) Este autor sostiene que “*Big Data* ha cambiado los matices para definir y cuantificar términos como valioso, importante y exitoso. Las nuevas fuentes de datos, junto con las nuevas posibilidades de analítica avanzada, dan lugar a respuestas más precisas y proporcionan un entendimiento más completo de las organizaciones.” (SCHMARZO, 2014) Este capítulo contiene el marco referencial que es un apoyo para el desarrollo del modelo que se muestra en el cuarto capítulo. Se analiza la aplicación de modelos de *Big Data Analytics* en las organizaciones y se estudia el caso de dos empresas trasnacionales: Starbucks y Foot Locker. En este marco, se destaca cómo empresas de clase mundial como *Starbucks* y *Foot Locker*, mediante el uso de *Big Data Analytics* han sabido sacar provecho de los datos para convertirlos en información valiosa para la toma de decisiones.

En el cuarto capítulo se analiza la metodología empleada en dicha investigación y su validez de constructo. Se muestra el desarrollo del marco teórico que comprende la identificación de variables que rige la administración de la tecnología y el uso de herramientas tecnológicas como *Big Data Analytics*. Asimismo, en dicho capítulo se muestra el modelo propuesto, su validación mediante diversos

instrumentos de medición y los ajustes necesarios provenientes de la validación de expertos.

El quinto y último capítulo refleja el modelo propuesto considerando la retroalimentación de los informantes.

Ulteriormente, se presentan las conclusiones de la investigación, la bibliografía consultada, el glosario y los apéndices del presente estudio.

En este marco, sin embargo, en el ámbito académico se cuenta con limitada información y/o bibliografía sobre el *Big Data Analytics* para analizarlo como una herramienta para la toma de decisiones. Se constató que existe una limitada información académica sobre este tema; por estas razones, se busca, a través de esta investigación, aportar mayor información a la ciencia de la administración respecto al uso de esta herramienta para la toma de decisiones. Por lo tanto desde la perspectiva de la administración y la tecnologías de información, considero el tema de *Big Data Analytics* para la toma de decisiones es un asunto interesante a profundizar y aplicar, no sólo para los directivos sino para la sociedad en general, la cual puede sacar provecho de estas nuevas herramientas y sin duda será próximamente un objeto de estudio insoslayable en la disciplina de administración de la tecnología para la toma de decisiones de las organizaciones.

Capítulo 1: Big Data: una herramienta indispensable para las organizaciones

“La naturaleza benigna provee de manera que en cualquier parte halles algo que aprender.”
Leonardo Da Vinci

En este primer capítulo se identifican los conceptos principales de la investigación. Se definen también los elementos que facilitan la comprensión sobre la importancia del uso de nueva técnicas de análisis de grandes volúmenes de datos, razón por la cual iniciamos con la definición de *Big Data*, su origen, los conceptos relacionados con el término, así como los objetivos y la importancia de su uso en las organizaciones.

Se termina estudiando el impacto del uso de Big Data en las organizaciones de manera que resulte útil y atractivo para la toma de decisiones a nivel directivo.

1.1 Concepto de Big Data

A continuación se presentan algunas de las más relevantes definiciones de Big Data para tener una mayor comprensión del concepto; el cual refleja de manera general, la necesidad de contar con una herramienta que transforme la gran cantidad de datos generados digitalmente en información veraz y eficaz para una óptima toma de decisiones a nivel directivo.

- La empresa pionera en Big Data, Meta Group, ahora Gartner, define en su glosario de TI, Big Data como: Big Data es de alto volumen, alta velocidad y los activos de información de gran variedad que exigen, formas innovadoras y rentables de procesamiento de la información para mejorar la comprensión y la toma de decisiones. (GARTNER Inc., 2016)

- Para el analista Doug LANEY, Big Data es de alto volumen, alta velocidad y los activos de información de gran variedad que exigen, formas innovadoras y rentables de procesamiento de la información para mejorar la comprensión y la toma de decisiones. (LANEY, 2001)
- Gloria PÉREZ de la Agencia de Salud Pública de Barcelona afirma que este término se refiere a los grandes volúmenes de información compleja y conectable que crece continuamente, de modo que la información parece duplicarse cada 2 años, y este fenómeno podría estar acelerando. (PÉREZ G. , 2015)
- La Universidad de Oxford define Big Data como la Informática de datos de un gran volumen, por lo general a través de su manipulación y su gestión presenta desafíos logísticos significativos; así como la rama de la informática que afectan a dichos datos. (Oxford University Press, 2016)
- La consultora estratégica global, McKinsey Global Institute, define Big Data como el conjuntos de datos cuyo tamaño va más allá de la capacidad de las herramientas típicas de software de bases de datos para capturar, almacenar, gestionar y analizar (McKinsey Global Institute (MGI), 2012)
- Los expertos en Big Data MAYER-SCHÖNBERGER y CUKIER, Big Data es la capacidad de la sociedad para aprovechar la información en nuevas formas de producir información útil o bienes y servicios de valor significativo " y" ... las cosas que uno puede hacer a gran escala que no se puede hacer en una más pequeña , para extraer nuevos conocimientos o crear nuevas formas de valor . (MAYER-SCHÖNBERGER & CUKIER, 2013)
- Tom DAVENPORT afirma que Big Data es la amplia gama de nuevos y masivos tipos de datos que han aparecido en la última década. (DAVENPORT, Big Data at Work: Dispelling the Myths, Uncovering the Opportunities, 2014)

Las definiciones anteriormente mencionadas coinciden en el volumen de los datos generados que requieren ser manipulados para crear nuevas formas de valor.

Cabe señalar que para esta tesis, se retoman las definiciones anteriormente mencionadas y se define *Big Data* como:

Herramienta que permite velozmente, capturar, almacenar, gestionar y analizar un conjuntos de datos de alto volumen, variabilidad, veracidad, visibilidad, valor y variedad cuyo tamaño va más allá de la capacidad de software de bases de datos relacionales para extraer nuevos conocimientos, crear nuevos formas de valor y exigir formas innovadoras y rentables de procesamiento de la información para mejorar la comprensión y la toma de decisiones.

1.2 Surgimiento de Big Data

El crecimiento del universo digital está considerado del 40% anual en la próxima década. (EMC MEXICO, 2014) No sólo por el incremento de la población mundial sino también por el hecho que las personas y las empresas realizan todas sus actividades en línea aunado a todas las “cosas” como dispositivos inteligentes que están conectados a internet y que están desatando una nueva ola de oportunidades de negocios alrededor del mundo. Al igual que el universo físico, el universo digital es enorme. Para el año 2020 se prevé que existan casi tantos bits digitales como estrellas en el universo.

El volumen de datos masivos crece constantemente y todos hacemos parte de dicho crecimiento. Así, debido al rápido avance tecnológico, las empresas se han tenido que enfrentar a nuevos desafíos como el analizar los grandes volúmenes generados para poder tomar decisiones rápidamente. Hoy, emplear herramientas tradicionales para analizar los datos generados segundo a segundo por las nuevas tecnologías han quedado obsoletas por lo que se requieren nuevos métodos para poder enfrentar las necesidades de los usuarios.

De la revolución digital que estamos viviendo en el siglo XXI surgen nuevos conceptos como Big Data, *Big Data Analytics*, Ciencia de Datos, entre otros, que hacen referencia a la acumulación masiva de datos y a las técnicas usadas para encontrar patrones dentro de esos datos.

Según Doug LANEY, Big Data se caracteriza por su *slogan* de las tres “V”s: volumen, velocidad y variedad. Estas tres “V”s surgen de un artículo titulado “3D Data Management Controlling Data: Volume, Velocity and Variety” (LANEY, 2001) publicado el 6 de febrero de 2001 por Meta Group (ahora Gartner) donde no se menciona explícitamente el término Big Data pero se constata del incremento de los formatos de datos, su volumen y se recomienda una mejor gestión de los datos. Doug Laney, analista de META Group, es uno de los pioneros del tema y subraya la importancia del análisis de grandes volúmenes de datos en congresos y presentaciones relacionadas con el tema; define el crecimiento constante de los datos como una oportunidad y un reto para investigar lo que más tarde, IBM denominó las 4V’s que hacen alusión al volumen, variedad, veracidad y visibilidad. Desde entonces muchos autores han intentado profundizar en el tema incluyendo otros tres elementos otros como la Veracidad, Valor y Visibilidad de los datos.

Figura 1 Las 7 V's de Big Data



Fuente: Elaboración propia

En efecto, el creciente volumen de información disponible implica retos ambicioso, técnicos y analíticos que los métodos tradicionales no logran vislumbrar en un tiempo de respuesta casi inmediato. En cuanto a la variedad, la facilidad de enviar un correo electrónico con datos de texto, mensaje de voz y fotos ha generado una miríada de información de datos totalmente heterogéneos. La veracidad hace mención de la calidad del dato como su predictibilidad; pero nada de esto sería extraordinario sin el elemento más importante que es la velocidad con la que es necesario analizar la información y sobre todo tomar decisiones acertadas

Los datos masivos nos impulsan a aferrarnos al potencial que supone analizar todos los datos ya sean internos o externos de la organización en tiempo real. Para realizar esto, se requiere de una infraestructura actualizada y moderna que pueda sacar partido de la riqueza de los datos. “Estamos ante una tecnología

sería, cuyo valor para crear nuevos modelos de negocio basados en petabytes² de datos en tiempo real, acompañados de analíticas avanzadas, ya han sido validados por varios sectores tan diversos como el minorista, los servicios financieros, las telecomunicaciones, la fabricación, la energía, el transporte y la hostelería.” (SCHMARZO, 2013) La gran cantidad de datos generados requiere de una infraestructura actualizada para aprovechar y analizar los datos y transformarlos en información.

1.3 Analíticos de Grandes Volúmenes de Datos (*Big Data Analytics*)

Para entender qué son los Analíticos de Grandes Volúmenes de Datos es importante definir lo que es la analítica.

De acuerdo a la Real Academia Española, la analítica significa “perteneciente o relativo al análisis” y “que procede descomponiendo, o que pasa del todo a las partes”. (Real Academia Española).

El Dr. Guerrero, experto científico de datos subraya que los datos se pueden fraccionar en dos categorías según su análisis:

- Análisis MICRO que es el análisis del dato al detalle cuyo acomodo es acorde a la trazabilidad, es decir con acceso a expedientes particulares de una persona.
- Análisis MACRO que es el análisis global del dato cuyo acomodo es para un estudio integral o global con estadísticas sobre los datos.

Por una parte, la capacidad para reconstruir el historial de un producto, es decir, la trazabilidad también se le conoce como búsqueda al átomo.

Por otra parte, el análisis global o macro, indica las estadísticas globales mediante el uso de estadística descriptiva. El análisis global ofrece gráficas descriptivas.

² Unidad de medida de almacenamiento informático simbolizado como PB. Un petabyte equivale a 1024 terabytes. (ALEGSA, 2016)

A su vez, la información del análisis macro o global se subdivide en 3 tipos:

1. **Análisis descriptivo:** En este punto, se pueden transformar y enriquecer los datos mediante técnicas matemáticas para establecer nuevas métricas en relación a los datos como: la frecuencia (¿Cada cuánto?), la proximidad (¿Cuánto hace?) y la secuencia (¿En qué orden?). Este análisis describe lo que ha pasado a través de la pregunta: ¿Cuántos...? Puede ser cuántos hombres hay, cuántas mujeres, cuántos viven en Miguel Hidalgo, etc.
2. **Análisis inferencial o predictivo**³: Hace uso de la estadística inferencial que permite inferir lo que puede pasar en un futuro según la tendencia que ha mostrado la información. Se determina qué podría pasar en el futuro si se continúa con la misma tendencia.
3. **Análisis analítico o prescriptivo:** Se utiliza la estadística analítica para correlacionar diversas variables que a simple vista no tienen un estrecho vínculo. “Mediante la prescripción, el propio sistema de análisis realiza recomendaciones sobre las acciones que se han de seguir para reducir costes o mejorar los beneficios, consiguiendo en todos los casos un incremento del ROI.” (Instituto de Ingeniería del Conocimiento, 2016) En este marco, se muestran, los problemas y posibles soluciones que se encuentran escondidos en los datos, para identificar patrones de comportamiento y generar recomendaciones.

El análisis descriptivo e inferencial se le conoce, también, como aprendizaje supervisado puesto que se cuenta con conocimiento de la clasificación a priori, es decir que ya están clasificados los datos.

Contrariamente, el análisis analítico o prescriptivo que se le conoce como de aprendizaje no supervisado, usa algoritmos para intentar construir patrones de un conjunto de datos del que no se cuenta con conocimiento *a priori* y que a simple

³ “La analítica predictiva proporciona herramientas para estimar aquellos datos de negocio que son desconocidos o inciertos, o que requieren de un proceso manual o costoso para su obtención.” (Instituto de Ingeniería del Conocimiento, 2016)

vista no tienen una relación directa entre ellos. En este caso es importante procesar los datos y agruparlos para permitir su clasificación y descubrir grupos de objetos con características afines.

En el ámbito de la estadística existe lo que se denomina como método estadístico que es el método científico aplicado a la estadística y se divide en dos fases: la descriptiva y la analítica.

“El método estadístico, dentro del método científico, consiste en una serie de pasos para llegar al verdadero conocimiento estadístico.

Etapas del método estadístico:

- a. Recogida de datos
- b. Ordenación y presentación de los datos en tablas simples o de doble entrada
- c. Determinación de medidas o parámetros que intenten resumir la cantidad de información
- d. Formular hipótesis sobre las regularidades que se presenten
- e. Por último, el análisis estadístico formal que permita verificar las hipótesis formuladas.” (GARCÍA RAMOS, RAMOS GONZÁLEZ, & RUIZ GARZÓN, 2007)

En este marco, por una parte, “la estadística descriptiva trata de organizar, representar y resumir un conjunto de datos de manera que pueda ser extraída la máxima información procedente de ellos”. (GARCÍA RAMOS, RAMOS GONZÁLEZ, & RUIZ GARZÓN, 2007) Por otra parte, la estadística analítica y la estadística inferencial “intenta sacar conclusiones respecto del fenómeno sobre el cual se ha tomado los datos”. (GARCÍA RAMOS, RAMOS GONZÁLEZ, & RUIZ GARZÓN, 2007)

Existen diversos tipos de analítica:

- **Analítica descriptiva⁴:** Se refiere al uso de métodos estadísticos para predecir el comportamiento de los usuarios con ayuda de la minería de datos para la predicción de probabilidades y tendencias. “La analítica predictiva pretende la mejora de decisiones y respuestas rápida al cambio y comprende técnicas de minería de datos, estadística y modelado; trata de analizar hechos actuales o históricos con el propósito de hacer predicciones sobre sucesos futuros. Desde un punto de vista corporativo, trata de predecir el comportamiento de sus diferentes categorías de usuarios, especialmente los clientes. » (Big Data Social, 2016)
- **Analítica de redes:** Se utiliza para el análisis de “las relaciones entre nodos discretos de un grado o de una red.” (Big Data Social, 2016) Su mayor utilización es en la redes sociales donde mediante la analítica de redes se pueden conocer las conexiones entre diversos objetos, personas u organizaciones. La analítica de redes ayuda a las organizaciones a medir su impacto en las redes sociales con los objetivos de la empresa y en las conversaciones relacionadas.
- **Analítica de sentimientos:** La minería de opiniones o analítica de sentimientos, es una herramienta utilizada por las empresas para identificar y extraer información subjetiva de contenido de diversos tipos como el análisis de medios sociales, el análisis de la voz del cliente a través de encuestas o el análisis de la voz del ciudadano. Hace uso de tecnologías de procesamiento del lenguaje natural, analítica de textos y lingüística computacional. El objetivo de la analítica de sentimientos es “deducir de este análisis, comportamiento de clientes, hábitos de compra, sistemas de recomendación, etc. Este tipo de analítica se asienta en la llamada minería de opinión que implica la construcción de un sistema para recolectar y examinar las opiniones relativas a un producto, hechas en entradas (posts)

⁴ “La analítica descriptiva consiste en almacenar y realizar agregaciones de datos históricos, visualizándolos de forma que puedan ayudar a la comprensión del estado actual y pasado del negocio. La analítica descriptiva nos cuenta cómo ha funcionado nuestro negocio hasta la fecha.” (Instituto de Ingeniería del Conocimiento, 2016)

de blogs, comentarios, revisiones o tuits (tweets). La minería de opinión automatizada utiliza técnicas de *machine learning*, uno de los componentes clave de la inteligencia artificial. » (Big Data Social, 2016)

- **Analítica de textos:** Se basa en la minería de textos para “descubrir, a partir de cantidades de texto grandes, el conocimiento que no está literalmente escrito en cualquiera de los documentos. Esto incluye buscar tendencias, promedios, desviaciones, dependencias, etc.” (TRIPOD, 2016)
- **Analítica web:** “Es la recogida y análisis de los datos que se registran cuando un usuario visita un sitio web y sirve para medir el comportamiento de los usuarios. Existen numerosas herramientas de analítica de datos: Google Analytics, Woopra, Omniture, Adobe Analytics, etc.” (Big Data Social, 2016)

El crecimiento constante de datos requiere que los datos sean analizados por expertos que se denominan científicos de datos para sacar conclusiones mediante distintos procesos como la minería de datos y los Analíticos de Grandes Volúmenes de Datos.

Una vez entendido el concepto de Big Data, es preciso dilucidar como analizar los grandes volúmenes de datos. Una empresa que ha sabido sacar provecho de los Analíticos de Grandes Volúmenes de Datos es Amazon. El gigante estadounidense de comercio electrónico, Amazon, usa el conocimiento que tiene de gustos y preferencias de sus clientes para ofrecer el producto indicado que están buscando. Usando *Big data Analytics*, las empresas pueden saber el comportamiento de compra de los clientes por lo que desde un punto de vista mercadológico ha tenido mucho éxito.

Es este aspecto, Nicoletti define los Analíticos de Grandes Volúmenes de Datos como “el proceso de examinar grandes conjuntos de datos que contienen una variedad de tipos de datos para descubrir patrones ocultos, correlaciones

desconocidas, tendencias del mercado, preferencias de los clientes, posibles riesgos y otra información útil relacionada”. (NICOLETTI, 2016)

Por su parte, los expertos de Tecnologías de información Davenport y Harris, afirman que *Big Data Analytics* se caracteriza por “el uso extensivo de datos, análisis estadístico y cuantitativo, modelos explicativos y predictivos y la gestión basada en los hechos para tomar decisiones y acciones” (DAVENPORT & G.HARRIS, *Competing on Analytics: The New Science of Winning*, 2007)

Big Data Analytics, que COX define como “el proceso científico de transformar los datos en conocimiento para tomar mejores decisiones.” (COX, 2015) , requiere conocer en un inicio cual es el problema que se desea resolver al utilizar esta herramienta para tener un mayor impacto. En este aspecto se puede catalogar en cuatro tipos de análisis: básico, avanzado, operacionalizado y monetizado (*basic analytics, advanced analytics, operationalized analytics, monetized analytics*).

Tabla 1 Tipos de análisis que se requieren para aplicar *Big Data Analytics*

Tipo de Análisis	Descripción
Analíticos básicos para comprensión	Reordenamiento de los datos , informes , visualizaciones simples , monitoreo básico
Analíticos avanzados para comprensión	Un análisis más complejo como el modelado predictivo y otras técnicas de búsqueda de patrones
Analíticos operacionalizados	Los analíticos se convierten en parte del proceso de negocio
Analíticos monetizados	Los analíticos se usan para generar ingresos.

Fuente: HURWITZ J., NUGENT A., HALPER F., KAUFMAN M. (2013). *Big Data*.

De manera más detallada, como lo describe Judith Hurwitz en su libro *Big Data*, los analíticos básicos pueden usarse para explorar datos si aún no se tiene identificado el proceso pero se conoce el valor de los datos. En este marco, se muestran visualizaciones simples de los datos utilizando estadísticas simples. El análisis básico, a menudo se utiliza cuando se tiene una gran cantidad de datos heterogéneos. (HURWITZ, NUGENT, HALPER, & KAUFMAN, 2013)

Los analíticos avanzados proporcionan algoritmos complejos para datos estructurados y no estructurados. Incluyen modelos estadísticos sofisticados, redes neuronales, analíticos de textos y otras técnicas avanzadas de minería de datos. Los analíticos avanzados se pueden implementar para encontrar patrones en los datos, predicción y previsión.

Cuando se operacionalizan los analíticos significa que hacen parte del proceso de negocio y se desarrolla un modelo para cada proceso. Es el caso en empresas de venta por teléfono (*call centers*) donde se tiene un modelo predictivo creado específicamente para la empresa. De esta manera, el modelo lanza un mensaje al operador para indicarle que cliente hace parte del mercado meta. Así, el modelo puede predecir cuales son los clientes que muy probablemente comprarán el producto adicional que les ofrecerá el operador. (HURWITZ, NUGENT, HALPER, & KAUFMAN, 2013)

Por último, *Big Data Analytics* puede usarse para monetizar los datos. En efecto, instituciones financieras actualmente utilizan esta herramienta para incrementar sus ingresos haciendo un cruce de diversas fuentes de información de sus clientes como: facturación, geolocalización y mensajes de texto que proveen información sobre patrones de comportamiento, no obstante, en este marco, entra en juego el aspecto de protección de datos personales que veremos más a continuación.

1.4 Herramientas para el análisis de grandes volúmenes de datos

1.4.1 Minería de datos

La minería de datos es un proceso que se utiliza para procesar grandes volúmenes de datos. Este proceso se utiliza en distintas áreas desde mercadotecnia, banca, seguros, educación, análisis de mercado, medicina, biología hasta sistemas, telecomunicaciones, procesos industriales entre muchos otros campos. Se utiliza un análisis matemático para descubrir patrones ocultos denominado descubrimiento de conocimiento en bases de datos (*knowledge discovery in databases*). Como lo define Gloria Pérez de la agencia de salud pública de Barcelona “La minería de datos es la exploración automática o semiautomática de los grandes conjuntos de datos con la intención de descubrir patrones. Es uno de los pasos que componen el proceso del *knowledge discovery in databases*, en el cual se incluyen la recolección y la preparación de los datos, la interpretación de los resultados y la información de estos.” (PÉREZ G. , 2015) Generalmente, los patrones ocultos no pueden descubrirse mediante las herramientas tradicionales por lo que se hace uso de la minería de datos.

Como lo afirman los autores del libro *Minería de datos: técnicas y herramientas*, “las técnicas de minería de datos persiguen el mismo descubrimiento automático del conocimiento contenido en la información almacenada de modo ordenado en grandes bases de datos. Estas técnicas tienen como objetivo descubrir patrones, perfiles y tendencias a través del análisis de los datos utilizando tecnologías de reconocimiento de patrones, redes neuronales, lógica difusa, algoritmos genéticos y otras técnicas avanzadas de análisis de datos”. (PÉREZ LÓPEZ & SANTÍN GONZÁLEZ, 2007)

Cabe señalar que si bien la minería de datos es una herramienta para extraer patrones y tendencias, suele necesitar de la existencia previa de un almacén de datos (*dataware house*). Como lo indican Pérez López y Santín González “el almacén de datos es el sistema de información central en todo este proceso. Un

almacén de datos es una colección de datos orientada a un dominio, integrada, no volátil y variante en el tiempo de ayuda en la toma de decisiones. Un almacén de datos es un conjunto de datos históricos, interno o externos y descriptivos de un contexto o área de estudio, que están integrados y organizados de tal forma que permiten aplicar eficientemente herramientas para resumir, describir y analizar los datos con el fin de ayudar en la toma de decisiones estratégicas”. (PÉREZ LÓPEZ & SANTÍN GONZÁLEZ, 2007)

Ahora bien, los almacenes de datos no son completamente indispensables para extraer el conocimiento a partir de los datos, ya que se puede hacer minería de datos con un fichero de datos. Sin embargo, conforme aumenta la cantidad de datos y se suman datos de fuentes heterogéneas, se van conociendo los beneficios de organizar un almacén de datos para realizar minería de datos.

Se pueden clasificar las técnicas de minerías de datos en tres:

1. Técnicas predictivas
2. Técnicas descriptivas
3. Técnicas auxiliares

Las técnicas predictivas se caracterizan por tener variables que pueden clasificarse inicialmente en dependientes e independientes. Como lo afirma Pérez López “las técnicas predictivas especifican el modelo para los datos en base a un conocimiento teórico previo. El modelo supuesto para los datos debe contrastarse después del proceso de minería de datos antes de aceptarlo como válido.

Formalmente, la aplicación de todo modelo debe superar las fases de identificación objetiva (a partir de los datos se aplican reglas que permitan identificar el mejor modelo posible que ajuste los datos), estimación (proceso de cálculo de los parámetros del modelo elegido para los datos en la fase de identificación), diagnóstico (proceso de contraste de la validez del modelo estimado) y predicción (proceso de utilización del modelo identificado, estimado y validado

para predecir valores futuros de las variables dependientes). En algunos casos, el modelo se obtiene como mezcla del conocimiento obtenido antes y después del *Data Mining* y también debe contrastarse antes de aceptarse como válido”. (PÉREZ LÓPEZ & SANTÍN GONZÁLEZ, 2007)

Lo que caracteriza las técnicas descriptivas es que todas las variables tienen inicialmente el mismo estatus. “En las técnicas descriptivas, no se asigna ningún papel predeterminado a las variables. No se supone la existencia de variables dependientes ni independientes y tampoco se supone la existencia de un modelo previo para los datos. Los modelos se crean automáticamente partiendo del reconocimiento de patrones. En este grupo se incluyen las técnicas de *clustering* y segmentación (que también son técnicas de clasificación en cierto modo), las técnicas de asociación y dependencia, las técnicas de análisis exploratorio de datos y las técnicas de reducción de la dimensión (factorial, componentes principales, correspondencias, etc.) y de escalamiento multidimensional”. (PÉREZ LÓPEZ & SANTÍN GONZÁLEZ, 2007)

En relación a las técnicas auxiliares, “son herramientas de apoyo más superficiales y limitadas. Se trata de nuevos métodos en técnicas estadísticas descriptivas, consultas e informes y enfocados en general hacia la verificación.” (PÉREZ LÓPEZ & SANTÍN GONZÁLEZ, 2007)

El siguiente cuadro muestra los sistemas de minerías de datos más utilizados en el mercado y las técnicas utilizadas.

Tabla 2 Sistemas de minerías de datos más utilizados en el mercado y las técnicas utilizadas

Producto	Compañía	Técnicas	Plataforma	Interfaz
Knowledge Seeker	Angoss	Árboles de Decisión	Win	ODBC
CART	Salford Systems	Árboles de Decisión	Win/UNIX	
Clementine	SPSS	Amplio abanico	Win/UNIX	ODBC
Data Surveyor	Data Distilleries	Amplio abanico	UNIX	ODBC
Gain Smarts	Urban Science	Gráficos-Ganancias	Win/UNIX	
Intelligent Miner	IBM	Amplio abanico	UNIX (AIX)	IBM, DB2
Micostrategy	Micostrategy	Datawarehouse	Win	Oracle
Polyanalyst	Megaputer	Simbólicas	Win	Oracle, ODBC
Darwin	Oracle	Amplio abanico	Win/UNIX	Oracle
Enterprise Miner	SAS Institute	Amplio abanico	Win/UNIX/Mac	
SGI MineSet	Silicon Graphics	Asociación y Clasificación	UNIX	Oracle, Sybase, Informix
Wizsoft/Wizwhy	Wizsoft			

Fuente: Pérez López y Santín González (2007) *Minería de datos: técnicas y herramientas*.

Cabe señalar que existen diversos sistemas que permiten realizar minería de datos. Los más conocidos en el mercado son las bases de datos Oracle y Microsoft SQL Server.

Oracle dispone de herramientas de minería de datos complejas que integran inteligencia de negocio (*Business Intelligence*) y *Data Mining* cuya orientación es mayormente empresarial y de sistemas de información.

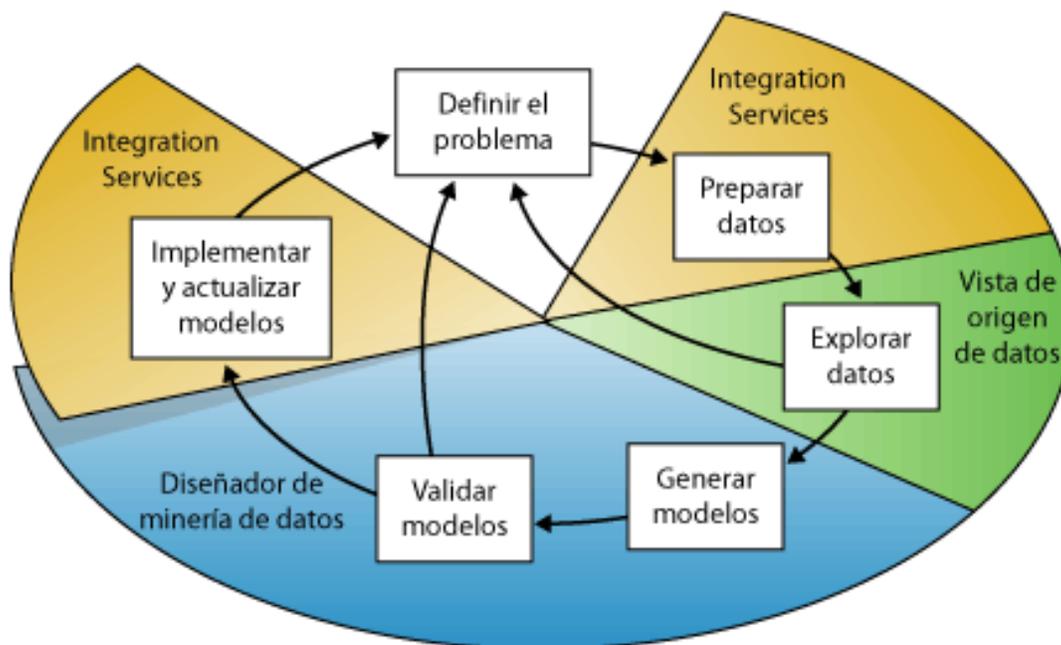
En la pagina oficial de Microsoft, se especifica que “la minería de datos de Microsoft SQL Server ofrece un entorno integrado para crear y trabajar con modelos de minería de datos. Este entorno incluye *SQL Server Development Studio*, que contiene algoritmos de minería de datos y herramientas de consulta que facilitan la generación de una solución completa para una serie de proyectos, y *SQL Server Management Studio*, que contiene herramientas que permiten examinar modelos y administrar objetos de minería de datos.

La generación de un modelo de minería de datos forma parte de un proceso mayor que incluye desde la formulación de preguntas acerca de los datos y la creación de un modelo para responderlas, hasta la implementación del modelo en un entorno de trabajo. Este proceso se puede definir mediante los seis pasos básicos siguientes:

1. Definir el problema
2. Preparar los datos
3. Explorar los datos
4. Generar modelos
5. Explorar y validar los modelos
6. Implementar y actualizar los modelos” (MICROSOFT, 2016)

A continuación se muestra el diagrama de las relaciones existentes entre cada proceso del modelo de minería de datos.

Figura 2 Diagrama de las relaciones existentes entre cada proceso del modelo de minería de datos y las tecnologías Microsoft SQL Server



Fuente: Microsoft (2016)

Aunado a las grandes empresas que tienen alta tecnología para el desarrollo de la minería de datos, también existen herramientas de minería de datos en el ámbito de software libre.

Por su parte, los autores del libro *Técnicas cuantitativas para la gestión en la ingeniería de software* afirman que “los algoritmos de aprendizaje automático se pueden englobar dentro de un grupo más general de técnicas denominadas minería de datos, cuyo propósito es extraer conocimiento no explícito a partir de grandes volúmenes de datos.” (TUYA, RAMOS ROMÁN, & DOLADO COSÍN, 2007) Efectivamente las técnicas de *data mining* o minería de datos se clasifican en dos categorías:

- Algoritmos supervisados o predictivos: “predicen el valor de un atributo (etiqueta) de un conjunto de datos, conocidos otros atributos (atributos descriptivos). A partir de datos cuya etiqueta se conoce se induce un modelo que relaciona dicha etiqueta y los atributos descriptivos. Y esas relaciones sirven para realizar la predicción en datos cuya etiqueta es desconocida.” (TUYA, RAMOS ROMÁN, & DOLADO COSÍN, 2007)
- Algoritmos no supervisados o de descubrimiento del conocimiento: “realizan tareas descriptivas como el descubrimiento de patrones y tendencias en los datos actuales (no utilizan datos históricos).” (TUYA, RAMOS ROMÁN, & DOLADO COSÍN, 2007)

Cabe subrayar que el revelar patrones de comportamiento no explícitos mediante algoritmos ya sea supervisados o predictivos es de gran utilidad para las organizaciones puesto que obtienen un beneficio tanto científico como de negocio.

1.4.1.1 Gobernanza de datos

El gobierno de datos o *data governance*⁵ es un tema de suma importancia para las organizaciones que manejan grandes volúmenes de datos. Jimmy Martínez, arquitecto de software de IBM subraya que “la correcta implementación de esta estrategia beneficia a las organizaciones en la generación de valor a partir de los datos y coordina y gestiona su sinergia con el área de negocio.” (MARTÍNEZ, 2012)

Jimmy Martínez, afirma que el Gobierno de Datos es “una disciplina encargada de la orquestación de gente, procesos y tecnología que permite habilitar a una compañía a apalancar la información como un recurso de valor empresarial, y al mismo tiempo, es la encargada de mantener a los usuarios, auditores y reguladores satisfechos, usando la mejora de la calidad de los datos para retener clientes, constituyendo y guiando a nuevas oportunidades en el mercado.” (MARTÍNEZ, 2012)

El arquitecto de software de IBM indica que existen seis pasos para el gobierno de datos:

1. **Establecer metas.** Sentencias principales que guían la operación y desarrollo de la cadena de suministro de información.
2. **Definir métricas.** Conjunto de medidas usadas para evaluar la efectividad del programa y los procesos de gobierno asociados.
3. **Tomar decisiones.** La estructura organizacional y el modelo de cambio ideológico para analizar y crear políticas de decisión.
4. **Comunicar políticas.** Herramientas, habilidades y técnicas usadas para comunicar decisiones políticas a la organización.
5. **Medir resultados.** Comparar resultados de las políticas con las metas, entradas, modelos de decisión y comunicación para proveer constante retroalimentación sobre la efectividad de la política.
6. **Auditar.** Herramienta usada para comprobar todo. (MARTÍNEZ, 2012)

⁵ “El gobierno de datos consiste en la capacidad de una organización para gestionar el conocimiento que tiene sobre su información.” (Power Data, 2015)

Las organizaciones requieren de una gobernanza de datos para administrar los datos con la finalidad de garantizar el uso de datos confiables y seguros que cumplan con las reglas y normativas necesarias. Una óptima gestión de los datos favorece el crecimiento del negocio puesto que los datos analizados arrojan información certera.

1.4.1.2 MapReduce

La tecnología *MapReduce* es un modelo de programación utilizado por Google desde el 2004 para simplificar el análisis y proceso de grandes volúmenes de datos. *MapReduce* es núcleo con *Hadoop* que es un sistema de código abierto para almacenar, procesar y analizar grandes volúmenes de datos. Como lo afirma Joyanes Aguilar, “el término *MapReduce* en realidad se refiere a dos procesos que *Hadoop* ejecuta. El primer proceso *Map*, el cual toma un conjunto de datos y lo convierte en otro, donde los elementos individuales son separados en tupias (pares de clave/valor). El proceso *Reduce* obtiene la salida de *map* como datos de entrada, y combina las tupias en un conjunto más pequeño.” (JOYANES AGUILAR, 2013)

La herramienta *MapReduce* ayuda a las organizaciones a una toma de decisiones rápida y eficiente dado que maneja fuentes de datos no estructuradas y semiestructuradas que no se pueden analizar con herramientas tradicionales, lo que hace de este paradigma de programación una ventaja competitiva.

1.4.2 Ciencia de datos

La ciencia de datos (*Data Science*) se refiere a una emergente área de trabajo relacionada con la recopilación, elaboración, análisis, visualización, gestión y conservación de grandes colecciones de información. Aunque el nombre de Ciencia de Datos parece conectar más estrechamente con áreas tales como

bases de datos y la informática, muchos tipos diferentes de habilidades – incluyendo habilidades no-matemáticas – son necesarias. (STANTON, 2012)

En los años 1980, según el experto científico de datos, Dr. Gabriel Guerrero, se desarrollan en primer lugar aplicaciones matemáticas para el análisis de datos (data mining). En segundo lugar, en la mitad de la década de 1990, surgen aplicaciones basadas en Internet que generan una amplia velocidad de datos. Y en tercer lugar, el adelanto de las técnicas computacionales, con el manejo de equipos de cómputo al mismo tiempo (cómputo distribuido) completa los tres elementos que son el punto de partida de la ciencia de datos.

Peter Naur, científico danés pionero en informática ofrece la siguiente definición de ciencia de datos: “se refiere a la ciencia del tratamiento de datos, una vez que se han generado mientras que la relación de los datos a lo que representan se deja a otros campos y ciencias.” (Communications of the ACM, 2016) En este aspecto, el experto con conocimientos analíticos suficientes para utilizar y manejar Big Data se le denomina científico de datos.

Joyanes Aguilar, en su libro titulado *Big Data, análisis de grandes volúmenes de datos en organizaciones*, afirma que “en la actualidad, al analista de datos tradicional se ha añadido una nueva profesión de científico de datos, que supone un grado superior al analista de datos, aunque hay escuelas que los engloban en una misma categoría, pero actualizado al tratamiento de grandes datos; y el científico de datos suma a las tareas típicas del analista de grandes datos unas nuevas responsabilidades como el analista algorítmico aplicado a técnicas estadísticas que aumenta la eficiencia en la toma de decisiones.

Los científicos de datos, tienden a utilizar diferentes conjuntos de herramientas y lenguajes de programación que facilitan manejar herramientas como *MapReduce* y *Hadoop*. Normalmente, la diferencia más clara entre un analista de datos tradicional y un científico de datos es la formación y el estudio. Naturalmente, si categorizamos las de dos profesiones, el analista de datos, tal vez, sea más profesional y el científico de datos se centrará más en la toma de decisiones con el

uso de lenguajes de programación como R, CSQL, lenguajes algorítmicos y las nuevas tecnologías y herramientas centradas en Big Data.” (JOYANES AGUILAR, 2013)

Los científicos de datos son una pieza clave en las organizaciones dada su escasez. “Según indica un estudio de McKinsey Global Institute, sólo en EE.UU. se necesitarán entre 140.000 y 190.000 expertos en analítica de datos en 2018, además de 1,5 millones de gestores capacitados para aplicar Big Data en la toma de decisiones. Por todo ello, la demanda de talento analítico será entre un 50 y un 60% más elevada que la oferta.” (MANYIKA, et al., 2011) En definitiva, los científicos de datos tienen un gran futuro en México por lo que múltiples universidades públicas y privadas han incorporado a su plan de estudios esta carrera y maestría.

1.4.3 Datos abiertos como política pública

La iniciativa de datos abiertos⁶ es una política decretada el pasado 20 de febrero de 2015 en el Diario Oficial de la Federación donde se establece que “Ejecutivo Federal tendrá a su cargo la política de inclusión digital universal, en la que se preverán los objetivos y metas en materia de infraestructura, accesibilidad y conectividad, tecnologías de la información y comunicación, y habilidades digitales, así como los programas de gobierno digital, gobierno y de datos abiertos” (Diario Oficial de la Federación, 2015) Asimismo, en el mismo texto, se afirma que “los datos abiertos del gobierno federal constituyen un activo usable y reutilizable por cualquier sector de la sociedad, que contribuye a impulsar el crecimiento económico, fortalecer la competitividad y promover la innovación, incrementar la transparencia y rendición de cuentas, fomentar la participación ciudadana, así como detonar una mayor eficiencia gubernamental y mejora de los servicios públicos, en apoyo a los objetivos de

⁶ “Los datos abiertos son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen.” (Open Knowledge International, 2011)

desarrollo, de buen gobierno y de generación de conocimiento” (Diario Oficial de la Federación, 2015)

Cabe señalar que los datos abiertos son datos que se caracterizan por ser:

- Accesibles para todos
- En línea
- Gratuitos
- Formatos abiertos y legibles por máquinas
- Licencia Abierta
- Primarios y oportunos

En este marco, por una parte podemos definir los datos abiertos como “datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen.” (Open Knowledge International, 2011)

Dado el impacto de datos abiertos a nivel mundial, el gobierno mexicano ha decidido impulsar esta iniciativa de datos abiertos para favorecer el crecimiento económico, fomentar la innovación pública así como la eficiencia gubernamental, el impacto social y la creación y mejora de servicios públicos basándose en los principios de buena gobernanza que son: la transparencia, la participación y la colaboración.

Por otra parte, el Gobierno de la República informó en 2015 mediante un boletín de la Dirección General de Información en Salud que “la iniciativa de Datos Abiertos del Gobierno de la República busca convertir un recurso público en un activo nacional para detonar un ecosistema de innovación pública.” (Dirección General de Información en Salud, 2015) Para ello, actualmente existe un sitio web de datos.gov.mx en beta.

Este sitio contiene historias que muestran el valor y uso de los datos abiertos, cuenta con más de 100 conjuntos de datos provenientes de 10 dependencias y entidades que

participaron en la primera generación del programa piloto Escuadrón de Datos enfocado en acelerar la implementación de esta iniciativa (SCT, SSA, SAGARPA, SEDESOL, SEP, PEMEX, CONAGUA, CONAPO, NAFIN y CONEVAL). Los conjuntos de datos contienen los elementos técnicos y legales necesarios para facilitar el acceso, uso, reuso y redistribución de los datos abiertos provenientes del gobierno. De manera incremental, los ciudadanos podrán encontrar cada vez más datos públicos provenientes de las dependencias y entidades de gobierno en este catálogo único de datos abiertos.

En datos.gob.mx, se pueden descargar y utilizar libremente los datos abiertos que el Gobierno de la República genera y recolecta, con el fin de innovar, emprender, incrementar la transparencia, eficientar al gobierno y promover la innovación cívica, entre otros usos que pueden ser generados y aprovechados por la sociedad.

También, en datos.gob.mx se encuentra información que utilizan los datos como insumos para crear nuevo conocimiento. Por ejemplo, a través del portal se informa que la Secretaría de Salud tiene 809 unidades de hospitalización y 30,565 médicos generales en el país; que el eje carretero más largo en México tiene 459 km; que la Secretaría de Educación Pública tiene 259,337 centros de trabajo en el territorio nacional; y que los recursos del Ramo 23 ascienden a 55,698 millones de pesos. Estos datos se pueden usar para mejorar el acceso a los servicios de salud, para entender cómo se gastan los recursos de la federación y saber los niveles de satisfacción de la calidad de estos servicios, o conocer esfuerzos innovadores que utilizan datos por parte de ciudadanos, empresas, organizaciones y entes de gobierno.

Un gobierno abierto es aquel que colabora con la sociedad y pone a su disposición las mejores herramientas para la toma de decisiones públicas.” (Dirección General de Información en Salud, 2015)

1.5 Big Data en la actualidad

Actualmente, el número de fotografías publicadas en Facebook es superior a la cantidad de píxeles que Kodak ha tratado en toda su historia. Diariamente se producen más videos que los que se produjeron durante los primeros cincuenta años de la televisión. Hoy en día vivimos en un fenómeno nunca antes visto donde la cantidad, velocidad y variedad de datos aumenta de manera colosal.

El periodo histórico que estamos viviendo denominado, la Era de la información, se relaciona con las tecnologías de la información; con la forma de comunicarse y las transformaciones de la sociedad. Según el informe anual: *Las Oportunidades del Universo Digital: los 'Rich Data'* y el creciente valor del Internet de las Cosas, realizado por la empresa *EMC Computer Systems Mexico*, expertos en temas relacionados con el universo digital y el principal proveedor mundial de inteligencia de mercado de tecnología de información, *International Data Corporation* conocido por sus siglas como IDC, las empresas están incrementando el volumen mundial de datos, los cuales se multiplican cada dos años desde 2013. (EMC Computer Systems México, 2014)

Así, de manera permanente el universo digital pasará de 4,4 millones de millones de gigabytes en 2013 hasta 44 millones de millones de gigabytes en 2020. Este informe confirma que en 2013, únicamente el 22% de la información era considerada útil, pero menos del 5% de los datos se analizaron realmente. Mientras que para 2020 se pronostica que más del 35% de los datos a nivel mundial podrían considerarse relevantes. (EMC Computer Systems México, 2014)

Este desarrollo de la tecnología genera una gran diversidad de fuentes de información y un enorme volumen de datos. El exceso de información y el vínculo entre datos, ha provocado la explosión de interpretaciones y predicciones de comportamiento no solo en las empresas, sino también en las instituciones gubernamentales y en el mundo de la sociedad civil. De esta manera, se ha

revolucionado no solo la forma de comunicarse entre las personas, sino ha cambiado sustancialmente, la toma de decisiones.

Por una parte, el creciente uso de teléfonos inteligentes y aparatos electrónicos conectados a internet ha creado un nuevo concepto llamado Internet de las cosas, IoT por sus siglas en inglés *Internet of Things*. El concepto de IoT fue creado por Kevin Ashton en el Auto-ID Center el Instituto Tecnológico de Massachusetts (MIT, por sus siglas en inglés) donde gracias al sistema RFID (siglas de *radio frequency identification*, es decir, “identificación por radiofrecuencia”), basta con integrar un chip de pocos milímetros en cualquier objeto del hogar, del trabajo o de la ciudad para poder procesar y transmitir información a partir de él constantemente. (SANZ, 2014)

El Internet de las cosas ha revolucionado el uso de los grandes volúmenes de datos generados por los aparatos con conexión a Internet, trayendo beneficios tanto económicos como ecológicos. Como lo afirma Peter Diamandis en su libro *Abundancia: el futuro es mejor de lo que piensas* “por muy potente que llegue a ser, el impacto que el internet de las cosas tenga en nuestras vidas, es eclipsado por su potencial como herramienta de negocio. Pronto las empresas serán capaces de casar perfectamente la demanda de productos con los pedidos de materia prima, haciendo más eficientes las cadenas de suministro y minimizando el gasto hasta un extremo increíble. Su eficiencia aumentará drásticamente. Solo por el potencial de ahorro de energía que supondría que los aparatos fundamentales solo se activarán cuando fueran necesarios (luces que se encienden cuando alguien se acerca a un edificio), merecería la pena el cambio. Y sería un ahorro a escala mundial. Hace años, Cisco se asoció con la NASA para colocar sensores en todo el planeta que proporcionaran información el tiempo real sobre los cambios del tiempo.” (DIAMANDIS & KOTLER, 2013)

Por otra parte, cabe señalar que, el surgimiento de aplicaciones del uso masivo de Internet como: los buscadores *Google* y *Yahoo*, las aplicaciones de correo *Gmail*,

Yahoo, Hotmail y las redes sociales como *Twitter, Facebook, Instagram* y *Netflix*, generan un nuevo concepto de negocio, donde el elemento estratégico es la venta de información.

Actualmente, casi todos los sectores económicos generan y tienen acceso a más datos de lo que se hubiera imaginado hace menos de una década. Las organizaciones requieren extraer valor de dichos datos, saber utilizarlos y poder realizar una toma de decisiones en menor tiempo y con mayor certidumbre.

En este marco, la ciencia de datos, la gobernanza de datos y *Big Data Analytics* ofrecen instrumentos necesarios para soportar este nuevo enfoque de negocio. Estos conceptos que serán definidos más adelante, se encuentran en el centro de desarrollo tecnológico y confrontan la problemática central de esta investigación que se refiere a la administración de grandes volúmenes de datos.

Cabe señalar que la explosión en el crecimiento de los datos y la rapidez con que se generan ha provocado un vacío en la forma de tomar decisiones. Las tecnologías de información han superado la posibilidad humana de realizar estadísticas y generar análisis predictivos por lo que se requiere de una herramienta que pueda dar soluciones instantáneas a la gran masa de datos de diferentes formatos para la toma de decisiones.

Es en este marco que, surge lo que se conoce como Big Data o grandes volúmenes de datos o datos masivos que áreas como la matemáticas, la econometría, la estadística, la administración, entre otras han creado modelos para interpretar los datos de tal forma que se puedan realizar análisis predictivos mejorando los procesos administrativos, operativos y productivos.

De esta manera, la palabra de moda Big Data está tomando protagonismo en el mundo de la analítica digital. La explosión de los datos surge del progreso tecnológico como las infraestructuras de intercambio de datos, aunado con la

evolución social hacia una creciente información compartida como el código abierto u *opendata*.

Calificar el valor de negocio es particularmente importante cuando la organización se dedica a vender un bien intangible por lo que comunicar de forma efectiva el valor de la institución mediante herramientas de *Big Data Analytics* se convierte inmediatamente en un diferenciador competitivo para la organización.

David Loshin, en su libro titulado *Big Data Analytics: from strategic planning to Enterprise integration with Tools, techniques, NoSQL, and Graph*, afirma que “las empresas que desarrollan perfiles de clientes basados en un muestreo limitado de datos se enfrentan a una situación de desventaja en comparación con las empresas que crean modelos integrales de clientes que abarcan todos los datos sobre el cliente para aumentar las ganancias, al mismo tiempo que mejora la experiencia del cliente.” (LOSHIN, 2013)

Big Data Analytics o Analíticos de Grandes Volúmenes de Datos es el proceso de examinar grandes conjuntos de datos con una gran diversidad de tipos de datos para descubrir patrones o estructuras ocultas o intrínsecas (*hidden patterns*) así como para encontrar correlaciones desconocidas entre los diferentes datos, tendencias de mercado, preferencias del cliente y en general, predecir fenómenos basados en los datos y/o analizar y descubrir hechos y estructuras de un conjunto de volúmenes de datos. (SAXSA, 2013)

En este ámbito, las organizaciones han instrumentado los Analíticos de Grandes Volúmenes de Datos para responder a tres principales objetivos: mejorar la experiencia del cliente, optimizar sus procesos, su desempeño operacional y reforzar o diversificar su modelo de negocio. No cabe duda que la implementación Big Data en las organizaciones ha marcado un cambio sustancial en la forma de entender los negocios.

1.6 Impacto de Big Data en la toma de decisiones en las organizaciones

Las organizaciones cuentan con tres sistemas que las caracterizan: el sistema de producción, el sistema informático y el sistema de apoyo a la toma de decisiones. “Según Ludwing Von Bertalanffy, considerado el padre de la teoría de sistemas: Un sistema es un complejo de elementos en interacción que tienen un fin común. Según Felipe Lara, uno de los investigadores que cultiva este campo en México: Un sistema es un conjunto de elementos que interactúan; en donde el comportamiento de uno de ellos afecta el comportamiento de la totalidad; y la forma como afecta el comportamiento depende de los demás elementos. Un sistema se caracteriza por ser holístico, trasdisciplinario y dinámico” (GARCÍA ACOSTA, 2002)

El sistema de producción muestra la forma en que se lleva a cabo la entrada de las materias primas así como el proceso dentro de la empresa para transformar los materiales y así obtener un producto terminado. (phpBB, 2008)

Un sistema informático se define como “un conjunto de partes que funcionan relacionándose entre sí para conseguir un objetivo preciso.” (GALLEGO, 2010) A lo largo de esta investigación se subraya la importancia de procesar la información con un objetivo en común ya que es clave para la toma de decisiones.

El sistema de apoyo a la toma de decisiones o DDS por sus siglas en inglés *Decision Support System*, “son sistemas de apoyo enfocados en el análisis de los datos de la empresa y sirven para ayudar a los directivos a tomar decisiones que a menudo son cambiantes y se deben tomar rápidamente.” (PerTutatis, 2015) En este marco, los sistemas de apoyo a la toma de decisiones permiten obtener información y aplicar herramientas de tratamiento de datos para una adecuada toma de decisiones en los niveles más altos de la organización.

Cada uno de estos sistemas reflejan datos que requieren ser analizados para la toma de decisiones. Por lo que la inversión tecnológica es un factor estratégico para alcanzar los objetivos empresariales. Efectivamente, el uso de Big Data fomenta la transformación de los sistemas y las infraestructuras de las tecnologías de información, así como favorece la mejora en los procesos de negocio, los productos y las soluciones innovadoras. Big Data es un elemento clave para la mejora del proceso de toma de decisiones y para definir la empresa con una ventaja competitiva frente a la competencia.

La analítica de grandes volúmenes de datos es de gran importancia para las organizaciones que manejan volúmenes importantes de datos que no pueden ser analizados con las herramientas tradicionales. *Big Data Analytics*, se implementa hoy en día en las empresas para:

- Mejorar la operación
- Incrementar ganancias
- Mejorar la venta a los clientes actuales, es decir vender más a los clientes actuales
- Buscar nuevos clientes
- Ampliar la cartera de productos

Ciertamente, son numerosas las empresas que hacen uso de herramientas Big Data para describir los acontecimientos en la organización como puede ser conocer el “¿Cuántos?” mediante la analítica descriptiva es decir: ¿Cuántos clientes hombres tiene la empresa?, ¿Cuántas mujeres?, ¿Qué rango de edad?.

Las organizaciones también implementan Big Data para poder hacer proyecciones a futuro basados en el histórico de los datos. Es decir, por ejemplo, si la tendencia de compra de las mujeres es de 2 vestidos al año, entonces para 2025, el nivel de ventas será de x %.

Ahora bien, la joya de la corona para las empresas al utilizar los analíticos prescriptivos⁷ de grandes volúmenes de datos. Es decir, es el descubrimiento de patrones ocultos, correlaciones desconocidas, entre otras informaciones útiles para la toma de decisiones.

Así, a lo largo de los años, ha existido una importante evolución del acceso a la información. Hace sesenta años, las computadoras digitales hicieron legible la información, hace veinte años, Internet la hizo alcanzable y hace diez años los primeros motores de búsqueda unificaron la información en una única base de datos.

A través de potentes herramientas analíticas se puede identificar si un proceso funciona, sin preguntarse el por qué funciona, como lo ha demostrado Google. Éste último ha conquistado el mundo publicitario empleando sólo matemática aplicada. No ha pretendido saber nada acerca de la cultura y las convenciones de la publicidad; simplemente ha apostado por los mejores datos y las mejores herramientas analíticas. Y ha acertado en su decisión. (ANDERSON, 2008). Más allá de las teorías psicológicas o sociales, el universo digital permite mediante algoritmos matemáticos representativos de Big Data, conocer qué hacen las personas y cómo lo hacen permitiendo rastrear y medir comportamientos con fidelidad sin precedentes. Son los datos lo que hablan por sí mismos.

La analítica avanzada en sinergia con Big Data, provoca una ola con grandes transformaciones en todo tipo de modelo de negocio, dado que permiten a las organizaciones considerar prioritariamente los materiales significativos y conocimientos aplicables que están sumergidos en los datos y que no han sido tomados en cuenta. La clave del rendimiento empresarial está en sacarle provecho a la analítica predictiva para invertir el proceso analítico tradicional y

⁷ “Mediante la prescripción, el propio sistema de análisis realiza recomendaciones sobre las acciones que se han de seguir para reducir costes o mejorar los beneficios, consiguiendo en todos los casos un incremento del ROI.” (Instituto de Ingeniería del Conocimiento, 2016)

reflejar conocimientos relevantes de las organizaciones. Este procedimiento se le denomina proceso analítico basado en conocimientos.

El proceso analítico basado en conocimientos significa, según Bill Schmarzo, que el proceso analítico del usuario puede partir de estos conocimientos enterrados, acompañados de recomendaciones específicas para mejorar el rendimiento empresarial. (SCHMARZO, 2013)

En este marco, la toma de decisiones gerenciales es un concepto administrativo que se define como “las decisiones que toman los órganos directivos de las organizaciones. Con carácter general, las decisiones gerenciales se orientan a las relaciones entre la organización y el entorno.” (CABAÑERO PISA, RODRÍGUEZ ARDURA, & SARRADELL LÓPEZ, 2002)

Así como en los años 1980, los escáneres de punto de venta cambiaron la visión sobre la información de fidelidad de los consumidores, en la década de los años 90, el comercio online revolucionó la manera de percibir los gustos y necesidades de los clientes a través de los registros de los Webs que reflejan los comportamientos de los consumidores. Hoy, estamos frente a una transformación del proceso de generación de valor de una organización dada la gran cantidad de datos que se generan en medios sociales, móviles, sensores o dispositivos.

Los medios sociales, ofrecen información creada por los usuarios sobre los intereses, pasiones, necesidades y asociaciones que son una fuente de valor significativa para las organizaciones. Asimismo, los datos generados por móviles sensores o dispositivos conectados a Internet ofrecen una visión detallada en tiempo real sobre los usuarios que permite hacer recomendaciones sobre bienes o servicios, tener un compromiso y una interacción con usuarios. Lo importante está en la velocidad de actuación y en la rapidez con la que se procesan los datos, se cruzan, se ordenan y se actúa tomando una decisión.

Según Joyanes Aguilar, Big Data permite a las organizaciones:

- ✓ Identificar y explotar oportunidades de negocio fugaces en tiempo real.
 - ✓ Actuar de forma veloz frente a cambios de estrategias
 - ✓ Tomar decisiones de rápidamente para interactuar con el cliente
 - ✓ Ofrecer el lugar adecuado y la forma más eficaz para comunicarse con el cliente
 - ✓ Desarrollar recomendaciones predictivas que determinen las decisiones de negocio
 - ✓ Revelar nuevos conocimientos sobre los clientes, los productos, las operaciones y los mercados.
 - ✓ Acelerar la velocidad de actuación frente al cliente, generando nuevas y únicas experiencias, personalizadas e interesantes en tiempo real.
- (JOYANES AGUILAR, 2013)

El uso de Big Data impacta en toda la organización ya que su buen uso ayuda a descubrir nuevas posibilidades de monetización. En efecto, con las herramientas Big Data se pueden identificar nuevos elementos clave de la empresa con una rapidez que permita tomar las medidas necesarias para la creación de valor con tal de obtener conocimientos empresariales en tiempo real.

El análisis de los grandes volúmenes de datos ofrece un gran potencial para la empresa porque supone analizar todos los datos generados en tiempo real con un nivel de detalle nunca antes visto. Transformar los datos en información es lo que genera un enriquecimiento en todos los aspectos para las organizaciones. Big Data permite descubrir patrones de comportamiento sobre los clientes, productos, campañas y operaciones de las empresas para la toma de decisiones rápida y efectiva y en caso de fallo o equivocación enmendar los errores velozmente.

Los Analíticos de Grandes Volúmenes de Datos permiten descubrir patrones ocultos, correlaciones desconocidas, tendencias del mercado, preferencias de los

clientes y posibles riesgos para optimizar el entorno empresarial de manera predictiva, alimentándose de datos obtenidos en tiempo real.

Uno de los grandes beneficios de Big Data es que permiten tomar decisiones personalizadas cuyos datos se obtienen de los patrones de comportamiento de los clientes, productos y operaciones realizadas dentro y fuera de la organización y que permiten optimizar las principales iniciativas empresariales y descubrir nuevas posibilidades de monetización. (SCHMARZO, 2013)

Los competidores innovan constantemente por lo que se necesita estar a la vanguardia y ofrecer a los clientes el producto adecuado en el momento apropiado con tal de identificar las oportunidades de negocio perdidas que no han sido tomadas en cuenta. Dichas oportunidades se pueden identificar al establecer el proceso administrativo⁸.

Según Alfredo Cipriano Luna, autor del libro titulado *Proceso administrativo*, la administración consiste en “la aplicación del Proceso Administrativo con eficacia y eficiencia, integrado por la planeación, organización, integración de recursos, dirección y control”. (LUNA GONZÁLEZ, 2015) Asimismo, Cipriano Luna subraya la importancia de todas y cada una de las fases ya que “deben funcionar en forma sinérgica para lograr resultados efectivos”. (LUNA GONZÁLEZ, 2015) Ahora bien, el proceso administrativo, según Alfredo Cipriano se clasifica de dos etapas, la etapa mecánica que incluye la planeación y la organización y la etapa dinámica que se integra de la dirección y el control.

La etapa de planeación es el inicio y la base del proceso administrativo. Esta base son los cimientos de lo que se realice en las siguientes etapas de organización, dirección y control puesto que sin la fase de planeación éstas últimas no tienen razón de ser. La planeación determina lo que la organización desea lograr, es

⁸ “Un proceso administrativo es una serie o una secuencia de actos regidos por un conjunto de reglas, políticas y/o actividades establecidas en una empresa u organización, con la finalidad de potenciar la eficiencia, consistencia y contabilidad de sus recursos humanos, técnicos y materiales.” (Significados, 2013)

decir el qué quiere la empresa, dónde se va a establecer, cuándo va a iniciar operaciones y cómo lo va a lograr. Según los autores Münch y García, la planeación es “la determinación de los objetivos y elección de los cursos de acción para lograrlos, con base en la investigación y elaboración de un esquema detallado que habrá de realizarse en el futuro”. (MÜNCH & MARTÍNEZ, 2012)

Según Luna González indica en su libro titulado “*Proceso Administrativo*” los elementos de la planeación que son la investigación, la misión, la visión, los valores, los objetivos, los pronósticos, las estrategias, las políticas, los procedimientos, los programas y los presupuestos.

La investigación es de suma trascendencia para la planeación puesto que se refiere al conocimiento del ambiente interno y externo de la organización. En cuanto al ambiente interno se considera al capital humano, los recursos financieros, los recursos materiales y los recursos tecnológicos. En cuanto a los factores externos, éstos están relacionados con la competencia, los clientes, el gobierno, los proveedores, la tecnología y el marco jurídico. (LUNA GONZÁLEZ, 2015)

El autor define la misión como la razón de ser de la organización, “un breve enunciado que sintetiza los principios, propósitos, estrategias y valores esenciales que deberían ser conocidos, comprendidos por todas las personas que colaboran en el desarrollo del negocio” (LUNA GONZÁLEZ, 2015)

En cuanto a la visión, González lo define como el estado futuro deseado, es decir como la aspiración de la organización. Los valores es “una convicción sólida sobre lo que es apropiado y lo que no lo es, que guía las acciones y conducta de los empleados en el cumplimiento de los propósitos de la organización”. (LUNA GONZÁLEZ, 2015) Por último, Luna González esclarece que los objetivos son los resultados que pretende alcanzar la organización en vías de dar cumplimiento a su

misión, visión y valores. Se relaciona con los ¿qué?, ¿qué queremos?, ¿qué deseamos? Y ¿qué vamos a hacer?

Los pronósticos, son también un elemento primordial de la planeación dado que “son los que deciden el futuro con base en determinados supuestos” (LUNA GONZÁLEZ, 2015) llamados expectativas relacionadas con el entorno económico y social que rodea a la organización. En cuanto a la estrategias, son las alternativas y la asignación de recursos para su ejecución.

Otro elemento de la planeación son las políticas que según el autor del libro “*Proceso Administrativo*” son “los conceptos que guían y orientan el pensamiento y acción de los ejecutivos en toma de decisiones para el logro de los objetivos”. (LUNA GONZÁLEZ, 2015) Ahora bien, los procedimientos indican de forma cronológica la manera de cómo debe realizarse el trabajo de acuerdo a las políticas establecidas.

Los programas son los planes que se rigen bajo un calendario que requieren lograrse en tiempo y forma con los recursos necesario para lograrlos exitosamente. Por último, los presupuestos, “ es la expresión de resultados esperados en términos numéricos” (LUNA GONZÁLEZ, 2015)

El proceso de organización reúne los recursos físicos y humanos para el cumplimiento de los objetivos establecidos en la etapa de planeación. La organización consiste, según los autores Koontz y Weihrich, en “la identificación y clasificación de las actividades requeridas, la agrupación de las actividades necesarias para el cumplimiento de los objetivos, la asignación de cada grupo de actividades a un administrador dotado de capacidad necesaria para supervisarlos y la estipulación de coordinación horizontal y vertical en la estructura organizacional”. (KOONTZ & WEIHRICH, 2004)

Una vez concluida la fase mecánica que integra el proceso de planificación y organización, el proceso administrativo se sumerge en una fase dinámica. Es decir que se pasa de una etapa de pensar a la etapa de actuar. La fase de actuar o dinámica está compuesta del proceso de dirección y control. La etapa de dirección, se le conoce también como el proceso de guiar y ejecutar para lograr la implementación. Como lo afirma Cuauhtémoc Anda, “un trabajo continuo de la dirección, es evaluar la efectividad de sus estrategias para tratar de obtener mejores resultados, revisar la estructura del mercado dentro del cual se opera, para determinar los cambios que deban introducirse, con el fin de preservar la misión o función básica del negocio”. (ANDA GUTIÉRREZ, 2004)

La cuarta y última etapa del proceso administrativo consiste en encontrar posibles fallas y supervisar que se cumplan con los objetivos. Como lo indica Anda Gutiérrez, el control “requiere de una serie de actividades que tienen por objetivo comparar un estado actual, como consecuencia de la acción, con otro estado planificado. Visto así, el control permite determinar las diferencias existentes entre estos dos estados. Se trata por lo tanto, de evaluar lo que se ha conseguido respecto a los objetivos planeados.” (ANDA GUTIÉRREZ, 2004) Esta última fase es de suma importancia puesto que una vez realizado el control apropiado, los altos directivos proceden a la toma de decisiones.

La toma de decisiones implica “identificar y seleccionar un curso de acción para enfrentar un problema específico y obtener ventajas cuando se presenta una oportunidad” (LUNA GONZÁLEZ, 2015) Ahora bien, el arte de tomar decisiones es un aspecto que poseen los directivos para el logro de los objetivos empresariales. En este marco, el proceso de planeación, organización, dirección y control realizado correctamente influirá en la persona tomadora de decisiones a elegir aquella alternativa que llevará al éxito el objetivo empresarial. “El recurso humano que es el administrador debe poseer la aptitud y actitud para que esa responsabilidad la lleve a cabo con eficiencia y eficacia para que se alcancen los objetivos de la empresa” (LUNA GONZÁLEZ, 2015)

Los grandes volúmenes de datos someten a prueba la infraestructura actual de las organizaciones, las cuales deben actuar de forma inmediata sin temor a equivocaciones y, en caso de error, actuar de manera inmediata, por lo que hoy más que nunca se requiere de una arquitectura diseñada para manejar en tiempo real *petabytes* de datos estructurados y no estructurados para la óptima toma de decisiones.

Para esta investigación, se consideran criterios de exclusión para las organizaciones que generan poca información por lo tanto pocos datos. Es decir que una organización pueda implementar un modelo de Big Data requiere tener la cantidad importante de datos que no puedan ser analizados por las herramientas tradicionales de análisis de datos como Excel ya que la toma de decisiones sería muy tardada.

1.7 Desafíos del uso de Big Data en las organizaciones

El uso de Big Data en las organizaciones tiene grandes beneficios para la población, no obstante, uno de los grandes retos se presenta en relación a la protección de datos personales. Entre más detallada sea la información del una persona más preciso será su análisis. Sin embargo, cuanto más minuciosos son los datos, la dificultad de acceder a ellos incrementa. En este ámbito aún hay un largo camino que recorrer para poder utilizar los datos personales con la finalidad de gobernarlos para encontrar soluciones y crear modelos para mejorar la vida de la sociedad.

El valor de la información reside en la interpretación que se le puede dar a los datos, el uso de Big Data tiene la capacidad de realizar predicciones acerca de acciones o eventos futuros como poder descubrir patrones de comportamiento oculto o identificar tendencias.

El tema de la protección de datos personales se remonta al año 1948, cuando la Asamblea General de las Naciones Unidas adopta el documento conocido

como *Declaración Universal de Derechos Humanos*, en este documento se expresan los derechos humanos conocidos como básicos y en especial. El artículo que se relaciona con la protección de datos personales es el artículo 12.

Artículo 12.

Nadie será objeto de injerencias arbitrarias en su vida privada, su familia, su domicilio o su correspondencia, ni de ataques a su honra y su reputación. Toda persona tiene derecho a la protección de la ley contra tales injerencias o ataques. (Organización de las Naciones Unidas, 1948)

En gran medida, el uso de Big Data en diversos sectores es percibido como un desafío para los principios de privacidad. En este aspecto, la utilización de los datos personales ha generado una gran preocupación respecto a la privacidad de los individuos, sobre todo en la actualidad dado que las redes sociales generan cantidades exorbitantes de datos que pueden ser extraídos de los perfiles. En este aspecto, existen actores interesados en la protección de datos y privacidad relacionados con Big Data que buscan salvaguardar la información utilizada por las nuevas tecnologías, como el Grupo de Trabajo Internacional sobre Protección de Datos en Telecomunicaciones, IWGDPT por sus siglas en inglés, o bien Grupo de Berlín.

El Grupo de Berlín, es un grupo de trabajo que “se formó en 1983, en el seno de la Conferencia Internacional de Comisarios de Protección de Datos y Privacidad, debido a la iniciativa del Comisario de Berlín para la Protección de Datos, el cual, desde entonces, preside el Grupo. El Grupo de Berlín ha adoptado numerosas recomendaciones relacionadas con la protección de la privacidad y las telecomunicaciones. Forman parte de este Grupo representantes de las autoridades de protección de datos de numerosos países del mundo. Desde principios de los años 90 el Grupo ha venido teniendo especial atención a la privacidad en Internet.” (ÁLVAREZ HERNANDO, 2011)

En la resolución sobre Big Data de la 36ª Conferencia Internacional de Autoridades de Protección de Datos y Privacidad, se “hace un llamado a todas las partes que utilizan Big Data para:

- Respetar el principio de especificación de finalidad.
- Limitar la cantidad de información recolectada y almacenada a un nivel que sea necesario para el propósito legítimo que pretende.
- Obtener, cuando sea apropiado, el consentimiento válido del titular de los datos en relación con el uso de información personal para fines de análisis y de creación de perfiles.
- Ser transparentes acerca de qué información se recolecta, cómo se procesa, con qué propósito serán utilizados y si será transferida a terceros.
- Dar a las personas acceso apropiado a los datos que han sido recolectados sobre ellas y a la información y decisiones que se han tomado con esos datos. Las personas deben ser avisadas de la fuente de sus datos personales y, cuando sea apropiado, de su derecho a corregir su información, así como de las herramientas para controlar esta información.
- Ofrecer a las personas, cuando sea apropiado, acceso a la información sobre los insumos principales y los criterios para la toma de decisiones (algoritmos) que se han utilizado como base para el desarrollo del perfil. La información debe presentarse en un formato claro y comprensible.
- Llevar a cabo una evaluación de impacto en la privacidad, especialmente cuando el análisis de Big Data implica usos novedosos o inesperados de los datos personales.
- Desarrollar y utilizar tecnologías de Big Data de acuerdo con los principios de la Privacidad por Diseño.
- Considerar cuándo los datos anónimos mejorarán la protección de la privacidad. La anonimización puede ayudar a mitigar los riesgos para la privacidad asociados con el análisis de Big Data, pero sólo si la anonimización está diseñada y gestionada apropiadamente. La solución óptima para anonimizar los

datos debe decidirse caso por caso, posiblemente utilizando una combinación de técnicas.

- Tener mucho cuidado, y actuar cumpliendo la legislación aplicable en materia de protección de datos, cuando se comparten o se publican conjuntos de datos con seudónimos o que pueden ser identificables indirectamente. El acceso debe ser limitado y controlado cuidadosamente si los datos contienen suficientes detalles, esto es, que pueden vincularse con otros conjuntos de datos o contienen datos personales.
- Demostrar que las decisiones respecto al uso de Big Data son justas, transparentes y responsables. Relacionado con el uso de datos para fines de creación de perfiles, tanto éstos como los algoritmos en que están basados requieren una valoración continua. Ésta necesita revisiones regulares para verificar si los resultados de la creación de perfiles son responsables, justos y éticos y si son compatibles y proporcionados con el propósito para el cual los perfiles son usados. Debe evitarse la injusticia con las personas debido a resultados completamente automatizados que arrojen un falso positivo o un falso negativo. Siempre debe estar disponible una valoración manual de resultados, con efectos significativos para los individuos.” (Agencia Andorrana de Protección de Datos, 2014)

Cabe señalar que en México, el 27 de abril de 2010, el Congreso de la Unión aprobó la Ley Federal de Protección de Datos Personales en Posesión de Particulares (LFPDPPP) que entró en vigor el 6 de Julio de 2010 y tiene como objetivo garantizar la privacidad y el derecho a la autodeterminación informativa de las personas, apuntado en el artículo 1.

Artículo 1.- La presente Ley es de orden público y de observancia general en toda la República y tiene por objeto la protección de los datos personales en posesión de los particulares, con la finalidad de regular su tratamiento legítimo, controlado e informado, a efecto de garantizar la privacidad y el derecho a la autodeterminación informativa de las personas. (LFPDPPP, 2010)

En dicha ley se indica que toda persona física o moral debe cumplir con la ley en relación a la protección de datos personales. En el artículo 3º transitorio de la ley se subraya que toda persona física o moral que recabe datos personales debe expedir su aviso de privacidad. “Los responsables designarán a la persona o departamento de datos personales a que se refiere el artículo 30 de la Ley y expedirán sus avisos de privacidad a los titulares de datos personales de conformidad a lo dispuesto por los artículos 16 y 17 a más tardar un año después de la entrada en vigor de la presente Ley.” (LFPDPPP, 2010)

En el Capítulo V de la LFPDPPP se señala la importancia de los Derechos de Acceso, Rectificación, Cancelación y Oposición, conocidos como Derechos ARCO, específicamente en el artículo 28.

Artículo 28.- El titular o su representante legal podrán solicitar al responsable en cualquier momento el acceso, rectificación, cancelación u oposición, respecto de los datos personales que le conciernen. (LFPDPPP, 2010)

Los derechos denominados ARCO, “son un conjunto de derechos que garantizan al ciudadano el poder de control de sus datos personales. Lo más importante en el ejercicio de este derecho es que sólo el titular de los datos personales puedes solicitar el acceso, la rectificación, cancelación u oposición, siempre que se encuentren en un sistema de datos personales. Por lo tanto en este tipo de solicitudes será requisito indispensable que sea acreditada la personalidad del solicitante o bien la del representante legal.” (Secretaría de Finanzas, 2016)

El organismo constitucional autónomo en México encargado de la protección de datos personales es el Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (por sus siglas, INAI) que tiene como misión “garantizar en el Estado mexicano los derechos de las personas a la información pública y a la protección de sus datos personales, así como promover una cultura de

transparencia, rendición de cuentas y debido tratamiento de datos personales para el fortalecimiento de una sociedad incluyente y participativa.” (INAI, 2015)

Los principales objetivos del INAI son:

- **Objetivo 1:** Garantizar el óptimo cumplimiento de los derechos de acceso a la información pública y la protección de datos personales.
- **Objetivo 2:** Promover el pleno ejercicio de los derechos de acceso a la información pública y de protección de datos personales, así como la transparencia y apertura de las instituciones públicas.
- **Objetivo 3:** Coordinar el Sistema Nacional de Transparencia y de Protección de Datos Personales, para que los órganos garantes establezcan, apliquen y evalúen acciones de acceso a la información pública, protección y debido tratamiento de datos personales.
- **Objetivo 4:** Impulsar el desempeño organizacional y promover un modelo institucional de servicio público orientado a resultados con un enfoque de derechos humanos y perspectiva de género. (INAI, 2015)

Con la nueva Ley Federal de Protección de Datos personales, los mexicanos cuentan con una legislación que coloca a las personas en el centro de la tutela del Estado de forma que tienen una legislación que protege la información personal de los ciudadanos que pueda encontrarse en las bases de datos de cualquier organización que recaude este tipo de datos. Asimismo, esta legislación muestra las reglas que deben seguir las empresas derivadas de principios internacionales. Además, la Ley también protege los datos que se consideran sensibles que están relacionados con el origen racial o étnico, estado de salud, información genética, creencias religiosas, filosóficas y morales, afiliación sindical, opiniones políticas y preferencias sexuales.

Otro de los desafíos que tienen las empresas es la falta de personal especializado y capacitado en Big Data. Si bien las empresas conocen los grandes beneficios de Big Data y tienen la intención de aplicarlo a sus procesos de negocio, se ven estancadas por la falta de recursos humanos especializados y conocimientos detallados. Según el estudio *Big Data Survey Europe 2012/2013 BARC*, patrocinado por *Teradata* realizado durante la segunda mitad de 2012 con empresas de los principales mercados de *software* europeos en Alemania, Austria, Suiza, Francia y Gran Bretaña, las empresas son muy conscientes del valor potencial de Big Data, así como de los retos que éstos conllevan.

Como razones principales que impulsan la demanda de soluciones de Big Data, el 72 por ciento de los participantes citan la necesidad de obtener el control sobre el crecimiento exponencial de los volúmenes de datos en sus organizaciones, el 75 por ciento mencionó las nuevas y mejores posibilidades para el análisis de datos que proporcionan dichas tecnologías. Por otro lado, dos tercios (66 por ciento) de los encuestados ya han identificado el valor de los llamados datos poliestructurados procedentes de distintas fuentes, como los social media, los sensores o los *weblogs*. Sólo el siete por ciento no vio una necesidad urgente de actuar en materia de Big data. (Teradata , 2013) A los desafíos antes mencionados, se agrega el principal que es el acceso a los datos.

1.8 Resumen de aportaciones

En este capítulo se analizaron las definiciones del concepto de Big Data y se determinó la definición que se le otorga al término Big Data para esta investigación.

En el siguiente punto, se presenta el origen de Big Data que dan fundamento a la importancia de esta revolución digital del siglo XXI. Se mencionan los pioneros de los grandes volúmenes de datos y la trascendencia que ha tenido Big Data.

Asimismo, después de esclarecer el surgimiento Big Data, se precisaron los conceptos relacionados que muestran un panorama más detallado sobre el tema como la ciencia de datos, la minería de datos y *Big Data Analytics*. Se puntualizó sobre la iniciativa del Gobierno de la República relacionada con los datos abiertos para tener una perspectiva más amplia sobre la importancia del tema a nivel Nacional.

Al ser un concepto que actualmente es muy demandado por las grandes universidades y empresas a nivel mundial, se especifica cual es la posición de Big Data en la actualidad. De manera que se logra puntualizar sobre el impacto de los grandes volúmenes de datos en las organizaciones y la importancia de su buen uso para la creación de valor en las empresas.

Finalmente, se especifican los desafíos del uso de Big Data mostrando los aspectos legales que se han creado a partir de esta revolución de las tecnologías de información.

Una vez esclarecidos los conceptos anteriores, es importante conocer los modelos que existen de Big Data, así como sus beneficios y sus técnicas que son puntos que se describen en el siguiente capítulo.

Capítulo 2: Modelos de Big Data

*“La idea era hacer investigación, buscar nuevos caminos a conquistar,
nuevas montañas que escalar”*

Gertrude Belle Elion

En este segundo capítulo se muestran dos modelos de Big Data para la toma de decisiones. El primer modelo, creado por Bill Schmarzo, muestra mediante un marco de trabajo las estrategias que debe realizar una organización para poder integrar un modelo de Big Data para lograr la estrategia empresarial establecida. El segundo modelo, creado por Gabriel Guerrero, refleja los pasos a seguir para implementar el modelo de Gobierno de Datos, *WebGestiones*, en una organización para el manejo de grandes volúmenes de datos.

Se concluye el capítulo realizando un comparativo de ambos modelos para retomar lo esencial de cada uno para el desarrollo del modelo propuesto.

2.1 Marco de trabajo de Bill Schmarzo

El Diccionario del Español de México del Colegio de México, define un modelo como una “presentación esquemática, abstracta, más pequeña o más simple que se hace de un objeto complejo para facilitar su comprensión: un modelo matemático, un modelo a escala, el modelo del átomo, el modelo del ADN” (Colegio de México, 2016)

En este sentido, un modelo de Big Data se puede definir como un proceso esquemático para identificar oportunidades de negocio y mejorar la toma de decisiones en las organizaciones.

Bill Schmarzo, experto del tema de Big Data y la analítica avanzada⁹, ofrece en su libro titulado *Big Data: el poder de los datos*, un marco de trabajo que contempla las

⁹ “El término inteligencia analítica o analítica avanzada, es una manera exhaustiva de utilizar los datos para definir estrategias y tomar decisiones óptimas.” (TIS Consulting Group, 2011)

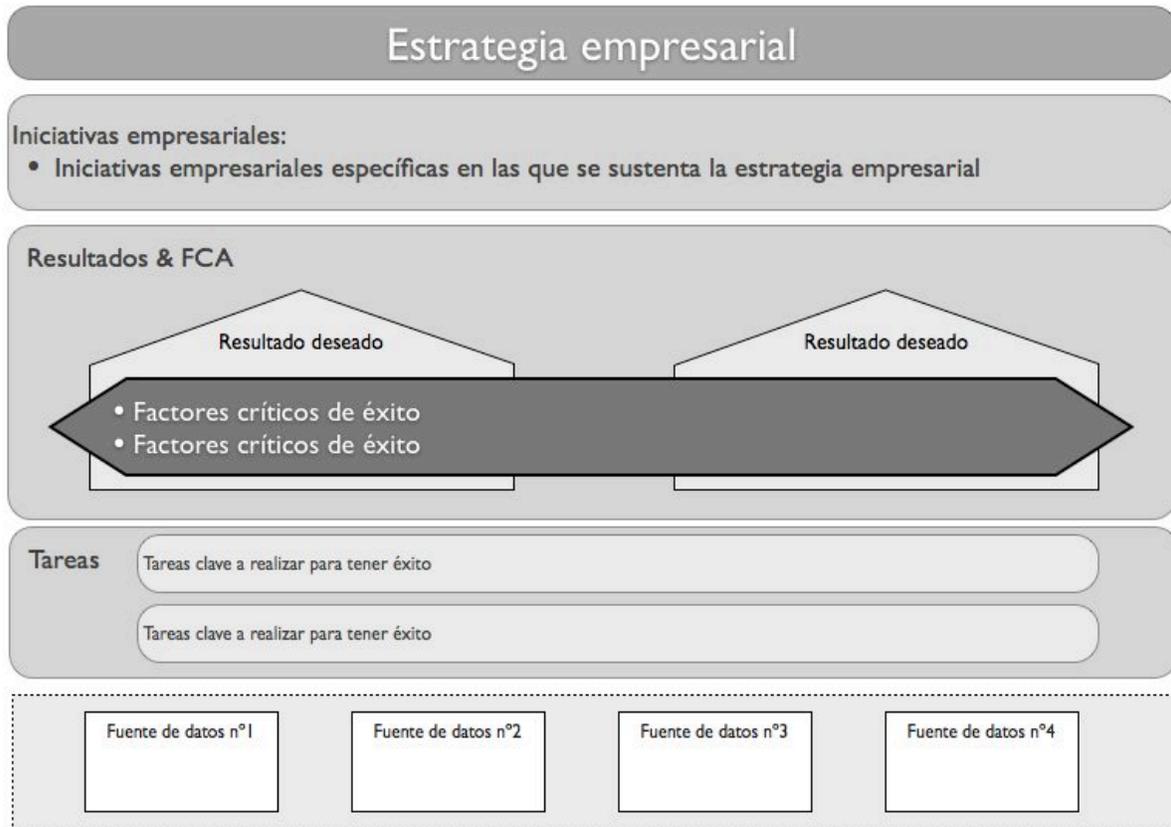
iniciativas clave para que cualquier organización, sin importar su giro, pueda aplicar las técnicas de Big Data. El documento se centra en el proceso de apoyo a la estrategia empresarial global de la organización de manera que se identifiquen las iniciativas empresariales de apoyo antes de pasar a los requisitos tecnológicos, analíticos, de arquitectura y de datos. El documento de estrategia para Big Data realizado por Schmarzo, se integra de cinco secciones:

- **Estrategia empresarial:** La estrategia empresarial elegida define claramente el ámbito en el que se centrará la iniciativa de Big Data. Schmarzo subraya que “el título no debería tener más de una frase, pero aun así debería proporcionar los detalles suficientes para identificar el objetivo empresarial global.” (SCHMARZO, 2013)
- **Iniciativas empresariales:** “Esta sección divide la estrategia empresarial en las iniciativas empresariales en las que se sustenta. Una iniciativa empresarial se define como un proyecto multidisciplinario que dura entre 9 y 12 meses, con objetivos económicos o empresariales claramente definidos en los que se mide el éxito de la iniciativa empresarial. No suele haber más de entre tres y cinco iniciativas empresariales por estrategia empresarial.” (SCHMARZO, 2013)
- **Resultados y factores críticos de éxito (FCE):** “Contiene los resultados y los factores críticos de éxito necesarios para llevar a cabo con éxito las iniciativas empresariales clave de la organización. Los resultados definen el estado final deseado o ideal. Los factores críticos de éxito definen lo que hay que hacer para que la iniciativa empresarial tenga éxito.” (SCHMARZO, 2013)
- **Tareas:** “Proporciona el siguiente nivel de detalle, pues documenta las tareas específicas que hay que realizar a la perfección para sustentar las iniciativas empresariales marcadas. Se trata de las tareas clave en las que

deberán colaborar las distintas partes de la organización para satisfacer las iniciativas empresariales. Esta es la sección de cómo se hace del documento, y es en este nivel de detalle en el que se definen, asignan y miden las tareas personales y los objetivos de la gestión. Lo normal es tener entre 8 y 12 tareas clave definidas y vinculadas a las iniciativas empresariales objetivo como parte del documento de estrategia para Big Data.” (SCHMARZO, 2013)

- **Fuentes de datos:** “Por último, el documento destaca las fuentes de datos clave necesarias para sustentar la estrategia empresarial y las iniciativas empresariales clave de apoyo. La definición de las tareas debería proporcionar un sólido conocimiento de las métricas y medidas clave, las dimensiones empresariales de relevancia, el nivel de granularidad y la frecuencia del acceso a los datos” (SCHMARZO, 2013)

Figura 3 Modelo de estrategia para Big Data por Bill Schmarzo



Fuente: Bill SCHMARZO (2014) *Big Data: El poder de los datos*.

El documento de estrategia para Big Data sirve para garantizar que las iniciativas de Big Data y los objetivos empresariales estén dirigidos a la misma meta. Asimismo, este documento sirve de guía para la toma de decisiones para directivos de las organizaciones y el área de tecnologías de información. Schmarzo afirma que “el documento de estrategia sirve para priorizar los requisitos tecnológicos al poner a prueba su capacidad de servir de apoyo a la estrategia empresarial elegida, las iniciativas empresariales clave, los factores críticos de éxito y las tareas clave.” (SCHMARZO, 2013)

2.1.1 Técnicas de creación de valor de Bill SCHMARZO

Bill Schmarzo, en su libro titulado *Big Data: el poder de los datos*, ofrece distintas técnicas para potenciar Big Data en las principales iniciativas empresariales de

manera que los usuarios profesionales puedan entender que tipo de preguntas se pueden responder y tomar las decisiones indicadas haciendo uso de Big Data con el objetivo de entender dónde y cómo pueden repercutir Big Data en los procesos de creación de valor empresarial.

Las técnicas que presenta Schmarzo permiten visualizar fácilmente las áreas específicas en la que Big data puede tener impacto sobre las empresas por lo que se apoyan en dos premisas básicas:

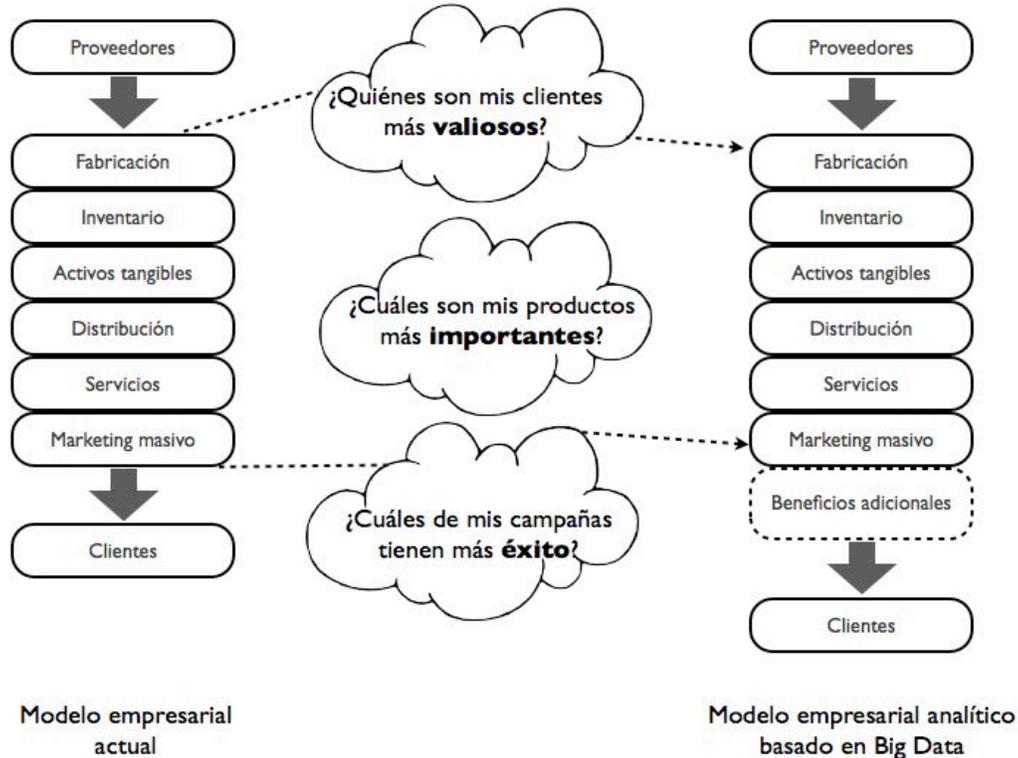
1. Los usuarios profesionales conocen el tipo de preguntas que intentan responder para sustentar sus principales procesos empresariales.
2. Los usuarios profesionales conocen el tipo de decisiones que intentan tomar actualmente para sustentar sus principales procesos empresariales.

Acorde a Schmarzo, “los usuarios profesionales necesitan responder este tipo de preguntas para poder:

- ✓ Descubrir nuevas posibilidades de obtener beneficios que puedan repercutir en su marketing y su organización de ventas.
- ✓ Reducir los costes de adquisición, fabricación, inventario, cadena de suministro, distribución, publicidad, ventas y mantenimiento.
- ✓ Atenuar el riesgo en todos los aspectos operativos y financieros de la cadena de valores de la organización.
- ✓ Conseguir nuevos clientes, productos y conocimientos operativos que puedan utilizar para lograr una ventaja competitiva frente a la competencia y sacar un mayor beneficio en el sector.

Lo que ha cambiado con Big Data es el modo en que podemos aprovechar las nuevas fuentes de datos y las nuevas capacidades analíticas para responder estas preguntas clave, con el fin de descubrir nuevos conocimientos sobre nuestros clientes, productos y mercados.” (SCHMARZO, 2013)

Figura 4 Big Data guía el proceso de creación de valor.



Fuente: Bill SCHMARZO (2014) *Big Data: El poder de los datos*.

Big Data permite “tomar decisiones con un nivel más de detalle para poder descubrir nuevos conocimientos sobre nuestros clientes, productos, operaciones,” (SCHMARZO, 2013) y aplicarlo para responder a las preguntas empresariales clave que se muestra en la figura arriba titulada “Big Data guía el proceso de creación de valor”. En efecto, Big Data perfecciona y acelera la capacidad para identificar áreas específicas del negocio y procesos empresariales específicos en los que puede devolver un valor empresarial inmediato, asegura Bill Schmarzo.

El proceso de creación de valor se puede explicar mediante la técnica de los factores motrices del valor de Big Data de Bill Schmarzo que se explica a continuación.

2.1.2 Factores motrices del valor de Big Data

Los cuatro factores motrices de Big Data permiten entender dónde y cómo puede repercutir Big Data en el proceso de creación de valor de una organización. La clave es conocer lo que Schmarzo denomina los factores motrices de Big Data que se pueden aplicar a las principales iniciativas empresariales o a los procesos empresariales de la organización para proporcionar nuevos conocimientos sobre el negocio y sirven para la toma de decisiones. La tabla a continuación muestra los factores motrices y su impacto en la monetización de los datos¹⁰.

¹⁰ “Por «monetización de datos» se entiende el empleo de los mismos como un activo directo para la generación de ingresos.” (REYES, 2015)

Tabla 3 Los cuatro factores motrices de Big Data

FACTORES MOTRICES DE BIG DATA	IMPACTO EN LA MONETIZACIÓN DE LOS DATOS
Datos estructurados: Datos transaccionales más detallados (por ejemplo: tarjetas de crédito, registros de llamadas)	Hace posible decisiones más detalladas y concretas (locales, estacionales, multidimensionales)
Datos sin estructurar: Datos diversos internos (por ejemplo: e-mails, comentarios de consumidores y externos (por ejemplo: medios sociales, móviles)	Hace posible decisiones no estructuradas más completas y precisas (con nuevas métricas, dimensiones y atributos dimensionales)
Velocidad de los datos: Acceso a datos de baja latencia (en tiempo real)	Hace posible tomar decisiones con más frecuencia y a tiempo (horas frente a semanas, bajo demanda)
Analítica predictiva: Causalidad, predictores, instrumentación, experimentación	Decisiones predictivas más aplicables (optimizar, recomendar, predecir, estimar, prever)

Fuente: *Bill SCHMARZO (2014) Big Data: El poder de los datos.*

- **Factor nº1: Acceso a datos transaccionales más detallados.**

El acceso a datos transaccionales como pueden ser registros de llamadas, identificadores de radio frecuencia, transacciones con tarjetas de crédito, transacciones de acciones, reclamaciones al seguro e opiniones médicas, entre otros, permiten conocer el grado de fidelidad del cliente. Estos datos detallados, granulares y estructurados, son para la mayoría de las organizaciones datos que recopilan pero que no les dan uso.

“Debido a las limitaciones de la tecnología actual y el coste de almacenamiento de datos, la mayoría de usuarios profesionales sólo tienen acceso a una cantidad de datos limitada en los que apoyar sus informes operativos y administrativos. Sin embargo, las tecnologías Big Data poseen la capacidad de acceder y analizar todos los datos transaccionales detallados de los usuarios profesionales, llevándoles a hacerse preguntas del tipo *y si (what if)*”, (SCHMARZO, 2013) afirma Schmarzo.

Efectivamente, con el acceso a los datos transaccionales más detallados, se pueden responder las siguientes preguntas:

- ¿Cuál es el valor empresarial potencial de dar un mayor apoyo a la adaptación local desarrollando predicciones y planes de producto a nivel de categoría, tienda o departamento?
- ¿Cuál es el valor empresarial potencial de dar un mayor apoyo a la toma de decisiones de temporada desarrollando planes de marketing a nivel de segmento del cliente, producto, código postal y temporada (Navidad, San Valentín o Día de las Madres)?
- ¿Cuál es el valor empresarial de priorizar los tipos de reclamación en las normas a nivel de cláusula, reclamación, semana o día?

Por lo tanto, analizar los datos transaccionales, como pueden ser la ubicación de tiendas, los productos, el día de la semana, la hora, el tipo de comportamiento del cliente, la categoría demográfica entre otros, puede mejorar sustancialmente la capacidad de una organización para descubrir posibilidades de negocio tangibles y aplicables.

- **Factor nº2: Acceso a datos sin estructurar**

“La capacidad de integrar el creciente volumen de datos sin estructurar con los datos transaccionales detallados que ya tenemos puede transformar radicalmente

el tipo de conocimientos que se puede extraer de los datos. Los datos sin estructurar pueden proporcionar nuevas métricas y dimensiones que los usuarios profesionales pueden utilizar para descubrir nuevos conocimientos sobre sus clientes, productos, operaciones y mercados”, indica Schmarzo. (SCHMARZO, 2013)

Las métricas extraídas de los datos sin estructurar favorece el análisis y toma de decisiones que pueden responder a preguntas como:

- ¿Cuál es el potencial empresarial de añadir los datos generados por sensores a nuestros modelos de mantenimiento predictivo de fabricación, cadena de suministro y productos?
- ¿Cuál es el potencial empresarial de integrar datos sin estructurar de fuentes de terceros (el tiempo, las noticias económicas o los eventos locales) en las decisiones de un hospital, un equipo médico o los cuidados sanitarios?
- ¿Cuál es el potencial empresarial de integrar datos de los medios sociales en nuestro análisis de fraude en las reclamaciones para identificar las reclamaciones inusuales que pertenezcan a comunidades de individuos conectados entre sí?
- **Factor nº3: Acceso a datos de baja latencia (en tiempo real)**

Actualmente, la gran cantidad de información generada se vuelve obsoleta muy rápidamente por lo que se requiere tener acceso a los datos en tiempo real para poder identificar nuevas posibilidades de monetización de los datos. Esta fugacidad de oportunidades del mercado que aparecen y desaparecen antes de que se tenga la ocasión de identificarlas y actuar, ha impuesto la necesidad de trabajar rápidamente.

En efecto, existe un gran potencial empresarial de tener acceso en tiempo real o con una baja latencia a fuentes de datos y métricas empresariales clave para poder reducir el tiempo entre que se da un evento del cliente o el producto y que esos datos estén disponibles para ser analizados y tomar una decisión. Por ejemplo, hay un gran potencial empresarial de los servicios que emplean la ubicación para comunicarse con sus clientes en tiempo real, mientras realizan sus compras.

Este tercer factor que explica Bill Schmarzo de acceso a datos de baja latencia y en tiempo real permite responder a las siguientes preguntas:

- ¿Cuál es el potencial empresarial de ser capaces de actualizar nuestros modelos analíticos del cliente, el producto, el riesgo y los operativos, a la carta, en base a negocios o mercados inmediatos dependientes de eventos externos (como que se prediga la llegada de un huracán devastador dentro de dos días)?
- ¿Cuál es el potencial empresarial de hacer nuevos clientes, el mantenimiento predictivo o tomar decisiones de optimización de la red en cuestión de minutos, horas o “según haga falta”?
- ¿Cuál es el potencial empresarial de actualizar los modelos analíticos en función de los eventos económicos, del mercado actual o locales (es decir, el tiempo, el tráfico, conciertos o partidos de fútbol)?
- ¿Cuál es el potencial empresarial de actualizar los requisitos del equipo o el inventario de un hospital en función de los problemas de salud locales o los brotes epidémicos?
- **Factor nº4: Integración de la analítica predictiva**

El cuarto y último factor pretende mostrar el potencial de transformar cualquier pregunta que los usuarios profesionales intenten responder o cualquier decisión que intenten tomar considerando verbos como predecir, prever, estimar, recomendar y optimizar. De esta forma, los usuarios profesionales pueden

visualizar un nuevo conjunto de preguntas sobre qué hacer acerca del potencial impacto empresarial de predecir qué podría ocurrir, recomendar una línea de acción específica o predecir el impacto de las distintas decisiones posibles.

Así, la integración de la analítica predictiva o avanzada en los principales procesos empresariales, permite producir respuestas predictivas a preguntas clave como:

- ¿Cuál es el potencial empresarial de aprovechar la analítica predictiva para optimizar la coordinación, los gastos publicitarios y las decisiones de equipo?
- ¿Cuál es el potencial empresarial de aprovechar la analítica predictiva para predecir el impacto financiero de los cambios en los precios, las rutas o los proveedores?
- ¿Cuál es el potencial empresarial de marcar a los socios en función de la fiabilidad de su calidad, distribución y servicio?
- ¿Cuál es el potencial empresarial de aprovechar la analítica predictiva para recomendar las mejores ofertas para mejorar la satisfacción del cliente, la retención o el tratamiento preventivo de pacientes?

A continuación se presenta un cuadro que resume lo citado anteriormente.

Tabla 4 Modelo de Bill SCHMARZO

¿Qué se desea lograr?		¿Cómo se logrará?	
Marco de trabajo de Bill SCHMARZO		Técnicas de creación de valor de Bill SCHMARZO	
Documento de estrategia con Iniciativas clave para aplicar Big Data en las organizaciones		Técnicas para entender dónde y cómo puede repercutir Big Data en la organización	
1	ESTRATEGIA EMPRESARIAL	<p>Detalles del objetivo global</p> <p>Se debe identificar alguna de las 2 premisas:</p> <ul style="list-style-type: none"> - Los usuarios conocen el tipo de preguntas que intentan responder para sustentar los principales procesos empresariales - Los usuarios conocen el tipo de decisiones que intentan tomar en los principales procesos empresariales 	<p>¿Qué objetivo desea cumplir la organización?</p> <ul style="list-style-type: none"> • Descubrir nuevas posibilidades de obtener beneficios que puedan repercutir en su marketing y su organización de ventas. • Reducir los costes de adquisición, fabricación, inventario, cadena de suministro, distribución, publicidad, ventas y mantenimiento. • Atenuar el riesgo en todos los aspectos operativos y financieros de la cadena de valores de la organización. • Conseguir nuevos clientes, productos y conocimientos operativos que puedan utilizar para lograr una ventaja competitiva frente a la competencia y sacar un mayor beneficio en el sector.
2	INICIATIVAS EMPRESARIALES	Proyecto multidisciplinario con objetivos definidos	Se crean en relación al objetivo que dese lograr la organización
3	RESULTADO Y FACTORES CRÍTICOS	Factores críticos de éxito para llevar a cabo las iniciativas empresariales	
4	TAREAS	Tareas específicas para sustentar las iniciativas empresariales	
5	FUENTES DE DATOS	Fuentes de datos necesarias para lograr la estrategia empresarial	<p>Se considera la técnica de los factores motrices del valor de Big data de Bill SCHMARZO</p> <p>Factor 1: Acceso a datos transaccionales detallados</p> <p>Factor 2: Acceso a datos sin estructurar</p> <p>Factor 3: Acceso a datos de baja latencia</p> <p>Factor 4: Integración de la analítica predictiva</p>

Fuente: Elaboración propia

2.2 Modelo de Gobierno de Datos WebGestiones de Dr. Gabriel Guerrero

El Modelo de Gobierno de Datos, *WebGestiones*, es un modelo de instrumentación para el manejo y la gobernanza de grandes volúmenes de datos elaborado por el Dr. Gabriel Guerrero, Director General de saXsa, empresa mexicana de tecnologías de información. El Modelo de Gobierno de Datos retoma los seis puntos de la gobernanza de los datos como eje para definir qué se desea lograr al implementar un modelo de Big Data. Recordemos que los seis pasos para el gobierno de datos son:

1. **Establecer metas.** Sentencias principales que guían la operación y desarrollo de la cadena de suministro de información.
2. **Definir métricas.** Conjunto de medidas usadas para evaluar la efectividad del programa y los procesos de gobierno asociados.
3. **Tomar decisiones.** La estructura organizacional y el modelo de cambio ideológico para analizar y crear políticas de decisión.
4. **Comunicar políticas.** Herramientas, habilidades y técnicas usadas para comunicar decisiones políticas a la organización.
5. **Medir resultados.** Comparar resultados de las políticas con las metas, entradas, modelos de decisión y comunicación para proveer constante retroalimentación sobre la efectividad de la política.
6. **Auditar.** Herramienta usada para comprobar todo. (MARTÍNEZ, 2012)

Para Dr. Guerrero, la gobernanza de los datos es la gestión integral de la disponibilidad, facilidad de uso y seguridad de los datos empleados en una empresa.

El Modelo de Gobierno de Datos WebGestiones refleja los doce pasos que se necesitan realizar para implementar Big Data en una organización con el objeto de mejorar los procesos productivos y por lo tanto lograr una mejora en los objetivos empresariales.

A continuación se muestran los pasos que integran el Modelo de Gobierno de Datos, *WebGestiones*:

1. **Asimilar la misión empresarial:** Se requiere entender la misión de la organización para conocer sus objetivos empresariales, saber cuál es su razón de ser (misión) y discernir lo que desea lograr (visión).
2. **Interpretar cada uno de los objetivos empresariales:** Se necesita conocer cada uno de los objetivos empresariales para lograr la misión.

3. **Determinar procesos productivos:** Para lograr los objetivos empresariales se requiere determinar los procesos productivos, es decir la secuencia de operaciones que genera una empresa para lograr su objetivos.
4. **Conocer los sistemas que integran los procesos:** En este punto, se analizan los sistemas de información que están asociados a cada proceso productivo.
5. **Determinar el conjunto de datos que genera cada sistema.**
6. **Definir resultados que desean medirse:** Se definen los resultados a corto plazo o tangibles que desean medirse para lograr el objetivo empresarial.
7. **Definir métricas:** Se establecen las métricas tangibles a corto plazo en los procesos productivos que desean un análisis descriptivo.
8. **Determinar fuentes relevantes:** Se definen las fuentes de datos relevantes que intervienen en los procesos que permiten el logro de los resultados a corto plazo.
9. **Definir el mecanismo de construcción de la métrica:** Se define el mecanismo a partir de las fuentes de información.
10. **Determinar visualización:** Se establece la herramienta preferida para la visualización de los resultado de las métricas.
11. **Desarrollar informes y tableros de control para cada métrica:** Se desarrolla el aspecto visual de los analíticos con la herramienta preferida para cada métrica a corto plazo (como tableros de control, tableros de mando, reportes de operación, tableros basado en estadística descriptiva).

12. **Establecer aspecto financiero:** Se define el aspecto financiero que se construye en base a cada una de las fases considerando el tiempo, recursos humanos e insumos físicos y lógicos para elaborar la solución Big Data.

El cuadro a continuación resume el modelo de Gobierno de Datos, WebGestiones considerando el qué se desea obtener y el cómo se obtendrá.

Tabla 5 Modelo de gobierno de datos WebGestiones de Dr. Gabriel Guerrero

Modelo de Gobierno de Datos Dr. Gabriel Guerrero			
Modelo de la instrumentación para el manejo y la gobernanza de grandes volúmenes de datos			
<i>¿Qué se desea obtener?</i>			<i>¿Cómo se obtendrá?</i>
Seis pasos para el gobierno de datos			Modelo de datos WebGestiones
1	ESTABLECER METAS	Premisas que guían la operación y desarrollo de la información	1. Asimilar la misión empresarial 2. Interpretar cada uno de los objetivos empresariales 3. Determinar procesos productivos 4. Conocer sistemas que integran los procesos 5. Determinar el conjunto de datos que genera cada sistema 6. Definir resultados que desean medirse 7. Definir métricas 8. Determinar fuentes relevantes 9. Definir el mecanismo de construcción de la métrica 10. Determinar visualización 11. Desarrollar informes 12. Establecer aspecto financiero
2	DEFINIR MÉTRICAS	Conjunto de medidas usadas para evaluar la efectividad	
3	TOMAR DECISIONES	Estructura organizacional para crear políticas de decisión	
4	COMUNICAR POLÍTICAS	Técnicas usadas para comunicar decisiones en la organización	
5	MEDIR RESULTADOS	Comparar resultados	
6	AUDITAR	Herramienta para comprobar todo	

Fuente: Elaboración propia

2.3 Comparativo del marco de trabajo de Bill Schmarzo y de Modelo de gobierno de datos de Dr. Gabriel Guerrero

El marco de trabajo de Schmarzo y el modelo de gobierno de datos de Guerrero tienen diversas similitudes que se resaltan al hacer un comparativo. En ambos casos se consideran dos aspectos sobre el uso de Big Data en las organizaciones, que son el qué se desea lograr y el cómo se pretende lograr.

El aspecto que considera el qué se desea lograr, toma en cuenta, la estrategia global de la organización. En este punto, Schmarzo toma en cuenta un modelo de estrategia empresarial donde consideran cinco pasos para lograr cumplir la estrategia empresarial, es decir el objetivo principal que pretende lograr la organización. En este aspecto, el Dr. Guerrero, define lo que se espera lograr tomando en cuenta los seis pasos para el gobierno de datos. Es decir que tanto Bill Schmarzo como Gabriel Guerrero centran su primer proceso del modelo en definir la meta u estrategia empresarial, que es el objetivo general que se desea lograr con la implementación de un modelo de Big Data en la organización.

Tabla 6 Comparativo del modelos

¿Qué se desea lograr?					
Marco de trabajo de Bill SCHMARZO			WebGestiones, Dr. Gabriel Guerrero		
Documento de estrategia con iniciativas clave para aplicar			Pasos para el Gobierno de Datos		
	DIMENSIONES	DESCRIPCIÓN		DIMENSIONES	DESCRIPCIÓN
1	ESTRATEGIA EMPRESARIAL	Detalles del objetivo global	1	ESTABLECER METAS	Premisas que guían la operación y desarrollo de la información
2	INICIATIVAS EMPRESARIALES	Proyecto multidisciplinario con objetivos definidos	2	DEFINIR MÉTRICAS	Conjunto de medidas usadas para evaluar la efectividad
3	RESULTADO Y FACTORES CRÍTICOS	Factores críticos de éxito para llevar a cabo las iniciativas empresariales	3	TOMAR DECISIONES	Estructura organizacional para crear políticas de decisión
4	TAREAS	Tareas específicas para sustentar las iniciativas empresariales	4	COMUNICAR POLÍTICAS	Técnicas usadas para comunicar decisiones en la organización
5	FUENTES DE DATOS	Fuentes de datos necesarias para lograr la estrategia empresarial	5	MEDIR RESULTADOS	Comparar resultados
			6	AUDITAR	Herramienta para comprobar todo

Fuente: Elaboración propia

Una vez que se tiene un completo entendimiento de qué desea lograr la organización implementando un modelo de Big Data, los expertos Schmarzo y Guerrero proceden a explicar el cómo se pretende implementar.

La tabla a continuación refleja la existencia de diversas formas de implementar Big Data en una organización. Por una parte, Bill Schmarzo subraya la importancia de identificar dónde está ubicada la empresa en relación al rumbo que desea tomar la organización al utilizar Big Data por que lo que ofrece dos premisas a considerar: si los usuarios conocen los tipos de preguntas que desean responder con el uso de Big Data y si los usuarios saben el tipo de decisiones que desean tomar. Para Schmarzo, es primordial entender cuales son las áreas con mayor valor para la empresa para, mediante el uso de Big Data, ofrecer beneficios adicionales. Ahora bien, para proporcionar nuevos conocimientos sobre el negocio, Schmarzo utiliza la técnica de los factores motrices del valor de Big Data. Los cuatro factores motrices del valor de Big Data permiten catalogar los datos según la utilización que se le puede dar a los datos para así convertirlos en información relevante para la empresa.

Por otra parte, el modelo de Gobierno de Datos *WebGestiones* analiza el conjunto de los datos sin importar su estructura. Lo que predomina en el modelo de Dr. Guerrero es la importancia de las métricas que se establecen una vez que se definen los resultados que desean medirse. Ahora bien, ambos modelos coinciden en la importancia de conocer perfectamente los procesos productivos de la organización con la finalidad de identificar las fuentes de datos relevantes y por lo tanto obtener resultados satisfactorios.

El cuadro a continuación resume el comparativo de los dos modelos citados anteriormente en relación al cómo se implementa Big Data en las organizaciones.

Tabla 7 Comparativo sobre cómo implementar un modelo de Big Data

<i>¿Cómo se logrará?</i>		
Técnicas de creación de valor de Bill SCHMARZO		Modelo de datos WebGestiones
Técnicas para entender dónde y cómo puede repercutir Big Data en la organización		Pasos para instrumentar el manejo y gobernanza de los grandes volúmenes de datos
<p>Se debe identificar alguna de las 2 premisas:</p> <ul style="list-style-type: none"> - Los usuarios conocen el tipo de preguntas que intentan responder para sustentar los principales procesos empresariales - Los usuarios conocen el tipo de decisiones que intentan tomar en los principales procesos empresariales 	<p>¿Qué objetivo desea cumplir la organización?</p> <ul style="list-style-type: none"> ▪ Descubrir nuevas posibilidades de obtener beneficios que puedan repercutir en su marketing y su organización de ventas. ▪ Reducir los costes de adquisición, fabricación, inventario, cadena de suministro, distribución, publicidad, ventas y mantenimiento. ▪ Atenuar el riesgo en todos los aspectos operativos y financieros de la cadena de valores de la organización. ▪ Conseguir nuevos clientes, productos y conocimientos operativos que puedan utilizar para lograr una ventaja competitiva frente a la competencia y sacar un mayor beneficio en el sector. 	<ol style="list-style-type: none"> 1. Asimilar la misión empresarial 2. Interpretar cada uno de los objetivos empresariales 3. Determinar procesos productivos 4. Conocer sistemas que integran los procesos 5. Determinar el conjunto de datos que genera cada sistema 6. Definir resultados que desean medirse 7. Definir métricas 8. Determinar fuentes relevantes 9. Definir el mecanismo de construcción de la métrica 10. Determinar visualización 11. Desarrollar informes 12. Establecer aspecto financiero
Se crean en relación al objetivo que dese lograr la organización		
<p>Se considera la técnica de los factores motrices del valor de Big data de Bill SCHMARZO</p> <p>Factor 1: Acceso a datos transaccionales detallados</p> <p>Factor 2: Acceso a datos sin estructurar</p> <p>Factor 3: Acceso a datos de baja latencia</p> <p>Factor 4: Integración de la analítica predictiva</p>		

Fuente: Elaboración propia

2.4 Resumen de aportaciones

En este capítulo, se presentaron dos modelos de Big data que utilizan distintas técnicas para identificar las áreas donde se puede aplicar Big Data para potenciar las estrategias empresariales. En un primer punto, se consideró el marco de trabajo de Bill Schmarzo para cerciorarse que los objetivos organizacionales e iniciativas Big Data

estén orientados hacia la misma finalidad. Una vez confirmado que los objetivos empresariales y las iniciativas de Big Data están dirigidas al mismo objetivo, se procede a utilizar la técnica de los factores motrices de creación de valor para identificar dónde y cómo puede repercutir Big Data en la organización.

En un segundo, se muestra el modelo de gobierno de datos llamado *WebGestiones* que es un modelo de Big Data elaborado por Gabriel Guerrero, experto científico de datos mexicano, en el cual se muestran los pasos a seguir para poder implementar *WebGestiones* en una organización sin importar el giro.

En un tercer y último punto se hace un comparativo de los dos modelos de Big Data en el que después del análisis se destaca la gran similitud que existe del uso de Big Data para la toma de decisiones. Ambos modelos consideran dos aspectos primordiales que son el qué y el cómo, es decir el qué ambiciona lograr la organización con la implementación de un modelo de Big Data para la toma de decisiones y el ¿cómo? pretende lograrlo. Este capítulo permite conocer los modelos que guiarán la construcción de un modelo de Big Data para la toma de decisiones que se desarrolla más adelante.

Capítulo 3: Marco referencial

“El valor de una educación universitaria no es el aprendizaje de muchos datos, sino el entrenamiento de la mente para pensar.”

Albert Einstein

En este tercer capítulo se muestra el marco de referencia donde se considera el documento de estrategia para instrumentar una solución Big Data para Starbucks realizado por Bill Schmarzo. Asimismo, se muestra la aplicación de la técnica de los cuatro factores motrices del mismo autor para ilustrar la forma de implementar estas técnicas para dar valor a los datos de una organización.

3.1 Documento de estrategia para instrumentar una solución Big Data para Starbucks por Bill Schmarzo

Starbucks Corporation es una empresa dedicada a la elaboración de café que cuenta con 17,800 franquicias en 50 países. La cadena de café internacional fundada en Seattle Washington cotiza en la bolsa NASDAQ desde el año de 1992.

A continuación se muestra un extracto del Informe Anual de Starbucks en 2011 que sirve como guía para el desarrollo del documento de estrategia para Big Data.

Nuestro objetivo como minoristas es ser los principales minoristas y la marca de café líder en cada uno de nuestros mercados objetivo, vendiendo café y los productos relacionados de mayor calidad, de modo que cada cliente tenga una experiencia única en Starbucks. Esta experiencia se construye en base a un mejor servicio al cliente, así como en unas tiendas gestionadas por la empresa y con un buen mantenimiento, que reflejen las personalidades de las comunidades en las que operan, consiguiendo de este modo clientes muy fieles. (Informe Anual de Starbucks de 2011)

Tomando en cuenta lo indicado en el informe, se observa que la estrategia empresarial es “ofrecer una experiencia única al cliente de Starbucks” por lo que se requieren identificar de dos a cuatro iniciativas empresariales clave para sustentar dicha estrategia. Ahora bien, en la cadena cafetera existe una “Tarjeta Oro” que consiste en una tarjeta personalizada gratuita que requiere ser registrada en la página de Starbucks www.starbucks.com.mx con los datos personales del usuario como nombre, dirección, fecha de nacimiento y otros. Esta tarjeta requiere ser deslizada por la terminal punto de venta en cada compra para acumular puntos y posteriormente canjearlos por algún producto de la marca. Considerando el punto anteriormente mencionado, Schmarzo ofrece dos iniciativas empresariales:

1. Aumentar el número de clientes “Tarjeta Oro”, pasando de seis millones a nueve millones en los próximos 12 meses.
2. Aumentar los beneficios y el compromiso de los clientes “Tarjeta Oro” incrementando las visitas a la tienda, los gastos por visita y la probabilidad de que cada miembro de la Tarjeta Oro recomiende a Starbucks.

El siguiente punto que especifica Bill Schmarzo en este ejemplo son los resultados deseados que Starbucks querría alcanzar para sustentar estas iniciativas empresariales, las cuales son:

- Desarrollar un conocimiento detallado de los comportamientos y las preferencias de los clientes Tarjetas Oro, así como de qué cosas les interesan y les apasionan.
- Actuar en base al conocimiento detallado de estos clientes para aumentar los beneficios de las tiendas, a la par que se mejora la experiencia global del cliente Starbucks.

El autor identifica los siguientes factores críticos de éxito:

- Ampliar el número de puntos de recopilación de datos de los clientes, aprovechando los medios sociales, la aplicación de Starbucks para móviles, el sitio Web y otras técnicas de comunicación más tradicionales como el correo directo, el correo electrónico y las encuestas.
- Aprovechar las transacciones de los miembros de Tarjeta Oro, sus reacciones (encuestas) y los datos de las redes sociales y los dispositivos móviles para obtener una imagen mucho más completa y precisa (perfil) de los distintos tipos o segmentos de clientes Tarjeta Oro.
- Descubrir y después integrar los conocimientos específicos de los clientes en los sistemas de operaciones (*call centers*, marketing directo, opiniones en tiendas), administración (informes de tiendas y cuadros de administración) y fidelización (servicio al cliente, marketing).

El siguiente punto es definir las tareas clave que son las operaciones del día a día de cada uno de los empleados de Starbucks. Son tareas específicas que la cadena necesita que ejecuten sus diferentes funciones empresariales. Es la sección que Schmarzo define como el “cómo se hace” del documento, por lo que para este ejemplo, enumera las siguientes tareas clave:

- Proporcionar más y mejores posibilidades de recopilar información sobre el compromiso del cliente a través de varios canales: en la tienda por la Web, en las redes sociales y en los dispositivos móviles.
- Perfilar y micro-segmentar los clientes para mejorar el marketing y ser más eficaces; haciendo hincapié en doblar el número de segmentos de clientes con perfil a finales de año.

- Analizar los datos de las redes sociales para identificar y monitorizar a los defensores de la marca; medir y monitorizar la probabilidad de que los clientes Tarjeta Oro recomienden a Starbucks. Es lo que se denomina LTR por sus siglas en inglés al proceso de *Likelihood to Recommend*.
- Monitorizar, medir y ajustar la efectividad del compromiso de los clientes, incluyendo el número de visitas, los beneficios por visita, el margen por visita, las tendencias de la cesta de compra y los defensores de la marca.

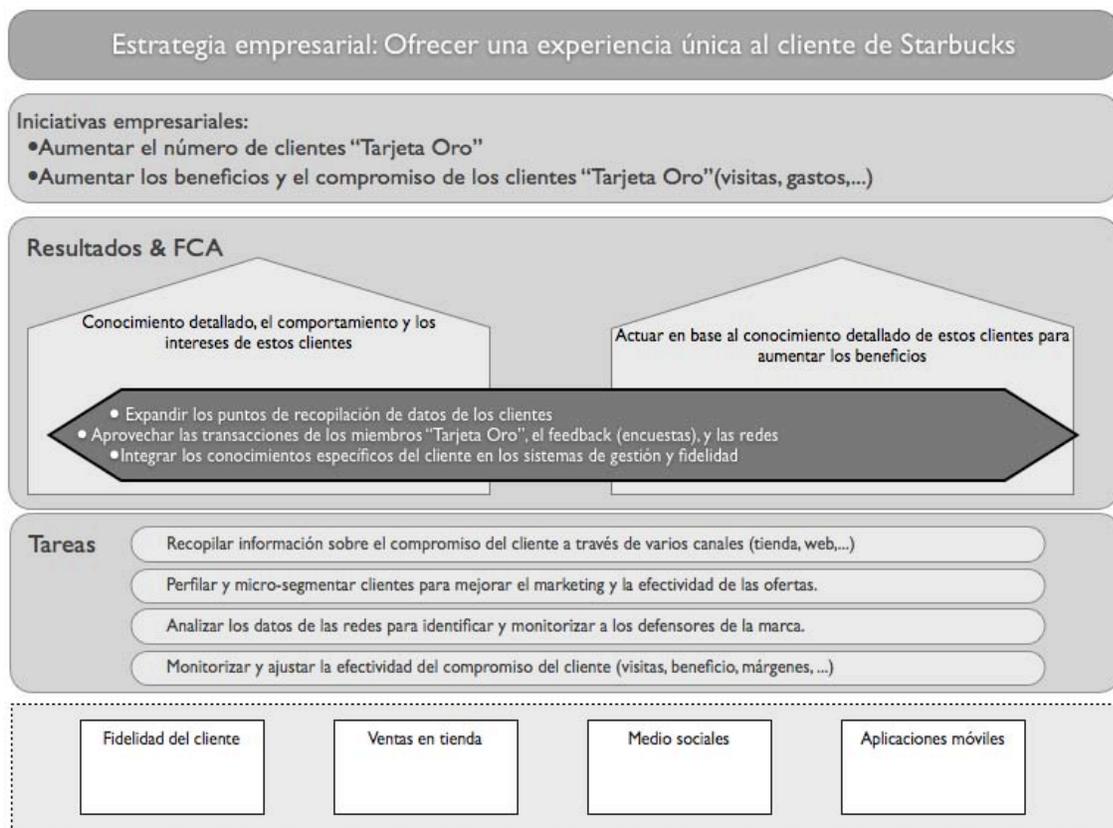
Por último, como subraya Schmarzo, “se determinan las fuentes de datos necesarias para sustentar la estrategia empresarial y las iniciativas empresariales clave de Starbucks”. (SCHMARZO, 2013) Las fuentes de datos que menciona el autor son:

- Datos demográficos, estilos de vida, preferencias de ubicación y datos de comportamientos asociados.
- Historial de compra de productos del cliente.
- Datos de ventas de productos.
- Publicaciones y opiniones en los medios sociales.
- Datos de localización de las aplicaciones móviles (unidos a los datos de compra de productos).

El experto en Big Data determina que “lo útil del proceso de desarrollo de este documento de estrategia es que además de definir las fuentes de datos y las capacidades analíticas en las que debe centrarse la organización para proporcionar un soporte completo a la estrategia empresarial, también define lo que las fuentes de datos y las capacidades analíticas no necesitan en este momento concreto. Hace de valla de contención para que la organización siga centrada, a la par que desarrolla simultáneamente las capacidades de Big Data a largo plazo, proporcionando la hoja de ruta empresarial pertinente para introducir nuevas tecnologías e innovaciones en Big Data.” (SCHMARZO, 2013)

A continuación se muestra el documento que reúne “los datos anteriormente mencionados que sirve de apoyo para identificar los requisitos de la administración de datos, la plataforma de datos, la inteligencia empresarial y los requisitos analíticos en los que se apoyará la estrategia empresarial” de la cadena que es: ofrecer una experiencia única al cliente de Starbucks. (SCHMARZO, 2013)

Figura 5 Documento de estrategia de Big Data para Starbucks



Fuente: Bill SCHMARZO (2014) Big Data: El poder de los datos.

3.2 Aplicación de las técnicas de creación de valor de Bill Schmarzo

A continuación se muestra un ejemplo de *merchandising* donde se aplica la técnica de creación de valor de Schmarzo con el objetivo de identificar las áreas de negocio

específicas en las que los cuatro factores motrices de Big Data pueden repercutir en la iniciativa empresarial clave de una organización.

“El *merchandising* consiste en poner precio, publicitar, empaquetar y colocar un producto para enganchar al cliente, ya sea en tiendas físicas o para su venta por Internet o una aplicación móvil. El *merchandising* aplica los cuatro principios del negocio minorista (presentación de los productos, colocación, promoción y precio) para generar ventas y beneficios tanto para el producto en sí como para las tiendas. El objetivo del *merchandising* es mostrar, destacar, promocionar y poner precio a productos, ya sea por separado o combinándolos (como por ejemplo, zapatos y calcetines) para lograr que el cliente los compre”, indica Bill Schmarzo. (SCHMARZO, 2013)

Para este ejemplo se considera la empresa Foot Locker , minorista de calzado y vestimenta deportiva para hombre y mujer que distribuye sus productos por Internet y en tiendas propias.

En el informe anual de 2010, Foot Locker anuncia que su principal iniciativa empresarial es convertirse en el principal vendedor de calzado y vestimenta deportiva proporcionando una clara visión de las marcas que venden. Asimismo, en dicho informe se subrayan las prioridades estratégicas de la empresa que son:

- Disponer de un buen surtido de vestimenta deportiva
- Convertir las tiendas y sitios Web en lugares atractivos para ir de compras
- Aumentar la productividad de todas las áreas

2.4.1 Cuatro factores motrices de Big Data: Ejemplo de *merchandising* en *Foot Locker*

La hoja de trabajo ofrece un entorno de trabajo en el que se aplican los cuatro factores motrices de Big Data para tener ideas sobre las áreas específicas donde Big Data

podría tener un impacto en la iniciativa empresarial principal del *merchandising* de Foot Locker.

- **Factor nº1: Acceso a datos transaccionales más detallados**

En este punto se responde a la pregunta: ¿Cómo se podrían utilizar los datos transaccionales de las Terminales Punto de Venta (TPV) para mejorar la segmentación de los clientes?

- Se pueden utilizar las transacciones por TPV detalladas, junto con los datos de fidelidad del cliente de Foot Locker, para que el número de micro segmentos pase de 50 a 500 en los comportamientos de compra y las tendencias por producto individuales y colectivas.
- Se pueden crear campañas publicitarias mucho más dirigidas y concretas, centrándose en los segmentos de clientes más fieles y realizando actividades publicitarias específicas para la temporada, la ciudad o el código postal.

- **Factor nº2: Acceso a datos sin estructurar**

El segundo factor contesta a la pregunta: ¿Cómo se podrían integrar los datos de los medios sociales en busca de intereses, pasiones, asociaciones y afiliaciones de los clientes relacionadas con los deportes, para de este modo crear modelos de *merchandising* más completos y dirigidos?

- Se pueden explorar los datos de los medios sociales en busca de intereses, pasiones, asociaciones y afiliaciones de los clientes relacionadas con deportes, para de este modo crear modelos de merchandising más completos y dirigidos.
- Se pueden utilizar los datos de las aplicaciones móviles para micro-segmentar clientes basados en las localizaciones, tiendas o deportes específicos.

- **Factor nº3: Acceso a datos de baja latencia (en tiempo real)**

Es decir: ¿Cómo se utilizarían los datos en tiempo real para mejorar la segmentación de clientes?

- Se podrían recalcular los modelos de *merchandising* justo después de eventos deportivos locales significativos
- Se podrían integrar los eventos deportivos locales para perfeccionar las campañas sobre la marcha.

- **Factor nº4: Integración de la analítica predictiva**

El cuarto y último factor contesta a la pregunta: ¿Cómo se podría utilizar la analítica avanzada o predictiva para mejorar la segmentación de los clientes?

- Se podrían desarrollar modelos analíticos que monitoricen y establezcan prioridades en el desarrollo de la campaña actual para poder determinar cuál es el mejor micro-segmento de clientes al que dirigir una campaña concreta para unos objetivos de venta establecidos.
- Se podría desarrollar un modelado de distribución multiplataforma para optimizar el gasto de *merchandising* en actividades orientadas al correo electrónico, el correo ordinario, la Web, los móviles o tiendas.

La hoja de trabajo para la visualización de Big Data considerando el ejemplo de *merchandising* de Foot Locker sirve como una herramienta para la generación de ideas acerca de cómo se podrían aplicar los cuatro factores motrices de Big Data a una iniciativa empresarial específica.

3.3 Resumen de aportaciones

En este capítulo se muestran ejemplos de la implementación las técnicas de Big Data expresadas en el capítulo dos. En relación al marco de trabajo, en el cual se extrae la información de la misión de la empresa para determinar una iniciativa empresarial, se toma el ejemplo de *Starbucks*. En este caso, el documento reúne la información de manera que la alta dirección pueda conocer de forma general los requisitos para llevar a cabo una estrategia empresarial para el logro de resultados y por ende la toma de decisiones.

Después, se muestra, mediante el ejemplo de *Foot Locker*, como se puede implementar la técnica de *Big Data* para dar valor a los datos de una organización. En este aspecto, se aplican los cuatro factores motrices para conocer las áreas en las que se puede aplicar Big Data la toma de decisiones.

Capítulo 4: Metodología

"Hay que perseverar y, sobre todo, tener confianza en uno mismo".
Marie Curie

En este cuarto capítulo se desarrolla la metodología para esta investigación. Efectivamente, se muestra el planteamiento del problema que se centra en la necesidad de contar con un modelo para la toma de decisiones en las organizaciones que manejan grandes volúmenes de datos. Asimismo se muestran las preguntas y los objetivos de esta investigación, la hipótesis de trabajo, el cuadro de congruencia, el diseño de la investigación para finalizar con el procedimiento empleado para la construcción del modelo propuesto, el desarrollo de dicho modelo así como su validación y por último los ajustes realizados al modelo propuesto.

4.1 Planteamiento del problema

Hoy en día, las empresas se han tenido que enfrentar a una creciente cantidad de datos generados por sus sistemas que requieren ser procesados para convertirlos en información y luego transformarlos en conocimiento listo para ser usado en la toma de decisiones. Los datos por si mismos, no necesariamente proveen información útil si son recolectados y organizados inapropiadamente. Asimismo, para que los datos sean convertidos en información requieren ser procesados en una forma que tengan significado para sus receptores.

Actualmente se cuenta con una gran diversidad de fuentes de información con grandes cantidades de datos de diferentes tipo que requieren ser analizados rápidamente para una eficiente toma de decisiones. Ciertamente, “la información sólo es útil cuando es relevante para su propósito y detallada, confiable, precisa y verificable, actualizada y entregada a tiempo, completa, inteligible, consistente, práctica y fácil de manejar y proteger”. (LÓPEZ-MARTÍNEZ, 2016) Muchas veces, la calidad de los datos no es

óptima por lo que requieren de un mecanismo que sea capaz de analizar los datos sin importar su calidad, tamaño o tipo.

El uso de datos abiertos y el análisis estadístico del histórico generado por las organizaciones ha favorecido la toma de decisiones a nivel directivo dado que con un estudio descriptivo, predictivo y prescriptivo de los datos generados es posible potenciar los objetivos organizacionales.

En la actualidad, son limitados los estudios analíticos que se han hecho a nivel mundial ya que aún existe una limitante para acceder a los datos. No obstante en una presentación pública de la iniciativa de datos abiertos del Gobierno de la República se muestran múltiples ejemplos del impacto social de los datos abiertos.

En cuanto al impacto social, *Big Data Analytics* sirve para transmitir información. Actualmente existe un nuevo término llamado “gobierno abierto” (Open government) que “significa impulsar un nuevo modelo de gobernanza, que requiere de un proceso de compromiso, colaboración y corresponsabilidad permanente y sostenible por parte del gobierno y todos los sectores de la sociedad.” (República, 2014) Entre más información se cuente y se pueda transmitir a más personas, entonces cada vez más la ciudadanía se siente involucrada y se fomenta la democracia.

El problema de la gran cantidad de información generada por las organizaciones es cada vez mayor aunado al creciente uso de dispositivos conectados a internet que producen grandes cantidades de datos que requieren ser analizados con prontitud para una eficiente toma de decisiones.

Cada vez más se suman organizaciones a la ola de las nuevas herramientas tecnológicas para poder comprender la manera en que la información, procesada correctamente permite conocer a fondo los movimientos, gustos y necesidades de los usuarios con la finalidad de conocer de qué manera la organización se adapta a este cambio tecnológico y cómo podría potenciar la generación de valor para la toma de

decisiones que son de urgente necesidad en un mundo donde las tecnologías de información han revolucionado la forma de actuar.

Existen diversos modelos que se han desarrollado para el uso de *Big Data Analytics* en las organizaciones con manejo grandes volúmenes de datos. Específicamente, en esta investigación se analizan dos modelos, el de Bill Schmarzo y el de Gabriel Guerrero. Dichos modelos se centran en el qué se desea lograr y en el cómo se pretende lograr desde un punto de vista informático. Ahora bien, los tomadores de decisiones, requieren conocer un modelo de análisis de grandes cantidades de información de manera gerencial de forma que el concepto de *Big Data Analytics* sea entendible para un administrador que toma decisiones a nivel directivo sin adentrarse en terminología técnica de dicha herramienta. Es por esto que resulta importante diseñar un modelo de *Big Data Analytics* que tome en cuenta el pensamiento administrativo para mejorar toma de decisiones a nivel directivo en las organizaciones con grandes volúmenes de datos.

4.2 Preguntas de investigación

4.2.1 General

¿Cómo una organización que genera grandes volúmenes de datos a través de *Big Data Analytics* puede mejorar la toma de decisiones?

4.2.2 Específicas

- ¿Cuál es la definición de *Big Data*?
- ¿*Big Data Analytics* sirve como un proceso de creación de valor para las organizaciones?

- ¿Cuáles son las etapas que debe incluir el desarrollo de un modelo de *Big Data Analytics* para la toma de decisiones en las organizaciones con grandes volúmenes de datos?
- ¿Qué modelos existe del uso de *Big Data Analytics* en las organizaciones?
- ¿Cuáles son los límites de un modelo de *Big Data Analytics* para la toma de decisiones?

4.3 Objetivos de la investigación

4.3.1 General

Diseñar un modelo de *Big Data Analytics* que se pueda aplicar a las organizaciones con manejo de grandes volúmenes de datos para mejorar la toma de decisiones a nivel directivo.

4.3.2 Específicos

- Investigar si *Big Data Analytics* sirve como un proceso de creación de valor para las organizaciones.
- Identificar las etapas que debe incluir el desarrollo de un modelo de *Big Data Analytics*.
- Revisar en la literatura algunos modelos de uso de *Big Data Analytics* en las organizaciones.
- Investigar la definición de *Big Data Analytics* para una toma de decisiones.

- Diseñar un modelo de *Big Data Analytics* para la mejora en la toma de decisiones.

4.4 Hipótesis de trabajo

El uso de un modelo de Big Data Analytics en las organizaciones con manejo de grandes volúmenes de datos mejora la toma de decisiones a nivel directivo.

4.5 Cuadro de Congruencia

Objetivo General	
Diseñar un modelo de <i>Big Data Analytics</i> que se pueda aplicar a las organizaciones con manejo de grandes volúmenes de datos para mejorar la toma de decisiones a nivel directivo.	
Pregunta de Investigación	
¿Cómo una organización que genera grandes volúmenes de datos a través de <i>Big Data Analytics</i> puede mejorar la toma de decisiones?	
Hipótesis de trabajo	
H: El uso de un modelo de Big Data Analytics en las organizaciones con manejo de grandes volúmenes de datos mejora la toma de decisiones a nivel directivo.	
Objetivos específicos	Pregunta de Investigación
<ul style="list-style-type: none"> ▪ Investigar la definición de <i>Big Data</i> ▪ Investigar si <i>Big Data Analytics</i> sirve como un proceso de creación de valor para las organizaciones. ▪ Identificar las etapas que debe incluir el desarrollo de un modelo de <i>Big Data Analytics</i>. ▪ Revisar en la literatura algunos modelos del uso de <i>Big Data Analytics</i> en las organizaciones. ▪ Diseñar un modelo de <i>Big Data Analytics</i> para la mejora en la toma de decisiones. 	<ul style="list-style-type: none"> ▪ ¿Cuál es la definición de <i>Big Data</i>? ▪ ¿<i>Big Data Analytics</i> sirve como un proceso de creación de valor para las organizaciones? ▪ ¿Cuáles son las etapas que debe incluir el desarrollo de un modelo de <i>Big Data Analytics</i> para la toma de decisiones en las organizaciones con grandes volúmenes de datos? ▪ ¿Qué modelos existe del uso de <i>Big Data Analytics</i> en las organizaciones?

4.6 Taxonomía del estudio

El estudio que se realizará será:

- ✓ **No Experimental:** El estudio es no experimental porque no se manipularán de manera intencional las variables para analizar las consecuencias de tal manipulación. En la investigación se analizarán las variables y se observarán los fenómenos en su ambiente natural para posteriormente analizarlos y realizar un modelo.
- ✓ **Exploratorio:** A través del estudio se investiga una solución para la toma de decisiones.
- ✓ **De corte transversal:** El modelo no pretende ser puesto en operación por lo que no se ambiciona la evaluación de los resultados a través del tiempo.
- ✓ **Cualitativo:** El modelo pretende ser evaluado mediante entrevistas cualitativas hechas a expertos en el tema de Big Data

4.7 Diseño de la investigación

El diseño de la investigación se realiza en 7 etapas:

1. **Identificación del problema:** Definición del problema actual que requiere una solución para una eficiente toma de decisiones en las organizaciones.
2. **Análisis bibliométrico:** Aplicación de métodos matemáticos y estadísticos para el análisis de publicaciones científicas.
3. **Formulación de hipótesis:** Se planteará una hipótesis para la solución del problema
4. **Elaboración del marco teórico:** Establecimiento del contexto teórico y conceptual bajo el cual será establecida la propuesta modelo *de Big Data Analytics* para la toma de decisiones en las organizaciones.

5. **Diseño del modelo:** Elaboración de una propuesta de modelo de *Big Data Analytics* para la toma de decisiones a nivel directivo en las organizaciones con grandes volúmenes de datos.
6. **Evaluación del modelo por expertos:** El modelo propuesto será presentado a 10 expertos de Big Data con la finalidad de recibir retroalimentación del modelo.
7. **Ajustes al modelo:** Con base en los comentarios recibidos por los expertos se llevaran a cabo los ajustes pertinentes al modelo propuesto.

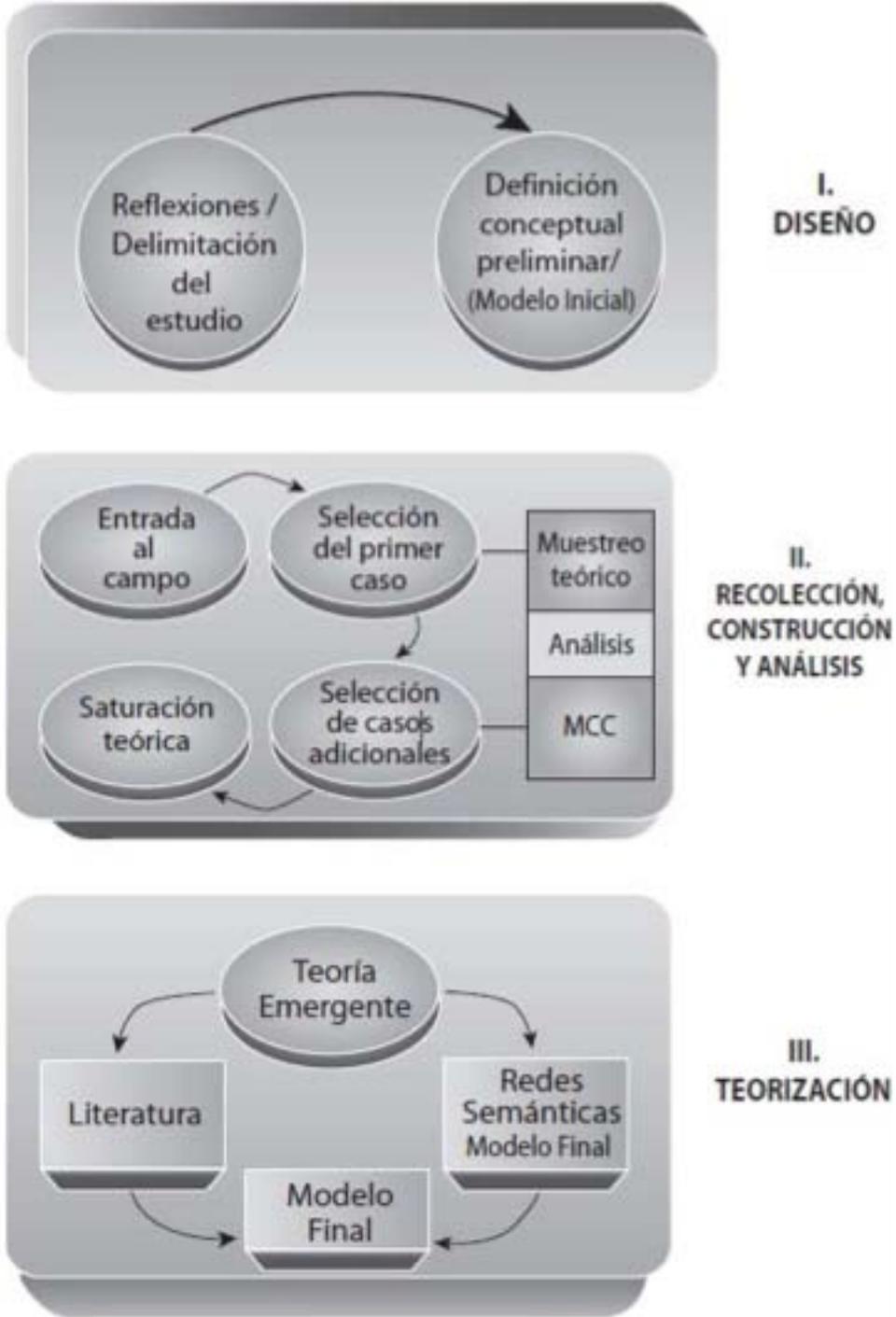
4.8 Procedimiento para la construcción del modelo

Los autores Milagros Bolseguí y Antonio Fuguet Smith ofrecen un procedimiento para la construcción de un modelo en su artículo titulado “Construcción de un modelo conceptual a través de la investigación cualitativa”. (BOLSEGUÍ & FUGUET SMITH, 2006). Los autores “plantearon un procedimiento para la construcción de un modelo conceptual por medio de la investigación cualitativa, justificándose en: que las características para el estudio de fenómenos sociales son la relevancia contextual, énfasis en todo el proceso y no solamente en el control y experimentación, ya que es necesario orientarse en el desarrollo de teorías que expliquen datos, y no en la búsqueda de datos que correspondan a una teoría previa. Bajo estas premisas el ser humano es el instrumento principal para la recolección de datos, ya que se reconoce la construcción del conocimiento tácito de los investigadores así como que la investigación se sustenta por el sistema de valores que caracterizan al investigador, el informante, el paradigma que se elija y la teoría sustantiva que se relaciona, la cual puede orientar tanto los métodos de recolección y análisis como las formas de interpretar los hallazgos.” (LIMA MACIEL, 2012)

Bajo una perspectiva cualitativa Bolseguí y Fuguet describen tres fases para la elaboración de un modelo que son:

- *Diseño*: se describe el Diseño General de la Investigación, se especifican los tipos de informantes claves seleccionados así como la naturaleza del método empleado: La entrevista cualitativa. Se describen también aspectos referidos al proceso de recolección de información.
- *Recolección, construcción y análisis*: se explica que tanto la categorización, como el análisis e interpretación son actividades conjuntas, estrechamente vinculadas entre sí.
- *Teorización y construcción del modelo*: Se argumenta que la investigación avanza desde una conceptualización preliminar denominada-modelo inicial- que adquiere mayor sentido y profundidad en fases posteriores –modelos intermedios- hasta llegar a una conceptualización final que se presenta como modelo final (modelo conceptual). (BOLSEGUÍ & FUGUET SMITH, 2006)

Figura 6 Fases del Proceso de Investigación



Fuente: Bolsegui, M., & Fuguet, S. A. (2006). *Construcción de un modelo conceptual a través de la investigación cualitativa*. Sapiens, 207 - 229.

4.8.2 Fase de inicio

Según Bolseguí y Fuguet, el diseño de una investigación cualitativa es emergente y en cascada, “se elabora en la medida que avanza la investigación, es decir, es un diseño flexible. El problema de investigación da lugar a un cuestionamiento y reformulación constante, que guía y establece los límites de la investigación. El diseño surge en la medida que la investigación se desarrolla.” (BOLSEGUÍ & FUGUET SMITH, 2006) En esta fase se define el objeto de estudio, se formulan las preguntas de investigación y se precisan los objetivos. Asimismo, se elabora un marco referencial preliminar que es enriquecido con literatura especializada para esta primera fase donde se incluyen ideas, autores y consultas. Además, se revisa literatura especializada para dilucidar interrogantes y delimitar categorías.

“Una vez iniciada la investigación, se van decidiendo otros aspectos y escenarios a estudiar. La incorporación progresiva de nuevos informantes o escenarios se relaciona con los intereses de la investigación, en términos de lo que se ha aprendido” afirman Bolseguí y Fuguet. En el siguiente gráfico se presenta un esquema del diseño general de investigación de los autores Bolseguí y Fuguet.

Figura 7 Diseño general de investigación



Fuente: Bolseguí, M., & Fuguet, S. A. (2006). *Construcción de un modelo conceptual a través de la investigación cualitativa*. Sapiens, 207 - 229.

4.8.3 Fase de Recolección, construcción y análisis

En esta segunda fase, Bolseguí y Fuguet firman que “en la investigación cualitativa el análisis y la interpretación son procesos vinculados, el propósito es el de interpretar el contenido de los relatos. Por ello se considera que el proceso de

análisis se desarrolla durante toda la investigación.” (BOLSEGUÍ & FUGUET SMITH, 2006) Durante el proceso de recolección de datos el investigador mantiene una interpretación fundamentada y extrae sus conclusiones a partir de las observaciones que realiza y de otros datos. Asimismo, el investigador presenta conclusiones atribuidas a una fuente real o una fuente genérica ya que le interesa preservar las realidades múltiples en las diferentes visiones de lo que sucede.

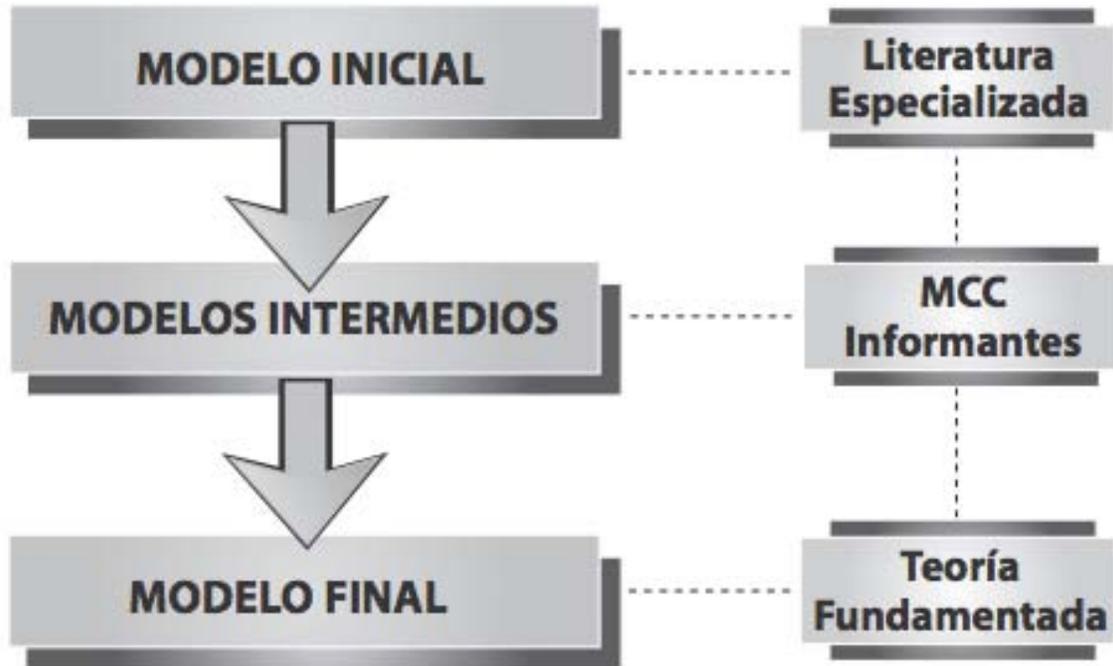
Como lo afirman Bolseguí y Fuguet, “en la investigación cualitativa el análisis y la interpretación son procesos vinculados, el propósito es el de interpretar el contenido de los relatos. Por ello se considera que el proceso de análisis se desarrolla durante toda la investigación. Durante el proceso de recolección de datos el investigador mantiene una interpretación fundamentada. El investigador extrae sus conclusiones a partir de las observaciones que realiza y de otros datos.” (BOLSEGUÍ & FUGUET SMITH, 2006). En esta fase, el investigador se interesa en conocer las realidades múltiples en las diferentes visiones de lo que sucede. Se recomienda que en el proceso de análisis de entrevistas, el analista documente empíricamente el proceso de construcción del significado con amplias ilustraciones y referencias a los registros, y describa las actividades discursivas complejas a través de las cuales los respondientes construyen significados. Por lo tanto, para facilitar el proceso de interpretación, el investigador categoriza, percibe, contrasta, compara, prioriza y establece nexos o propiedades.

4.8.4 Fase de teorización y construcción del modelo

El modelo desarrollado por Bolseguí y Fuguet propone, “en primer lugar, consultar diversas fuentes: bibliográficas, documentales, electrónicas, documentos institucionales, boletines informativos. En segundo lugar, entrevistar a diferentes investigadores y expertos en el área de evaluación; registrar y categorizar sus opiniones, para luego proceder a una triangulación de las diversas fuentes consultadas, con el interés de avanzar en la conceptualización de la cultura de evaluación.” (BOLSEGUÍ & FUGUET SMITH, 2006) En esta última fase de teorización y construcción del modelo, se retoma el modelo inicial que es la

conceptualización preliminar que fue adquiriendo mayor sentido y profundidad en fases posteriores, lo que coincide con una continua interpretación y reinterpretación de la información protocolar disponible que da pie a la segunda fase, es decir al diseño de modelos intermedios mediante la aplicación el método comparativo continuo (MCC) en el cual se aplican entrevistas y se contraste la visión del entrevistado con el modelo propuesto de manera que se incorporan las aportaciones y se formula una teoría emergente para concluir con la conceptualización final que es el modelo final.

Figura 8 Modelos



Fuente: Bolsegui, M., & Fuguet, S. A. (2006). *Construcción de un modelo conceptual a través de la investigación cualitativa*. Sapiens, 207 - 229.

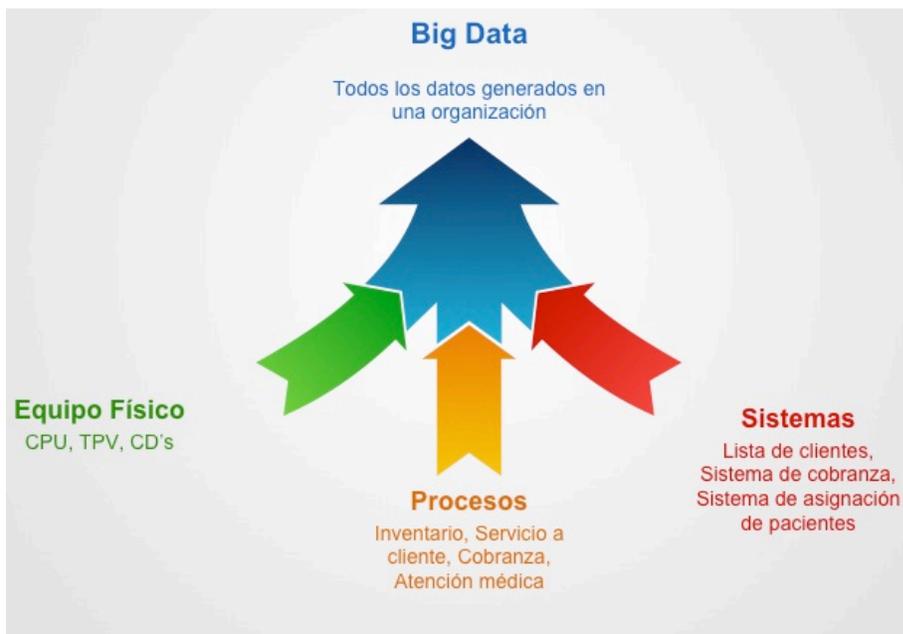
4.9 Modelo propuesto

El modelo que se propone a continuación, consta de diez fases. Se parte de la premisa que toda organización con grandes volúmenes de datos cuenta con tres componentes generadores de datos: los equipos físicos, los procesos y los sistemas. El equipo físico hace referencia a los “dispositivos físicos de un sistema informático. Incluye CPU, terminales, placas de memoria, impresoras, unidades de disco, etc. A veces se utiliza en este sentido el término máquina. » (Diccionario LID de Empresa y Economía, 2016)

Los procesos comprenden todas aquellas actividades de la organización que le permiten alcanzar su misión, como por ejemplo, el área de pensiones en una institución de salud pública, cobranza, atención a clientes, inventario, almacenamiento, mantenimiento, entre otras. Son las áreas de la empresa u organización que sirven para definir la razón de ser de la misma.

Un sistema en una organización permite almacenar la información generada por los procesos. Por lo que tanto, el equipo físico como los procesos y los sistemas son factores que influyen en la generación de grandes volúmenes de datos en las organizaciones.

Figura 9 Fuentes generadoras de grandes volúmenes de datos



Fuente: Elaboración propia

El modelo propuesto consta de diez fases que se explican a continuación.

Fase I: Identificar el eje de aplicación en la empresa

En esta fase se busca conocer cuáles son los procesos clave de la organización. En este aspecto, se debe estudiar el modo de operación de la organización para identificar las áreas de la empresa con un rendimiento inferior para conocer cuáles requieren mayor atención. A través de esta gestión del rendimiento empresarial que utiliza la analítica básica, se busca identificar los procesos y los indicadores empresariales clave. En este aspecto es primordial no perder de vista el principal objetivo empresarial.

Se recomienda conocer:

- Tendencias
- Comparaciones históricas

- Índices de la empresa: satisfacción del cliente, rendimiento de productos
- Indicadores de rendimiento clave: porcentaje de mercado, retorno de inversión
- Limitantes y recursos disponibles: fuentes de datos disponibles

Mediante el análisis de la gestión del rendimiento empresarial, se pretende conocer cuál es la situación actual de la empresa, hacia dónde desea llegar, identificar los problemas, las oportunidades y los procesos que requieren mayor atención.

Una vez conociendo los procesos de la organización, se debe seleccionar un solo proceso y conocer cuál será el eje de aplicación.

Los ejes de aplicación pueden ser:

- Mejorar la operación
- Incrementar ganancias
- Mejorar la venta a los clientes actuales, es decir vender más a los clientes actuales
- Buscar nuevos clientes
- Ampliar la cartera de productos

El área directiva de la empresa debe definir los siguientes aspectos:

- el proceso que desea mejorar
- el eje que desea aplicar a dicho proceso

FASE II: Ubicar los procesos asociados al eje de aplicación

Una vez identificados los procesos, se deben asociar con los ejes de aplicación con tal de detallar las tareas específicas que hay que realizar para sustentar las iniciativas empresariales.

Para ilustrar esta fase, se ofrece el ejemplo de una institución de salud pública, que tiene como procesos principales, la atención médica, la cobranza a empresas y el otorgamiento de pensiones. El directivo de la institución selecciona el proceso de mejorar de atención a médica y define como eje de aplicación, mejorar la operación de la atención al paciente.

FASE III: Identificar la lista de sistemas de soportan el proceso

Cada proceso tiene una serie de sistemas asociados al mismo. En esta etapa se requiere identificar la lista de sistemas que soportan el proceso a mejorar.

Al retomar el ejemplo de la fase II, los sistemas del proceso de atención médica serían:

- Sistema de alta de derechohabientes
- Sistema de vigencia de derechos
- Sistema de asignación de Unidad Médico Familiar

FASE IV: Analizar la lista de tablas de datos generados para cada sistema

Los sistemas de cada proceso tienen sus propias tablas con datos específicos. Un sistema puede tener un sinnúmero de tablas. Por ejemplo, el sistema de asignación de Unidad Médico Familiar posee 3 tablas:

- Tabla 1: Catálogo de pacientes, que indica la lista de los pacientes de la Unidad Médico Familiar
- Tabla 2: Datos de diagnósticos de pacientes, donde se indican las notas médicas de los pacientes
- Tabla 3: Datos de somatometría que se refiere a peso, talla e índice de masa corporal de cada paciente

En el supuesto que el objetivo es instrumentar una forma más eficiente de atención personalizada al paciente para evitar acciones médicas repetidas, el uso de Big Data permitirá la creación de un solo expediente clínico electrónico que sin

importar en qué clínica, hospital o centro sea leído, éste contenga toda la información del paciente.

FASE V: Recepción de datos en crudo (raw)

En esta quinta fase del modelo propuesto, se recaban los datos de cada una de las tablas tal cual como los entrega el sistema predictivo, sin ningún tipo de procesamiento. Esta forma de datos se les conoce como datos crudos (*raw data*) o datos originales (*as is data*).

Los datos originales se integran a un almacén de datos llamado *Big Data Box* para su procesamiento.

FASE VI: Gobernanza de los datos

El tratamiento de los datos mejor conocido como gobernanza de los datos (Data Governance) en términos informáticos, es el paso que prosigue al de recepción de datos en crudo. En efecto, una vez que se integran los datos crudos al *Big Data Box*, este almacén realiza un proceso de gobernanza de los datos, a través de un análisis de estructura y calidad de la información.

Para esta fase se hace uso de herramientas como Hadoop HDFS y Spark SQL para simplificar el análisis y proceso de grandes volúmenes de datos con tal de asegurar la calidad, el valor y la riqueza de los datos.

FASE VII: Definición de las reglas de negocio

En este séptimo punto, se retoman los aspectos de los modelos de Schmarzo y Guerrero en cuanto al establecimiento de métricas y se definen las reglas de negocio o métricas de calidad donde se determinan los metadatos que contienen la información que indica qué tipo y qué formato tiene cada dato. Se confirma la información de los datos con la finalidad de conocer cuántos campos tiene cada

tabla, qué tipo de dato contiene cada campo y se determinan las reglas de negocio.

Para ilustrar esta fase se retoma el ejemplo de la institución de salud pública. En el supuesto que la tabla de nombres de pacientes tiene 6 campos, estos podrían ser:

- El campo 1: nombre de pila del paciente
- El campo 2: apellido paterno del paciente
- El campo 3: apellido materno del paciente
- El campo 4: fecha de nacimiento del paciente
- El campo 5: lugar de nacimiento del paciente
- El campo 6: sexo del paciente

Cada campo tiene al menos una regla de negocio. Al considerar el ejemplo citado anteriormente, las reglas de negocio serían:

- Regla de negocio del campo 1: Sólo pueden ser letras
- Regla de negocio del campo 2: Sólo pueden ser letras
- Regla de negocio del campo 3: Sólo pueden ser letras
- Regla de negocio del campo 4: Es una fecha con 2 dígitos para el día seguidos (DD) de 2 dígitos para el mes (MM) , seguidos de 4 dígitos para el año (AAAA)
- Regla de negocio del campo 5: Debe coincidir con alguna ciudad del mundo
- Regla de negocio del campo 6: Sólo puede ser femenino (F) o masculino (M)

FASE VIII: Verificación de reglas de negocio

En esta fase se verifica que los datos que se encuentran dentro de la Big Data Box cumplan con las reglas de negocio establecidas.

Este análisis ofrece dos resultados:

1. Datos que sí cumplen con las reglas de negocio

2. Datos que no cumplen con las reglas de negocio

En ocasiones, en el campo sexo, suele haber más de una forma de definir femenino/masculino, como puede ser hombre/mujer o simplemente F/M o H/M lo que conlleva a una confusión a la hora de determinar el sexo del paciente. En este contexto se deben alinear los campos a un estándar previamente elegido.

Después del análisis, los datos que no cumplen con las reglas de negocio se extraen de la *Big Data Box* y se devuelven al cliente para su análisis y corrección puesto que esta propuesta de modelo se maneja bajo la noción de GIGO. El GIGO es la abreviatura de *Garbush In Garbush Out* que significa que si entra basura sale basura. Es decir que el éxito de esta propuesta de modelo depende de la garantía de calidad de los datos que entran. La *Big Data Box* incluye los datos que fueron seleccionados minuciosamente considerando su calidad, fiabilidad y valor para poner en relieve su capacidad predictiva con la finalidad de asegurar un correcto desarrollo de un modelo analítico.

FASE IX: Definir tipología del análisis de los datos

Una vez que está comprobada la calidad de los datos y que se extrajeron los datos que no cumplen con las reglas de negocio, la *Big Data Box* sólo debe incluir los datos con garantía de calidad. Dichos datos se fraccionan en dos categorías: micro y macro, según su análisis como se indicó en el primer capítulo.

Recuperando el ejemplo del paciente, mediante la trazabilidad del análisis micro o la búsqueda al átomo se puede conocer el historial del paciente, es decir con qué médico se atendió, en qué clínica u hospital, en qué fecha, cuál fue su diagnóstico, etc.

Por otra parte, el análisis global o macro, indica las estadísticas globales mediante el uso de estadística descriptiva. Por ejemplo, una empresa puede conocer qué

porcentaje de sus clientes son hombres o son mujeres, cuántos viven en la delegación Miguel Hidalgo, etc. El análisis global ofrece gráficas descriptivas.

A su vez, la información del análisis macro o global se subdivide en 3 tipos como se mencionó en el primer capítulo:

1. **Análisis descriptivo**
2. **Análisis inferencial o predictivo**
3. **Análisis analítico o prescriptivo**

En este punto, el área directiva selecciona el tipo de análisis que desea llevar a cabo ya sea micro o macro.

Por ejemplo, si una institución de salud tiene muchas quejas de pacientes, entonces se requiere mejorar el proceso de atención médica. En este caso se realiza un análisis micro para conocer la trazabilidad del paciente, de manera que se tiene el conocimiento integral del historial del paciente para entender la razón de ser de su queja.

Otro caso es el de una empresa que desea mejorar el nivel de ventas de un producto e nivel nacional. Se recomienda el uso de aprendizaje automatizado es decir que primero se realice un análisis descriptivo para conocer cuántos productos se venden en cada estado. Posteriormente, se aconseja la realización del análisis inferencial para realizar proyecciones de ventas.

En el supuesto que una empresa no conozca las razones de su carencia en ganancias, se recomienda la realización de un análisis analítico para que los datos se integren en la *Big Data Box* y se encuentren las relaciones entre las diferentes variables para encontrar patrones de comportamiento ocultos.

FASE X: Visualización de los resultados

Una vez que se tiene el modelo, se procede a la visualización del modelo. La visualización es una representación digital del modelo para la toma de decisiones.

En la representación de los resultados, los tomadores de decisiones pueden estudiar en tiempo real lo que sucede con su negocio.

Para la visualización de los resultado se pueden utilizar herramientas comerciales como *Tableau* para analizar y descubrir información sobre los datos en tiempo real.

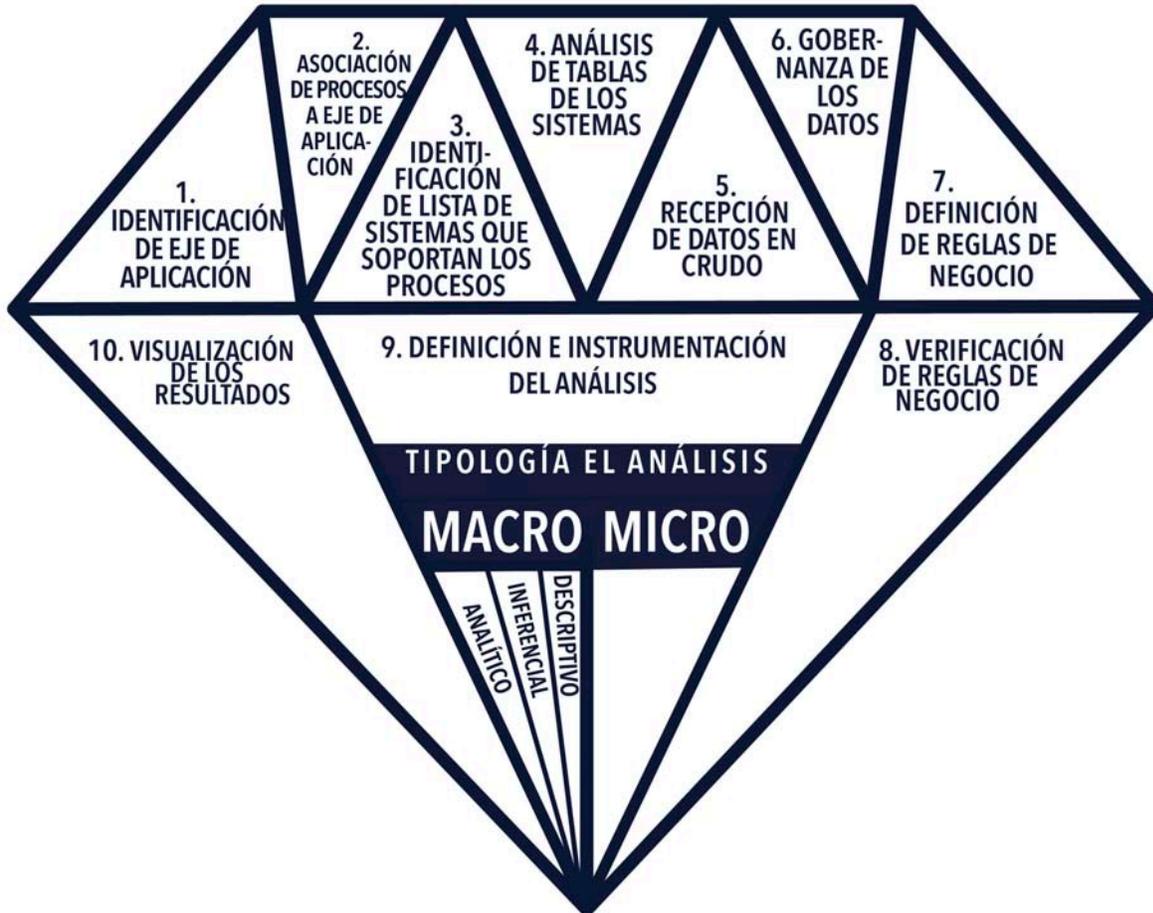
Tableau es una plataforma de análisis empresarial que es fácil de implementar y administrar garantizando el control y la seguridad de los datos. La información que refleja *Tableau* se muestra mediante tableros interactivos que ayudan a descubrir la información oculta al instante mediante el análisis visual en vivo que alimenta la exploración de datos ilimitada. (Tableau, 2016)

También es posible utilizar bibliotecas de programación en el programa R o Python para realizar las gráficas. Actualmente se cuenta con una excelente biblioteca para la generación de gráficas como Bokeh. “Bokeh es una biblioteca de visualización interactiva de Python que apunta a navegadores web modernos para su presentación. Su objetivo es proporcionar una construcción elegante y concisa de gráficos novedosos al estilo de D3.js, y ampliar esta capacidad con interactividad de alto rendimiento en conjuntos de datos muy grandes o en streaming. Bokeh puede ayudar a cualquiera que quiera crear rápida y fácilmente gráficos interactivos, cuadros de mando y aplicaciones de datos.” (BOKEH, 2015)

Los resultados se comunican a través de gráficas las cuales muestran los resultados y recomendaciones que provienen del modelo de Big Data para la toma de decisiones.

Al complementar todas las fases del modelo propuesto, queda una forma de diamante por lo que se nombra al modelo Big Data Diamond.

Figura 10 Big Data Diamond



Fuente: Elaboración propia basada en las 10 facetas de *data governance* (Informatica, 2017)

4.9.2 Comparativo del modelo propuesto respecto a los modelos existentes

El modelo propuesto se basa en los modelos creados por Bill Schmarzo y por Gabriel Guerrero para mejorar el proceso de toma de decisiones en las organizaciones con manejo de grandes volúmenes de datos. Dicho modelo ofrece un panorama con una visión administrativa de una herramienta de *Big Data Analytics* para la toma de decisiones puesto que considera los pasos a seguir de forma detallada. Asimismo, se estudia la empresa no sólo desde un punto de vista de un objetivo o meta específica que lograr sino que, el modelo propuesto pretende conocer a la organización como un todo para poder ofrecer, según los procesos clave de la organización un análisis micro y/o macro de la información. El cuadro a continuación resume el comparativo entre el modelo propuesto y los modelos de Schmarzo y Guerrero.

Tabla 8 Cuadro comparativo

	FASE	MODELO PROPUESTO		Marco de trabajo, SCHMARZO		Modelo de datos WebGestiones, GUERRERO		
		DIMENSIONES	DESCRIPCIÓN	DIMENSIONES	DESCRIPCIÓN	DIMENSIONES	DESCRIPCIÓN	
¿QUÉ?	1	Identificar el eje de aplicación en la empresa	Conocer los procesos clave de la organización	ESTRATEGIA EMPRESARIAL	¿Qué objetivo desea cumplir la organización? • Descubrir nuevas posibilidades de obtener beneficios que puedan repercutir en su marketing y su organización de ventas. • Reducir los costes de adquisición, fabricación, inventario, cadena de suministro, distribución, publicidad, ventas y mantenimiento. • Atenuar el riesgo en todos los aspectos operativos y financieros de la cadena de valores de la organización.	ESTABLECER METAS	1. Asimilar la misión empresarial 2. Interpretar cada uno de los objetivos empresariales 3. Determinar procesos productivos	
	2	Ubicar los procesos asociados al eje de aplicación	Identificar procesos y asociarlos con ejes de aplicación	INICIATIVAS EMPRESARIALES				
¿CÓMO?	3	Identificar la lista de sistemas de soportan el proceso	Identificar la lista de sistemas que soportan el proceso a mejorar	FUENTES DE DATOS	• Conseguir nuevos clientes, productos y conocimientos operativos que puedan utilizar para lograr una ventaja competitiva frente a la competencia y sacar un mayor beneficio en el sector	DEFINIR MÉTRICAS	4. Conocer sistemas que integran los procesos 5. Determinar el conjunto de datos que genera cada sistema	
	4	Analizar la lista de tablas de datos generados para cada sistema	Identificar las tablas de datos creadas en cada sistema					
	5	Recepción de datos en crudo (raw)	Recabar los datos de cada una de las tablas	TAREAS		6. Definir resultados que desean 8. Determinar fuentes relevantes 7. Definir métricas		
	6	Gobernanza de los datos	Tratar datos en Big data Box					
	7	Definición de las reglas de negocio	Definir métricas o reglas de negocio	RESULTADO Y FACTORES CRÍTICOS		AUDITAR	9. Definir el mecanismo de construcción de la métrica	
	8	Verificación de reglas de negocio	Validar reglas de negocio					
	9	Definir tipología del análisis de los datos	Definir tipo de análisis: macro (descriptivo, inferencial o analítico) o micro			Aplicar factores motrices	MEDIR RESULTADOS	10. Determinar visualización
	10	Visualización de los resultados	Establecer forma de visualización de la información			Visualización resultados	TOMAR DECISIONES	11. Desarrollar informes 12. Establecer aspecto financiero

Fuente: Elaboración propia

4.10 Validación del modelo propuesto

4.10.2 Instrumento de medición

Para la validación del modelo se toma en cuenta la entrevista cualitativa como un instrumento de medición. En efecto, según Bolsegui y Fuguet, “la Entrevista Cualitativa se define como una conversación de carácter profesional guiada por un tema central. En esta conversación el entrevistador supone que el entrevistado elabora significados en relación con otras personas, por cuanto el sujeto entrevistado pertenece a distintas redes donde se habla del tema abordado.” (BOLSEGUÍ & FUGUET SMITH, 2006)

La entrevista cualitativa se presenta a expertos del tema de Big Data (entrevistados) con la finalidad de medir, mejorar y enriquecer el modelo propuesto. En este marco, se extendió una invitación por correo electrónico a 10 expertos Big Data donde se presentó el resumen general del modelo (contenido en el apéndice) junto con un guión de 11 preguntas que puedan provocar tanto respuestas de grandes narrativas como otras que solo indican sí o no. Este tipo de entrevista “no descansa sobre el control riguroso del método, puesto que no se considera como un procedimiento o un momento sino como un proceso, donde tanto el entrevistador como el entrevistado cambian y queda el conocimiento.” (BOLSEGUÍ & FUGUET SMITH, 2006)

Con la finalidad de lograr un acompañamiento adecuado con los entrevistados, una vez enviada la información se acordó una entrevista con cada uno de los expertos para conocer su retroalimentación. La diversidad de giros en los que se desempeñan los expertos ofrece a la entrevista cualitativa una visión flexible y dinámica que enriquece la validación del modelo propuesto.

4.10.3 Selección de los informantes

En esta investigación, la selección de informantes se realizó minuciosamente con la finalidad de extraer conclusiones trascendentes a partir de las entrevistas cualitativas. Para realizar las entrevistas individuales se diseñó un guión (apéndice) que fue enviado a los informantes (expertos) y para su posterior análisis y entrevista. El método de recolección de información utilizado fue el de la entrevista a expertos del tema de Big Data. Efectivamente, se invitó a 10 expertos en Big Data que tuvieran más de 5 años de experiencia en el manejo de grandes volúmenes de datos, que sean tomadores de decisiones y que desarrollen sus labores en la Ciudad de México. Siete de los diez entrevistados accedieron a dar su punto de vista sobre el modelo propuesto. Los tres restantes no fue posible localizarlos ni dieron respuesta por ningún medio. A continuación se presenta el perfil de los entrevistados.

Tabla 9 Selección de informantes

Expertos	Puesto	Giro de la empresa	Años de experiencia
Experto 1	Director de Tecnología	Desarrollo de soluciones Big Data	5
Experto 2	Director de Innovación y Desarrollo	Consultoría	6
Experto 3	Responsable de software	Consultoría de proyectos	12
Experto 4	Investigador	Educación	35
Experto 5	Director General	Franquicias	16
Experto 6	Director Comercial	Consultoría de proyectos	11
Experto 7	Desarrollador	Consultoría de TIC's	7

Fuente: Elaboración propia

4.10.4 Resultados de las entrevistas

Una vez concluido el proceso de entrevistas, se procede a la recolección, construcción y análisis de los resultados que derivan en las siguientes aseveraciones:

- De los 10 entrevistados, 3 no contestaron favorablemente a la invitación.
- Los 7 informantes indicaron que les fue fácil entender el modelo propuesto y no tuvieron ninguna duda respecto a los conceptos.
- Las fortalezas del modelo propuesto que identificaron los expertos son:
 - Es un modelo integral y comprehensivo además de ser simple.
 - Es una propuesta interesante el conocer la forma en que puede ser aplicable a cualquier organización con grandes volúmenes de datos y en beneficio de la toma de decisiones.
 - El modelo permite un excelente entendimiento del proceso de trabajo con datos.
 - El modelo propuesto es un marco teórico bien construido y acorde con las tendencias actuales.
 - El modelo que se propone posee una excelente adaptación a una lógica de negocio.
 - La posibilidad de integrar muchos sistemas y manejarlos como una sola entidad es un aspecto atractivo del modelo.
 - La propuesta facilita la lectura e interpretación de los grandes volúmenes de datos.
- En cuanto a las debilidades, los informaron destacaron las siguientes áreas de oportunidad:
 - Se percibe un modelo abstracto.
 - Es necesario considerar que no todas las empresas tienen estructurada su información minuciosamente como para identificar qué sistemas soportan qué procesos.
 - A pesar de que el modelo está muy bien construido se considera importante aplicar en un caso práctico o detallar las fases a través del estudio de un caso de éxito en el manejo de grandes volúmenes de datos para la generación de valor.

- Seis informantes indicaron que el modelo se considera aplicable a cualquier organización que maneje grandes volúmenes de datos. No obstante, un experto indicó que no todas las empresas que generan grandes volúmenes de datos pueden identificar fácilmente los procesos asociados al eje de aplicación.
- Los informantes que trabajan en un entorno empresarial afirmaron que el modelo propuesto es aplicable a su ámbito laboral.
- Las mejoras que los informantes proponen para el modelo son las siguientes:
 - Desarrollar un ejemplo (ficticio o real) de su aplicación.
 - Entre las fases V (raw data) y la VII (gobernanza de los datos) incluir un análisis preliminar de los datos principalmente para identificar de manera temprana alguna inconsistencia en los mismos.
 - Reforzar la sección de analítica con algunos ejemplos de aplicación de los tres tipos de analíticos: los descriptivos, los inferenciales y los predictivos.
- De forma unánime, los informantes contestaron que el modelo favorece la toma de decisiones.

4.11 Ajustes al modelo

A partir de los aportes de los informantes, se realizan los siguientes ajustes para enriquecer el modelo propuesto:

- El planteamiento general del modelo propuesto se considera que está bien explicado, no obstante, se considera abstracto por lo que se recomienda incluir aspectos administrativos.
- Se debe detallar el análisis preliminar de los requerimientos de la organización.

- Se recomienda incluir una figura que refleje el ciclo de vida general del modelo. Dicha figura debe incluir el diagrama titulado “Fuentes generadoras de grandes volúmenes de datos”
- Es necesario agrupar las fases para una mayor comprensión de la relación entre cada una.
- Se recomienda agrupar las fases para evitar redundancias y favorecer el flujo entre cada etapa del modelo.

El modelo enriquecido con las opiniones de los expertos se muestra en el siguiente capítulo.

Capítulo 5 Modelo Final

"El cerebro nunca debe jubilarse, sino trabajar noche y día, porque a cierta edad - como la mía- ya no es necesario dormir, es una pérdida de tiempo."

Rita Levi-Montalcini

Hoy, los datos abundan y necesitan ser analizados adecuadamente para evitar eliminar información que puede ser útil para la empresa. Las organizaciones generan una gran cantidad de información que requiere ser analizada con herramientas no tradicionales para la toma de decisiones. Desafortunadamente, la capacidad de convertir los datos en información confiable y útil para la empresa no es una tarea fácil. En siglo XXI, las organizaciones se han tenido que enfrentar al exceso de datos que generan grandes desafíos. La velocidad con que se generan los datos requiere de un modelo que pueda responder a las necesidades actuales con la misma rapidez. Frecuentemente, al desear incluir un nuevo modelo de operación en la organización, éste toma mucho tiempo en implementarse por lo que es altamente probable que al estar operando ya no solucione el problema inicial.

Asimismo, muchas empresas desean integrar modelos para la toma de decisiones, no obstante, si los datos carecen de valor, de una métrica adecuada y de un repositorio central de datos, la información extraída del modelo será inconsistente y arrojará resultados inconsistentes e inexactos lo que conllevará a una mala toma de decisiones.

Cada organización tiene sus propios modelos y formas de validarlo que generalmente sólo unos cuantos científicos de datos son capaces de implementarlo. Por lo tanto, la pérdida de conocimiento importante en los modelos creados individualmente es elevada dado que "cuando la persona deja la empresa, el conocimiento se va con ella." (SAS México, 2015)

Las organizaciones que han implementado modelos analíticos han conocido una mejora en la toma de decisiones además que han mejorado sustancialmente la

relación con los clientes, los procesos y han aumentado sus ingresos. Sin embargo, aún existen múltiples organizaciones que desconocen cómo implementar un modelo de *Big Data Analytics*.

En los capítulos anteriores, se muestra la importancia del uso de *Big Data Analytics* en las organizaciones que generan grandes cantidades de información. El modelo final consta de cuatro dimensiones: recopilación de datos, preparación de la Big Data Box, evaluación de los resultados y toma de decisiones. Es importante subrayar que tanto el modelo de Schmarzo como el de Guerrero no integran los cimientos del pensamiento administrativo que consiste en aplicar el proceso administrativo integrando las fases de planeación, organización, dirección y control. Por lo tanto, las cuatro dimensiones del modelo se integran de 9 fases cuya integración conforman el proceso administrativo para una óptima toma de decisiones.

Para el modelo final, se retoman las cuatro etapas del proceso administrativo de manera que las 9 fases del modelo de *Big Data Analytics* se integran a dicho modelo.

La etapa de planeación del modelo final incluye las tres primeras fases que son: identificar las necesidades de la empresa, ubicar los procesos clave y definir las métricas de calidad. En efecto, se requiere conocer ampliamente la empresa para ubicar sus necesidades, establecer los objetivos y definir las reglas o métricas para la implementación de un modelo de *Big Data Analytics* una óptima toma de decisiones. Una correcta planeación determina el éxito del modelo puesto que tiene un enfoque hacia el futuro de mejoramiento y progreso. Asimismo, esta etapa permite a los directivos evaluar alternativas y tomar la mejor decisión. Una vez establecida la planeación, se procede al proceso de organización.

En el modelo propuesto la organización está relacionada con la clasificación de los datos para el logro de los objetivos establecidos en el proceso de planeación. En

este marco, la etapa de organización está compuesta de 2 fases respecto a la recolección y preparación que la *Big Data Box* que implica el análisis, la recepción de datos en crudo, la gobernanza y la estructuración de los datos.

Efectivamente, la etapa de organización, en términos del modelo final, está enfocado en la estructura, la limpieza y la preparación de la *Big data Box*; más adelante se detalla cada una de las fases integradas en el proceso de organización. Esta etapa es importante puesto que es el medio para lograr lo planeado y determinar la estructura para que las fases de dirección y control tengan bases confiables establecidas con orden y organización. Una vez cubierta la etapa de organización se procede al proceso de dirección.

El proceso de dirección esta integrado por dos fases: la elaboración del análisis global de los datos y la visualización de los resultados. Al identificar la tipología del análisis y posteriormente la visualización de los resultados, los tomadores de decisiones podrán lograr los objetivos planificados siempre que el proceso de dirección se realice mediante liderazgo, comunicación, motivación, desarrollo, integración de los equipos y capacitación. Ahora bien, para cerciorarse que los resultados arrojados sean los esperados es necesario cubrir el proceso de control.

En el modelo, el proceso de control se integra de la fase de evaluación de los resultados donde instrumentos de control se identifican posibles fallas o el éxito del cumplimiento del objetivo empresarial. De esta forma se procede a la última fase que está relacionada con la toma de decisiones.

La tabla a continuación facilita el entendimiento del proceso administrativo integrado del modelo final.

Tabla 10 Proceso Administrativo integrado al modelo final

FASE DEL PROCESO ADMINISTRATIVO	MODELO FINAL	
	Nº FASE	FASE
PLANEAR	1	Identificar las necesidades de la empresa
	2	Ubicar los procesos clave
	3	Definir las métricas de calidad
ORGANIZAR	4	Recabar los datos
	5	Preparar Big Data Box
DIRIGIR	6	Elaborar análisis global de los datos
	7	Visualizar los resultados
CONTROLAR	8	Evaluar los resultados
TOMAR DE DECISIONES	9	Tomar decisiones

Fuente: Elaboración propia.

Fase I: Conocer las necesidades de la organización

En un primer punto, Schmarzo indica la importancia de definir claramente el ámbito en el que se centrará la iniciativa *Big Data* para lo cual se debe identificar el objetivo empresarial global. Es decir, como indica Dr. Gabriel Guerrero, en una primera fase es importante asimilar la misión empresarial, conocer cuál es la misión y visión de la empresa. El segundo paso del modelo *WebGestiones* destaca la necesidad de interpretar cada uno de los objetivos empresariales. Estos procesos de conocer y asimilar la misión empresarial y de interpretar cada uno de los objetivos empresariales, equivalen a los dos primeros pasos del modelo de Guerrero. Dichos pasos se retoman para el modelo propuesto y se integran en una sola fase correspondiente a la primera.

En este primer punto es imprescindible identificar la misión y visión de la empresa. Asimismo, se debe estudiar el modo de operación de la organización para identificar las áreas de la empresa con un rendimiento inferior para conocer cuáles requieren mayor atención. En este aspecto es primordial establecer un objetivo empresarial el cuál determina el fin por el que se busca implementar un modelo de *Big Data Analytics* en la organización. Para establecer dicho objetivo se requiere

tomar en cuenta los principales elementos de la planeación indicados en el primer capítulo que se refieren a la misión, visión, valores y objetivos. En este aspecto, como lo define Luna González, es importante considerar los ¿qué?, ¿qué queremos?, ¿qué deseamos?, y ¿qué vamos a hacer?

En esta primera fase se da respuesta a las siguientes interrogantes:

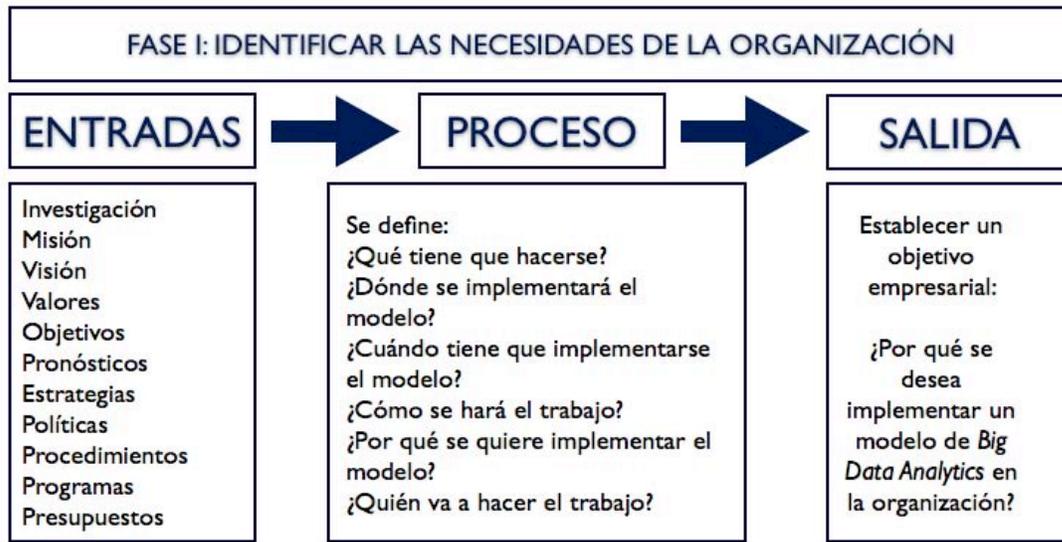
- ¿Qué tiene que hacerse?
- ¿Dónde se implementará el modelo?
- ¿Cuándo tiene que implementarse el modelo?
- ¿Cómo se hará el trabajo?
- ¿Por qué se quiere implementar el modelo?
- ¿Quién va a hacer el trabajo?

Para dar respuesta a estas preguntas se lleva a cabo el proceso de planeación donde se aplican los elementos citados que el primer capítulo que son: investigación, visión, misión, valores, objetivos, pronósticos, estrategias, políticas, procedimientos, programas y presupuestos.

En esta fase se consideran las expectativas de los directivos en cuanto a los resultados esperados posteriores a la aplicación de un modelos de *Big Data Analytics* en la organización. Esta fase pretende conocer cuál es la situación actual de la empresa, hacia dónde desea llegar, identificar los problemas, las oportunidades y los procesos que requieren mayor atención.

La siguiente figura resume las entradas y las salidas de esta primera fase.

Figura 12 FASE I: entradas y salidas



Fuente: Elaboración propia

FASE II: Ubicar los procesos clave

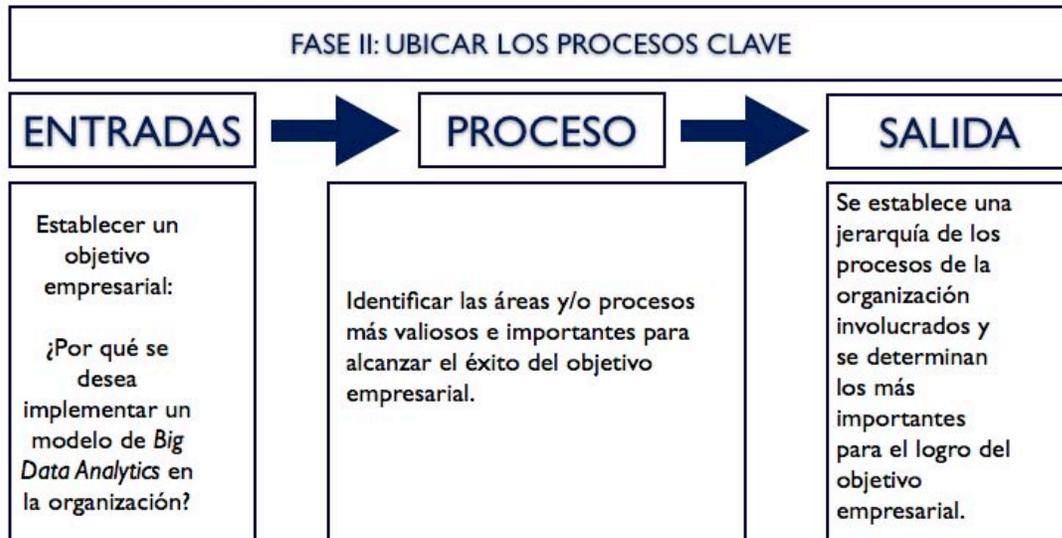
Para esta segunda fase se retoma el tercer paso del modelo WebGestiones que indica la importancia de determinar los procesos productivos. Guerrero indica que para lograr los objetivos empresariales se requiere determinar la secuencia de operaciones que genera la organización para el logro de sus objetivos. En esta segunda fase se retoma el tercer punto del modelo mexicano de Dr. Guerrero, así como el aspecto de tareas del modelo de Schmarzo.

En esta fase se busca conocer cuales son las áreas clave de la organización basados en las necesidades de la organización. Es decir que una vez que se establece el objetivo empresarial y se determina la razón por la cual la organización desea implementar un modelo de *Big Data Analytics*, se procede a conocer las áreas clave para el logro del objetivo empresarial. Se establece una jerarquía de los procesos de la organización involucrados y se determinan los más importantes para el logro del objetivo empresarial. En este aspecto se debe hacer un análisis del modelo empresarial actual de la organización y determinar cual es

el área y/o proceso más valioso e importante para alcanzar el éxito del objetivo empresarial.

La siguiente figura resume las entradas y las salidas de esta segunda fase.

Figura 13 FASE II: entradas y salidas



Fuente: Elaboración propia

FASE III: Definir las métricas de calidad

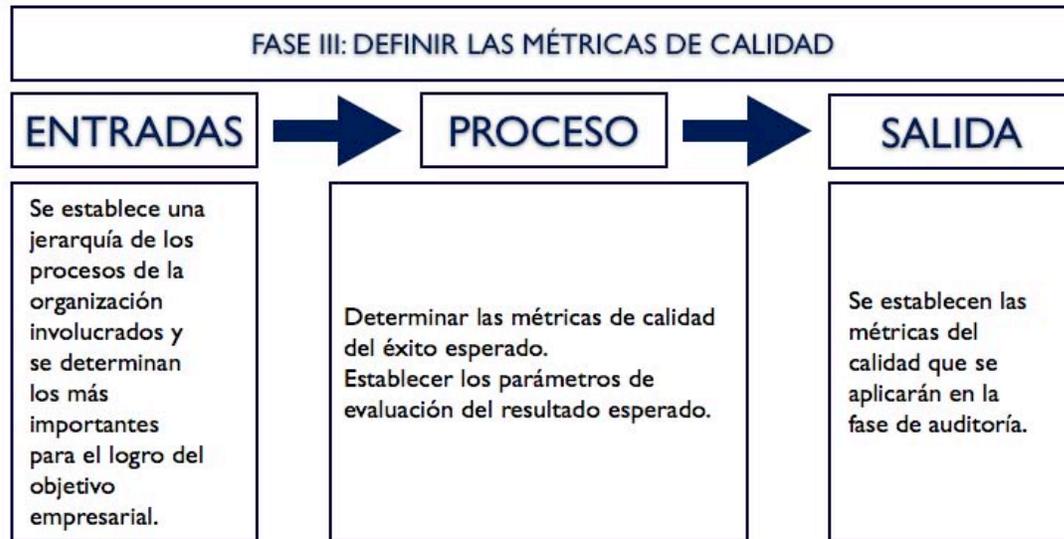
En esta fase se definen los factores críticos de éxito necesarios para llevar a cabo con éxito las iniciativas empresariales es lo que Schmarzo denomina como resultados y factores críticos. Schmarzo define que los factores críticos de éxito como las reglas que se deben llevar a cabo para que la iniciativa empresarial tenga éxito. El modelo WebGestiones determina esta fase como aquella donde se deben definir los resultados que desean medirse, determinar las métricas definir el mecanismo de construcción de la métrica.

En este punto, se definen las reglas de negocio o métricas de calidad donde se determinan los parámetros de evaluación del resultado deseado. En esta fase

únicamente se definen las métricas de calidad puesto que su aplicación será posterior.

La figura a continuación resume las entradas y las salidas de esta tercera fase.

Figura 14 FASE III: entradas y salidas



Fuente: Elaboración propia

FASE IV: Recabar los datos

Para esta cuarta fase se retoma lo que Schmarzo subraya como fuentes de datos. Es decir que se destacan las fuentes de datos clave necesarias para sustentar la estrategia empresarial y las iniciativas empresariales clave de apoyo. Asimismo se retoman los puntos cuarto y ocho del modelo de Guerrero donde indica la necesidad de conocer los sistemas que integran los procesos y determinar las fuentes de datos relevantes.

Cada proceso tiene una serie de sistemas asociados al mismo. En esta etapa se requiere identificar la lista de sistemas que soportan el proceso a mejorar. Esta fase marca el inicio del proceso de organización. Para determinar los datos que se deben recabar se debe analizar los datos que están involucrados en los procesos

clave. Una vez determinados cuales datos están relacionados con dichos procesos, se deben estudiar aquellos datos que estén involucrados en el logro del objetivo empresarial establecido en la primera fase.

En este punto se identifica el conjunto de datos que están involucrados en los procesos clave y por ende que están relacionados con el objetivo final. En este aspecto se recaban los datos tal cual como los tiene la organización, sin ningún tipo de procesamiento. Esta forma de datos se les conoce como datos crudos (*raw data*) o datos originales (*as is data*).

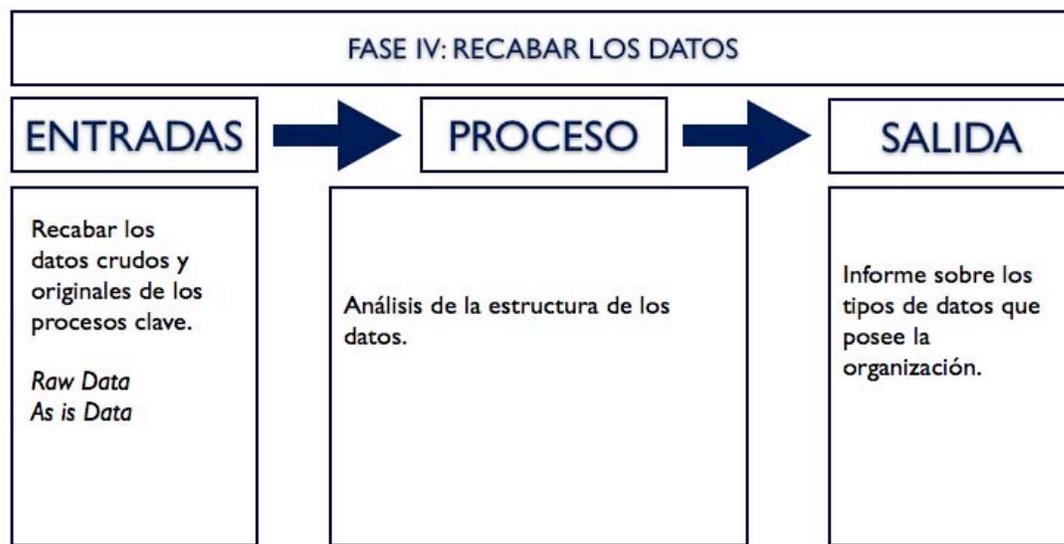
A partir de esta fase se preparan los datos es decir, como indica Guerrero en el quinto punto de su modelo, se determina el conjunto de datos de cada sistema.

Los datos de cada proceso tienen sus propias características por lo que se analiza la información para determinar que información se puede determinar con el análisis de dichos datos de forma que se puedan estructurar para una mejor comprensión de la información que posee la empresa.

Esta fase se centra en estudiar qué tipo de datos se tienen y cómo se podrían catalogar. Se realiza un informe donde se indica los tipos de datos que posee la organización

La figura a continuación resume las entradas y las salidas de la cuarta fase.

Figura 15 FASE IV: entradas y salidas



Fuente: Elaboración propia

FASE V: Preparar Big Data Box

Esta fase corresponde al proceso de preparación de los datos que Schmarzo indica como fuentes de datos. En este punto se retoman los cuatro factores motrices del valor de Big Data para catalogar los datos.

El tratamiento de los datos mejor conocido como gobernanza de los datos (Data Governance) en términos informáticos, es el paso que prosigue al de recepción de datos en crudo. Se realiza un proceso de gobernanza de los datos, a través de un análisis de estructura y calidad de la información.

Para asegurarse que los datos cumplen con las métricas de calidad de la tercera fase, se realiza una evaluación. En esta fase se hace uso de herramientas como Berkeley DB, Hadoop HDFS y Spark SQL para simplificar el análisis y proceso de grandes volúmenes de datos con tal de asegurar la calidad, el valor y la riqueza de los datos.

En este punto, el modelo propuesto se adentra al proceso de evaluación de los resultados. Retomando los seis pasos de la gobernanza de datos del modelo *WebGestiones*, esta fase corresponde al punto 5 que coincide con la manera de medir los resultados.

En esta fase se verifica que los datos cumplan con las reglas de negocio establecidas.

Este análisis ofrece dos resultados:

1. Datos que sí cumplen con las reglas de negocio
2. Datos que no cumplen con las reglas de negocio

Después de estructurar los datos, los datos que no cumplen con las reglas de negocio se extraen de la *Big Data Box* y se devuelven al cliente para su análisis y corrección puesto que esta propuesta de modelo se maneja bajo la noción de GIGO. El GIGO es la abreviatura de *Garbasha In Garbasha Out* que significa que si entra basura sale basura. Es decir que el éxito de esta propuesta de modelo depende de la garantía de calidad de los datos que entran. La *Big Data Box* incluye los datos que fueron seleccionados minuciosamente considerando su calidad, fiabilidad y valor para poner en relieve su capacidad predictiva con la finalidad de asegurar un correcto desarrollo de un modelo analítico.

La siguiente figura resume las entradas y las salidas de la quinta fase.

Figura 16 FASE V: entradas y salidas



Fuente: Elaboración propia

FASE VI: Elaborar análisis global de los datos

En este punto se retoman los cuatro factores críticos de Schmarzo para poder catalogar los datos y definir el tipo de análisis que se puede realizar. Los datos analizados se integran a un almacén de datos llamado *Big Data Box* para su procesamiento y análisis global.

Una vez que está comprobada la calidad de los datos y que se extrajeron los datos que no cumplen con las reglas de negocio, la *Big Data Box* sólo debe incluir los datos con garantía de calidad. Dichos datos se fraccionan en dos categorías según su análisis: micro y macro. En este punto, el área directiva selecciona el tipo de análisis que desea llevar a cabo ya sea micro o macro.

En esta fase, el modelo se adentra al proceso de dirección puesto que se afirma la dirección para alcanzar el objetivo empresarial. En esta fase se toman las decisiones para lograr lo que la organización desea en un futuro.

Si la organización desea realizar un análisis global, se enfocara en el aspecto macro del análisis de los datos. Posteriormente, con dicho análisis se podrán

hacer estudios descriptivos, predictivos y analíticos. De este análisis global surgen las directrices que seguirá la organización para una óptima toma de decisiones puesto que los resultados arrojados por el Big Data Box ofrecerán información valiosa y mostrarán posibles correlaciones ocultas y patrones de comportamiento.

La siguiente figura resume las entradas y las salidas de la sexta fase.

Figura 17 FASE VI: entradas y salidas



Fuente: Elaboración propia

FASE VII: Visualizar de los resultados

En esta última fase se retoma los puntos 10 y 11 del modelo *WebGestiones* donde se establece la herramienta preferida para la visualización de los resultados. Asimismo, se auditan los resultados para verificar que se cumpla con la estrategia empresarial establecida en la primera fase del modelo propuesto.

Una vez que se tiene el modelo, se procede a la visualización del modelo. La visualización es una representación digital del modelo para la toma de decisiones. En la representación de los resultados, los tomadores de decisiones pueden estudiar en tiempo real lo que sucede con su negocio.

Para la visualización de los resultado se pueden utilizar herramientas comerciales para analizar y descubrir información sobre los datos en tiempo real. Las herramientas de visualización son plataformas de análisis empresarial que son fáciles de implementar y administrar garantizando el control y la seguridad de los datos. La información que reflejan las herramientas de visualización se muestra mediante tableros interactivos que ayudan a descubrir la información oculta al instante mediante el análisis visual en vivo que alimenta la exploración de datos ilimitada.

También es posible utilizar herramientas de programación para realizar las gráficas, cuadros de mando y aplicaciones de datos de forma rápida que se tenga una construcción elegante y concisa de gráficos novedosos y así ampliar la capacidad de interactividad de alto rendimiento en conjuntos de datos muy grandes o en tiempo real (*streaming*).

Los resultados se comunican a través de gráficas las cuales muestran los resultados y recomendaciones que provienen del modelo de *Big Data* para la toma de decisiones. La siguiente figura resume las entradas y las salidas de la séptima fase.

Figura 18 FASE VII: entradas y salidas



Fuente: Elaboración propia

FASE VIII: Evaluar los resultados

En esta fase se evalúa si los resultados arrojados en la fase anterior corresponden con las necesidades de la organización expresadas en la primera fase. Si después de la visualización de los resultados, los altos directivos consideran que el modelo sirve como una herramienta para cubrir las necesidades de la empresa expresadas en la primera fase, entonces el modelo será exitoso y se procede a la siguiente fase. En caso de que no se cumpla con el objetivo empresarial por el que se implementó el modelo de *Big Data Analytics*, se deben evaluar las fallas del modelo.

Entre los múltiples instrumentos de control que existen en la administración, se resaltan las auditorías administrativas que sirven para evaluar la eficacia y eficiencia de los procesos de planeación, organización, dirección y control inherentes en una organización. (LUNA GONZÁLEZ, 2015) Asimismo, para la medición del impacto de la implantación del modelo en la organización, se pueden aplicar instrumentos de control como las gráficas de Gantt, que es una “técnica de planeación y control desarrollada por Henry L. Gantt, mediante una gráfica de barras, los requisitos de tiempo para diversas tareas o acontecimientos de una producción o algún programa”. (LUNA GONZÁLEZ, 2015) Asimismo, se puede determinar si se cumplió con el objetivo empresarial indicado en la primera fase a través de cuadros de mando como el *Balanced Score Card* para estimar el impacto del modelo con el objetivo establecido.

Para establecer el control del impacto del modelo, se establecen formatos donde se indican las metas cuantificadas alcanzadas, si hay una variación entre los resultados, esperados, estimados y reales, las variaciones de los resultados y la justificación de dichas variaciones.

Figura 19 FASE VIII: entradas y salidas



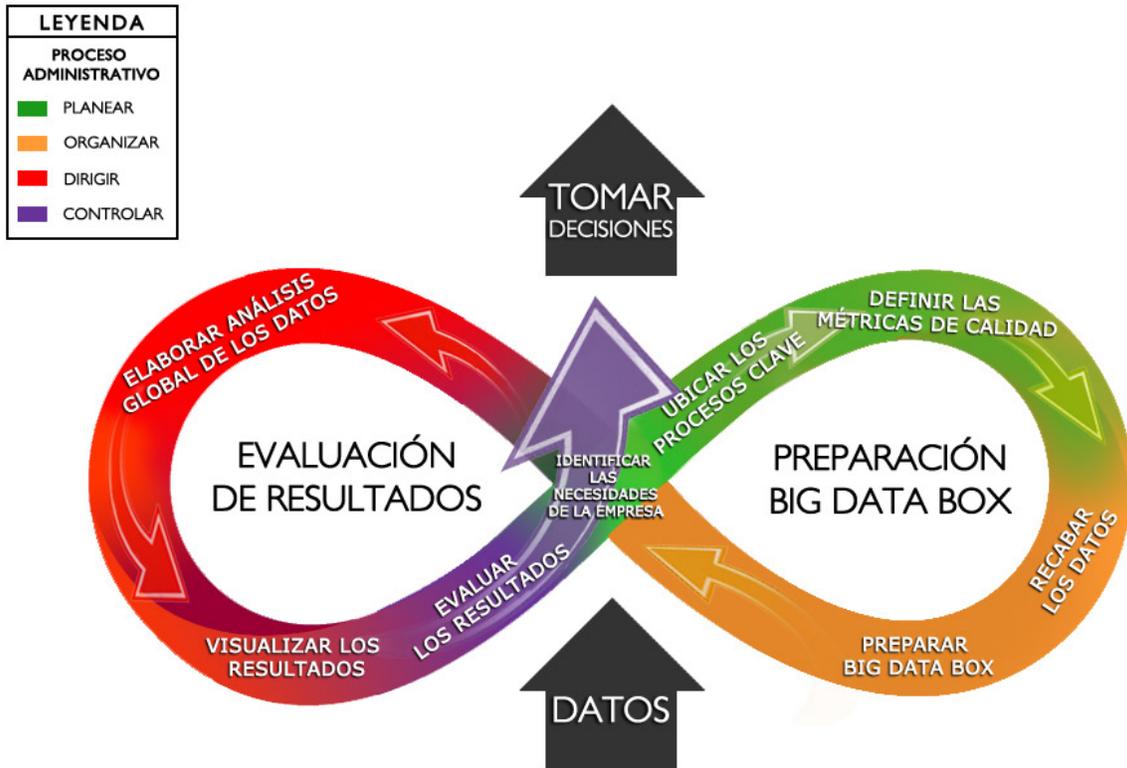
Fuente: Elaboración propia

FASE IX: Tomar decisiones

Esta última fase del modelo involucra a los altos directivos de las organizaciones que, una vez concluidas las fases anteriores, proceden a la toma de decisiones.

El modelo final se muestran en la figura a continuación. Se eliminó el *Big Data Diamond* puesto que de acuerdo a la retroalimentación de los expertos y del tutor de esta tesis, se recomienda el uso de la figura de infinito para una mejor visualización del flujo y relación entre las múltiples fases.

Figura 20 Modelo de *Big Data Analytics* para la toma de decisiones



Fuente: Elaboración propia basado en el White paper: Gestionando el Ciclo de vida analítico para las Decisiones a Escala (SAS México, 2015)

Conclusiones

El creciente papel que juegan las nuevas tecnologías a nivel global, ha acelerado la introducción de nuevas herramientas sofisticadas que deben ser integradas a las estrategias generales de las organizaciones. El uso de Big Data en las organizaciones no debe considerarse como el instrumento que resolverá todas las deficiencias y dificultades, sino como una herramienta que generará una transformación empresarial en la toma de decisiones. La velocidad de la era del conocimiento y el cúmulo de información con que se cuenta, imponen una modificación de los modelos de negocio de las organizaciones, para sobrevivir ante la avalancha de nuevas tecnologías e información que utilizan los competidores.

Actualmente, uno de los grandes retos que se enfrentan los tomadores de decisiones es el identificar como ingresar a la ola de nuevas tecnologías, para no quedar rezagados y estar a la vanguardia. Las organizaciones requieren conocer los grandes beneficios de pasar de un estudio de los datos de forma descriptiva a un estudio predictivo y prescriptivo, para descubrir así correlaciones ocultas, nuevos patrones de comportamiento de los consumidores y ofrecer recomendaciones a los clientes. De esta forma la alta dirección pasará de conocer su clientela únicamente centrada en el “cuántos” (¿Cuántos hombres?, ¿Cuántas mujeres?, etc.) a adelantarse a las futuras necesidades no sólo de sus clientes actuales, sino también de nuevos consumidores. Integrar nuevas tecnologías a la organización se ha convertido en una necesidad ante el reto de la competencia global.

Se afirma que la información es poder, pero es inservible tener un exceso de datos si no se define un objetivo sobre el uso que se le quiere dar a dichos datos. Por lo que la respuesta a la pregunta general de investigación ¿Cómo una organización que genera grandes volúmenes de datos a través de *Big Data Analytics* puede mejorar la toma de decisiones? Tenemos la siguiente respuesta: una organización

que genera grandes volúmenes de datos puede mejorar la toma de decisiones mediante la implementación de un modelo de Analíticos de Grandes Volúmenes de Datos considerando que para que dicho modelo sea exitoso, se debe tener previamente definido el objetivo del uso de dicha herramienta.

En esta investigación se plantean diversas interrogantes relacionadas con la definición de Big Data. En este aspecto, Big Data se define como una herramienta que permite velozmente, capturar, almacenar, gestionar y analizar un conjuntos de datos de alto volumen, variabilidad, veracidad, visibilidad, valor y variedad cuyo tamaño va más allá de la capacidad de software de bases de datos relacionales para extraer nuevos conocimientos, crear nuevas formas de valor y exigir formas innovadoras y rentables de procesamiento de la información para mejorar la comprensión y la toma de decisiones.

Otra pregunta que se analiza es respecto a saber si: *¿Big Data Analytics* sirve como un proceso de creación de valor para las organizaciones? Al respecto, se puede afirmar que *Big Data Analytics* sirve como un proceso de creación de valor puesto que, mediante el uso de dicha herramienta, las organizaciones son capaces de aprovechar las fuentes de datos para descubrir nuevos conocimientos.

Asimismo, en esta investigación se da respuesta a la interrogante: *¿Cuáles son las etapas que debe incluir el desarrollo de un modelo de Big Data Analytics para la toma de decisiones en las organizaciones con grandes volúmenes de datos?* En esta investigación se explica que un modelo de *Big Data Analytics* para la toma de decisiones en las organizaciones con grandes volúmenes de datos incluye 9 fases integradas al proceso administrativo. Las fases son: identificar las necesidades de la empresa, ubicar los procesos clave, definir las métricas de calidad, recabar los datos, preparar la Big Data Box , elaborar el análisis global de los datos, visualizar los resultados, evaluar los resultados y tomar decisiones.

En relación a la pregunta: ¿Cuáles son los límites de un modelo de *Big Data Analytics* para la toma de decisiones? *Big Data Analytics* es una herramienta para el análisis de datos. Es un apoyo en la toma de decisiones de las organizaciones, pero no puede reemplazar el juicio de los tomadores de decisiones. La analítica tiene limitantes, uno se puede perder en la marea de datos e ir demasiado lejos en aspectos cuantitativos. Es importante conocer las necesidades de los clientes y las áreas de oportunidad de las organizaciones para poder ver más allá de lo que reflejan los datos y las herramientas. Los datos deben ser complementados con las palabras y el raciocino de los tomadores de decisiones, para poder hacer una sinergia entre estos dos aspectos y sacar conclusiones más acertadas.

En cuanto a conocer: ¿Qué modelos existen del uso de *Big Data Analytics* en las organizaciones? Se puede afirmar que existen diversos modelos en el mercado de las Tecnologías de la Información, no obstante este estudio se centra en el marco de trabajo de los expertos Bill Schmarzo y en el modelo *WebGestiones* de Gabriel Guerrero.

Un buen uso de las nuevas tecnologías garantiza la supervivencia de la organización ante la feroz competencia nacional e internacional. Mediante la literatura analizada se comprueba la hipótesis de esta investigación. Efectivamente, el uso de un modelo de *Big Data Analytics* en las organizaciones con manejo de grandes volúmenes de datos mejora la toma de decisiones a nivel directivo. Las organizaciones que han sabido sacar provecho de las nuevas tecnologías, han conocido una mejora en los procesos de toma de decisiones y por consecuencia, tienen un incremento de productividad y una mejora en la explotación del conocimiento.

En esta investigación, se afirma que el modelo no sólo mejora los resultados empresariales, sino que es una ventaja competitiva para la organización puesto que con su implementación, la compañía tiene una característica diferencial respecto a sus competidores.

Los resultados de este estudio reflejan que Big Data es un tema de investigación muy novedoso y sobre todo en México. Las publicaciones científicas son mayormente estadounidenses y no han cesado de aumentar a partir de 2013. En este aspecto, en México, aún hay un largo camino que recorrer para no sólo aplicar Big Data en las organizaciones sino también publicar los resultados y extender los beneficios del uso de estas herramientas novedosas para la toma de decisiones.

Asimismo, esta investigación muestra que existen diversos modelos de Big Data en el mercado pero están enfocados principalmente en aspectos informáticos y científicos. Sin embargo, la aplicación de Big Data puede generar grandes beneficios para las Ciencias Sociales, por lo que si se integra una perspectiva administrativa al uso de los Analíticos de Grandes Volúmenes de Datos en ámbitos diversos como las Ciencias Sociales podría generarse grandes estudios de las transformaciones sociedad.

Se concluye que *Big Data Analytics* sirve como un proceso de creación de valor para las organizaciones puesto que se pueden tomar decisiones estratégicas para aprovechar al máximo el volumen y la variedad de datos. Esta toma de decisiones es precisa y facilita una rápida actuación. Integrar la tecnología a la organización perpetúa el liderazgo en innovación y por ende, permite una reinversión constante de una ventaja competitiva. Por lo tanto, incorporar un modelo de *Big Data Analytics* en la organización es una forma de producir información valiosa para la toma de decisiones. Así, a medida que incrementa la información, aumenta el conocimiento que se tiene de la organización y todos los actores que la rodean.

Los beneficios del uso de Big Data son diversos puesto que con esta herramienta se ha facilitado el procesamiento de los datos sin importar el tamaño ni la diversidad de éstos, obteniendo resultados detallados en tiempo real para una óptima toma de decisiones.

Es necesario que las organizaciones, sin importar su tamaño, estén a la vanguardia frente al uso de tecnologías para mejorar y facilitar la toma de decisiones, hacer más eficiente los procesos y crecer en su ámbito. En este marco, las instituciones públicas y los gobiernos han implementado políticas públicas mediante el análisis de los grandes volúmenes de datos generados por la ciudadanía. El uso de Big Data ha favorecido el estudio de datos abiertos para el análisis de indicadores de diversas índoles sociales, como puede ser la creación de ciudades inteligentes que garantizan un incremento de la calidad de vida de los ciudadanos mediante la aplicación de una infraestructura basada en las tecnologías de la información y comunicación para un desarrollo sostenible.

Con la finalidad de reafirmar el modelo propuesto sería pertinente aplicarlo a un ejemplo real en el marco de una investigación doctoral. De esta manera se podría demostrar con una aplicación real del modelo, que la recopilación, comparación, almacenamiento y la utilización de los datos en una sociedad cada vez más interconectada y desarrollada tecnológicamente, tiene grandes beneficios sociales, económicos, culturales, políticos y desde luego tecnológicos. Big Data puede emplearse no sólo en organismos con y sin fines de lucro, sino también se emplea para el desarrollo de políticas públicas.

La aplicación de Big Data, no sólo es el ámbito empresarial sino también en las Ciencias de la Salud y las Ciencias Sociales. Estos grandes avances tecnológicos prometen transformar la humanidad reduciendo el impacto ambiental. Efectivamente, son crecientes los proyectos que hacen uso de los datos abiertos para, mediante el análisis de las grandes cantidades de datos, apoyar a los ciudadanos y organizaciones civiles a analizar los problemas del país y desarrollar una cultura cívica, para lograr cambios en la sociedad.

“Las computadoras son increíblemente rápidas, precisas y estúpidas; los humanos son increíblemente lentos, imprecisos y brillantes; juntos son poderosos más allá de la imaginación”
Albert Einstein

Bibliografía

PÁGINAS WEB

- LÓPEZ-MARTÍNEZ, R. E. (15 de 08 de 2016). Conceptos teóricos sobre ICyVT. (N. Guerrero, Interviewer)
- CABAÑERO PISA, C., RODRÍGUEZ ARDURA, I., & SARRADELL LÓPEZ, E. (2002). *Toma de decisiones estratégicas*. Cataluña: Editorial UOC.
- LANEY, D. (2001). *Data Management: Controlling Data Volume, Velocity, and Variety*. Stamford: META Group Inc.
- LAZER, D., & KENNEDY, R. (10 de 01 de 2015). *WIRED*. Retrieved 20 de 03 de 2016 from WIRED: <http://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/>
- LFPDPPP. (05 de 07 de 2010). *CÁMARA DE DIPUTADOS DEL H. CONGRESO DE LA UNIÓN LFPDPPP*. Retrieved 12 de 04 de 2016 from CÁMARA DE DIPUTADOS DEL H. CONGRESO DE LA UNIÓN: <http://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>
- LIMA MACIEL, R. (2012). *Modelo de ingeniería de software con base en directrices de administración del conocimiento*. D.F.: UNAM. Colegio de México. (SR de SR de 2016). *Diccionario del Español de México*. Retrieved 20 de 08 de 2016 from Diccionario del Español de México: <http://dem.colmex.mx/>
- Communications of the ACM. (19 de 03 de 2016). *Communications of the ACM*. Retrieved 01 de 08 de 2016 from Communications of the ACM: <http://cacm.acm.org/magazines/2016/3/198858-peter-naur/fulltext>
- LOSHIN, D. (2013). *Big Data Analytics. From Strategic planning to enterprise integration with tools, techniques, noSQL, and graphs*. Waltman: MK Publications.
- COX, L. A. (2015). *Breakthroughs in Decision Science and Risk Analysis*. Denver: Wiley.
- CRISP-DM. (SR de SR de 1996). *CRISP-DM*. Retrieved 02 de 06 de 2016 from CRISP-DM: <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>
- LUNA GONZÁLEZ, A. C. (2015). *Proceso Administrativo*. Ciudad de México: Grupo Editorial Patria.
- ÁLVAREZ HERNANDO, J. (2011). *Guía Práctica sobre protección de datos*. Valladolid: Lex Nova.
- AGENCE TECHNIQUE DE L'INFORMATION SUR L'HOSPITALISATION ATIH. (SR de SR de 2013). *ATIH*. Retrieved 14 de 02 de 2016 from ATIH: www.atih.sante.fr/mco/presentation
- Agencia Andorrana de Protección de Datos. (SR de 02 de 2014). *Agencia Andorrana de Protección de Datos*. Retrieved 09 de 07 de 2016 from Agencia Andorrana de Protección de Datos: https://www.apda.ad/system/files/2014_2_resolution-big-data-es.pdf
- ANDA GUTIÉRREZ, C. (2004). *Administración y Calidad*. Ciudad de México: Editorial Limusa.

- ANDERSON, C. (23 de 06 de 2008). *WIRED*. Retrieved 01 de 06 de 2016 from WIRED: <http://www.wired.com/2008/06/pb-theory/>
- Big Data Social. (2016). *Big Data Social*. Retrieved 16 de 11 de 2016 from Big Data Social: <http://www.bigdata-social.com/analitica-big-data/>
- BOLSEGUÍ, M., & FUGUET SMITH, A. (06 de 2006). Construcción de un modelo conceptual a través de la investigación cualitativa. *Revista Universitaria de Investigación* , 207-229.
- BOKEH. (2015). *BOKEH*. Retrieved 13 de 12 de 2016 from BOKEH: <http://bokeh.pydata.org/en/latest/>
- Business Insider. (16 de 02 de 2012). The incredible story of how target exposed a teen girl's pregnancy .
- DATAPRIX. (27 de 05 de 2014). *DATAPRIX*. Retrieved 22 de 01 de 2017 from DATAPRIX: <http://www.dataprix.com/blog-it/big-data/big-data-gestion-datos-no-estructurados>
- DAVENPORT, T. (2014). *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*. Boston: Harvard Business School Publishing Corporation.
- DAVENPORT, T., & G.HARRIS, J. (2007). *Competing on Analytics: The New Science of Winning*. Boston: Harvard Business School Publishing Corporation.
- Diccionario LID de Empresa y Economía. (2016). *Diccionario LID de Empresa y Economía*. Retrieved 10 de 11 de 2016 from Diccionario LID de Empresa y Economía: <http://www.diclib.com/equipo%20f%C3%ADsico/show/es/alkonaeconomia/2597#ixzz4QOZtzB27>
- DIAMANDIS, P., & KOTLER, S. (2013). *Abundancia: El futuro es mejor de lo que piensas*. Barcelona: Antoni Bosh Editor.
- Diario Oficial de la Federación. (20 de 02 de 2015). *Diario Oficial de la Federación*. From Diario Oficial de la Federación: http://www.dof.gob.mx/nota_detalle.php?codigo=5382838&fecha=20/02/2015
- Dirección General de Información en Salud. (27 de 01 de 2015). *Dirección General de Información en Salud*. Retrieved 10 de 07 de 2016 from Dirección General de Información en Salud: http://www.dgis.salud.gob.mx/contenidos/basesdedatos/da_boletin.html
- EMC Computer Systems México. (SR de 04 de 2014). *México EMC*. Retrieved 02 de 08 de 2015 from Las oportunidades del Universo Digital: Los Rich Data y el creciente valor del Internet de las Cosas: <http://mexico.emc.com/leadership/digital-universe/2014iview/index.htm>
- EMC MEXICO. (1 de 04 de 2014). *IDC Analyze the future*. Retrieved 20 de 05 de 2016 from IDC Analyze the future: <http://mexico.emc.com/leadership/digital-universe/2014iview/index.htm>
- GALLEGO, J. C. (2010). *PCPI - Mantenimiento de Sistemas Microinformaticos*. España: Editex.
- GARCÍA ACOSTA, G. (2002). *La ergonomía desde la visión sistémica*. Bogotá: Universidad Nacional de Colombia.
- GARCÍA RAMOS, J. A., RAMOS GONZÁLEZ, C. D., & RUIZ GARZÓN, L. (2007). *Estadística empresarial* . Cádiz: Universidad de Cádiz.
- GARTNER Inc. (s/r de s/r de 2016). *Technology Research*. Retrieved 01 de 04 de 2016 from IT Glossary: <http://www.gartner.com/it-glossary/big-data/>

- GUTIÉRREZ, R. (14 de 01 de 2014). Big data: Oportunidades para empresas en 2015 . *Forbes México* , pp. 01-02.
- Harvard Business Review. (2013). *Big Data: The future of Information and Business* . Massachusets: Harvard Business Review Publishing.
- HOOD L., G. D. (12 de 12 de 2008). *Computing Community Consortium* . Retrieved 01 de 07 de 2016 from CCC-Led White Papers: <http://cra.org/ccc/resources/ccc-led-whitepapers/>
- HURWITZ, J., NUGENT, A., HALPER, F., & KAUFMAN, M. (2013). *Big Data*. New Jersey: John Wiley & Sons, Inc.
- INAI. (2015). *Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales*. Retrieved 01 de 02 de 2016 from Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales: <http://inicio.ifai.org.mx/SitePages/misionViosionObjetivos.aspx>
- Informatica. (01 de 01 de 2017). *Informatica.com*. From Informatica.com: <https://www.informatica.com/solutions/data-governance.html#fbid=tYISBoClaPA>
- JARAMILLO ANTILLÓN, J. (1998). *Principio de gerencia y administración de servicios médicos y hospitales*. San José: Editorial Nacional de Salud y Seguridad Social.
- JOYANES AGUILAR, L. (2013). *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. Ciudad de México: Alfaomega Grupo Editor.
- KOONTZ, H., & WEHRICH, H. (2004). *Administración: Una Perspectiva Global*. Ciudad de México: McGraw-Hill Interamericana.
- McKinsey Global Institute (MGI). (2012). *Big Data: The next frontier for innovation, competition, and productivity* . McKinsey & Company.
- MÜNCH, & MARTÍNEZ, G. (2012). *Fundamentos de la Administración*. Ciudad de México: Trillas.
- MANYIKA, J., CHUI, M., BROWN, B., BUGHIN, J., DOBBS, R., ROXBURGH, C., et al. (SR de 05 de 2011). *Mckinsey&Company*. Retrieved 01 de 08 de 2016 from Mckinsey&Company: <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>
- MARTÍNEZ, J. (23 de 0 de 2012). *Seis pasos para el gobierno de datos*. Retrieved 20 de 11 de 2016 from IBM: <http://www.ibm.com/developerworks/ssa/data/library/techarticle/gobierno-datos/>
- MAYER-SCHÖNBERGER, V., & CUKIER, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Nueva York: Houghton Mifflin Harcourt Publishing Company.
- MICROSOFT. (SR de SR de 2016). *MSDN Library*. Retrieved 01 de 08 de 2016 from MSDN Library: <https://msdn.microsoft.com/es-es/library/ms174949.aspx>
- NICOLETTI, B. (2016). *Digital Insurance: Business Innovation in the Post-Crisis Era*. New York: Palgrave Macmillan.
- NOTIMEX. (02 de 03 de 2014). Esperan inversiones de 215 mdd en tecnología big data en México. *El Financiero* , p. 01.
- Open Data Handbook. (SR de SR de SR). *Open Data Handbook*. Retrieved 20 de 07 de 2016 from Open Data Handbook: <http://opendatahandbook.org/guide/es/what-is-open-data/>
- Organización de las Naciones Unidas. (10 de 12 de 1948). *Declaración Universal de los Derechos Humanos*. Retrieved 22 de 06 de 2016 from Departamento de

Información Pública de las Naciones Unidas:
<http://www.un.org/es/documents/udhr/>
 Oxford University Press. (s/r de s/r de 2016). *Oxford English Dictionary*. Retrieved 05 de 04 de 2016 from Oxford English Dictionary:
<http://www.oed.com/view/Entry/18833#eid301162177>
 PÉREZ LÓPEZ, C., & SANTÍN GONZÁLEZ, D. (2007). *Minería de datos: técnicas y herramientas*. Madrid: THOMSON.
 PÉREZ, G. (2015). *Peligros del uso de los Big Data en la investigación en salud pública y en epidemiología*. Agencia de Salud pública de Barcelona. España: Elsevier.
 PÉREZ, J., & GARDEY, A. (SR de SR de 2008). *Definición.de*. Retrieved 20 de 09 de 2016 from Definición.de: <http://definicion.de/eficiencia/>
 PerTutatis. (20 de 01 de 2015). *PerTutatis*. Retrieved 01 de 02 de 2017 from PerTutatis: <http://pertutatis.cat/que-son-los-sistemas-de-apoyo-a-la-toma-de-decisiones-dds/>
 phpBB. (15 de 09 de 2008). *Administración de Operaciones*. Retrieved 01 de 02 de 2017 from Administración de Operaciones: <http://aspc-admon-op.forosactivos.net/t3-concepto-de-sistema-de-produccion>
 PRITCHARD, A. (1969). Statistical Bibliography or Bibliometrics? *Journal of documentation* , 348-349.
 RAE. (SR de SR de 2016). *Real Academia Española*. Retrieved 20 de 09 de 2016 from Real Academia Española: <http://dle.rae.es/?id=EPVwpUD>
 Real Academia Española. (n.d.). *Real Academia Española*. Retrieved 10 de 10 de 2016 from Real Academia Española: <http://www.rae.es>
 RISKIN, D. (1 de 10 de 2012). *Forbes*. Retrieved 13 de 02 de 2016 from Forbes: <http://www.forbes.com/sites/singularity/2012/10/01/the-next-revolution-in-healthcare/#62f448b46e1e>
 SCHMARZO, B. (2013). *Big Data: El poder de los datos*. Madrid: Anaya Multimedia.
 SÁNCHEZ DÍAZ, A. M. (12 de 06 de 2015). *Siete Días Médicos revista de atención primaria*. From Siete Días Médicos: <http://www.sietediasmedicos.com/zona-franca/bandeja-de-entrada/item/5333-hacia-un-nuevo-paradigma-en-medicina#.V1ZAICPhAb0>
 SÁNCHEZ PÉREZ, G., & ROJAS GONZÁLEZ, I. (2012). Leyes de protección de datos personales en el mundo y la protección de datos biométricos – Parte I. *Seguridad Cultura y prevención para TI* , 04-08.
 SANSONI, S. (06 de 10 de 2015). *SOLON*. Retrieved 15 de 02 de 2016 from SOLON: <http://www.linkedin.com/pulse/solon-la-plateforme-intelligente-dibm-renforce-lutte-contre-sansoni>
 SANZ, E. (2014). *¿Qué es el "Internet de las cosas"?* Retrieved 2016 from Muy Interesante: <http://www.muyinteresante.es/curiosidades/preguntas-respuestas/ique-es-el-qinternet-de-las-cosasq>
 SAS México. (2015). *Gestionando el ciclo de vida analítico para las decisiones a escala*. Ciudad de México: SAS Institute Inc. .
 SAS México. (2015). *Gestionando el Ciclo de Vida Analítico para las decisiones a escala*. Ciudad de México: SAS México.

- SAXSA. (2013). Big Data: Perspectivas y Alcances. *Big Data: Perspectivas y Alcances*. SR.
- Secretaría de Finanzas. (05 de 08 de 2016). *Secretaría de Finanzas*. Retrieved 05 de 08 de 2016 from Secretaría de Finanzas:
<http://www.finanzas.df.gob.mx/oip/arco/index.html>
- SHORTLIFFE, E., PERREAULT, L., WIEDERHOLD, G., & FAGAN, L. (2001). *Medical Informatics*. New York: Springer Science + Business Media.
- SOBRADILLO, P., POZO, F., & AGUSTI, Á. (04 de 09 de 2010). *Archivos de Bronconeumología*. From Archivos de Bronconeumología:
<http://www.archbronconeumol.org/index.php?p=watermark&idApp=UINPBA00003Z&piitem=S0300289610002887&origen=bronco&web=bronco&urlApp=http://www.archbronconeumol.org&estadotem=S300&idiomaltem=es>
- STANTON, J. (2012). *An Introduction to Data Science*. Syracuse: Syracuse University.
- Tableau. (2016). *Tableau*. Retrieved 12 de 11 de 2016 from Tableau:
<http://www.tableau.com/es-es>
- Teradata . (11 de 04 de 2013). *Teradata* . Retrieved 08 de 06 de 2016 from Teradata : <http://www.teradata.com/News-Releases/2013/BARC-Big-Data-Survey-Lack-of-Experts-and-Know-how-A-Main-Obstacle-to-Monetizing-Big-Data/>
- TRIPOD. (2016). *TRIPOD*. Retrieved 10 de 11 de 2016 from TRIPOD:
<http://mineriadetextos.tripod.com/>
- TUYA, J., RAMOS ROMÁN, I., & DOLADO COSÍN, J. (2007). *Técnicas cuantitativas para la gestión en la ingeniería de software*. Oleiros: Netbiblo.
- Universidad Nacional de Colombia. (18 de 11 de 2016). *Universidad Nacional de Colombia*. Retrieved 18 de 11 de 2016 from Sistema de Información de la Amazonia Colombiana: <http://www.unal.edu.co/siamac/sig/metadatos1.html>
- Universidad Politécnica de Cartagena. (2016). *Universidad Politécnica de Cartagena*. Retrieved 16 de 11 de 2016 from Universidad Politécnica de Cartagena: <https://www.upct.es/~gio/trazabilidad.htm>
- UnixTime.Info. (18 de 01 de 2011). *UnixTime.Info*. Retrieved 02 de 02 de 2017 from UnixTime.Info:
<http://web.archive.org/web/20110722004843/http://unixtime.info/>

TESIS

El impacto del cambio tecnológico en las organizaciones y su administración “La emergencia de las organizaciones virtuales y del enfoque integral de redes”. Luis Roberto Vega González. Asesor: M.A. Miguel Sánchez Boy. México 2001. Programa de Posgrado en Ciencias de la Administración, FCA-UNAM.

Impacto de la tecnología de información sobre las organizaciones mexicanas. Tesis maestría. Fernando Mar Olivares. Asesor: MAI. Héctor Horton Muñoz. México 2006. Programa de Posgrado en Ciencias de la Administración, FCA-UNAM.

La tecnología como factor de competitividad en una microempresa. Tesis maestría. Mario Bonilla Gutiérrez. Asesor: Dr. Luis Alfredo Valdés Hernández. México 2008. Programa de Posgrado en Ciencias de la Administración, FCA-UNAM.

Modelo administrativo para la gestión de la innovación tecnológica en México. Tesis maestría. Ma. Del Carmen Tamayo Díaz. Asesor: José Luis Solleiro Rebolledo. México 2002. Programa de Posgrado en Ciencias de la Administración, FCA-UNAM.

Planeación estratégica aplicada a las empresas de base tecnológica. Tesis maestría. Adriana Oviedo Pacheco. Asesor: Dr. Luis Alfredo Valdés Hernández. México 2009. Programa de Posgrado en Ciencias de la Administración, FCA-UNAM.

Tecnología para la Pequeña y Mediana Empresa Mexicana. Tesis maestría. Cuauhtémoc Hernández Suárez. Asesor: M.A. Carlos Puga Murguía . México 2002. Programa de Posgrado en Ciencias de la Administración, FCA-UNAM.

LIBROS

ALBA ALDAVE Ma. Cristina (2005) Franquicias: una perspectiva Mundial, Primera edición, Fondo Editorial FCA, D.F. México.

ALBA ALDAVE Ma. Cristina (2004) Las franquicias en México: una nueva visión, Primera edición, Fondo Editorial FCA, D.F. México.

BABINET Gilles (2015) Big Data, penser l'homme et le monde autrement. Primera Edición, Editeur Le passeur. Francia

BALAGUÉ Christine et FAYON David (2012) Facebook, Twitter et les autres...Intégrer les réseaux sociaux dans une stratégie d'entreprise. Segunda edición. Ed. Pearson. Francia

BODIET Stéphane (2011) Que sais-je? Le Web Marketing. Segunda edición. Presses Universitaires de France, Paris, France

DELORT Pierre (2015) Que sais-je? Le Big Data. Presses Universitaires de France, Paris, France

FOURMENTRAUX Jean-Paul (2015) Identités numériques. Expressions et traçabilité. Primera edición. Les Essentiels d'Hermès. CNRS Éditions. Paris, Francia.

KAWASAKI Guy (2015) Twitter power. How to dominate your market one twwt at a time. Infomedia, Inc. And Dave Taylor. New Jersey, U.S.A.

LOSHIN David (2013) Big Data Analytics: from strategic planning to Enterprise integration with Tools, techniques, NoSQL, and Graph, Elsevier Inc., Waltham, M.A., USA

MAYER-SCHÖNBERGER Viktor & CUKIER Kenneth (2013), A Revolution that Big Data will transform how we Data live, work, and think, Houghton Mifflin Harcourt Publishing Company, New York, USA.

NIEL Xavier et ROUX Dominique (2008) Que sais-je? Les 100 mots de l'internet. Tercera edición. Presses Universitaires de France, Paris, France

PUISIEUX Guillaume (2008) Le marketing des systèmes d'information. EYROLLES Éditions d'organisation. Paris, Francia

SCHMARZO Bill (2013) Big Data: El poder de los datos ANAYA Multimedia. Madrid, España

SIMON Phil (2013), John Wiley & Sons, Too Big To Ignore. Inc. Hoboken, New Jersey, USA.

THOMAS Rob & McSHARRY Patrick (2015) Big Data Revolution, what farmers, doctors and insurance agents teach us about discovering big Data Patterns. Primera edición. John Willey and sons Ltd. West Sussex, United Kingdom

Glosario

AVISO DE PRIVACIDAD: Documento físico, electrónico o en cualquier otro formato generado por el responsable que es puesto a disposición del titular, previo al tratamiento de sus datos personales, de conformidad con el artículo 15 de la Ley Federal de Protección de Datos Personales en Posesión de Particulares. (LFPDPPP, 2010)

DATO PERSONAL: Se refiere a toda aquella información asociada a una persona o individuo que lo hace identificable del resto de las personas y/o como parte de un grupo determinado de individuos, por ejemplo: nombre, domicilio, teléfono, fotografía, huellas dactilares, sexo, nacionalidad, edad, lugar de nacimiento, raza, filiación, preferencias políticas, fecha de nacimiento, imagen del iris del ojo, patrón de la voz, etc. La idea central de este concepto es común en las legislaciones de protección de datos que distintos países han redactado. (SÁNCHEZ PÉREZ & ROJAS GONZÁLEZ, 2012)

DATOS NO ESTRUCTURADOS: Son aquellos datos no almacenados en una base de datos tradicional. La información no estructurada no puede ser almacenada en estructuras de datos relacionales predefinidas. (DATAPRIX, 2014)

DATOS SEMI ESTRUCTURADOS: Los datos semiestructurados serían aquellos datos que no residen de bases de datos relacionales, pero presentan una organización interna que facilita su tratamiento, tales como documentos XML y datos almacenados en bases de datos NoSQL. (DATAPRIX, 2014)

DATOS PERSONALES SENSIBLES: Todos aquellos datos que se relacionan con el nivel más íntimo de su titular y cuya divulgación pueda ser causa de discriminación o generar un severo riesgo para su titular. De manera general, se

consideran datos sensibles aquellos que revelen características como origen étnico o racial, estado de salud, creencias religiosas, opiniones políticas, preferencia sexual, pertenencia a sindicatos, creencias filosóficas y morales, entre otras. Esta clase de información debe ser tratada con mayor responsabilidad y establecer medidas de protección más estrictas. (SÁNCHEZ PÉREZ & ROJAS GONZÁLEZ, 2012)

DATOS BIOMÉTRICOS: Por definición común, los datos biométricos son aquellos rasgos físicos, biológicos o de comportamiento de un individuo que lo identifican como único del resto de la población. Aquellos sistemas informáticos en los que se mide algún dato biométrico, como parte del proceso de identificación y/o autenticación de un sujeto, son conocidos como sistemas de seguridad biométrica o simplemente sistemas biométricos. (SÁNCHEZ PÉREZ & ROJAS GONZÁLEZ, 2012)

METADATOS: Los metadatos son datos altamente estructurados que describen información, describen el contenido, la calidad, la condición y otras características de los datos. (Universidad Nacional de Colombia, 2016)

TIEMPO UNIX: Tiempo Unix o Tiempo POSIX es un sistema para la descripción de instantes de tiempo: se define como la cantidad de segundos transcurridos desde la medianoche UTC del 1 de enero de 1970, sin contar segundos intercalares. (UnixTime.Info, 2011)

TRAZABILIDAD: Según la Unificación de Normativas Españolas 66.901-92, la trazabilidad “es la capacidad para reconstruir el historial de la utilización o la localización de un artículo o producto mediante una identificación registrada.” (Universidad Politécnica de Cartagena, 2016)

Apéndices

Análisis Bibliométrico

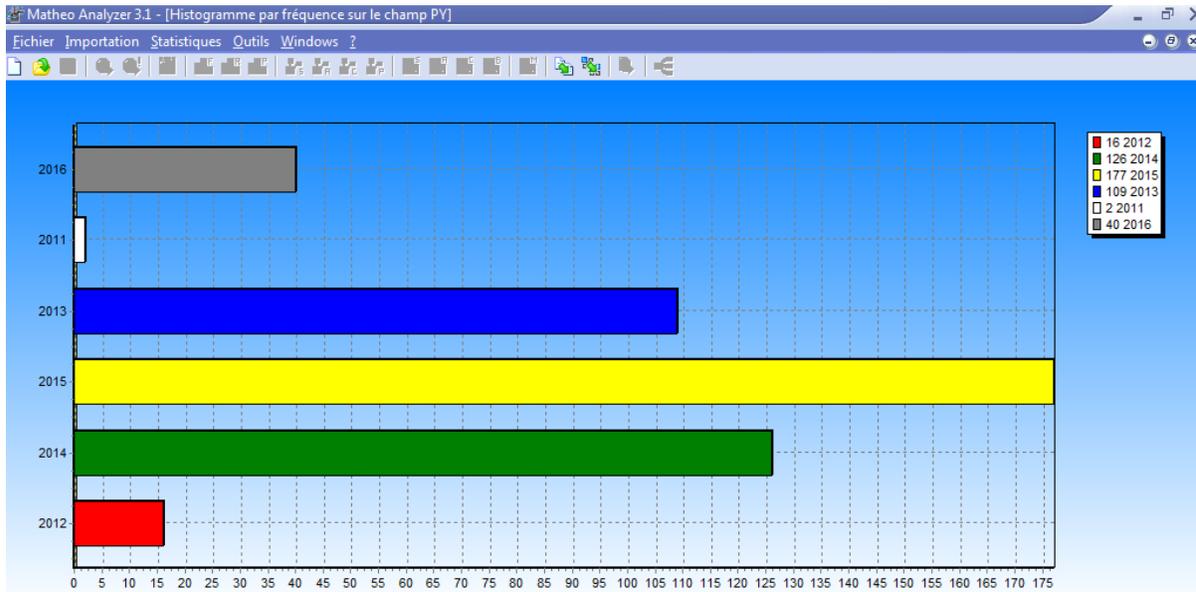
Introducción

El análisis bibliométrico es la “aplicación de métodos matemáticos y estadísticos para el análisis de libros y otros medios de comunicación”. (PRITCHARD, 1969) El estudio que se obtiene con el análisis bibliométrico permite conocer la tendencia que prevalece en el tema de Big Data que fue analizado por búsquedas realizadas en la base de datos “*Web of knowledge*” que brinda la biblioteca digital de la UNAM. Los resultados de las búsquedas se exportaron en texto plano para su posterior tratamiento en lo software: Matheo Analyzer, Sci2 Tool y Gephi. Estos dos últimos softwares muestran el análisis proporciona diversas graficas entre las que se resaltan: la gráfica de coocurrencia de palabras clave originales de Big Data, la gráfica temporal para la identificación de términos emergentes de palabras clave originales del tema de Big Data.

Búsqueda del término “Big Data”

Se realizó la búsqueda de Big Data con 470 registros de la palabra Big Data por lo que se obtuvieron los resultados siguientes.

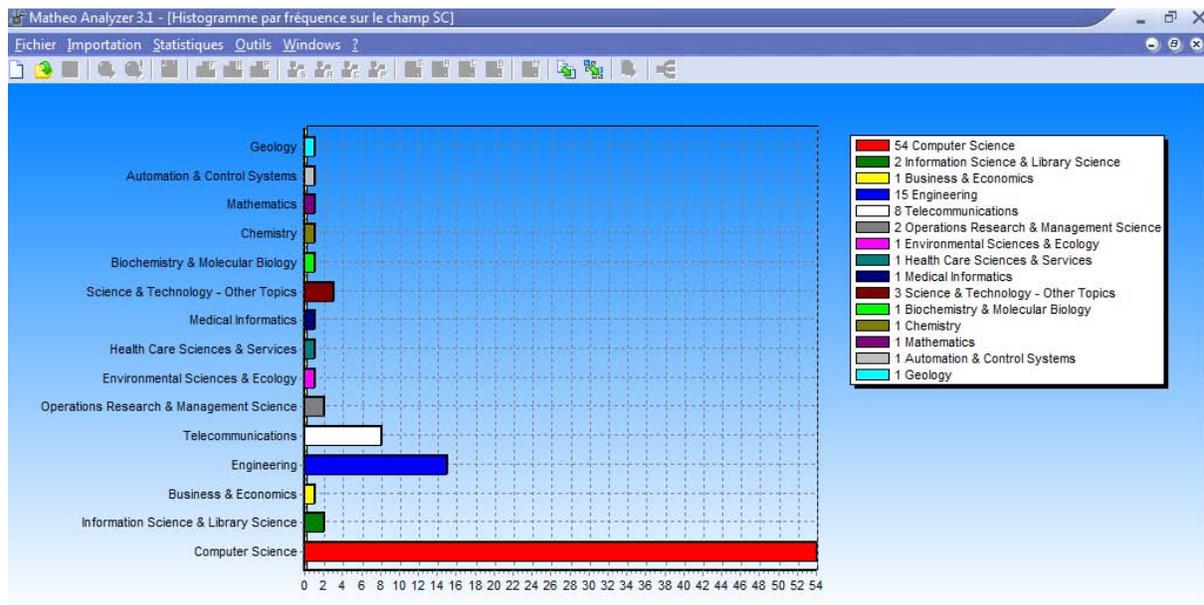
Figura 21 Histograma de publicaciones por año de Big Data



Fuente: Elaboración propia

El estudio del histograma de publicaciones por año del tema de Big Data refleja que a partir de 2013 el número de publicaciones no ha cesado de aumentar. Efectivamente, en 2015, a nivel mundial se habla sobre la importancia de Big Data para el análisis de las grandes cantidades de datos generados por segundo a nivel mundial. Es un tema muy novedoso que está en su apogeo.

Figura 22 Histograma por frecuencia sobre el campo SC de Big Data



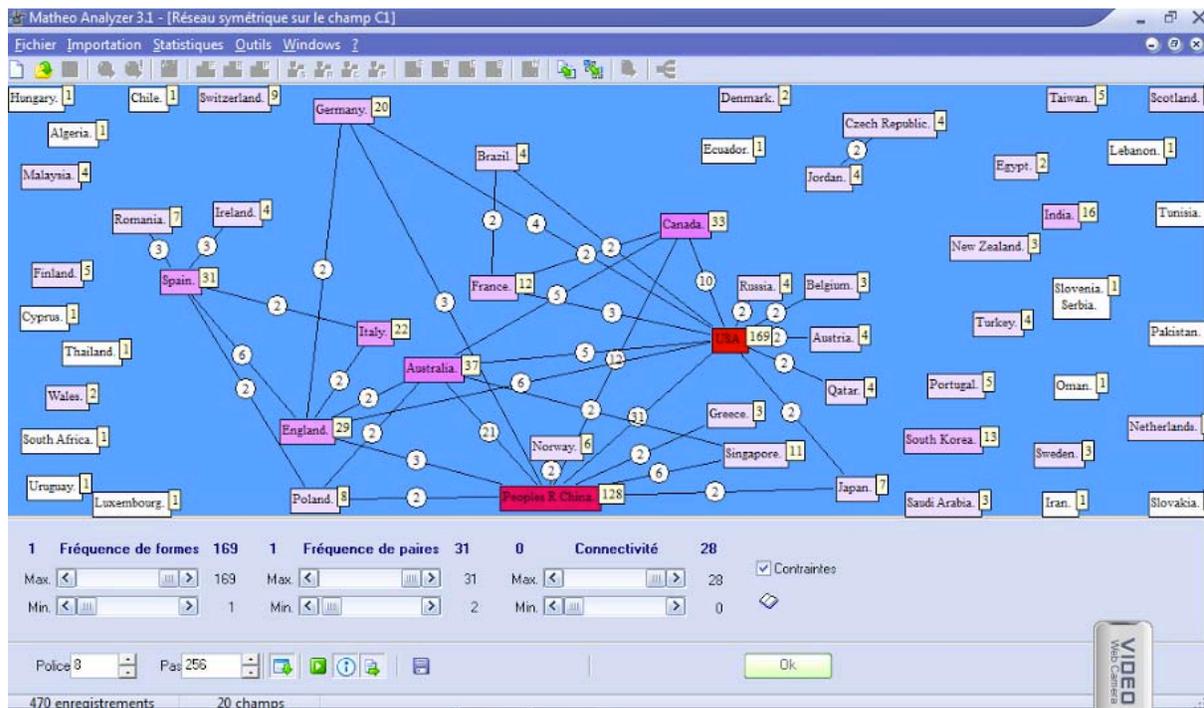
Fuente: Elaboración propia

El termino Big Data está completamente relacionado con Ciencias de la Computación y con las innovaciones tecnológicas por lo que no es de asombrarse que la principal categoría temática será Ciencias de la computación. Si se consideran las 3 principales categorías temáticas, se obtiene:

1. Ciencias de la computación
2. Ingeniería
3. Telecomunicaciones

Big data es un tema que está relacionado con la computación dado que es una herramienta que facilita el análisis de los datos; está relacionado con la ingeniería porque incluye métodos matemáticos para su estudio y por último está relacionado con las telecomunicaciones dado que el auge de los dispositivos conectados a internet ha provocado este universo digital con una gran cantidad de datos que requieren ser analizados.

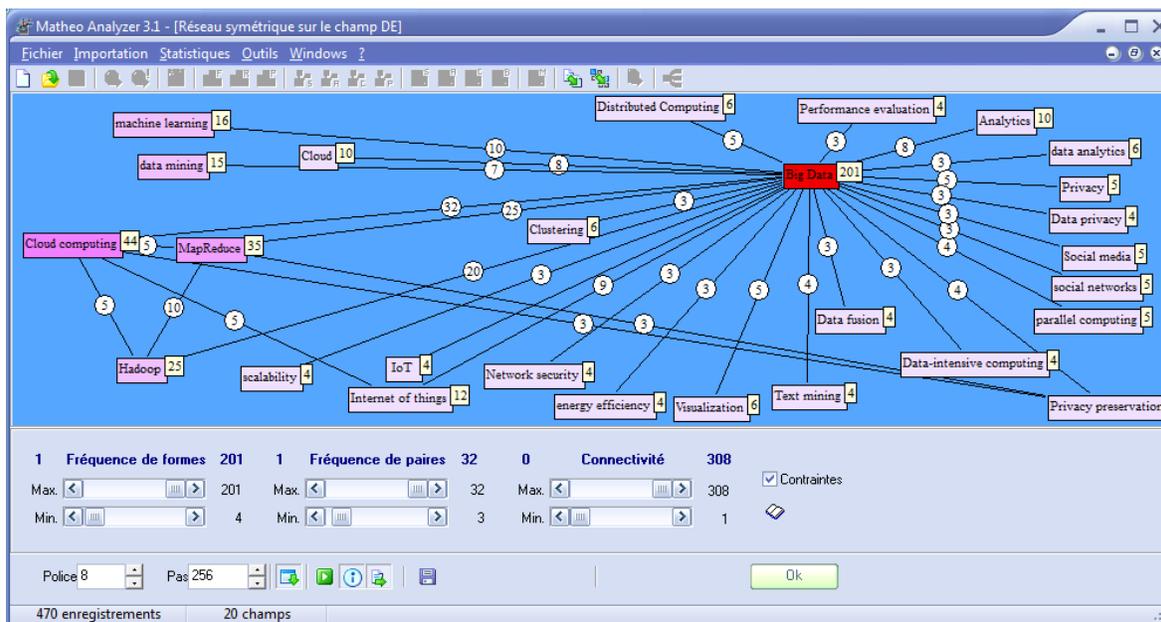
Figura 23 Red simétrica sobre el campo C1 de Big Data



Fuente: Elaboración propia

Esta red muestra que Estados Unidos con 169 referencias, seguido de China con 128 formas, seguidos de Australia con 37 referencias, Canadá con 33 referencias y España con 31 formas con los países que más publican sobre Big Data. Un aspecto sorprendente es que indique que Rumanía, con 7 referencias, es un país que publica artículos relacionados con Big Data ya que a nivel mundial no es un país que se considere como especialista en temas de innovación y tecnologías como Japón que tiene 7 formas, que asombrosamente tiene el mismo número de publicaciones. Qatar es un nodo con ruta corta que tiene 4 referencias, como Jordania y República Checa, no obstante entre estos dos últimos países se tienen 2 pares de referencias. Corea del Sur es un nodo aislado con 13 referencias que carece de vínculos con algún país.

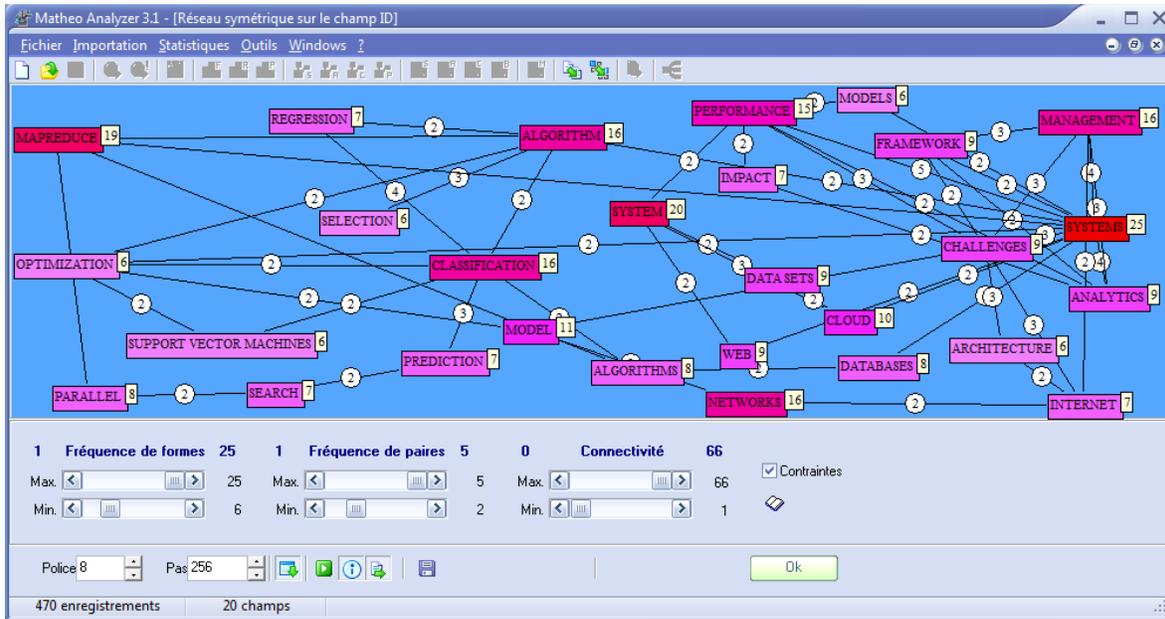
Figura 24 Red simétrica sobre el campo (DE) de Big Data



Fuente: Elaboración propia

La búsqueda por palabras clave (DE) refleja la importancia de la computación en la nube (44 referencias), *MapReduce* (35 referencias), *Hadoop* (25 referencias), *Machine Learning* (16 referencias), e internet de las cosas (12 referencias), para conocer el término de Big Data ya que las referencias o número de formas son las más elevadas. Si bien, la gran mayoría de las personas relaciona Big Data con redes sociales (4 formas), en este marco, el estudio demuestra si existe una relación importante entre Big Data y las redes sociales pero no es tan trascendente como el Cloud Computing (32 formas). Una cuestión que es importante resaltar es en relación a la seguridad. En este aspecto, mucho se ha comentado sobre la carente seguridad del Internet de las cosas y Big Data pero en esta red se muestra que las publicaciones arbitradas no se resalta esa relación dado que no hay un vínculo directo entre IoT o *Internet of Things* y *Privacy preservation* y/o *Network security* y por ende no tiene ningún par de referencias o formas en común. Por último, el termino Big Data tiene en su mayoría rutas cortas con los términos relacionados.

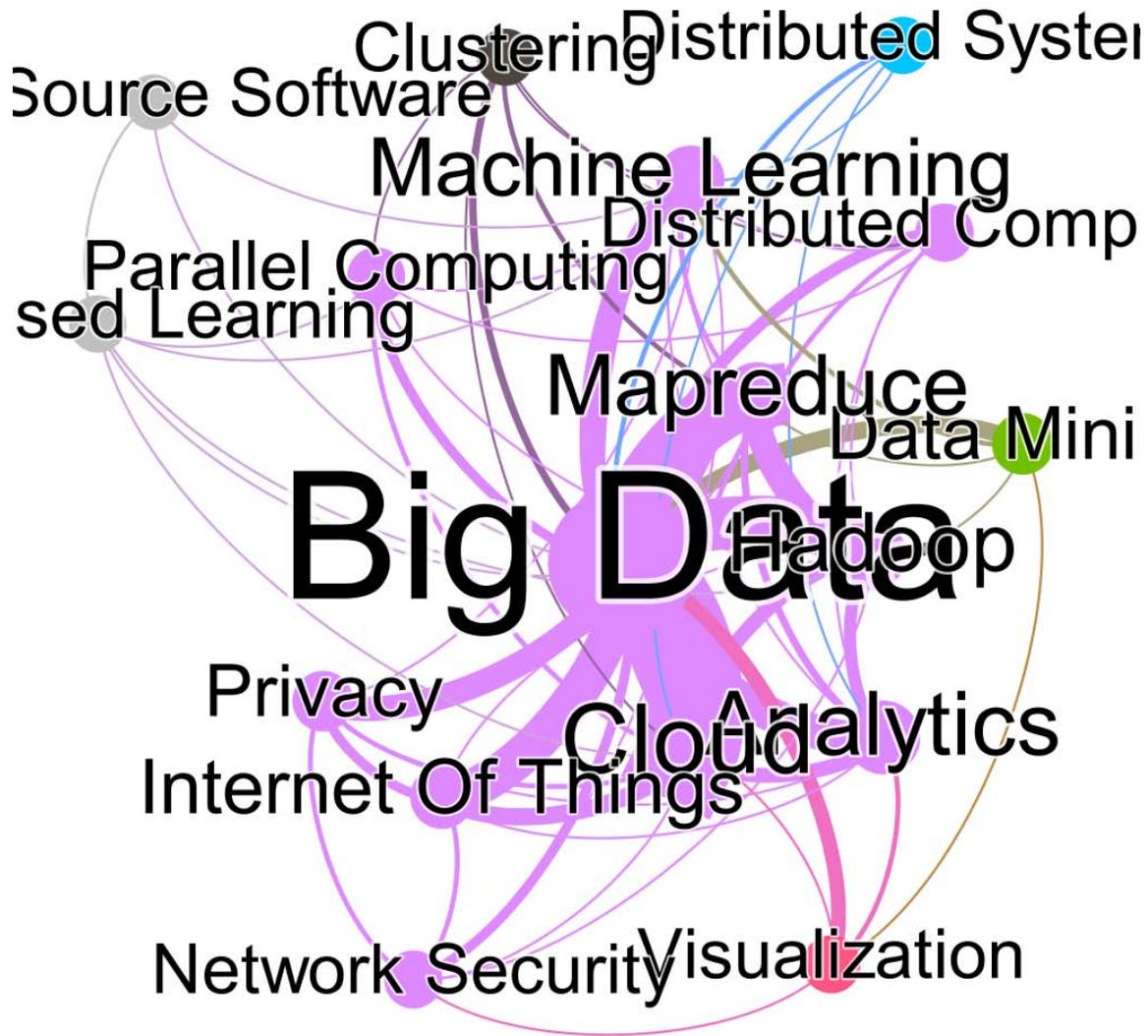
Figura 25 Red simétrica sobre el campo (ID) de Big Data



Fuente: Elaboración propia

Esta red muestra las palabras clave relacionadas con Big Data. Destacan las palabras: Sistemas (25 referencias), Sistema (20 referencias), MapReduce (19 referencias) y Algoritmos, Administración, Clasificación y redes (16 referencias). Estas mismas palabras clave tienen una conectividad superior a 3 lo que refleja su importancia respecto al tema de Big data y a su vez reflejan una centralidad destacable.

Figura 27 Gráfica de coocurrencia de palabras clave de Big Data



Fuente: Elaboración propia

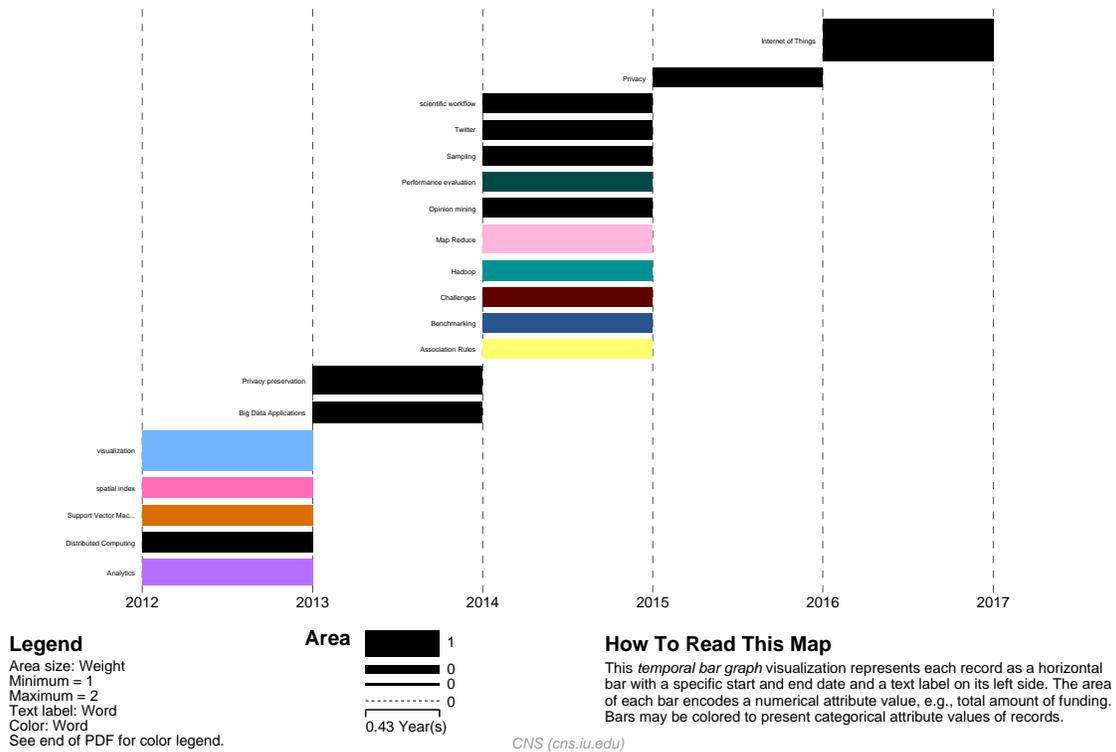
En relación a la coocurrencia de palabra clave del tema de Big Data se recalca que las palabras con una numerosa coocurrencia son: *Cloud* y *Analytics*. No obstante, la grafica no es similar a la de escenarios ya que en este caso casi todas las palabras clave tienen el mismo tamaño por lo que se infiere que los nodos tienen el mismo peso a excepción de *Cloud* y *Analytics*. Esta afirmación se podrá comprobar con la gráfica de visualización temporal.

Visualización temporal para la identificación de términos emergentes de palabras clave originales el tema Big Data

Figura 28 Gráfica temporal para la identificación de términos emergentes de palabras clave de Big Data

Temporal Visualization

(Generated from CSV file: C:\Users\PEDROB-1\AppData\Local\Temp\temp\Preprocessed-maximum burst level 16915788742772764548-6688789401312555245.csv)
September 27, 2016 | 12:27 AM CDT



Fuente: Elaboración propia

Temporal Visualization

(Generated from CSV file: C:\Users\PEDROB-1\AppData\Local\Temp\temp\Preprocessed-maximum burst level 18915788742772764548-6688789401312555245.csv)
September 27, 2016 | 12:27 AM CDT

- Analytics
- Association Rules
- Benchmarking
- Challenges
- Hadoop
- Map Reduce
- Performance evaluation
- Support Vector Machine
- spatial index
- visualization
- Big Data Applications
- Distributed Computing
- Internet of Things
- Opinion mining
- Privacy
- Privacy preservation
- Sampling
- Twitter
- scientific workflow

GNS (gns.li.edu)

Después de analizar la gráfica de coocurrencia de palabras clave se identificó que la mayoría de los términos tienen el mismo peso, dato que se confirma con la gráfica temporal para la identificación de términos emergentes de palabras clave. En efecto, para el tema de Big Data, los términos emergentes tienen un promedio de área de 1 y una duración de 1 año entre el año 2012 y el año 2017, lo que demuestra que es un término innovador y en constante cambio. La gráfica refleja la innovación del tema dado que muestra datos de 2017 y la duración de cada término es de únicamente 1 año. Los términos emergentes que destacan en esta gráfica son: Internet de las cosas (2016-2017) y visualización (2012-2013) lo que se infiere es que el término de Internet de las cosas, a futuro, seguirá con la misma tendencia de tener un peso importante sobre Big Data. Al no existir un término emergente que tenga una duración mayor a 1 año, se infiere que es necesario tener mayor investigación sobre Big Data y sobretodo estudios que sean actuales y publicado con premura dado lo efímero del tema.

Resumen general enviado a los informantes

Modelo de *Big Data Analytics* para la toma de decisiones en las organizaciones con manejo de grandes volúmenes de datos.

Resumen

La inversión tecnológica es un factor estratégico para alcanzar los objetivos empresariales. Hoy, el uso de Big Data fomenta la transformación de los sistemas y las infraestructuras de las tecnologías de información, así como favorece la mejora en los procesos de negocio, las soluciones innovadoras y los productos. Las organizaciones precisan incluir, como apoyo en la toma de decisiones un modelo analítico adaptable a cualquier tipo de empresa que maneje grandes volúmenes de datos para identificar nuevas oportunidades de negocio, ayudar en una eficiente toma de decisiones, incrementar las ganancias, mejorar las relaciones con los clientes y proveedores, optimizar la operación, ampliar la cartera de productos y sobre todo gestionar los datos para convertirlos rápidamente en información valiosa para una rápida, eficiente y automática toma de decisiones. *Big Data Analytics* es un elemento clave para la toma de decisiones y para tener en la empresa u organización una ventaja competitiva frente a la competencia.

Palabras clave: *Big Data*, *Big Data Analytics*, Toma de decisiones

Introducción

La analítica de grandes volúmenes de datos (*Big Data Analytics*) es de gran importancia para las organizaciones que manejan volúmenes importantes de datos que no pueden ser analizados con las herramientas tradicionales. *Big Data Analytics*, se implementa hoy en día en las empresas para:

- Mejorar la operación
- Incrementar ganancias
- Mejorar la venta a los clientes actuales, es decir vender más a los clientes actuales
- Buscar nuevos clientes
- Ampliar la cartera de productos

Actualmente, son numerosas las empresas que hacen uso de herramientas Big Data para describir los acontecimientos en la organización como puede ser conocer el “¿Cuántos?”: ¿Cuántos clientes hombres tiene la empresa?, ¿Cuántas mujeres?, ¿Qué rango de edad?. Las organizaciones también implementan Big Data para poder hacer proyecciones a futuro basados en el histórico de los datos.

Sin embargo, una de las aplicaciones más deseadas para las empresas es utilizar los analíticos de grandes volúmenes de dato para el descubrimiento de patrones ocultos, correlaciones desconocidas, elaborar sistemas de recomendaciones, entre otras informaciones útiles para la toma de decisiones.

A lo largo de los años, ha existido una importante evolución del acceso a la información. Hace sesenta años, las computadoras digitales hicieron legible la información, hace veinte años, Internet la hizo alcanzable y hace diez años los primeros motores de búsqueda unificaron el acceso a la información por medio de aplicaciones web.

La analítica avanzada en sinergia con Big Data, provoca una ola con grandes transformaciones en todo tipo de modelo de negocio, dado que permiten a las organizaciones considerar prioritariamente los elementos significativos y conocimientos aplicables que están sumergidos en los datos y que no han sido tomados en cuenta. La clave del rendimiento empresarial está en sacarle provecho a la analítica predictiva para mejorar el proceso analítico tradicional y reflejar conocimientos relevantes de las organizaciones.

Los medios sociales, ofrecen información creada por los usuarios sobre los intereses, pasiones, necesidades y asociaciones que son una fuente de valor significativa para las organizaciones. Asimismo, los datos generados por móviles sensores o dispositivos conectados a Internet ofrecen una visión detallada, en tiempo real sobre los usuarios que permite hacer recomendaciones sobre bienes o servicios, tener un compromiso y una interacción con usuarios. Un aspecto importante está en la velocidad de actuación y en la rapidez con la que se procesan los datos, se cruzan, se ordenan y se actúa para tomar una decisión.

El análisis de los grandes volúmenes de datos ofrece un gran potencial para la empresa porque supone analizar todos los datos generados en tiempo real, con un nivel de detalle nunca antes visto. Trasformar los datos en información es lo que genera un enriquecimiento en todos los aspectos para las organizaciones. Big Data permite descubrir patrones de comportamiento sobre los clientes, productos, campañas y operaciones de las empresas para la toma de decisiones de manera rápida y efectiva y en caso de fallo o equivocación, enmendar los errores velozmente.

Los Analíticos de Grandes Volúmenes de Datos permiten descubrir patrones ocultos, correlaciones desconocidas, tendencias del mercado, preferencias de los clientes y posibles riesgos para optimizar el entorno empresarial de manera predictiva, alimentándose de datos obtenidos en tiempo real.

Uno de los grandes beneficios de Big Data es que permite tomar decisiones personalizadas basándose en datos que se obtienen de los patrones de comportamiento de los clientes, productos y operaciones realizadas dentro y fuera de la organización y que permite optimizar las principales iniciativas empresariales y descubrir nuevas posibilidades de monetización. (SCHMARZO, 2013) Los competidores innovan constantemente por lo que se necesita estar a la vanguardia y ofrecer a los clientes el producto adecuado en el momento apropiado con tal de identificar las oportunidades de negocio perdidas que no han sido tomadas en cuenta.

Los grandes volúmenes de datos someten a prueba la infraestructura actual de las organizaciones, las cuales deben actuar de forma inmediata sin temor a equivocaciones y, en caso de error, actuar de manera inmediata, por lo que hoy más que nunca se requiere de una arquitectura diseñada para manejar en tiempo real grandes cantidades de datos estructurados y no estructurados para la óptima toma de decisiones.

Objetivo

Diseñar un modelo de Big Data Analytics que se pueda aplicar a las organizaciones con manejo de grandes volúmenes de datos para mejorar la toma de decisiones a nivel directivo.

Hipótesis

El uso de un modelo de Big Data Analytics en las organizaciones con manejo de grandes volúmenes de datos mejora la toma de decisiones a nivel directivo.

Antecedentes

El crecimiento del universo digital está considerado del 40% anual en la próxima década. (EMC MEXICO, 2014) No sólo por el incremento de la población mundial sino también por el hecho que las personas y las empresas realizan todas sus actividades en línea, aunado a todas las “cosas” como dispositivos inteligentes que están conectados a internet y que están generando una nueva ola de oportunidades de negocios alrededor del mundo. Al igual que el universo físico, el universo digital es enorme. Para el año 2020 se prevé que existan casi tantos bits digitales como estrellas en el universo.

El volumen de datos masivos crece constantemente y todos formamos parte de dicho crecimiento. Así, debido al rápido avance tecnológico, las empresas se han tenido que enfrentar a nuevos desafíos como el analizar los grandes volúmenes generados para poder tomar decisiones de forma rápida y eficaz. Hoy, emplear herramientas tradicionales para analizar los datos generados segundo a segundo por las nuevas tecnologías, han quedado obsoletas por lo que se requieren nuevos métodos para poder enfrentar las necesidades de los usuarios.

De la revolución digital que estamos viviendo en el siglo XXI surgen nuevos conceptos como Big Data, *Big Data Analytics*, Ciencia de Datos, entre otros, que hacen referencia a la acumulación masiva de datos y a las técnicas usadas para encontrar patrones dentro de esos datos.

Según Doug LANEY, Big Data se caracteriza por su *slogan* de las tres “V”s: volumen, velocidad y variedad. Estas tres “V”s surgen de un artículo titulado “3D Data Management Controlling Data: Volume, Velocity and Variety” (LANEY, 2001) publicado el 6 de febrero de 2001 por Meta Group (ahora Gartner) donde no se menciona explícitamente el término Big Data pero se constata del incremento de los formatos de datos, su volumen y se recomienda una mejor gestión de los datos. Doug Laney, analista de META Group, es uno de los pioneros del tema y subraya la importancia del análisis de grandes volúmenes de datos en congresos y presentaciones relacionadas con el tema; define el crecimiento constante de los datos como una oportunidad y un reto para investigar lo que más tarde, IBM denominó las 4V’s que hacen alusión al volumen, variedad, veracidad y visibilidad. Desde entonces muchos autores han intentado profundizar en el tema incluyendo otros tres elementos otros como la Veracidad, Valor y Visibilidad de los datos.

En efecto, el creciente volumen de información disponible implica retos ambiciosos, técnicos y analíticos que los métodos tradicionales no logran vislumbrar en un tiempo de respuesta casi inmediato. En cuanto a la variedad, la facilidad de enviar un correo electrónico con datos de texto, mensaje de voz y fotos ha generado una enorme cantidad de información de datos totalmente heterogéneos. La veracidad hace mención de la calidad del dato como su predictibilidad; pero nada de esto sería extraordinario sin el elemento más importante que es la velocidad con la que es necesario analizar la información y sobre todo tomar decisiones justo a tiempo.

Los datos masivos nos impulsan a centrarnos en el potencial que supone analizar todos los datos ya sean internos o externos de la organización en tiempo real. Para realizar esto, se requiere de una infraestructura actualizada y moderna que

pueda sacar partido de la riqueza de los datos. “Estamos ante una tecnología seria, cuyo valor para crear nuevos modelos de negocio basados en petabytes de datos en tiempo real, acompañados de analíticas avanzadas, ya han sido validados por varios sectores tan diversos como el minorista, los servicios financieros, las telecomunicaciones, la fabricación, la energía, el transporte y la hostelería.” (SCHMARZO, 2013) La gran cantidad de datos generados requiere de una infraestructura actualizada para aprovechar y analizar los datos y transformarlos en información. El mundo de los grandes volúmenes de datos impacta en todos los sectores, no obstante, por ejemplo, en el campo de la salud, tiene grandes beneficios para la sociedad, como veremos más adelante.

Marco teórico

El concepto de Big Data

Después de analizar algunas de las definiciones de Big Data más relevantes, para tener una mayor comprensión del concepto (GARTNER Inc., 2016), (LANEY, 2001), (PÉREZ G. , 2015), (Oxford University Press, 2016), (McKinsey Global Institute (MGI), 2012), (MAYER-SCHÖNBERGER & CUKIER, 2013), (DAVENPORT, Big Data at Work: Dispelling the Myths, Uncovering the Opportunities, 2014), se opta, para esta investigación, considerar la siguiente definición:

Conjuntos de datos de alto volumen, variabilidad, veracidad, visibilidad, valor, velocidad y variedad cuyo tamaño va más allá de la capacidad de las herramientas típicas de software de bases de datos para capturar, almacenar, gestionar y analizar datos para extraer nuevos conocimientos, crear nuevas formas de valor y exigir formas innovadoras y rentables de procesamiento de la información para mejorar la comprensión y la toma de decisiones.

Este concepto refleja de manera general, la necesidad de contar con una herramienta que transforme la gran cantidad de datos generados digitalmente en información veraz y eficaz para una óptima toma de decisiones a nivel directivo.

Tipos de analítica

Para entender qué son los Analíticos de Grandes Volúmenes de Datos es importante definir lo que es la analítica.

De acuerdo a la Real Academia Española, la analítica significa “perteneciente o relativo al análisis” y “que procede descomponiendo, o que pasa del todo a las partes”. (Real Academia Española).

En el ámbito de la estadística existe lo que se denomina como método estadístico que es el método científico aplicado a la estadística y se divide en dos fases: la descriptiva, la inferencial y la analítica.

En este marco, por una parte, “la estadística descriptiva trata de organizar, representar y resumir un conjunto de datos de manera que pueda ser extraída la máxima información procedente de ellos”. (GARCÍA RAMOS, RAMOS GONZÁLEZ, & RUIZ GARZÓN, 2007) Por otra parte, la estadística analítica y la estadística inferencial “intenta sacar conclusiones respecto del fenómeno sobre el cual se ha tomado los datos”. (GARCÍA RAMOS, RAMOS GONZÁLEZ, & RUIZ GARZÓN, 2007)

Existen diversos tipos de analítica:

- **Analítica predictiva:** Se refiere al uso de métodos estadísticos para predecir el comportamiento de los usuarios con ayuda de la minería de datos para la predicción de probabilidades y tendencias. “La analítica predictiva pretende la mejora de decisiones y respuestas rápida al cambio y comprende técnicas de minería de datos, estadística y modelado; trata de analizar hechos actuales o históricos con el propósito de hacer predicciones sobre sucesos futuros. Desde un punto de vista corporativo, trata de predecir el comportamiento de sus diferentes categorías de usuarios, especialmente los clientes. » (Big Data Social, 2016)
- **Analítica de redes:** Se utiliza para el análisis de “las relaciones entre nodos discretos de un grado o de una red.” (Big Data Social, 2016) Su mayor utilización es en la redes sociales donde mediante la analítica de redes se pueden conocer las conexiones entre diversos objetos, personas u organizaciones. La analítica de redes ayuda a las organizaciones a medir su impacto en las redes sociales con los objetivos de la empresa y en las conversaciones relacionadas.
- **Analítica de sentimientos:** La minería de opiniones o analítica de sentimientos, es una herramienta utilizada por las empresas para identificar

y extraer información subjetiva de contenido de diversos tipos como el análisis de medios sociales, el análisis de la voz del cliente a través de encuestas o el análisis de la voz del ciudadano. Hace uso de tecnologías de procesamiento del lenguaje natural, analítica de textos y lingüística computacional. El objetivo de la analítica de sentimientos es “deducir de este análisis, comportamiento de clientes, hábitos de compra, sistemas de recomendación, etc. Este tipo de analítica se asienta en la llamada minería de opinión que implica la construcción de un sistema para recolectar y examinar las opiniones relativas a un producto, hechas en entradas (posts) de blogs, comentarios, revisiones o tuits (tweets). La minería de opinión automatizada utiliza técnicas de *machine learning*, uno de los componentes clave de la inteligencia artificial. » (Big Data Social, 2016)

- **Analítica de textos:** Se basa en la minería de textos para “descubrir, a partir de cantidades de texto grandes, el conocimiento que no está literalmente escrito en cualquiera de los documentos. Esto incluye buscar tendencias, promedios, desviaciones, dependencias, etc.” (TRIPOD, 2016)
- **Analítica web:** “Es la recogida y análisis de los datos que se registran cuando un usuario visita un sitio web y sirve para medir el comportamiento de los usuarios. Existen numerosas herramientas de analítica de datos: Google Analytics, Woopra, Omniture, Adobe Analytics, etc.” (Big Data Social, 2016)

Metodología

El estudio que se realizará será:

- ✓ **Cualitativo:** El modelo pretende ser evaluado mediante entrevistas cualitativas hechas a expertos en el tema de Big Data
- ✓ **Exploratorio:** A través del estudio se investiga una solución para la toma de decisiones.

- ✓ **De corte transversal:** El modelo no pretende ser puesto en operación por lo que no se ambiciona la evaluación de los resultados a través del tiempo.
- ✓ **No Experimental:** El estudio es no experimental porque no se manipularán de manera intencional las variables para analizar las consecuencias de tal manipulación. En la investigación se analizará las variables y se observarán los fenómenos en su ambiente natural para posteriormente analizarlos y realizar un modelo.

El diseño de la investigación se realiza en 7 etapas:

8. **Identificación del problema:** Definición del problema actual que requiere una solución para una eficiente toma de decisiones en las organizaciones.
9. **Análisis bibliométrico:** Aplicación de métodos matemáticos y estadísticos para el análisis de publicaciones científicas.
10. **Formulación de hipótesis:** Se planteará una hipótesis para la solución del problema
11. **Elaboración del marco teórico:** Establecimiento del contexto teórico y conceptual bajo el cual será establecida la propuesta modelo de *Big Data Analytics* para la toma de decisiones en las organizaciones.
12. **Diseño del modelo:** Elaboración una propuesta de modelo de *Big Data Analytics* para la toma de decisiones a nivel directivo en las organizaciones con grandes volúmenes de datos.
13. **Evaluación del modelo por expertos:** El modelo propuesto será presentado a 10 expertos de Big Data con la finalidad de recibir retroalimentación del modelo.
14. **Ajustes al modelo:** Con base en los comentarios recibidos por los expertos se llevaran a cabo los ajustes pertinentes al modelo propuesto.

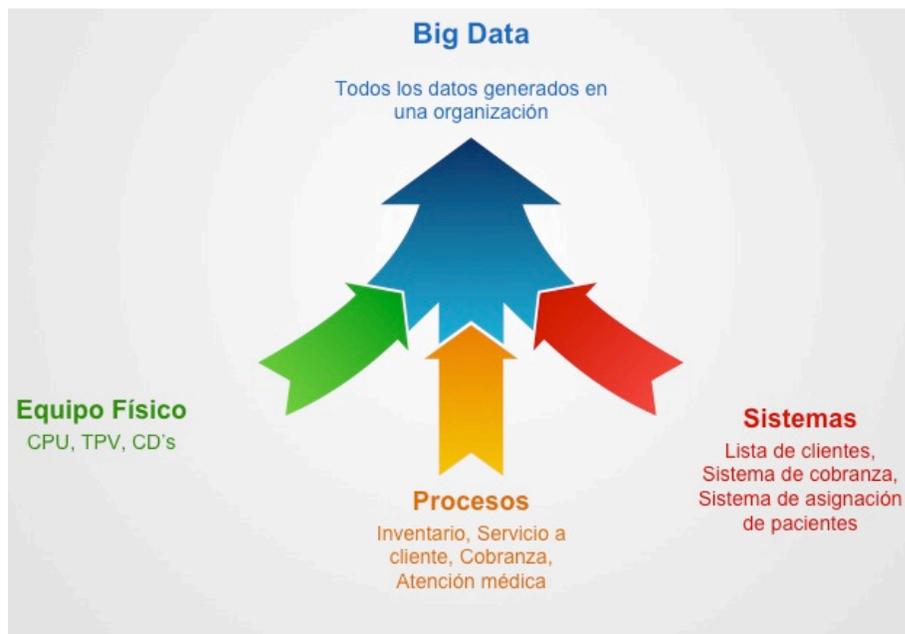
Resultados preliminares

El modelo que se propone a continuación, consta de diez fases. Se parte de la premisa que toda organización con grandes volúmenes de datos cuenta con tres componentes generadores de datos: los equipos físicos, los procesos y los sistemas. El equipo físico hace referencia a los “dispositivos físicos de un sistema informático. Incluye CPU, terminales, placas de memoria, impresoras, unidades de disco, etc. A veces se utiliza en este sentido el término máquina. » (Diccionario LID de Empresa y Economía, 2016)

Los procesos comprenden todas aquellas actividades de la organización que le permiten alcanzar su misión, como por ejemplo, el área de pensiones en una institución de salud pública, cobranza, atención a clientes, inventario, almacenamiento, mantenimiento, entre otras. Son las áreas de la empresa u organización que sirven para definir la razón de ser de la misma.

Un sistema en una organización permite almacenar la información generada por los procesos. Por lo que tanto, el equipo físico como los procesos y los sistemas son factores que influyen en la generación de grandes volúmenes de datos en las organizaciones.

Figura 29 Fuentes generadoras de grandes volúmenes de datos



Fuente: Elaboración propia

Fase I: Identificar el eje de aplicación en la empresa

En esta fase se busca conocer cuáles son los procesos clave de la organización. En este aspecto, se debe estudiar el modo de operación de la organización para identificar las áreas de la empresa con un rendimiento inferior para conocer cuáles requieren mayor atención. A través de esta gestión del rendimiento empresarial que utiliza la analítica básica, se busca identificar los procesos y los indicadores empresariales clave. En este aspecto es primordial no perder de vista el principal objetivo empresarial.

Mediante el análisis de la gestión del rendimiento empresarial, se pretende conocer cuál es la situación actual de la empresa, hacia dónde desea llegar, identificar los problemas, las oportunidades y los procesos que requieren mayor atención.

Una vez conociendo los procesos de la organización, se debe seleccionar un solo proceso y conocer cuál será el eje de aplicación.

Los ejes de aplicación pueden ser:

- Mejorar la operación
- Incrementar ganancias
- Mejorar la venta a los clientes actuales, es decir vender más a los clientes actuales
- Buscar nuevos clientes
- Ampliar la cartera de productos

El área directiva de la empresa debe definir los siguientes aspectos:

- el proceso que desea eficientar
- el eje que desea aplicar a dicho proceso

FASE II: Ubicar los procesos asociados al eje de aplicación

Una vez identificados los procesos, se deben asociar con los ejes de aplicación.

Para ilustrar esta fase, se ofrece el ejemplo de una institución de salud pública, que tiene como procesos principales, la atención médica, la cobranza a empresas y el otorgamiento de pensiones. El directivo de la institución selecciona el proceso de eficientar de atención a médica y define como eje de aplicación, mejorar la operación de la atención al paciente.

FASE III: Identificar la lista de sistemas de soportan el proceso

Cada proceso tiene una serie de sistemas asociados al mismo. En esta etapa se requiere identificar la lista de sistemas que soportan el proceso a eficientar.

Al retomar el ejemplo de la fase II, los sistemas del proceso de atención médica serían:

- Sistema de alta de derechohabientes

- Sistema de vigencia de derechos
- Sistema de asignación de Unidad Médico Familiar

FASE IV: Analizar la lista de tablas de datos generados para cada sistema

Los sistemas de cada proceso tienen sus propias tablas con datos específicos. Un sistema puede tener un sinnúmero de tablas. Por ejemplo, el sistema de asignación de Unidad Médico Familiar posee 2 tablas:

- Tabla 1: Catálogo de pacientes, que indica la lista de los pacientes de la Unidad Médico Familiar
- Tabla 2: Datos de diagnósticos de pacientes, donde se indican las notas médicas de los pacientes

En el supuesto que el objetivo es instrumentar una forma más eficiente de atención personalizada al paciente, para evitar acciones médicas repetidas, el uso de Big Data permitirá la creación de un solo expediente clínico electrónico que sin importar en qué clínica, hospital o centro sea leído, éste contenga toda la información del paciente.

FASE V: Recepción de datos en crudo (raw)

Se recaban los datos de cada una de las tablas tal cual como los entrega el sistema predictivo, sin ningún tipo de procesamiento. Esta forma de datos se les conoce como datos crudos (*raw data*) o datos originales (*as is data*).

Los datos originales se integran a un almacén de datos llamado *Big Data Box* para su procesamiento.

FASE VI: Gobernanza de los datos

El tratamiento de los datos mejor conocido como gobernanza de los datos (Data Governance) en términos informáticos, es el paso que prosigue al de recepción de

datos en crudo. En efecto, una vez que se integran los datos crudos al *Big Data Box*, este almacén realiza un proceso de gobernanza de los datos, a través de un análisis de estructura y calidad de la información.

FASE VII: Definición de las reglas de negocio

En este punto, se definen las reglas de negocio o métricas de calidad donde se determinan los metadatos que contienen la información que indica qué tipo y qué formato tiene cada dato. Se confirma la información de los datos con la finalidad de conocer cuántos campos tiene cada tabla, qué tipo de dato contiene cada campo y se determinan las reglas de negocio.

Para ilustrar esta fase se retoma el ejemplo de la institución de salud pública. En el supuesto que la tabla de nombres de pacientes tiene 2 campos, estos podrían ser:

- El campo 1: nombre de pila del paciente
- El campo 2: fecha de nacimiento del paciente

Cada campo tiene al menos una regla de negocio. Al considerar el ejemplo citado anteriormente, las reglas de negocio serían:

- Regla de negocio del campo 1: Sólo pueden ser letras
- Regla de negocio del campo 2: Es una fecha con 2 dígitos para el día seguidos (DD) de 2 dígitos para el mes (MM) , seguidos de 4 dígitos para el año (AAAA)

FASE VIII: Verificación de reglas de negocio

En esta fase se verifica que los datos que se encuentran dentro de la Big Data Box cumplan con las reglas de negocio establecidas.

Este análisis ofrece dos resultados:

1. Datos que sí cumplen con las reglas de negocio
2. Datos que no cumplen con las reglas de negocio

Después del análisis, los datos que no cumplen con las reglas de negocio se extraen de la *Big Data Box* y se devuelven al cliente para su análisis y corrección. La *Big Data Box* incluye los datos que fueron seleccionados minuciosamente considerando su calidad, fiabilidad y valor para poner en relieve su capacidad predictiva con la finalidad de asegurar un correcto desarrollo de un modelo analítico.

FASE IX: Definir tipología del análisis de los datos

Una vez que está comprobada la calidad de los datos y que se extrajeron los datos que no cumplen con las reglas de negocio, la *Big Data Box* sólo debe incluir los datos con garantía de calidad. Dichos datos se fraccionan en dos categorías según su análisis.

Existen dos tipos de análisis:

- Análisis MICRO que es el análisis del dato al detalle cuyo acomodo es acorde a la trazabilidad, es decir con acceso a expedientes particulares de una persona.
- Análisis MACRO que es el análisis global del dato cuyo acomodo es para un estudio integral o global con estadísticas sobre los datos.

Por una parte, la capacidad para reconstruir el historial de un producto, es decir, la trazabilidad también se le conoce como búsqueda al átomo.

Recuperando el ejemplo del paciente, mediante la trazabilidad o búsqueda al átomo se puede conocer el historial del paciente, es decir con qué médico se atendió, en qué clínica u hospital, en qué fecha, cuál fue su diagnóstico, etc.

Por otra parte, el análisis global o macro, indica las estadísticas globales mediante el uso de estadística descriptiva. Por ejemplo, una empresa puede conocer qué porcentaje de sus clientes son hombres o son mujeres, cuántos viven en la delegación Miguel Hidalgo, etc. El análisis global ofrece gráficas descriptivas.

A su vez, la información del análisis macro o global se subdivide en 3 tipos:

4. **Análisis descriptivo:** Como se mencionó anteriormente en esta investigación se hace uso de la estadística descriptiva para “organizar, representar y resumir un conjunto de datos de manera que pueda ser extraída la máxima información procedente de ellos” (GARCÍA RAMOS, RAMOS GONZÁLEZ, & RUIZ GARZÓN, 2007) En este punto, se pueden transformar y enriquecer los datos mediante técnicas matemáticas para establecer nuevas métricas en relación a los datos como: la frecuencia (¿Cada cuánto?), la proximidad (¿Cuánto hace?) y la secuencia (¿En qué orden?). Este análisis describe lo que ha pasado a través de la pregunta: ¿Cuántos...? Puede ser cuántos hombres hay, cuántas mujeres, cuántos viven en Miguel Hidalgo, etc.
5. **Análisis inferencial o predictivo:** Hace uso de la estadística inferencial que permite inferir lo que puede pasar en un futuro según la tendencia que ha mostrado la información. Se determina qué podría pasar en el futuro si se continúa con la misma tendencia.
6. **Análisis analítico o prescriptivo:** Se utiliza la estadística analítica para correlacionar diversas variables que a simple vista no tienen un estrecho vínculo. En este marco, se ofrecen, mediante técnicas de analítica predictiva, los problemas y posibles soluciones que se encuentran escondidos en los datos, para identificar patrones de comportamiento.

En este punto, el área directiva selecciona el tipo de análisis que desea llevar a cabo ya sea micro o macro.

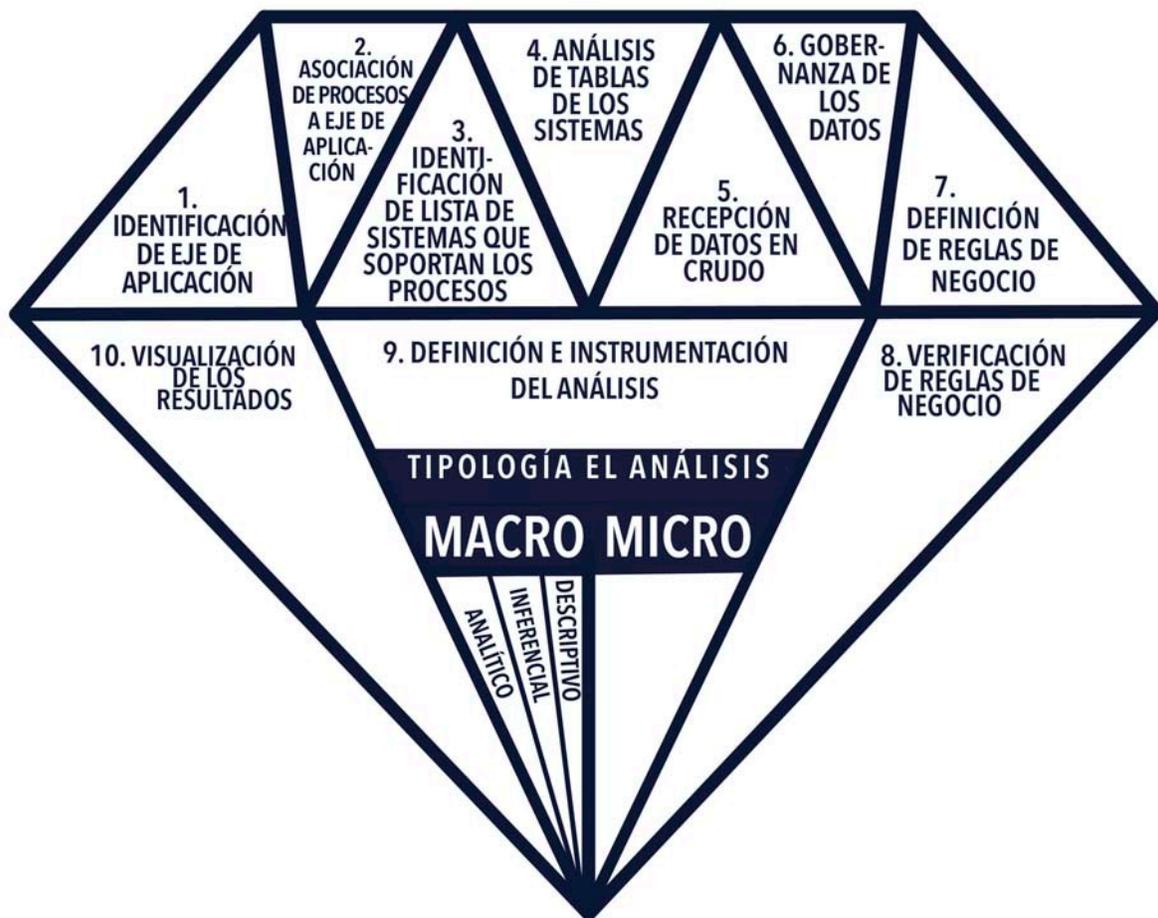
FASE X: Visualización de los resultados

Una vez que se tiene el modelo, se procede a la visualización del modelo. La visualización es una representación digital del modelo para la toma de decisiones. En la representación de los resultados, los tomadores de decisiones pueden estudiar en tiempo real lo que sucede con su negocio.

Los resultados se comunican a través de gráficas las cuales muestran los resultados y recomendaciones que provienen del modelo de Big Data para la toma de decisiones.

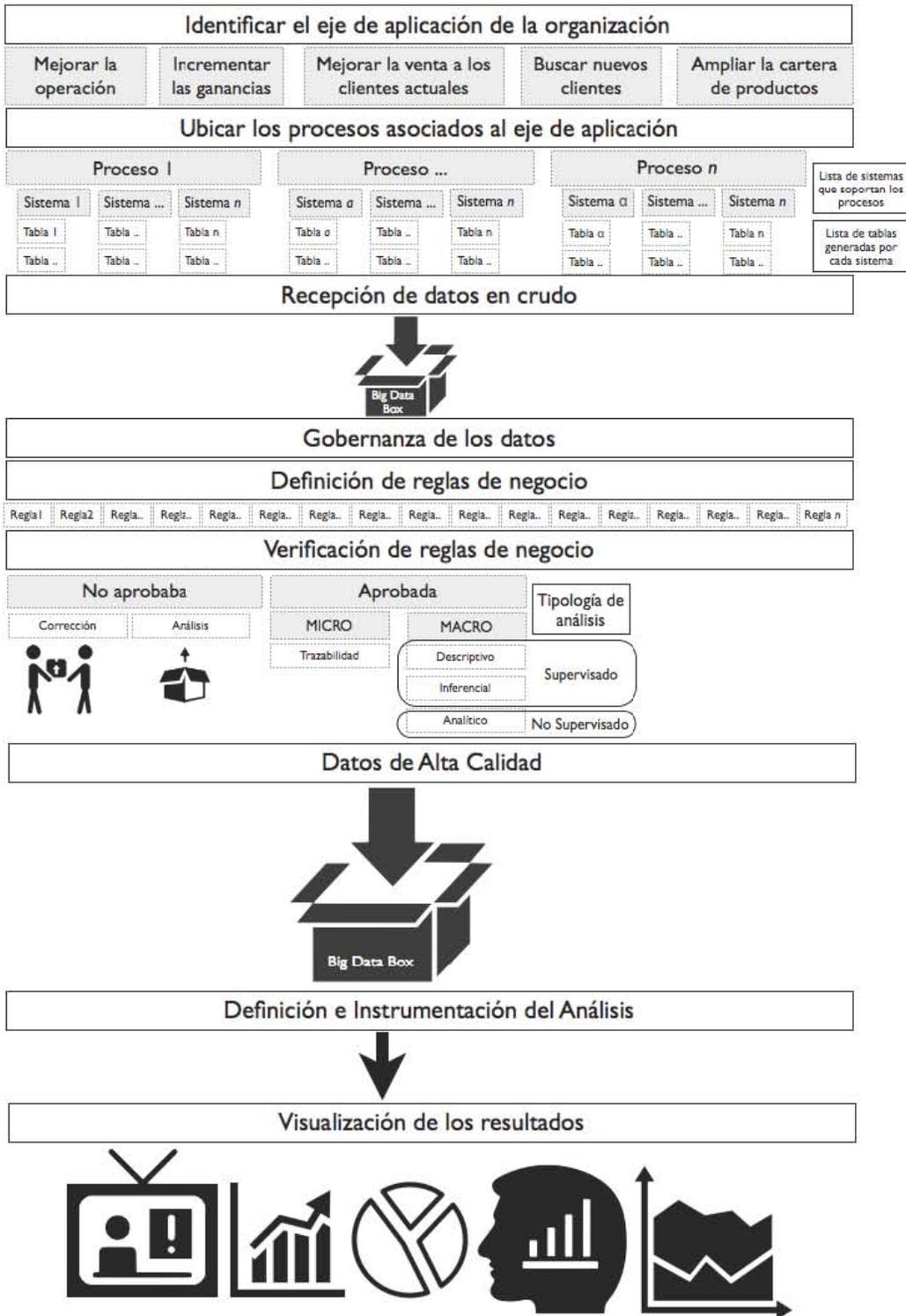
Al complementar todas las fases del modelo propuesto, queda una forma de diamante por lo que se nombra al modelo Big Data Diamond.

Figura 30 Big Data Diamond



Fuente: Elaboración propia basada en las 10 facetas de *data governance* de (Informatica, 2017)

Figura 31 Fases del modelo propuesto



Fuente: Elaboración propia

Bibliografía

- LÓPEZ-MARTÍNEZ, R. E. (15 de 08 de 2016). Conceptos teóricos sobre ICyVT. (N. Guerrero, Interviewer)
- CABAÑERO PISA, C., RODRÍGUEZ ARDURA, I., & SARRADELL LÓPEZ, E. (2002). *Toma de decisiones estratégicas*. Cataluña: Editorial UOC.
- LANEY, D. (2001). *Data Management: Controlling Data Volume, Velocity, and Variety*. Stamford: META Group Inc.
- LAZER, D., & KENNEDY, R. (10 de 01 de 2015). *WIRED*. Retrieved 20 de 03 de 2016 from WIRED: <http://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/>
- LFPDPPP. (05 de 07 de 2010). *CÁMARA DE DIPUTADOS DEL H. CONGRESO DE LA UNIÓN LFPDPPP*. Retrieved 12 de 04 de 2016 from CÁMARA DE DIPUTADOS DEL H. CONGRESO DE LA UNIÓN: <http://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>
- LIMA MACIEL, R. (2012). *Modelo de ingeniería de software con base en directrices de administración del conocimiento*. D.F.: UNAM. Colegio de México. (SR de SR de 2016). *Diccionario del Español de México*. Retrieved 20 de 08 de 2016 from Diccionario del Español de México: <http://dem.colmex.mx/>
- Communications of the ACM. (19 de 03 de 2016). *Communications of the ACM*. Retrieved 01 de 08 de 2016 from Communications of the ACM: <http://cacm.acm.org/magazines/2016/3/198858-peter-naur/fulltext>
- LOSHIN, D. (2013). *Big Data Analytics. From Strategic planning to enterprise integration with tools, techniques, noSQL, and graphs*. Waltman: MK Publications.
- COX, L. A. (2015). *Breakthroughs in Decision Science and Risk Analysis*. Denver: Wiley.
- CRISP-DM. (SR de SR de 1996). *CRISP-DM*. Retrieved 02 de 06 de 2016 from CRISP-DM: <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>
- LUNA GONZÁLEZ, A. C. (2015). *Proceso Administrativo*. Ciudad de México: Grupo Editorial Patria.
- ÁLVAREZ HERNANDO, J. (2011). *Guía Práctica sobre protección de datos*. Valladolid: Lex Nova.
- AGENCE TECHNIQUE DE L'INFORMATION SUR L'HOSPITALISATION ATIH. (SR de SR de 2013). *ATIH*. Retrieved 14 de 02 de 2016 from ATIH: www.atih.sante.fr/mco/presentation
- Agencia Andorrana de Protección de Datos. (SR de 02 de 2014). *Agencia Andorrana de Protección de Datos*. Retrieved 09 de 07 de 2016 from Agencia Andorrana de Protección de Datos: https://www.apda.ad/system/files/2014_2_resolution-big-data-es.pdf
- ANDA GUTIÉRREZ, C. (2004). *Administración y Calidad*. Ciudad de México: Editorial Limusa.
- ANDERSON, C. (23 de 06 de 2008). *WIRED*. Retrieved 01 de 06 de 2016 from WIRED: <http://www.wired.com/2008/06/pb-theory/>

- Big Data Social. (2016). *Big Data Social*. Retrieved 16 de 11 de 2016 from Big Data Social: <http://www.bigdata-social.com/analitica-big-data/>
- BOLSEGUÍ, M., & FUGUET SMITH, A. (06 de 2006). Construcción de un modelo conceptual a través de la investigación cualitativa. *Revista Universitaria de Investigación* , 207-229.
- BOKEH. (2015). *BOKEH*. Retrieved 13 de 12 de 2016 from BOKEH: <http://bokeh.pydata.org/en/latest/>
- Business Insider. (16 de 02 de 2012). The incredible story of how target exposed a teen girl's pregnancy .
- DATAPRIX. (27 de 05 de 2014). *DATAPRIX*. Retrieved 22 de 01 de 2017 from DATAPRIX: <http://www.dataprix.com/blog-it/big-data/big-data-gestion-datos-no-estructurados>
- DAVENPORT, T. (2014). *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*. Boston: Harvard Business School Publishing Corporation.
- DAVENPORT, T., & G.HARRIS, J. (2007). *Competing on Analytics: The New Science of Winning*. Boston: Harvard Business School Publishing Corporation.
- Diccionario LID de Empresa y Economía. (2016). *Diccionario LID de Empresa y Economía*. Retrieved 10 de 11 de 2016 from Diccionario LID de Empresa y Economía: <http://www.diclib.com/equipo%20f%C3%ADsico/show/es/alkonaeconomia/2597#ixzz4QOZtzB27>
- DIAMANDIS, P., & KOTLER, S. (2013). *Abundancia: El futuro es mejor de lo que piensas*. Barcelona: Antoni Bosh Editor.
- Diario Oficial de la Federación. (20 de 02 de 2015). *Diario Oficial de la Federación*. From Diario Oficial de la Federación: http://www.dof.gob.mx/nota_detalle.php?codigo=5382838&fecha=20/02/2015
- Dirección General de Información en Salud. (27 de 01 de 2015). *Dirección General de Información en Salud*. Retrieved 10 de 07 de 2016 from Dirección General de Información en Salud: http://www.dgis.salud.gob.mx/contenidos/basesdedatos/da_boletin.html
- EMC Computer Systems México. (SR de 04 de 2014). *México EMC*. Retrieved 02 de 08 de 2015 from Las oportunidades del Universo Digital: Los Rich Data y el creciente valor del Internet de las Cosas: <http://mexico.emc.com/leadership/digital-universe/2014iview/index.htm>
- EMC MEXICO. (1 de 04 de 2014). *IDC Analyze the future*. Retrieved 20 de 05 de 2016 from IDC Analyze the future: <http://mexico.emc.com/leadership/digital-universe/2014iview/index.htm>
- GALLEGO, J. C. (2010). *PCPI - Mantenimiento de Sistemas Microinformaticos*. España: Editex.
- GARCÍA ACOSTA, G. (2002). *La ergonomía desde la visión sistémica*. Bogotá: Universidad Nacional de Colombia.
- GARCÍA RAMOS, J. A., RAMOS GONZÁLEZ, C. D., & RUIZ GARZÓN, L. (2007). *Estadística empresarial* . Cádiz: Universidad de Cádiz.
- GARTNER Inc. (s/r de s/r de 2016). *Technology Research*. Retrieved 01 de 04 de 2016 from IT Glossary: <http://www.gartner.com/it-glossary/big-data/>
- GUTIÉRREZ, R. (14 de 01 de 2014). Big data: Oportunidades para empresas en 2015 . *Forbes México* , pp. 01-02.

- Harvard Business Review. (2013). *Big Data: The future of Information and Business*. Massachusetts: Harvard Business Review Publishing.
- HOOD L., G. D. (12 de 12 de 2008). *Computing Community Consortium*. Retrieved 01 de 07 de 2016 from CCC-Led White Papers: <http://cra.org/ccc/resources/ccc-led-whitepapers/>
- HURWITZ, J., NUGENT, A., HALPER, F., & KAUFMAN, M. (2013). *Big Data*. New Jersey: John Wiley & Sons, Inc.
- INAI. (2015). *Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales*. Retrieved 01 de 02 de 2016 from Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales: <http://inicio.ifai.org.mx/SitePages/misionViosionObjetivos.aspx>
- Informatica. (01 de 01 de 2017). *Informatica.com*. From Informatica.com: <https://www.informatica.com/solutions/data-governance.html#fbid=tYISBoClaPA>
- JARAMILLO ANTILLÓN, J. (1998). *Principio de gerencia y administración de servicios médicos y hospitales*. San José: Editorial Nacional de Salud y Seguridad Social.
- JOYANES AGUILAR, L. (2013). *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. Ciudad de México: Alfaomega Grupo Editor.
- KOONTZ, H., & WEHRICH, H. (2004). *Administración: Una Perspectiva Global*. Ciudad de México: McGraw-Hill Interamericana.
- McKinsey Global Institute (MGI). (2012). *Big Data: The next frontier for innovation, competition, and productivity*. McKinsey & Company.
- MÜNCH, & MARTÍNEZ, G. (2012). *Fundamentos de la Administración*. Ciudad de México: Trillas.
- MANYIKA, J., CHUI, M., BROWN, B., BUGHIN, J., DOBBS, R., ROXBURGH, C., et al. (SR de 05 de 2011). *Mckinsey&Company*. Retrieved 01 de 08 de 2016 from Mckinsey&Company: <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>
- MARTÍNEZ, J. (23 de 0 de 2012). *Seis pasos para el gobierno de datos*. Retrieved 20 de 11 de 2016 from IBM: <http://www.ibm.com/developerworks/ssa/data/library/techarticle/gobierno-datos/>
- MAYER-SCHÖNBERGER, V., & CUKIER, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Nueva York: Houghton Mifflin Harcourt Publishing Company.
- MICROSOFT. (SR de SR de 2016). *MSDN Library*. Retrieved 01 de 08 de 2016 from MSDN Library: <https://msdn.microsoft.com/es-es/library/ms174949.aspx>
- NICOLETTI, B. (2016). *Digital Insurance: Business Innovation in the Post-Crisis Era*. New York: Palgrave Macmillan.
- NOTIMEX. (02 de 03 de 2014). Esperan inversiones de 215 mdd en tecnología big data en México. *El Financiero*, p. 01.
- Open Data Handbook. (SR de SR de SR). *Open Data Handbook*. Retrieved 20 de 07 de 2016 from Open Data Handbook: <http://opendatahandbook.org/guide/es/what-is-open-data/>
- Organización de las Naciones Unidas. (10 de 12 de 1948). *Declaración Universal de los Derechos Humanos*. Retrieved 22 de 06 de 2016 from Departamento de Información Pública de las Naciones Unidas: <http://www.un.org/es/documents/udhr/>

- Oxford University Press. (s/r de s/r de 2016). *Oxford English Dictionary*. Retrieved 05 de 04 de 2016 from Oxford English Dictionary: <http://www.oed.com/view/Entry/18833#eid301162177>
- PÉREZ LÓPEZ, C., & SANTÍN GONZÁLEZ, D. (2007). *Minería de datos: técnicas y herramientas*. Madrid: THOMSON.
- PÉREZ, G. (2015). *Peligros del uso de los Big Data en la investigación en salud pública y en epidemiología*. Agencia de Salud pública de Barcelona. España: Elsevier.
- PÉREZ, J., & GARDEY, A. (SR de SR de 2008). *Definición.de*. Retrieved 20 de 09 de 2016 from Definición.de: <http://definicion.de/eficiencia/>
- PerTutatis. (20 de 01 de 2015). *PerTutatis*. Retrieved 01 de 02 de 2017 from PerTutatis: <http://pertutatis.cat/que-son-los-sistemas-de-apoyo-a-la-toma-de-decisiones-dds/>
- phpBB. (15 de 09 de 2008). *Administración de Operaciones*. Retrieved 01 de 02 de 2017 from Administración de Operaciones: <http://aspc-admon-op.forosactivos.net/t3-concepto-de-sistema-de-produccion>
- PRITCHARD, A. (1969). Statistical Bibliography or Bibliometrics? *Journal of documentation* , 348-349.
- RAE. (SR de SR de 2016). *Real Academia Española*. Retrieved 20 de 09 de 2016 from Real Academia Española: <http://dle.rae.es/?id=EPVwpUD>
- Real Academia Española. (n.d.). *Real Academia Española*. Retrieved 10 de 10 de 2016 from Real Academia Española: <http://www.rae.es>
- RISKIN, D. (1 de 10 de 2012). *Forbes*. Retrieved 13 de 02 de 2016 from Forbes: <http://www.forbes.com/sites/singularity/2012/10/01/the-next-revolution-in-healthcare/#62f448b46e1e>
- SCHMARZO, B. (2013). *Big Data: El poder de los datos*. Madrid: Anaya Multimedia.
- SÁNCHEZ DÍAZ, A. M. (12 de 06 de 2015). *Siete Días Médicos revista de atención primaria*. From Siete Días Médicos: <http://www.sietediasmedicos.com/zona-franca/bandeja-de-entrada/item/5333-hacia-un-nuevo-paradigma-en-medicina#.V1ZAICPhAb0>
- SÁNCHEZ PÉREZ, G., & ROJAS GONZÁLEZ, I. (2012). Leyes de protección de datos personales en el mundo y la protección de datos biométricos – Parte I. *Seguridad Cultura y prevención para TI* , 04-08.
- SANSONI, S. (06 de 10 de 2015). *SOLON*. Retrieved 15 de 02 de 2016 from SOLON: <http://www.linkedin.com/pulse/solon-la-plateforme-intelligente-dibm-renforce-lutte-contre-sansoni>
- SANZ, E. (2014). *¿Qué es el "Internet de las cosas"?* Retrieved 2016 from Muy Interesante: <http://www.muyinteresante.es/curiosidades/preguntas-respuestas/ique-es-el-qinternet-de-las-cosasq>
- SAS México. (2015). *Gestionando el ciclo de vida analítico para las decisiones a escala*. Ciudad de México: SAS Institute Inc. .
- SAS México. (2015). *Gestionando el Ciclo de Vida Analítico para las decisiones a escala*. Ciudad de México: SAS México.
- SAXSA. (2013). Big Data: Perspectivas y Alcances. *Big Data: Perspectivas y Alcances*. SR.

Secretaría de Finanzas. (05 de 08 de 2016). *Secretaría de Finanzas*. Retrieved 05 de 08 de 2016 from Secretaría de Finanzas:

<http://www.finanzas.df.gob.mx/oip/arco/index.html>

SHORTLIFFE, E., PERREAULT, L., WIEDERHOLD, G., & FAGAN, L. (2001). *Medical Informatics*. New York: Springer Science + Business Media.

SOBRADILLO, P., POZO, F., & AGUSTI, Á. (04 de 09 de 2010). *Archivos de Bronconeumología*. From Archivos de Bronconeumología:

<http://www.archbronconeumol.org/index.php?p=watermark&idApp=UINPBA00003Z&piitem=S0300289610002887&origen=bronco&web=bronco&urlApp=http://www.archbronconeumol.org&estadoitem=S300&idiomaltem=es>

STANTON, J. (2012). *An Introduction to Data Science*. Syracuse: Syracuse University.

Tableau. (2016). *Tableau*. Retrieved 12 de 11 de 2016 from Tableau:

<http://www.tableau.com/es-es>

Teradata . (11 de 04 de 2013). *Teradata* . Retrieved 08 de 06 de 2016 from Teradata : <http://www.teradata.com/News-Releases/2013/BARC-Big-Data-Survey-Lack-of-Experts-and-Know-how-A-Main-Obstacle-to-Monetizing-Big-Data/>

TRIPOD. (2016). *TRIPOD*. Retrieved 10 de 11 de 2016 from TRIPOD:

<http://mineriadetextos.tripod.com/>

TUYA, J., RAMOS ROMÁN, I., & DOLADO COSÍN, J. (2007). *Técnicas cuantitativas para la gestión en la ingeniería de software*. Oleiros: Netbiblo.

Universidad Nacional de Colombia. (18 de 11 de 2016). *Universidad Nacional de Colombia*. Retrieved 18 de 11 de 2016 from Sistema de Información de la Amazonia Colombiana: <http://www.unal.edu.co/siamac/sig/metadatos1.html>

Universidad Politécnica de Cartagena. (2016). *Universidad Politécnica de Cartagena*. Retrieved 16 de 11 de 2016 from Universidad Politécnica de Cartagena: <https://www.upct.es/~gio/trazabilidad.htm>

UnixTime.Info. (18 de 01 de 2011). *UnixTime.Info*. Retrieved 02 de 02 de 2017 from UnixTime.Info:

<http://web.archive.org/web/20110722004843/http://unixtime.info/>

Guión empleado en la entrevista

¿Qué puesto tiene actualmente?

Giro de la organización

Años de experiencia

¿Le fue fácil entender el modelo propuesto?

¿Tiene alguna duda respecto a los conceptos?

¿Qué fortalezas considera que tiene el modelo?

¿Considera que el modelo es aplicable a cualquier organización que maneja grandes volúmenes de datos sin importar el giro?

¿Considera aplicable el modelo a su ámbito laboral?

¿Qué mejoras propondría al modelo?

¿Considera que el modelo favorece la toma de decisiones?