



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**FACULTAD DE ESTUDIOS SUPERIORES
ARAGÓN**

**CREACIÓN DE DATA WAREHOUSE Y BUSINESS
INTELLIGENCE PARA SISTEMA DE ALUMNOS,
ESCUELAS Y MAESTROS**

**TRABAJO ESCRITO
EN LA MODALIDAD DE DESARROLLO DE
UN CASO PRÁCTICO PARA
OBTENER EL TÍTULO DE:
INGENIERO EN COMPUTACIÓN
P R E S E N T A :
MICHEL OMAR ZARCO USCANGA**

ASESORA: MTRA. SILVIA VEGA MUYTOY



FES Aragón

MÉXICO, 2013.

Índice

INTRODUCCIÓN	3
Capítulo1.Problemática.....	5
1.1 Problemática de la SEP.....	5
Capítulo 2. Solución propuesta y diseño	9
2.1 Solución propuesta.....	9
2.1.1 Alcance de los servicios que deben ejecutar las herramientas de la solución integral	9
2.1.2 El número de usuarios y roles que se esperan en los dos primeros años.....	11
2.1.3 Herramientas que se deben a implementar como parte de la solución.....	12
2.2 Diseño.....	18
2.2.1 Almacén de datos.....	18
2.2.2 Modelado de bases de datos	18
2.2.3 Diseño de la información	19
2.2.4 Diseño data mart de Alumnos.....	21
2.2.5 Diseño data mart de Escuelas.....	23
2.2.6 Diseño de los modelos para la creación de los tableros de control de Inteligencia de Negocios	24
3. Desarrollo e Implementación.....	27
3.1 Desarrollo	27
3.1.1 Desarrollo DWH.....	27
3.1.2 Desarrollo de Inteligencia de negocios	30
3.1.2.1 Página Principal	33
3.1.2.2 Nomina	34
3.1.2.3 Alumnos	37
3.1.2.4 Escuelas	41

3.2 Implementación	43
3.2.1 Seguridad en los servidores de producción	43
4. Consideraciones y reestructuración del modelo.....	45
4.1 Modificaciones al modelo del DWH	45
4.1.1 Esquema DWH_STAGE	46
4.1.2 ESQUEMA DWH_ROLAP	47
4.1.3 ESQUEMA DWH_MOLAP.....	50
4.2 Flujo de información	51
4.3 Actualización de tableros de control.....	52
4.3.1 Construcción de los tableros de control	52
4.3.2 Lista de fuentes empleadas.....	52
4.3.3 Documentación de las reglas utilizadas	54
4.3.4 Vistas	57
Conclusión	65
Glosario.....	66
Referencias.....	69

INTRODUCCIÓN

La Secretaría de Educación Pública a través de la Unidad de Planeación y Evaluación de Políticas Educativas (UPEPE), requiere un Modelo de Operación para generar, explotar y mantener actualizada la Base de Datos del RNAME: Registro Nacional de Alumnos de Educación Básica, Registro Nacional de Maestros y El Registro Nacional de Escuelas.

La misión, visión y objetivos de la UPEPE son las siguientes:

MISIÓN

Coordinar el desarrollo de las acciones estratégicas del Programa Sectorial de Educación y el Sistema Nacional de Evaluación Educativa orientando los esfuerzos al cumplimiento de los objetivos de la Institución.

VISIÓN

Articular el Sistema Nacional de Evaluación Educativa, desarrollando las mejores estrategias, acciones y prácticas que contribuyan a la mejora de resultados en el sistema educativo, generando así la formulación de políticas educativas.

OBJETIVO GENERAL

Fortalecer el proceso general de planeación y evaluación del Sistema Educativo Nacional, mediante el mejoramiento de las capacidades del personal y de las técnicas y herramientas que se utilizan en las diferentes unidades responsables del sector, relacionadas con estos procesos, así como el impulso de las innovaciones y modernización de estos procesos.

OBJETIVOS ESPECÍFICOS

Promover las mejoras en el ejercicio de planeación, con el fin de que los resultados de ésta, se reflejen en la formulación de las políticas y en la integración del presupuesto para el mejoramiento del Sistema Educativo Nacional (SEN).

Crear las condiciones adecuadas para propiciar una participación más efectiva de todos los actores en el proceso de planeación y evaluación, tanto a nivel federal como estatal.

Mejorar la capacidad de respuesta en las demandas de información, tanto al interior del sector como a la población abierta, mediante el mejoramiento de la calidad y oportunidad de la información, así como mediante procesos de modernización de métodos y procedimientos.

Impulsar la operación del Sistema Nacional de Evaluación Educativa que permita medir con objetividad los avances del sector y el impacto de las políticas públicas, así como la atención eficiente y eficaz para una mayor transparencia y rendición de cuentas.

Para la UPEPE la visión del RNAME se sintetiza en:

“Contar en el año 2012 con la información detallada de Alumnos, Maestros, Escuelas y Plazas Federalizadas, integrada en una plataforma tecnológica única que facilite el acopio, la administración y la distribución de información estratégica del Sistema Educativo Nacional. Con la integridad, consistencia y seguridad suficientes, para dar soporte a los requerimientos de información de forma oportuna y congruente. Convirtiéndola en “la referencia” para los interesados en datos del sector; para producir información estadística a partir de sistemas registrales; para apoyar los procesos de planeación y toma de decisiones; y para fortalecer y propiciar las mejores prácticas en términos de comunicación y transparencia ante la sociedad.”

La iniciativa de la UPEPE, motivo del presente proceso, contempla la integración de la plataforma tecnológica (licencias de herramientas) y servicios de procesamiento de la información así como el desarrollo e implementación de las soluciones tecnológicas que permitan:

1. La extracción de Información de Alumnos, Maestros, Escuelas y Plazas proveniente de diversas fuentes (Disponibles en Archivos con formato de texto plano, en bases de datos relacionales o en sistemas operando en línea)
2. El procesamiento de la información mediante herramientas de calidad de datos (higiene, estandarización, des-duplicación y consolidación de registros)
3. La integración de registros únicos de Alumnos, Maestros, Escuelas y Plazas en herramientas de administración de datos Maestros
4. La integración de una base de datos Nacional del RNAME
5. La creación de un Data Warehouse; al cual se dará acceso a las entidades federativas para la integración, validación y seguimiento de la información reportada sobre plazas y erogaciones, así como a la información estadística de alumnos, maestros y escuelas y el detalle de los registros que la componen con diferentes niveles de agregación.
6. El análisis de la información consolidada a través del uso de tableros de control, mapas geográficos y herramientas para realizar minería de datos.

Aunque en el capítulo 2 del documento se describirá el proceso completo del sistema, el trabajo va a estar enfocado en la parte del data warehouse y de la herramienta de Business Intelligence.

Para ello en el Capítulo 1 hablaremos de la problemática que tenía la SEP y la UPEPE.

En el Capítulo 2 hablaremos de la solución propuesta y el diseño de ésta.

En el Capítulo 3 hablaremos del desarrollo y la implementación de la solución.

Y en el Capítulo 4 hablaremos de las modificaciones que se le realizaron al diseño.

Capítulo 1. Problemática

1.1 Problemática de la UPEPE

A continuación se presenta un extracto de la licitación que hizo la SEP y la UPEPE para la creación del Registro Nacional de Alumnos, Maestros y Escuelas (RNAME), donde se plantea la problemática que la institución tenía.

“La SEP requiere de la propuesta de la plataforma tecnológica y de servicios para la integración y explotación de la información del Registro Nacional de Alumnos (Educación Básica), Registro Nacional de Maestros (que integra el Registro Nacional de Plazas) y El Registro Nacional de Escuelas; y que sea la base para la Estrategia de Inteligencia de Negocios en toda la información que utiliza la UPEPE.

Las razones que motivan este proyecto son la multiplicidad de fuentes de información para concentrar la información solicitada por el Artículo Noveno del Presupuesto de Egresos de la Federación y la integración de los registros de maestros, alumnos y escuelas.

La SEP integra la información de alumnos maestros y escuelas proveniente de las 32 entidades federativas, por lo que existe una gran diversidad de plataformas tecnológicas y formatos en los que se almacena la información, esta información deberá someterse a procesos de higiene, estandarización y deduplicación de la información

Cada entidad federativa maneja su propio sistema de bases de datos, por lo cual no se tiene un control adecuado sobre la información de alumnos, escuelas y maestros. Esto genera problemas ya que con ese sistema es difícil detectar irregularidades, un ejemplo sería ubicar maestros que cobren por plazas en diferentes estados de la república al mismo tiempo. Para esto se necesita una base de datos única a nivel nacional.

Para esto, la UPEPE considera la arquitectura de la figura 1.1, basada en mejores prácticas, para satisfacer la consulta de Información dentro de la Dependencia.

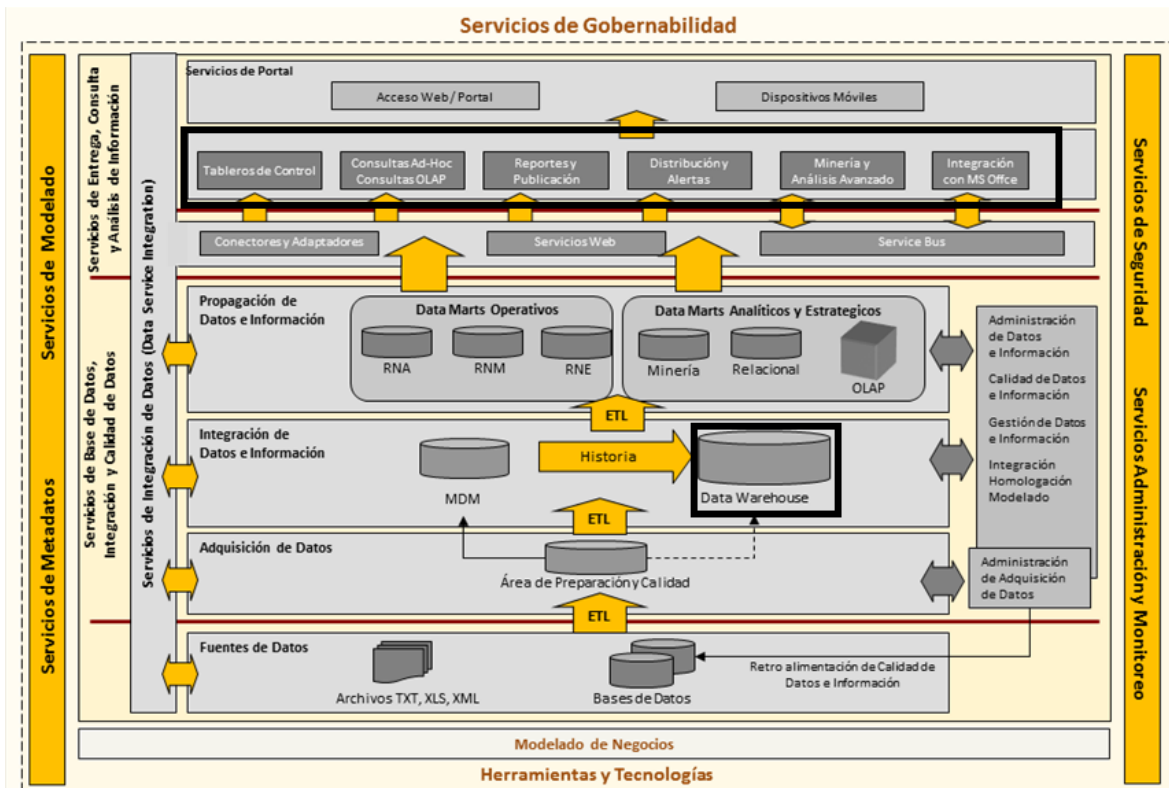


Figura 1.1 Arquitectura requerida del sistema, lo resaltado dentro de rectángulos es en lo que enfocaremos este documento.

Los objetivos principales de este proyecto se resumen en los siguientes puntos.

- Se requiere contar con una plataforma tecnológica para la constitución, integración y explotación de la información del Registro Nacional de Alumnos, Maestros y Escuelas (RNAME), que en su vertiente de maestros, comprende el Registro Nacional de Plazas.
- Que dicha plataforma tecnológica propuesta sirva de base para la Estrategia de Inteligencia de Negocios en toda la UPEPE.
- Del mismo modo, esta plataforma tecnológica debe ser capaz de cubrir los crecientes requerimientos de recepción, limpieza, generación, explotación y análisis de información de la dependencia.
- La propuesta deberá basarse en la arquitectura referida y proponer la combinación óptima de herramientas para cubrir el requerimiento de los servicios.
- Se debe lograr la integración, validación y seguimiento de la calidad de la información reportada sobre Alumnos, Maestros, Escuelas, Plazas y Erogaciones que permita la consulta, evaluación y fiscalización, a través de una exhaustiva revisión de la calidad de datos.

- Esta revisión de la calidad de datos debe incluir procesos de estandarización de catálogos, tipo de datos y tablas que garanticen la confiabilidad de la información eliminando errores y/o duplicidad de registros.
- Estandarizar la forma en la que los usuarios estarán accediendo la información permitiendo que los usuarios realicen consultas eficientes y fáciles de construir.
- La solución deberá permitir crear, consultar y modificar reportes ad-hoc, tableros de control y alertas que puedan ser accedidos mediante:
 - Un portal web
 - Documentos de uso cotidiano como son MS Office
 - Exploradores de datos
 - Paneles de visualización para datos multidimensionales
 - Dispositivos móviles
 - Ser consumidos por WebServices

En la figura 1.2 se presenta la infraestructura que se tiene a disposición para el proyecto.

Es importante comentar que la Capa de Procesamiento es virtualizada, por lo que los servidores de Data Quality y de MDM, si en la solución que se preponga para estas herramientas están integradas, estos equipos pueden unirse para conformar uno solo.

Por otra parte, la capa de presentación incluye dos servidores WEB a fin de tener un balanceo de cargas de las peticiones al portal, y se agrega un servidor de aplicaciones a fin de considerarlo para las posibles aplicaciones que soporten la operación de la solución y que el proveedor licencie o desarrolle, estos equipos también son virtualizados

La capa de servicios de datos también es virtualizada, la distribución de aplicaciones en cada uno de los dos servidores (DWH, BI, RDBMS y MDM) es la propuesta por la UPEPE; sin embargo, puede ésta distribuirse de la forma que el proveedor de la solución lo requiera; siempre y cuando, no implique cambios en la arquitectura mostrada.

A fin de garantizar la operación de las diversas herramientas que los licitantes propongan, toda la arquitectura es virtual basada en procesadores de 64 Bits y se contempla la posible instalación de cualquiera de los siguientes sistemas operativos: Windows Server 2008 R2, LINUX compatible con RHEL 5 o UNIX 64 Bits, en todo caso el licitante deberá manifestar en cuál de las plataformas opera cada uno de sus productos.

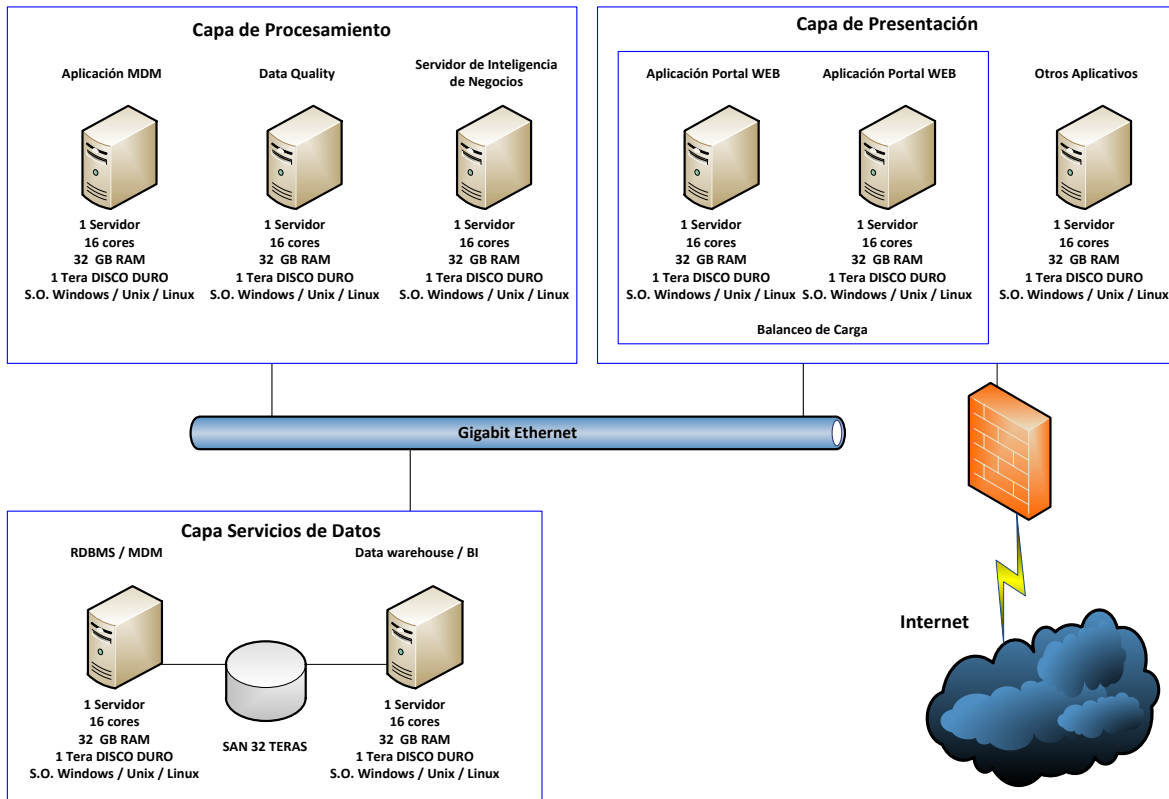


Figura 1.2 Infraestructura disponible”

Para esta licitación participaron varias empresas pero fue la empresa donde yo trabajaba la que presentó la mejor propuesta de solución y gano la licitación para realizar el proyecto utilizando herramientas de Oracle.

Mi puesto dentro de la empresa era la de consultor de la Herramienta de Oracle Business Intelligence, para este proyecto se asignaron a más de 20 consultores, pero el equipo de trabajo específico al área de BI (Business Intelligence) fue de 3 personas.

Como el equipo de trabajo de BI trabajaba directamente con el área de Data Warehouse, se tuvo cierta participación con el diseño del Data Warehouse ya que sin un buen diseño los resultados obtenidos con la herramienta de BI serian incorrectos

Capítulo 2. Solución propuesta y diseño

En este capítulo veremos la solución que propuso la empresa donde yo trabajaba y que abarca todo el proceso de solución que se vio en la figura 1.1, además de los alcances que deben tener las herramientas de Oracle que mencionamos en el capítulo anterior, más adelante se aclarara en que parte de todo el proyecto tuvo participación y en eso se basará el presente trabajo.

2.1 Solución propuesta

La propuesta que dimos proporciona una solución de implementación de herramientas de integración calidad, explotación y publicación de información, que contemplan:

- Licenciamiento de las Herramientas de Software de Oracle que satisfagan los servicios solicitados.
- Configuración y adecuación de las herramientas para la implementación de un proceso automatizado que cubra desde la entrada de datos, calidad, limpieza, administración de datos maestros, hasta la incorporación a un Data Warehouse y la explotación mediante herramientas de inteligencia de negocio y publicación en web.
- Servicios de procesamiento de la información
- Servicios de Transferencia Tecnológica.

2.1.1 Alcance de los servicios que deben ejecutar las herramientas de la solución integral

La SEP a través de la UPEPE determinó los alcances que deben de tener cada una de las herramientas que se solicitan en la licitación presentada en el capítulo anterior, a continuación se muestran los alcances que debían cubrir cada una de las herramientas:

- **Servicios de Integración de Datos (Data Service Integration)**

Estos servicios realizan principalmente: la adquisición e integración de Datos de las diferentes fuentes de información de Alumnos, Escuelas, Maestros y Plazas así como la población de las distintas estructuras Diseñadas, entre ellas las correspondientes a las requeridas por las diferentes herramientas de limpieza de Datos, Maestro de Datos, Data Warehouse e Inteligencia de Negocio.

Resulta esencial contar con un ETL para poder obtener, integrar y entregar datos almacenados en diversas fuentes, cubriendo los requerimientos solicitados para este tipo de servicios en la sección correspondiente de esta licitación. Lo anterior permitirá mejorar los tiempos de procesamiento al crear procesos automatizados que garanticen la total fiabilidad de los datos.

- **Servicios de Calidad de Datos (Data Quality)**

La solución que se proporciona permite la definición de reglas de negocio para evitar la integración de datos inconsistentes, incorrectos, incompletos o ambiguos,

Para ello se deberán proporcionar los catálogos que sirven para guiar a los procesos de calidad de datos, tales como: Nombres, Apellidos, Calles, tipo de Vialidad, Colonias, Delegaciones, Municipios, Asentamientos y Tipos de Asentamiento, así como un conjunto de reglas iniciales que guíen los procesos de calidad sobre nombres y domicilios de personas en México.

- **Servicios de Administración de Datos Maestros (MDM)**

Master Data Management (MDM) ofrece a los procesos de recopilación, agregación, consolidación y aseguramiento de la calidad, la persistencia y difusión de dichos datos en toda la organización para asegurar la consistencia y el control en el mantenimiento y uso de la aplicación de esta información.

La aplicación de las soluciones tecnológicas especializadas en MDM, permitirá gestionar y consolidar los datos maestros dispersos, ofreciendo una visión única de alumnos, maestros, escuelas y plazas.

La principal fortaleza que se busca es eliminar información duplicada generando referencias a registros únicos, facilitando el manejo de los datos desde una visión estratégica e institucional.

El MDM permitirá garantizar que no se utilicen múltiples versiones de los mismos datos maestros.

El funcionamiento de un MDM implica una integración de los datos de identificación y ubicación de los datos de Alumnos, Maestro, Escuelas y Plazas provenientes de diferentes fuentes.

Se tiene la necesidad de contar con un repositorio de los datos maestros que permita identificar de forma única a la misma persona y que incorpore mecanismos de integración y consolidación de datos provenientes de diferentes sistemas y fuentes de información basados en reglas del negocio sobre la persistencia, precedencia y actualización del registro maestro.

Por otra parte la herramienta mantiene un registro del origen de los datos y puede rastrear en todo momento de donde han surgido y los cambios realizados.

- **Servicios de Repositorio de Datos (Database)**

Para este servicio debe considerarse el almacenamiento de:

- Una base de datos relacional que integre la información proveniente de las diferentes fuentes de información en la base de datos nacional RNAME en un Instancia Nacional.
- Un área de base de datos de staging que reúna las bases de datos provenientes de las diferentes fuentes de información.
- Un área de preparación de datos.
- Una base de datos de Data Warehouse.

- **Servicios de Inteligencia del Negocio**

Este servicio contempla de manera general:

- El análisis de la información consolidada a través del uso de tableros de control, mapas geográficos y herramientas para realizar minería de datos.
- Estandarizar la forma en la que los usuarios estarán accediendo la información permitiendo que los usuarios realicen consultas eficientes y fáciles de construir.
- Creación, consulta y modificación de reportes ad-hoc tableros de control y alertas que puedan ser accedidos mediante:
 - Un portal web
 - Documentos de uso cotidiano como son MS Office
 - Exploradores de datos
 - Paneles de visualización para datos multidimensionales
 - Dispositivos móviles
 - Ser consumidos por WebServices

- **Servicios de Entrega y Consulta de Información**

Este servicios servirá como acceso a la información obtenida del sistema, cumplirá con: desplegar información importante del mismo, analíticos provenientes de éste, así como guardar documentos importantes, administración y segregación de roles de usuarios para consulta de la información. Así como la integración con fuentes o sistemas que complementen la información manejada por la solución.

2.1.2 El número de usuarios y roles que se esperan en los dos primeros años

Mediante de un estudio hecho por la UPEPE y con la ayuda del gerente del proyecto de la empresa donde yo trabajaba y tomando en cuenta el número de entidades de la republica que existen además del número de usuarios que tendrían que usar cada herramienta dio como resultado los siguientes números de usuarios y roles:

Herramientas de Bases de Datos:

- 1 Súper Usuario
- 3 Administradores
- 3 Respaldo y recuperación
- 15 Personas de Desarrollo
- 100 Personas acceso total a la base de datos
- 20 Cuentas conexión de perfil público para consulta en Internet

Herramientas de Data Warehouse e Inteligencia de Negocio:

- 1 Súper Usuario
- 3 Administradores
- 5 Personas de Desarrollo
- 100 Personas acceso total (reporteo y Drill Down)
- 20 Cuentas conexión de perfil público para consulta en Internet

Herramientas de Calidad de Datos y Administración de Datos Maestros:

- 5 desarrolladores/configuradores
- 10 Personas corriendo Procesos
- 25 Personas resolviendo inconsistencias

2.1.3 Herramientas que se deben implementar como parte de la solución

Estas herramientas de Oracle deben ejecutar lo siguiente:

- Servicios de Integración de Datos

En la tabla 2.1 se pueden ver los servicios que deben proveer los diferentes componentes que integren la solución de integración de datos:

REQUERIMIENTO
Extracción y entrega de información desde y hacia diversas fuentes: bases de datos y/o archivos planos.
Procesamiento en tiempo real y batch para integración de datos.
Conectividad con diferentes tipos de bases de datos.
Flujos de extracción, transformación y carga

Flujos de validación de datos en el punto de entrada
Transformación de información a través de: Filtros Sumarizaciones Agregaciones Desnormalización Consolidación Generación de métricas derivadas
Capacidad de definir, suscribir, modificar y lanzar eventos y alertas.

Tabla 2.1 Requerimientos de integración de datos

- **Servicios de Administración de Datos Maestros y Calidad de Datos**

La intención de la Dependencia es contar con una base de datos consolidada y homologada que pueda operarse institucionalmente y analizarse conjuntamente con otras fuentes de información del Sector Educativo.

Por lo anterior es necesaria la inclusión de una herramienta informática para la Administración de Datos Maestros Centralizada (MDM). Esta herramienta deberá:

- Permitir la generación y mantenimiento de un registro único para Alumnos, Maestros, Escuelas y Plazas que sirva de referencia y pueda centralizar la información de todas las entidades federativas y que provienen de una o más fuentes de información.
- Administrar información de Alumnos y Maestros como personas físicas que pueden estar asociadas a una Escuela.
- Ofrecer funcionalidades de extracción de los datos de referencia de las diferentes fuentes y cargarlos en el Padrón Único centralizado.
- Representaciones consistentes de los datos compartidos.
- Información de alumnos y maestros limpia, acertada y desduplicada.
- Integración con las herramientas de calidad de datos en su ciclo completo.
- Una vista completa de 360° (Cubos de información)
- Capacidades completas de integración para adquirir, sincronizar y federar datos.

La herramienta MDM soporta la creación de un modelo de datos con las características presentadas en la tabla 2.2

REQUERIMIENTO
La definición de entidades tipo "Alumno" y "Profesor"; o bien una entidad "Sujeto", "Persona" o equivalente que permita diferenciar a los alumnos de los maestros
La definición de una entidad tipo "Escuela", "Propiedad", "Inmueble", "Activo" o equivalente que permita modelar a los Centros de Trabajo (en general corresponden a "Escuelas")
La definición de una entidad tipo "Cuenta", "Plaza" o equivalente que permita el modelar plazas.
Modelado de la relación entre las entidades "Maestros", "Alumnos" y "Escuelas"
Modelado de la relación entre las entidades "Maestros", "Plazas" y "Escuelas"
Modelado de la relación entre las entidades "Escuelas" y "Maestros"
Modelado de la relación entre las entidades "Maestros" y "Alumnos"
Capacidad de extender el modelo permitiendo la configuración de la información a la medida de las necesidades y características de cada grupo de datos maestros (Alumnos, Maestros, Escuelas, Plazas)

Tabla 2.2 Requerimientos de administración de datos maestros

- **Calidad de adquisición de datos (Data Quality)**

La tabla 2.3 muestra los requerimientos para la calidad de adquisición de datos

OBJETIVOS
Validación de campos vacíos o con valores nulos
Espacios en blanco en campos obligatorios
Frecuencia y porcentaje de falta de información para determinar si se efectúa la carga de la información o se rechaza.
Frecuencia y valores de datos capturados para determinar si se efectúa la carga de la información o se rechaza.

Tabla 2.3 Requerimientos de calidad de adquisición de datos

En todos estos casos se requiere el manejo de cifras de control que permitan conocer la calidad de la información enviada para decidir, con ésta u otras herramientas si los archivos se reciben o se rechazan.

- **Base de Datos**

La solución cuenta con un manejador de base de datos que permite maximizar la disponibilidad del sistema, confiable y que permite realizar un gran número de actividades de administración en línea, brindando estabilidad y seguridad de operación, la tabla 2.4 muestra los requerimientos de la base de datos.

OBJETIVOS
El manejador de base de datos es el mismo como motor relacional y multidimensional.
La base de datos opera en el ambiente descrito en la plataforma de infraestructura de que dispondrán las herramientas propuestas para la licitación presentada
El manejador de base de datos permite portabilidad binaria entre múltiples plataformas de cómputo
Capacidad del manejador de base de datos de manejar un modelo de concurrencia de usuarios que consulten y actualicen información al mismo tiempo garantizando el desempeño de las transacciones y la precisión de la información.
Se requiere que el motor de base de datos disponga de un Diccionario de datos nativo en la propia herramienta.
Se requiere que el manejador de Base de Datos permita la administración de los recursos utilizados por ésta (CPU y memoria), basada en las prioridades de acceso determinadas para cada usuario
Se requiere que el ambiente de administración en una interfaz gráfica.
Capacidad para monitorear el comportamiento de cualquier objeto de la base de datos.
Capacidad de las herramientas de administración para asegurar alta disponibilidad de las aplicaciones simplificando las tareas de diagnóstico, por lo que debe permitir el monitoreo de las n instancias del cluster como una misma instancia de la base de datos.
Capacidad de la base de datos para permitir: detectar y ajustarse en cambios dinámicos en CPU y memoria.
Se requiere que el motor de base de datos disponga de un esquema de persistencia de los resultados de una consulta sin necesidad de la definición de un objeto por cada consulta, evitando sobrecarga (accesos innecesarios).
Soporte de la base de datos para ser creada con múltiples tamaños de bloque de sistema operativo.
Capacidad de la base de datos para mejorar el desempeño en las transacciones detectando y limitando el consumo de recursos (cpu y memoria) en un tiempo específico
Capacidad de crecimiento para soportar clúster de sistema operativo y de base de datos y operar en clúster activo-activo.
Se requiere que la Base de Datos permita realizar el particionamiento lógico de tablas (sub-tablas) basado en la definición de criterios.
Capacidad del manejador de Base de Datos para reconstrucción de índices en línea
Capacidad de la base de datos para realizar cambios a la estructura de tablas en línea (columnas pueden ser agregadas o borradas).
Capacidad para segmentar volúmenes de datos y de índices.
Conectividad a la Base de Datos con cualquier otro motor de bases de datos incluyendo ODBC, OLE-DB y JDBC
Capacidad para soportar procedimientos almacenados

Tabla 2.4 Requerimientos de la base de datos

- Inteligencia de negocios (Business intelligence)

Esta es la parte del proyecto en la que yo estuve directamente involucrado y en los siguientes capítulos es lo que se verá, esto es la parte que se encuentra en los rectángulos en la figura 1.1

La tabla 2.5 muestran los servicios de Inteligencia de Negocio que se requieren:

REQUERIMIENTO
La solución efectúa su autenticación y autorización empleando directorios compatibles con el estándar LDAP en su versión 2 o 3.
La solución deberá estar certificada para operar en la plataforma de infraestructura de que dispondrán las herramientas propuestas en licitación.
Soporte a tableros que permitan realizar análisis de manera sencilla e interactiva.
Capacidad para desarrollar Tableros de Control que brinden un vistazo rápido a la información y que provean la funcionalidad de navegación (drill) a detalle de la información dentro de la misma plataforma.
Deberá de permitir la exploración intuitiva de resultados sin necesidad de conocer el modelo de datos, las búsquedas se realizarán a partir de criterios que puedan soportar textos comunes.
Deberá tener la capacidad de desplegar detalles relevantes de la información buscada (capacidad de navegar en las diferentes jerarquías de la información)
Generación de gráficas para representar la información buscada, lo que ofrecerá versatilidad de acceso e interpretación de los datos.
Capacidad para compartir los resultados vía e-mail.
Capacidad de guardar la consulta en Web y posteriormente compartir el resultado de las búsquedas con tan solo una liga.
El motor OLAP deberá ser lo suficientemente flexible para que el usuario pueda hacer una navegación o exploración de los datos.
Capacidad para reutilizar el contenido de la exploración de datos relacionales, no relacionales y OLAP.
La herramienta deberá permitir el uso de mapas geo referenciados y su visualización a través de Google Maps o ESRI de forma nativa

Tabla 2.5 Requerimientos de inteligencia de negocios (Business Intelligence)

- Entrega y consulta de información

Se requiere de una interfaz que cumpla con las siguientes características (ver tabla 2.6):

REQUERIMIENTO
Permitirá la navegación entre diferentes páginas donde se desplegará la información (uso de breadcrumbs) informando así al usuario que ubicación tiene en ese momento.
Deberá realizar como parte de la solución, la conversión de imágenes a un formato ligero para que este pueda desplegarse fácilmente en un navegador.
Contará con herramientas para que se puedan crear menús de acceso rápido a diferentes páginas. Ajustando los accesos a la información de acuerdo al perfil definido.
Permitirá hacer “refresh” por componentes específicos dentro de cada una de las páginas con información (no actualizar toda la información, únicamente aquella en donde se esté haciendo énfasis).
Permitirá conocer cuáles son las páginas más visitadas y generar recomendaciones a los usuarios, basadas en los indicadores que se monitorean.
Permitirá el manejo de los documentos tipo Microsoft Office y existirá la alternativa de conversión de archivos a formato HTML y PDF.
Contará con la facilidad para personalizar los elementos de la estructura visual, modelos, colores, fuentes, imágenes e iconos.
Cada página podrá tener sus propios portlets y estos podrán ser añadidos o eliminados por aquellos usuarios que estén autorizados para tal fin.
La solución deberá permitir transformar contenido de múltiples formatos a formatos visibles en web: HTML, XML, PDF o formatos de Microsoft Office.
Permitirá la integración con desarrollos de otras plataformas: Java, .Net, Sharepoint y aplicaciones de terceros basados en estándares internacionales.
La solución deberá permitir relacionar reportes entre sí, manejar links al contenido y dirigir la navegación entre contenidos.
La solución deberá permitir al administrador de contenido calendarizar la fecha de publicación y expiración del mismo (gráficas, reportes, documentos).
Permitirá la creación de portlets.

Tabla 2.6 Requerimientos de entrega y consulta de información

Conforme a los requerimientos vistos en este punto, se analizaron los objetivos que debían cumplir las herramientas de Business Intelligence y el Data Warehouse, y se dio paso a la elaboración del diseño que cumplan esos objetivos marcados, esto se verá en el punto 2.2

2.2 Diseño

Esté punto y posteriores capítulos estarán dedicados únicamente a la parte del Almacén de datos e Inteligencia de Negocios (Data Warehouse and Business Intelligence), ya que es la parte del proyecto que interesa a este documento.

2.2.1 Almacén de datos

Un Data Warehouse (DWH) es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta. La creación de un data warehouse representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de Inteligencia de Negocios.

La ventaja principal de este tipo de bases de datos radica en las estructuras en las que se almacena la información (modelos de tablas en estrella, en copo de nieve, cubos relacionales... etc). Este tipo de persistencia de la información es homogénea y fiable, y permite la consulta y el tratamiento jerarquizado de la misma (siempre en un entorno diferente a los sistemas operacionales).

Tomar decisiones teniendo como base información confiable, es uno de los objetivos para la UPEPE, por lo que se requiere de la implementación de un Data Warehouse (DWH) que integre información relevante de la Nómina de los maestros, resultados de la prueba Enlace a los alumnos de primaria y secundaria y programas de ayuda que tienen las escuelas. Que sirva como plataforma para la explotación de información de Plazas, Alumnos, Maestros y Escuelas, mediante Datamarts o Cubos de Información.

2.2.2 Modelado de bases de datos

Para la construcción de un DWH existen varios esquemas de desarrollo, los más comunes son Tipo Estrella y Copo de Nieve; cada uno de ellos cuenta con diferentes características que dependiendo del volumen y detalle de información que se requiera manipular se elegirá el modelo adecuado.

Para el modelo de información de la SEP, se utilizó el esquema Tipo Estrella, es decir, se tuvo una visión multidimensional de un proceso que se mide a través de varios indicadores. A nivel de diseño, consiste en una tabla de hechos (Facts) en el centro de la estructura, y varias tablas de dimensión, por cada dimensión de análisis que participa de la descripción de ese hecho.

En la tabla de hechos se encuentran los atributos destinados, mientras que en las tablas de dimensión, los atributos se destinan a elementos de nivel (que representan los distintos niveles de las jerarquías de dimensión) y a atributos de dimensión (encargados de la descripción de estos elementos de nivel).

2.2.3 Diseño de la información de nómina

En esta sección del documento, se describe el modelo general de Data Warehouse que se diseñó para la SEP.

Se diseñó basándose en layouts o archivos de Excel que fueron entregados por la SEP para su análisis, esta información es referente a Nómina, Trabajadores comisionados y con licencia. Cada tabla de hechos viene especificada con sus dimensiones y la imagen del diseño en estrella resultante.

A continuación se muestra el ejemplo de la tabla de hechos llamada “Fact_OrdinariaExtraor”, la cual trae información de las percepciones y deducciones de los trabajadores. También se muestran las dimensiones que tienen esta tabla de hechos y el diagrama de la estrella resultante.

Cabe mencionar que las demás tablas referente a nómina fueron hechas de la misma manera.

Tabla de Hechos: Fact_OrdinariaExtraor

Su principal objetivo es determinar el número total de percepciones, deducciones y montos de las mismas y está dimensionado por las siguientes tablas (ver diagrama 2.1):

- Tipo Nómina= Se refiere al tipo de nómina al que corresponde el pago del trabajador
- CCT= El centro del trabajo donde laboró el trabajador
- Empleado= Los datos del trabajador que recibe el pago
- Pagaduría= Se utiliza para realizar las aportaciones al ISSSTE y al SAR
- Nivel de Puesto= Describe el nivel de puesto del trabajador
- Zona Económica= Clasificador del territorio nacional que considera el tabulador de sueldos
- Fuente de financiamiento= Origen del financiamiento por concepto de pago
- Pago retroactivo= Describe si el pago es retroactivo o no
- Entidad= Entidad donde laboró el trabajador
- Tiempo= Fecha del pago

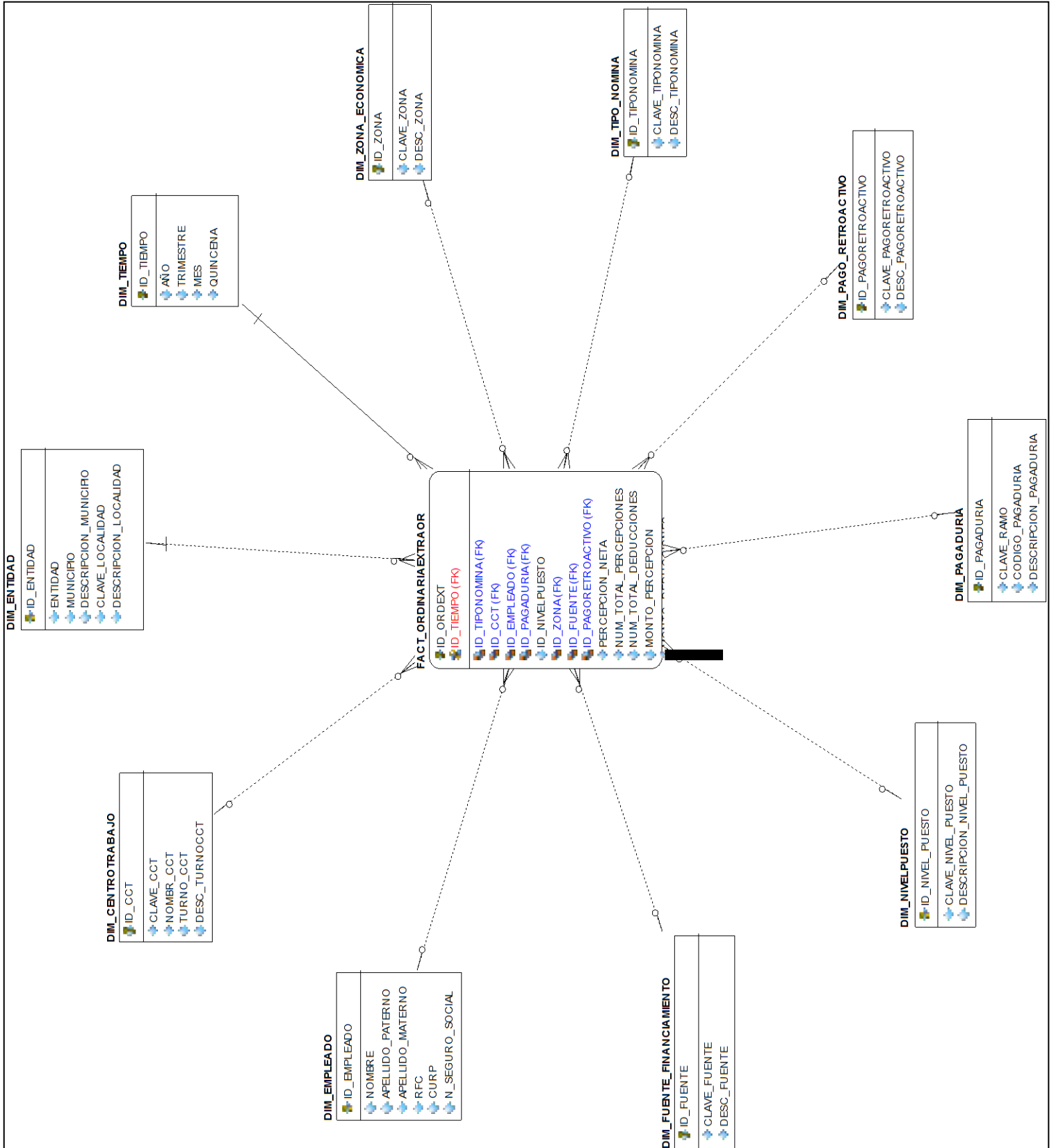


Diagrama 2.1 Esquema Tipo Estrella Fact_OrdinariaExtraor

2.2.4 Diseño data mart de Alumnos

Un Data mart es una versión especial de almacén de datos (data warehouse). Son subconjuntos de datos con el propósito de ayudar a que en un área específica dentro del negocio se pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades.

Se diseñó un data mart de la información de Alumnos proporcionada por la SEP

Diseño de información

Descripción de tablas entregadas por la SEP para ser integradas al Modelo DWH:

La tabla de hechos “FACT_ENLACE” tiene información sobre los resultados de la prueba llamada “ENLACE” a los niños de primaria y secundaria en todo el país, contiene información del alumno, la escuela donde estudia y su puntaje obtenido.

A continuación se presentan las descripciones de las tablas que forman parte de las dimensiones de la FACT_ENLACE dando como resultado la estrella mostrada en el diagrama 2.2. Estas tablas fueron hechas como dimensiones porque están relacionadas con la tabla de hechos.

- Dim_ciclo_enlace = el ciclo escolar que cursa el alumno que presenta la prueba enlace
- Dim_rango_edades_enlace = Guarda rangos de edades de los alumnos
- Dim_folio_enlace = número de folio de la prueba enlace
- Dim_materia_enlace = Guarda el nombre de las materias que contienen la prueba enlace
- Dim_grado_nivel_enlace = es el grado en el que están los alumnos
- Dim_tipo_localidad = tipo de localidad donde está el centro de trabajo (urbano o rural)
- Dim_nivel_logro = resultado de la prueba enlace (Excelente, bueno, regular, insuficiente)
- Dim_ct = centro de trabajo donde estudian los alumnos que presentan la prueba
- Dim_grado_marginación = grado de marginación de la localidad donde está el ct
- Dim_programas = los programas de ayuda que tiene cada centro de trabajo

ESQUEMA ENLACE

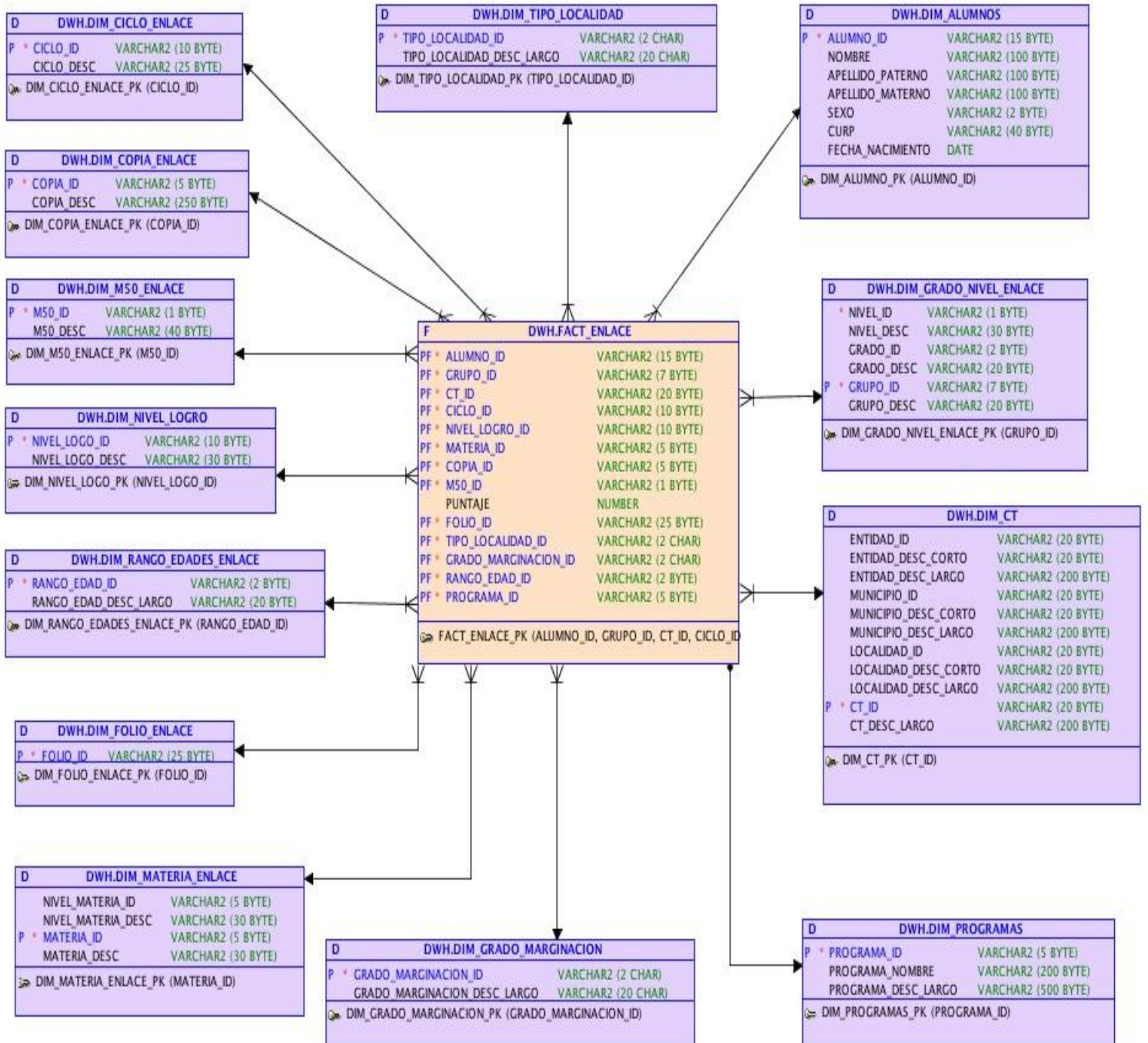


Diagrama 2.2 Esquema tipo estrella Fact_Enlace

2.2.5 Diseño data mart de Escuelas

El data mart de escuelas contiene información sobre los programas de ayuda a los que están inscritos cada centro de trabajo.

Para cada centro de trabajo se menciona su respectivo nivel educativo, en cuantos programas está inscrito y cuál es el monto económico y el número de beneficiados de ese centro de trabajo (ver diagrama 2.3).

ESQUEMA DATAMART PROGRAMAS

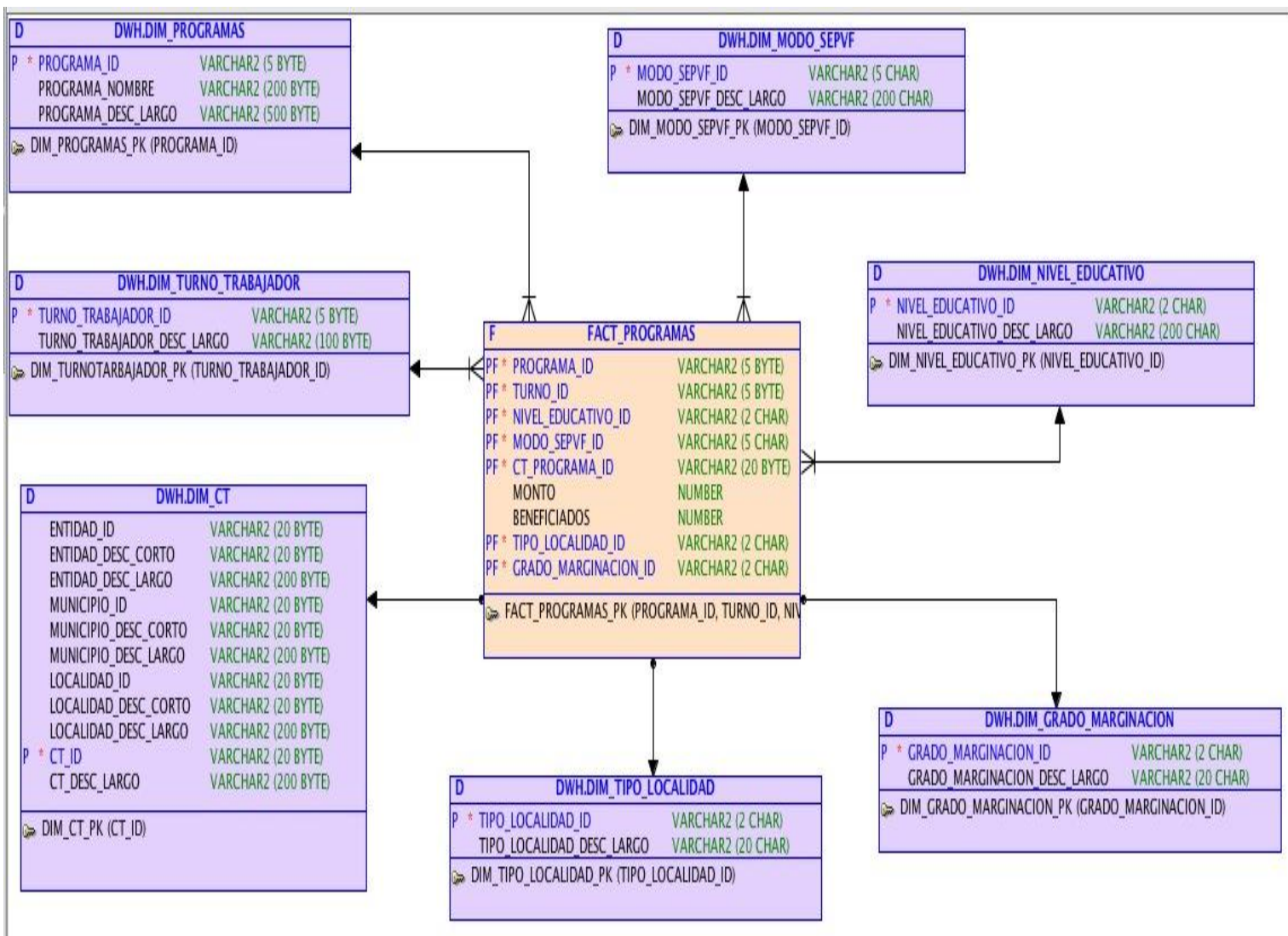


Diagrama 2.3 Esquema tipo estrella de escuelas

2.2.6 Diseño de los modelos para la creación de los tableros de control de Inteligencia de Negocios

Este es el primer paso para la creación de reportes y tableros. Se utilizó la herramienta de Oracle Business Intelligence, el cual tiene un componente llamado "Administration Tool", dicha herramienta crea 3 tipos de capas necesarias para hacer un modelo de negocio que cumpla con las necesidades que la Secretaría requiere.

- Capa Física.- Esta capa contiene la conexión a los repositorios físicos de las fuentes de información (tablas, columnas, dimensiones). En la figura 2.1 se ve que se puede conectar a diferentes tipos de fuentes de información, el requisito es que las fuentes sean consistentes entre sí para generar un modelo estable. Las fuentes que alimentan a la herramienta son dos:
 - Datawarehouse de Oracle
 - Cubos OLAP de Oracle

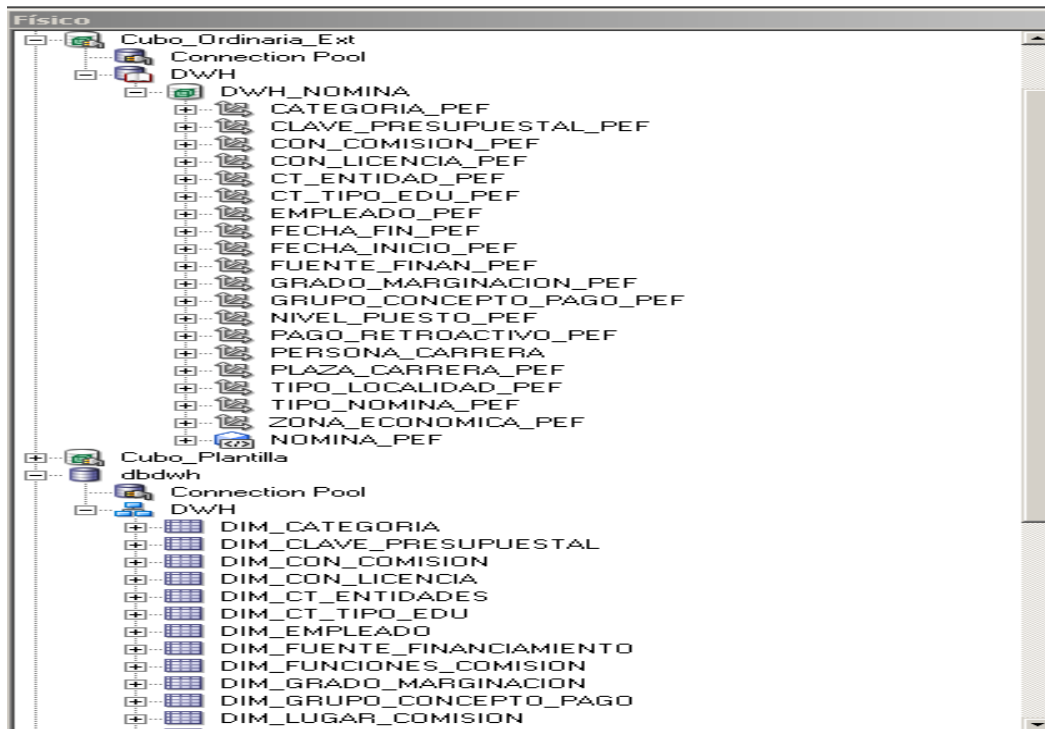


Figura 2.1 Capa física de la herramienta de bi

- Modelado.- Esta capa se encarga como su nombre lo dice de modelar las fuentes de información a las que se realizó la conexión en la capa física. Aquí se crean jerarquías, cálculos, combinaciones de fuentes según se requiera para cubrir los requerimientos de la Secretaría. Para plazas se crearon 5 modelos principales: Nómina, Cancelados, Plantilla, Comisionados, Licencias, (Ver figura 2.2).



Figura 2.1 Capade modelado de la herramienta de BI

- Presentación.- Esta capa se utiliza para mostrar el modelo definido en la capa anterior de manera entendible y práctica para el usuario encargado de hacer los reportes con la herramienta llamada Answers (Ver figura 2.3).

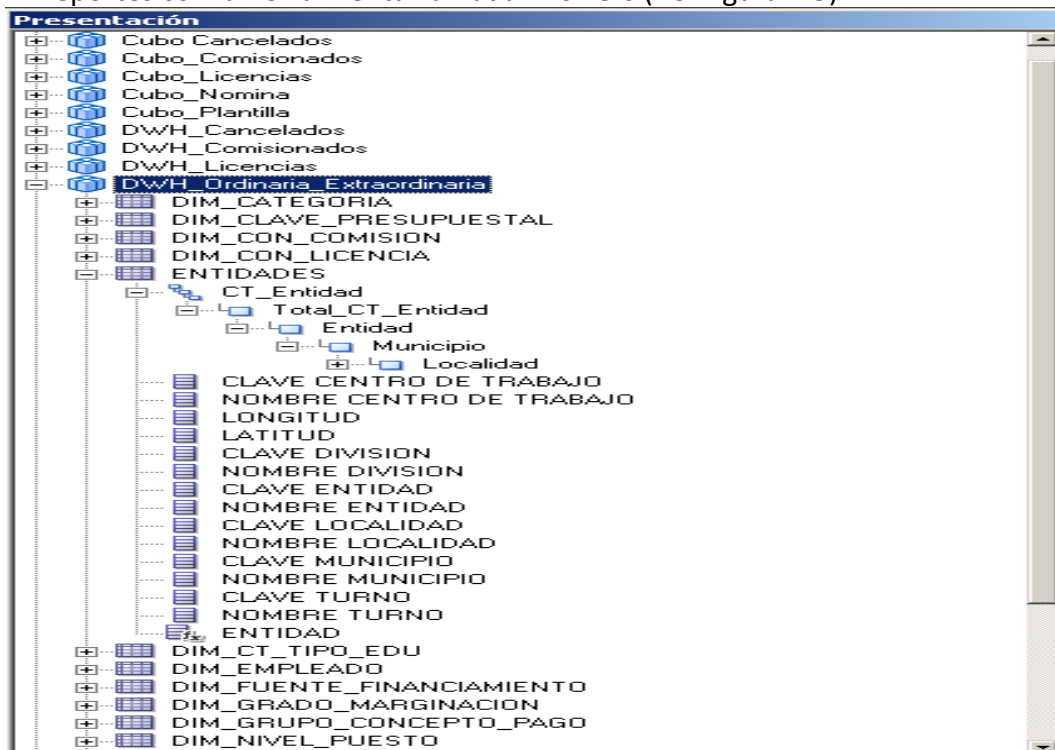


Figura 2.3 Capa de presentación de la herramienta de BI

OBIEE Answers. Desde aquí se pueden hacer consultas a la medida, se selecciona el área temática a utilizar y se seleccionan los campos que así se requiera, este elemento se verá en el siguiente capítulo.

Es importante resaltar que se debe de tener un buen análisis y diseño tanto del DWH como de la herramienta de BI, ya que sin un buen diseño no se podría llegar a los resultados deseados. Una vez teniendo el diseño, se realiza el desarrollo del mismo para su posterior implementación.

Capítulo 3. Desarrollo e Implementación

3.1 Desarrollo

Aquí se detallará el desarrollo del Data Warehouse y de los reportes y Tableros de control de inteligencia de negocios, se usaron herramientas de Oracle para realizar estos desarrollos, en el caso del Data Warehouse se usó la herramienta llamada “Analytic Workspace Administrator” y para el desarrollo de BI se utilizó la herramienta llamada “Oracle Business Intelligence”.

3.1.1 Desarrollo DWH

El modelo está creado para cubrir las necesidades de negocio a un nivel de granularidad muy bajo, lo que quiere decir que cualquier pregunta que el usuario se pueda formular el DWH podrá contestarla.

También es importante mencionar que para que en este modelo se puedan hacer comparaciones contra años anteriores usando la herramienta de Oracle Business Intelligence por lo que será necesario cargar un historial de información que se va dando con el tiempo.

OBIEE u Oracle Business Intelligence es la herramienta de Oracle utilizada para crear los reportes y tableros de control de inteligencia de negocios.

Para llevar a cabo los procesos de carga Extracción y transformación fue necesario realizar una serie de procesos los cuales precisamente hacen que la información pueda ser almacenada de tal forma que pueda ser explotada en un formato más visual en OBIEE.

Una vez que se tiene limpia la información el siguiente paso es el desarrollo y generación de los cubos de información, para ellos se utilizó la herramienta Analytic Workspace Administrator

Esta herramienta permite generar de manera física los cubos de información, para cada cubo se creó un área analítica de trabajo esto con el fin de que cada uno tenga sus propias dimensiones y sus propias métricas.

Se hizo de esta forma para tener un mejor control y más estructurada la generación y administración de los mismos.

A continuación se muestran las pantallas de la herramienta así como la estructura de cada una de las áreas de trabajo con sus dimensiones y cubos correspondientes:

Pantalla principal del Analytic Workspace Administrator: para poder crear las áreas de trabajo lo primero que se tiene que hacer es crear una conexión a nuestro repositorio de datos, en este caso el repositorio se llama datawarehouse.

Se le pasan los parámetros de conexión como se muestra en la figura 3.1:

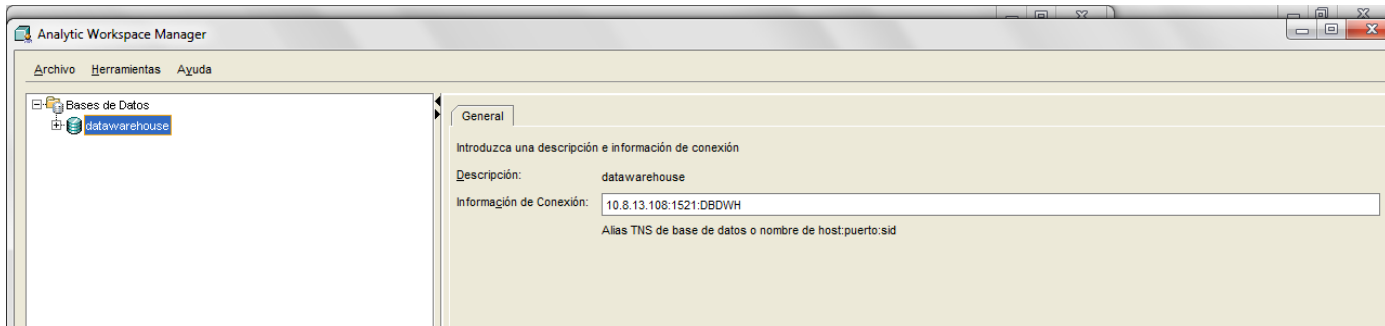


Figura 3.1 Conexión a repositorio de datos

Una vez creada la conexión al querer acceder pedirá un usuario y un contraseña (ver figura 3.2).

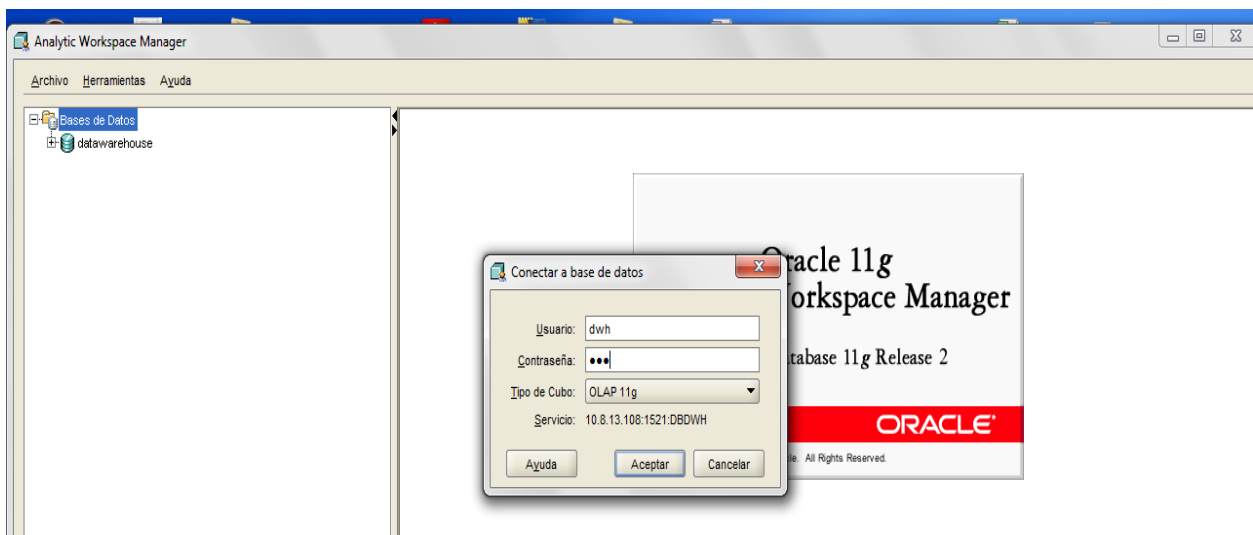


Figura 3.2. Conexión a esquema de base de datos.

Una vez adentro se podrán generar las áreas analíticas de trabajo, en este caso ya se cuentan con ellas como se muestra en la figura 3.3

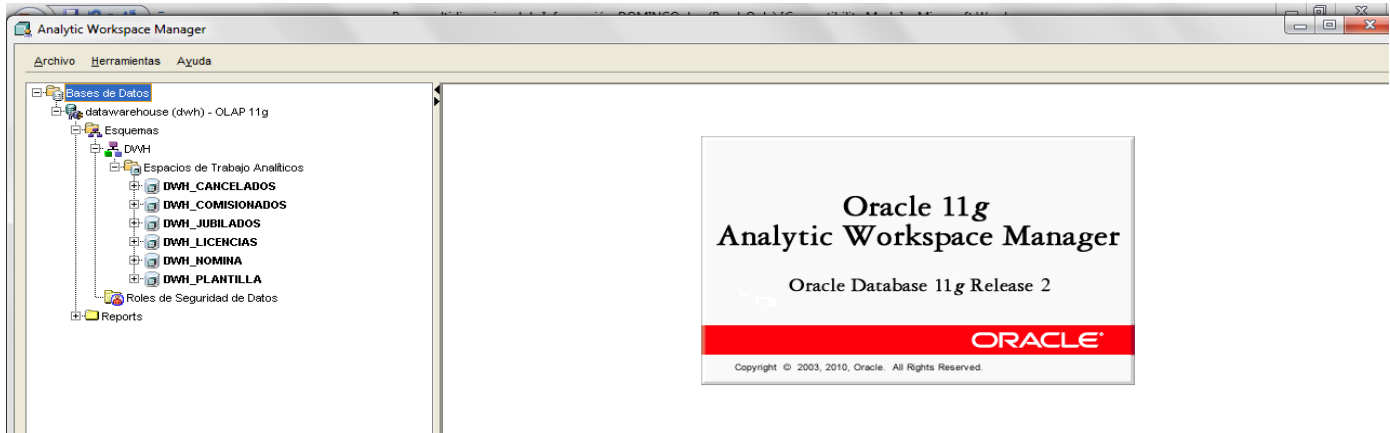


Figura 3.3. Áreas de trabajo

Como se mencionó en párrafos anteriores cada área de trabajo cuenta con un apartado de dimensiones y un apartado de cubos como lo muestra la figura 3.4

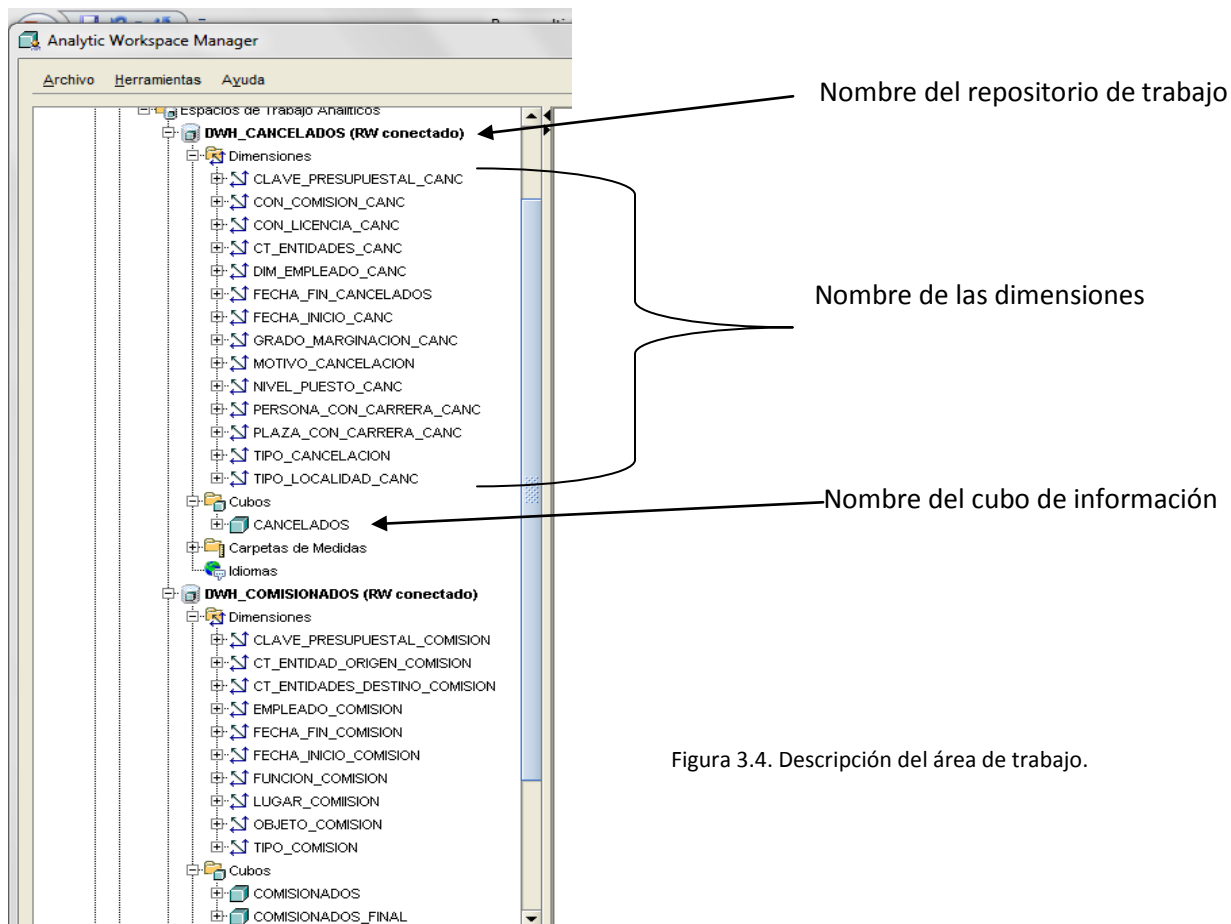


Figura 3.4. Descripción del área de trabajo.

3.1.2 Desarrollo de Inteligencia de negocios

En la herramienta de Oracle Business Intelligence hay una parte que se llama Answers.

Desde aquí se pueden hacer consultas a la medida, se selecciona el área temática a utilizar y se seleccionan los campos que así se requiera, cabe mencionar que los valores aquí presentados no son los reales, esto es por petición de la SEP.

El área temática es la parte del negocio del cual te interesa saber información, en este documento se mostrara el área temática de Nomina.

Como se mencionó en la capa anterior se tienen 5 modelos principales extraídos del Data Warehouse:

- Ordinaria Extraordinaria (Nómina): Contiene información referente a la nómina ordinaria, extraordinaria y honorarios como:
 - Montos de Percepciones
 - Montos de Deducciones
 - Categorías
 - Entidades
 - Centros de Trabajo
 - Nivel de Puesto del Empleado
 - Empleados
 - Zona Económica

Periodos de Pago

- Cancelados: Contiene información de pagos cancelados determinada por:
 - Entidad
 - Centros de Trabajo
 - Categorías
 - Nivel de Puesto del Empleado
 - Empleados
 - Zona Económica
 - Periodos de Pago
 - Montos de Percepciones
 - Montos de Deducciones
- Comisionados: Este modelo presenta información perteneciente a personal que se encuentra comisionado. A través de este modelo se podrá saber:
 - En qué entidad esta comisionado el personal
 - Centro de Trabajo origen al que pertenece
 - Centro de Trabajo destino a dónde se fue de comisión
 - Función de la comisión
 - Objetivo de la comisión
 - Lugar donde se encuentra comisionado (en caso de ser un lugar fuera del sector educativo)

- Tipo de comisión
- Fechas de Inicio y Fin de la comisión

- Licencias: Se podrá analizar información referente al personal que se encuentra en licencia. La información que se podrá conocer de la persona con licencia es:
 - Entidad en la cual trabaja el empleado
 - Centro de Trabajo del empleado
 - Tipos de Licencia
 - Fechas de inicio y fin de la licencia
 - Percepción en caso de tratarse de una licencia con goce de sueldo

- Plantilla: Aquí se podrá encontrar información relacionada con el empleado como:
 - Tipo de Categoría
 - Clave Presupuestal (Plaza)
 - Entidad-Centro de Trabajo donde labora
 - Nivel educativo en el que se encuentra
 - Estatus de activo o inactivo
 - Función que desempeña
 - Grado al que se encuentra asignado (Docente)
 - Grupo al que se encuentra asignado (Docente)
 - Materia a la que se encuentra asignado (Docente)
 - Nombramiento
 - Turno del Trabajador

A continuación se presenta el Tablero de Plazas, su navegación, elementos que lo constituyen.

Íconos de Navegación



Este ícono simboliza enlaces a otros tableros o reportes.



Este ícono indica que el reporte o tablero tiene un mapa.



Este ícono se utiliza como parte de los títulos o encabezados de las secciones de tableros con cantidades.



Este ícono representa una tabla con montos.



Al dar clic en este ícono se regresa al tablero o reporte inmediato superior.



Este ícono se refiere a que ese reporte o tablero tiene información con mayor detalle.

3.1.2.1 Página Principal

En esta primera página se presenta información General tal como cantidad de empleados, plazas y montos por tipo de nómina (figura 3.5).

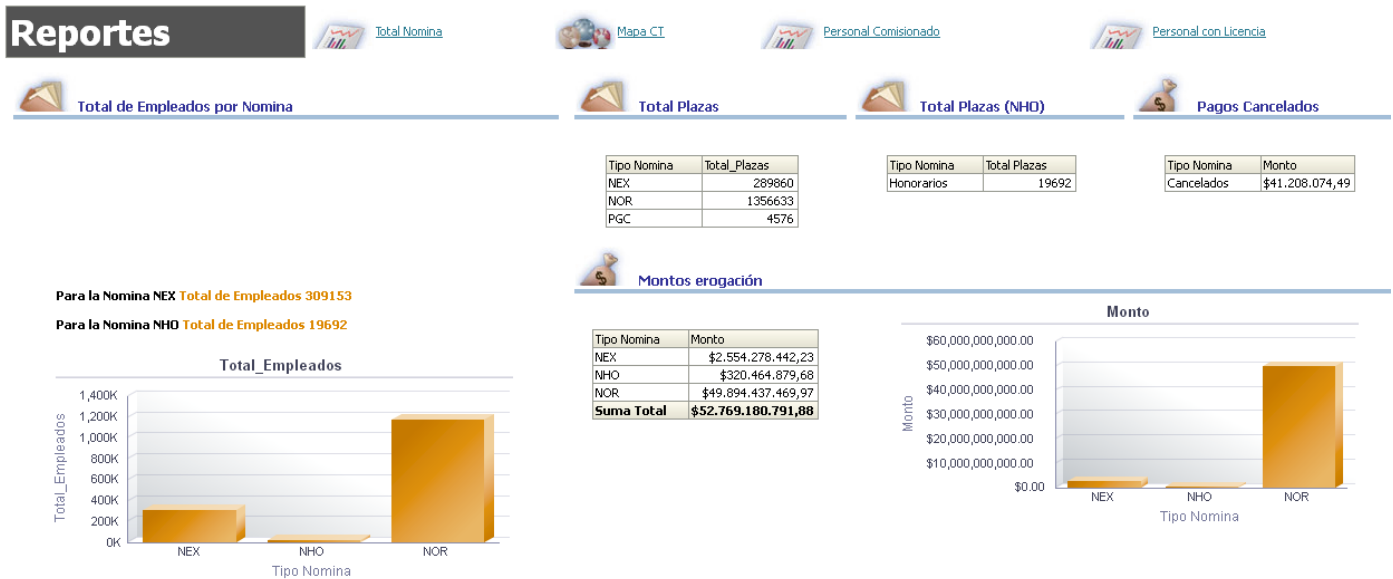


Figura 3.5 Página principal

Se divide en:

1. Reportes: Son iconos que permitirán navegar a través del tablero.



2. Área de Trabajo: Es el lugar donde se presenta la información, tablas y gráficos (ver figura 3.6).



Total Plazas

Tipo Nomina	Total_Plazas
NEX	289860
NOR	1356633
PGC	4576



Total Plazas (NHO)

Tipo Nomina	Total Plazas
Honorarios	19692



Pagos Cancelados

Tipo Nomina	Monto
Cancelados	\$41.208.074,49



Montos erogación

Tipo Nomina	Monto
NEX	\$2.554.278.442,23
NHO	\$320.464.879,68
NOR	\$49.894.437.469,97
Suma Total	\$52.769.180.791,88

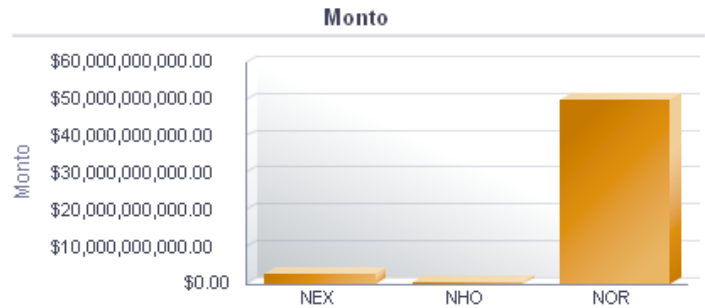


Figura 3.6. Área de Trabajo

3.1.2.2 Nomina

Se accede a esta parte del tablero al hacer clic sobre el ícono que se encuentra en la parte superior que dice Total Nómina.



Total Nomina

En esta página se podrá encontrar el total de empleados, centros de trabajo, horas de trabajo, percepciones y deducciones por Entidad.



Figura 3.7 Reportes de nómina

La figura 3.7 muestra lo siguiente:

1. Contiene enlaces a otros reportes relacionados con la nómina: Personal por Honorarios, por nómina ordinaria, nómina extraordinaria, Plantilla.
2. En esta sección de la página se puede observar dos gráficos que muestran algunas métricas relacionadas con la Nómina y su comparación con las entidades. Tienen una barra horizontal en la parte superior donde podrá elegir el tipo de nómina que se desea comparar.
3. Esta sección contiene una tabla con el total de empleados, centros de trabajo, percepciones y deducciones por Entidad. En la parte superior tiene un filtro donde podrá seleccionar el tipo de nómina que se desea consultar (ver figura 3.8).

En la columna de Entidad podrá hacerse el análisis hasta centro de trabajo. Al ir abriendo el árbol se puede hacer un recorrido por los municipios de cada entidad, pasar por las localidades del municipio seleccionado, verificar si se trata de alguna de las dos divisiones FAEB o FAETA y llegar hasta el centro de trabajo, como se observa en la imagen siguiente:

CT_Entidad
[-] Total_CT_Entidad 1
[-] AGUASCALIENTES
[-] CAMPECHE 2
[-] CHIAPAS
[-] CHIHUAHUA
[-] COAHUILA
[-] DISTRITO FEDERAL
[-] Benito Juárez 3
[-] Cuauhtémoc
[-] Cuauhtémoc 4
[-] FAEB 5
DOLORES ANGELA CASTILLO
DOLORES ANGELA CASTILLO
JOSE LUIS VIEYRA GONZALEZ 6
JOSE LUIS VIEYRA GONZALEZ
JOSE MARIA PINO SUAREZ

Figura 3.8 Columna de centros de trabajo

La figura 3.8 muestra lo siguiente

1. Total Nacional. Es la suma de todos los Estados
2. Nombre de la Entidad
3. Nombre del Municipio o Delegación
4. Nombre de la Localidad
5. Nombre de la División (Básica, Conalep, INEA)
6. Nombre del Centro de Trabajo

- **Nomina Ordinaria - Extraordinaria**

Aquí se podrá analizar con mayor detalle al empleado y su comportamiento en la nómina, se tiene la opción para realizar pivotes de columnas y aplicar los filtros según requiera el usuario que este analizando la nómina (ver figura 3.9).

AÑO: 2011
 TRIMESTRE: Q1
 ENTIDAD: Distrito Federal
 TIPO NOMINA: NOR
 TIPO CONTRATACION: Jornada
 TIPO LOCALIDAD: (Todos los Valores de...)

CON COMEDOR: NO
 CON LICENCIA: NO

PERSONA CON CARRERA: NO
 PLAZA CON CARRERA: SI

RFC empresa por: _____
 CURP empresa por: _____
 APELLIDO PATERNO empresa por: _____
 APELLIDO MATERNO empresa por: _____
 NOMBRE empresa por: _____

1

Aplicar Restablecer

ORDINARIA-EXTRAORDINARIA(EXCEL)

ENTIDAD	CLAVE PRESUPUESTAL	TIPO PLAZA	EMPLEADO	NIVEL PUESTO	PERIODO DE	PERIODO AL	ZONA ECONOMICA	FUENTE FINANCIAMIENTO	PAGO RETROACTIVO	TIPO EDUCATIVO	PERCEPCION NETA	TOTAL PERCEPCION	TOTAL DEDUCCION	D
DF	071352A01E05000111112	Directivo (personal docente en función de dirección ó supervisión educativa)	GAMBICA VERONICA	NO DEFINIDA	201101	201101	NO DEFINIDA	Estatal	No especificado u Otro	EDUCACION BASICA	5,604	5,096	483	
					201102	201102	NO DEFINIDA	Estatal	No especificado u Otro	EDUCACION BASICA	2,747	3,230	483	
					201103	201103	NO DEFINIDA	Estatal	No especificado u Otro	EDUCACION BASICA	2,747	3,230	483	
					201104	201104	NO DEFINIDA	Estatal	No especificado u Otro	EDUCACION BASICA	2,747	3,230	483	
					201105	201105	NO DEFINIDA	Estatal	No especificado u Otro	EDUCACION BASICA	2,747	3,230	483	
					201106	201106	NO DEFINIDA	Estatal	No especificado u Otro	EDUCACION BASICA	2,747	3,230	483	
	0719531A08026000710002	Directivo (personal docente en función de dirección ó supervisión educativa)		NO DEFINIDA	201101	201101	NO DEFINIDA	Estatal	No especificado u Otro	EDUCACION BASICA	9,141	10,005	864	
					201102	201102	NO DEFINIDA	Estatal	No especificado u Otro	EDUCACION BASICA	3,582	4,445	864	
					201103	201103	NO DEFINIDA	Estatal	No especificado u Otro	EDUCACION BASICA	3,582	4,445	864	
					201104	201104	NO DEFINIDA	Estatal	No especificado u Otro	EDUCACION BASICA	3,582	4,445	864	
					201105	201105	NO DEFINIDA	Estatal	No especificado u Otro	EDUCACION BASICA	3,582	4,445	864	
					201106	201106	NO DEFINIDA	Estatal	No especificado u Otro	EDUCACION BASICA	3,582	4,445	864	
	072001A01E05000111126	Directivo (personal docente en función de dirección ó supervisión educativa)	COLIN FUENTES ERIC	NO DEFINIDA	201101	201101	NO DEFINIDA	Estatal	No especificado u Otro	EDUCACION BASICA	5,753	6,262	609	

2

Figura 3.9 Nomina Oraditaria_Extraordinaria

La Figura 3.9 muestra lo siguiente:

1. Contiene una serie de diferentes tipos de filtros a seleccionar según la información que se desee consultar. Después de hacer la selección deseada dar clic en aplicar para ver los resultados.
2. Sección de la tabla donde se refleja el resultado de las selecciones correspondientes en los filtros. Así mismo aquí podrá hacerse un pivoteo de columnas, y análisis a profundidad (drill down).

Las columnas que contienen miembros en azul son aquellas en las que podrá hacerse análisis a profundidad.

Los demás reportes de nomina tienen una navegación y presentación igual a los reportes vistos anteriormente, por lo que continuaremos con la demás información

- Mapas

Para este tablero se cuenta con dos mapas (Google Maps) donde se puede observar los estados con mayores sueldos (ver figura 3.10 y 3.11)

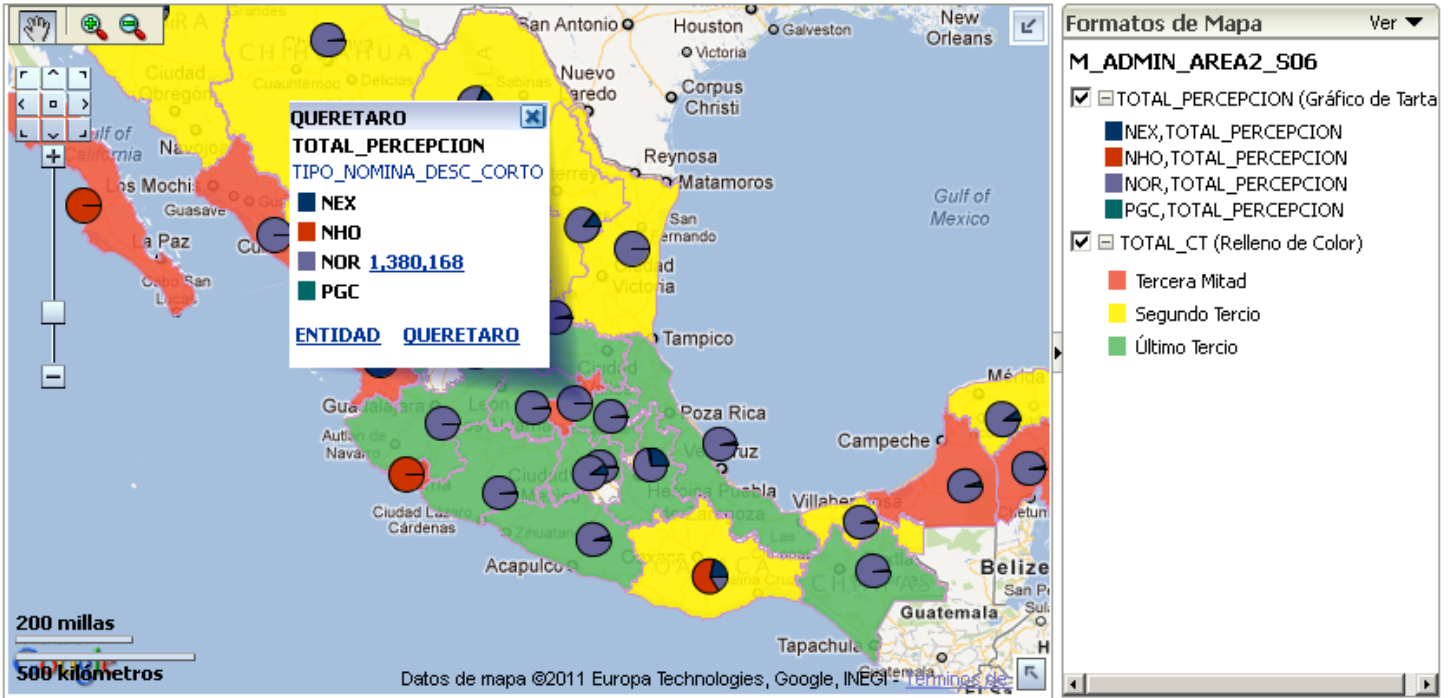


Figura 3.10 Total de percepciones por estado

Los centros de trabajo, donde se localizan y número de empleados que tiene.

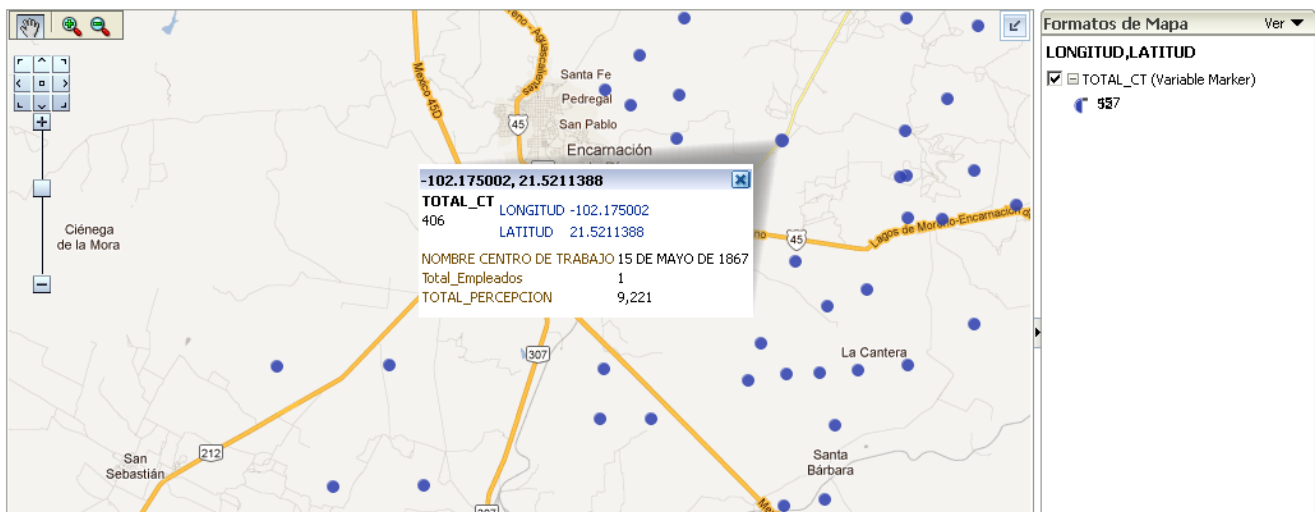


Figura 3.11 Localización de los centros de trabajo

3.1.2.3 Alumnos

Los reportes de Alumnos eran básicamente para observar los resultados de la prueba llamada ENLACE, que se aplica a los alumnos de primaria y secundaria en todo el país.

Los resultados que pueden obtener los alumnos en el examen son:

- Excelente
- Bueno
- Elemental
- Insuficiente

El reporte de la figura 3.12 muestra los resultados de la prueba ENLACE de cada entidad de la república.

Alumnos Nivel de Logro Por Materia

Entidad	MATERIA	NIVEL DE LOGRO	ALUMNOS	ALUMNOS	PUNTAJE PROMEDIO
Total_Entidad	Matemáticas	BUENO	249,784	26.8%	641
		ELEMENTAL	442,901	47.6%	504
		EXCELENTE	105,206	11.3%	754
		INSUFICIENTE	133,499	14.3%	370
AGUASCALIENTES	Matemáticas	BUENO	18,127	1.9%	642
		ELEMENTAL	31,451	3.4%	505
		EXCELENTE	7,687	0.8%	756
		INSUFICIENTE	9,387	1.0%	371
BAJA CALIFORNIA	Matemáticas	BUENO	42,239	4.5%	639
		ELEMENTAL	88,441	9.5%	503
		EXCELENTE	13,491	1.4%	750
		INSUFICIENTE	27,835	3.0%	369
BAJA CALIFORNIA SUR	Matemáticas	BUENO	7,125	0.8%	640
		ELEMENTAL	14,649	1.6%	502
		EXCELENTE	2,368	0.3%	749
		INSUFICIENTE	4,337	0.5%	373
CAMPECHE	Matemáticas	BUENO	5,904	0.6%	641
		ELEMENTAL	8,945	1.0%	507
		EXCELENTE	3,069	0.3%	754
		INSUFICIENTE	2,080	0.2%	375
CHIAPAS	Matemáticas	BUENO	4,818	0.5%	645
		ELEMENTAL	7,126	0.8%	503
		EXCELENTE	2,807	0.3%	756
		INSUFICIENTE	2,266	0.2%	370
CHIHUAHUA	Matemáticas	BUENO	46,143	5.0%	643
		ELEMENTAL	75,600	8.1%	505
		EXCELENTE	21,291	2.3%	755
		INSUFICIENTE	22,508	2.4%	371
COAHUILA	Matemáticas	BUENO	48,560	5.2%	639
		ELEMENTAL	88,105	9.3%	504

Figura 3.12 Resultados de la prueba ENLACE por entidad

El siguiente reporte muestra los resultados de la prueba ENLACE por rango de edades, se muestra el puntaje promedio y el número de alumnos que realizaron la prueba (figura 3.13).

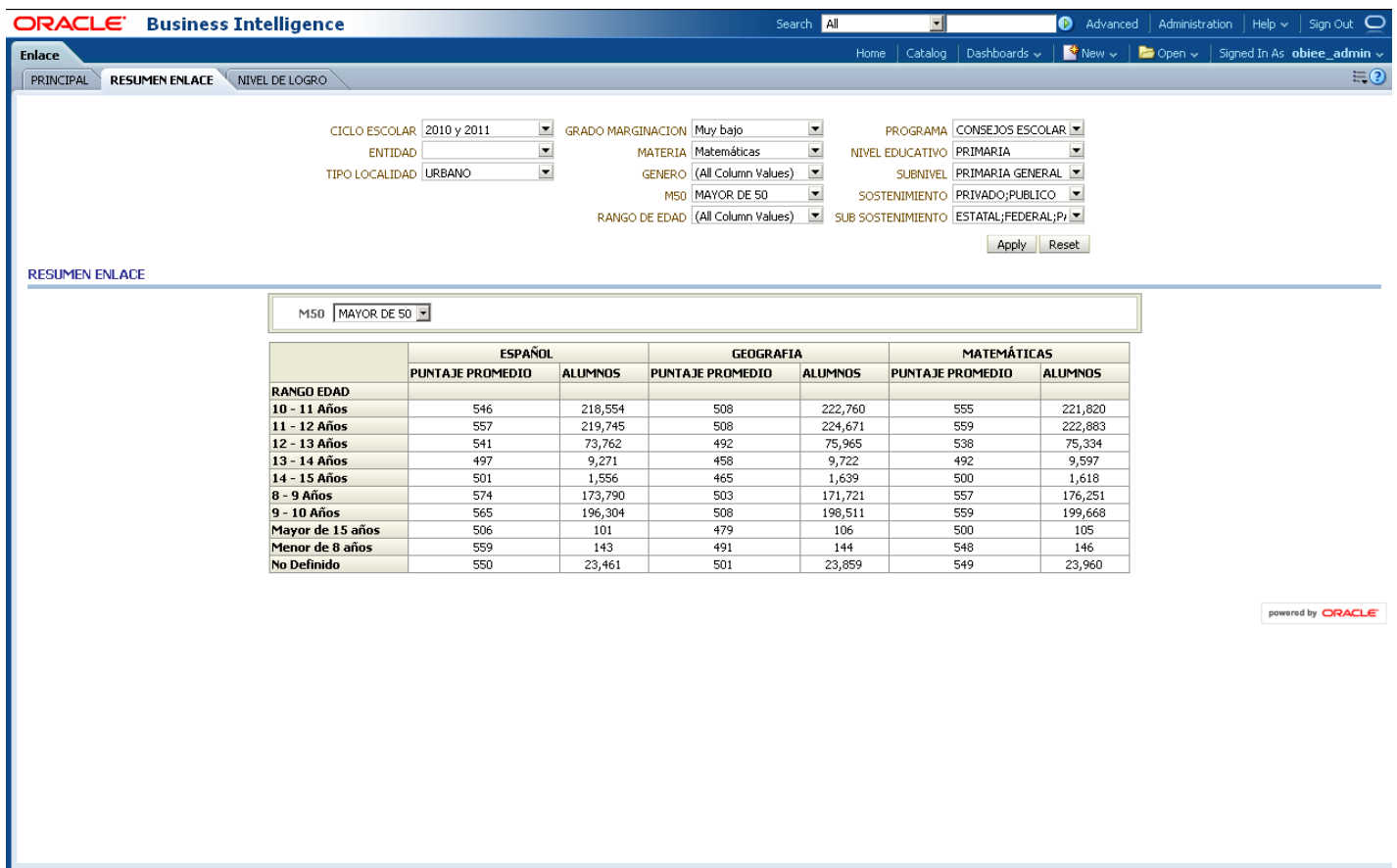


Figura 3.13 Resultado de la prueba ENLACE por rango de edades

El nivel de detalle que se puede conseguir llega incluso al nivel de poder saber el resultado de la Prueba ENLACE de un alumno en particular como lo muestra la figura 3.14

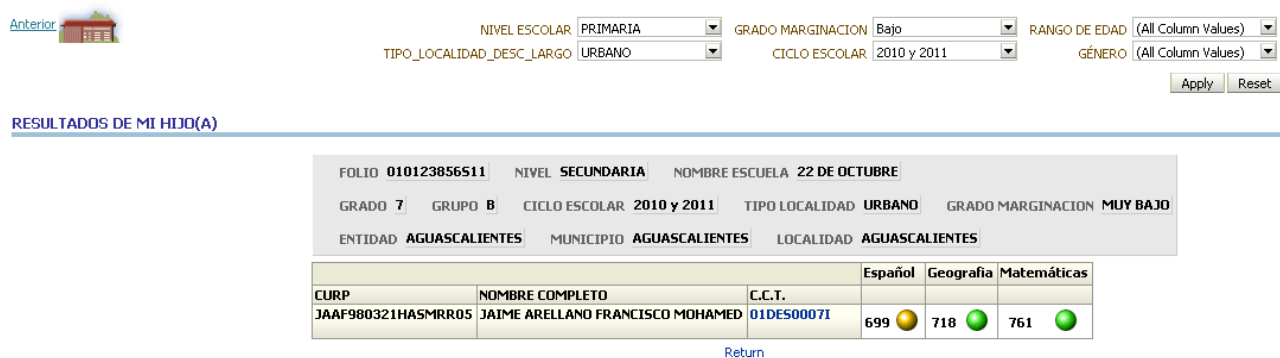


Figura 3.14 Resultados de la prueba enlace por alumno

3.1.2.4 Escuelas

El diseño de data warehouse de escuelas permite generar reportes que nos digan los programas de beneficios que otorga la SEP a las escuelas, ¿cuántas escuelas tienen cuantos programas de beneficios?, y esto para cada Entidad Federativa del país. Esto se muestra en la figura 3.15 donde podemos ver el número de beneficiados por Entidad y el monto económico de esos beneficios.



[Anterior](#)



[Detalle de Programas](#)



[Programas de Estimulos por CT](#)



[Ver Mapa](#)



[Cifras de control](#)

Beneficiados por Programas de Estimulos

Entidad	BENEFICIADOS	MONTO
Total_entidad	226,319	881,565,086
AGUASCALIENTES	2,713	10,041,060
BAJA CALIFORNIA	10,024	31,347,011
BAJA CALIFORNIA SUR	1,210	4,528,283
CAMPECHE	2,027	8,490,378
CHIAPAS	11,358	49,349,474
CHIHUAHUA	7,528	30,333,277
COAHUILA	6,674	22,988,957
COLIMA	1,775	7,106,169
DISTRITO FEDERAL	15,110	59,287,453
DURANGO	4,960	19,399,446
GUANAJUATO	9,724	42,173,117
GUERRERO	4,972	24,386,169
HIDALGO	5,660	23,808,125
JALISCO	16,820	63,076,365
MEXICO	26,304	99,814,182
MICHOACAN	4,450	17,919,439
MORELOS	2,817	11,367,643
NAYARIT	3,118	11,446,300
NUEVO LEON	10,335	43,665,578
OAXACA	499	2,113,658
PUEBLA	10,026	40,559,532
QUERETARO	3,290	12,302,002
QUINTANA ROO	2,583	10,335,875
SAN LUIS POTOSI	7,802	31,664,648

Cantidad de Programas por Entidad

Entidad	CANTIDAD DE PROGRAMAS APLICADOS	CANTIDAD DE PROGRAMAS DE ESTIMULOS APLICADOS
Total_entidad	327,682	40,423
AGUASCALIENTES	4,847	394
BAJA CALIFORNIA	10,091	874
BAJA CALIFORNIA SUR	2,829	188
CAMPECHE	5,146	424
CHIAPAS	29,403	3,034
CHIHUAHUA	11,408	1,260
COAHUILA	11,111	791
COLIMA	3,096	246
DISTRITO FEDERAL	21,938	1,668
DURANGO	13,529	1,105
FALTAN	19	0
GUANAJUATO	20,294	2,229
GUERRERO	18,806	1,657
HIDALGO	15,962	1,238
JALISCO	14,938	3,250
MEXICO	21,625	3,344
MICHOACAN	7,656	980
MORELOS	3,530	484
NAYARIT	4,396	609
NUEVO LEON	13,840	2,402
OAXACA	7,704	437

Figura 3.15 Programas de beneficios por entidad

También se puede ver la información con muchos más detalle, en la figura 3.16 se tiene un reporte donde podemos ver por cada escuela cuantos programas de beneficios tiene. Se observa el nombre de la entidad, el nombre de cada centro de trabajo y el nombre de cada programa que tiene el centro de trabajo.

ENTIDAD Centro de Trabajo

ENTIDAD_DESC_LARGO	CT_ID	CT_DESC_LARGO	PROGRAMA_NOMBRE
Aguascalientes	01CPRO1855	POR DEFINIR	ESCUELAS APTITUDES SOBRESALIENTES
	01DE50001O	LIC. BENITO JUAREZ	PROGRAMA DE FORTALECIMIENTO A LA INFRAESTRUCTURA Y EL EQUIPAMIENTO
			ESCUELAS APTITUDES SOBRESALIENTES
			CONSEJOS ESCOLARES DE PARTICIPACIÓN SOCIAL REGISTRADOS
			ESCUELAS CON MODELO DE GESTIÓN PARTICIPATIVA SIN FINANCIAMIENTO
			PROGRAMA ESCUELA SEGURA
	01DE50002N	MOISES SAENZ	ESCUELAS DISCAPACIDAD
			PROGRAMA HABILIDADES DIGITALES PARA TODOS
			CONSEJOS ESCOLARES DE PARTICIPACIÓN SOCIAL REGISTRADOS
			ESCUELAS CON MODELO DE GESTIÓN PARTICIPATIVA SIN FINANCIAMIENTO
	01DE50003M	LIC. ADOLFO LOPEZ MATEOS	PROGRAMA ESCUELA SEGURA
			PROGRAMA DE FORTALECIMIENTO A LA INFRAESTRUCTURA Y EL EQUIPAMIENTO
			ESCUELAS APTITUDES SOBRESALIENTES
			PROGRAMA HABILIDADES DIGITALES PARA TODOS
	01DE50004L	IZCOATL	CONSEJOS ESCOLARES DE PARTICIPACIÓN SOCIAL REGISTRADOS
			ESCUELAS CON MODELO DE GESTIÓN PARTICIPATIVA SIN FINANCIAMIENTO
			PROGRAMA ESCUELA SEGURA
			ESCUELAS DISCAPACIDAD
	01DE50004L1	POR DEFINIR	PROGRAMA ECD 2010 INDIVIDUAL
	01DE50005K	IGNACIO ALLENDE	PROGRAMA DE FORTALECIMIENTO A LA INFRAESTRUCTURA Y EL EQUIPAMIENTO

Figura 3.16 Programas por escuela

En la figura 3.17 esta página podemos visualizar:

- Dentro del mapa, la distribución geográfica de los Centros de Trabajo, programas y beneficiados

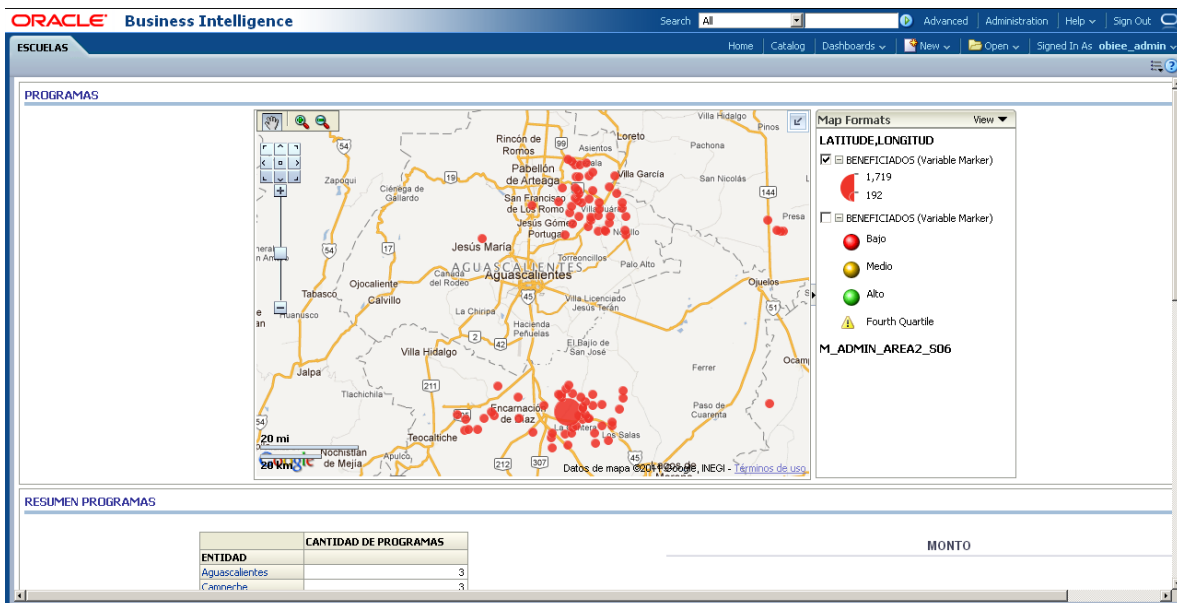


Figura 3.17 Distribución geográfica de los centros de trabajo

3.2 Implementación

En la implementación o paso a producción realizamos un proceso en el que se lleva el desarrollo previamente realizado a los servidores donde la aplicación estará ejecutándose y estará disponible para el usuario final.

Estos servidores tienen las mismas herramientas y capacidades que los de desarrollo, por lo que se realiza una copia exacta del servidor de desarrollo al servidor de producción.

3.2.1 Seguridad en los servidores de producción

Como los servidores de producción estarán disponibles para el usuario final, se necesita agregar seguridad para acceder a la aplicación, por esa razón se crearon diferentes usuarios con diferentes privilegios para ver información dependiendo de su nivel de acceso.

La tabla 3.1 muestra los tipos de usuario que tienen acceso a la aplicación de Business Intelligence

Tipo de usuario	Características del Usuario
Administrador	Usuario con privilegios para realizar administración de la herramienta, consulta, creación de reportes y tableros.
Upepe	Usuario con privilegios de lectura de reportes y tableros. Puede consultar la información de todas las entidades.
Público	Usuario con acceso a tableros específicos a través de internet.
Entidad	Usuario con privilegios de lectura de reportes y tableros. Solo puede ver información de su entidad.

Tabla 3.1 Características de los usuarios

Se delegó la administración de los servidores de producción a un tercero, una empresa llamada KIO Networks, esto es un fragmento sacado de su página de internet:

“El servicio de hospedaje dedicado de KIO Networks es el punto de acceso hacia nuestra robusta y confiable infraestructura:

- Ambiente aislado por medio de racks y jaulas perimetrales.
- Soluciones de hospedaje con equipamiento en hardware y software de distintas plataformas.
- Hospedaje en nuestros centros de datos de misión crítica de alta disponibilidad con accesos biométricos, detectores de iris y de reconocimiento facial que aseguran total privacidad y seguridad.

KIO Networks te puede ofrecer hosting con servicios de internet redundante por múltiples carriers, con redes administradas y balanceadas de última tecnología.

Equipos de profesionales especializados y certificados en todas las plataformas tecnológicas, equipos activos de comunicaciones, redes, internet y seguridad.”

Después de terminar el desarrollo y la implementación del DWH y de la herramienta de BI, se le realizaron pruebas a éstos, dando como resultado cifras diferentes a las obtenidas por la UPEPE, esto se debió a muchas circunstancias pero el uso de diferentes reglas de negocio y algunas fuentes de información que ellos usaban y nosotros no teníamos, fueron los principales motivos.

Por ello se decidió realizar cambios a la estructura del DWH y a por lo tanto a la herramienta de BI, esto se puede ver en el capítulo 4.

Capítulo 4. Consideraciones y reestructuración del modelo

Después someter a prueba lo analizado en los capítulos anteriores, los resultados que se obtuvieron con el sistema y los resultados que la SEP obtenía eran diferentes, esto se debía principalmente a la diferencia de reglas de negocios usadas para la extracción de datos al data warehouse.

Además al conocer mejor la información con la que se trabaja, se decidió hacer una reestructuración del modelo del data warehouse, para un mejorar el rendimiento y el procesamiento de datos.

4.1 Modificaciones al modelo del DWH

Debido a que para llevar a cabo un buen funcionamiento de BI (Inteligencia de Negocios), es necesario gestionar datos guardados en diversos formatos, fuentes y tipos, para luego depurarlos e integrarlos, además de almacenarlos en un solo destino que permita su posterior análisis y explotación, es importante contar con una buena arquitectura que satisfaga todas estas necesidades. Es por eso que se presenta a continuación la arquitectura de Datawarehouse.

La arquitectura del Datawarehouse está conformado por cuatro esquemas, los cuales cada uno de ellos tienen funciones específicas y almacenan de manera particular los datos (ver figura 4.1).

- DWH_STAGE. En este esquema se almacena los datos actuales, es decir, en el caso de PEF el último trimestre reportado por las Entidades. Este esquema tiene un tamaño considerable, se debe a que los datos residentes poseen el más bajo nivel de granularidad, o sea, se almacenan a nivel detalle.
- DWH_HISTORICOS. Los datos que vivirán en este esquema son datos antiguos, que no son frecuentemente consultados. También se almacenan a nivel de detalle. Se almacenaran todos los trimestres menos el trimestre actual.
- DWH_ROLAP. Este esquema almacenará los datos ligeramente resumidos de todos los trimestres. Cabe recalcar que en este esquema se aprecia la arquitectura de estrella por cada data mart.
- DWH_MOLAP. Es esquema almacena datos altamente resumidos, estos datos se compactan aún más a los datos ligeramente resumidos, y se almacenan en objetos llamados cubos.

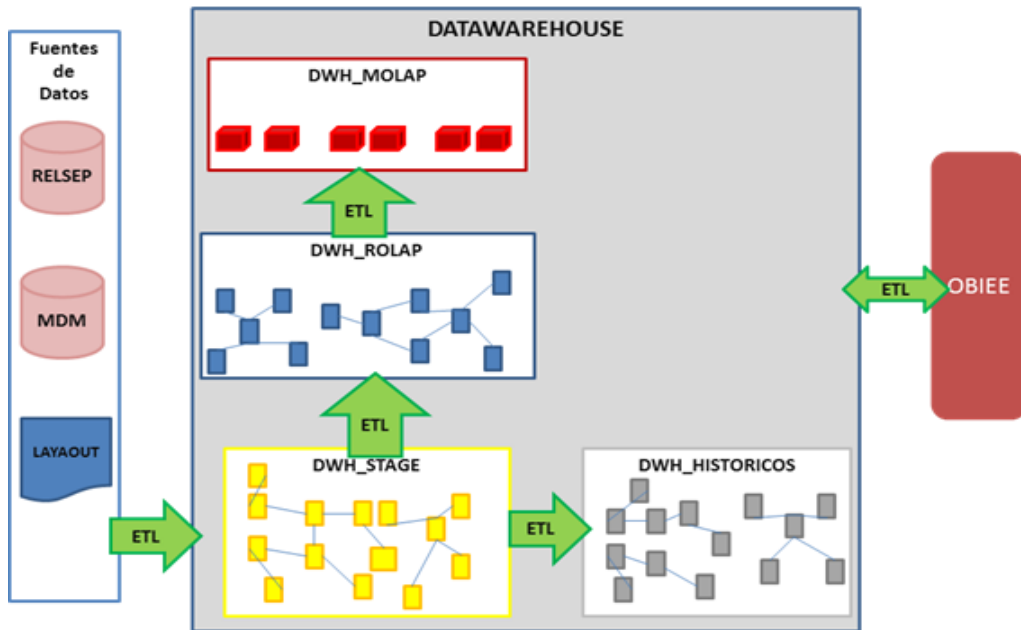


Figura 4.1 Nuevo modelo DWH

4.1.1 Esquema DWH_STAGE

Este esquema cuenta con tres tipos de tablas:

Intermedias. Estas tablas guardan las claves únicas de las diferentes fuentes de datos que existen, y guarda la nueva clave única generada y que utilizará en el DWH.

Temporales. Estas tablas forman parte de alguna de las fuentes de datos que se utilizará para poblar el Datawarehouse.

Registro. En estas tablas almacenan toda la información de PEF, de manera detallada, estas tablas se encuentran normalizadas hasta la tercera forma normal.

Tablas intermedias se componen de cuatro columnas principalmente:

- Columna que almacena el ID generado para el DWH y como tipo de datos es entero.
- Columna que almacena el ID externo, es decir, el ID de la fuente de datos.
- Columna que hace referencia a la tabla DWH_FUENTES_DATOS donde podemos saber el origen de datos del ID externo.
- Columna que almacena la fecha en que fue cargado ese registro.

4.1.2 ESQUEMA DWH_ROLAP

Este esquema contiene los modelos de estrella por cada data mart. Se describen 2 ejemplos a continuación, Los demás modelos fueron hechos de la misma manera.

DATA MART LICENCIAS.

Este modelo de datos almacena la información de PEF que hace referencia a los trabajadores que tiene una licencia, las métricas que se pueden explotar de este modelo son las siguientes:

- Percepciones federales.
- Percepciones Otras.
- Deducciones.
- Número de trabajadores = `count(distinct empleado_id)`.
- Número de plazas = `count(distinct plaza_id)`

Cuenta el modelo con las siguientes tablas de tipo dimensión con las que se pueden agrupar y filtrar la información (ver figura 4.2).

- DIM_ANIO.
- DIM_CATEGORIA.
- DIM_ENTIDAD_CT.
- DIM_GRADO_MARGINACION.
- DIM_PLAZA.
- DIM_TRIMESTRE.
- DIM_TIPO_LOCALIDAD.
- DIM_TIPO_LICENCIA.
- DIM_TIPO_PLAZA.
- DIM_FECHA.
- DIM_EMPLEADO.
- DIM_DIVISION.

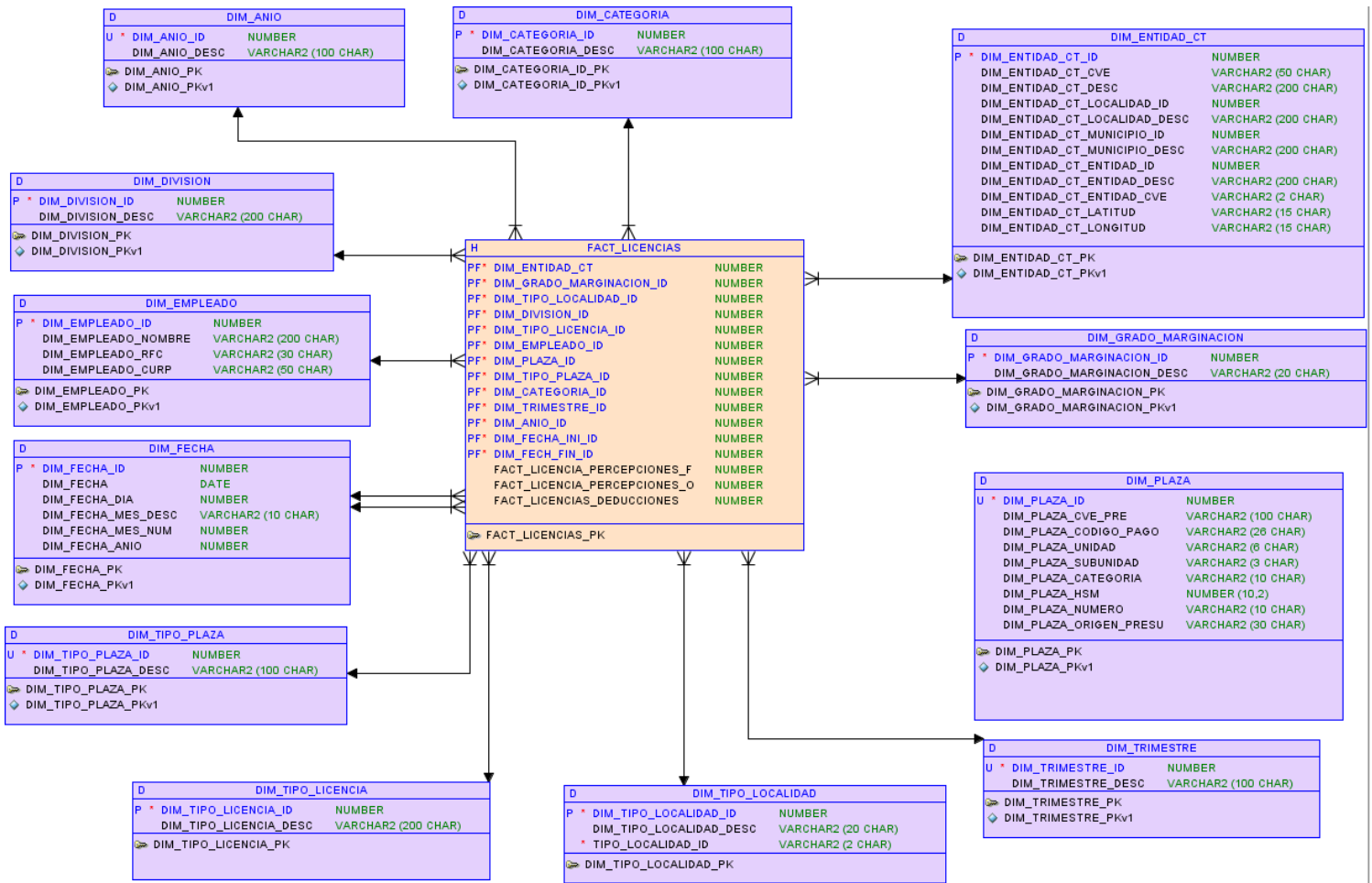


Figura 4.2 Data mart Licencias.

DATA MART NOMINA MONTOS

El modelo de nómina montos almacena las cantidades pagadas y que fueron registradas por las entidades federativas a la SEP. Este modelo cuenta con la siguiente métrica:

- Monto

Y para agrupar y filtrar la información cuanta con las siguientes dimensiones (ver figura 4.3):

- Dim_anio.
- Dim_concepto_pago.
- Dim_empleado.
- Dim_entidad_ct.
- Dim_grado_marginacion.
- Dim_plaza.
- Dim_quincena.

- Dim_tipo_localidad.
- Dim_subnivel_edu.
- Dim_sistema.
- Dim_tipo_nomina.
- Dim_tipo_pago.
- Dim_trimestre.
- Dim_zona_economica.
- Dim_pago_retroactivo.
- Dim_nivel_puesto.
- Dim_fuente_financiamiento.
- Dim_division.

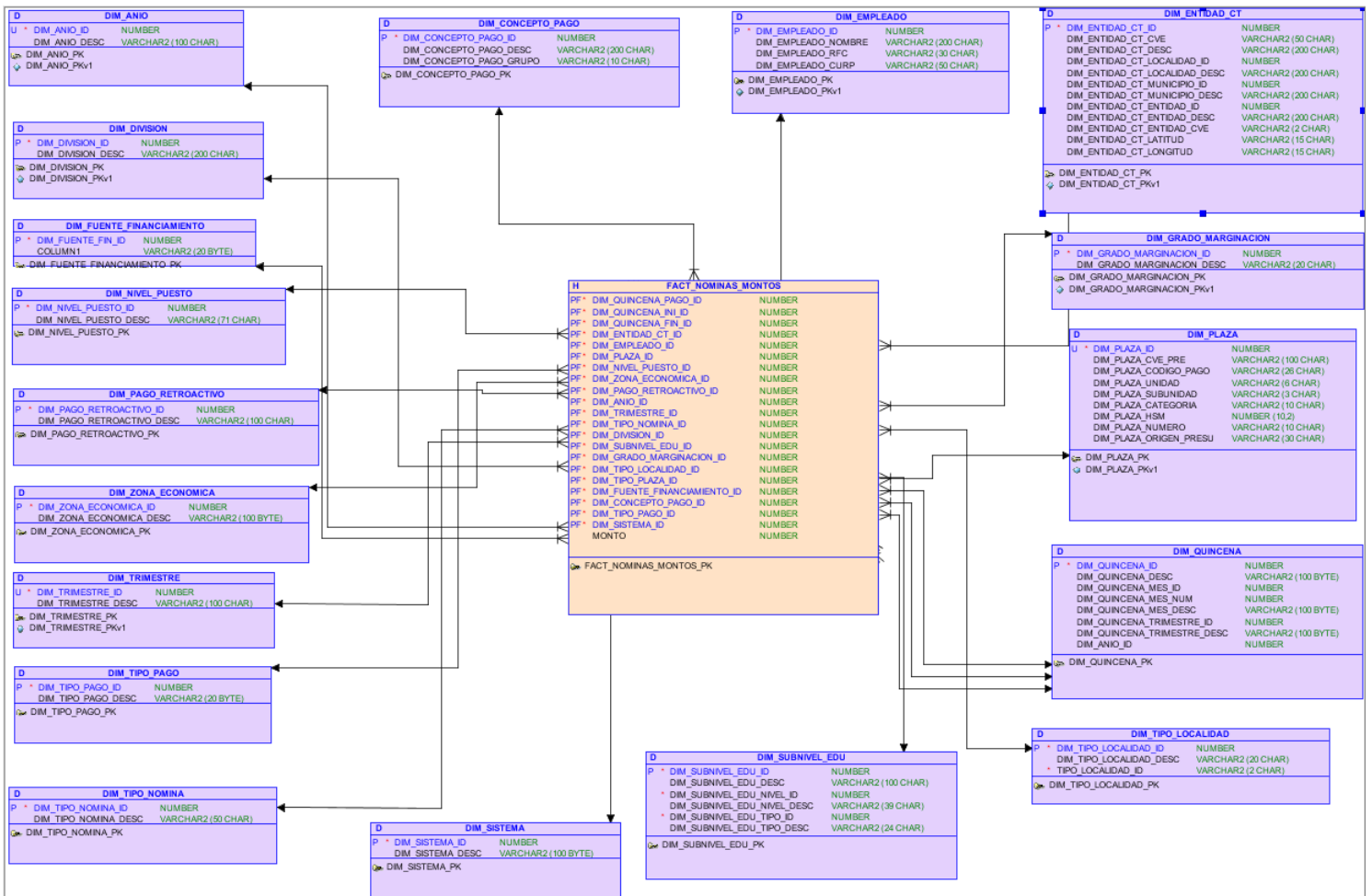


Figura 4.3 Data mart Nomina montos.

4.1.3 ESQUEMA DWH_MOLAP

En este esquema se encuentran los siguientes objetos (ver tabla 4.1):

CUBO	DESCRIPCION
CANCELADOS	Almacena los montos de los pagos cancelados a un nivel altamente resumidos
LICENCIAS	Almacena los montos pagado de los trabajadores con licencia a un nivel altamente resumidos
COMISIONADOS	Almacena los montos pagados de los trabajadores con comisión a un nivel altamente resumidos
HONORARIOS	Almacena los montos pagados por honorario con un nivel altamente resumido
NÓMINAS	Almacena los montos pagados de las nóminas ordinaria y extraordinaria a un nivel altamente resumido.
PLANTILLA	Almacena las horas reales según lo registrado por la plantilla.

Tabla 4.1 Cubos de información

Estos cubos comparten las dimensiones que se ven en la tabla 4.2

Dimensión	Descripción
ANIO	Dimension que indica el año reportado por la entidad federativa.
TRIMESTRE	Dimension que indica el trimestre en el que fue reportada la información por las entidades federativas.
ENTIDAD CT	Dimensión que indica la Entidad, municipio, localidad y centro de trabajo a la que pertenece la información.

Tabla 4.2 Dimensiones de los cubos

4.2 Flujo de información

Los procesos de extracción, transformación y carga han cambiado, esto debido a que la fuente de información ha sufrido cambios y la arquitectura del almacén de datos también tiene cambios.

El flujo de datos desde la extracción de las fuentes de información hasta la carga de los cubos MOLAP se muestra en la figura 4.4

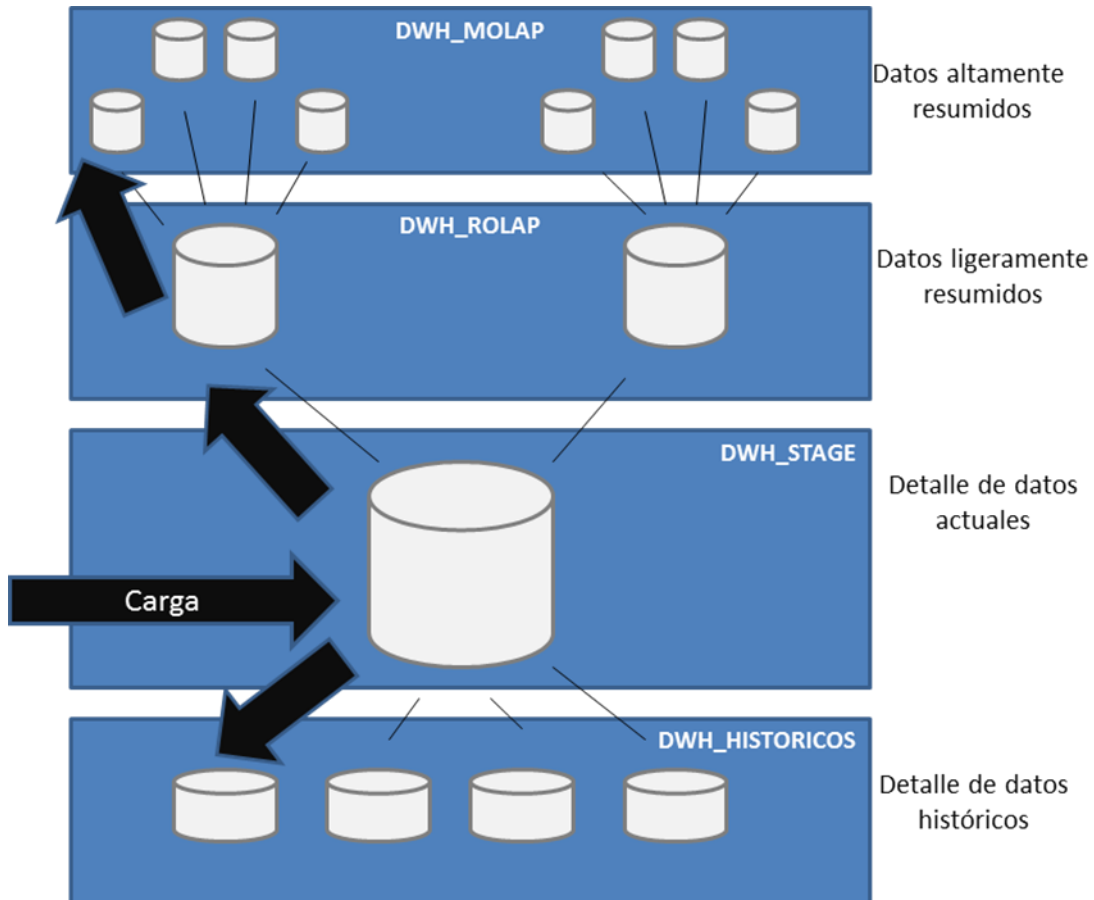


Figura 4.4 Flujo de información

4.3 Actualización de tableros de control

Después de hacer los ajustes al Data Warehouse se necesitó actualizar los tableros de control de Business Intelligence, esto con la finalidad de ver reflejados las nuevas cifras en los reportes.

Es importante mencionar que los tableros serán los mismos para cada uno de los trimestres y que el contenido de éstos se actualiza al momento de agregar información a la fuente única de explotación (Data Warehouse), por parte de la herramienta de Inteligencia de Negocios (OBIEE).

4.3.1 Construcción de los tableros de control

El tablero de plazas contiene información referente:

- Nómina Empleados (Ordinaria, Extraordinaria, Honorarios, Pagos Cancelados)
- Personal Comisionado
- Personal con Licencia
- Categoría y Plazas por Centro de Trabajo
- Fuente de Financiamiento (Federal y Estatal)

Herramientas Utilizadas:

- Oracle Business Intelligence Enterprise Edition (OBIEE):
 - Administration Tool
 - Answers
 - Interactive Dashboards
 - MapViewer

4.3.2 Lista de fuentes empleadas

Este es el primer paso para la creación de reportes y tableros. Se utilizó la herramienta de Oracle llamada "Administration Tool", dicha herramienta crea 3 tipos de capas necesarias para hacer un modelo de negocio que cumpla con las necesidades que la SEP requiere.

- **Capa Física.**- Esta capa contiene la conexión a los repositorios físicos de las fuentes de información (tablas, columnas, dimensiones). Ésta se puede conectar a diferentes tipos de fuentes, el requisito es que las fuentes sean consistentes entre sí para generar un modelo estable. Las fuentes que alimentan a la herramienta son dos (Ver figura 4.5):
 - Datawarehouse de Oracle
 - Cubos OLAP de Oracle

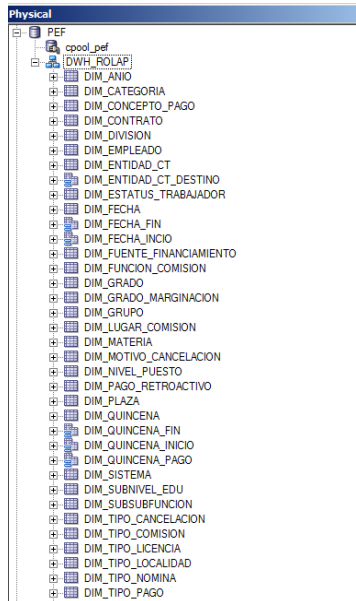


Figura 4.5 Capa física de la herramienta de BI.

- Modelado.-** Esta capa se encarga, como su nombre lo dice, de modelar las fuentes de información a las que se realizó la conexión en la capa física. Aquí se crean jerarquías, cálculos, combinaciones de fuentes según se requiera para cubrir los requerimientos de la Secretaría. Para plazas se crearon 5 modelos principales: Nómina(PEF), Cancelados, Plantilla, Comisionados, Licencias (Ver figura 4.6).

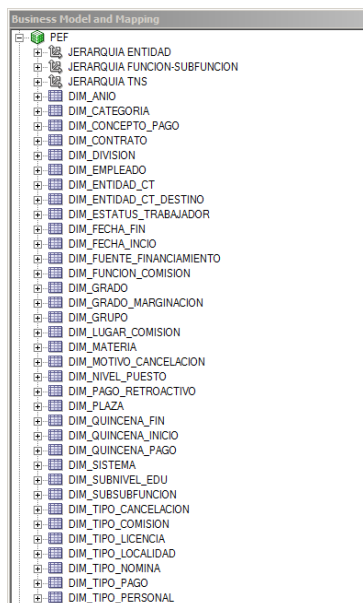


Figura 4.6 Capa de modelado de la herramienta de BI

4.3.3 Documentación de las reglas utilizadas

La tabla 4.3 presenta los cambios más significativos que se hicieron en los cálculos realizados por cada modelo (son 5 modelos principales):

Nombre Modelo	Columna que contiene la métrica	Cálculo	Descripción	Función
PEF	a)FACT_ORDINARIA_EXTRAORDINARIA b)FACT_PGC_CANCELADOS	Total_Plazas	Número de Plazas	count(distinct FACT_ORDINARIA_EXTRAORDINARIA.CL AVE_PRESUPUESTAL_ID)
PEF	a)FACT_ORDINARIA_EXTRAORDINARIA b)FACT_PGC_CANCELADOS	TOTAL_CT	Número de Centros de Trabajo	Count (Distinct DIM_CT_ENTIDADES.CT_ENTIDAD_ID)
PEF	a)FACT_ORDINARIA_EXTRAORDINARIA b)FACT_PGC_CANCELADOS	Total_Empleados	Número de Empleados	count(distinct DIM_EMPLEADO.EMPLEADO_ID)
PEF	a)FACT_ORDINARIA_EXTRAORDINARIA b)FACT_PGC_CANCELADOS	Total_Horas	Número de Horas que Trabaja el Empleado	sum(FACT_ORDINARIA_EXTRAORDINARIA.TOTAL_HORAS)
PEF	a)FACT_ORDINARIA_EXTRAORDINARIA b)FACT_PGC_CANCELADOS	Total_Percepción	Percepción quincenal	sum(FACT_ORDINARIA_EXTRAORDINARIA.TOTAL_PERCEPCION)

PEF	a)FACT_ORDINARIA_EXTRAORDINARIA b)FACT_PGC_CANCELADOS	Total_Deducción	Deducción Quincenal	sum(FACT_ORDINARIA_EXTRAORDINARIA.TOTAL_DEDUCCION)
PEF	FACT_COMISIONADOS	DEDUCCIÓN	Deducción que tiene el comisionado	sum(FACT_COMISIONADOS.DEDUCCION)
PEF	FACT_COMISIONADOS	PERCEPCIÓN	Percepción que recibe el comisionado	sum(FACT_COMISIONADOS.PERCEPCION)
PEF	FACT_COMISIONADOS	Total_Personas	Número de personas comisionadas	count(distinct DIM_EMPLEADO.EMPLEADO_ID)
PEF	FACT_LICENCIAS	Total_Plazas	Cantidad de plazas con licencia	COUNT(DISTINCT "DIM_CLAVE_PRESUPUESTAL"."CLAVE_PRESUPUESTAL_ID")
PEF	FACT_LICENCIAS	Total_Personas	Cantidad de personas con licencias	COUNT(DISTINCT "DIM_EMPLEADO"."EMPLEADO_ID")
PEF	FACT_LICENCIAS	Percepción	Percepción que recibe el personal con licencia	SUM("FACT_LICENCIAS"."PERCEPCION")

Tabla 4.3 Reglas utilizadas

- **Presentación.-** Esta capa se utiliza para mostrar el modelo definido en la capa anterior de manera entendible y práctica para el usuario encargado de hacer los reportes con la herramienta llamada Answers (ver figura 4.7).

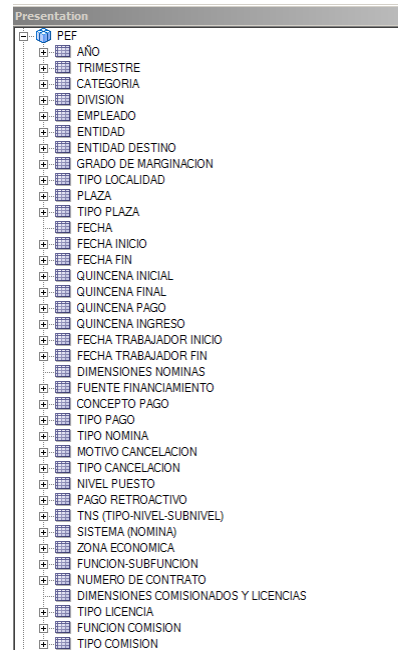


Figura 4.7 Capa de presentación de la herramienta de BI

Como se mencionó en la capa anterior se tienen 5 modelos principales:

- **Ordinaria Extraordinaria (Nómina):** Contiene información referente a la nómina ordinaria, extraordinaria y honorarios. A continuación se pueden visualizar los datos que se muestran ahí:
 - Montos de Percepciones
 - Montos de Deducciones
 - Categorías
 - Entidades
 - Centros de Trabajo
 - Nivel de Puesto del Empleado
 - Empleados
 - Zona Económica

- Cancelados: Contiene información de pagos cancelados determinados por:
 - Entidad.
 - Centros de Trabajo.
 - Categorías.
 - Nivel de Puesto del Empleado.
 - Empleados.
 - Zona Económica.
 - Periodos de Pago.
 - Montos de Percepciones.
 - Montos de Deducciones.

- Comisionados: Este modelo presenta información perteneciente al personal que se encuentra comisionado. A través de este modelo se puede saber:
 - En qué entidad está comisionado el personal.
 - Centro de Trabajo origen al que pertenece.
 - Centro de Trabajo destino a dónde se fue de comisión.
 - Función de la comisión.
 - Objetivo de la comisión.
 - Lugar donde se encuentra comisionado (en caso de ser un lugar fuera del sector educativo).
 - Tipo de comisión.
 - Fechas de Inicio y Fin de la comisión.

- Licencias: Se podrá analizar información referente al personal que se encuentra con permisos. La información que se podrá conocer de la persona es:
 - Entidad en la cual trabaja el empleado.
 - Centro de Trabajo del empleado.
 - Tipos de Licencia.
 - Fechas de inicio y fin de la licencia.
 - Percepción en caso de tratarse de una licencia con goce de sueldo.

Los reportes que se utilizan para la construcción del tablero están definidos en las actas correspondientes en los anexos presentados en la etapa anterior.

4.3.4 Vistas

A continuación se presenta el **Tablero de Plazas**, su navegación y elementos que lo constituyen.

Íconos de Navegación



Este ícono simboliza enlaces a otros tableros o reportes.



Al dar clic en este ícono se regresa al tablero del índice General.

Texto

Un texto en color significa un enlace a otro tablero o para realizar un análisis a profundidad.

I. Página Principal

En esta primera página (Figura 4.8) se presenta el índice general para acceder a los diferentes tableros.

The screenshot displays the Oracle Business Intelligence interface. At the top, there is a navigation bar with 'ORACLE Business Intelligence' and a search bar. Below this, the 'INDICE GENERAL' page is shown. It features a 'BIENVENIDO' section with a welcome message and a 'TABLEROS RNAME+' section with a list of categories: PEF (with sub-items: Nóminas, Comisionados, Licencias), ALUMNOS (with sub-item: Enlace Control Escolar), MAESTROS (with sub-item: Evaluación Universal, CUP), and 911 (with sub-item: Cifras por Entidad). To the right, there are sections for 'TABLEROS DE RECEPCION' (Recepción de Archivos Actual, Recepción de Archivos Histórico) and 'TABLEROS SNIE' (INDICADORES, RNAME, PORTAL RNAME). Below the dashboard, there is a banner for the SEP (Secretaría de Educación Pública) with a logo and a navigation menu: ACERCA DE LA SEP, EDUCACIÓN POR NIVELES, SEP EN LOS ESTADOS, BECAS, DOCENCIA, PROGRAMAS, SALA DE PRENSA. The banner also includes a large image of the SEP building and a text box with the headline 'La SEP aceptó recomen...' and sub-headlines 'Regresan a clases millon...' and 'Millones de estudiantes o vacaciones de invierno'.

Figura 4.8 Pagina pincipal

Al dar clic sobre algún tablero de PEF (Nomina) se presentará una pantalla similar a la siguiente (figura 4.9):

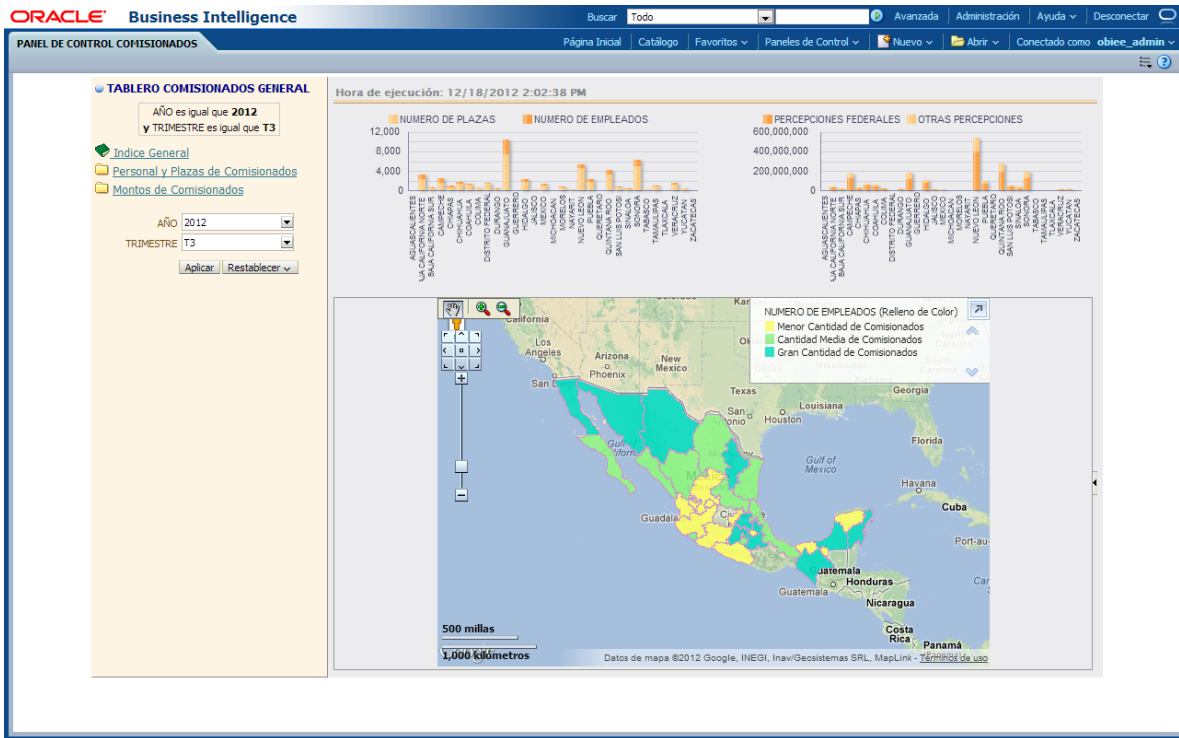


Figura 4.9 Tablero de Nomina

II. Nóminas

Desde este tablero se tiene acceso a información de las 4 nóminas:

- Ordinaria
- Extraordinaria
- Honorarios
- Pagos Cancelados



La información más importante con la que se podrá contar en estas nóminas será respecto a: Montos, número de personas, números de plazas, cantidad horas, información por categorías y tipos de nómina. El nivel de detalle va de Entidad bajando por Municipio, Localidad y Centro de Trabajo (Figura 4.10).

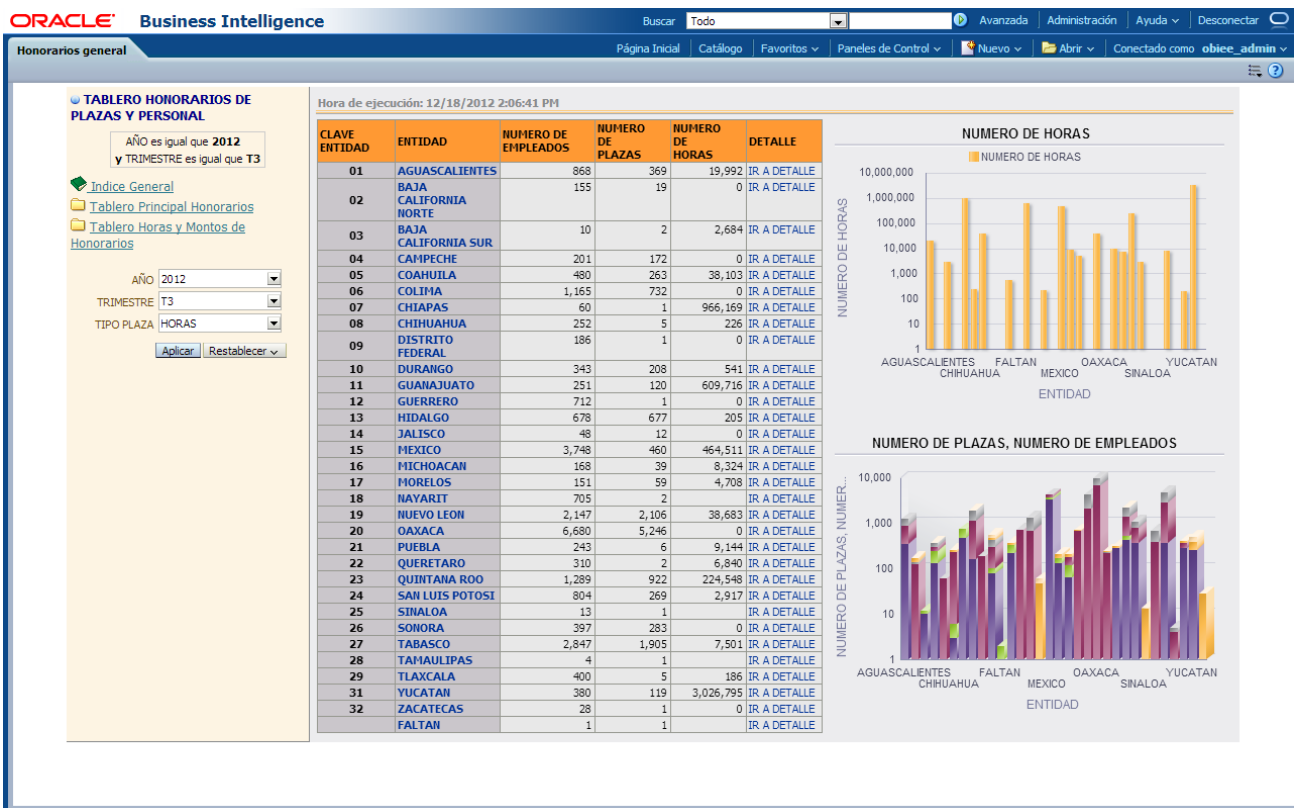


Figura 4.10 Información de nominas

III. Ordinaria/ Extraordinaria

Aquí se podrá analizar con mayor detalle al empleado y su comportamiento en la nómina, se tiene la opción para realizar pivotes de columnas y aplicar los filtros según requiera el usuario que esté analizando la nómina (Figura 4.11).

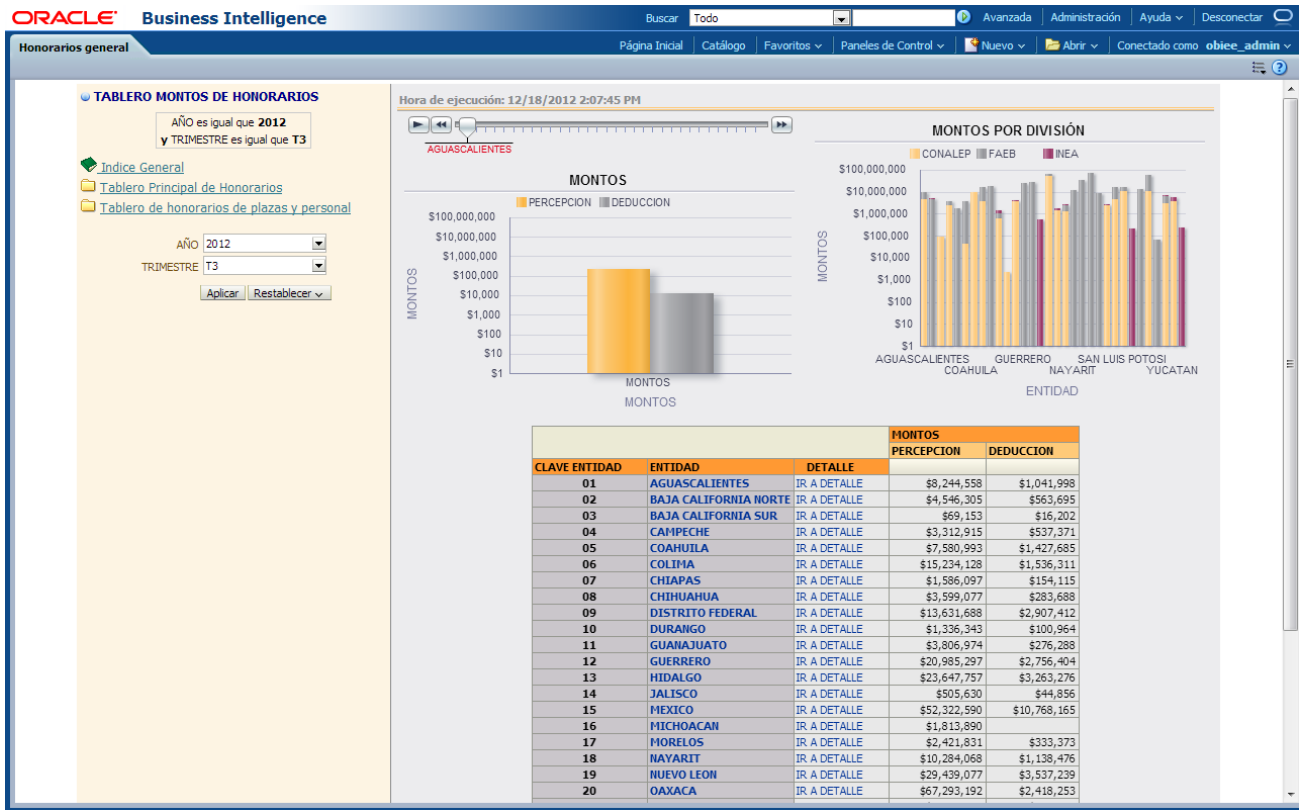


Figura 4.11 Nomina ordinaria/Extraordinaria

V. Personal Comisionado

Aquí se reportan exclusivamente las Categoría(s)/Plaza(s) reportadas por la EF en que está comisionado un trabajador (No todas su plazas).

Resumen Comisionados (Ver figura 4.12 y 4.13)

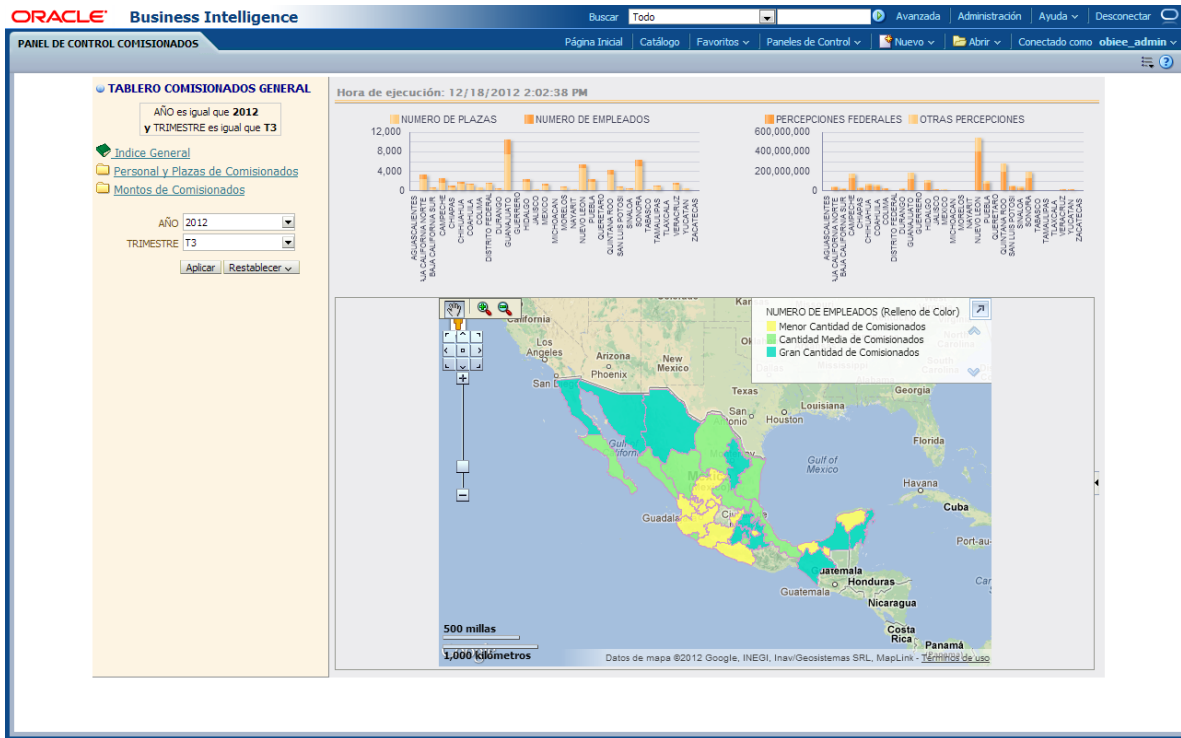


Figura 4.12 Personal Comisionado

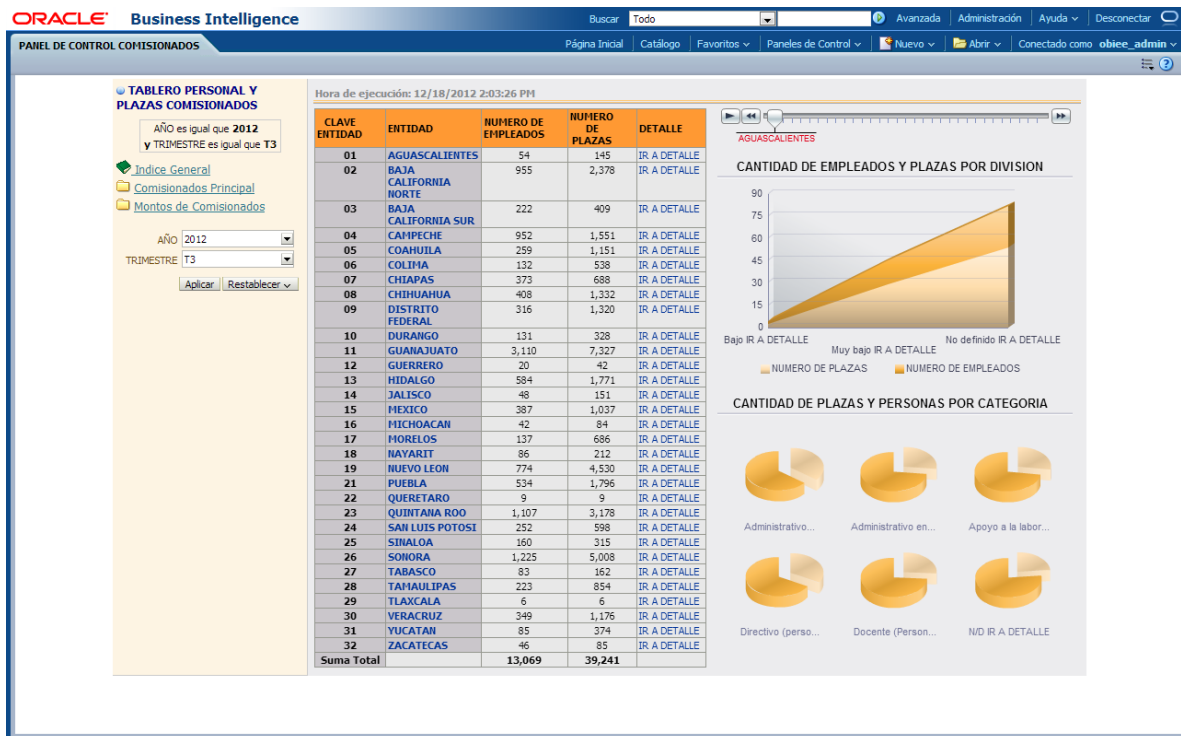


Figura 4.13 Personal comisionado detalle

VI. Personal con Licencia

Aquí se reporta exclusivamente las Categoría(s)/Plaza(s) reportadas por la EF en las cuales se autorizó una licencia al trabajador (No todas su plazas, ver figura 4.14).

Resumen Personal con Licencia

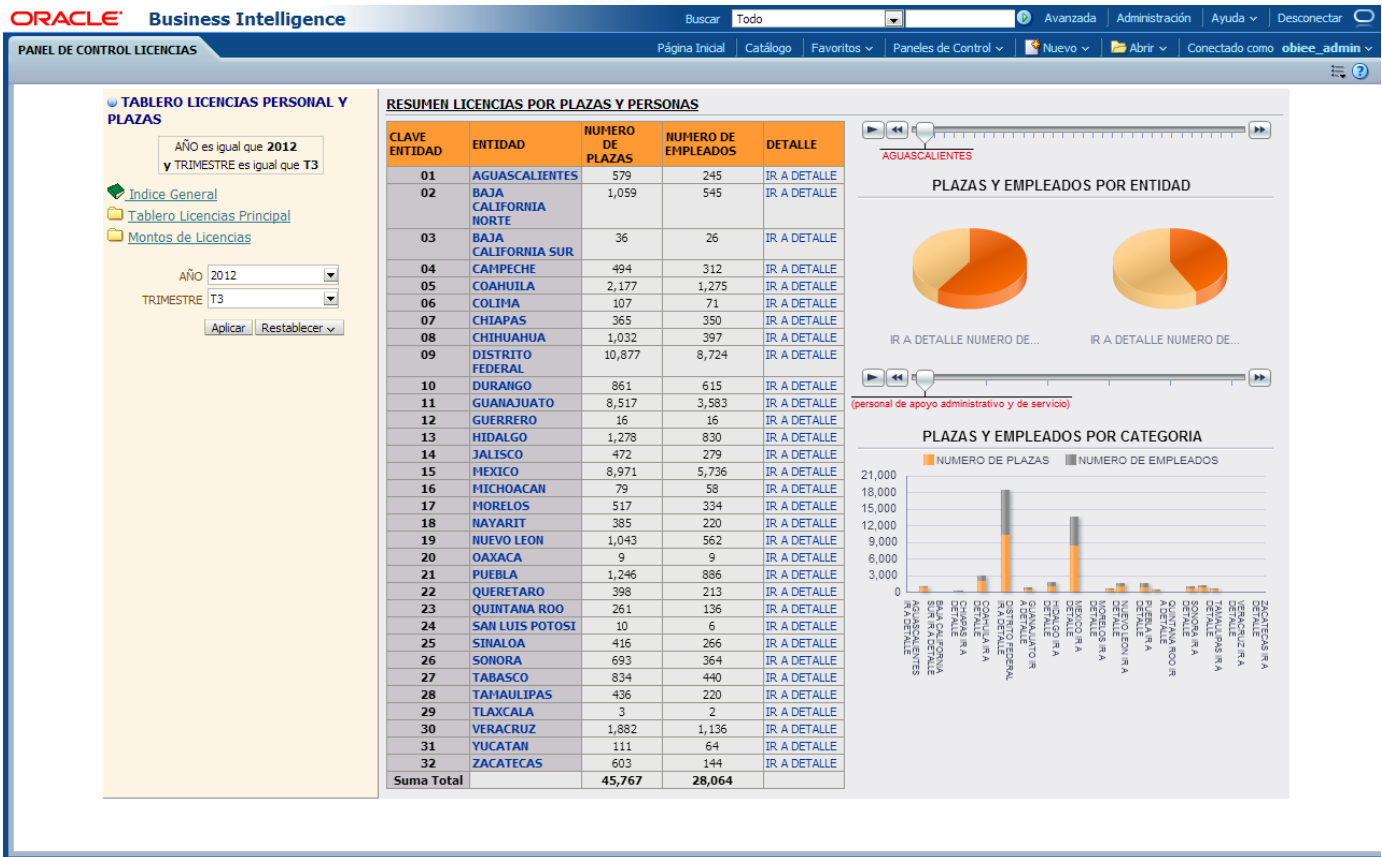


Figura 4.14 Personal con licencia

VII. Mapas

Los tableros principales de PEF(nómina) presentan un mapa donde pueden identificarse a nivel entidad los valores de montos, número de plazas y personas (figura 4.15).



Figura 4.15 Mapas

Después de realizar estos cambios, La SEP logro obtener de la herramienta de BI datos más precisos, y que concordaban con los datos que ellos obtenían usando su antiguo sistema, por lo que ahora el sistema del RNAME (Registro Nacional de Alumnos, Escuelas y Maestros) ha sido catalogado como de interés nacional, lo que demuestra la importancia y la confianza que la SEP tiene en el sistema que le desarrollamos.

Conclusión

Este desarrollo de un Data Warehouse y Business Intelligence se logró convertir en un proyecto de interés nacional, ya que logra sus objetivos integrando todas las bases de datos de las 32 entidades en una base de datos nacional. Por lo que ahora se pueden hacer análisis de la información de alumnos escuelas y maestros que antes hubiera sido muy difícil poder conseguir.

Esto me ayudo bastante en lo profesional porque ahora cuento con un buen entendimiento de la estructura que un data warehouse debe tener, y el entendimiento y habilidad de manejar la herramienta de Oracle Business Intelligence, lo cual me dio un buen puesto en la empresa donde trabajaba durante este proyecto.

Para la SEP el sistema que le proporcionamos le ayudo bastante, y ahora es éste el que utilizan para basar sus decisiones, y el que les sirve para tener un mejor entendimiento de los resultados que como entidad o secretaría de gobierno obtienen.

Ahora la SEP tiene una visión global de en qué se gasta el dinero que le da a cada entidad de la república, cuantos profesores tienen, cuantos profesores están comisionados o en licencia, gracias a la base de datos nacional que le fue entregada en donde puede realizar estos análisis.

También ahora son capaces de detectar irregularidades que antes no podían detectar, como profesores que cobren por más plazas de las que pueden tener, profesores que cobren por plazas en diferentes estados de la república entre varias cosas más, lo cual les ayuda a tomar las decisiones correspondientes basadas en un mejor conocimiento.

En cuanto a los Alumnos la SEP tiene un mejor control sobre los resultados de la prueba ENLACE de una manera gráfica e intuitiva, de una manera rápida puede saber indicadores como: ¿en qué nivel se encuentran los alumnos de primaria y secundaria?, ¿Afecta el grado de marginación del área geográfica donde viven a al nivel de los alumnos en la prueba ENLACE?, etc.

Todas las herramientas lograron cumplir sus objetivos por los que fueron requeridos y ahora la SEP ya tiene un sistema para lograr una mejor toma de decisiones basados en sus reportes e indicadores solicitados por ellos mismos y lograr que esas decisiones puedan aumentar el nivel de la educación en México.

Se puede esperar que en un futuro la SEP pueda integrar más fuentes de información a esta base de datos nacional para realizar una mejor gestión de sus recursos y lograr la toma de decisiones más acertadas posibles.

Glosario

Termino	Definición	Página
Batch.	Es un archivo de procesamiento por lotes,	12
Base de datos	Se define una base de datos como una serie de datos organizados y relacionados entre sí, los cuales son recolectados y explotados por los sistemas de información de una empresa o negocio en particular.	3
BI o Business Intelligence	Conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización o empresa.	4
Clúster	Se aplica a los conjuntos o conglomerados de computadoras construidos mediante la utilización de hardwares comunes y que se comportan como si fuesen una única computadora.	15
Cubo de información	Es una base de datos multidimensional, en la cual el almacenamiento físico de los datos se realiza en un vector multidimensional. Los cubos OLAP se pueden considerar como una ampliación de las dos dimensiones de una hoja de cálculo.	18
Data mart	Un Data mart es una versión especial de almacén de datos (data warehouse). Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones.	21
Data Quality	Calidad de datos se refiere a los procesos, técnicas, algoritmos y operaciones encaminados a mejorar la calidad de los datos existentes en empresas y organismos.	7
Data warehouse	Es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza.	4
Drill down	Significa pasar de datos sumariados o resumidos a una información más detallada	11

ETL	Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio.	9
HTML	Lenguaje de marcado predominante para la elaboración de páginas web que se utiliza para describir y traducir la estructura y la información en forma de texto, así como para complementar el texto con objetos tales como imágenes.	17
Layout	La noción de layout suele utilizarse para nombrar al esquema de distribución de los elementos dentro un diseño	19
LDAP	Protocolo a nivel de aplicación que permite el acceso a un servicio de directorio ordenado y distribuido para buscar diversa información en un entorno de red.	16
MDM	Consiste en un conjunto de procesos y herramientas que define y gestiona de forma consistente las entidades de datos no transaccionales de una organización. Busca, por lo tanto, recopilar, agregar, identificar, asegurar la calidad y la persistencia y distribuir los datos de forma uniforme en dicho contexto.	7
MOLAP	Se trata de una alternativa a la tecnología ROLAP (OLAP-Relacional). Aunque ambos tipos de herramientas están diseñadas para realizar análisis de datos a través de un modelo de datos multidimensional, MOLAP se diferencia significativamente en que requiere un preprocesamiento y almacenamiento de la información contenida en el cubo OLAP. MOLAP almacena estos datos en una matriz de almacenamiento multidimensional optimizada, más que en una base de datos relacional (o en un ROLAP).	50
OLAP	Procesamiento analítico en línea (On-Line Analytical Processing). Es una solución utilizada en el campo de la llamada Inteligencia empresarial (o Business Intelligence) cuyo objetivo es agilizar la consulta de grandes cantidades de datos.	16
Portal Web	Un Portal Web está basado en un Sistema de Administración de Contenidos, el cual permite publicar información en la web de manera fácil y rápida, sin tener conocimientos de programación ni de diseño de páginas web.	7

Portlets	Son componentes modulares de las interfaces de usuario gestionadas y visualizadas en un portal web. Los portlets producen fragmentos de código de marcado que se agregan en una página de un portal.	17
RDBMS	Sistema de gestión de bases de datos relacionales	7
RHEL	Red Hat Enterprise Linux	7
ROLAP	Significa Procesamiento Analítico OnLine Relacional, es decir, se trata de sistemas y herramientas OLAP (Procesamiento Analítico OnLine) construidos sobre una base de datos relacional.	50
SQL	El lenguaje de consulta estructurado o SQL (por sus siglas en inglés structured query language) es un lenguaje declarativo de acceso a bases de datos relacionales que permite especificar diversos tipos de operaciones en ellas.	15
Tabla de hechos	En un data warehouse, una tabla de hechos (o tabla fact) es la tabla central de un esquema dimensional (en estrella o en copo de nieve) y contiene los valores de las medidas de negocio. Cada medida se toma mediante la intersección de las dimensiones que la definen, dichas dimensiones estarán reflejadas en sus correspondientes tablas de dimensiones que rodearán la tabla de hechos y estarán relacionadas con ella.	18
Web service	Es una tecnología que utiliza un conjunto de protocolos y estándares que sirven para intercambiar datos entre aplicaciones. Distintas aplicaciones de software desarrolladas en lenguajes de programación diferentes, y ejecutadas sobre cualquier plataforma, pueden utilizar los servicios web para intercambiar datos en redes de ordenadores como Internet	11
XML	es un lenguaje de marcas desarrollado por el World Wide Web Consortium (W3C). Deriva del lenguaje SGML y permite definir la gramática de lenguajes específicos (de la misma manera que HTML es a su vez un lenguaje definido por SGML) para estructurar documentos grandes. A diferencia de otros lenguajes, XML da soporte a bases de datos, siendo útil cuando varias aplicaciones se deben comunicar entre sí o integrar información.	17

Referencias

<http://www.oracle.com/us/solutions/business-analytics/business-intelligence/overview/index.html>

<http://www.oracle.com/us/solutions/business-analytics/business-intelligence/enterprise-edition/overview/index.html>

<http://www.oracle.com/us/products/database/overview/index.html>

<http://www.oracle.com/us/products/applications/master-data-management/overview/index.html?origref=http://www.oracle.com/us/products/database/overview/index.html>

<http://www.oracle.com/us/products/middleware/data-integration/overview/index.html>

<http://www.rname.sep.gob.mx/>

http://pic.dhe.ibm.com/infocenter/idm/docv3/index.jsp?topic=%2Fcom.ibm.datatools.dimension.al.ui.doc%2Ftopics%2Fc_dm_star_schemas.html

Kimball, Ralph, *The data warehouse life cycle toolkit*.

<http://www.kimballgroup.com>