



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Facultad de **C**ontaduría y **A**ministración

Métodos para adicionar nodos de manera rápida y efectiva a clusters de alto rendimiento y casos prácticos

Diseño de un Sistema o Proyecto

Omar Alejandro Pompa García



México Cd.Mx.

2017



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Facultad de **C**ontaduría y **A**ministración

Métodos para adicionar nodos de manera rápida y efectiva a clusters de alto rendimiento y casos prácticos

Diseño de un Sistema o Proyecto

Que para obtener el título de:
Licenciado en Informática

Presenta:
Omar Alejandro Pompa García

Asesor:
M. I. Lourdes Yolanda Flores Salgado



México Cd.Mx.

2017

Índice general

Dedicatoria	1
Agradecimientos	2
Introducción	7
Problema de investigación	7
Justificación	8
Objetivos	8
Alcance	8
Limitaciones	9
Aportaciones	9
Estructura Capitular	9
1. Dirección General de Cómputo y de Tecnologías de Información y Comunicación	11
1.1. Antecedentes	11
1.2. Misión	14
1.3. Visión	14
1.4. Organigrama de la DGTIC	14
1.5. Organigrama de la Coordinación de Supercómputo	15
2. Marco Teórico y Necesidad	16
2.1. Supercomputadora	16
2.2. Cómputo paralelo	16
2.2.1. <i>Massively Parallel Processors</i> (MPP)	16
2.2.2. <i>Symetric Multiprocesors</i> (SMP)	17
2.2.3. <i>Cache-Coherent Nonuniform Memory Access</i> (CC-NUMA)	17
2.2.4. Cómputo Distribuido	17
Clústeres	18
Tipos de Clústeres	19
Grids	20
2.3. GNU/Linux	20
2.4. Imagen de disco y Clonación	20
2.5. Estructura lógica de un disco duro	21
2.6. Herramientas para imágenes y clonación	21
2.6.1. dd	21
2.6.2. Clonezilla	22
Características	23

ÍNDICE GENERAL

Ventajas	23
Desventajas	23
2.6.3. Mondo Rescue	24
Características	24
Requerimientos	24
2.6.4. Redo Backup and Recovery	25
Ventajas	25
Desventajas	25
Requerimientos	25
2.6.5. Comparativa de las herramientas	26
2.6.6. Necesidad la Coordinación de Supercómputo en DGTIC	26
3. Guías de instalación y propuesta para la solución del problema	28
3.1. Guía de Mondo Rescue	28
3.1.1. Instalación y configuración	28
3.1.2. Restablecimiento de la imagen	34
3.2. Guía para Redo Backup and Recovery	35
3.2.1. Configuración	35
3.2.2. Restauración de la imagen	38
3.3. Guía para Clonezilla Server Edition	39
3.3.1. Instalación y configuración	39
Configuración de DRBL	41
Configuración en el servidor para guardar una imagen	48
Configuración en el servidor para restaurar una imagen	52
Guardar imagen desde el cliente	56
Restaurar la imagen desde el cliente	57
4. Resultados	59
4.1. Comparación de tiempos en el proceso de generación y recuperación	59
4.2. Elección de <i>software</i> para implementación	61
Conclusiones	62
Anexos	64
Glosario	64
Referencias	66

Dedicatoria

Este proyecto está dedicado principalmente a mi familia.

A mi mamá, que siempre ha estado para mí, siendo de las personas más importantes en mi vida.

Agradecimientos

A mis padres:

Por los cuidados y el apoyo que nunca han dejado de brindarme.

A mis hermanos:

Ivan:

Por guiarme como hermano mayor y apoyarme en mi vida académica y laboral siempre que lo he necesitado.

Miriam:

Por estar al pendiente de mí cuando hace falta alguien y apoyarme durante el trayecto de mi carrera.

Edgar y Paco:

Por todo el apoyo brindado y compartir gustos personales.

Mauricio:

Por ser el nuevo miembro de la familia y mejorar nuestras vidas.

A mi asesora, M. Yolanda Flores Salgado:

Por apoyarme durante todo el periodo de este trabajo con sus enseñanzas y conocimientos transmitidos.

A la Universidad Nacional Autónoma de México:

Por darme la oportunidad de ser parte de ella desde el bachillerato, aprendiendo a ser mejor profesional y persona. Por brindarme los mejores momentos que he vivido.

A mis profesores y a la Facultad de Contaduría y Administración:

Por la formación académica obtenida para ingresar al entorno laboral.

Al M. Alberto López Vivas:

Por el apoyo recibido laboralmente. Por darme las facilidades técnicas y prácticas para la realización de las pruebas clave para este proyecto.

Al LI Eduardo Iván Ortega Alarcón y a la DGTIC:

Por el tiempo dedicado.

Al Instituto de Investigaciones en Materiales:

Por dejarme formar profesionalmente y realizar las pruebas de los elementos necesarios.

«UNIX is very simple, it just needs a genius to understand its simplicity.»

Dennis Ritchie

Introducción

Con el paso del tiempo, los seres humanos utilizamos cada vez más las computadoras para diversos usos, teniendo una demanda considerable. El aprovechamiento de estos recursos es importante para muchas áreas de estudio, ya que el procesamiento de datos es mucho más rápido y eficaz que al hacerlo manualmente (operaciones matemáticas, por ejemplo).

Incluso en estos tiempos, las computadoras personales o servidores especializados pueden ser muy tardados en concluir trabajos que necesiten procesar mucha información. Si se requiere que se reduzca el tiempo en concluir los trabajos, se puede optar por utilizar supercomputadoras. El problema es que estas máquinas son muy costosas y una alternativa es el uso de clústeres que pueden estar constituidos por muchas máquinas conectadas entre sí y ser muy potentes.

Problema de investigación

Los clústeres actualmente pueden llegar a estar formados por miles de computadoras en las cuales, cada equipo debe de tener sistema operativo, aplicaciones necesarias y configuraciones indispensables para su correcto funcionamiento. Esto sin duda se lleva mucho tiempo (dependiendo del número de máquinas) y es propenso a errores al momento de instalaciones o configuraciones que puede llegar a repercutir en más de un equipo.

La Coordinación de Supercómputo de la Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC) de la UNAM administra varias supercomputadoras que deben de contar con alta disponibilidad, por lo que al agregar o al fallar uno o varios equipos, se debe de resolver el problema en un periodo de tiempo corto, aunando que las

computadoras pueden ser difíciles de configurar.

Justificación

Actualmente, en la DGTIC se tiene un *software* de paga que realiza la función de copias enteras de discos duros, solo que la licencia permite un determinado número de copias y restablecimiento de la imagen y renovar o adquirir una licencia nueva es muy costoso.

Otro inconveniente es que el programa está hecho para trabajar con equipos de marca HP, lo que dificulta la restauración en dado caso de querer utilizar un equipo de otra marca, ya que se debe de configurar manualmente el equipo del que se quiera sacar copia o restaurar.

Las pruebas de diferentes herramientas para clonación e imágenes de discos duros como Mondo Rescue, Redo Backup y Clonezilla nos darán un panorama para poder elegir cuál es el *software* más apto para implementar y resolver la problemática que actualmente existe.

Objetivos

- Estudiar cómo está constituido un clúster
- Instalar las herramientas en un clúster de pruebas
- Configurar adecuadamente las herramientas mencionadas anteriormente para poder comparar los beneficios que tiene cada una para elegir la más capaz y eficiente
- Implementar de un herramienta en un clúster de la DGTIC

Alcance

Dejar en funcionamiento una herramienta para poder clonar o hacer una imagen del disco duro de una computadora de clúster y automatizar su uso, configurando la aplicación para que no haya necesidad de hacerlo más de una vez.

Limitaciones

- Las pruebas se realizarán en equipos de cómputo de menor capacidad a los de la DGTIC
- Se trabajará con tres herramientas
 - Mondo Rescue
 - Redo Backup
 - Clonezilla

Aportaciones

- Tener una nueva fuente de información sobre el supercómputo y sus bases, ya que la mayoría de la información relacionada con este tema solo se puede consultar en inglés
- Aplicar el *software* especializado para la creación de imágenes de discos duros, para conocer sus debilidades, fortalezas y sus especificaciones y así poder contemplar la herramienta adecuada para alguna necesidad
- Resolver la problemática que actualmente hay tanto en DGTIC como en alguna otra institución que pueda tomar como referencia esta investigación, tratado de evitar los errores, el tiempo excesivo, la complejidad y querer reducir costos a la hora de adquirir un *software* para cuando se requiera añadir o instalar algún equipo de supercómputo o cualquier computadora de la que desee tener una copia de seguridad de un disco duro completo

Estructura Capítular

Capítulo I

Recuento histórico de los orígenes y de cómo se fue transformando de acuerdo a sus necesidades lo que hoy es la Dirección General de Tecnologías de Información y Comunicación. Se cita su misión, visión y la estructura orgánica tanto de la DGTIC, como de la Coordinación de Supercómputo.

Capítulo II

Fundamentación de la práctica con el marco teórico. Descripción de las herramientas a utilizar, así como la comparación para la selección de la más acorde con las necesidades. Puntualización de la problemática y los aspectos a mejorar, contemplando la resolución del problema

Capítulo III

Guías de uso, instalaciones y configuraciones de Redo Backup, Mondo Rescue y Clonezilla Server Edition.

Capítulo IV

Los resultados obtenidos en las pruebas y en la implementación, comparando tiempos y características

CAPÍTULO 1

Dirección General de Cómputo y de Tecnologías de Información y Comunicación

1.1 Antecedentes

En la década de los cincuenta, los cálculos matemáticos eran muy complejos y tardaban meses en resolverse, por esta razón, el Ingeniero Sergio Beltrán López fue el encargado de visitar la Universidad de California en Los Ángeles (UCLA) para ver el funcionamiento de las computadoras IBM al resolver los diversos problemas de ecuaciones.

Años más tarde, en 1955 se creó el Departamento de Cálculo Electrónico en la planta baja de la Facultad de Ciencias (lo que hoy es la Torre II de Humanidades) para hacer investigación en diversas áreas de la ciencia como la física, las matemáticas, la actuaría, entre otras. Ese mismo año, la Dirección de Servicios Escolares ya cuenta con la Sección de Máquinas, aunque otras fuentes dicen que su creación fue en 1962¹.

Tres años después, el 8 de junio de 1958, gracias al éxito y recursos obtenidos, paso de ser un departamento a ser el Centro de Electrónico de Cálculo, que tiempo después cambió su nombre a Centro de Cálculo Electrónico (CCE) y adquiriendo la primera computadora en toda América Latina, el equipo IBM 650 o Cerebro Electrónico, que funcionaba con válvulas de vacío (bulbos), un tambor magnético que sirve para almacenar datos e incluía una perforadora y una lectora de tarjetas. La IBM 650 se adquirió por el esfuerzo que realizaron el Rector de la UNAM, Nabor Carrillo, el Dr. Alberto Barajas, Coordinador del Consejo Técnico

¹<http://www.historiadelcomputo.unam.mx/cronologia.html>

de la Investigación Científica, el Director de la Facultad de Ciencias, Carlos Graeff y del Ingeniero Sergio Beltrán que en ese entonces era el Director del Centro de Cálculo Electrónico.

La UNAM quería obtener un equipo más potente y nuevo, la IBM 704 con un descuento del 60 % pero por el presupuesto que se dio en ese entonces, no era suficiente.

En 1966, se obtiene la computadora IBM 1440 para ser usada en la contabilidad y la administración. A cargo del equipo estuvo el Departamento de Sistemas del Patronato Universitario.

La Dirección General de Sistematización de Datos, se funda a partir de la unificación de la Sección de Máquinas y del Departamento de Sistemas del Patronato Universitario en el año de 1967. Gracias a esto, el área de administración podía contar con soporte de cómputo.

Años después, la Coordinación de Ciencias que estaba a cargo del CCE, decide en 1970, convertirlo en el Centro de Investigaciones en Matemáticas Aplicadas, Sistemas y Servicios (CIMASS) como consecuencia de la demanda que había para el Centro. El Rector Pablo González Casanova proporcionó los recursos necesarios para que fuera posible el cambio. El CIMASS se dedicó a investigar lo relacionado con la computación.

El 15 de marzo de 1973 se divide el CIMASS en el Centro de Investigaciones en Matemáticas Aplicadas y en Sistemas (CIMAS) que se encargaría de la investigación en matemáticas y computación y en el Centro de Servicios de Cómputo (CSC) que atendería a la comunidad de la Universidad y para la administración, siendo sus directores el Dr. Tomás Garza Hernández y el Ing. Francisco Martínez Palomo respectivamente.

El CIMAS se convirtió en el Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) el 10 de marzo de 1976 cuando fue aprobado por el Consejo Universitario, quedándose como director el Dr. Garza Hernández, siendo su principal objetivo hacer investigación en ciencia e ingeniería de la computación, matemáticas y sistemas.

Pensando en la mejora del cómputo y en el constante cambio tecnológico, la UNAM se dio a la tarea de crear el Programa Universitario de Cómputo (PUC) que fue aprobado por el Rector Dr. Octavio Rivero Serrano el 15 de octubre de 1981. El PUC desarrolló infraestructura para el apoyo tanto para México como para la Universidad, incluyendo cursos para bachillerato y licenciatura. Apoyaba en áreas tan diversas como la investigación, la docencia y la administración.

A mediados de mayo de 1985, por acuerdo del Rector Jorge Carpizo, se crea el Consejo Asesor de Cómputo, la Dirección General de Servicios de Cómputo para la Administración y la Dirección General de Servicios de Cómputo Académico (DGSCA). Esta última sustituye y adquiere el presupuesto y el personal del Programa Universitario de Cómputo.

La DGSCA contaba con una Secretaría General y se dividía en la Dirección de Cómputo para la Administración Académica, la Dirección de Cómputo para la Docencia y la Dirección de Cómputo para la Investigación. Este cambio ocurrió porque el PUC no contaba con la infraestructura y características adecuadas para la Universidad y debido al papel fundamental del cómputo, se optó por crear esta nueva Dirección para solventar estas deficiencias. Algunas acciones destacables de la DGSCA fueron la adquisición de CRAY YMP 432, la primera supercomputadora en toda América Latina; la obtención de certificación por el Forum Incident Response Security Team (FIRST)² para los incidentes de seguridad en cómputo; la inauguración del Observatorio de Visualización Inmersiva IXTLI; la implementación de la red inalámbrica (RIU); y la adquisición de la súpercomputadora KanBalam.

El 27 de septiembre de 2010, el Rector Dr. José Narro Robles a través de un acuerdo, cambia el nombre de la DGSCA a la Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC) como consecuencia de la evolución de las tecnologías. El nombre de la DGSCA no iba acorde a lo que desarrollaba y estudiaba la institución, así, el nuevo nombre que se le dio, cubre tanto todos los servicios que ofrece como lo que se estudia.

La DGTIC se caracteriza por dar cursos de preparación para mantenerse actualizados con las nuevas tecnologías; hacer crecer la red de la Universidad con telefonía y videoconferencia; digitalizar la información de la UNAM y que todo tipo de personas puedan tener acceso, incluyendo personas con discapacidades físicas; apoyar en el área de la investigación, la academia y la cultura.

La DGTIC se encarga de todo lo relacionado con el cómputo en la UNAM, desde soporte técnico, redes, servidores, desarrollo de infraestructura, supercómputo, entre otras cosas.

La Coordinación de Supercómputo en la Universidad, tiene sus orígenes en 1991 con la llegada de Sirio (CRAY YMP 432 mencionada anteriormente) y posteriormente se adquirieron otras cuatro supercomputadoras, Berenice, Bakliz, KanBalam y Miztli. Este departamento coordina y administra lo relacionado a esta área para poder brindar apoyo a la comunidad

²Organización reconocida mundialmente que detecta incidencias en los sistemas de información y elabora métodos para contrarrestarlos o prevenirlos.

de la UNAM en temas de investigación y educación.

1.2 Misión

«La Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC), contribuye al logro de los objetivos de la UNAM, como punto de unión de la comunidad universitaria, para aprovechar los beneficios que las tecnologías de la información y las comunicaciones pueden aportar a la docencia, la investigación, la difusión de la cultura y la administración universitaria.» (UNAM, 2015). [Recuperado de http://transparencia.unam.mx/files/documentos/Metas_Objetivos_2015.pdf]

1.3 Visión

«La Dirección General de Cómputo y de Tecnologías de Información y Comunicación, es la dependencia que crea las innovaciones en tecnologías de información y comunicación, para impulsar a la Universidad Nacional Autónoma de México, como líder en Latinoamérica.» (UNAM, 2015). [Recuperado de http://transparencia.unam.mx/files/documentos/Metas_Objetivos_2015.pdf]

1.4 Organigrama de la DGTIC

En la Figura (1.1) podemos ver cómo está conformada la DGTIC en sus diferentes direcciones.

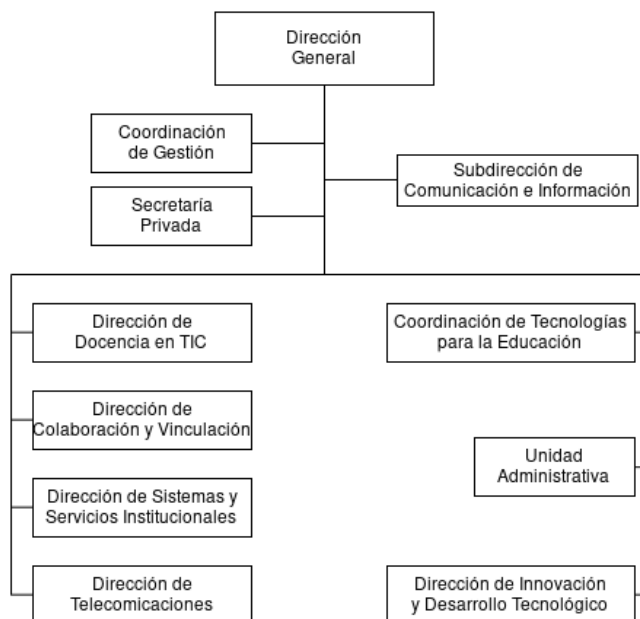


Figura 1.1: Organigrama de la DGTIC [Recuperado de <http://tic.unam.mx/organigrama.html>]

1.5 Organigrama de la Coordinación de Supercómputo

La Coordinación de Supercómputo de la DGTIC, está conformada por la siguiente estructura, como se muestra en la Figura (1.2)

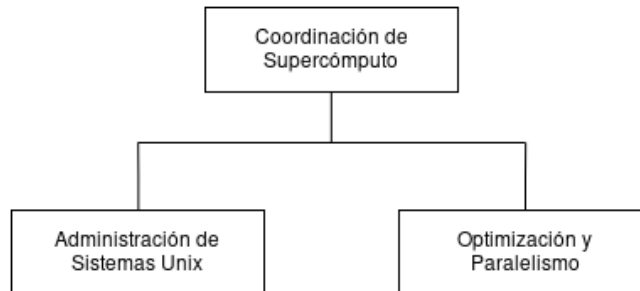


Figura 1.2: Organigrama de la Coordinación de Supercómputo

Marco Teórico y Necesidad

2.1 Supercomputadora

Es una computadora con desempeño sobresaliente. Cuenta con una enorme capacidad de memoria, disco duro y procesadores más potentes que los convencionales. De la misma manera, es más rápida la comunicación entre el *hardware* mencionado anteriormente pudiendo hacer millones de operaciones por segundo. Las supercomputadoras sirven para poder trabajar con gran cantidad de información terminando sus tareas en un periodo de tiempo corto (a comparación con una computadora personal), usándose principalmente en problemas matemáticos.

2.2 Cómputo paralelo

Es el uso de dos o más procesadores que trabajan simultáneamente con el fin de completar un solo trabajo. Esto se logra distribuyendo el trabajo en tareas para cada procesador.

Para poder clasificar las diversas arquitecturas, se toma en cuenta cómo se distribuye la memoria, los procesadores y las conexiones para poder comunicarse. Los más conocidos son los siguientes:

2.2.1 *Massively Parallel Processors* (MPP)

Está compuesto por cientos de nodos conectados entre sí por algún *switch* o una red de alta velocidad. Los nodos son totalmente independientes ya que pueden contar con *hardware* distinto al de los demás y con diversos periféricos conectados, de igual manera, el *software* puede variar al resto de los demás nodos y tener diferentes sistemas operativos.

Debe de existir un nodo que se encargue de mandar datos y órdenes al resto para poder realizar sus tareas, a esto se le llama arquitectura *shared nothing*.

2.2.2 *Symmetric Multiprocessors (SMP)*

Tiene una arquitectura *shared everything* en la cual, se comparten todos sus recursos tales como memoria, procesadores, almacenamiento y el sistema operativo es el mismo en todos los nodos configurados.

2.2.3 *Cache-Coherent Nonuniform Memory Access (CC-NUMA)*

Este tipo de cómputo paralelo cuenta con memoria NUMA (*Non-Uniform Memory Access*) que se usa para el multiproceso en el que la lectura de los datos es más rápida en el propio procesador que en la de los demás. También está constituido con coherencia de caché que se define como la consistencia de los datos guardados en los cachés de los diferentes procesadores, evitando que los datos almacenados en un caché sean usados en lugar de utilizar el caché de otro procesador con datos íntegros o recientes, tal como se muestra en la Figura (2.1).

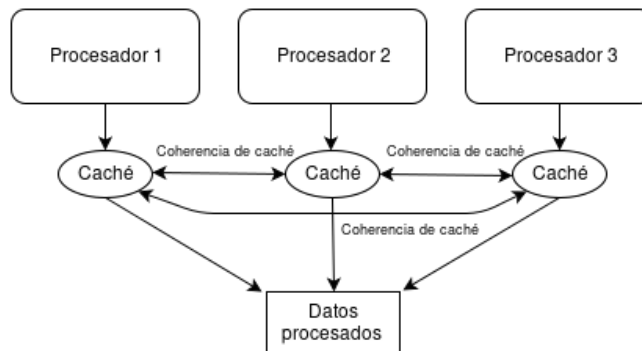


Figura 2.1: Coherencia de caché

2.2.4 *Cómputo Distribuido*

Este tipo de sistema paralelo, se refiere a que la información que se procesa se reparte entre varias máquinas que son independientes. Sus características son:

- **Concurrencia.** Diferentes procesos pueden estar ejecutándose simultáneamente y comunicándose entre sí
- **Escalabilidad.** Si se requiere aumentar la potencia, se pueden agregar dispositivos para tener una capacidad mayor
- **Recursos compartidos.** El *software* y *hardware* puede ser usado por varios equipos como los discos duros o archivos

Los tipos de cómputo distribuido más conocidos son los siguientes:

Clústeres

Son un conjunto de computadoras que se pueden comunicar entre sí por medio de una red LAN y funcionan como una sola máquina, combinando sus recursos, como memoria RAM, almacenamiento y procesadores.

Se componen de un nodo principal (o nodo maestro) que es el encargado de distribuir procesos, de enviar y recibir los datos procesados al término de un trabajo; y de los nodos esclavos (comúnmente llamados nodos) que aunque tienen sistemas operativos independientes, solo son usados para recibir y procesar las diversas tareas.

El nodo maestro se caracteriza por tener *middleware* que interactúa entre el usuario, el sistema operativo y las aplicaciones, para que tanto el usuario como el sistema, puedan ver el conjunto de computadoras conectadas, como una sola.

El *middleware* está compuesto por:

- *Job Management System (JMS)*
Se encarga de monitorear, controlar y obtener los resultados de los trabajos o procesos. Por otra parte, se puede configurar el reparto igualitario de los procesos y así tener el control de lo mencionado anteriormente.
- *Cluster Management System*
Controla todos los nodos así como su configuración, es por esto que es una parte importante del *middleware*.
- *Cluster Monitoring System*
Revisa constantemente todos los nodos para que el usuario sepa si existe algún problema, para obtener estadísticas o simplemente para saber la carga y el rendimiento del equipo, obteniendo información de la memoria libre, almacenamientos, carga de procesadores y los procesos ejecutados.
- *Parallel Execution Libraries*
Estas bibliotecas son usadas para que las aplicaciones puedan ser paralelizadas y puedan distribuirse entre los diversos procesadores y nodos. MPI es el *software* más conocido.

- *Global Process Space*

Complementa al JMS, controlado en todo el clúster todos los procesos con un identificador único para poder ser manejados.

Todos los nodos deben de contar con una tarjeta de red para que se puedan comunicar, recibir datos y así estar conectados a un *switch* o algún otro dispositivo que permita la comunicación entre ellos.

En la Figura (2.2) podemos ver cómo está constituido un clúster con los mínimos requerimientos:

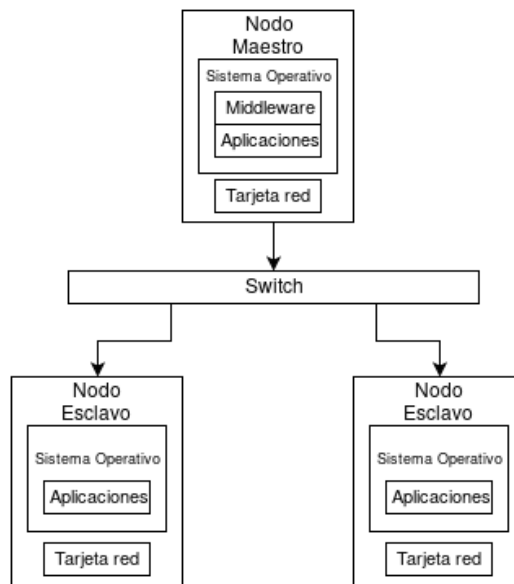


Figura 2.2: Estructura de un clúster

Tipos de Clústeres

Se pueden clasificar de acuerdo a sus características y dependiendo del uso o el tipo de servicio para el que se vaya a utilizar.

Alto Rendimiento (*High Performace Cluster*) Se utilizan para procesar trabajos que requieren bastante potencia y recursos, que por lo general, tardan gran cantidad de tiempo en concluir el proceso de cálculo llegando a tardar hasta años.

Alta Disponibilidad (*High Availability*) Este tipo de clústeres se caracterizan por ser confiables, ya que siempre podrán ser utilizados, esto se logra teniendo redundancia tanto en el *software* como en el *hardware*. En el caso del *software*, se tienen programas especializados para poder detectar las irregularidades y tratar de resolver algún problema. En el *hardware*, si llega a fallar un dispositivo, entra otro que está de respaldo como discos duros, fuentes de poder, etcétera.

Alta Eficiencia en Grandes Volúmenes de Datos o Tareas (*High Throughput*) Se utiliza para procesar gran cantidad de información en el menor periodo de tiempo posible. Sus tareas son independientes y esto ayuda a que al estar procesando datos de una tarea, las demás pueden seguir ejecutándose y así terminar en un tiempo menor.

Mallas Computaciones (*Grids*)

Son computadoras que pueden variar tanto en su *hardware* como en su *software*, ya sea en el sistema operativo o aplicaciones, y su estructura es prácticamente la misma que la de un clúster, excepto que su ubicación física puede variar, estando en diferentes lugares de una empresa, por ejemplo.

2.3 GNU/Linux

En la mayoría de los clústeres se utiliza este sistema operativo por su estabilidad y sus diversas aplicaciones.

También conocido como Linux, es un sistema operativo multiusuario (pueden estar conectados varios usuarios a la vez), multitarea, (ejecución de varios comandos al mismo tiempo) portable (es soportado por varias arquitecturas), entre otros. Está inspirado en MINIX, un sistema creado para la enseñanza que a su vez está basado en Unix del cual, toma la mayoría de sus características. El nombre de GNU/Linux proviene del *kernel* programado por Linus Torvalds y las aplicaciones del proyecto GNU de la *Free Software Foundation*.

2.4 Imagen de disco y Clonación

Al hacer una copia o un respaldo de información, no hacemos más que copiar los archivos de nuestro interés a otra parte del disco o a un dispositivo.

La imagen de disco es el proceso mediante el cual se hace una copia exacta de toda la estructura lógica y archivos del disco duro, reproduciendo byte por byte a uno o varios archivos (depende de la aplicación usada) alojado en la misma computadora o en un dispositivo extraíble.

La clonación de un disco duro toma este concepto, pero en vez de solo crear el archivo, la imagen se transfiere a una o varias computadoras conectadas a través de alguna conexión de red dejando el equipo listo para usarse.

Para entender mejor qué es lo que se copia al hacer una clonación o una imagen, a continuación se menciona la estructura lógica de un disco duro.

2.5 Estructura lógica de un disco duro

Para poder acceder a los datos, se necesita de una organización en el disco que se componen de los siguientes elementos:

- Sector de arranque. Comprende toda la información, comandos y opciones para el inicio adecuado de un sistema operativo. Se compone del código de arranque, la firma del disco, la tabla de particiones y la firma de arranque.
- Sistema de archivos. Administra la información que está almacenada, contando con procedimientos y funciones para leer, escribir, recuperar y organizar dicha información.

2.6 Herramientas para imágenes y clonación

Existen diversos programas para poder llevar a cabo esta función, los más conocidos de *software* privativo¹ son Acronis, AOEMI Backuper y Paragon.

El *software* libre es el que una vez que se adquiere se puede ejecutar, estudiar, redistribuir y distribuir el código modificado sin necesidad de un permiso del creador, por ejemplo, GNU/Linux.

A continuación se mencionan las aplicaciones más importantes y conocidas para la clonación e imágenes de disco para el sistema operativo GNU/Linux.

2.6.1 dd

Originalmente escrito para Unix, es un comando que viene instalado por defecto en todas las distribuciones² de GNU/Linux. Su función es la de leer y escribir datos desde un dispositivo de almacenamiento, haciendo una copia de todo el dispositivo, incluyendo tabla de particiones y su estructura, pudiendo hacer una imagen de disco, cabe mencionar que no comprime el archivo. Por ejemplo, para hacer una imagen de un dispositivo (`/dev/sdb`), el comando adecuado es el siguiente:

```
dd if=/dev/sdb of=/home/usuario/imagen.img
```

¹ *Software* al que no se tiene acceso al código fuente y por lo tanto, no se puede modificar ni ser compartido, por lo general se debe de pagar por una licencia para tener acceso a todas sus funcionalidades.

² Sistema operativo con *kernel* de Linux que cuenta con un conjunto de aplicaciones en específico que es lo que lo diferencia de las demás.

En donde:

- dd - Comando
- if= - opción para decirle al comando dd de dónde va a leer la información seguido de la dirección del dispositivo
- of= - opción en la que se le informa al comando el destino en dónde se escribirán los datos

Si se requiere hacer una clonación en otra computadora que esté conectada vía LAN, es necesario utilizar otro comando en conjunto y tener los permisos suficientes para escribir en el dispositivo de destino, el más conocido es ssh que sirve para poder conectarse mediante la red. La instrucción sería la siguiente:

```
dd if=/dev/sdb | ssh usuario@hostname dd of=/dev/sda
```

2.6.2 Clonezilla

Es un *software* libre desarrollado por la *National Center for High-performance Computing* y diseñado por Steven Shiau. Sirve para poder crear imágenes de discos o de alguna partición. Otras de sus funciones es la de recuperar archivos y hacer respaldos.

Clonezilla utiliza distintas aplicaciones que vienen en en las distribuciones de GNU/Linux, las más importantes son:

- dd - incluido en el paquete coreutils, indispensable para poder instalar cualquier distribución
- bzip2 - Compresor de archivos con un algoritmo en específico
- dosfstools - Herramienta para crear, poder leer y escribir en sistemas de archivos tipo FAT
- gzip - Utilidad para comprimir archivos de GNU
- parted - Programa para poder manipular particiones de disco
- ntfs-3g - Utilidades para los sistemas NTFS
- partimage - Guarda imágenes de disco en varios formatos

Características

Entre características de Clonezilla se pueden destacar las siguientes:

- Soporta particiones como MBR y GPT
- Sistemas de archivos soportados: ext2, ext3, ext4, reiserfs, reiser4, xfs, jfs, btrfs, f2fs, nilfs2, FAT12, FAT16, FAT32, NTFS HFS+, UFS, minix VMFS3 y VMFS5
- Puede copiar volúmenes lógicos
- Opción de imagen cifrada
- Protocolos para guardar copia remotamente: SSH, Samba, NFS, WebDAV
- Tipos de compresión: pgzip, gzip, pbzip2, bzip2, lzop, lzma

Ventajas

- Se puede usar con requerimientos mínimos en una computadora de 196 MB en RAM y procesador de 32 o 64 bits
- Cuenta con una versión que funciona como servidor para poder clonar o almacenar imágenes
- No se necesita tener un sistema operativo instalado ya que también funciona como *Live CD*³

Desventajas

- Por el formato de la imagen, no se puede montar para poder revisar archivos o explorar las diversas carpetas
- La imagen solo puede ser con un dispositivo que no esté montado y no se esté usando
- El disco de destino no puede ser de menor capacidad que el tamaño de la imagen

También hay una versión llamada Clonezilla Server Edition que funciona como servidor, almacenando imágenes o restaurando uno o varios equipos automáticamente.

Existen varias alternativas para que el cliente pueda arrancar y así comunicarse con el servidor. Dependiendo del tipo de *hardware* con el que se cuente, se puede generar un

³Es un CD o DVD que carga al arranque un sistema operativo sin necesidad de una instalación o escribir en el disco duro.

medio arrancable como un disquete, un disco compacto o en un disco duro con un sistema GNU/Linux instalado anteriormente.

Otro método es usando DRBL (*Diskless Remote Boot in Linux*) para gestionar el arranque de un sistema operativo en un cliente por medio de la red. Emplea los siguiente:

- PXE (*Preboot Execution Environment*). Es un código contenido en el BIOS que al iniciarse, busca en la red un servidor que le proporcione un sistema operativo para poder arrancar
- NFS (*Network File System*). Es un sistema de archivos que deja compartir archivos o carpetas dentro de una red
- NIS (*Network Information Services*). Servicio en cual, un servidor comparte información para evitar ser duplicada en los clientes, como cuentas de usuarios, particiones de disco, grupos, etcétera

2.6.3 Mondo Rescue

Programa de *software* libre para la recuperación de datos en caso de desastres, es por esto que se puede hacer una copia exacta del sistema operativo y de todos los datos almacenados. Utiliza Mindi, un programa para poder hacer la copia de seguridad en un CD o DVD y partir el archivo creado por si hay problemas de espacio. Es necesario tener un sistema operativo tipo GNU/Linux para instalado.

Al igual que Clonezilla, depende de comandos de GNU/Linux como `afio`, `gzip`, `mkisoft` y utilidades especiales de `perl`.

Características

- Sistemas de archivo soportados: LVM 1 y 2, RAID, ext2, ext3, ext4, JFS, XFS, ReiserFS y FAT
- Soporta UEFI
- Imagen creada con el formato ISO
- Compresión en bzip, gzip y lzo

Requerimientos

128 MB en RAM, al menos 800 MB de espacio libre en disco duro, un dispositivo en donde se pueda pasar la imagen o directorios compartidos por la red.

2.6.4 Redo Backup and Recovery

Es un *Live CD* basado en Ubuntu que sirve para hacer respaldos y restablecerlos, creando una imagen del disco duro completo y cuenta con herramientas para recuperación de archivos borrados.

Ventajas

- Interfaz gráfica amigable
- Compatible con sistemas Windows
- No es necesario instalar sistemas o programas
- Acepta hacer respaldos por red
- Sistemas de archivos soportados: btrfs, ext2, ext3, ext4, reiserfs, reiserfs4, XFS, JFS, NTFS, FAT12, FAT16, FAT32, exFAT, HFS plus, UFS2, VMFS, F2FS y NILFS2

Desventajas

- No cuenta con opción de compresión

Requerimientos

- 512 en RAM
- Unidad de DVD o CD

2.6.5 Comparativa de las herramientas

A continuación se muestra una tabla para poder comparar las diferentes herramientas.

	Clonezilla	Mondo Rescue	Redo Backup and Recovery
Sistemas de archivos soportados	ext2, ext3, ext4, reiserfs, reiser4, xfs, jfs, btrfs, f2fs, nilfs2, FAT12, FAT16, FAT32, NTFS, HFS+, UFS, minix, VMFS3 y VMFS5	LVM 1 y 2, RAID, ext2, ext3, ext4, JFS, XFS, ReiserFS y FAT	Btrfs, ext2, ext3, ext4, reiserfs, reiserfs4, XFS, JFS, NTFS, FAT12, FAT16, FAT32, exFAT, HFS plus, UFS2, VMFS, F2FS y NILFS2
Cliente/Servidor	Sí	Sí	No
Tipos de compresión	gzip, bzip2, lzop y lzma	Bzip, gzip y lzo	Ninguna
Compresión en paralelo	Sí	No	No
<i>Live CD</i>	Sí	Sí	Sí
¿Se puede instalar en un SO?	Sí	Sí	No
Imagen en caliente ⁴	No	Sí	No

2.6.6 Necesidad la Coordinación de Supercómputo en DGTIC

Actualmente se cuentan con varios clústeres funcionales que ya están en producción. La instalación y configuración de clústeres lleva mucho tiempo debido a la complejidad y la preparación del *software* necesario. Otra consideración importante, es que un solo clúster puede llegar a tener cientos de nodos.

El problema surge cuando se requiere añadir uno o varios nodos a un clúster, entre más nodos se requieran, mayor será el trabajo a realizar y el tiempo en que estén listos para su uso y debido a la demanda que existe para poder utilizarlo, es necesario tenerlos listos lo más rápido posible.

⁴Hacer una copia cuando el sistema o la partición esté montada o siendo usada.

Como se mencionó anteriormente, ya se tiene una herramienta que puede resolver el problema, pero debido a las limitantes que se tiene, no se puede disponer o automatizar debidamente y puede llevar más tiempo configurando opciones para poder realizar el proceso.

Es por esto que se necesita la implementación de algún procedimiento para solucionar las siguientes problemáticas:

- Tiempo excesivo
- Repetición de instalaciones y configuraciones
- Configuraciones complicadas
- Errores en las configuraciones
- Complicaciones para detectar equipo de marcas diferentes
- Renovación de licencia

Guías de instalación y propuesta para la solución del problema

Dada la situación actual de la Coordinación de Supercómputo y su problemática, es necesario implementar un método que pueda cubrir dichas necesidades. Existen diversos tipos de *software* especializados mencionados en el capítulo anterior que pueden ayudar a resolverlas. A continuación se mencionan sus configuraciones, instalaciones, cómo se usan, etcétera.

Algunas de las siguientes imágenes fueron tomadas desde máquinas virtuales debido a que no se pueden tomar capturas de pantalla sin entorno gráfico pero siguiendo el mismo procedimiento.

3.1 Guía de Mondo Rescue

La versión utilizada es la 3.2.2-r3578.

3.1.1 Instalación y configuración

Agregar a los repositorios el archivo para poder obtener los paquetes de instalación:

```
wget -P /etc/yum.repos.d ftp://ftp.mondorescue.org/rhel/6/x86_64/afio
.repo
```

A continuación se debe de actualizar la base de datos del gestor de paquetes para que pueda leer y acceder al repositorio agregado anteriormente:

```
yum update
```

Para instalar Mondo Rescue, hay que ejecutar el siguiente comando:

```
yum install -y mondo
```

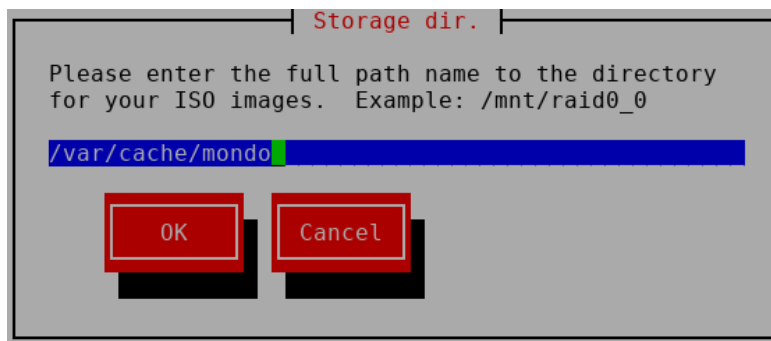
Una vez terminada esta operación, el comando a ejecutar es el siguiente:

```
mondoarchive
```

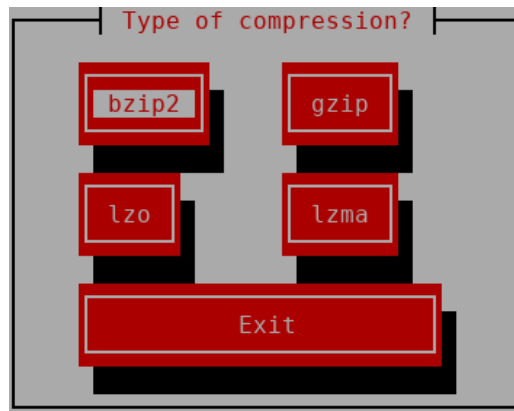
En la pantalla principal se puede elegir el dispositivo en donde queremos guardar la imagen a crear



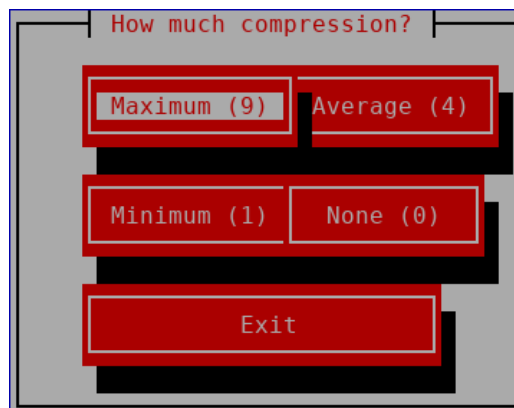
Ingresar la ruta para guardar la imagen ISO que será creada



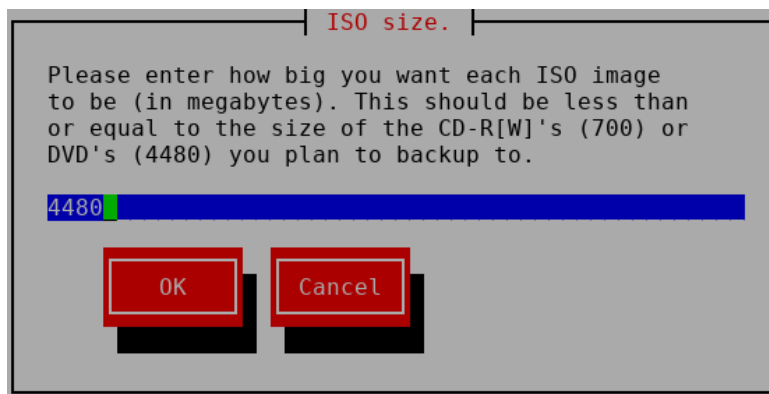
Seleccionar el tipo de compresión



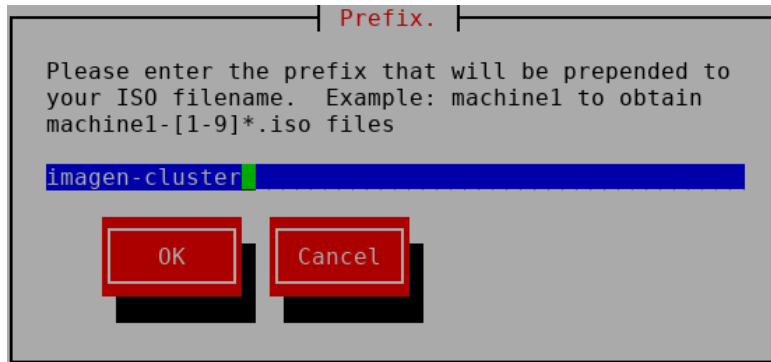
Elegir el nivel de compresión deseado



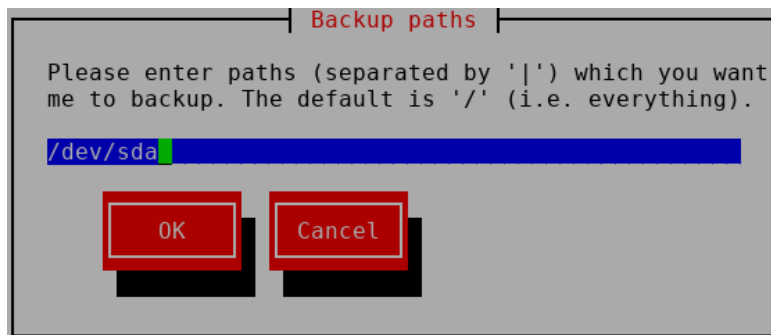
Si la imagen es muy grande se tiene la opción de dividirla, ingresando el tamaño en Megabytes de cada parte. Para que solo sea un archivo, teclear un número muy grande, por ejemplo 1000000



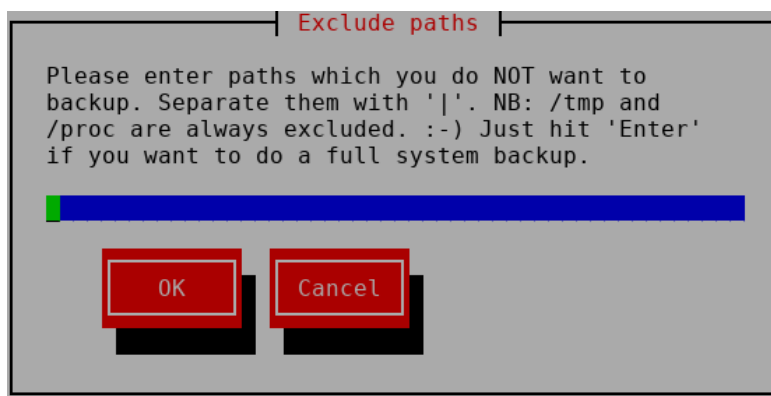
La siguiente opción es solo para el nombre de la imagen, por ejemplo imagen-cluster



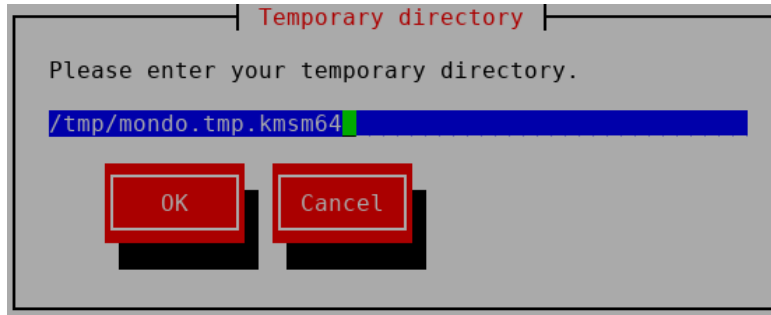
Se debe de poner la ruta a respaldar o el dispositivo, por ejemplo /dev/sda



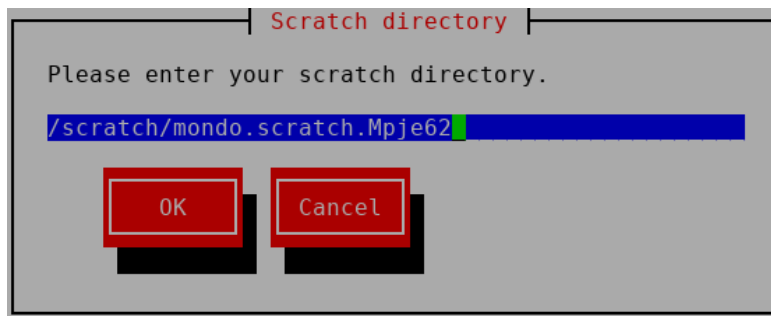
Se pedirá que se ingresen las rutas a excluir en la imagen, por defecto no se toman ni /tmp que solo guarda archivos temporales y al reiniciar se borran, ni /proc que contiene información de procesos e información de *hardware* que toma el sistema durante el arranque



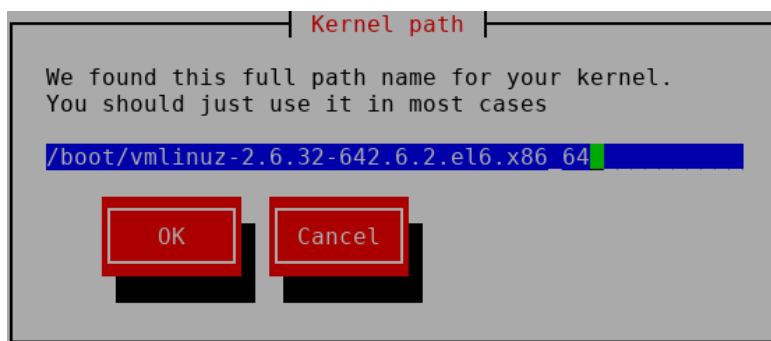
Poner la ruta de los temporales



La siguiente ruta que pide es la de *scratch*, que se utiliza cuando se generan grandes volúmenes de información y al finalizar, solo se hace una copia de lo que se pretende utilizar. Cuando el proceso ya está terminado, lo que está dentro de este directorio ya no es servible



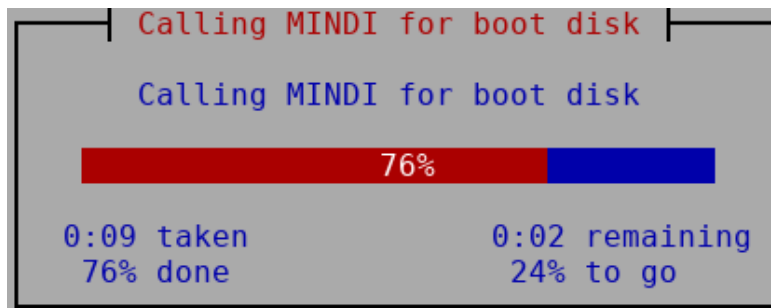
Se detectará automáticamente la ruta del núcleo del sistema y solo hay que presionar «OK»



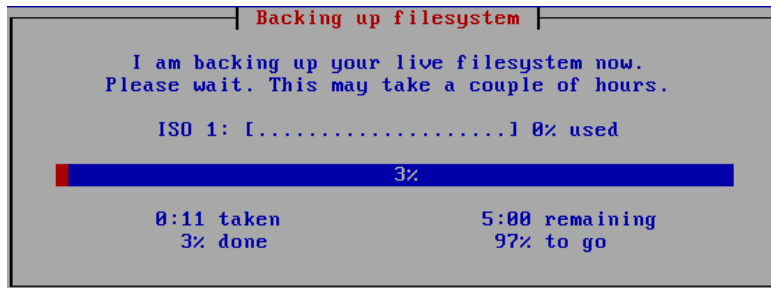
Otra opción con la que se cuenta, es la de verificar la imagen para asegurarnos de que la imagen se creó correctamente



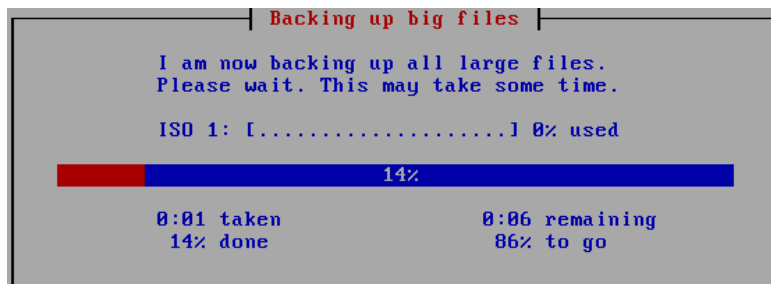
Se creará una lista de todos los archivos y el arranque del disco



El siguiente proceso generará una copia idéntica del sistema de archivos del dispositivo seleccionado

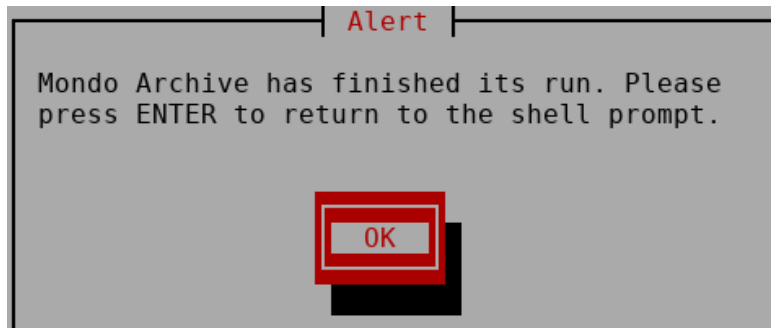


El proceso más tardado es el de copiar todos los archivos contenidos en el disco



Ya terminado, se correrá el comando `mkisofs` que sirve para generar archivos ISO y se verificará el archivo creado.

Cuando finalice, saldrá la siguiente ventana indicando que el proceso ya está terminado



3.1.2 Restablecimiento de la imagen

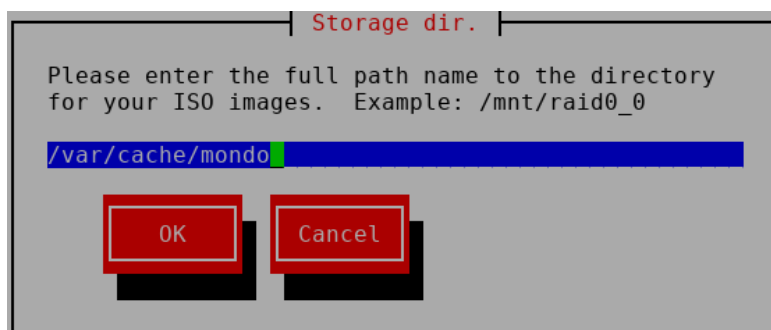
El siguiente comando, es para iniciar el programa para poder restaurar la imagen

```
mondorestore
```

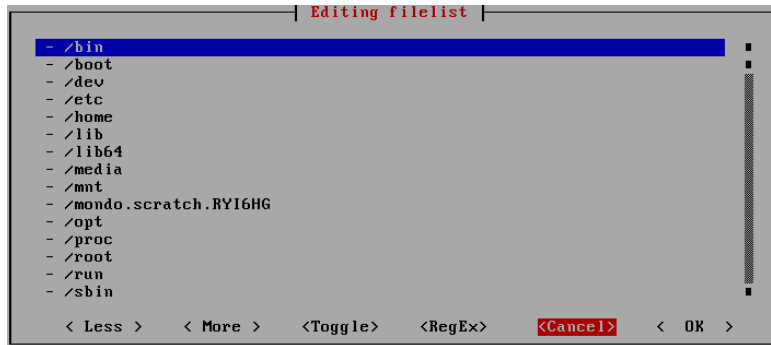
Nos preguntará en dónde tenemos alojada nuestra imagen



Ingresar la ruta de la imagen



Elegir las carpetas a recuperar



Colocar la ruta para restaurar



Ese es todo el procedimiento, solo hay que esperar a que termine el proceso.

3.2 Guía para Redo Backup and Recovery

La versión utilizada es la 1.0.4.

3.2.1 Configuración

Descargar la imagen de la siguiente dirección electrónica

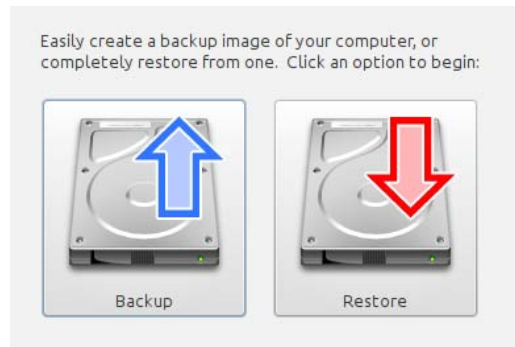
```
https://sourceforge.net/projects/redobackup/files/latest/download
```

Cuando esté listo, se puede grabar un disco compacto, DVD o directo a una USB, indicando la ruta del dispositivo y el nombre del archivo de la siguiente forma:

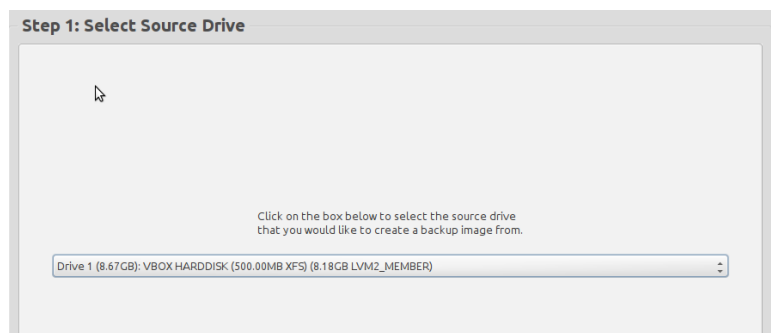
```
dd if=redobackup-livecd-1.0.4.iso of=/dev/sdb
```

Cambiar en el BIOS el dispositivo de arranque y esperar a que inicie el sistema operativo

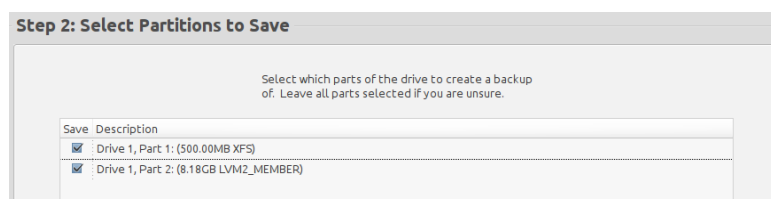
Ya que haya iniciado el sistema, automáticamente se abrirá una ventana con dos opciones a elegir, «*Backup*» y «*Restore*», elegir la primera



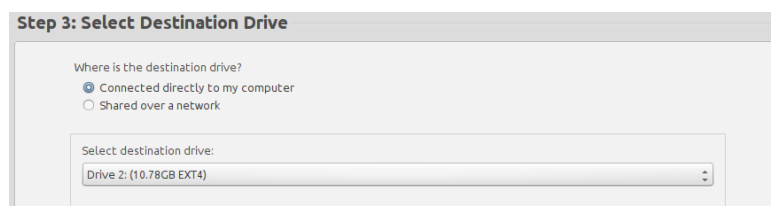
Seleccionar el dispositivo del que se quiera sacar la imagen



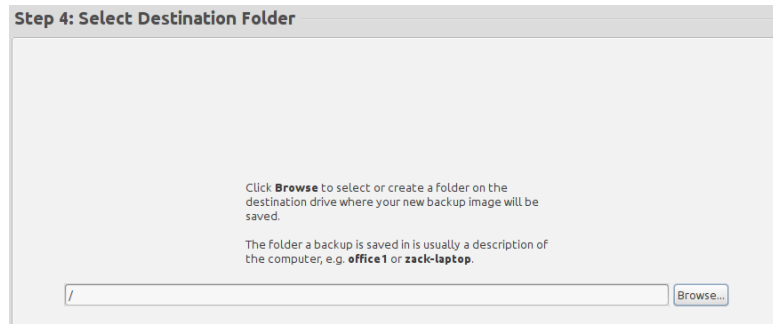
Marcar las particiones para crear la imagen. Para crearla del disco completo, seleccionar todas las opciones



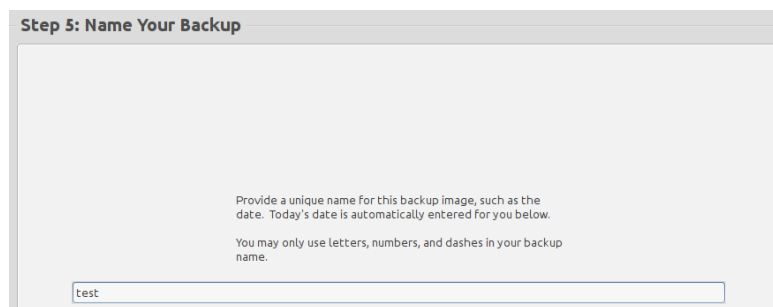
Escoger en dónde guardar la imagen que se va a crear, por red o en algún dispositivo conectado



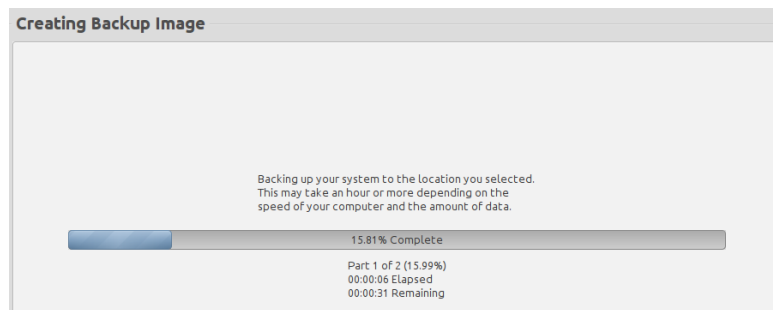
Ingresar la carpeta en donde se guardará la imagen



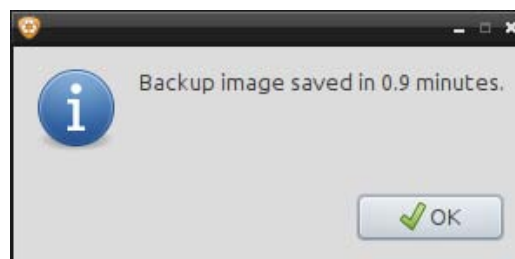
Darle un nombre a la imagen



Se iniciará el proceso de creación de la imagen

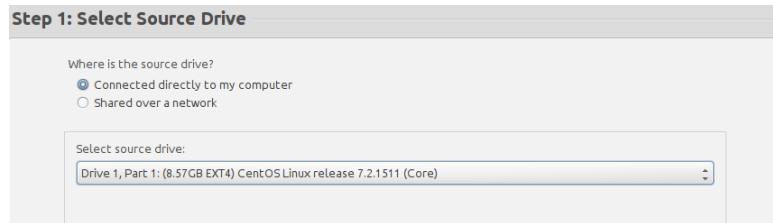


Al finalizar saldrá una ventana diciendo que la imagen fue guardada y el tiempo que tomó

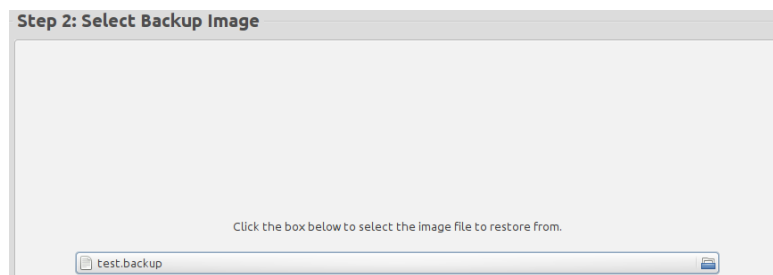


3.2.2 Restauración de la imagen

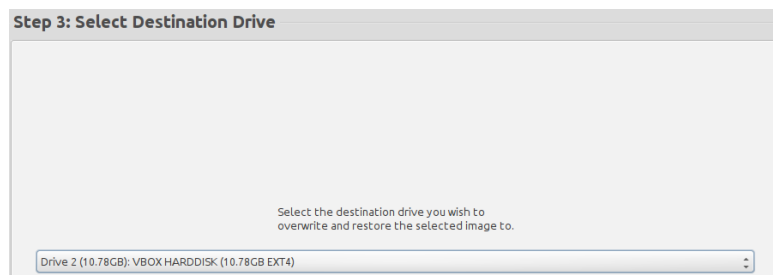
Para este proceso, en la pantalla principal se debe de seleccionar «Restore»
Seleccionar el disco en donde se encuentra la imagen



Seleccionar el archivo creado anteriormente



Agregar el disco de destino



Al finalizar se mostrará el tiempo que tardó en hacer el proceso.

3.3 Guía para Clonezilla Server Edition

3.3.1 Instalación y configuración

Instalar las dependencias para DRBL

```
yum -y install perl perl-Digest-SHA1 perl-Data-Dumper-Names
```

Con el siguiente comando se descarga e instala DRBL

```
rpm -ivh https://superb-dca2.dl.sourceforge.net/project/drbl/  
drbl_stable/2.22.5/drbl-2.22.5-drb11.noarch.rpm
```

Para empezar a usar este *software*, primero hay que seguir unos pasos e instalar paquetes necesarios en el servidor de arranque remoto, ejecutando el siguiente comando

```
drblsrv -i
```

Se preguntará si queremos usar las versiones de prueba o la inestable (que no son muy recomendables debido a que pueden presentar fallas) y es recomendable utilizar siempre las versiones estables

```
*****.  
Use the CO6.8 repository settings in /etc/yum.repos.d/.  
Seting the DRBL yum repository in /etc/yum.repos.d/... done!  
*****.  
Do you want to use the DRBL "unstable" packages which are much more  
powerful but may be very unstable? (For the super brave!!!)(If you  
downloaded the DRBL rpm packages from the unstable directory, say  
"yes" here)  
[y/N]  
Do you want to use the DRBL "testing" packages which are more  
powerful but may be buggy? (For the brave!!)(If you downloaded  
the DRBL rpm packages from the testing directory, say "yes" here)  
[y/N]
```

Tiene la opción de poder instalar distribuciones de GNU/Linux nuevas a través de la red

```
Do you want to install the network installation boot images so that  
you can let the client computer install some GNU/Linux distributions  
(Debian, Ubuntu, RedHat Linux, Fedora Core, Mandriva, CentOS and  
OpenSuSE...) via a network connection?  
!!NOTE!!
```

```
This will download a lot of files (Typically > 100 MB) it might
take a few minutes.
If the client computer has a hard drive that you may install
GNU/Linux onto, put a Y here. If you answer "no" here, you can run
"drbl-netinstall" to install them later.
[y/N]
```

Esta parte es para habilitar la consola, así se puede ver el proceso de la imagen y con opción de actualizar el sistema

```
Do you want to use the serial console output on the client
computer(s)?
If you do NOT know what to pick, say "N" here, otherwise the
client computer(s) may show NOTHING on the screen!
[y/N]
The CPU arch option for your clients: 2
The optimization for your system is set to be the same as this
server computer.
*****.
*****.
Do you want to upgrade the operating system?
[y/N]
```

A continuación empezará con el proceso de instalación de los paquetes requeridos

```
Use kernel modules from /tftpboot/node_root//lib/modules/
2.6.32-642.11.1.el6.x86_64.
Trying to include network card firmwares if they exist in
/tftpboot/node_root//lib/firmware/...
Calling hook udev...
Creating the initRAMFS image...
Initramfs, remove ramdisk_size/ramdisk_block in
/tftpboot/nbi_img/pxelinux.cfg/default if exists...
Finished!
Done!
*****.
Done!
```

Configuración de DRBL

Correr el siguiente comando que sirve para ingresar información al servidor

```
drblpush -i
```

Pedirá el dominio a utilizar

```
Hint! When a yes/no option is available, the default value
is uppercase, Ex. (y/N), the default is "N", when you press
"Enter", it will use "N". If you are not sure which one to
choose, you can just press "Enter" key.
```

```
*****
```

```
Searching the installed packages for DRBL server...This
might take several minutes...
```

```
Finished searching the installed packages for DRBL server.
```

```
*****
```

```
-----
```

```
The interactive mode let you supply the information of your
DRBL environment.
```

```
-----
```

```
dnsdomainname: Unknown host
```

```
-----
```

```
Please enter DNS domain (such as drbl.sf.net):
```

```
[drbl.org]
```

Ingresar el nombre de dominio de NIS para configurarlo y generar automáticamente los nombres de las máquinas a clonar, solo habría que agregar un prefijo

```
Please enter NIS/YP domain name:
```

```
[iim]
```

```
Set DOMAIN as iim
```

```
-----
```

```
Please enter the client hostname prefix:
```

```
This prefix is used to automatically create hostname for clients.
```

```
If you want to overwrite some or all automatically created
hostnames, press Ctrl-C to quit this program now, edit
/etc/drbl/client-ip-hostname, then run this program again.
```

```
[ritchie-]
```

Escribir la interfaz de red que tenga salida a internet

```
The public IP address of this server is NOT found.
Which ethernet port in this server is for public Internet access,
not for DRBL connection?
Available ethernet ports in this server:
eth0 (192.168.14.200), eth1 (10.0.0.1),
[eth0]
```

Se pueden guardar las direcciones MAC para poder reconocer los equipos posteriormente

```
Now we can collect the MAC address of clients!
If you want to let the DHCP service in DRBL server offer same
IP address to client every time when client boot, and you
never did this procedure, you should do it now!
If you already have those MAC addresses of clients, you can
put them into different group files (These files number is the
same number of networks cards for DRBL service). In this case,
you can skip this step.
This step helps you to record the MAC addresses of clients,
then divide them into different groups. It will save your
time and reduce the typos.
The MAC addresses will be recorded turn by turn according to
the boot of clients, and they will be put into different files
according to the network card in server, file name will be like
macadr-eth1.txt, macadr-eth2.txt... You can find them in
directory /etc/drbl.
Please boot the clients by order, make sure they boot from
etherboot or PXE!
Do you want to collect them?
[y/N]
```

El servidor de DHCP instalado por DRBL, puede asignar la misma IP a los clientes por medio de sus direcciones MAC. Ingresar la interfaz de red que se comunicará con los demás nodos

```
OK! Let's continue...
*****
Hostmin: 10.0.0.1
Do you want to let the DHCP service in DRBL server offer same IP
```



```
address to the client every time when client boots (If you want this
function, you have to collect the MAC addresses of clients, and save
them in file(s) (as in the previous procedure)). This is for the
clients connected to DRBL server's ethernet network interface eth1 ?
[y/N]
```

Ahora hay que ingresar el primer número que se utilizará en el último octeto de las direcciones de IP a asignar

```
OK! Let's continue, we will set the IP address of clients by "first
boot gets IP first" instead of fixed one!
*****
What is the initial number do you want to use in the last set of
digits in the IP (i.e. the initial
value of d in the IP address a.b.c.d) for DRBL clients connected to
this ethernet port eth1.
[1]
```

Lo que sigue es poner el número de clientes a los que se restaurará la imagen, preferentemente un número mucho mayor, ya que el servidor de DHCP no asigna IPs secuencialmente y esto podría provocar que no se inicie el sistema. También preguntará si es correcta la asignación de IPs

```
How many DRBL clients (PC for students) connected to DRBL server's
ethernet network interface eth1 ?
Please enter the number:
[100]
*****
We will set the IP address for the clients connected to DRBL
server's ethernet network interface eth1
as: 10.0.0.1 - 10.0.0.100
Accept ? [Y/n]
```

Saldrá un menú para elegir cómo el cliente debe de arrancar el sistema

```
In the system, there are 3 modes for diskless linux services:
[0] Full DRBL mode, every client has its own NFS based /etc
and /var.
[1] DRBL SSI (Single system image) mode, every client uses
tmpfs based /etc and /var. In this mode, the loading and
necessary disk space of server will be lighter. NOTE! (a)
```

```
The client machine memory is recommended at least 256 MB.
(b) The setting and config files of client will not be
saved to the DRBL server! They are just used once and will
vanish after the machine shutdowns! Besides, if you modify
any file in the template client (located in /tftpboot/nodes),
you have to run drbl-gen-ssi-files to create the template
tarball in /tftpboot/node_root/drbl_ssi/.
(c) If you want to provide some file to overwrite the
setting in the template tarball when client boots, check
/tftpboot/node_root/drbl_ssi/clients/00_README for more
details.
[2] I do NOT want to provide diskless Linux service to
client.
Which mode do you prefer?
[0]
```

Menú muy similar pero en este caso es para Clonezilla

```
In the system, there are 4 modes available for clonezilla:
[0] Full Clonezilla mode, every client has its own NFS based /etc
and /var.
[1] Clonezilla box mode, every client uses tmpfs based /etc and
/var.
In this mode, the loading and necessary disk space of server will
be lighter than that in Full Clonezilla mode. Note! In Clonezilla
box mode, the setting and config files of client will not be saved
to the DRBL server! They just use once and will vanish after the
machine shutdowns!
[2] I do NOT want clonezilla.
[3] Use Clonezilla live as the OS (Operating System) of clients
(Testing).
Which mode do you prefer?
[0]
```

Preguntará la ruta para guardar las imágenes creadas, si es el caso

```
When using clonezilla, which directory in this server you want
to store the saved image (Please use absolute path, and do NOT
assign it under /mnt/, /media/ or /tmp/)?
[/scratch]
```

Para poder crear una partición swap en el cliente por si es necesario utilizar más memoria de la que se tiene

```
If there is a local harddrive with swap partition or writable file
system in your client machine, do you want to use that swap
partition or create a swap file in the writable filesystem so that
client has more memory to use? (This step will NOT destroy any data
in that harddisk)
[Y/n]
```

Máximo espacio en Megabytes para swap

```
What's the maximun size (Megabytes) for the swap space?
We will try to allocate the swap space for you, if it's not enough,
60% of the free space will be used.
[128]
```

Escoger si iniciar los clientes con entorno gráfico

```
Which mode do you want the clients to use after they boot?
"1": Graphic mode (X window system) (default),
"2": Text mode.
[1]
```

Es opcional ingresar una nueva contraseña a los clientes, en vez de utilizar la que se copió al crear la imagen y antes de que arranque el sistema

```
Do you want to set the root's password for clients instead of
using same root's password copied from server?
(For better security)
[y/N]
OK! Let's continue...
-----
Do you want to set the pxelinux password for clients so that when
client boots, a password must be entered to startup
(For better security)
[y/N]
```

En caso de que sea requerido, hay opción de acceder a la consola en los clientes al arranque y el tiempo de espera en segundos antes de que automáticamente arranque el cliente

```
Do you want to set the pxelinux password for clients so that when
client boots, a password must be entered to startup
(For better security)
[y/N]
OK! Let's continue...
-----
Do you want to set the boot prompt for clients?
[Y/n]
How many 1/10 sec is the boot prompt timeout for clients?
[70]
```

Menú gráfico en el arranque del cliente y a continuación opción de configurar IPs públicas para los clientes

```
Do you want to use graphic background for PXE menu when client
boots?
Note! If you use graphical PXELinux menu, however client fails
to boot, you can switch to text mode by running "switch-pxe-bg-mode
-m text".
[y/N]
Use text PXE Linux menu for the client.
-----
-----
By using alias interface, every client can have 2 IPs,
one of them is private IP for clients connected to DRBL server, and
the other is public IP for clients directly connected to WAN from
switch!
Do you want to setup public IP for clients?
[y/N]
```

Se puede utilizar el servidor como NAT para los demás clientes

```
OK! Let's continue...
-----
-----
Do you want to let DRBL server as a NAT server? If not, your DRBL
client will NOT be able to access Internet.
```

```
[Y/n]
```

Esta pregunta es para saber si se quiere guardar una copia de la configuración del *firewall* (iptables) ya que será modificado

```
We are now ready to deploy the files to system!
Do you want to continue?
Warning! If you go on, your firewall rules will be overwritten during
the setup!
The original rules will be backuped as iptables.drblsave in system
config directory (/etc/sysconfig or /etc/default).
[Y/n]
```

A continuación configurará lo necesario

```
Enjoy DRBL!!!
http://drbl.org; http://drbl.nchc.org.tw
NCHC Free Software Labs, Taiwan. http://free.nchc.org.tw
*****.
If you like, you can reboot the DRBL server now to make sure
everything is ready...(This is not necessary, just an option)
*****.
The DRBL server is ready! Now set the client machines to boot from
PXE. (refer to http://drbl.org for more details)
P.S. The config file is saved as /etc/drbl/drblpush.conf. Therefore
if
you want to run drblpush with the same config again, you may run it
as:
drblpush -c /etc/drbl/drblpush.conf
```

Configuración en el servidor para guardar una imagen

Ahora hay que correr el servicio con el siguiente comando

```
dcS
```

Elegir todos los clientes o parte de ellos para poder trabajar

```
DRBL, developed by NCHC Free Software Labs
//Hint! From now on, if multiple choices are available, you have to press
space key to mark your selection. An asterisk (*) will be shown when the
selection is done//
Do you want to set the mode for all the clients or just some of them?
Select mode:

All Select all the clients
Part Select client(s) by IP or MAC address

<Ok> <Cancel>
```

En el menú que se despliega, elegir la sexta opción para iniciar Clonezilla

```
DRBL, developed by NCHC Free Software Labs
Switch the mode:

remote-linux-gra Client_remote_Linux_graphic_mode_powerful_client
remote-linux-txt Client_remote_Linux_text_mode_powerful_client
terminal Client_remote_Display_Linux_terminal_mode
remote-memtest Client_remote_boot_to_run_Memtest86+
remote-fdos Client_remote_boot_to_run_FreeDOS
clonezilla-start Start clonezilla mode
clonezilla-stop Stop_clonezilla_mode
netinstall Client_install_Linux_via_network
local Client_boots_its_local_OS
reboot Reboot_client_now
shutdown Shutdown_client_now
Wake-on-LAN Turn_on_client_by_Wake-on-LAN_now
more More_modes_or_commands

<Ok> <Cancel>
```

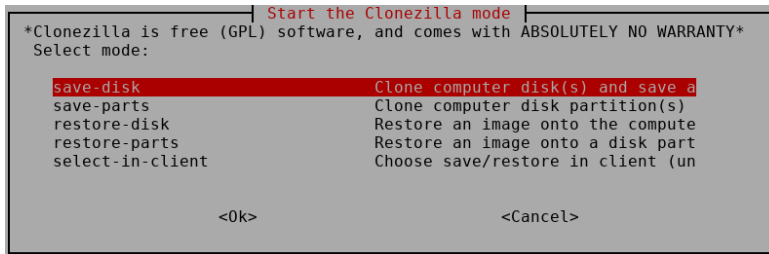
Escoger el modo de configuración para las diversas opciones. El modo experto es mejor, ya que hay libertad de elegir lo más conveniente para cada situación

```
Clonezilla - Opensource Clone System (OCS)
Choose the mode to run the following wizard about advanced parameters:

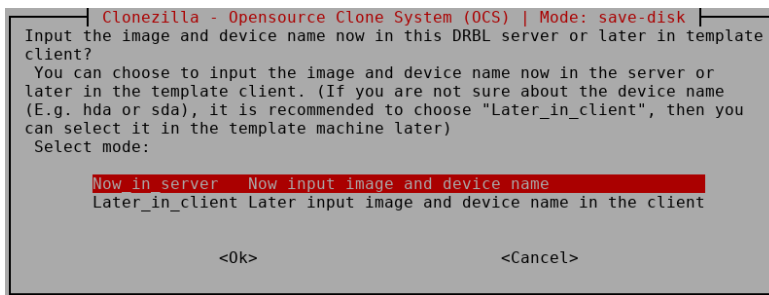
Beginner Beginner mode: Accept the default options
Expert Expert mode: Choose your own options
Exit Exit. Enter command line prompt

<Ok> <Cancel>
```

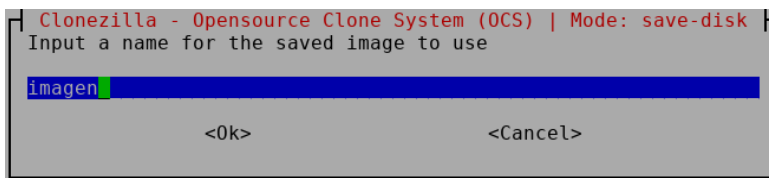
Para crear la imagen seleccionar «*save-disk*»



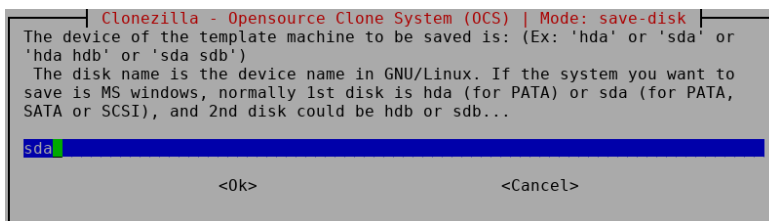
Para escoger entre ingresar el nombre de la imagen ahora o que se haga directo en el cliente



Escribir el nombre del archivo de imagen



Ingresar si el disco es SATA (sda) o PATA (hda)



En caso de que alguna herramienta de imágenes llegara a fallar, se puede elegir una prioridad en la lista

```

Clonezilla advanced extra parameters | Mode: save-disk
Which clone program(s) and what priority do you prefer? Program priority
means that if the file system is not supported by the first program, the
next program will be used. E.g. If you choose "Priority: ntfsclone >
partimage > dd", then if the file system is xfs, Clonezilla will try to use
ntfsclone first, and of course, xfs is not supported by ntfsclone, so
Clonezilla will try to use partimage with dd being the last program to try
if that does not work.
The default settings are optimized. If you have no idea, keep the default
value and do NOT change anything and continue.

-q2 Priority: partclone > partimage > dd
-q1 Priority: Only dd (supports all filesystem, but inefficient)
-q Priority: ntfsclone > partimage > dd
-q Priority: partimage > dd (no ntfsclone)

<Ok> <Cancel>
    
```

Marcar las opciones dependiendo de las necesidades de cada usuario

```

Clonezilla advanced extra parameters | Mode: save-disk
Set advanced parameters (multiple choices available). If you have no idea, keep the default values
and do NOT change anything. Just press Enter. (Press space key to mark your selection. An asterisk
(*) will be shown when the selection is done):

[*] -c Client waits for confirmation before cloning
[*] j2 Clone the hidden data between MBR and 1st partition
[ ] -nogui Use text output only, no TUI/GUI output
[ ] -a Do NOT force to turn on HD DMA
[ ] -batch Run clone in batch mode (DANGEROUS!)
[ ] -f Restart NFS service
[ ] -s Client skips the hardware detection when booting
[ ] -rm-win-swap-hib Remove page and hibernation files in Win if exists
[ ] -ntfs-ok Skip checking NTFS integrity, even bad sectors (ntfsclone only)
[ ] -rescue Continue reading next one when disk blocks read errors
[ ] -gs Generate image MD5 checksums
[ ] -gs Generate image SHA1 checksums
[ ] -gmf Generate checksum for files in device after saving
[ ] -o0 Run script in $OCS_PRERUN_DIR before clone starts
[ ] -o1 Run script in $OCS_POSTRUN_DIR as clone finishes

<Ok> <Cancel>
    
```

Esta opción es de gran utilidad y es muy recomendada debido a que si hay algún error en el sistema de archivos o en el disco, la imagen no será creada

```

Clonezilla advanced extra parameters | Mode: save-disk
Choose if you want to check and repair the file system before saving it.
This option is only for certain file systems which are well supported by
fsck on GNU/Linux, like ext2/3/4, reiserfs, xfs, jfs, vfat. Not for NTFS,

-fsck Skip checking/repairing source fi
-fsck-y Interactively check and repair so
-fsck-y Auto (Caution!) check and repair

<Ok> <Cancel>
    
```

La imagen creada puede ser revisada para saber si se puede restaurar o existe algún error

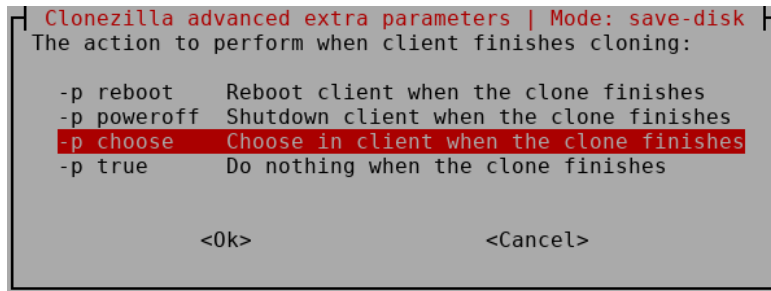
```

Clonezilla advanced extra parameters | Mode: save-disk
After the image is saved, do you want to check if the image is restorable?
///NOTE/// This action will only check the image is restorable, and it will
not write any data to the harddrive.

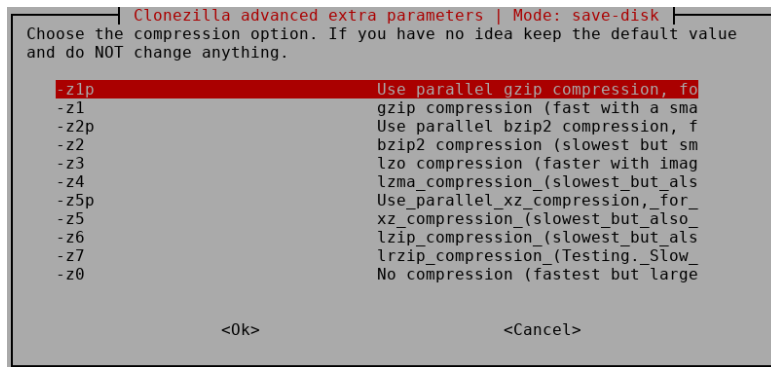
Yes, check the saved image
-scs No, skip checking the saved image

<Ok> <Cancel>
    
```

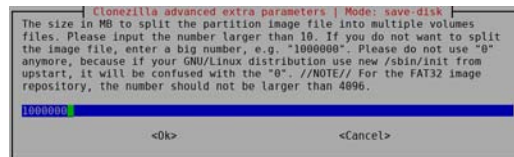

Elegir la opción más conveniente cuando termine el proceso



Seleccionar el método de compresión, lo más conveniente es que sea paralelo ya que utiliza todos los procesadores con los que cuenta el equipo, si es el caso



Si se requiere hacer en varias partes la imagen, especificar el tamaño que se necesite para cada una. Para la imagen en una sola parte, poner un número muy grande



Esta es una forma amigable de configurar el servidor para empezar a crear imágenes de disco, si por alguna razón el servicio está parado o desconfigurado, se pueden evitar estos pasos ya que al finalizar nos muestra en consola la instrucción para solo correrla

```
*****.
Now set up the client machines to boot from PXE or Etherboot
(refer to http://drbl.org for more details).
Then boot the client to make the template image! Note: If the
partition you want to save is a NTFS filesystem, it is recommended to
defrag that partition first.
NOTE! (1) If the cloned OS is MS windows, and it fails to boot with
an error message like "Missing Operating System" or "Invalid System
```

```

Disk", then you can try to (1) change the IDE hard drive settings
in the BIOS to use LBA mode instead of AUTO mode. (2) Or you can try
to use parameter -t1 when restoring. This is for all clients, so we
remove other host-based PXE config files in
/tftpboot/nbi_img/pxelinux.cfg/ and keep
/tftpboot/nbi_img/pxelinux.cfg/default.
Clean all the previous saved PXELINUX config files if they
exist...done!
Clean all the previous saved GRUB EFI NB config files if they
exist...done!
PS. Next time you can run this command directly:
drbl-ocs -b -q2 -j2 -scs -p reboot -z1p -i 1000000 -l en_US.UTF-8
startdisk save 2017-01-23-15-img sda
This command is also saved as this file name for later use if
necessary:
/tmp/ocs-2017-01-23-15-img-2017-01-23-15-35
done!

```

El comando que más importa es el siguiente

```

drbl-ocs -b -q2 -j2 -scs -p reboot -z1p -i 1000000 -l en_US.UTF-8
startdisk save 2017-01-23-15-img sda

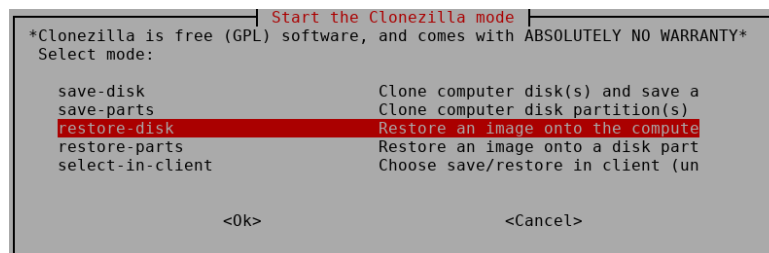
```

Configuración en el servidor para restaurar una imagen

De igual manera, correr el siguiente comando

```
dcs
```

Seleccionar «*restore-disk*»



Opciones a elegir en caso de que se requieran

```

Clonezilla advanced extra parameters | Mode: restore-disk
Set advanced parameters (multiple choices available). If you have no idea, keep the default values and do NOT change anything. Just press Enter.
(Press space key to mark your selection. An asterisk (*) will be shown when the selection is done)

* -a auto Reinstall grub in client disk MBR (only if grub config exists)
* -i auto Automatically adjust filesystem geometry for a NTFS boot partition if exists
* -z *fdisk uses ID of hard drive from ID01for non-grub boot loader)
* -k * Use Full-duplex network when multicast clone
* -noqul Use text output only, no TUI/SGUI output
* -noip PC Change MS Win hostname (based on IP address) after clone
* -noip PC Change MS Win hostname (based on MAC address) after clone
* * Prints verbose messages (especially for updates!)
* * Client waits for confirmation before cloning
* -w * Select the image to restore on client (only for unicast restoring)
* -t * Client skip restoring the mbr (Master Boot Record)
* -tl Client restores the prebuilt bootloader from syslinux (for Windows only)
* -tr Client skip restoring the EMB (Extended Boot Record)
* -tr Try to resize the filesystem to fit partition size
* -na Put ntfsclone temp file in image dir on server
* -n * fdisk uses the CHS value of hard drive from the saved image
* -icrr Ignore CRC checking of partitions
* -lrr Do not remove Linux user hardware record after restoring
* -lrr Do not remove NTFS volume dirty flag after it is restored
* -lrs Do not update syslinux-related files after restoring
* -lss Skip checking destination disk size before creating partition table
* -lefl Skip updating boot entries in EFI/VRAM after restoring
* -ll Write MBR ID01 ID again after image is restored. Not OK for partition table differs from that of the image
* -ll Write the hidden data between MBR and 1st partition
* -cb Check image by MD5 checksums
* -cs Check image by SHA1 checksums
* -cmf Inspect checksum for files in device after restoring
* -t Revert MBR service
* -t Client skips the hardware detection when booting
* -a Do NOT force to turn on HD DMA
* -m Run script in $OCS_POSTMBR_DIR when clone starts
* -ml Run script in $OCS_POSTMBR_DIR as clone finishes
* -ret Save restoring error log in image dir.

<Ok> <Cancel>
    
```

Si se quiere que el disco quede igual, dejar la opción por defecto

```

Clonezilla advanced extra parameters | Mode: restore-disk
Set advanced parameters. If you have no idea, keep the default values and do NOT change anything. Just press Enter. Choose the mode to create the partition table on the target disk: **ATTENTION**(1) TO CREATE A NEW PARTITION TABLE ON THE TARGET DISK. ALL THE DATA ON THE TARGET DEVICE WILL BE ERASED!!! (2) Clonezilla will not restore an image from a large disk (partition) to a smaller disk (partition). However, it can restore an image from a small disk (partition) to a larger disk (partition). (3) If you do NOT want Clonezilla to create a partition table, check -k:

Use the partition table from the image
-k Do NOT create a partition table on the target disk
-k1 Create partition table proportionally
-k2 Enter command line prompt to create partition manually later
-j0 Use dd to create partition (NOT OK if logical drives exist)
exit Exit

<Ok> <Cancel>
    
```

Revisa si la imagen creada se puede restaurar en en el equipo de destino

```

Clonezilla advanced extra parameters | Mode: restore-disk
Before restoring the image, do you want to check if the image is restorable on this server? ///NOTE/// This action will only check if the image is restorable or not, and it will not write any data to the harddrive.

Yes, check the image before restoring
-sc0 No, skip checking the image before restoring

<Ok> <Cancel>
    
```

Elegir qué hacer cuando termine de restaurar

```

Clonezilla advanced extra parameters | Mode: restore-disk
The action to perform when client finishes cloning:

-p reboot Reboot client when the clone finishes
-p poweroff Shutdown client when the clone finishes
-p choose Choose in client when the clone finishes
-p true Do nothing when the clone finishes

<Ok> <Cancel>
    
```

Seleccionar la imagen que se quiere restaurar

```
Clonezilla - Opensource Clone System (OCS) | Mode: resto
Choose the image file to restore (only unencrypted images
are shown):

2016-11-18-21-img 2016-1118-2116 sda 107GB

<Ok> <Cancel>
```

Escoger el disco de destino

```
Clonezilla - Opensource Clone System (OCS) | Mode: resto
Choose the target disk(s) to be overwritten (ALL DATA ON
THE ENTIRE DISK WILL BE LOST AND REPLACED!!) (Press space
key to mark your selection. An asterisk (*) will be shown
when the selection is done):

[*] sda disk(sd) disk(a)

<Ok> <Cancel>
```

Elegir el modo en que se enviará la información

```
Clonezilla - Opensource Clone System (OCS) | Mode: resto
Choose the mode to restore client disk

multicast multicast restore
broadcast broadcast restore
unicast unicast restore

<Ok> <Cancel>
```

Elegir la opción más conveniente

```
Clonezilla - Opensource Clone System (OCS) | Mode: resto
Choose the method for multicast clone:

clients+time-to-wait Set the number of client
time-to-wait Set the time for clients
clients-to-wait Set the number of client

<Ok> <Cancel>
```

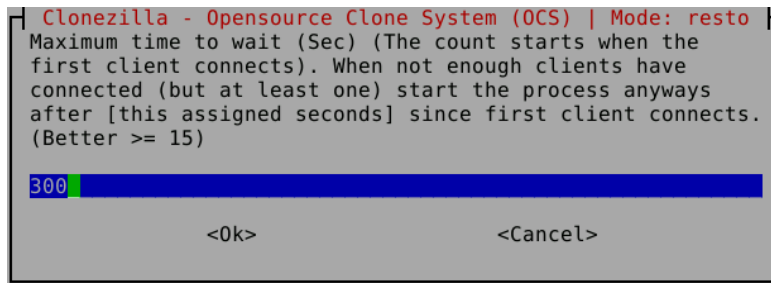
Número de clientes para restaurar la imagen

```
Clonezilla - Opensource Clone System (OCS) | Mode: resto
How many clients to restore ?

100

<Ok> <Cancel>
```

Tiempo de espera para iniciar el proceso



Se configurará y saldrá un mensaje parecido a este

```

Now set the client machines to boot from PXE or Etherboot
(refer to http://drbl.org for more details).
Then boot those clients, so that the template image can be cloned to
them!
NOTE! (1) If the cloned OS is MS windows, and it fails to boot with
an error message like "Missing Operating System" or
"Invalid System Disk", then you can try to (1) change the IDE hard
drive settings in the BIOS to use LBA mode instead of AUTO mode.
(2) Or you can try to use parameter -t1 when restoring.
This is for all clients, so we remove other host-based PXE config
files in /tftpboot/nbi_img/pxelinux.cfg/ and keep
/tftpboot/nbi_img/pxelinux.cfg/default.
Clean all the previous saved PXELINUX config files if they
exist...done!
Clean all the previous saved GRUB EFI NB config files if they
exist...done!
PS. Next time you can run this command directly:
drbl-ocs -b -g auto -e1 auto -e2 -r -x -j2 -sc0 -p reboot
--clients-to-wait 5 --max-time-to-wait 300 -l en_US.UTF-8 startdisk
multicast_restore node-img sda This command is also saved as this
file name for later use if necessary:
/tmp/ocs-node-img-2017-01-23-19-50
done!

```

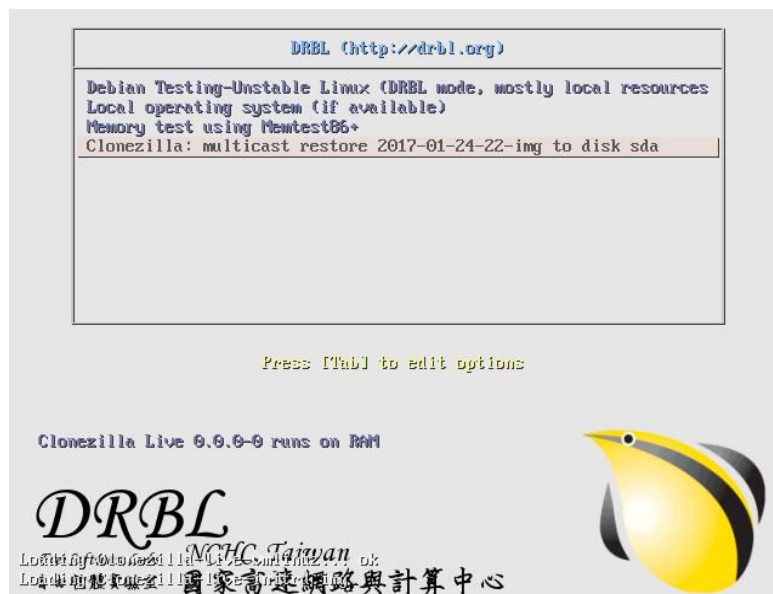

En el servidor aparecerá un mensaje diciendo que la operación se terminó con éxito

```
Client 192.168.100.1 (xx:xx:xx:xx:xx:xx) finished cloning. Stats:
success, 8.2 kB, 8.0 KiB, 0 mins; /dev/sda5, success, 72.7 MB = 35522
Blocks, .119 mins; /dev/sda6, success, 85.1 MB = 166180 Blocks,
.101 mins; /dev/sda7, success, 618.6 MB = 302068 Blocks, .360 mins;
/dev/sda8, success, 5.4 MB = 2658 Blocks, .118 mins;
```

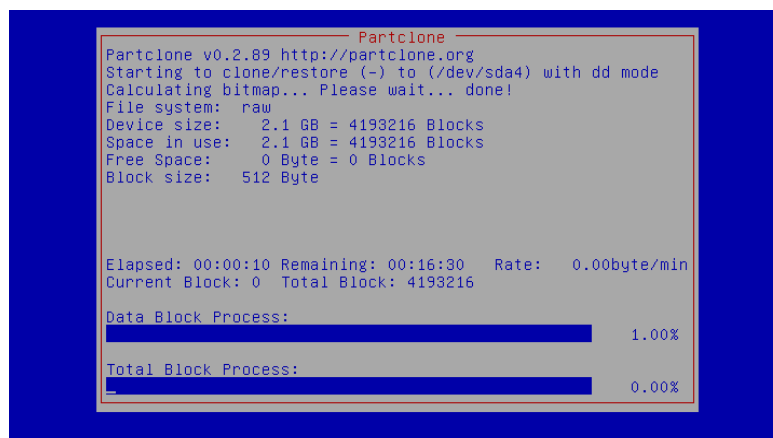
Restaurar la imagen desde el cliente

Configurar el cliente para iniciar desde red.

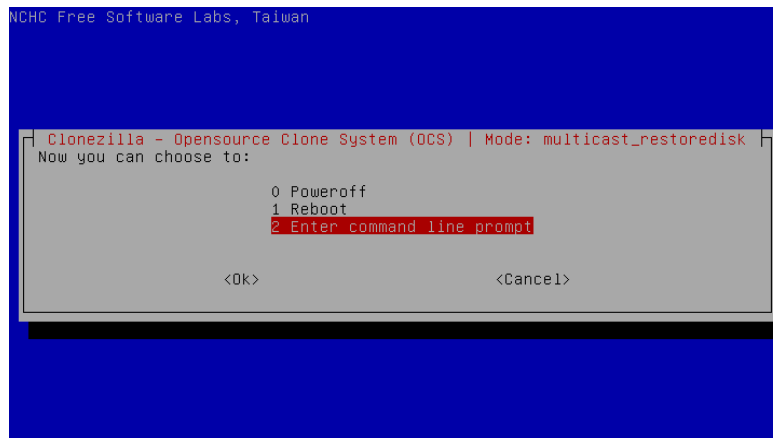
Aparecerá una pantalla similar y hay que esperar unos segundo o dar enter



Después de unos segundos, se puede ver cómo empieza el proceso de restauración



Al finalizar, aparecerá una pantalla con opciones



De la misma manera que al guardar la imagen, en el servidor se pueden ver los datos de la restauración

```
Client 192.168.100.1 (xx:xx:xx:xx:xx:xx) finished cloning. Stats:  
/dev/sda4, success, 2.1 GB = 4193216 Blocks, 1.301 mins; /dev/sda5,  
success, 72.7 MB = 35522 Blocks, .636 mins; /dev/sda6, success,  
85.1 MB = 166180 Blocks, .605 mins; /dev/sda7, success, 618.6 MB  
= 302068 Blocks, .873 mins; /dev/sda8, success, 5.4 MB = 2658  
Blocks, .601 mins;
```


4.1 Comparación de tiempos en el proceso de generación y recuperación

Para esta medición, la imagen se fue creando remotamente mediante NFS, compartiendo la carpeta de destino alojada en el servidor de Clonezilla, a excepción de Redo Backup, en donde se utilizó sshfs para la transferencia de archivos.

Las características de los equipos utilizados son los siguientes:

- Procesador Intel® Xeon® Processor X5550 a 2.67GHz
- Memoria RAM de 24 GB
- El disco duro tiene capacidad de 1 TB con espacio utilizado de 20 GB
- Sistema Operativo CentOS 6.8 con software mínimo instalado
- Switch 3COM Baseline 2126-G

	Clonezilla Server	Mondo Rescue	Redo Backup
Creación de la imagen	21.9 minutos	40.43 minutos	29 minutos
Tamaño de la imagen	12 GB	15 GB	13 GB
Restauración (1 nodo)	25.6 minutos	50 minutos	28.9 minutos
Restauración múltiple (4 nodos)	21.6 minutos	No soportado	No soportado
Tipo compresión	gzip	gzip	gzip
Paralelización en compresión	Sí	No soportado	No soportado
Protocolo de transferencia	NFS	NFS	sshfs

Con las tres herramientas probadas, se pudieron detectar las capacidades y los inconvenientes de cada una.

Redo Backup se utiliza sin tener que hacer instalaciones utilizando un *live CD* y es muy sencillo de usar sin tener la necesidad de utilizar la línea de comandos, teniendo una interfaz gráfica en la que se puede seleccionar el destino y el origen para la crear la imagen de disco. Puedes disponer de la imagen desde un directorio compartido por red y estas configuraciones se pueden hacer en cuestión de minutos. Aunque puede servir si se quiere tener una imagen de disco y restablecerla una vez o en un periodo tiempo largo en alguna computadora personal, no es apto para restablecer varios equipos a la vez por lo que no se tomó como opción para utilizarlo en un clúster.

Con Mondo Rescue se necesita tener un sistema operativo de tipo GNU/Linux para poder instalar sus paquetes. Su configuración es un poco tediosa porque cuenta con muchas opciones para crear la imagen, aunque puedes dejar prácticamente todo por defecto sin efectos negativos, ya que lo más importante a configurar es la ruta de origen, de destino y si es el caso, el tipo y nivel de compresión, dependiendo del espacio que se tenga para alojar la imagen que se va a crear. Gracias a las pruebas realizadas, se puede decir que su mayor ventaja, es la de crear imágenes en caliente, lo que la mayoría de este tipo de software no puede hacer. El tiempo tardado y que no pueda reestablecer varios equipos a la vez, fueron factor para no ser tomado en cuenta para implementarlo.

Por último, Clonezilla Server Edition tiene la opción de ser instalado en la mayoría de las distribuciones de GNU/Linux o poder iniciar desde un *live CD*. Su configuración puede ser un poco tardada cuando se hace la primera vez, pero vale la pena si se va a estar en constante uso, ya que tan solo correr un par de comandos en el servidor, puede iniciar el proceso de creación o restauración de una imagen. Una de sus mayores ventajas es que funciona como cliente/servidor, pudiendo guardar una imagen creada desde un cliente detectado por red LAN y teniendo como opción paralelizar la compresión y llevarse menos tiempo en la creación de la imagen. De la misma manera, puede restablecer la imagen en varios clientes a la vez tomándose prácticamente el mismo tiempo en terminar, que si lo hiciéramos en un solo cliente. Lo mencionado anteriormente, es lo que se tomó en cuenta para elegir este programa y ser una herramienta eficaz para añadir nodos a clústeres de alto rendimiento.

4.2 Elección de *software* para implementación

Las guías del capítulo 3, sirven para tener un panorama de todas las funciones que manejan cada una de las herramientas analizadas. En conjunto con el cuadro comparativo del capítulo 2, se opta por utilizar Clonezilla Server Edition, considerando los siguientes aspectos:

- Automatización y facilidades en las configuraciones y en su uso
- Soporte de múltiples sistemas de archivos
- Por tener arquitectura cliente/servidor, se puede almacenar la imagen creada de un nodo y cuando se requieran añadir extras, solo hay que configurar el cliente para proceder con la copia de la imagen
- Si por alguna razón llegara a fallar el servidor en donde se aloja Clonezilla, se puede utilizar el *Live CD* para restaurar o agregar un nodo, teniendo un respaldo de la imagen
- Cuenta con varias opciones de compresión
- Se puede paralelizar el tipo de compresión para que sea mucho más rápido el proceso
- No tiene limitaciones con marcas de computadoras
- Es totalmente gratuito

Conclusiones

En la actualidad, las copias de seguridad son recomendadas por expertos en el tema u organizaciones importantes en el mundo de la informática y aunque solo se respaldan los archivos, son de gran ayuda en caso de borrar información o cualquier circunstancia que pueda comprometer nuestros datos almacenados.

Por otra parte, existe *software* que puede ayudar a respaldar tanto los archivos almacenados, como toda la estructura de un disco duro, evitando volver a instalar un sistema operativo completo, controladores y configuraciones que pueden ser muy tardadas.

Este proyecto se enfocó principalmente en el uso de herramientas para crear imágenes de disco y así reducir el tiempo que se toma en las instalaciones y configuraciones, empleando aplicaciones sin costo.

A continuación se mencionan los objetivos con sus respectivas conclusiones y así poder saber si se cumplieron o no.

Objetivo: Estudiar cómo está constituido un clúster

Conclusión:

La investigación dio como resultado los conocimientos necesarios para saber la definición de clúster, cuáles son sus componentes y sus piezas fundamentales para que puedan funcionar como tal. También se puntualizaron los tipos más conocidos de clústeres para poder dar una idea de cuál se podría utilizar para determinada tarea y con cuál estamos trabajando. Abarcando estos puntos, podemos decir que el objetivo fijado se cumplió al desarrollar el tema de clústeres.

Objetivo: Instalar las herramientas en un clúster de pruebas

Conclusión:

Este objetivo se pudo cumplir, gracias a las pruebas realizadas en el Instituto de Investigaciones en Materiales con equipos prestados, instalando satisfactoriamente las tres herramientas y realizando guías de instalación para poder basarse en ellas cuando se requiera tener estos programas sin necesidad de estar buscando documentación. También pueden ser tomadas en cuenta por cualquier persona u organización.

Objetivo: Configurar adecuadamente las herramientas mencionadas anteriormente para poder comparar los beneficios que tiene cada una para elegir la más capaz y eficiente

Conclusión:

De la misma manera, se hicieron guías para poder configurar Mondo Rescue, Redo Backup y Clonezilla, basándose en la práctica. Las configuraciones fueron exitosas y no se presentó ningún error. Se hicieron cuadros para poder comparar las herramientas y así poder elegir la más idónea, teniendo en cuenta como principal factor, la rapidez en que hacen su trabajo de sacar una imagen y restaurarla en uno y varios nodos, cumpliendo este objetivo.

Objetivo: Implementar de un herramienta en un clúster de la DGTIC

Conclusión:

Actualmente, la DGTIC está dando mantenimiento a la infraestructura, por lo que varios equipos están apagados e imposibilitó que se pudieran hacer pruebas y dejar funcionando el servidor de Clonezilla. Sin embargo, con el material generado como las guías se podrá dar continuidad a este proyecto.

Glosario

Clonación. Proceso en el cual se realiza una copia exacta de un disco a otro disco directamente.

Clúster. Conjunto de computadoras que están conectadas mediante una red para comunicarse y funcionar como una sola, siendo más potente.

Cómputo distribuido. Repartición de trabajos en varias computadoras para concluirlos en un tiempo menor.

Cómputo paralelo. Utilización de varios procesadores para completar una tarea que se reparte para concluir más rápido.

DHCP (*Dynamic Host Configuration Protocol*). Servicio que se encarga de asignar direcciones IP.

Imagen de disco. Archivo o archivos que contienen toda la estructura y archivos de un disco o partición creado con anterioridad.

LAN (*Local Area Network*). Red de dispositivos conectados entre sí en un área pequeña como una casa o un negocio.

Middleware. Aplicación que tiene la función de que el usuario como las computadoras puedan ver un clúster como un solo equipo.

NAT (*Network Address Translation*). Encargado de que las direcciones IP privadas tengan salida a internet.

NFS (*Network File System*). Comparte archivos dentro una red por medio de una carpeta.

NIS (*Network Information Services*). Servicio que se comparte por la red para evitar información duplicada en los demás equipos.

Supercomputadora. Es una computadora que cuenta con *hardware* más potente que el las computadoras personales y servidores, siendo mucho más rápida al procesar datos.

Referencias

- [1] Alegre, M. (2010). *Sistemas operativos monopuesto*. España: Paraninfo. pp. 172.
- [2] Andrews, J. (2016). *Lab Manual*. Estados Unidos: Cengage Learning. pp. 588.
- [3] Arch Linux. (s.f.). *Clonezilla 3.21.13-1*. Recuperado el 18 de septiembre de 2016, de <https://www.archlinux.org/packages/community/any/clonezilla/>
- [4] Arch Linux. (s.f.). *Dosfstools 4.0-3*. Recuperado el 18 de septiembre de 2016, de <https://www.archlinux.org/packages/core/i686/dosfstools/>
- [5] Arch Linux. (s.f.). *Ntfs-3g 2016.2.22-1*. Recuperado el 18 de septiembre de 2016, de https://www.archlinux.org/packages/extra/x86_64/ntfs-3g/
- [6] Arch Linux. (s.f.). *Parted 3.2-5*. Recuperado el 18 de septiembre de 2016, de https://www.archlinux.org/packages/extra/x86_64/parted/
- [7] Arch Linux. (s.f.). *Partimage 0.6.9-9*. Recuperado el 18 de septiembre de 2016, de https://www.archlinux.org/packages/community/x86_64/partimage/
- [8] Berral, I. (2010). *Operaciones auxiliares de montaje de componentes informáticos*. España: Paraninfo. pp. 106-107, 120-121.
- [9] Buyya, R. (1999). *High Performance Cluster Computing, Volume 1*. Estados Unidos: Prentice Hall. pp. 4-13.

- [10] Colobran, M., Arqués J., Galindo, E. (2008). *Administración de sistemas operativos en red*. España: Editorial UOC. pp. 36-37.
- [11] Cruz, E. & Gordillo, J. *Supercómputo en la UNAM*. Entérate [en línea]. Noviembre de 2002, no. 14. Recuperado el 15 de septiembre de 2016, de <http://www.enterate.unam.mx/Articulos/2002/noviembre/supercom.htm>
- [12] DGTIC, UNAM. (s.f.). *Acuerdo TIC*. Recuperado el 31 de agosto de 2016, de <http://www.tic.unam.mx/pdfs/AcuerdoTIC.pdf>
- [13] DGTIC, UNAM. (s.f.). *Cronología*. Recuperado el 2 de septiembre de 2016, de <http://www.historiadelcomputo.unam.mx/cronologia.html>
- [14] DGTIC, UNAM. (s.f.). *Organización*. Recuperado el 25 de septiembre de 2016, de <http://www.super.unam.mx/index.php/home/acerca-de/organizacion>
- [15] DGTIC, UNAM. (s.f.). *¿Qué es Supercómputo?*. Recuperado el 2 de septiembre de 2016, de <http://www.super.unam.mx/index.php/quesc>
- [16] Fernández, R. *En el principio*. Revista Digital Universitaria [en línea]. 31 de marzo de 2000, vol. 1, no. 0. Recuperado el 1 de septiembre de 2016, de <http://www.revista.unam.mx/vol.0/art4/princi.html>
- [17] Forum Incident Response Security Team. (s.f.). *FIRST is the global Forum for Incident Response and Security Teams*. Recuperado el 21 de diciembre de 2016, de <https://www.first.org/>
- [18] Free Software Foundation. (s.f.). *¿Qué es el software libre?*. Recuperado el 12 de noviembre de 2016, de <https://www.gnu.org/philosophy/free-sw.es.html>
- [19] Guy, G., & Fugazzi, T. *Method of and apparatus for duplicating, upgrading and configuring hard disk drives*. Estados Unidos, US 6,179,492 B1. (G06F8/63; G06F11/1446; G06F2003/0697; G06F3/0601), 30 de enero de 2001. Appl. US09097558, 15 de junio de 1998. p. 1.
- [20] Jorba, J. & Suppi, R. (2004). *Administración avanzada de GNU/Linux*. España: UOC. pp. 21-22.

- [21] Lyons, D. (2011). *Cluster Computing for Robotics and Computer Vision*. Estados Unidos: World Scientific. p. 4.
- [22] Mondo Rescue. (s.f.). *About Mondo Rescue*. Recuperado el 30 de octubre de 2016, de <http://www.mondorescue.org/about.shtml>
- [23] Mondo Rescue. (s.f.). *Download Mondo Rescue*. Recuperado el 30 de octubre de 2016, de <http://www.mondorescue.org/downloads.shtml>
- [24] NCHC Free Software Labs. (s.f.). *Clonezilla*. Recuperado el 28 de octubre de 2016, de <http://clonezilla.org/>
- [25] NCHC Free Software Labs. (s.f.). *DRBL-Installation*. Recuperado el 10 de noviembre de 2016, de <http://drbl.org/installation/04-netboot-client.php>
- [26] NCHC Free Software Labs. (s.f.). *Partclone Features*. Recuperado el 10 de noviembre de 2016, de <http://partclone.org/features/>
- [27] NCHC Free Software Labs. (s.f.). *What are the compression options in Clonezilla? What are the differences?*. Recuperado el 29 de octubre de 2016, de http://drbl.org/faq/fine-print.php?path=./2_System/78_clonezilla_compression_option.faq
- [28] Nelson, B., Phillips, A. & Steuart, C. (2014). *Guide to Computer Forensics and Investigations*. Estados Unidos: Cengage Learning. p. 104.
- [29] Ortiz, D., Rodríguez, F. & Coello, C. *La computación en México: los primeros años*. Revista Digital Universitaria [en línea]. 10 de septiembre de 2008, vol. 9, no. 9. Recuperado el 31 de agosto de 2016, de <http://www.revista.unam.mx/vol.9/num9/art63/int63-1.htm>
- [30] Pan, J., Chen, S. & Nguyen, N. (2012). *Intelligent Information and Database Systems: 4th Asian Conference, ACIIDS 2012, Kaohsiung, Taiwan, March 19-21, 2012, Proceedings, Part II*. Estados Unidos: Springer Science & Business Media. pp.75-76
- [31] Price, A. (2010). *Easy PC Care: Be Your Own Expert: Teach Yourself*. Gran Bretaña: Hachette. pp. 83-84.
- [32] Rodríguez, L. & Carnota, R. (2015). *Historias de las TIC en América Latina y el Caribe: Inicios, desarrollos y rupturas*. Barcelona, España: Fundación Telefónica. pp. 149-151.

- [33] Sommerville, I. (2005). *Ingeniería del software*. España: Pearson Educación. p. 242.
- [34] UNAM. *Acuerdo por el cual se reestructuran los servicios de cómputo en la Universidad Nacional Autónoma de México*. Gaceta UNAM [en línea]. 16 de mayo de 1985, vol. 1, no. 10. Recuperado el 30 de agosto de 2016, de <http://acervo.gaceta.unam.mx/index.php/gum80/article/download/22266/22264>
- [35] UNAM. *Acuerdo por el que cambia de denominación la Dirección General de Servicios de Cómputo Académico a Dirección General de Cómputo y de Tecnologías de Información y Comunicación de la Universidad Nacional Autónoma de México*. Gaceta UNAM [en línea]. 27 de septiembre de 2010, no. 4279. Recuperado el 30 de agosto de 2016, de <http://acervo.gaceta.unam.mx/index.php/gum10/article/download/64606/64576>
- [36] UNAM. (s.f.). *Catálogo de Publicaciones Periódicas del Subsistema de la Investigación científica de la UNAM*. Recuperado el 31 de agosto de 2016, de http://www.paginaspersonales.unam.mx/files/175/Publica_20150727161502.pdf
- [37] UNAM. *Nacen los Centros de Investigación Matemáticas y de Servicios de Cómputo*. Gaceta UNAM [en línea]. 16 de marzo de 1973, vol. 5, no. extraordinario. Recuperado el 30 de agosto de 2016, de <http://acervo.gaceta.unam.mx/index.php/gum70/article/download/6970/6968>
- [38] UNAM. (s.f.). *Programa Universitario de Cómputo*. Recuperado el 2 de septiembre de 2016, de <http://historiadeldcompu.unam.mx/files/acervos/PUC.pdf>
- [39] Vergara, C. *Computación en la UNAM y México, una retrospectiva*. Entérate [en línea]. Agosto de 2008, no. 71. Recuperado el 29 de agosto de 2016, de <http://www.enterate.unam.mx/artic/2008/agosto/art4.html>
- [40] Welsh, M. (2003). *Running Linux*. Estados Unidos: O'Reilly. pp. 562-563.
- [41] Wolf, G., Ruiz, E., Bergero F. & Meza E. (2015). *Fundamentos de sistemas operativos*. Distrito Federal: Universidad Nacional Autónoma de México, Instituto de Investigaciones Económicas y Facultad de Ingeniería. p 62.
- [42] Zebradots. (s.f.). *Redo Backup & Recovery*. Recuperado el 9 de noviembre de 2016, de <http://redobackup.org/>

[43] Zeller, T. (2014). *Safe Circuit*. Recuperado el 5 de noviembre de 2016, de <http://www.admin-magazine.com/Archive/2014/22/Redo-Backup>