

LA VIRTUALIZACIÓN, UN ENFOQUE EMPRESARIAL HACIA EL FUTURO

VIRTUALIZATION, A BUSINESS APPROACH TO THE FUTURE

ABSTRACT

This paper aims to establish a serie of pillars to help guide more clearly when and how to run a virtualization project. The first part consists of a conceptual explanation to a better understanding of the importance of virtualization and cloud computing. Then a series of key questions to profile the best type of virtualization, the most appropriate provider are formulated, and finally, we present a case report to illustrate virtualization.

Key words: apache hadoop, cloud computing, guide, hardware, infrastructure, software, virtualization.

RESUMEN

Este documento tiene como objetivo establecer una serie de pilares para ayudar a orientar con mayor claridad cuándo y cómo ejecutar un proyecto de virtualización. La primera parte consta de una explicación conceptual, para comprender mejor la importancia de la virtualización y el cloud computing. A continuación presentamos una serie de preguntas claves para perfilar el mejor tipo de virtualización, el proveedor mas indicado y finalmente se presenta un reporte de caso para ilustrar la virtualización.

Palabras claves: apache hadoop, computación en la nube, guía, hardware, infraestructura, software, virtualización.

Nancy Yaneth Gélvez García

Magister en Ciencias de la Información y las Comunicaciones
Docente planta de la Universidad Distrital Francisco José de Caldas
nygelvezg@udistrital.edu.co
Bogotá, Colombia

Carlos Andrés Moreno Giraldo

Estudiante de Ingeniería de Sistemas
Universidad Distrital Francisco José de Caldas
camorenog@correo.udistrital.edu.co
Bogotá, Colombia

Diego Giovanni Ruiz Zambrano

Estudiante de Ingeniería de Sistemas
Universidad Distrital Francisco José de Caldas
dgruizz@correo.udistrital.edu.co
Bogotá, Colombia

Tipo: Artículo reporte de caso

Fecha de Recepción: Febrero 27 de 2013

Fecha de Aceptación: Mayo 2 de 2013

1. INTRODUCCIÓN

Luego de un proceso de recopilación de fuentes de información relativamente divergentes; un filtro y análisis reflexivo; presentamos en este artículo un conjunto de ideas que pretende orientar a los lectores hacia el cuestionamiento de lo poco que conocemos con respecto al mundo de la virtualización y su gran potencial como nicho de proyectos y de un gran accionar profesional por parte de la Ingeniería.

No se pretende promover el uso de una determinada marca de aplicación para virtualización, sino más bien orientar la toma de una decisión de este tipo mediante un sustento técnico y estratégico.

2. CONTEXTO TEÓRICO

Como preámbulo, es completamente necesario definir los temas base del presente documento, para comprender mejor el marco general de referencia, capturar la esencia de lo que se pretende resaltar y sobre todo generar inquietudes que permitan la ampliación del conocimiento en el tema central.

La infraestructura (tecnológica) es el conjunto de elementos que se consideran necesarios para el funcionamiento de una organización o para el desarrollo de una actividad [1]. En general se identifican dos clases de infraestructura: infraestructura hardware (física) e infraestructura software (lógica). La primera, consta de elementos tan diversos como aires acondicionados, sensores, cámaras, servidores, routers, firewalls, computadores portátiles, impresoras, teléfonos, etc.

El conjunto de elementos lógicos o de software comprende desde los sistemas operativos (Linux, Windows, etc.) hasta aplicaciones del ámbito general que permiten el funcionamiento de otros sistemas informáticos concretos de los servicios, como las bases de datos, servidores de aplicaciones o las herramientas ofimáticas.

Por otra parte, el término virtualización se pue-

de entender como la creación a través de software de una versión virtual de algún recurso tecnológico, como por ejemplo una plataforma de hardware, un sistema operativo, un dispositivo de almacenamiento u otros recursos de red [2].

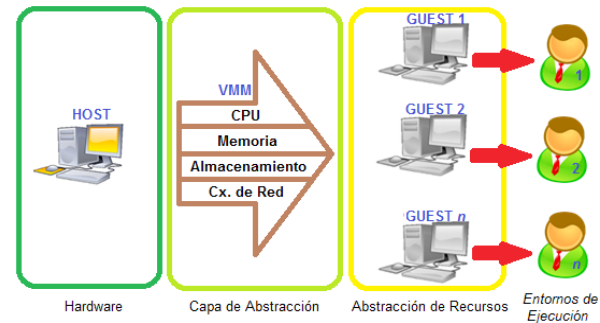


Figura 1. Representación del proceso de virtualización.

En la figura 1, se representa el proceso de virtualización, en el que el Hypervisor o VMM (Virtual Machine Monitor) crea una capa de abstracción entre el hardware del host y el SO de la(s) máquina(s) virtual(es), administrando de manera adecuada sus 4 recursos principales (CPU, memoria, almacenamiento y conexiones de red).

3. ESTADO DEL ARTE

La consolidación del modelo de la computación en la nube ha encaminado a la virtualización, hoy en día, como un requisito cotidiano dentro de los recursos tecnológicos que todo tipo de empresas requiere para su funcionamiento, accediendo permanentemente a su información y reduciendo sus gastos de capital.

Las ventajas de este modelo son resumibles en tres factores: economía, flexibilidad y seguridad. Según Efraín Soler, CEO de O4IT (empresa experta en estos temas y con una fuerte presencia en Colombia), “una solución en la nube reemplaza los gastos de capital por gastos operativos, que permiten una mejor planificación del presupuesto y un mejor retorno sobre la inversión. Con una solución en la nube se pueden agregar o remover puestos y servidores, así como modificar sus prestaciones o configura-

ción de forma casi inmediata” [3].

Y que no es posible hablar de virtualización sin hablar un poco más del modelo de computación en la nube, ya que según las propias predicciones de Efraín Soler, el avance en este sentido ocupa varios escenarios prometedores basados en la investigación y el desarrollo [3].

3.1. Consolidación del hardware

Lo primero es la consolidación tecnológica en lo concerniente al hardware necesario para la operación de proveedores de servicios en la nube.

3.2. Necesidad de estandarización

También se verán iniciativas de estandarización de protocolos, servicios y modelos de trabajo en la nube que repercutirán en beneficios para aquellas empresas que usen estos servicios, ya que podrán hacer transiciones entre países sin preocuparse por migraciones o adaptaciones tecnológicas.

3.3. Usuarios del hogar

Para el usuario final es muy posible que la experiencia de cómputo sea integrada aún más en la oficina y en el hogar, con equipos que operen cada vez menos como una computadora tradicional y más como un componente entre los necesarios para un trabajo productivo o para entretenimiento.

Dichos escenarios coinciden completamente en los vacíos que se van haciendo más notorios dentro del campo de los proyectos de virtualización. En primer lugar, la consolidación del hardware es un inconveniente dentro del dimensionamiento inicial en un proceso de virtualización ya que se vuelve completamente dependiente de la herramienta a elegir (o esto es de lo que comercialmente pretenden vencer).

Estrechamente relacionado con lo anterior se encuentra que posiblemente no exista un patrón que permita identificar características co-

munes de un proyecto de virtualización a otro; enfrentando a un universo nuevo de interrogantes por cada proyecto que se emprenda.

Finalmente, la virtualización es una tendencia tan fuerte, que ya no es un tópico solamente de grandes corporaciones u organizaciones mega-tecnológicas sino también es un tema de personas del común, con necesidades bastante puntuales para las cuales un proyecto de virtualización puede ser la solución.

4. CLARIFICANDO EL PANORAMA

A continuación y según lo investigado es atreverse a presentar la serie de pasos o hitos que se consideran fundamentales para emprender de una manera más transparente, un proyecto de virtualización de infraestructura tecnológica.

Lo primero que se debería hacer, es preguntarse si realmente necesita virtualizar parte o todo su proceso tecnológico actual. Entre más preguntas de las que a continuación se listan conteste afirmativamente, más se acercará a una respuesta afirmativa para virtualizar [4].

- ¿Su compañía o el proceso del cual hace parte está sufriendo o puede sufrir de un crecimiento descontrolado?
- ¿Sus servidores y sistemas de almacenamiento de información están desorganizados?
- ¿Los activos de TI de su organización están subutilizados?
- ¿La disponibilidad de energía es insuficiente?
- ¿Los costos de administración de su equipo/infraestructura IT son demasiado altos?
- ¿El retorno de la inversión en IT es bajo?
- ¿Posee un downtime excesivo?
- ¿Las infraestructuras que maneja son inflexibles?
- ¿Sus datos son inmóviles?

Una vez se ha confirmado que se debe virtualizar, lo siguiente que debería hacerse es fijar un alcance al proyecto de virtualización; es decir ¿qué se va a virtualizar? Las diferentes

compañías que ofrecen estos servicios dividen de acuerdo a la arquitectura o filosofía de su herramienta y dan una guía diferenciada muy acorde a la misma. Sin embargo son completamente identificables los diferentes tipos de clasificación que existen (tabla 1).

Tabla 1. Tipos de clasificación [5].

Nombre	Descripción
Por hardware	Anillo de privilegio "-1" usado para aislar todas las capas superiores de software de las operaciones de virtualización.
De almacenamiento	Abstrae el almacenamiento lógico del almacenamiento físico, y es comúnmente usado en SANs ("storage area network" red de área de almacenamiento).
Particionamiento	Es la división de un solo recurso, como en espacio de disco o ancho de banda de la red, en un número más pequeño y con recursos del mismo tipo.
Máquina virtual	Compartición de recursos locales físicos entre varios dispositivos virtuales.
Hypervisor de almacenamiento	Es un pack portátil de gestión centralizada, utilizado para mejorar el valor combinado de los sistemas de disco de almacenamiento múltiples.

Es muy importante también identificar en qué momento se encuentra la organización objeto del proyecto de virtualización: surgimiento, expansión o transformación.

Para una empresa que apenas se está consolidando, lo más recomendable es que le dé prioridad a no invertir demasiado en capital de TI sino en asegurar la información, que es el insumo más importante para asegurar su crecimiento. La virtualización enfocada al almacenamiento y en corto tiempo es la recomendada.

Si la organización está en proceso de crecimiento, lo mejor sería pensar en una solución que le permita integrar diferentes plataformas (ya sea porque el crecimiento se está dando por adquisición de otras compañías) o una que le brinde mayor escalabilidad (por diversificación de mercado o cobertura), pero en ambos casos

asegurando muy bien el tema de seguridad.

Finalmente, para aquellas organizaciones en proceso de algún tipo de transformación, se recomienda una armonización entre los beneficios de una virtualización enfocada al almacenamiento y otra enfocada a los servidores de aplicación. Las organizaciones que experimentan este estado, son especialmente sensibles a la pérdida de información y no es nada deseable experimentar interrupciones demasiado largas que afecten el negocio.

5. ESCALANDO A UN NIVEL MÁS TÉCNICO

Los cuatro requerimientos para que una arquitectura pueda ser considerada como virtualizable [4].

1. Dos modos de operación.
2. Un método para que programas no privilegiados, llamen rutinas privilegiadas del sistema.
3. Un mecanismo de reubicación o protección de memoria como la segmentación o la paginación.
4. Interrupciones asíncronas para permitir al sistema E/S comunicarse con el procesador.

6. IMPORTANTE: MARCAS

VMWare es mundialmente reconocido como la marca líder en productos de virtualización, pero como se ha venido analizando, elegir el software adecuado para la virtualización del centro de datos es una tarea compleja que no debe ser acotada solo por este tema. Hyper-V de Microsoft se ha convertido en un competidor formidable para VMware, especialmente con Windows Server 2008 R2 y Windows Server 2012. Por lo que se puede decir que en la actualidad, estas dos son las principales opciones de virtualización de servidores; es posible conformar una terna o mejor, un podio con VirtualBox la opción OpenSource.

¿Cómo elegir cuál?. Existen varios tips o claves para hacer una buena elección mirando marcas; por ejemplo por costo; si la organización

ya tiene una cuenta de Windows Server 2012 o la plataforma de SO Windows Server 2008, se puede descargar del servidor Hyper-V sin costo. El único costo es el marco de gestión de System Center. Microsoft incluye la gestión de los entornos físicos y virtuales con Hyper-V y VMware. Hyper-V proporciona capacidades de migración: la migración en vivo se incluye en Windows Server sin cargo adicional. Según varias declaraciones (casos de éxito) dan fe de la solidez de VMware. Con VMware, VMotion, tanto en Foundation y ediciones Standard, hay un cargo adicional si se desea agregar capacidades de migración.

El tema de marcas, también va de la mano con la envergadura de la organización o proceso, así pues, las corporaciones maduras y con un buen capital pueden optar por VMWare o Hiper-V de Microsoft. Para usuarios “hogar” o con poca experiencia VirtualBox es una excelente alternativa, gracias a su interfaz intuitiva, potencia funcional y ser SW libre. Para decidir entre VMWare y Hiper-V, se puede retomar el primer párrafo de este apartado.

Para aquellas organizaciones más arriesgadas o con más poder científico existen alternativas como XEN y/o Qemu que pueden cumplir con las expectativas a un mayor costo técnico.

7. EJEMPLO PRÁCTICO

Para una mejor asimilación de lo descrito hasta este momento y una aproximación más acertada de lo que se requiere para organizar y crear un plan de virtualización, se presenta a continuación un ejemplo práctico, compuesto de la siguiente manera:

7.1. Objetivo

Montar un clúster con Apache Hadoop (de ahora en adelante llamado simplemente Hadoop) en 3 nodos para desplegar dos aplicaciones que permitan descargar imágenes de forma distribuida y ser guardadas en las diferentes máquinas que hacen parte del clúster, aprovechando el procesamiento de Hadoop.

7.2. Marco teórico

Sistema distribuido: se define como una colección de computadoras separadas físicamente y conectadas entre sí por una red de comunicaciones distribuida; cada máquina posee sus componentes de hardware y software que el usuario percibe como un solo sistema. El usuario accede a los recursos remotos (RPC) de la misma manera en que accede a recursos locales, o un grupo de computadores que usan un software para conseguir un objetivo en común.

Hadoop: es un framework Java de código abierto para el procesamiento y consulta de grandes cantidades de datos sobre grandes grupos de productos de hardware. Hadoop es un proyecto de Apache nivel superior, iniciado y dirigido por Yahoo!. Se apoya en una comunidad activa de colaboradores de todo el mundo por su éxito.

Con una inversión significativa en tecnología de Yahoo!, Apache Hadoop se ha convertido en una tecnología de cloud computing empresarial y se está convirtiendo en el estándar de facto de la industria para grandes procesamientos de datos [7].

Cluster: conjunto o conglomerado de computadoras construido mediante la utilización de hardwares comunes y que se comporta como si fuese una única computadora.

Nodo: Puede ser un simple computador, sistema multiprocesador o una estación de trabajo (workstation). En lo que a este tema respecta, de forma muy general, un nodo es un punto de intersección o unión de varios elementos que confluyen en el mismo lugar.

Hadoop Distributed File System (HDFS): es un sistema de archivos distribuido, escalable y portátil escrito en Java para el framework Hadoop. Cada nodo en una instancia Hadoop típicamente tiene un único nodo de datos; un clúster de datos forma el clúster HDFS.

MapReduce: consiste en un Job Tracker (rastreador de trabajos), para el cual las aplicacio-

nes cliente envían trabajos MapReduce.

El rastreador de trabajos (Job Tracker) impulsa el trabajo fuera a los nodos Task Tracker disponibles en el clúster, intentando mantener el trabajo tan cerca de los datos como sea posible.

7.3. Infraestructura utilizada

SO: el nodo maestro utiliza Windows Vista (32 bits), un nodo esclavo utiliza Windows Vista (32 bits) y otro nodo esclavo instalado como máquina virtual utiliza Windows 7.

Hardware: maestro, procesador Turion 2.20 Ghz, 1406 RAM, DD 120 GB y Windows Vista 32. Esclavo 1, procesador Turion 2 Ghz, 2000 RAM, DD 120 GB y Windows Vista 32. Esclavo 2, Virtual 960 RAM, DD 10 GB y Windows 7 32.

Software: sobre estas máquinas se ha instalado CYGWIN [8]. Cada máquina tiene instalado el JDK con las variables de entorno JAVA_HOME, PATH y CYGWIN_HOME, las cuales fueron configuradas como se muestra en la figura 2.

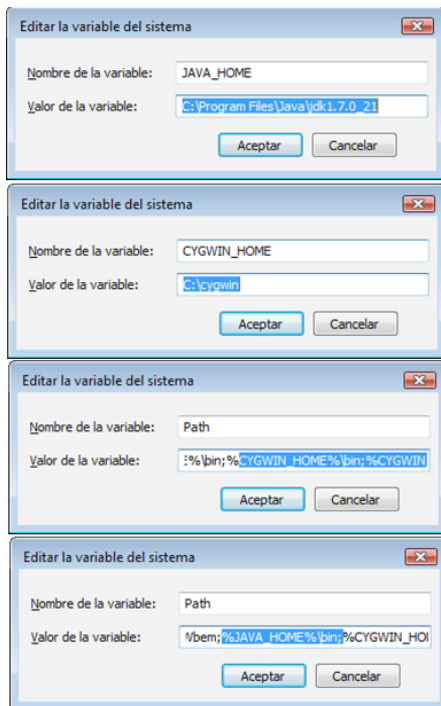


Figura 2. Configuración de las variables de entorno en Windows.

7.4. Parametrizaciones previas

En las tres máquinas se desactivó el antivirus y los cortafuegos de Windows, para evitar problemas de acceso a través de los puertos.

Se tuvieron en cuenta también incidencias como el uso de Hadoop 0.20.2 y cuestiones de formato para un correcto funcionamiento como el uso de cuentas sin caracteres especiales o rutas con espacios [9].

7.5. Configuración de máquinas

Se asignaron direcciones IP listadas en la tabla 2, para las tres máquinas utilizando un AP inalámbrico como gateway (192.168.0.1) que también opera como servidor DNS preferido; además se estableció el mismo grupo de trabajo para las tres máquinas, con su respectivo nombre.

Tabla 2. Configuración de red de los nodos Hadoop.

Nombre máquina	Usuario	Descripción	Dirección IP
Nelson1	Nelson	Master - nodo maestro	192.168.0.100
Laptop Andres	Carlos	Slave1 -nodo esclavo 1	192.168.0.101
GT1-PC	GT1	Slave2 -nodo esclavo 2	192.168.0.102

Adicionalmente se ha hecho la respectiva referencia de acceso de cada nodo en el archivo HOSTS de Windows en cada máquina, como se muestra en la figura 3.

```

19 127.0.0.1 localhost
20 192.168.0.100 master
21 192.168.0.101 slave1
22 192.168.0.102 slave2

```

Figura 3. Archivo HOSTS modificado.

7.6. Configuración con CYGWIN

A través de la consola de CYGWIN se configura el servicio SSH en todas las máquinas:

```
$ ssh-host-config
```

Se generan las claves públicas y privadas:

```
$ ssh-keygen -t dsa -P "" -f ~/.ssh/id_dsa
$ cat ~/.ssh/id_dsa.pub >> ~/.ssh/authorized_keys
```

Y se hace el intercambio de claves entre las máquinas. ES muy importante tener presente que en los nombres de los usuarios no se usen caracteres especiales ni espacios.

Desde el maestro se copia la clave a un archivo en la carpeta .ssh de cada computador esclavo

```
$ scp ~/.ssh/id_dsa.pub Carlos@LaptopAndres:~/.ssh/master-key.pub
```

```
$ scp ~/.ssh/id_dsa.pub GT1@GT1-PC:~/.ssh/master-key.pub
```

Y en los computadores esclavos se copia esta clave a la carpeta de de claves autorizadas

```
$ cat ~/.ssh/master-key.pub >> ~/.ssh/authorized_keys
```

Esto permitirá al nodo maestro tener acceso (figura 4) a los dos esclavos sin tener que escribir una contraseña.

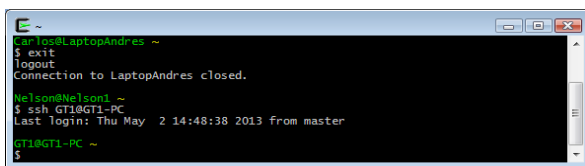


Figura 4. Acceso desde el maestro a los esclavos sin contraseña.

De forma similar se hace el intercambio de claves de los esclavos al maestro. Desde cada computador esclavo se copia la clave a la carpeta .ssh del computador maestro:

```
$ scp ~/.ssh/id_dsa.pub Nelson@Nelson1:~/.ssh/slave-key.pub
```

Y en el computador maestro se copian estas claves a la carpeta de claves autorizadas.

```
$ cat ~/.ssh/slave-key.pub >> ~/.ssh/authorized_keys
```

Con esto se logra que los nodos esclavos tengan acceso al nodo maestro. Además se ha creado el archivo config para el SSH en las máquinas. Este archivo contiene los nombres de los Host y usuarios del clúster que se quiere implementar, como se muestra en la figura 5.

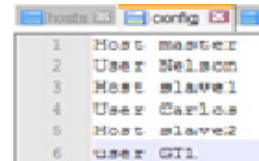


Figura 5. Archivo SSH modificado.

7.7. Instalación de Hadoop

Se ha descargado Apache Hadoop en su versión 0.20.2 -esta versión permite trabajar la API de HIPI-. Para utilizarla en las tres máquinas simplemente se ha descomprimido el contenido en la ruta \usr\local\hadoop en la carpeta donde se encuentra instalado CYGWIN, en cada máquina.

En la carpeta /config del hadoop que se ha extraído se cambian los archivos "core-site.xml", "hdfs-site.xml" y mapred-site.xml, tal como se observa en la figura 6.

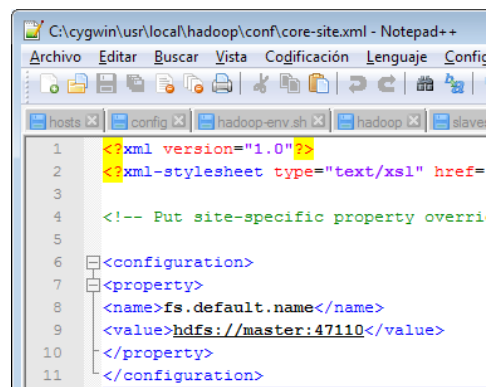


Figura 6. Configuración del núcleo de Hadoop.

Como se observa en las figuras 7 y 8, se establece replicación en los tres nodos, maestro, esclavo 1 y esclavo 2.

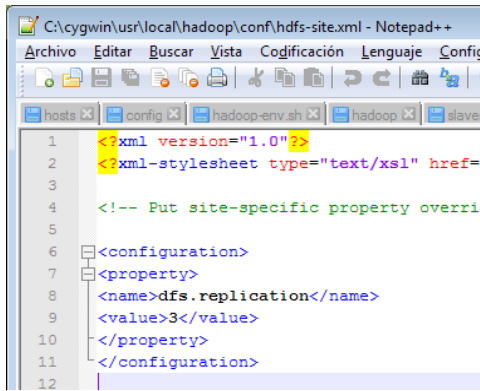


Figura 7. Hdfs (sistema de ficheros) de Hadoop.

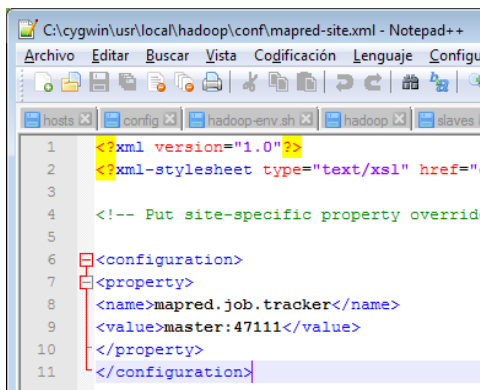


Figura 8. Configuración del MapReduce.

Además se configuran los archivos “hadoop-env.sh”, “hadoop” y hadoop-env.sh para definir correctamente las rutas de la variable de entorno del Java, JAVA_HOME. En la consola de CYGWIN se comprueba esta instalación de Hadoop (figura 9).

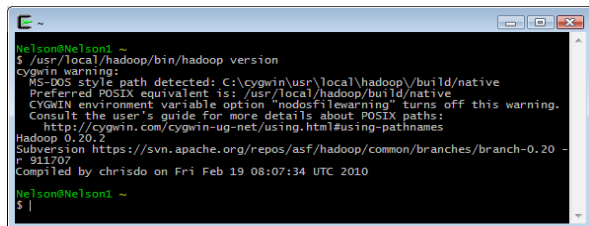


Figura 9. Comprobación de la Instalación de Hadoop.

Además, es necesario ubicar el archivo “slaves” en donde se deben inscribir los esclavos que hacen parte del clúster.

Es importante dar formato al *namenode* con el comando:

```
$ /usr/local/hadoop/bin/hadoop name-  
node -format
```

Ahora se puede iniciar el clúster con hadoop utilizando el archivo start-all.sh. Esto inicia el HDFS y el MapReduce (figura 10).

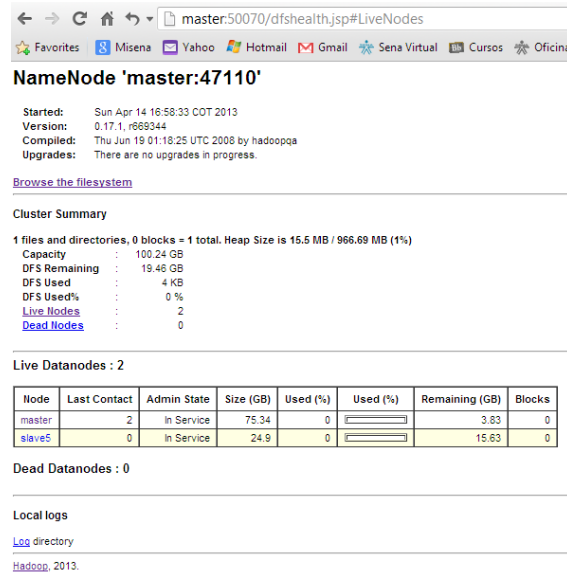


Figura 10. Comprobación de Hadoop.

8. APLICACIÓN DE DESCARGA DE IMÁGENES

Las aplicaciones utilizadas para la descarga distribuida de imágenes y extracción de las mismas son ejemplos tomados de la API de HIPI [10], los ejemplos son llamados Distributed Downloader y JpegFromHib.

Estos utilizan la API y el framework MapReduce; se compilaron utilizando eclipse y el plugin de Map Reduce que ofrece hadoop (agregar configuración de eclipse y compilación de ejemplos HIPPI) [11]. En este caso no se profundiza en explicar o detallar la descarga y la compilación en eclipse de HIPI ya que no es necesariamente el tema principal de interés.

9. APLICACIÓN CORRIENDO EN EL CLUSTER

A través del DistributedDownloader se leerá un archivo con una serie de url referentes a las

imágenes; este proceso “mapeará” las imágenes en el clúster y las “reducirá” en un archivo disponible en el HDFS para todos los nodos.

Para esto se ha creado una carpeta local (url) en el nodo maestro con un archivo .txt que contiene las url de las imágenes que se quieren descargar (figura 11), en este caso hacen referencia a la dirección de la aplicación en el Tomcat del maestro (*master:8080/web-page/index.html*).

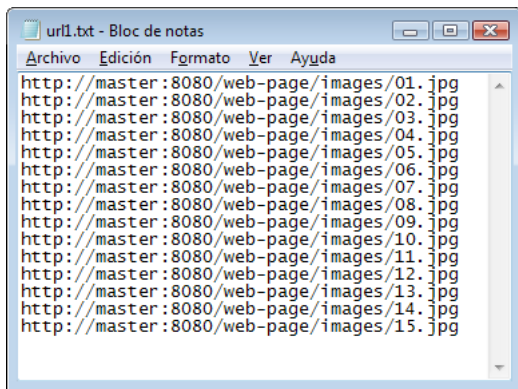


Figura 11. Archivo de URL de imágenes.

Primero, desde el nodo maestro se debe crear una carpeta también llamada url en el DFS del clúster, tal cual como se muestra en la figura 12.

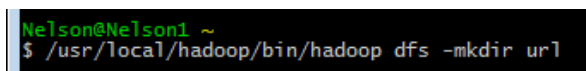


Figura 12. Creación de la carpeta 'usr' en el DFS.

También se creará una carpeta llamada “download” en el DFS para almacenar el archivo comprimido de imágenes (HIB). En esta carpeta se copiará el TXT con la lista de rutas para descargar las imágenes guardadas en el directorio local.

Se ejecuta el JAR creado previamente para DistributedDownloader (ver ejecución del JAR con Hadoop) pasando parámetros de entrada (<Clase principal> <txt con urls> <ruta-dfs/archivo.hib> <#nodos>) para guardar las imágenes en un comprimido .HIB (HIPI Image Bundle) similar a tar y disponerlas en el HDFS (figura 13).

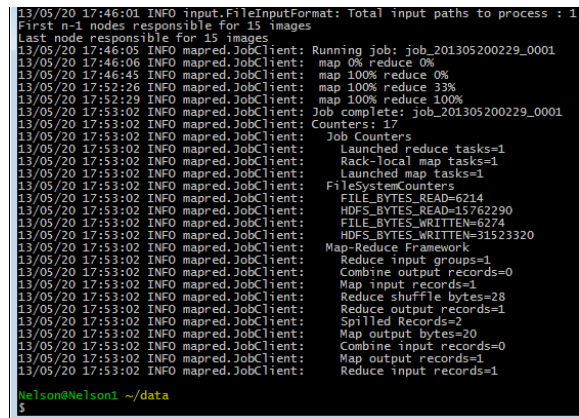


Figura 13. Se puede observar el procesamiento con MapReduce.

Ya que el DFS es el sistema de archivos de todo el clúster, se puede ver el HIB (figura 14) creado desde cualquier esclavo [12].

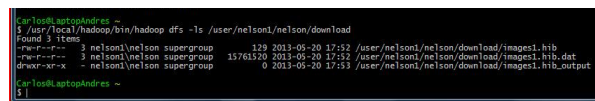


Figura 14. Vista del HIB desde el esclavo 1.

Hecha esta descarga se hace la extracción de las imágenes que están en el archivo HIB (también en el DFS), ejecutando JpegFromHib –también desde el nodo maestro, y de manera similar con parámetros de entrada (<Clase principal> <ruta-dfs/archivo.hib> <carpeta-donde-se guardarán>).

Para copiarlos a cualquier equipo (en la figura se utiliza desde el nodo esclavo 1) se puede utilizar el comando CopyToLocal de hadoop (ver ejemplo en la figura 15).

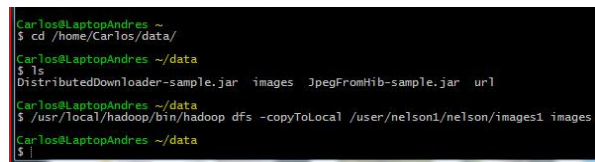


Figura 15. Uso del comando "CopytoLocal".

Las imágenes quedan disponibles para el uso local, en el esclavo 1, como se observa en la figura 16.

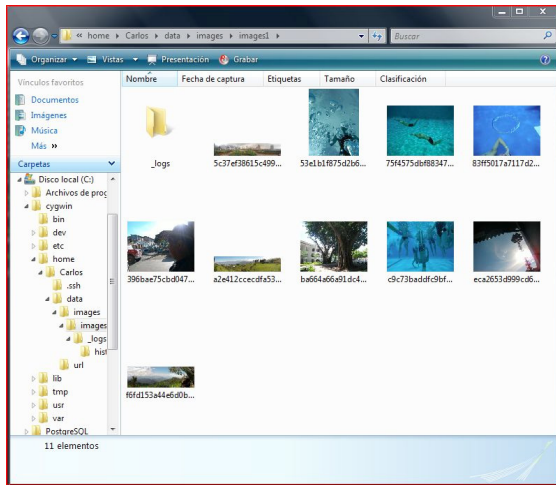


Figura 16. Imágenes visibles y disponibles desde cualquier parte del cluster.

10. CONCLUSIONES

A medida que el modelo de Cloud Computing se extiende, la virtualización se hace más fuerte e indispensable; así como la necesidad de crecer en estándares de diseño e integración que la regulen.

Las soluciones de infraestructura virtual son ideales para entornos de producción en parte debido a que se ejecutan en servidores y escritorios estándar de la industria y son compatibles con una amplia gama de sistemas operativos y entornos de aplicación, así como de infraestructuras de red y almacenamiento.

Como una solución en la nube se pueden agregar o remover puestos y servidores, así como modificar sus prestaciones o configuración de forma casi inmediata.

Las tecnologías de virtualización pueden optimizar el rendimiento y simplificar la administración de la infraestructura de información.

Es gratificante tener la oportunidad y la orientación para indagar sobre temas de vanguardia, que nos permiten mantener un enfoque investigativo y reflexivo hacia el mundo y hacia dónde debemos apuntar los ingenieros en el ejercicio de nuestra profesión.

Referencias Bibliográficas

- [1] Universitat Oberta de Catalunya (2013). [En línea], consultado en Febrero 1 de 2013, disponible en: http://www.uoc.edu/portal/es/tecnologia_uoc/infraestructures/index.html
- [2] Turban, E; King, D; Lee, J; Viehland, D (2008). «Chapter 19: Building E-Commerce Applications and Infrastructure». *Electronic Commerce A Managerial Perspective* (5th edición). Prentice-Hall, pp. 27.
- [3] Artículo de la Revista Portafolio 8 de Febrero de 2013. Transcrito en la página Oficial de O4IT S.A.S. [En línea], consultado en Diciembre 15 de 2012, disponible en: <http://o4it.com/es/la-nube-es-el-aqui-y-el-ahora/>
- [4] EMC, guía para virtualización de su infraestructura de Información. [En línea], consultado en Febrero 12 de 2013, disponible en: <http://colombia.emc.com/collateral/software/15-min-guide/h2977-virtualization.pdf>.
- [5] "Infrastructure Planning and Design Guides for Virtualization". Disponible en <http://technet.microsoft.com/en-us/solutionaccelerators/ee395429>.
- [6] Golberg, R.P. *Virtual Machines – Semantic and Examples*. Cambridge, Massachusetts, Estados Unidos: Proceedings of the workshop on virtual computer systems, 1973- p 74-112.
- [7] Hadoop at Yahoo! (2013). [En línea], consultado en Noviembre 20 de 2012, disponible: <http://developer.yahoo.com/hadoop/>.
- [8] Instalaciones y parametrizaciones basadas en las notas de clase y el artículo: Running Hadoop on Ubuntu Linux (Multi-Node Cluster). [En línea], consultado en Agosto

- 11 de 2012, disponible en: <http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-multi-node-cluster/>.
- [9] Hadoop Configuration on Windows through Cygwin. [En línea], consultado en Septiembre 18 de 2012, disponible: <http://stackoverflow.com/questions/12378199/hadoop-configuration-on-windows-through-cygwin>.
- [10] HIPI, Hadoop Image Processing Interface. [En línea], consultado en Agosto 5 de 2012, disponible en <http://hipi.cs.virginia.edu/>.
- [11] HIPPI, MapReduce Tutorial. . [En línea], consultado en Diciembre 7 de 2012, disponible en http://hadoop.apache.org/docs/r1.0.4/mapred_tutorial.html.
- [12] HDFS Architecture Guide. . [En línea], consultado en Noviembre 22 de 2012, disponible en http://hadoop.apache.org/docs/r1.0.4/hdfs_design.html.