

## “Creación de un Cluster de Linux utilizando Knoppix”

**Autores:** Alicia Guadalupe Valdez Menchaca  
Maestra en Tecnologías de Información  
Facultad de Ingeniería Mecánica y Eléctrica  
Universidad Autónoma de Coahuila  
[avm10190@mail.uadec.mx](mailto:avm10190@mail.uadec.mx)

MI Gloria Campos Posada  
Facultad de Ingeniería Mecánica y Eléctrica  
Universidad Autónoma de Coahuila  
[gecposada@yahoo.com.mx](mailto:gecposada@yahoo.com.mx)

### **Resumen:**

En la Facultad de Ingeniería de la Universidad Autónoma de Coahuila se ha integrado un grupo de maestros para crear un Cluster de Computadoras utilizando software libre, en este caso Knoppix el cual es una distribución Linux, para el apoyo de los estudiantes de Sistemas Operativos.

### **Introducción**

El cómputo con clusters surge como resultado de la convergencia de varias tendencias actuales que incluyen la disponibilidad de microprocesadores económicos de alto rendimiento y redes de alta velocidad, el desarrollo de herramientas de software para cómputo distribuido de alto rendimiento, así como la creciente necesidad de potencia computacional para aplicaciones que la requieran.

**Cluster** es un grupo de múltiples [computadoras](#) unidas mediante una [red](#) de alta velocidad, de tal forma que el conjunto es visto como una única computadora, más potente que las comunes de escritorio.

El cómputo con clusters surge como resultado de la convergencia de varias tendencias actuales que incluyen la disponibilidad de microprocesadores económicos de alto rendimiento y redes de alta velocidad, el desarrollo de herramientas de software para cómputo distribuido de alto rendimiento, así como la creciente necesidad de potencia computacional para aplicaciones que la requieran.

## Desarrollo

Los cluster permiten aumentar la escalabilidad, disponibilidad y fiabilidad de múltiples niveles de red.

## **Características**

A continuación se definen las características más importantes:

- *Escalabilidad*: capacidad de un equipo de hacer frente a volúmenes de trabajo cada vez mayores, sin dejar por ello de prestar un nivel de rendimiento aceptable.

- *Disponibilidad*: es la capacidad de estar presente, de estar listo en un determinado momento en el que se quiere hacer uso.

- *Fiabilidad*: es la probabilidad de funcionamiento correcto.

Un cluster puede presentarse como una solución de especial interés sobre todos a nivel de empresas, las cuales pueden aprovecharse de estas especiales características de computación para mantener sus equipos actualizados por un precio bastante más económico que el que les supondría actualizar todos sus equipos informáticos y con unas capacidades de computación que en muchos casos pueden llegar a superar el hardware de última generación.

## **Tipos**

Básicamente se distinguen tres tipos de cluster atendiendo al uso que se le da:

- **High Availability**: [Alta disponibilidad](#) Consiste en la conexión de una o varias computadoras conectadas en red utilizándose una conexión *heartbeat* para monitorear cual de sus servicios está en uso, así como la sustitución de una máquina por otra cuando uno de sus servicios haya caído.

- **Load-balancing**: [Equilibrio de carga](#) Utilizado en los servidores web, el cluster verifica cual de las máquinas de éste posee mayores recursos libres y así, asignarle el trabajo pertinente. Actualmente, los *cluster load-balancing* son también *fail-over* con el extra de balanceo de carga y número de nodos.

- **High Performance computing:** Alto rendimiento Cluster destinado al alto rendimiento, capacidad muy alta de proceso para cómputo de grandes volúmenes de datos.

De un cluster se espera que presente combinaciones de los siguientes servicios:

1. Alto rendimiento (*High Performance*)
2. Alta disponibilidad (*High Availability*)
3. Equilibrio de carga (*Load Balancing*)
4. Escalabilidad (*Scalability*)

La construcción de las computadoras del cluster es más fácil y económica debido a su flexibilidad: pueden tener todas la misma configuración de hardware y sistema operativo (cluster homogéneo), diferente rendimiento pero con arquitecturas y sistemas operativos similares (cluster semi-homogéneo), o tener diferente hardware y sistema operativo (cluster heterogéneo).

Para que un cluster funcione como tal, no basta solo con conectar entre sí las computadoras, sino que es necesario proveer un sistema de manejo del cluster, el cual se encargue de interactuar con el usuario y los procesos que corren en él para optimizar el funcionamiento.

### ***Componentes de un Cluster***

En general, un cluster necesita de varios componentes de software y hardware para poder funcionar:

- **Nodos:** Las computadoras o servidores
- **Sistemas Operativos**
- **Conexiones de Red**
- **Middleware**

Es el que recibe los trabajos entrantes al cluster y los redistribuye de manera que el proceso se ejecute más rápido y el sistema no presente sobrecargas en un servidor. Esto se realiza mediante políticas definidas en el sistema (automáticamente o por un administrador) que le indican dónde y cómo debe distribuir los procesos, por un sistema de monitorización, el cual controla la carga de cada CPU y la cantidad de procesos en él.

- Protocolos de Comunicación y servicios.
- Aplicaciones



Figura1: Clúster Linux "Colmena" instalado en la UABC

Cómputo paralelo es la división de una gran tarea o trabajo computacional de cierta cantidad de partes para ejecutarse de manera simultanea en todos los nodos. La gran mayoría de las tareas son fraccionables.

La división de las tareas puede hacerse de 2 formas:

- De manera manual. Esto es indicando explícitamente en el programa como debe de dividirse, se requiere para ello tener instaladas las librerías para el paso de mensajes entre los procesadores (MPI - interfase para el Paso de Mensajes o PVM - Máquina virtual Paralela).
- De manera automática. Esto es configurando el Cluster con OpenMosix para crear una sola imagen del sistema operativo común a todos los nodos. De esta forma los nodos se comunican entre si sin la necesidad de la intervención del usuario manteniendo un optimo balanceo en la división y distribución de la carga. En otras palabras el paralelismo transparente al usuario quien trabaja en la maquina como si esta fuese de un solo procesador pero de mucha mayor capacidad.

### **Aplicaciones.**

El cómputo paralelo se vuelve indispensable cuando se trabaja con algoritmos complejos o con manejo intensivo de datos, donde una computadora con un solo procesador tardaría horas o días y que requerimos de los resultados a la mayor brevedad posible.

Gracias al poder de cómputo que se logra con un cluster, la simulación ha pasado a ser una herramienta mas para la investigación. Y mediante esta se resuelven problemas de gran escala, por ejemplo:

Aerodinámica: simulaciones de vuelo.

Astrofísica: modelado de la evolución de las galaxias, dinámica estelar (el Instituto de astronomía de la UNAM en Ensenada cuenta con un cluster para esto).

Biología: modelado de nuevos componentes genéticos, simulación de redes neuronales (gracias al uso de supercomputadoras fue posible descifrar el genoma humano).

Ciencias de la computación - simulación de circuitos integrados y de nuevos microprocesadores, minería de datos, inteligencia artificial.

Química: predicción de reacciones químicas.

Ingeniería: análisis de estructuras, cálculos, cómputo numérico.

Geofísica: exploración sísmica, filtrado de imágenes satelitales.

Física Nuclear: simulación de reacciones nucleares.

El objetivo de la utilización de un cluster en el área de la enseñanza es en apoyo a los alumnos de maestría y a los alumnos de licenciatura.

## **Tecnologías de Implementación**

### **MOSIX**

Es un paquete de software que mejora el kernel de Linux con capacidades de computación de clusters. El kernel mejorado permite cualquier cluster de estaciones de trabajo y servidores X86/Pentium/AMD trabajar coordinadamente como parte de un sólo sistema.

MOSIX es una extensión del kernel de Linux que permite ejecutar aplicaciones "normales" (no paralelizadas) en un Cluster. Una de las posibilidades de MOSIX es la "migración de procesos", que permite migrar procesos de nodo en nodo. Si por ejemplo, cierto proceso está dominando la carga de un nodo, este será movido a otro que tiene más recursos. Una de las características de MOSIX es que, a diferencia de otros clusters, no es necesario modificar las aplicaciones ni tampoco utilizar librerías especiales. De hecho, tampoco es necesario asignar "a mano" los procesos a los diferentes nodos que componen el cluster.

La idea es que después de la creación de un nuevo proceso (fork), MOSIX intenta asignarlo al mejor nodo disponible en ese entonces. MOSIX

monitorea constantemente los procesos, y si fuera necesario, migrará un proceso entre los nodos para maximizar el rendimiento promedio.

MOSIX realiza todo esto automáticamente, bajo el concepto de "**fork and forget**" al igual que en un sistema SMP. Esto significa que sólo algunas aplicaciones se beneficiarán de un cluster MOSIX, básicamente:

- Procesos que requieren de mucho CPU, aplicaciones científicas, de ingeniería, etc.
- Procesos paralelos, especialmente los que tienen tiempos de ejecución impredecibles.
- Clusters con nodos de diferentes velocidades y/o distintas cantidades de memoria.
- Entornos multiusuario y de tiempo compartido.
- Servidores WEB escalables.

MOSIX funciona silenciosamente. Sus operaciones son transparentes para las aplicaciones. Los usuarios no necesitan saber dónde se están ejecutando los procesos, tampoco necesitan preocuparse de lo que están haciendo otros usuarios.

Como MOSIX está implementado en el kernel de Linux, sus operaciones son totalmente transparentes para las aplicaciones. Esto permite definir distintos tipos de clusters, incluso un cluster con diferentes CPU's o velocidades LAN.

Otra característica de MOSIX, es que sus algoritmos son descentralizados - esto significa que cada nodo puede ser el maestro de los procesos creados localmente, y un servidor de los procesos remotos que migraron desde otros nodos. Esto permite agregar o remover nodos desde el cluster en cualquier momento.

Los nodos destinados a MOSIX deben tener instalados Linux La instalación automática de MOSIX es compatible con RedHat 5.1, 6.0, 6.2, 7.0, 7.1 y SuSE 6.0, 6.1, 6.2, 6.3, 7.0. En caso de preferir otra distribución, simplemente se debería seguir un procedimiento de instalación manual que se encuentra documentado en los fuentes de MOSIX, siempre y cuando se cumplan los requisitos: **make** versión 3.77 o superior, **gcc** versión 2.95.3, 2.95.4 o 2.96.74+, y **binutils** 2.9.1.0.25 o superior.

## Knoppix

Knoppix es una [distribución](#) de [GNU/Linux](#) . Está desarrollada por el consultor de [GNU/Linux Klaus Knopper](#).

### *Distribuciones derivadas de Knoppix*

- [ClusterKnoppix](#) es una distribución basada en Knoppix y que utiliza [Linux Terminal Server Project](#) y [OpenMosix](#). Es una vía conveniente para probar configuraciones en cluster de computadores cluster.

### *Posibilidades*

Se puede usar de distintas formas como:

- Para enseñar y demostrar de manera sencilla el sistema GNU/Linux, especialmente como sistema operativo.
- Probar rápidamente la compatibilidad de hardware bajo Linux antes de comprarlo o utilizarlo, especialmente para tarjeta de vídeo.
- Utilizar las herramientas incluidas para restaurar un sistema corrupto o sus datos perdidos.

Según la siguiente grafica de sistemas operativos utilizados se tiene:

Cuota de Mercado	Sistema Operativo	Composición
GNU/Linux	28.5%	GNU/Linux
Windows	24.4%	Todos los Windows combinados (incluyendo 95, 98, NT)
Sun	17.7%	Sun Solaris o SunOS
BSD	15.0%	Familia BSD (FreeBSD, NetBSD, OpenBSD, BSDI, ...)
IRIX	5.3%	SGI IRIX

Tabla 1 - Sistemas operativos utilizados

### **Instalación:**

1. Configuración básica del ClusterKnopixx.
2. Configuración de la Red.
3. Consola Root (root shell).
4. Asignación de una IP al nodo.
5. Verificación de la Conectividad.
6. Inicio de los servicios de OpenMosix
7. Verificación de los servicios de OpenMosix.



## Conclusiones

El trabajo expuesto en este artículo nos ha mostrado que el software GNU es una valiosa herramienta en la enseñanza de los sistemas operativos.

Actualmente se ha instalado exitosamente el cluster por lo que estamos en desarrollo de aplicaciones para ser manejadas por el cluster.

## Futuras aplicaciones

Los cluster en el futuro serán utilizados en un grid, que vendrá siendo un cluster de cluster, que facilitará la investigación y desarrollo tecnológico de las grandes comunidades académicas.

En el 2002 se lanzó la propuesta en la UNAM sobre la creación de un macrocluster con las siguientes palabras: " La UNAM tiene un gran potencial para convertirse en la universidad de la información y la computación al servicio de la producción y la transmisión del conocimiento. Ninguna otra institución nacional suma el capital intelectual, la fuerza computacional, la conectividad y el acceso a la gran cantidad de acervos de información científica y cultural del mundo. Ejercer un liderazgo a nivel nacional en estas áreas significará un importante servicio para México. Aceptando el desafío de conjuntar todos los recursos universitarios pertinentes, en el marco de este proyecto se plantea la creación de una red académica de colaboración, con el objetivo de integrar capacidades, a fin de hacer investigación y crear, tanto tecnología como cultura, para la transformación del campus universitario en un modelo del aprovechamiento de la información digital, la conectividad y las técnicas emergentes de la computación y la inteligencia computacional.

## Bibliografía

Gallego Catalina, UNIX, Linux Iniciación y Referencia, 2ª. Edición ISBN: 8448140095, Editorial McGraw-Hill

García Tomás Jesús, Santiago Fernando, Redes para proceso distribuido, ISBN: 970-15-0339-2, 2000, Editorial Alfa-Omega.

Cárdenas Haro José Antonio, Artículo: "Clúster Linux "Colmena", Revista "Gaceta" editada por la Universidad Autónoma de Baja California. Edición 27 de Marzo 2004.

Sitio Web, Noviembre 2007, <http://www.sourcepole.com/sources/software/gis-knoppix/index.xml>



Buyya, R. (1999). High Performance Cluster Computing: Programming and Applications, Volume II, *Prentice Hall*, Upper Saddle River, New Jersey.

# **SOCIEDAD DE LA INFORMACION**

[www.sociedadelainformacion.com](http://www.sociedadelainformacion.com)

Edita:



Director: José Ángel Ruiz Felipe  
Jefe de publicaciones: Antero Soria  
Luján  
D.L.: AB 293-2001  
ISSN: 1578-326x