

Instalación y Configuración de un Sistema de Supercomputos a
Medida (Cluster de Computadoras) para Desarrollar Tareas de
Computación en Paralelo Bajo Linux

Sergio Arteaga Molina

Fundación Universitaria San Martín - Escenarios Múltiples

Facultad de Ingeniería de Sistemas

Barranquilla

2005

INSTALACIÓN Y CONFIGURACIÓN DE UN SISTEMA DE SUPERCÓMPUTO A
MEDIDA (CLUSTER DE COMPUTADORAS) PARA DESARROLLAR TAREAS
DE COMPUTACIÓN EN PARALELO BAJO LINUX

SERGIO ARTEAGA MOLINA

FUNDACION UNIVERSITARIA SAN MARTIN – ESCENARIOS MULTIPLES

FACULTAD DE INGENIERIA DE SISTEMAS

BARRANQUILLA

2005

**INSTALACIÓN Y CONFIGURACIÓN DE UN SISTEMA DE SUPERCÓMPUTO A
MEDIDA (CLUSTER DE COMPUTADORAS) PARA DESARROLLAR TAREAS
DE COMPUTACIÓN EN PARALELO BAJO LINUX**

SERGIO ARTEAGA MOLINA

Proyecto de Grado para optar al título de

INGENIERO DE SISTEMAS

Director

SERGIO JIMÉNEZ MARTINEZ

INGENIERO DE SISTEMAS

FUNDACION UNIVERSITARIA SAN MARTIN – ESCENARIOS MULTIPLES

FACULTAD DE INGENIERIA DE SISTEMAS

BARRANQUILLA

2005

NOTA DE ACEPTACIÓN

Firma presidente del jurado

Firma del jurado

Firma del jurado

Barranquilla, Julio de 2005

TABLA DE CONTENIDO

	Pág
0. ANTECEDENTES DEL PROYECTO	7
0.1 IDENTIFICACION DEL PROBLEMA	7
0.1.1 AREA DEL CONOCIMIENTO	7
0.1.2 AREA TEMÁTICA	7
0.1.3 TEMA	7
0.1.4 TITULO	7
0.2 PLANTEAMIENTO DEL PROBLEMA	7
0.2.1 DESCRIPCIÓN	7
0.2.2 FORMULACION	8
0.3 JUSTIFICACIÓN	9
0.4 OBJETIVOS	10
0.4.1 OBJETIVO GENERAL	10
0.4.2 OBJETIVO ESPECIFICO	10
0.5 MARCO REFERENCIAL	10
0.5.1 MARCO TEORICO	10
0.5.1.1 El concepto de cluster	10
0.5.1.2 Características de un cluster	12
0.5.1.3 Acoplamiento de un cluster	14
0.5.1.4 Esquema y otras características	16
0.5.1.5. Homogeneidad de un cluster	19
0.5.1.6 Clasificación según el servicio prioritario	19
0.5.1.7 Alto rendimiento (HP, high performance)	20
0.5.1.8 Alta disponibilidad (HA, high availability)	20
0.1.5.9 Alta confiabilidad (HR, high reliability)	21

0.5.2	MARCO CONCEPTUAL	21
0.6	METODOLOGÍA	26
0.6.1	TIPO DE INVESTIGACIÓN	26
0.6.2	FUENTES DE INFORMACIÓN BASICA	26
1.0	CONSTRUCCIÓN DEL CLUSTER	28
1.1	HARDWARE DEL CLUSTER	27
1.1.1	TOPOLOGÍA DE LA RED	27
1.1.2	CONFIGURACIÓN NODOS	28
1.1.3	CONFIGURACIÓN NODO PRINCIPAL	28
1.1.4	CONEXIONES FÍSICAS DE LAS MAQUINAS EN LA RED	28
1.2	SOFTWARE BASE	29
1.2.1	INSTALACIÓN DE LA DISTRIBUCIÓN MADRAKE LINUX 9.1 SERVIDOR Y NODOS	29
1.2.2	ADMINISTRACION DE LINUX MANDRAKE 9.1	38
1.2.3	CONFIGURACIONES DE RED	39
1.3	PROTOCOLO SSH	43
1.3.1	FUNCIONAMIENTO DEL SSH	44
1.3.2	METODOS DE AUTENTICACION DE USUARIOS	45
1.3.3	AUTENTICACION CON CONTRASENAS	45
1.3.4	AUTENTICACION CON CLAVE PUBLICA	46
1.3.5	CONFIGURACION DE OPENSSSH	47
1.3.6	CONFIGURACION DEL ENTORNO	48
1.4	SERVICIO NFS	50
1.4.1	CONFIGURACION DEL SERVIDOR NFS	50
1.4.2	PUESTA EN MARCHA DEL SERVICIO	51
1.4.3	CONFIGURACION DEL CLIENTE NFS	52
1.5	OPENMOSIX	52
1.5.1	CARACTERISTICAS DE OPENMOSIX	52

1.5.2	SUBSISTEMAS DE OPENMOSIX	53
1.5.3	INSTALACION DEL CLUSTER OPENMOSIX	54
1.5.3.1	INSTALACION DEL KERNEL DE OPEN LINUX	54
1.5.3.2	CONFIGURACION DE LA TOPOLOGIA DEL CLUSTER	57
1.5.3.3	SCRIPT DE INICIALIZACION	58
1.5.4	MIGRAION DE TAREAS	58
1.6	OPENMOSIXVIEW	60
1.6.1	INSTALACION DE OPENMOSIXVIEW	61
1.6.2	ENTORNO DEL OPENMOSIXVIEW	62
1.7	PRUEBA DEL OPENMOSIX Y DEL OPENMOSIXVIEW	64
	BIBLIOGRAFÍA	70

0. ANTECEDENTES DEL PROYECTO

0.1 IDENTIFICACION DEL PROBLEMA

0.1.1 Área del Conocimiento: Redes de computador- Sistemas Distribuidos.

0.1.2 Área Temática: Sistemas distribuidos y redes de computadoras.

0.1.3 Tema: La construcción del cluster de computadora con Linux.

0.1.4 Título: Instalación y configuración de un sistema de supercómputo a medida (cluster de computadoras) para desarrollar tareas de computación en paralelo bajo Linux.

0.2 PLANTEAMIENTO DEL PROBLEMA

0.2.1 Descripción: En la actualidad el comercio está basado en nuevas tiendas y negocios en Internet que funcionan 24 horas al día, 7 días a la semana, que solo necesitan personal, técnico, para su puesta en marcha y al que se accede remotamente para su administración. Con este nuevo tipo de negocio, muchas empresas hacen inversiones en equipo y personal técnico, para ofrecer a nivel

mundial soluciones que de otra manera podrían ser inviables por precio, tiempo u organización, pero exigiéndole a estas nuevas tecnologías lo siguiente:

- Máximo rendimiento, mínimo costo.
- Máximo aprovechamiento de los recursos existentes.
- Disponibilidad máxima.
- Confiabilidad máxima.
- Adaptación a los cambios, tanto en forma de carga para el sistema como en forma de nuevo planteamiento del negocio. El sistema debe ser flexible y escalable.

Este último punto es importante por motivos claramente económicos (no solo a nivel de empresa) al suponer un gran reto en el diseño de sistemas para que estos puedan adaptarse de manera eficiente a nuevas exigencias.

Basados en la realidad de nuestra región, las empresas poseen un problema aún mas grande, y es encontrar una asesoría técnica para diseñar e implementar este tipo de soluciones debido a que en la ciudad no se cuenta con instituciones educativas que ofrezcan programas de capacitación a nivel específico.

0.2.2 Formulación: ¿Como construir un cluster de computadora de alto rendimiento para ejecutar operaciones en paralelas a bajo costo?

0.3 JUSTIFICACIÓN

El diseño de un sistema de supercómputo “a medida” o clusters de computadoras es la herramienta que ha revolucionado la teoría de la computación de alto rendimiento y alta disponibilidad, debido a que se beneficia de la utilización de ordenadores convencionales que pueden ser adquiridos a un bajo costo por su producción en serie. Por otro lado emplea como plataforma el sistema operativo Linux, un derivado del UNIX de libre distribución que posee una gran estabilidad y robustez, y herramientas GNU

La utilización del cluster de computadoras se ha venido dando hace unos cuantos años, por lo cual son pocas las personas que poseen la formación necesaria para instalar y administrar este tipo de soluciones, es por esto que se ha generado la necesidad de capacitar personal técnico competente para apoyar a la mediana y gran empresa, así como también a las universidades.

Como consecuencia de lo anterior es importante el desarrollo de este proyecto para dejar por sentado una metodología para la construcción y administración del cluster de computadoras, que sirva como material de apoyo a la formación del estudiante y personal técnico capaz de seleccionar el hardware y configurar software que más se adapte a las necesidades del sistema de supercómputo requeridas por la empresa.

0.4 OBJETIVOS

0.4.1 Objetivo General: Elaborar un manual metodológico para la instalación y configuración de un cluster de ordenadores de alto rendimiento que permita la ejecución de tareas de computación en paralelo utilizando herramientas GNU /Linux.

0.4.2 Objetivos Específicos:

- Analizar diferentes tipos de clusters de ejecución en paralelo.
- Configurar servidor y nodos clientes con sistema operativo Linux.
- Instalar diferentes herramientas de configuración de cluster.
- Analizar las diferentes herramientas de administración del cluster.
- Realizar pruebas de ejecución en paralelo.

0.5 MARCO REFERENCIAL

0.5.1 Marco Teórico:

0.5.1.1 El concepto de cluster:

Un cluster podemos entenderlo como:



Nota

Este trabajo se encuentra disponible únicamente en **formato físico**, ya que no cuenta con versión digital. Por tal motivo, en este repositorio solo serán visibles las páginas preliminares con fines informativos. Para su consulta completa, el usuario deberá verificar la ubicación o sede donde se encuentra el ejemplar físico, o bien solicitar información adicional escribiendo al correo institucional:

repositoriofusm@sanmartin.edu.co