



Universidad
Tecnológica
de Pereira

UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTAD DE INGENIERÍAS
Programa de Ingeniería de Sistemas y Computación



1. Datos Personales del Estudiante	
Código	1115421345
Nombres	Héctor Fabio
Apellidos	Jiménez Saldarriaga
email	hfjimenez@utp.edu.co
Teléfonos	+573175548245

Nombre del Director	Ramiro Andrés Barrios Valencia
VoBo Director(Firma)	
VoBo Comité Curricular	

2. Modalidad		
1. Trabajo de Investigación Formativa	Proyecto de Investigación	
	Proyecto de Aplicación	X
	Monografía	
2. Práctica de Extensión	Práctica Universitaria	
	Emprendimiento Empresarial	
	Proyecto Social	



Universidad
Tecnológica
de Pereira

UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTAD DE INGENIERÍAS
Programa de Ingeniería de Sistemas y Computación



3. Título del Trabajo de Grado
Despliegue Automático de Infraestructuras HPC multinucleo y multimaquina con maquinas virtuales utilizando Vagrant



4. Introducción

La computación de alto rendimiento o HPC de sus siglas en ingles *High Performance Computing* es un campo especial e importante de las ciencias de la computación que implica resolver problemas complejos que demandan grandes capacidades de recursos computacionales⁽¹⁾(*como ejemplo un gran de numero de procesadores, alta cantidad de Memoria, alta capacidad de almacenamiento*), es una de las mejores maneras de computar algoritmos en diferentes campos de la ciencia como la física, medicina, bioinformatica, mecánica, química u otros problemas comunes de Big Data. Estas tareas y procesos son grandes para ser realizados en una sola maquina, por ello se hace necesario construir lo que se denomina un *Clúster* de computadores definido como un grupo de múltiples nodos o computadores que trabajan de manera conjunta y distribuida para cumplir con una meta. Debido a la naturaleza de múltiples computadores conectados entre si, existen diferentes modelos de computación paralela y distribuida que han sido desarrollados e implementados por la comunidad científica para la coordinación de los computadores pertenecientes a un Clúster, de manera que su uso sea eficiente y efectivo. Un modelo muy utilizado y popular es el *paso de mensajes*, una técnica empleada en programación concurrente para aportar sincronización entre procesos y permitir la exclusión mutua, su diseño esta pensando para algoritmos y programas que exploten la existencia de múltiples procesadores en un Cluster. MPI es el estándar de facto preferido por la comunidad HPC a la hora de realizar implementaciones de paso de mensajes.

Tradicionalmente, configurar un cluster de computadores, por ejemplo, un cluster MPI, es una tarea desafiante que requiere que estudiantes, aficionados, novatos y avanzados pasen un tiempo configurando los diferentes elementos del sistema y la red. Sin embargo, en los últimos años el avance del Cloud Computing y la implementación de metodologías ágiles que impulsan culturas como la DevOps han hecho que la automatización en todas las etapas sea masiva, gracias a esto diferentes proyectos de software libre y de código abierto han visto la luz, como Vagrant, una herramienta de código abierto que permite la creación y configuración de ambientes virtuales portables y ligeros. En este proyecto se presenta una implementación utilizando Vagrant para realizar la construcción automática de un cluster MPI multinucleo y multimaquina.



5. Planteamiento del Problema

Descripción

La infraestructura como código o IaC (Infrastructure as Code) es uno de los pilares fundamentales de la cultura DevOps, la IaC hace referencia a la práctica de poder administrar, aprovisionar, actualizar, monitorear y configurar la infraestructura TI de las empresas, se utilizan scripts en diferentes lenguajes de programación en lugar de configurar las máquinas de manera manual. La IaC trata todas las configuraciones de la infraestructura TI como código permite a las máquinas virtuales gestionarse de manera programada y automática, lo que elimina la necesidad de realizar configuraciones manuales por parte del personal a cargo de componentes individuales de hardware. Esto hace que la infraestructura sea muy moldeable, es decir, escalable y replicable siendo agnóstico a el hardware, proveedor de servidor cloud, además se utilizan las mismas practicas implementadas para la gestión y versionamiento del código de aplicaciones de software.

La implementación y construcción de un cluster MPI implica tiempo y conocimiento de todos los elementos que lo componen, es necesario conocer un completo detalle de como realizar la instalación de cada uno de los componentes del cluster, algunas tareas que son repetitivas se puede automatizar implementando IaC ejemplo de ello es la creación de usuarios con sus niveles de privilegios, las configuraciones de seguridad, la configuración de servicios, las configuraciones de sistemas de archivos compartidos entre otras tareas repetitivas. Es necesario en un cluster MPI asegurar la inmutabilidad y homogeneidad en las configuraciones presentes en los nodos que lo componen.

6. Objetivo General

Diseñar e implementar un conjunto de programas para el aprovisionamiento automático de un cluster computacional sobre máquinas virtuales, que soporten procesos con MPI.



7. Objetivos Específicos

- Obtener la lista de requerimientos, limitaciones, especificaciones para el diseño del aprovisionamiento automático utilizando Vagrant.
- Desarrollar e implementar un sistema de archivos distribuido en el Clúster MPI.
- Realizar pruebas de ejecución y funcionamiento del cluster.
- Realizar pruebas de ejecución y desempeño multimaquina en el Clúster MPI compuesto por maquinas virtuales.

8. Metodología

Fases de la Metodología

1. Fase 1 : Recolección de requerimientos y análisis de la información
2. Fase 2 : Diseño e implementación de despliegues automáticos
3. Fase 3 : Diseño e implementación de pruebas de despliegue
4. Fase 4 : Documentación y conclusiones



3. Pruebas

Hipótesis Será posible construir un conjunto de programas que realicen el aprovisionamiento e instalación automática de un cluster MPI y sus nodos ?, Será que este entregable ayudara a los estudiantes del curso de HPC a entender como instalar un cluster MPI?

Variables

- Costos de Hardware
- Tiempo de aprovisionamiento
- Performance del cluster dado que son maquinas virtuales.

Diseño Metodológico

- El enfoque y método seguido para la ejecución de este proyecto es “*aprender haciéndolo*”, parte desde el aprendizaje de las características de Vagrant y la programación de los scripts en Ruby hasta hasta la parametrización y configuración de cada una de las tecnologías y software implicados para dejar el cluster de MPI funcionando, todo ello llevado a cabo desde un punto de vista práctico.

- Tipo de investigación

Basados en los puntos anteriores se considera que éste es un Proyecto de aplicación; ya que se pretende que por medio de la obtención de nuevos conocimientos por medio de investigación, se obtenga un resultado o entregable después de la culminación.

En esta investigación se utilizará un enfoque cuantitativo. Además de que se empleará el método científico.

- Estrategias

Se llevará a cabo un conjunto de pruebas sobre el sistema, para establecer si cumple con las estimaciones del proyecto y si es funcional.



9. Cronograma

A continuación se describe un esquema de las actividades que se desarrollarán durante el proyecto, cada una con su respectiva duración:

Cronograma de Actividades						
Actividad / Tiempo(Meses)	1	2	3	4	5	6
Consultas e investigación sobre MPI, cluster y computación de alto rendimiento.	X					
Investigación y Consulta de Infraestructura como Código, entendimiento del funcionamiento interno de Vagrant	X					
Pruebas de aprovisionamiento automático de diferentes elementos de infraestructura con Vagrant	X					
Implementación de algoritmos y scripts para la instalación de utilidades		X				
Pruebas de funcionamiento comunicación master-nodos			X			
Ejecución de algoritmo en cluster				X		
Elaboración de informe final					X	
Entrega						X

10. Presupuesto y Fuentes de Financiación

Recursos Humanos

- Director de proyecto.

Recursos Físicos

Laboratorio Sirius CIDT

Cluster Lovelace, Bloque L.

Computador Tesista, Servidores Físicos a disposición tesista.

Locaciones Grupo de Investigación Sirius.

Ingeniería de Sistemas y Computación. Universidad Tecnológica de Pereira.



A continuación un valor estimado para los recursos considerados en el proyecto:

Recurso Humano	Costo(Peso Colombiano) / Hora	Número Horas	Total	Fuente Financiadora
Investigador (Tesisista)	6.500	200	1.300.000	Universidad
Director / Asesor	38.000	80	3.040.000	Universidad
Compra o Alquiler de Maquinaria y Equipos	Costo(Peso Colombiano)	Cantidad	Total	Fuente Financiadora
Computador Thinkpad X1 6TH, procesador i7 8GEN, 1TB DD y 16GB RAM	8.100.000	1	8.100.000	Tesisista
Fungibles	Costo(Peso Colombiano)	Cantidad	Total	Fuente Financiadora
Papelería y otros	50.000		50.000	Tesisista
Servicios públicos	130.000		130.000	Tesisista
Costo Total del Proyecto			12.620.000	



11. Bibliografía

- Resource selection and allocation for dynamic adaptive computing in heterogeneous clusters, John U. Duselis The Donald Bren School of Information and Computer Science, University of California, Irvine, Irvine, CA 92697; E. Enrique Cauich ; Richert K. Wang ; Isaac D. Scherson
- Leveraging a Cluster-Booster Architecture for Brain-Scale Simulations, Conference Paper, Pramod Kumbhar, Michael Hines, Aleksandr Ovcharenko, Damian A. Mallon, James King, Florentino Sainz, Felix Schürmann, Fabien Delalondre
- Infrastructure As Code, Kief Morris
- Vagrant: Up and Running: Create and Manage Virtualized Development Environments
- Infrastructure as Code: Managing Servers in the Cloud
- Site Reliability Engineering: How Google Runs Production Systems