

PROPUESTA DE LABORATORIO HPC APLICANDO SOFTWARE LIBRE PARA DISEÑO DE UN CLUSTER

HPC LABORATORY PROPOSAL APPLYING FREE SOFTWARE FOR DESIGNING A CLUSTER

Noé Ramón Rosales Morales¹, Juan Manuel Bernal Ontiveros², Marisela Palacios Reyes³,
Margarita Bailón Estrada⁴, Francisco Zorrilla Briones⁵

¹Maestría en Software Libre, ²Maestría en Ciencias en Ingeniería Industrial, ³Maestría en Ciencias en Ciencias de la Computación, ⁴Ingeniería en Sistemas Computacionales en Programación, ⁵Maestría en Ingeniería Industrial, ⁶Estudiante de la Carrera en Ingeniería en Sistemas Computacionales

^{1,2,3,4,5,6}Instituto Tecnológico de Ciudad Juárez, Departamento de Sistemas y Computación, Avenida Tecnológico No. 1340 Fraccionamiento El Crucero Código Postal 32500, Teléfono (656) 6882500, Ciudad Juárez, Chihuahua México

¹nrosales@itcj.edu.mx, ²jbernal@itcj.edu.mx, ³mpalacios@itcj.edu.mx, ⁴mbailon@itcj.edu.mx, ⁵fzorrilla@itcj.edu.mx,

Resumen—El HPC o Computo de Alto Desempeño no es nuevo en México ya que existen 16 centros de Supercomputo registrados en el CONACYT, no son muchos sin embargo 9 de estos se concentran al interior de la República en estados como Puebla, Guadalajara, San Luis Potosí, en el Estado de México, y la ciudad de México también dos están al Norte-Oeste en Baja California, otros dos están en Yucatán, hay uno en Chiapas y otro en Sonora, se dedican a diversas tareas de aplicación científica, académica, y algunos inclusive comercial[3]. Pudiera haber más como el CIMAV ubicado en Chihuahua capital sin embargo no estaba en la lista al momento de esta investigación.

Supercomputo como el termino lo indica permite realizar cálculos muy sofisticados y complejos que requieren procesamiento en paralelo, algo que las computadoras PC no pueden realizar.

El siguiente trabajo propone la necesidad que tiene el ITCJ en crear un sistema de HPC mediante un “clúster” aplicando software libre para ser utilizado en la capacitación de profesores en estas áreas y en el desarrollo de prácticas de laboratorio para diversas materias de especialidad de las carreras de Ingeniería en Sistemas Computacionales como en Tics, ya que actualmente no se cuenta con dicho laboratorio de infraestructura física se realizan las practicas mediante la creación de un “clústerbeuwolf” la llamada supercomputadora de los pobres[7] y en este caso con maquinas virtuales y se lograr cumplir con los programas académicos de las materias de especialidad de dichas carreras. Sin embargo

el ITCJ pierde la oportunidad de ser un NODO de HPC para apoyar a maestros investigadores que tienen proyectos donde necesitan aplicar HPC, como también pierde la oportunidad de vinculación y colaboración en dar solución a problemas que las empresas pudieran presentar. A continuación se profundizara más en esto.

Palabras Clave: HPC, Computo de Alto Desempeño, Clúster, Computo Avanzado.

Abstract – HPC or High Performance Computing is not new in Mexico since there are 16 Supercomputing centers registered in CONACYT, these aren't many, however 9 of these are mostly in the middle of the Republic in states such as Puebla, Guadalajara, San Luis Potosi, Mexico State, and Mexico City, two are also to the North-West in Baja California, another two in Yucatan, there is one in Chiapas and another in Sonora, they are dedicated to various scientific and academic tasks, and some including commercial[3]. There could be more, like the CIMAV located in the capital of Chihuahua; however it was not on the list at the time of this investigation.

Supercomputing as the term indicates allows very sophisticated and complex computations that require parallel processing, something that PC computers cannot do.

The following work proposes the need for ITCJ to create an HPC system with a "cluster" applying free software to be used in the training of professors in these areas and in the development of laboratory practices for various subjects for the major in Computer Systems Engineering and TICs, since there is currently no such physical infrastructure laboratory, the practices are carried out with the creation of a "Beowulf cluster" the so-called poor man's supercomputer[7] and in this case with virtual machines so it is possible to comply with the academic programs of the major's courses for said careers. However ITCJ loses the opportunity to be an HPC NODE and support other professors doing research or have projects where they need to apply HPC, as well as the opportunity to link and collaborate in solving problems that companies may present. This will be further explored.

Key words – HPC, High Performance Computing, BeowulfCluster, Advance Computing.

INTRODUCCIÓN

El (HPC) "High Performance Computing" es decir el Computo de Alto desempeño es la nueva infraestructura tecnológica que tanto la industria privada como el

gobierno consideran una necesidad por que sobre esta están dependiendo muchas otras propuestas comerciales, industriales y políticas nacionales como lo son abastecimiento de energía limpia y renovable, los sistemas de logística para distribución de bienes en tiempo real, la toma de decisiones automatizadas, así como la telemedicina en zonas rurales y para atender a estas nuevas áreas se requiere que las IES en particular del sistema del TecNM cuenten con los laboratorios y la capacitación continua de los profesores que se encargaran de formar a los próximos profesionistas que darán soporte a estas nuevas áreas de trabajo.

Por parte del Estado, la (HCTI) iniciativa de la Ley General en materia de Humanidades, Ciencias, Tecnologías e Innovación "reconoce el derecho constitucional de toda persona a gozar de los beneficios del desarrollo científico y la innovación tecnológica como fundamento de la política pública en materia de HCTI" y reconoce su potencial para garantizar otros derechos humanos[2]. A la vez consideran necesarios la implementación de sistemas de visualización de geoposición aplicando grandes volúmenes de datos o Big Data por dependencias como PEMEX, CFE, IMSS, ISSSTE, SEDATU, INEGI y otras del gobierno mexicano requieren de sistemas innovadores para la visualización de datos geoposicionados para detectar tendencias en los datos que mejoren la toma de decisiones en la alta dirección en dichas dependencias [3].

"En el sistema socio-estructural de las IES emerge la formación docente como uno de los principales vehículos para su transformación, pues en ésta recaen las esperanzas para formar nuevo talento..."[10]

La (IESP) International Exascale Software Project, el organismo Internacional de Proyectos de Software para Computo de muy Alto Desempeño solicita que se motive y facilite colaboración en educación y entrenamiento, debido a la magnitud de los cambios en los modelos de programación, en la infraestructura y las herramientas que trae consigo estas nuevas arquitecturas de muy alto rendimiento en el área de la educación y capacitación [9] aplicando software Open Source [11].

El presente trabajo es precisamente una propuesta de lo que la Academia de Sistemas Computacionales del Instituto Tecnológico de Ciudad Juárez requiere para encarar los demandas del (HPC) Computo de Alto Desempeño planteados por parte de gobierno federal y la iniciativa privada. Diseñar un "Beowulf cluster" para capacitación y entrenamiento de docentes y estudiantes del área para lograr resolver algoritmos del tipo HPC. Un "Beowulf cluster" es una agrupación de computadoras interconectadas en una red donde los nodos (cada computadora) y la red son dedicados al grupo, es decir al "cluster"[14], los nodos pueden ser equipo común como (PC) Computadoras Personales, sin embargo si se emplean maquinas "work stations" estaciones de trabajo o servidores el cluster tendrá la capacidad de ofrecer un desempeño en los cálculos más potencializado, en cuestión de licenciamiento de software para que sea

asequible se implementa con software libre o de código abierto

Con el Objetivo de presentar las necesidades de infraestructura tecnológica para fomentar la Investigación, el Desarrollo y la Innovación tecnológica con aplicación de (HPC) Computo de Alto Desempeño, beneficiando así al sistema de TecNM, fortaleciendo su comunidad tecnológica los docentes, y estudiantes al contar con una plataforma de HPC que sea capaz de satisfacer las necesidades de capacitación de profesores en las nuevas tecnologías como implementación de la Inteligencia Artificial en Machine Learning, apoyado en el Big Data, y formando el talento de profesionistas que podrán satisfacer las necesidades de la iniciativa privada comercial, industrial así como a las diferentes dependencias estatales, resolviendo las problemáticas nacionales planteados en el Plan Nacional de Desarrollo

Descripción del problema

La nueva industria 4.0, el gobierno federal y la nueva evolución de la Educación 4.0 están demandando implementación y enseñanza de nuevas tecnologías emergentes y avanzadas sin embargo el sistema del TecNM, y el Instituto Tecnológico de Ciudad Juárez carecen del presupuesto necesario para contar con un sistema de HPC "High Performance Computing" o Computo de Alto Desempeño, que apoye tanto a docentes como a los estudiantes a adquirir capacitación a través de la experiencia práctica necesaria para enfrentar los nuevo retos tecnológicos en las áreas de Big Data, Inteligencia Artificial y Computo de Alto Desempeño ya que no se cuenta con las infraestructura y capacitación necesarias.

Definición del Problema

Falta de la infraestructuratecnológica avanzada en la forma de servidores, "switches" de comunicación rápida para la creación de un "clúster" de computo para el análisis de problemas complejos impide la capacitación continua de profesores así como el aprendizaje basado en proyectos o de experiencia aplicada como plantea la nueva Educación 4.0.

Preguntas de la investigación

El presente trabajo de investigación busca ofrecer una propuesta para responder a la siguiente pregunta ¿Cómo capacitar a los docentes y estudiantes del Instituto Tecnológico de Cd. Juárez en la tecnologías de HPC, sino se cuentan con equipos de Computo de Alto Desempeño?

Otras preguntas a resolver

¿Como crear un clúster de HPC?

¿Que tipo de equipos de computo se requieren para crear el clúster?

¿Cuantas computadoras se pueden conectar para sumar su poder computo, antes de que se empiece a degradar el desempeño obtenido?

¿Que tipo de software se utiliza para interconectar los equipos de computo?

¿Que tipo de aplicaciones se le da a un clúster de HPC?

Hipótesis / Supuesto

El diseño de un laboratorio basado en Software Libre y Código Abierto aplicando un sistema gestor de máquinas virtuales permitirá que los estudiantes adquieran conocimiento práctico para prepararlos para los nuevos retos de HPC que presentan tanto el gobierno como la iniciativa privada.

Antecedentes

La cuarta revolución Industrial está en curso, la llamada Industria 4.0 y está demandando recursos humanos calificados en las últimas tecnologías emergentes como en Inteligencia Artificial, Big Data, Computo de Alto Rendimiento, por ende la respuesta es la actual evolución de la Educación en su versión 4.0 que debe responder a estas demandas a la par, si la nueva forma industrial requiere líneas de producción flexibles entonces se requiere un aprendizaje a la medida o personalizado, automatización de tareas, se requiere una enseñanza en el diseño de sistemas de tutores inteligentes, la nueva industria necesita una producción a la medida, de igual forma la nueva política de educación debe ser flexible, en función de los intereses y necesidades de cada alumno a su ritmo de aprendizaje que sea independiente de su edad y curso, si la industria necesita analítica de datos, se requiere una analítica de aprendizaje y si la industria necesita una larga vida de aprendizaje se requiere una capacitación continua de los profesores, estos son solo algunos de los ejemplos que propuso la doctora Ana Franzoni, del ITAM en el Foro Interinstitucional de Educación 4.0 llevado a cabo el 17 de junio del 2019. La Dra. Franzoni argumenta la necesidad inmediata de aumentar la colaboración entre las instituciones de Educación Superior y la Industria, para desarrollar una metodología basada en aprender de forma práctica y ver la tecnología como una herramienta de transformación del aprendizaje para formar las competencias de negocio digital, innovación, creatividad, comunicación, trabajo en equipo y liderazgo todo con el propósito de satisfacer las necesidades de la Industria 4.0 [8].

La actual Constitución Mexicana en la reforma del 15 de Mayo del 2019 el Artículo 3o se modificó en su fracción V para colocar el desarrollo de la ciencia, y la innovación tecnológica como un derecho que toda persona mexicana deberá gozar indicando además que el Estado apoyara la investigación científica, humanística y tecnológica y que garantiza acceso abierto al conocimiento que de allí emane, aun más que proveerá los recursos, y estímulos suficientes así como también alentar el fortalecimiento y la difusión de nuestra cultura [1]. Adicionalmente se crea la iniciativa de la Ley General en materia de Humanidades, Ciencias, Tecnologías e Innovación (HCTI) el pasado 13 de diciembre del 2022 donde se reafirma la garantía del derecho humano a la ciencia, que fortalece el ahora CONAHCYT Consejo Nacional de Humanidades, Ciencias y Tecnología que en su estructura busca promover la libertad de investigación, el acceso universal al conocimiento científico, a otorgar financiamiento público a través de programas

presupuestarios, para el apoyo y fomento de actividades HCTI que deberán canalizarse de preferencia de manera directa a los becarios, humanistas, científicos, tecnólogos e innovadores, adicionalmente implementar un Sistema Nacional de Posgrados para facilitar y promover la creación y consolidación de programas de posgrado orientados a la investigación, en todas las ciencias, humanidades, para profesionalizar en temas de interés público nacional [2]. La ley como tal aun está siendo revisada de forma conjunta por las cámaras de diputados y senadores, sin embargo en su forma de iniciativa de ley se establece la política pública de “contribuir al avance del conocimiento universal, al fortalecimiento de la soberanía nacional, al desarrollo integral y sostenible del país, al bienestar de las generaciones presentes y futuras, a la preservación, restauración, protección y mejoramiento del ambiente” por mencionar algunos de los objetivos del Estado mexicano.

Así mismo en su título cuarto del fomento y financiamiento entre otros objetivos se propone destinar recursos para operar Programas Nacionales Estratégicos capaces de diagnosticar, y proponer soluciones a las problemáticas nacionales específicas en relación a salud, agua, educación, cultura, vivienda, soberanía alimentaria, agentes tóxicos y contaminantes, seguridad, sistemas socio-ecológicos, energía, cambio climático y la transición energética por mencionar los más relevantes y que están de inmersos en la agenda 2030.

Aun más en Consejo Nacional en materia de HCTI, que incluye el Sistema Nacional de Información, tiene contemplado crear ecosistemas nacionales informáticos, de repositorios de información.

Desde el 2018 existe una política de gobierno para impulsar la investigación pasando de solo investigación teórica a lograr un equilibrio en el desarrollo tecnológico con impacto social. Desde la actual administración federal las distintas dependencias como “Salud, Energía, Hidrocarburos, Educación, Atención Ciudadana, Geografía y Estadística están haciendo modificaciones a sus entornos legales para incorporar procesos para captura, procesamiento y administración de grandes volúmenes de datos. Esto es recabando, procesando y analizando datos para contar con bases de datos para la toma de decisiones más acertadas en la solución a las problemáticas de las citadas áreas.

El estado del (High Performance Computing) HPC por sus siglas en inglés, es decir el Computo de Alto Desempeño o Rendimiento aquí en México es y ha sido una necesidad desde años, solo que antes únicamente algunas empresas y centros de investigación contaban con este tipo de laboratorios y sus hallazgos solo han sido para favorecer a los intereses de algunos cuantos, mientras que ahora bajo la nueva óptica de libre acceso al conocimiento y la tecnología se busca llevar las nuevas plataformas de computo avanzado o de alto desempeño a más instituciones académicas para que docentes, estudiantes e investigadores tengan la posibilidad de acceder a estas infraestructuras tecnológicas y ponerlas a

resolver las problemáticas nacionales para otorgar los beneficios de las nuevas investigaciones a los sectores de la sociedad más vulnerables y necesitados, los pobres. Según Castañeda [3] “la actividad de HPC en México se ha concentrado en el fortalecimiento de las capacidades instaladas de hardware y software en las instituciones educativas y centros de investigación bajo la cobertura del Programa de Laboratorios Nacionales CONACYT”.

Así como el gobierno federal también los gobiernos estatales están haciendo su esfuerzo para apoyar la ciencia, la tecnología y en especial a la ciencia de datos, la inteligencia artificial, la automatización de procesos administrativos por medio de software como también buscar formas de implementar la nueva industria 4.0 en temas de interés del Estado. Algunos ejemplos de éxito son los gobiernos de Jalisco, Guanajuato, Baja California, Nuevo León y Veracruz. En dichos estados se han aplicado programas para capacitar en las nuevas herramientas y tecnológicas que utilizan el HPC.

El sector empresarial impulsa el Cómputo de Alto Desempeño, para la industria 4.0, en procesos de cadena de suministros y logística así como en sistemas predictivos basados en ciencia de datos e inteligencia artificial. La Cámara Nacional de la Industria Electrónica, de Telecomunicaciones y Tecnologías de la Información, en su reporte del 2020, indica que existen una necesidad de más de 900,000 empleos en México relacionados con la industria de las (TIC) Tecnologías de Información y Comunicación requiriéndose más de 550,000 profesionistas en el ámbito de hardware y algo más de 350,000 en el ámbito de software, esta es otra de las muchas consecuencias de la pandemia por COVID-19 del 2020. “la pandemia del Covid-19, ha traído de manera mundial un aceleramiento en el uso e implementación de las tecnologías de la información y comunicación, tanto de forma personal como dentro de las organizaciones... hay grandes retos y oportunidades con este acontecimiento, este brinco cuántico hacia la transformación digital apoyará a nuestros sectores a ser claves en el fortalecimiento y crecimiento de otras industrias” [4].

El CIMAV Centro de Investigación de Materiales Avanzados, cuya sede está en la ciudad de Chihuahua cuenta con un sistema HPC Cómputo de Alto Desempeño llamado Prometeo, es un “clúster” tipo Beuwolf que actualmente consta con 99 nodos de cómputo operando para que investigadores de diversas instituciones públicas y privadas en el estado realicen investigación, según comenta el encargado del área de Cómputo Científico, el maestro Rodrigo Domínguez, quien cuenta con 14 años apoyando a alumnos e investigadores en aplicar el sistema Prometeo de HPC en aplicaciones de modelado computacional tanto para física y química de materiales en diversas y emergentes áreas de la ciencia como Cogeneración termosolar, Nanoquímica Médica Computacional, Nanobiotecnología como también para

apoyar en la solución de problemas relacionados con la contaminación ambiental, así como en proyectos de la industria automotriz y aeroespacial, a la vez ofrecen estudios de Posgrado en dichas ramas a profesionistas interesados [5].

Siendo que existe mucho interés tanto por parte del gobierno, empresas y demás instituciones en las tecnologías del Cómputo de Alto Desempeño es imperativo que el Sistema de Tecnológico Nacional de México “TecNM” a través de sus campus los Institutos Tecnológico que cuenten con carreras tecnológicas como Ingeniería en Sistemas Computacionales, TICS, y afines formen a sus estudiantes en estos conocimientos tanto teóricos como prácticos, para que sean capaces de afrontar los retos y la problemática de la actual y futura sociedad buscando las oportunidades de vinculación y empleabilidad a través de programas como estadías, prácticas, o residencias profesionales en dichas instituciones públicas así como en la iniciativa privada.

La formación de estos profesionistas competentes y capaces el sistema del TecNM lo hace manteniendo los programas educativos y sus contenidos temáticos vigentes a través de reuniones de seguimiento curricular que incorporan los nuevos conocimientos emergentes y estará la vanguardia en las tendencias siempre cambiantes del área de Tics, el Tecnológico de Cd. Juárez como parte de esta enorme red educativa no es la excepción asistiendo a dichas reuniones para uniformizar, replantear y consolidar los planes de estudio que han de impartirse en este caso tanto en la carrera de Ingeniería de Sistemas Computacionales como en la carrera de Ingeniería de Tecnologías de la Información y Comunicación y sus Academias representativas para ofrecer contenidos actualizados para preparar a los estudiantes de dichas carreras para los retos de HPC Cómputo de Alto Desempeño que gobierno e industria demanda.

La realidad es que la mera revisión de los planes de estudio y la definición de guías para las prácticas no son suficientes sino se cuenta adicionalmente con la infraestructura necesaria para llevar a cabo las prácticas en cuyas materias se cubren los temas de Cómputo Avanzado o el HPC. A la par es necesaria la capacitación y actualización del personal docente que impartirá esta curricular.

No es suficiente que estos temas sean tratados de forma teórica sino que es imprescindible acercarse a los estudiantes a contar con experiencia que se obtiene de la práctica sobre conocer, instalar, configurar, administrar equipos para Cómputo de Alto Desempeño como servidores de doble microprocesador Xeon, servidores con arreglos de RAID que sumen Terabytes, configurar conectividad para sistemas distribuidos, sin embargo la adquisición de equipo con estas especificaciones está fuera del presupuesto, ya que según comenta el maestro Domínguez cuando el CIMAV instaló el “clúster” Prometeo, este empezó con una inversión inicial de

\$1,200,000 pesos que se invirtieron en nueve nodos de computo un servidor maestro y 8 nodos esclavos, para trabajar de forma conjunta en lo que se conoce como un “Beuwolf clúster” (R. Domínguez, comunicación personal, 21 de abril del 2023). A un “Beuwolf clúster” se le conoce también como la súper computadora de los pobres, porque contrario a las supercomputadoras conocidas de la lista top500.org que pueden costar billones de dólares y realizar un quintillón de operaciones de punto flotante por segundo como el sistema Frontera que desde hace un año ocupa el lugar #1 de la lista de las 500 más potentes supercomputadoras [6].

Un sistema basado en un “clúster Beuwolf” se puede crear con diferentes tipos de equipos que no necesariamente deben ser los más nuevos o veloces, sino al contrario tiene la flexibilidad de permitir emplear equipo de computo ordinario y mediante la interconectividad y aplicación de software libre como una distribución GNU/Linux da la oportunidad de crear una computadora más potente por trabajar todas las maquinas en paralelo[7][13], y este es el tipo de experiencia práctica que adquieren los estudiantes de las materias como Taller de Sistemas Operativos, Tópicos Avanzados de Virtualización, Tópicos Selectos de Seguridad Informática, Tópicos Avanzados de Bases de datos.

Desarrollo

Se propone la creación de un “clústerbeowulf” con cuatro equipos de tipo servidor blade con doble microprocesador de tecnología XEON de mínimo seis núcleos de doble banco de memoria RAM de hasta 256GB con múltiplesbahías de disco duro para adaptar a configuración RAID, conectados estos vía un Smith de rápida transferencia infiniband, sin embargo como se menciono por la actual carencia de infraestructura se implementa un sistema de virtualización llamado VirtualBox actualmente se encuentra en su versión 7. Este gestor de maquina virtuales requiere equipo de computo que en su microprocesador posea la tecnología de VT-X para el manejo de virtualización, por ende se instala de momento en equipo tipo PC con un microprocesador Intel iCore 5 con el cual para efecto de la practica aunque sea virtualizado la creación del clúster donde se configura un nodo maestro que controla la orquestación del sistema y el acceso a los nodos esclavos.

El sistema operativo empleados el GNU/Linux distribución Ubuntu Server versión 16.04, se emplea esta distribución en vez de la más reciente la 23.04 debido a que se encuentra mucha documentación tanto escrita como en los foros de internet, si ocurriesen problemas de algún tipo existe una alta confiabilidad en que ya fue resuelto y la solución es rápidamente aplicable, sin embargo se puede emplear cualquier distribución incluidas las basadas en Red Hat como Fedora o OpenSuse.

Como software de conectividad para formar el clúster se emplea se requiere un programa especial de paso de mensajes entre sistemas operativos (MPI) “Message Passing Interface”, el primer estándar se logro crear en

1994, con el tiempo este se publicaría y se forma un estándar de dominio público e independiente de plataforma tanto en lenguajes como Fortran y C para crear programas de computo en paralelo y que sea aprovechado por sistemas tipos clúster.

El montado del clústerconsistirá entonces en:

La Red en este caso el gestor de virtual box permite crear una red interna entre los nodos conectados,

Un nodo maestro, o principal,

Una serie de nodos esclavos, debido a que esta implementación la cantidad de nodos a virtualizar depende de la memoria RAM disponible en la maquina anfitriona es decir la maquina principal, para no gastar mucha RAM de los 4 GB disponibles se les asigna a los nodos esclavos 512 MB de RAM y 1024 MB al nodo maestro.

Como puerto de salida, para este sistema se implementa el mismo nodo maestro como puerto de salida o “Gateway”.

Todos los nodos incluyendo el nodo maestro deberán tener instalado el siguiente software aparte de su GNU/Linux Ubuntu:

El NFS, “Network File Sistema”, para formato de la partición

el SSH Secure Shell, para la comunicación remota

El sistema de MPI “Message Passing Interface” y el MPICH. El MPI es un estándar para el intercambio de mensajes que ofrece portabilidad hacia la arquitectura de computo en paralelo, este estándar permite definir la sintaxis y semántica de las rutinas para las bibliotecas o librerías de funciones entre lenguajes como C, C++ y Fortran [15].

De la configuración:

Siguiente se configuran los nodos, se conectan mediante la asignación de una IP fija en red de área local 192.168.1.100 y así sucesivante los otros nodo esclavo.

Se debe crear un usuario para cada nodo que se conecte al clúster, para este tipo de usuario es importante su configuración manual por tal el número de identificación debe ser menor a 1000 para que no aparezca en la pantalla de login del sistema. Para cada uno de los nodos emplear el mismo número de ID para que el sistema del MPI.

El firewall deberá ser reconfigurado para permitir acceso a una subredespecífica.

De la instalación del Sistema de Archivos

En el nodo maestro se instala en una porción del disco duro que será accedida por todos los nodos que forman el clúster

De la autenticación automatizada entre los nodos

Para que el clúster funcione el nodo maestro debe comunicarse con los demás nodos y viceversa para esto se emplea ssh que permite realizar la conectividad, se requiere generar el par de llaves pública y privada.

De la preparación del gestor de procesos

En esta parte del proceso se instala el MPICH, que se emplea para la gestión de los procesos en paralelo que estará ejecutando el clúster, se debe configurar el hydra, una aplicación de gestor de usuarios del MPICH. El MPI, se puede emplear como una biblioteca para el paso de mensajes que implementa la especificación MPICH, la cual se usa tanto para crear, mantener y probar algoritmos complejos basados en el “cluster”[16]. Con este se pueden desarrollar aplicaciones que corran en sistemas conectados. Permite la entrega inmediata de mensajes entre los procesos del “cluster”.

Para correr trabajos en el clúster

El sistema de MPICH2 es un paquete de aplicaciones de ejemplo para probar el clúster para correrlos se debe contar con acceso a internet para descargarlos. Contiene una herramienta cpi que es útil para mostrar que nodos están corriendo que aplicación.

El sistema ya está listo para correr cualquier tipo de software que requiera procesamiento en paralelo en el clúster. Cabe mencionar que se favorece la implementación de software libre y de Código Abierto sobre el tipo propietario debido a que el primero su licenciamiento es asequible para todos los niveles de usuarios, mientras que el propietario limita al usuario a la adquisición del licenciamiento pagado por periodos específicos de tiempo de otra forma no se permite su uso o lo limita a ciertas características básicas [12].

Observaciones

Los estudiantes de la materia de Tópicos Avanzados de Virtualización se verán beneficiados de este tipo de configuración inclusive si es mediante máquinas virtuales porque tendrán la oportunidad de instalar, configurar y administrar el clúster de a primera mano, podrán ejecutar programas de algoritmos sofisticados que requieren múltiples microprocesadores para que sus cálculos sean realizados. En particular en el tema de clustering podrán comprobar las ventajas y desventajas de utilizar virtualización contra emplear equipos físicos para crear el clúster al medir desempeño. Tendrán oportunidad de comparar software libre y de open source en algunos casos con software propietario y hasta con software como servicio en la nube, sin embargo para los servicios en la nube como el Amazon EC2 cloud solicita disponer de tarjeta de crédito o débito para registro y pruebas lo cual puede ser limitante para la mayoría de estudiantes ya que la población estudiantil no pertenecen a la estratificación socio-económica alta para disponer de recursos de esta forma.

Para la materia de Tópicos Selectos de Seguridad informática al emplear un clúster pueden probar las destrezas de ataque y contra-ataque en pruebas de penetración a sistemas red organizacionales simulados, de nueva forma aquí se beneficiarían más si se contara con un clúster elaborado con equipo físico como se describe anteriormente, esto debido al poder de cómputo requerido para dichas pruebas las cuales pueden

requerirse horas o días en un sistema de cómputo PC, pero en un HPC formado por un clúster físico se puede reducir en un 50% el tiempo. En materias como la Tópicos Avanzados de Base de Datos se pueden simular bases de datos distribuidas, y se pueden cargar bases de datos de gran volumen para simular Big Data y en caso de requerir toma de decisiones se requieren conjuntos de datos grandes para su análisis mediante Machine Learning aplicando ciencia de datos, de nueva cuenta con un clúster virtualizado se pueden hacer pruebas sin embargo se necesita un clúster físico con las especificaciones mencionadas para lograr la demostración de aceleración de los cálculos en tiempo real, así como la reducción de tiempo requerido en su totalidad.

CONCLUSIONES

Las materias de Tópicos Avanzados de Virtualización, Tópicos Selectos de Seguridad informática, Tópicos Avanzados de Bases de Datos que son materias de las especialidades tanto de las carreras de Ingeniería en Sistemas Computacionales como de la Ingeniería en Tecnologías de la Información y Comunicaciones se han estado impartiendo tal cual en plataforma de virtualización desde que las especialidades fueron autorizadas, en dichos cursos se cubren de la mejor forma posible los saberes propios de los programas educativos sin embargo sin la infraestructura propia para HPC, no se puede apreciar las características del Cómputo de Alto Desempeño y sus beneficios en la aceleración de procesos. Tampoco es posible apoyar trabajos de investigación que realizan algunos profesores del área de Posgrado y se ven en la necesidad de rentar un servicio en la nube de HPC con 24 núcleos para procesamiento paralelo de algoritmos avanzados, cuyo enfoque en un caso particular es investigación básica en el área de materiales y sus propiedades fisicoquímicas bajo diferentes condiciones climáticas bien deberán participar en la convocatoria anual del Centro Nacional de Supercómputo auspiciado por el Instituto Potosino de Investigación Científica y Tecnológica, A.C. (IPICYT) sin embargo aunque este servicio si es gratuito, con frecuencia está siendo utilizado por múltiples investigadores y hay fila de utilización del sistema de supercómputo.

RECOMENDACIONES

El Instituto Tecnológico de Ciudad Juárez es una IES que cuenta con 60 años desde su fundación, es uno de los campus más grandes tanto en población estudiantil como planta docente del sistema. Como IES forma una población estudiantil que cae en la estratificación socio-económica media hacia baja, ofrecer tecnología de punta a nuestros estudiantes les permitirá a que ellos al colocarse en empresas con mejores sueldos por el talento que lograron cosechar en su estancia en el Instituto, esto les permitirá cerrar esas brechas socio-económicas en sus comunidades como en sus familias. Por otra parte contar con la infraestructura de HPC permitirá al Instituto a través de la vinculación y creación de convenios de colaboración trabajar con las empresas de la región para

apoyar en investigación y solución a problemas industriales complejos ya que no hay otras instituciones que ofrezcan soluciones de HPC sino hasta la capital del Estado, lo cual para las empresas como las maquiladoras suele ser inconveniente por costos de transporte o logística y terminan subarrendando servicios en la nube sin embargo el costo por hora suele ser disuasorio. A la vez se le puede apoyar a los 3 niveles de gobierno en problemas de investigación humanística, ciencia y tecnología que consideren necesarios y así el ITCJ a 60 años de su fundación sigue impulsando el estado del arte en educación tecnológica pertinente, estando actualizado, y siendo relevante en la región norte del país.

BIBLIOGRAFÍA

[1] Constitución Política de los Estados Unidos Mexicanos [Const.] Art. 3. 11 Nov. 2022 (México)

[2] Resumen Ejecutivo Iniciativa de Ley General en materia de Humanidades, Ciencia, Tecnología e Innovación [HCTI] 13 de Diciembre 2022 (México) https://conacyt.mx/wp-content/uploads/Ley%20HCTI/Iniciativa/Resumen_ejecutivo_HCTI.pdf

[3] Castañeda, D. (9 febrero 2023). Estado del HPC en México. CONECTA. Blog de la Red Nacional de Investigación y Educación en México CUDI. <https://conecta.blog/estado-del-hpc-en-mexico/#:~:text=En%20consecuencia%2C%20la%20actividad%20de,%2DCONACYT%2C%20marzo%202021>).

[4] Flores, L. (21 Julio 2020). Pandemia impulsó a empresas a la transformación digital: Canieti noreste. El economista. <https://www.eleconomista.com.mx/estados/Pandemia-impulso-a-empresas-a-la-transformacion-digital-Canietinoreste-20200721-0023.html>

[5] CIMAV. (s.f). Investigación. Centro de Investigación en Materiales Avanzados. <https://cimav.edu.mx/investigacion/>

[6] Peckham, O. (30 Mayo 2022). The Final Frontier: US has its first Exascale <https://www.hpcwire.com/2022/05/30/the-final-frontier-us-has-its-first-exascale-supercomputer/>

[7] Rajaraman, V. (feb. 2023). Beowulf Cluster-ApoorMan's Supercomputer. *Resonance*. 28(2). 249-250 <https://www.ias.ac.in/article/fulltext/reso/028/02/0247-0253>

[8] Mendoza, J. (2019). Hoy tenemos el poder de la transformación del proceso educativo. *Docencia Politécnica*, 1(1), 10-13. <https://www.ipn.mx/assets/files/innovacion/docs/docencia-politecnica/docencia-politecnica-1/Docencia-Politecnica-No-1.pdf>

[9] Dongarra, J. (2011). The International Exascale Software Project roadmap. *The International Journal of High Performance Computing Applications* 25(1) 3-60 <https://journals.sagepub.com/doi/epdf/10.1177/1094342010391989>

[10] Cruz, C. (2020). Formación docente para el desarrollo de habilidades digitales en el contexto de la educación 4.0. *Docencia Politécnica*, 1(4), 5-10 <https://www.ipn.mx/assets/files/innovacion/docs/docencia-politecnica/docencia-politecnica-4/Docencia-Politecnica-No-4.pdf>

[11] Williams, S. (2002). Free as in Freedom Richard Stallman's Crusade for Free Software, O'Reilly. <https://www.oreilly.com/openbook/freedom/ch01.html>

[12] Perens, B., Raymond, E. & Stallman, R. (1999). Open Sources: Voices from the Open Source Revolution. 1er Ed. O'Reilly. USA. <https://www.oreilly.com/openbook/opensources/book/>

[13] Kumar, S. (2009). Building a Beowulf Cluster in just 13 steps. Linuxdotcom <https://www.linux.com/training-tutorials/building-beowulf-cluster-just-13-steps/>

[14] Brown, R. (2004). Engineering a Beowulf-Style compute Custer. Duke University, Physics Department. USA. https://webhome.phy.duke.edu/~rgb/Beowulf/beowulf_book/beowulf_book/index.html

[15] Pereira, S. (2013). Building a simple Beowulf cluster with Ubuntu. group. University of York. https://www-users.york.ac.uk/~mjf5/pi_cluster/src/Building_a_simple_Beowulf_cluster.html

[16] INTEL. (s.f). Intel Mpi Library. Deliver flexible, efficient, and scalable cluster messaging. <https://www.intel.com/content/www/us/en/developer/tools/oneapi/mpi-library.html#gs.y92dh3>

ROLES DE CONTRIBUCIÓN

Rol	Autor (es)
Conceptualización	Noé Ramón Rosales Morales, Marisela Palacios Reyes
Redacción	Noé Ramón Rosales Morales, Margarita Bailon Estrada
Administración del Proyecto	Noé Ramón Rosales Morales
Curación de datos	Juan Manuel Bernal Ontiveros
Metodología	Margarita Bailón Estrada
Recursos	Marisela Palacios Reyes
Visualización	Francisco Zorrilla Briones
Software	Joseph Gabriel Gardner Martínez
Validación	Juan Manuel Bernal, Marisela Palacios Reyes



Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.