2.0 2.7

PROYECTO DE INVESTIGACIÓN CUANTITATIVO

TITULO

Análisis, Diseño y Despliegue de un Clúster de Servidores para alto procesamiento matemático

RESUMEN

En el presente proyecto, desarrollaremos el análisis, diseño y despliegue de la construcción de un cluster de servidores para alto procesamiento matemático con la finalidad de cubrir la necesidad de los proyectos de investigación en lo que concierne a ordenamiento de bases de datos, búsquedas, ejecución de algoritmos, simulación entre otros; la aplicación de un cluster reducirá los costos en infraestructura en sistemas de alto desempeño computacional y su sostenibilidad en el tiempo, haciendolo que sea incremental, podremos analizar los costos de operatividad y eficiencia en el consumo de energía.

Palabras claves

Cluster de servidores, eficiencia, operatividad, procesamiento matemático

Abstract

In this project, we will develop the analysis, design and deployment of the construction of a cluster of servers for high mathematical processing in order to cover the need for research projects regarding database ordering, searches, execution of algorithms, simulation among others; the application of a cluster will reduce infrastructure costs in high-performance computing systems and its sustainability over time, making it incremental, we will be able to analyze the costs of operation and efficiency in energy consumption

Keywords

Server cluster, efficiency, operability, mathematical processing

I. PLANTEAMIENTO DEL PROBLEMA.

Existe una necesidad de contar con sistemas de alta capacidad de procesamiento matemático por una constante y creciente demanda en simulaciones, cálculo matemático, búsquedas por similitud y en trabajos de computación gráfica, para ello hay equipo denominados de alto procesamiento computacional, pero son muy costosos, existiendo en el Perú solo 9 de estos equipos, ninguna de ellas en nuestra región; existe una alternativa de desarrollar un cluster de servidores medial el cual podremos sumar la capacidad de procesamiento de cada uno de estos servidores para convertirlo en una máquina que pueda hacer estos cálculos, sin embargo aun no contamos con la experiencia de un cluster en nuestra región para cubrir esta demanda.

II. JUSTIFICACIÓN DEL PROYECTO

No existe en nuestra universidad y en la Región un equipo de alta capacidad de procesamiento matemático que ayude en el procesamiento de información de las investigaciones, en la simulaciones y procesamiento gráfico, existiendo la necesidad de analizar, diseñar y construir un equipo (cluster) con este perfil, asi como definir el tipo de sistema operativo, tipo de compiladores y lenguajes

de programación que se utilizara en este equipo

52 53 54

51

п

55 56 57

5 / 5 8 5 9 6 0

61 62 63 64 65

66

67 68 69

70

| 7 | 1 |
|---|---|
| 7 | 2 |
| 7 | 3 |
| 7 | Δ |

75 76 77

78 79 80

88

89

95 96 97

98

99

100

III.HIPOTESIS

. HIPÓTESIS GENERAL

El uso de un Clúster de servidores influye significativamente en el requerimiento de alto procesamiento matemático .

HIPÓTESIS ESPECIFICAS

- Es posible determinar el análisis y diseño arquitectónico de un clúster de servidores para alto procesamiento matemático.
- El rendimiento computacional y la eficiencia del consumo de energía afecta significativamente un clúster de servidores para alto procesamiento matemático

IV. OBJETIVOS

4.1. Objetivo General

Determinar en qué medida un Clúster de servidores influye en el requerimiento de alto procesamiento matemático

4.2. Objetivos Específicos

- Determinar el análisis y diseño arquitectónico de un clúster de servidores para alto procesamiento matemático .
- Determinar en qué medida el rendimiento computacional y la eficiencia del consumo de energía varia en un clúster de servidores para alto procesamiento matemático

V. ANTECEDENTES

Wilson A. Acero, Jorge A. Aguilar, y R. David Mejía (2016), "Cluster de Alta Disponibilidad para Virtualizar los Servidores de Adquisición y Procesamiento de Datos del Instituto Geofísico". Un cluster de alta disponibilidad permite garantizar que los servicios que un sistema informático brinda estén disponibles aunque un fallo de hardware o software suceda. documento presenta de forma resumida el diseño e implementación de de alta disponibilidad basada en Pacemaker para brindar alta disponibilidad a los sistemas de adquisición y procesamiento del Instituto Geofísico: SeisComp3, Earthworm y ShakeMap. La solución está formada por dos clusters: el primero al que se denomina clusterwa es un cluster de alta disponibilidad del tipo activo/activo formado por tres servidores físicos. Este cluster administra tres servidores virtuales KVM y los módulos de los sistemas de adquisición y procesamiento que fueron instalados y configurados en los servidores virtuales. El segundo es un cluster de almacenamiento del tipo activo/pasivo al que se denomina clusteralmawa formado por dos servidores físicos. Este cluster administra un dispositivo de almacenamiento replicado que utiliza tecnología DRBD; este dispositivo contiene los discos duros de los tres servidores virtuales KVM. Pacemaker monitoriza y controla, mediante agentes de recursos, el funcionamiento de los servidores físicos, de los servidores virtuales y de los módulos que forman los sistemas mencionados y en caso de detectar alguna falla de hardware o software

Pacemaker realiza las acciones necesarias, en base a políticas y reglas configuradas por el administrador, para que los sistemas operen sin interrupción.

Sebastián A., Fernández J. M. (2014) Computación paralela y clústeres de cálculo (Parallel computing and computer clusters), Las principales razones para el uso de m'aguinas de c'omputo paralelo son: disminuir el tiempo totalde ejecuci´on de una aplicaci´on, conseguir resolver problemas m´as complejos, de grandes dimensionesy permitir la ejecuci'on simult'anea de tareas concurrentes. Existen adem'as motivos econ'omicos y quese han alcanzado unos l'imites f'isicos muy dif'iciles de mejorar tecnol'ogicamente en transmisi'on dedatos y velocidad de CPU de los procesadores actuales. Por ello, la 'unica forma de seguir aumentandoel rendimiento de los ordenadores es mediante el cambio del paradigma de computaci´on secuencial alparalelo, con m'ultiples procesadores, n'ucleos e hilos (ver Subsecci'on 1.2.1). La ley de Moore postulaque aproximadamente cada dos a nos se duplica el n'umero de transistores que se pueden meter enun circuito integrado de tama no est andar. Esta lev es emp´ırica, formulada por el cofundador delntel, Gordon E. Moore. el 19 de abril de 1965, cuyo cumplimiento se ha podido constatar hasta hoy[5]. Las habilidades de muchos dispositivos electr´onicos digitales est´an ligadas a esta ley: velocidady capacidad de procesamiento, capacidad de memoria, resoluci´on de sensores, etc... Como se puedecomprobar en la Figura 1.1, la ley de Moore ligada a la capacidad de procesamiento se mantienevigente gracias a la tecnolog'ia multin'ucleo (m'ultiples n'ucleos en un 'unico procesador), dado quedesde al a no 2006 existen muy pocos avances respecto a la velocidad de las CPUs.

Cáceres G. (2012), estrategia de implementación de un clúster de alta disponibilidad de n nodos sobre linux usando software, libre. Un clúster de alta disponibilidad permite el acceso continuo a los servicios y aplicaciones más dentro de una organización. El siguiente trabajo presenta una estrategia para la implementación de un clúster de alta disponibilidad de N nodos sobre Linux usando software libre. Este documento menciona los componentes más importantes del clúster y expone los conceptos más importantes. También explica la función que cumple cada uno de los componentes y como deben ser integrados dentro del clúster. Por último muestra como instalar, configurar y probar un clúster de alta disponibilidad con tres nodos para los servicios de transferencia de archivos, resolución de nombres, y servidor de páginas Web.

VI. MARCO TEÓRICO

Cluster de Servidores

(Borges, 2019), "Un servidor en clúster, también conocido como clúster de servidores, se trata de la unión de varios sistemas informáticos (servidores) que funcionan como si fueran uno solo.

Cuando hablamos de «unir» nos referimos a que comparten recursos de hardware y software, funcionando así como si fueran un solo sistema unificado. Esta unión de recursos se realiza con diversos fines, aunque la razón más popularidad es ofrecer velocidad y por sobre todo alta disponibilidad ante fallos.

Proveedores pioneros en el uso de servidores en clúster fueron el proyecto



153 154

155 156

157 158

159 160

161 162

163 164

165

166 167 168

169 170 171

172 173 174

176 177 178

175

179 180 181

182 183

184 185

186

187 188

189 190 191

192 193

194 195 196

197 198 199

200

201

202

ARPANET en 1969, así como Sun Microsystems o IBM. Luego va en la década de los 90 y posteriores años se popularizó el concepto gracias al uso masivo de servicios de empresas como Microsoft con productos como el viejo MSN, el servicio de servidores de correo Hotmail, Yahoo con Yahoo Mail, o los cientos de servicios de Google que conocemos hoy en día.

En palabras claras, un clúster de servidores se puede definir como un grupo de equipos informáticos que comparten recursos balanceados y están conectados a una red de alta velocidad.

Ventaias de los servidores en clúster

Existen múltiples beneficios y ventajas del uso de servidores en clúster:

Alta disponibilidad: este es uno de los grandes beneficios de la tecnología de clustering, es decir, ofrecer un mejor uptime de servidores gracias a que si falla uno, el resto tomará el trabajo ofreciendo así una alta disponibilidad del servicio. Alta velocidad de despacho: se obtiene gracias al equilibrio de las cargas, haciendo que los servidores que reciben las solicitudes de los clientes puedan despachar los datos de forma inmediata, sin delays ni retrasos de ningún tipo. Balanceo de carga: se establecen diferentes métodos de balanceo de carga, haciendo que una o varias peticiones se deriven a determinados hosts, mientras que otras irán destinadas a otro grupo de servidores bajo el clúster. Esto hace que ningún servidor se vea saturado.

Escalabilidad: la mayoría de los clúster de servidores soportan escalabilidad de forma horizontal, permitiendo agregar más servidores de forma masiva según se necesite debido a una posible creciente demanda de parte de los usuarios. Resistencia ante ataques DDOS: tener un clúster de servidores ofrece un beneficio adicional de soportar mejor las cargas de sistema que se originan ante ataques DDOS masivos, donde se suelen inundar los servidores con peticiones no deseadas. El uso de múltiples servidores puede ayudar a resistir mejores diferentes tipos de ataque destinados a agotar los recursos de sistema o red."

Computación de alto rendimiento

(«HPC», 2022), "La computación de alto rendimiento (HPC) representa la capacidad de procesar datos y realizar cálculos complejos a velocidades muy altas. Para ponerlo en perspectiva, un equipo portátil o de sobremesa con un procesador de 3 GHz puede realizar unos 3.000 millones de cálculos por segundo. Aunque esto es mucho más rápido de lo que puede lograr cualquier humano, palidece en comparación con las soluciones HPC que pueden realizar cuadrillones de cálculos por segundo.

Uno de los tipos de soluciones HPC más conocidos es el superordenador. Un superordenador contiene miles de nodos de computación que trabajan juntos para completar una o varias tareas. Esto se denomina procesamiento paralelo. Es similar a tener miles de equipos conectados en red, combinando la potencia computacional para completar tareas más rápidamente.

¿Por qué es importante la computación de alto rendimiento? Es a través de los datos que se hacen descubrimientos científicos innovadores,

se alimentan las innovaciones que cambian los juegos y se mejora la calidad de vida de miles de millones de personas en todo el mundo. HPC es la base de los avances científicos, industriales y sociales.

A medida que evolucionan tecnologías como el Internet de las cosas (IoT), la inteligencia artificial (AI) y las imágenes 3D, el tamaño y la cantidad de datos con los que las organizaciones deben trabajar crece exponencialmente. Para muchos fines, como la transmisión de un evento deportivo en directo, el seguimiento de una tormenta en desarrollo, la prueba de nuevos productos o el análisis de tendencias de stock, la capacidad de procesar datos en tiempo real es crucial.

Para seguir un paso por delante de la competencia, las organizaciones necesitan una infraestructura de TI muy fiable y rápida para procesar, almacenar y analizar cantidades masivas de datos.

¿Cómo funciona HPC?

Las soluciones HPC tienen tres componentes principales:

Computación

Red

Almacenamiento

Para crear una arquitectura de computación de alto rendimiento, los servidores de computación se conectan en un clúster. Los programas de software y los algoritmos se ejecutan simultáneamente en los servidores del clúster. El clúster está conectado en red al almacenamiento de datos para capturar la salida. Juntos, estos componentes funcionan sin problemas para completar un conjunto diverso de tareas.

Para funcionar al máximo rendimiento, cada componente debe seguir el ritmo de los demás. Por ejemplo, el componente de almacenamiento debe poder alimentar e ingerir datos hacia y desde los servidores informáticos tan rápidamente como se procesa. Asimismo, los componentes de red deben ser capaces de admitir el transporte de datos de alta velocidad entre los servidores de computación y el almacenamiento de datos. Si un componente no puede mantener el ritmo del resto, el rendimiento de toda la infraestructura HPC se resiente."

(«prototicad3D», 2022), "Una simulación computacional es el estudio de fenómenos físicos cuyas ecuaciones gobernantes se resuelven con el uso de métodos numéricos y la capacidad de cálculo de las computadoras. Las simulaciones se dividen en dos grandes ramas de análisis: el análisis termo-estructural, basado en el método del elemento finito (MEF) y el análisis de fluidos, basado en la dinámica de fluidos computacional (CFD).

El proceso mediante el cual se realiza una simulación computacional suele constar de tres partes fundamentales: a) preprocesamiento, b) solución de ecuaciones (o procesamiento) y c) postprocesamiento (ver figura 1). A continuación, se hace una breve descripción de estas etapas:

En el preprocesamiento se introducen los datos geométricos, las condiciones de borde, las excitaciones del sistema, el mallado y el tipo de análisis a realizar.

Para poder llevarlo a cabo se necesita conocer de antemano las condiciones de trabajo iniciales, lo que implica muchas veces tomar datos en campo o suposiciones sobre el comportamiento preliminar del componente en estudio, y de aquí el discernimiento sobre qué análisis se debe hacer para obtener la respuesta buscada.

El procesamiento, o solución de las ecuaciones, se logra por medio del uso de un computador y algoritmos especializados en el cálculo numérico. Aunque este paso lo realiza la computadora de forma independiente en base a la configuración previa, la calidad de la información inicial suministrada determinará a su vez la validez y precisión de los resultados a obtener.

Finalmente se realiza el postprocesamiento, en donde se visualizan y analizan con más detalle los resultados. De ser necesario se solicitan resultados adicionales que complementen el estudio que se está llevando a cabo, todos en base al archivo de solución resuelto en la etapa anterior. En caso de que los resultados no sean los esperados, o se observen incongruencias en los mismos, se decide si se debe volver a la etapa de preprocesamiento o si se está conforme con la solución.

Mediante la simulación computacional se pueden estudiar una gran variedad de fenómenos físicos, lo que permite al diseñador estudiar diversos escenarios sin necesidad de llevar a cabo numerosos experimentos de laboratorio. Usar este tipo de tecnología para el desarrollo de productos permite alcanzar soluciones de forma rápida y precisa. Desde la ingeniería conceptual hasta la ingeniería de detalles se podrá tener una amplia comprensión de la física involucrada en el problema y de las diferentes respuestas de los sistemas diseñados.

Otra característica importante es que permiten el estudio de productos, equipos y/o procesos existentes, lo que conlleva a las empresas a mejorar sus actividades al aplicar esta tecnología en distintos ámbitos, tales como: mantenimiento, procesos de fabricación, análisis de fallas, investigación, optimización y mejora de productos.

En el caso de la dinámica de fluidos computacional, se puede estudiar el comportamiento de un fluido incluyendo todas sus características físico-químicas reales, permitiendo a los ingenieros evaluar el comportamiento de sus diseños antes de ser construidos, y sin la necesidad de costosos experimentos. Por mencionar algunos ejemplos, con apoyo del CFD se pueden realizar estudios aerodinámicos para evaluar el comportamiento del flujo externo sobre vehículos y aeronaves (ver figura 3). El estudio proporciona valores como coeficientes de arrastre y sustentación, perfiles de presiones, perfiles de velocidad, etc. De igual forma en otro ramo industrial, en la simulación de turbomáquinas se puede evaluar la eficiencia de consumo de estos equipos y sus curvas de operación

En el campo de los análisis estructurales por el método del elemento finito, hacen posible obtener valores de esfuerzo – deformación en cualquier elemento de una estructura compleja. Otros análisis importantes obtenidos con este método son: estudios térmicos, análisis de vibraciones, evaluaciones de

daño acumulado y vida a fatiga, optimización estructural, por mencionar sólo algunos. Con el uso de simulaciones computacionales se estudia el comportamiento de mecanismos complejos y de la interacción entre sus partes. Por ejemplo, luego de un análisis dinámico, es posible obtener valores de desplazamientos, velocidades y aceleraciones en un ensamblaje representando partes de los componentes de la suspensión de un vehículo.

Como hemos visto, todas las simulaciones computacionales se pueden llevar a cabo si se ejecutan de forma adecuada cada una de sus tres etapas fundamentales, tanto en aplicaciones FEA o CFD. Existen una gama de estudios muy variados en otras físicas que no fueron mencionadas, como por ejemplo estudios electromagnéticos de alta y baja frecuencia. En todo caso, esta publicación buscar dar información adicional a todos aquellos técnicos y especialistas en áreas de diseño sobre los beneficios técnicos, económicos y de tiempo cuando se implementan las herramientas de simulación en los ciclos convencionales de diseño."

VII. METODOLOGÍA

7.1. Lugar de estudio

Se desarrollara en el Taller de redes y telecomunicaciones de la Facultad de Ingeniería de Sistemas e Ingeniería Civil de la Universidad Nacional de Ucayali..

7.2. Población y tamaño de muestra

Población.

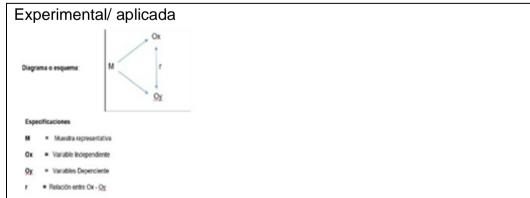
La población estará conformada Un clúster de servidores.

Muestra

La muestra será igual a la población por ser una investigación de corte experimental/ Aplicada.

7.3. Descripción detallada de los métodos, uso de materiales, equipos o insumos.

a) Diseño de muestreo



b) Descripción detallada del uso de materiales, equipos, insumos, entre otros.

 Se utilizará en el desarrollo de la presente investigación 4 servidores HP Gen 9, un Switch Administrable 10 Gb, un Rack de 68 RU, Patch, 5 Patch Cord y un Power Rack de 6 tomas,

c) Descripción de variables a ser analizados en el objetivo específico

Un Cluster de Servidores (conformado por 4 servidores HP Gen 8(Intel Xeon Silver, (10 cores, 20 Hilos Windows y Linux reconoce 20 CPUs internos, 8 Gb RAM y SSD 240 GB)

Alto procesamiento matemático (Serán algoritmos que requieran gran capacidad de procesamiento, ordenamiento de bases de datos, búsquedas, simulaciones, se considerarán la complejidad, tiempo y velocidad de procesamiento)

d) Aplicación de prueba estadística inferencial.

Se aplicaran en muestras independientes, tomando en consideración los aspectos cuantitativos y distribución normal, siendo el coeficiente de correlación de Pearson para la prueba de hipótesis

7.4. Tabla de recolección de datos por objetivos específicos.

se establecera la relación de las dos variables cuantitativas continuas con distribución normal (un cluster de servidores y su aplicación en requerimientos de alto procesamiento matemático) se utilizará el coeficiente de correlación de Pearson (r de Pearson).

VIII. CRONOGRAMA DE ACTIVIDADES

| Actividad | Trimestres | | | | | | | | | | | |
|---|------------|---|---|---|---|---|---|---|---|---|---|---|
| Actividad | | | | | | | | | | | | X |
| PLANIFICACION Y APROBACION DEL PROYECTO | | | | | | | | | | | | |
| Elección del tema | Χ | | | | | | | | | | | |
| Revisión bibliográfica | | Χ | | | | | | | | | | |
| Elaboración del proyecto | | Χ | Χ | | | | | | | | | |
| Aprobación del proyecto | | | Χ | | | | | | | | | |
| Reestructuración del proyecto | | | Χ | Χ | | | | | | | | |
| EJECUCIÓN DEL PROYECTO | | | | | | | | | | | | |
| Marco teórico | | | | Χ | | | | | | | | |
| Validación del instrumento | | | | Χ | | | | | | | | |
| Implementación del Software | | | | Χ | Χ | | | | | | | |
| Procesamiento de datos | | | | | Χ | | | | | | | |
| Análisis e interpretación de datos | | | | | | Χ | Χ | | | | | |
| INFORME FINAL | | | | | | | | | | | | |
| Sistematización final del informe | | | | | | | | Χ | | | | |
| Conclusiones y sugerencias | | | | | | | | | Χ | | | |
| Presentación del informe final | | | | | | | | | | Χ | | |
| Aprobación de la tesis | | | | | | | | | | | Χ | |
| Sustentación del informe final | | | | | | | | | | | | Χ |

| Descripción | Unidad de medida | Costo Unitario (S/.) | Cantidad | Costo total (S/.) |
|--------------------------------------|---------------------|----------------------|----------|-------------------|
| Alimentos y Bebidas | ración | 100.00 | 4 | 400.00 |
| Materiales y útiles de Oficina | Unidad | 200.00 | 1 | 200.00 |
| Equipos informáticos | Equipo | 3,400.00 | 1 | 3,400.00 |
| Servicios Diversos | Unidad | 2,000.00 | 1 | 2,000.00 |
| | | | Total | 5,000.00 |

VIII. BIBLIOGRAFÍA

 Abarca F. (2020), Universidad Complutense de Madrid, Análisis de rendimiento de un clúster heterogéneo formado por placas Raspberry Pi

Aliaga E., Román W. y Pinedo R. Universidad Nacional de Ucayali (2016), sistema automatizado para el control y monitoreo del comportamiento de alevinos de paiche en cautiverio

Borges, E. (2019). Servidor en cluster. Recuperado 15 de agosto de 2022, de infranetworking website: https://blog.infranetworking.com/servidor-en-cluster/

Cariño S (2006) Métodos de Investigación I, Editorial Limusa, México

Carrasco S. (2009:243). Metodología de la Investigación Científica Lima.

HPC. (2022). Recuperado 15 de agosto de 2022, de NetApp website: <a href="https://www.netapp.com/es/data-storage/high-performance-computing/what-is-hpc/#:~:text=La%20computaci%C3%B3n%20de%20alto%20rendimiento%20(HPC)%20representa%20la%20capacidad%20de,millones%20de%20c%C3%A1lculos%20por%20segundo.

Méndez, C. E. Metodología Diseño y Desarrollo del proceso de investigación, Tercera Edición, Editorial Mc Graw Hill, 2005. Colombia.

Nugroho, S & Widiyanto, A. (2020). Designing parallel computing using raspberry pi clusters for IoT servers on apache Hadoop. Journal of Physics: Conference Series

Prototicad3D. (2022). Recuperado 15 de agosto de 2022, de SIMULACION COMPUTACIONAL website: https://www.prototicad3d.com/2022/03/simulacion-computacional-una.html

Sánchez H., Carlessi C., Reyes. Metodología y diseños en la Investigación Científica., pp.83-118

Sierra D. (2006) Metodología de Investigación Científica.

Salas, E. (2000) Una Introducción a la Investigación Científica. Ed. Gráfica Educativa, p.288

IX. ANEXO

| PROBLEMA | OBJETIVO GENERAL | HIPOTESIS | VARIABLES | DIMENCIONES | METODOLOGIA |
|--|--|---|--|---|-----------------------------|
| PROBLEMA GENERAL ¿En qué medida un Clúster de servidores influye en el requerimiento de alto procesamiento matemático? PROBLEMAS ESPECIFICOS ¿Es posible determinar el análisis y diseño arquitectónico de un clúster de servidores para alto procesamiento matemático? ¿En qué medida el rendimiento computacional y la eficiencia del consumo de energía varia en un clúster de servidores para alto procesamiento matemático? | OBJETIVO GENERAL Determinar en qué medida un Clúster de servidores influye en el requerimiento de alto procesamiento matemático OBJETIVOS ESPECÍFICOS • determinar el análisis y diseño arquitectónico de un clúster de servidores para alto procesamiento matemático . • Determinar en qué medida el rendimiento computacional y la eficiencia del consumo de energía varia en un clúster de servidores para alto procesamiento matemático | HIPÓTESIS GENERAL El uso de un Clúster de servidores influye significativamente en el requerimiento de alto procesamiento matemático . HIPÓTESIS ESPECIFICAS • Es posible determinar el análisis y diseño arquitectónico de un clúster de servidores para alto procesamiento matemático. • El rendimiento computacional y la eficiencia del consumo de energía afecta significativamente un clúster de servidores para alto procesamiento matemático | Clúster de servidores Alto procesamiento matemático | Análisis Diseño Despliegue Complejidad Tiempo Velocidad de procesamiento | ESQUEMA DE LA INVESTIGACIÓN |