

## *Arquitecturas Escalables*

### *computación de alto rendimiento*

Alexy Yessenia Rincón Capera  
Universidad Industrial de Santander, UIS  
Bucaramanga, Colombia  
Alexyr93@gmail.com

Harver Andrey Cordero Duarte  
Universidad Industrial de Santander, UIS  
Bucaramanga, Colombia  
harvercordero@gmail.com

**Abstract**—Este documento electrónico se enfocara en dar una breve descripción de terminología sobre supercomputación para concluir con una comparación entre la supercomputadora GUANE-1 de la Universidad Industrial de Santander UIS y la actual mejor computadora según el top500 de supercomputo, se tratara una descripción de cada ordenador dando datos técnicos y su arquitectura correspondiente. (*Abstract*)

**Abstract**— This electronic document will focus on giving a brief description of terminology supercomputing to conclude with a comparison between the GUANE -1 supercomputer of the Industrial University of Santander UIS and the current best computer according to the Top 500 Supercomputing , a description of each computer were giving technical data and corresponding architecture. (*Abstract*)

La escalabilidad esta intrínsecamente relacionada al diseño. Dicha característica influye en el rendimiento, de tal forma que se hace indispensable al momento de ser constituido dicho Sistema.

Como ya se menciona con anterioridad que la escalabilidad es un factor importante, se debe tener en cuenta que la misma supone un factor crítico en su crecimiento, ahora, pensemos que la finalidad de un sistema sea crecer en la cantidad de usuarios que hagan uso del mismo, conservando el rendimiento actual, para ello se debe tener en cuenta 2 posibles opciones:

- Tener un Hardware de mayor potencia
- Tener una mejor combinación tanto de hardware como de software.

Aclarados los anteriores conceptos es importante resaltar que existen 2 tipos de escalabilidad.

#### *1. Escalabilidad Vertical*

Este tipo de escalabilidad, también llamada escalar hacia arriba, consta en añadir más recurso de hardware a un nodo en particular, el cual será más potente y eficaz que el actual. Ya configurado el sistema, se procede a realizar una serie de validaciones y respectivas copias de seguridad para ponerlo de nuevo en funcionamiento; además la migración de las aplicaciones soportadas por los anteriores componentes no sufrirá más que unas leves modificaciones en el código de ser necesarias.

#### *2. Escalabilidad Horizontal*

La escalabilidad horizontal se caracteriza por potenciar el rendimiento de cualquier sistema de manera global, diferenciándose así de la escalabilidad vertical que busca aumentar potencia en una parte o nodo en concreto, es decir, agrega más nodos al sistema de tal manera que funcione a modo espejo.

Teniendo en cuenta los tipos de escalabilidad, no basta solo con esto, cuando se diseña un sistema en torno a compartir recursos, es necesario considerar el cómo se va a balancear la carga del trabajo, es decir dividir el trabajo entre varios procesos, nodos u otros recursos. Este concepto surge de la necesidad de evitar el efecto cuello de botella y para ello las

## I. INTRODUCCIÓN

EL crecimiento exponencial de la tecnología trae como consecuencia un estado de cambio y continuo crecimiento en lo que a computación se refiere. Partiendo de esta premisa nace la necesidad de que las nuevas tecnologías puedan ofrecer servicios de calidad al mismo tiempo que poseen la capacidad de reaccionar a los cambios continuos y adaptarse al crecimiento acelerado, con la característica de no perder la calidad en su rendimiento y el servicio ofrecido; esta característica recibe el nombre de escalabilidad, la cual desempeña un papel fundamental en las arquitecturas de computación de alto rendimiento.

## II. MARCO TEORICO

### A. Escalabilidad

Se define por escalabilidad a la capacidad de adaptación de un Sistema en torno al rendimiento del mismo a medida que se ve un aumento significativo en la cantidad de usuarios del mismo. Aunque es sencillo entender el concepto de escalabilidad, se torna como un aspecto complejo, fundamental e importante en el diseño de la arquitectura de una supercomputadora o un Sistema en general.

<sup>1</sup> .Es una medida del rendimiento del equipo, útil en los campos de los cálculos científicos; es una más exacta de las instrucciones generadas por segundo.

<sup>2</sup> .Reduced Instruction Set Computer, es un tipo de diseño de CPU generalmente usada en microprocesadores.

tareas se reparte a través de un algoritmo encargado de analizar las peticiones y redireccionarlas al nodo indicado para su mejor funcionamiento y procesamiento.

### B. Balance de carga

Al momento de diseñar un sistema el cual va a compartir recursos, se hace necesario el considerar como se va a balancear la carga del trabajo. Este concepto obtiene sentido si se ve como la técnica utilizada para dividir el trabajo en varias tareas o procesos, nodos o cualquier otro recurso disponible. Esto a su vez se relaciona con los sistemas multiprocesales, los cuales trabajan o podrían hacerlo con múltiples unidades para cumplir con su funcionalidad. Como se mencionó anteriormente se debe tener en cuenta evitar el efecto cuello de botella, razón por la cual el balance de carga del trabajo se reparte de manera uniforme por medio de un algoritmo que analiza las repeticiones del sistema y las redirecciona a la opción que le parece más eficiente.

Entendido el concepto de balance de carga, ahora debemos tener en cuenta que este consta de dos partes, el balance de carga se puede aplicar tanto por hardware, como por software.

#### 1) Balance de carga por Hardware

Este tipo de balance presenta las siguientes características:

- A partir de un algoritmo llamado (Round Robin, LRU), examina peticiones de protocolo HTTP entrantes y selecciona el que considera más apropiado entre los diferentes del sistema.
- La elección del clon del sistema se basa en un algoritmo de sustitución que es aleatorio.
- Teniendo en cuenta lo anterior, provoca algunas dificultades en el diseño, pues no se garantiza que si un usuario realiza múltiples peticiones, estas sean atendidas por el mismo clon del sistema; motivo por el cual no hay mantenimiento en la sesión del usuario en servidor y esto condiciona en una medida el diseño.
- Hablando de la sesión, esta necesariamente debe ser mantenida por el desarrollador.
- Finalmente como esto es netamente un proceso de hardware, dependiendo del mismo este puede hacerse muy rápido.

#### 2) Balance de carga por software

- Toma el paquete y lo examina a nivel de protocolo HTTP con la finalidad de garantizar el mantenimiento de la sesión de usuario.
- Distintas peticiones del mismo usuario son servidas por el mismo clon del servidor.
- Son algo más lentos que los balanceados por hardware

- Tienen la característica de ser soluciones que conllevan un gasto menor que las de hardware.

### 3) Clusters sobre servidores

El término conceptualizado de clustering abre una nueva característica, permite la capacidad de unir servidores con la finalidad de que estos trabajen en paralelo, esto quiere decir, que trabajan como si fueran uno solo. En las etapas primigenias del clustering, tema que se tocará más a fondo con el desarrollo de este documento, sus diseños presentaban problemas muy serios que se han podido solventar con el paso del tiempo y gracias a la evolución presente en este campo. En la actualidad se crean clusters en función de ciertas necesidades como unión de hardware, clusters de software y alto rendimiento en las bases de datos.[1]

## III. COMPUTACION DE ALTO RENDIMIENTO

La computación de alto rendimiento (HCP) se ha convertido en una herramienta de gran importancia para la resolución de problemas complejos por medio de simulaciones computacionales. Su uso se presenta a manera de resolver grandes problemas en diferentes áreas de conocimiento ya sea en campos de la ciencia, la ingeniería o negocio.

En la práctica algunos problemas pueden solucionarse con solo un procesador, pero hay muchos casos donde es necesario el uso de varios procesadores para la solución de problemas complejos introduciéndose así en la computación en paralelo. [2]

### A. Computación en Paralelo

Esta es una técnica de programación que permite la ejecución de muchas instrucciones en forma simultánea, aplicando cierto principio que menciona que los problemas grandes en muchas circunstancias pueden dividirse en problemas más pequeños y luego ser resueltos de forma simultánea.

### B. Hardware

La memoria principal en un ordenador en paralelo puede ser compartida entre todos los elementos de procesamiento en un único espacio de direcciones, o distribuida, cada elemento de procesamiento tiene su propio espacio local de direcciones. El término memoria distribuida se refiere al hecho de que la memoria se distribuye lógicamente, pero a menudo implica que también se distribuyen físicamente. La memoria distribuida y la visualización de memoria combinan los dos enfoques, donde el procesador tiene su propia memoria local y permite acceso a la memoria de los procesadores que no son locales. Los accesos a la memoria local suelen ser más rápidos que los accesos a memoria no local. [3]

<sup>1</sup> .Es una medida del rendimiento del equipo, útil en los campos de los cálculos científicos; es una más exacta de las instrucciones generadas por segundo.

<sup>2</sup> .Reduced Instruction Set Computer, es un tipo de diseño de CPU generalmente usada en microprocesadores.

Existen diferentes tipos de computación en paralelo:

- Paralelismo a nivel de bit:

Computadora esto reduce el número de instrucciones necesarias que el procesador deba ejecutar para realizar una operación en variables las cuales sus tamaños son más grandes que la longitud de la palabra. Por ejemplo, cuando un procesador de 8 bits debe sumar dos enteros de 16 bits, el procesador primero debe adicionar los 8 bits de orden inferior de cada número entero con la instrucción de adición, luego, añadir los 8 bits de orden superior utilizando la instrucción de adición con acarreo que tiene en cuenta el bit de acarreo de la adición de orden inferior, en este caso un procesador de 8 bits requiere dos instrucciones para completar una sola operación, en donde un procesador de 16 bits necesita una sola instrucción para poder completarla.

- Paralelismo a nivel de instrucciones:

Un programa de computador es una secuencia de instrucciones que son ejecutadas por un computador, las cuales pueden reordenarse y combinarse en grupos y ser ejecutadas paralelamente sin cambiar el resultado final del programa. Este tipo de paralelismo fue predominante desde mediados de la década de los 80's hasta mediados de la década de los 90's.

- Paralelismo de tareas:

Una característica de un programa en paralelo es la que cálculos completamente diferentes se pueden realizar en cualquier conjunto igual o diferente de datos. Por lo general el paralelismo de tareas no escala con el tamaño de un problema.

- Paralelismo de Datos:

Distribuye los datos a través de diferentes nodos de cómputo para que los procesen en paralelo y las tareas que realizan son comunes y el resultado es uno solo. [4]

#### IV. TAXONOMIA DE FLYNN

En 1966 Michael Flynn propuso un mecanismo de clasificación de las computadoras. El método de Flynn se basa en el número de instrucciones y de la secuencia de datos que la computadora utiliza para procesar información. Puede haber secuencias de instrucciones sencillas o múltiples y secuencias de datos sencillas o múltiples. Esto da lugar a 4 tipos de computadoras, de las cuales solamente dos son aplicables a las computadoras paralelas.

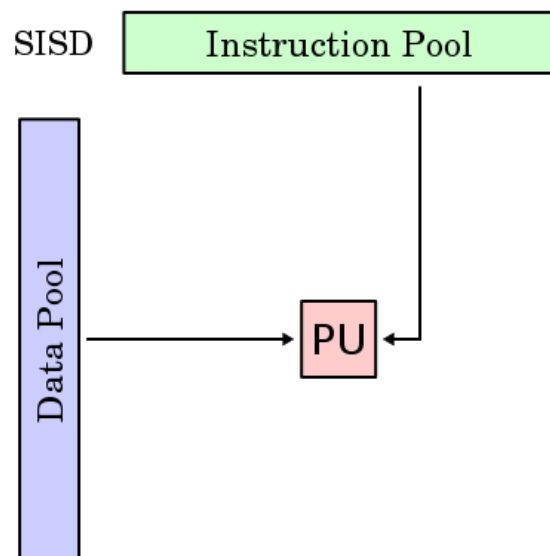
- **Una instrucción, un dato (SISD)**

En computación, SISD (Single Instruction, Single Data, en español: "una instrucción, un dato") es un término que se refiere a una arquitectura computacional en la que un único procesador ejecuta un sólo flujo de instrucciones, para operar sobre datos almacenados en una única memoria. Se corresponde con la arquitectura de Von Neumann.

Según Michael J. Flynn, SISD puede tener características de procesamiento concurrente. La carga de instrucciones y la ejecución segmentada de instrucciones son ejemplos comunes encontrados en las computadoras SISD más modernas.

##### A. Características

- La CPU procesa únicamente una instrucción por cada ciclo de reloj
- Únicamente un dato es procesado en cada ciclo de reloj
- Es el modelo más antiguo de computadora y el más extendido



- **Múltiples instrucciones, un dato (MISD):**

En computación, MISD (Multiple Instruction, Single Data, en español: "múltiples instrucciones, un dato") es un tipo de arquitectura computacional (particularmente de computación paralela) donde muchas unidades funcionales realizan diferentes operaciones en los mismos datos. Las arquitecturas segmentadas pertenecen a este tipo, aunque en un extremo se podría llegar a decir que los datos son diferentes después de ser procesados por cada etapa en el pipeline, con lo cual no entraría en esta categoría.

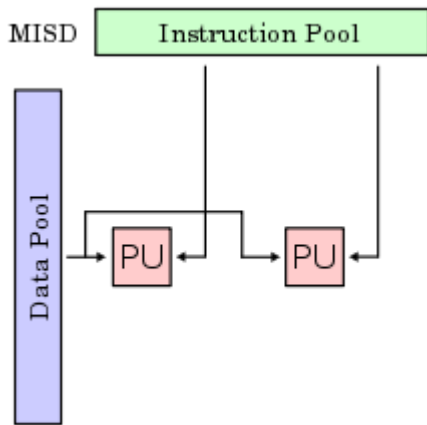
<sup>1</sup> .Es una medida del rendimiento del equipo, útil en los campos de los cálculos científicos; es una más exacta de las instrucciones generadas por segundo.

<sup>2</sup> .Reduced Instruction Set Computer, es un tipo de diseño de CPU generalmente usada en microprocesadores.

Las máquinas tolerantes a fallos ejecutan la misma instrucción redundantemente para detectar y corregir errores, utilizando task replication, son consideradas de este tipo. No existen muchos ejemplos de esta arquitectura dado que las técnicas más comunes de procesamiento de datos en paralelo suelen ser más apropiadas para MIMD y SIMD. Específicamente, facilitan el escalamiento y el uso de recursos computacionales mejor que MISD.

**Características del modelo MISD:**

- Cada unidad ejecuta una instrucción distinta
- Cada unidad procesa el mismo dato
- Aplicación muy limitada en la vida real



**Una instrucción, múltiples datos (SIMD)**

En computación, SIMD (Single Instruction, Multiple Data, en español: "una instrucción, múltiples datos") es una técnica empleada para conseguir paralelismo a nivel de datos.

Los repertorios SIMD consisten en instrucciones que aplican una misma operación sobre un conjunto más o menos grande de datos. Es una organización en donde una única unidad de control común despacha las instrucciones a diferentes unidades de procesamiento. Todas éstas reciben la misma instrucción, pero operan sobre diferentes conjuntos de datos. Es decir, la misma instrucción es ejecutada de manera sincronizada por todas las unidades de procesamiento.

**Características del modelo SIMD:**

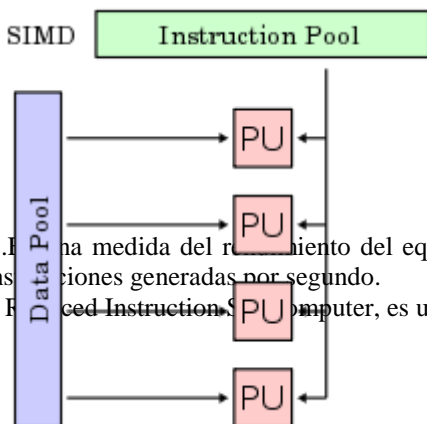
- Todas las unidades ejecutan la misma instrucción
- Cada unidad procesa un dato distinto
- Todas las unidades operan simultáneamente

**Múltiples instrucciones, múltiples datos (MIMD)**

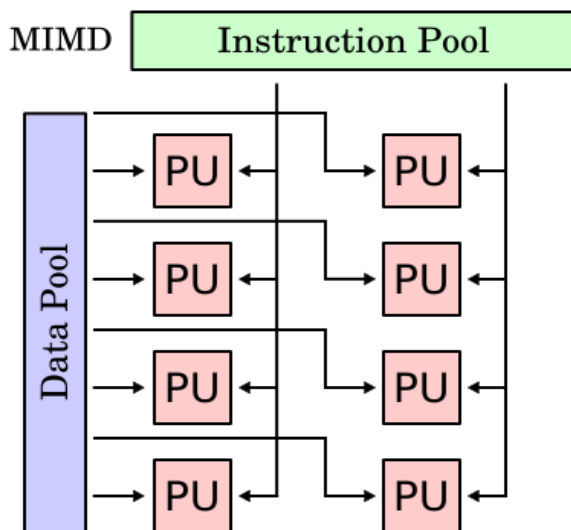
En computación, MIMD (del inglés Multiple Instruction, Multiple Data, en español "múltiples instrucciones, múltiples datos") es una técnica empleada para lograr paralelismo. Las máquinas que usan MIMD tienen un número de procesadores que funcionan de manera asíncrona e independiente. En cualquier momento, cualquier procesador puede ejecutar diferentes instrucciones sobre distintos datos. La arquitectura MIMD pueden utilizarse en una amplia gama de aplicaciones como el diseño asistido, simulación, modelado y en interruptores. Las computadoras MIMD se pueden categorizar por tener memoria compartida o distribuida, clasificación que se basa en cómo el procesador MIMD accede a la memoria. La memoria compartida de las máquinas puede estar basada en buses, extensiones, o de tipo jerárquico. Las máquinas con memoria distribuida pueden tener esquemas de interconexión en hipercubo o malla

**Características del modelo MIMD:**

- Cada unidad ejecuta una instrucción distinta
- Cada unidad procesa un dato distinto
- Todas las unidades operan simultáneamente [5]



1. Una medida del rendimiento del equipo, útil en los campos de los cálculos científicos; es una más exacta de las instrucciones generadas por segundo.  
 2. Single Instruction, Multiple Data computer, es un tipo de diseño de CPU generalmente usada en microprocesadores.



## B. Ventajas

Algunas ventajas son que permite resolver problemas que no pueden resolverse en una sola unidad central de procesamiento CPU, también permite resolver problemas que no se resuelven en un tiempo razonable.

Los algoritmos deben ser diseñados de forma eficiente, ya que no habría ganancia en caso de que estos no sean diseñados adecuadamente. El diseño de computadores paralelos debe proporcionar alta escalabilidad. Los lenguajes para computación en paralelo deben ser flexibles para permitir una implementación eficiente, y facilidad para su programación. [6]

En los últimos años se ha venido empleado con mayor interés en la computación de alto rendimiento, dadas a las limitaciones de la arquitectura las cuales impiden el aumento en la frecuencia, la cual provocaría un mayor consumo de energía.

Para el uso de sistemas con múltiples procesadores se podemos ver que normalmente toman por uno de dos caminos posibles, uno de ellos es la computación distribuida que utiliza la potencia de procesamiento de muchos computadores, tanto organizadas como distribuidas, con muchos dominios administrativos utilizándola siempre que un ordenador este disponible. Por otro lado utilizar un gran numero de procesadores próximos entre si, por ejemplo un cluster de computadores. En un sistema paralelo masivo centralizado es de gran importancia la velocidad y flexibilidad de interconexión.

Es una computadora cuya capacidad es de alto nivel computacional a comparación de una orientada a propósitos generales. Su rendimiento es medido en operaciones de punto flotante por segundo (FLOPS) en lugar de millones de instrucciones por segundo. Estas supercomputadores juegan un papel muy importante en las ciencias de la computación donde se utiliza para realizar tareas de cálculo en múltiples campos, incluida la computación cuántica y también permite realizar simulaciones físicas, abordan problemas muy complejos o que no pueden realizarse en el mundo físico debido a que pueden ser muy peligrosos, también porque pueden involucrar cosas o muy grandes o demasiado pequeñas. Algunos ejemplos pueden ser el moldeamiento de explosiones nucleares, el estudio del sol o el clima espacial, moldeamiento del clima para predecir el clima futuro, etc.

HPC usa supercomputadores y cluster de supercomputadores para resolver problemas avanzados de computación.[7]

## A. Clúster

Podemos ver en la actualidad diferentes formas de HPC, Debido a un alto rendimiento y bajo costo, el clúster es, con mucho, la forma más popular de la computación HPC. Clúster es un conjunto o conglomerado de ordenadores unidos entre sí por medio de una red de alta velocidad donde comparten sus características para formar una única computadora. Usualmente son utilizados para mejorar el rendimiento y disponibilidad muy por encima al que proveería un computador individual, resaltando en su rendimiento, disponibilidad, balanceo de carga y escalabilidad. Debido a su flexibilidad es económico y fácil de construir [8]

El clúster HPC de los productos básicos: Durante los últimos diez años, el clúster HPC ha perturbado todo el mercado de la supercomputación. Construido a partir de servidores off-the-shelf estándar e interconexiones de alta velocidad, un sistema HPC típica puede entregar, el rendimiento rentable líder en la industria. Un grupo típico puede emplear a cientos, miles e incluso decenas.

Los clústeres pueden clasificarse según sus características:

- HPC (High Performance Computing Clústeres: clústeres de alto rendimiento).

### Alto rendimiento:

Son clústeres en los cuales se ejecutan tareas que requieren de gran capacidad computacional, grandes cantidades de memoria, o ambos a la vez. El llevar a cabo estas tareas puede comprometer los recursos del clúster por largos periodos de tiempo.

- HA o HACC (High Availability Computing Clusters: clústeres de alta disponibilidad).

## V. SUPERCOMPUTADORA

<sup>1</sup> .Es una medida del rendimiento del equipo, útil en los campos de los cálculos científicos; es una más exacta de las instrucciones generadas por segundo.

<sup>2</sup> .Reduced Instruction Set Computer, es un tipo de diseño de CPU generalmente usada en microprocesadores.

**Alta disponibilidad:** Son clústeres cuyo objetivo de diseño es el de proveer disponibilidad y con fiabilidad. Estos clústeres tratan de brindar la máxima disponibilidad de los servicios que ofrecen. La con fiabilidad se provee mediante software que detecta fallos y permite recuperarse frente a los mismos, mientras que en hardware se evita tener un único punto de fallos.

- HT o HTCC (High Throughput Computing Clusters: clústeres de alta eficiencia).

**Alta eficiencia:** Son clústeres cuyo objetivo de diseño es el ejecutar la mayor cantidad de tareas en el menor tiempo posible. Existe independencia de datos entre las tareas individuales. El retardo entre los nodos del clúster no es considerado un gran problema. [9]

## B. Grid Computing

Es un paradigma de computación distribuida que une la potencia de múltiples equipos para lograr su escalabilidad y capacidad. Esta infraestructura integra y usa en forma colectiva los recursos computacionales en diferentes organizaciones conectadas en red

Permite compartir recursos no centralizados geográficamente para resolver problemas de gran escala. Ofrece muchas ventajas a comparación de otras tecnologías alternativas.

La potencia que ofrece la infraestructura grid al conectar múltiples computadores es prácticamente ilimitada, y ofrece una perfecta integración de sistemas y dispositivos heterogéneos sin generar algún problema en las conexiones entre las diferentes máquinas.

Capacidad de balanceo de sistemas: no habría necesidad de calcular la capacidad de los sistemas en función de los picos de trabajo, ya que la capacidad se puede reasignar desde la granja de recursos a donde se necesite;

- Alta disponibilidad. con la nueva funcionalidad, si un servidor falla, se reasignan los servicios en los servidores restantes;
- Reducción de costes: con esta arquitectura los servicios son gestionados por "granjas de recursos". Ya no es necesario disponer de "grandes servidores" y podremos hacer uso de componentes de bajo coste.

Cada sistema puede ser configurado siguiendo el mismo patrón;

Se relaciona el concepto de grid con la nueva generación del protocolo IP. El nuevo protocolo de Internet IPv6 permitirá trabajar con una Internet más rápida y accesible. Una de las ideas clave en la superación de las limitaciones actuales de Internet IPv4 es la aparición de nuevos niveles de servicio que harán uso de la nueva capacidad de la red para intercomunicar los ordenadores.

Este avance en la comunicación permitirá el avance de las ideas de grid computing al utilizar como soporte la altísima conectividad de Internet. Es por ello que uno de los campos de mayor innovación en el uso del grid computing, fuera de los conceptos de supercomputación, es el desarrollo de un estándar para definir los *Grid Services* frente a los actuales *Web Services*. [10]

## C. Cloud Computing

Es un modelo de servicios escalables bajo demanda para la asignación y el consumo de recursos de cómputo. Describe el uso de infraestructura, aplicaciones, información y una serie de servicios compuestos por reservas de recursos de computación, redes, información y almacenamiento.

La prestación de servicios de cloud computing puede asociarse a tres modelos específicos:

### 1) *Infrastructure as a Service(IaaS)*

Ofrece al consumidor la provisión de procesamiento, almacenamiento, redes y cualquier otro recurso de cómputo necesario para poder instalar software, incluyendo el sistema operativo y aplicaciones. El usuario no tiene control sobre el sistema de nube subyacente pero si del Sistema operativo y aplicaciones. Ejemplo: Amazon Web Services EC2.

### 2) *Platform as a Service(PaaS)*

Ofrece al consumidor la capacidad de ejecutar aplicaciones por éste desarrolladas o contratadas a terceros, a partir de los lenguajes de programación o interfaces provistas por el proveedor. El usuario no tiene control ni sobre el sistema subyacente ni sobre los recursos de Infraestructura de nube. Ejemplo: Microsoft Azure.

### 3) *Software as a Service(SaaS)*

Ofrece al consumidor la capacidad de utilizar las aplicaciones del proveedor que se ejecutan sobre la infraestructura en la nube. Las aplicaciones son accedidas desde los dispositivos cliente a través de interfaces, por

<sup>1</sup> .Es una medida del rendimiento del equipo, útil en los campos de los cálculos científicos; es una más exacta de las instrucciones generadas por segundo.

<sup>2</sup> .Reduced Instruction Set Computer, es un tipo de diseño de CPU generalmente usada en microprocesadores.

ejemplo un navegador web. En este caso, el usuario solo tiene acceso a una interfaz de configuración del software provisto. Ejemplo: Salesforce.[11]

El uso de un gran numero de computadores organizados en clusters incrustados con una infraestructura de telecomunicaciones distribuida para la resolución de problemas de computación masiva es a lo que se llama computación distribuida. Y su aplicando se a ha dado utilizando muchas de las ideas de la computación de alto rendimiento.

Un sistema distribuido se puede definir como un conjunto de computadoras que se encuentran separadas físicamente y están conectadas entre sí por medio de una red de comunicaciones; donde cada máquina posee sus propios componentes de hardware y software los cuales son percibidos como un solo sistema.

Esto con el objetivo de proporcionar con fiabilidad al sistema, ya que encaso de que un componente se descomponga otro componente debe ser capaz de remplazarlo.

La administración del calor es uno de los mayores problemas en los complejos dispositivos electrónicos y afecta el potencial de los sistemas informáticos en varias formas. El ensamblaje de miles de procesadores juntos genera inevitablemente cantidades significativas de densidad de calor que necesitan ser tratados. La habilidad del sistema para remover el calor es un factor limitante. Existen muchas supercomputadoras que tienen mas capacidad de infraestructura para el diseño general conservando la potencia y la infraestructura de refrigeración para manejar mas que los picos eléctricos de potencia teórica consumida por la supercomputador.[12]

#### D. Sunway TaihuLight vs Guane Cluster

##### 1) Sunway TaihuLight

El nuevo supercomputador que se estima es el más potente actualmente en el mundo, originario de China en su totalidad, ocupando el primer lugar en el top 500 de supercomputo. Es una Es una supercomputadora diseñada totalmente con procesadores hechos en china, pues estados unidos vetó la exportación a este país; con una potencia de 93 petaflops<sup>1</sup> en el punto de referencia LINPACK.

Fue desarrollada por el Centro Nacional de Investigacion de Ingenieria y Tecnologia en paralelolo (NRCPC)y fue instalado en el centro Nacional de Supercomputación en Wuxi.

Sunway TaihuLight cuenta con 10.649.600 de nucleos de coputacion que comprenden 40.960 nodos y es dos veces más rápida y tres veces tan eficiente como su predecesora Tianhe-2, supercomputadora china que tambien ocupo el primer lugar en el top500 de supercomputacion que registro un rendimiento de 33.86 petaflops/s.

<sup>1</sup> .Es una medida del rendimiento del equipo, útil en los campos de los cálculos científicos; es una más exacta de las instrucciones generadas por segundo.

<sup>2</sup> -Reduced Instruction Set Computer, es un tipo de diseño de CPU generalmente usada en microprocesadores.

Respecto al consumo de potencia de pico bajo se encuentra alrededor de los 15.37 MW, lo cual permite a Sunway tomar un lugar en el Green500 en cuanto a su métreica de rendimiento/potencia.

##### a) Arquitectura Sunway

Como se menciona anteriormente, la supercomputadora Sunway usa un total de 40.960 procesadores SW26010 manycore de 64-bit RISC<sup>2</sup>, cada chip del procesador contiene 256 nucleos de procesamiento aparte de contar con 4 nucleos auxiliares destinados a la administración del sistema, esto para tener un total de 10.649.600 nucleos de CPU a través del sistema entero.

A modo de resumen que es más representativo cuenta con las siguientes características de acuerdo a su arquitectura:

- Sunway SW26010 260C a 1.45 GHz de 64-bit RISC con 260 nucleos.
- 40.960 procesadores con 10.649.600 nucleos.
- Memoria total de 1.310 TiB (Tebibyte 2<sup>40</sup> Bytes).
- Sistema Operativo Raise OS version 2.0.5 en base Linux.
- Precio Básico 1.8 millardos de Yuan (US\$273 millones).[13]

##### 2) GUANE-1

GUANE-1 es una plataforma dirigida especialmente al desarrollo e investigación, se encuentra ubicada físicamente en los Edificios de Investigación del Parque Tecnológico de Guatiguará.

##### a) Características de los nodos

Es un clúster compuesto por 16 nodos ProLiant SL390s G7

- 8 nodos con
  - 2 procesadores Intel(R) Xeon(R) CPU E5645 @ 2.40GHz. (12 Cores 2 Hebra)
  - 104 GB en RAM
  - 1 disco SAS de 200GB
  - 8 GPU Tesla M2075
- 3 nodos con
  - 2 procesadores Intel(R) Xeon(R) CPU E5645 @ 2.40GHz. (12 Cores 2 Hebra)

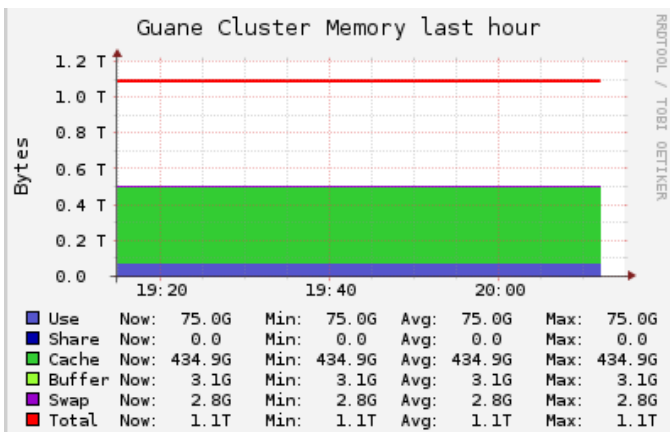
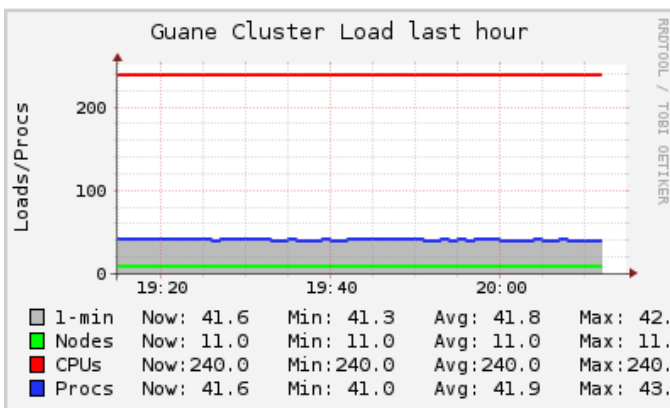
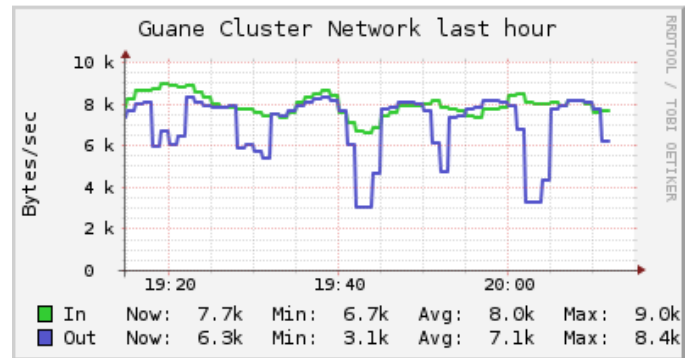
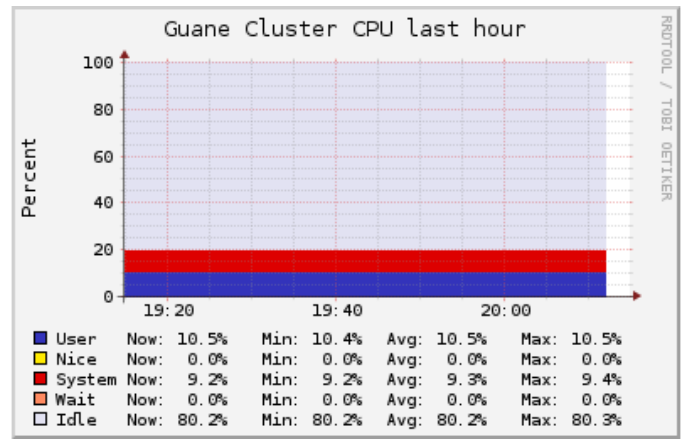


- 104 GB en RAM
- 1 disco SAS de 200GB
- 8 GPU Tesla S2050
- 5 nodos con
  - 2 procesadores Intel(R) Xeon(R) CPU E5640 @ 2.67GHz (8 Cores 2 Hebra)
  - 104 GB en RAM
  - 1 disco SAS de 200GB
  - 8 GPU Tesla S2050

b) Características de la Red

- 1 Red Giga Ethernet de Administración
- 1 Red 10Gbps
- 1 Red Infiniband[14]

c) Tablas estadísticas GUANE



E. Conclusiones SUNWAY vs GUANE-1

A modo de información se presentará la siguiente información extraída del top500 de supercomputo:

El top 10 de la lista sería:

- Sunway TaihuLight (China): 93.01 PFLOP/S
- Tianhe-2 (China): 33,86 PFLOP/S
- Titan (USA): 17,59 PFLOP/S
- Sequoia (USA): 17,17 PFLOP/S
- K computer (Japón): 10,51 PFLOP/S
- Mira (USA): 8,59 PFLOP/S
- Trinity (USA): 8,10 PFLOP/S
- Piz Daint (Suiza): 6,27 PFLOP/S
- HLRS (Alemania): 5,64 PFLOP/S
- Shaheen II (Arabia Saudi): 5,53 PFLOP/S

<sup>1</sup> .Es una medida del rendimiento del equipo, útil en los campos de los cálculos científicos; es una más exacta de las instrucciones generadas por segundo.

<sup>2</sup> .Reduced Instruction Set Computer, es un tipo de diseño de CPU generalmente usada en microprocesadores.



Habiendo entregado este dato técnico se puede proceder a dar unas breves conclusiones acerca de estos dos equipos de supercomputo expuestos en el documento:

- Respecto a la arquitectura podemos apreciar claramente que la computadora Sunway supera grandemente al nodo GUANE-1, no solo en potencia, sino tambien en velocidad de computo.
- El modulo GUANE a pesar de estar diseñado para la investigación, podemos apreciar claramente en su arquitectura que cuenta con nodos con características técnicas específicas para diferentes actividades.

## VI. REFERENCES/BIBLIOGRAFIA

- [1] <http://www.juntadeandalucia.es/servicios/madeja/contenido/recurso/220>
- [2] <https://conceptosarquitecturadecomputadoras.wordpress.com/computacion-paralela/>
- [3] [https://es.wikipedia.org/wiki/Computaci%C3%B3n\\_paralela#Taxonom.C3.ADA\\_de\\_Flynn](https://es.wikipedia.org/wiki/Computaci%C3%B3n_paralela#Taxonom.C3.ADA_de_Flynn)
- [4] <http://itcv-arquitectura-de-computadoras.blogspot.com.co/2014/11/42-tipos-de-computacion-paralela.html>
- [5] <http://archtectcomp.blogspot.com.co/2014/10/taxonomia-de-flynn.html>
- [6] [http://informatica.uv.es/iiguia/ALP/materiales/1\\_1\\_a\\_ComputacionParalela.pdf](http://informatica.uv.es/iiguia/ALP/materiales/1_1_a_ComputacionParalela.pdf)
- [7] <https://en.wikipedia.org/wiki/Supercomputer>
- [8] [https://es.wikipedia.org/wiki/CI%C3%BAster\\_\(inform%C3%A1tica\)](https://es.wikipedia.org/wiki/CI%C3%BAster_(inform%C3%A1tica))
- [9] [https://es.wikipedia.org/wiki/Computaci%C3%B3n\\_grid](https://es.wikipedia.org/wiki/Computaci%C3%B3n_grid)
- [10] [http://www.internetsano.gob.ar/archivos/cloudcomputing\\_empresas.pdf](http://www.internetsano.gob.ar/archivos/cloudcomputing_empresas.pdf)
- [11] [https://es.wikipedia.org/wiki/Computaci%C3%B3n\\_distribuida](https://es.wikipedia.org/wiki/Computaci%C3%B3n_distribuida)
- [12] [https://en.wikipedia.org/wiki/Supercomputer#Hardware\\_and\\_architecture](https://en.wikipedia.org/wiki/Supercomputer#Hardware_and_architecture)
- [13] <https://www.top500.org/news/new-chinese-supercomputer-named-worlds-fastest-system-on-latest-top500-list/>
- [14] <http://www.sc3.uis.edu.co/servicios/hardware/>

<sup>1</sup> .Es una medida del rendimiento del equipo, útil en los campos de los cálculos científicos; es una más exacta de las instrucciones generadas por segundo.

<sup>2</sup> .Reduced Instruction Set Computer, es un tipo de diseño de CPU generalmente usada en microprocesadores.