Universidad Central “Marta Abreu” de Las Villas.
Facultad Matemática, Física y Computación
Licenciatura en Ciencia de la Computación

TRABAJO DE DIPLOMA

Estudio de la factibilidad de la implementación de un Centro de Ciencias de un área específica en un cluster de computadoras.

Autor: **Fernando Ibargollín Fajardo.**

Tutor: **Lic. Romel Vázquez Rodríguez.**

SANTA CLARA

- 2008-

Hacemos constar que el presente trabajo fue realizado en la Universidad Central Marta Abreu de Las Villas como parte de la culminación de los estudios de la especialidad de Ciencias de la Computación, autorizando a que el mismo sea utilizado por la institución, para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos ni publicado sin la autorización de la Universidad.

_____

Firma del autor

Los abajo firmantes, certificamos que el presente trabajo ha sido realizado según acuerdos de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

_____ _____

Firma del tutor Firma del jefe del Laboratorio

Pensamiento

“El mundo camino hacia la electrónica… Todo indica que esta ciencia se constituirá en algo así como una medida del desarrollo; quien la domine será un país de vanguardia. Vamos a volcar nuestros esfuerzos es este sentido con audacia revolucionaria.”

Ernesto Guevara de la Serna.

Dedicatoria

A mi madre y a mi padre

A mi hermano y a mi abuela

A los que ayudaron a formar mi personalidad

A los que me quieren y me aceptan

Agradecimientos

A mi tutor Romel Vázquez por estar siempre disponible

A todos mis compañeros que son muy importantes para mi

A Donibell por su comprensión y apoyo

A mi familia que de una manera u otra forman parte de este logro

A todos mis profesores por la enseñanza brindada

A mis amigos

Resumen

El surgimiento de los centros de ciencia, donde se almacenan y manipulan grandes volúmenes de datos del orden de los peta bytes, compulsa el desarrollo de nuevas tecnologías, técnicas y escenarios para manipular estos datos. La visualización científica se ha desarrollado en los últimos años como un área importante de la Computación Gráfica, que simplifica el análisis, la comprensión y la comunicación de modelos, datos y conceptos en la ciencia y la ingeniería. Las diferentes técnicas de visualización proporcionan a los científicos representaciones visuales adecuadas para mostrar las correlaciones internas de los datos y constituyen una alternativa eficaz para el análisis visual de grandes volúmenes de datos. La concentración de datos de dominios específicos en centros de ciencia permite el acceso a ellos y su manipulación de forma masiva por científicos de todo el mundo, incrementando las posibilidades de intercambio de información y resultados. En este trabajo se presentan las principales características de los centros de ciencia de mayor relevancia en la actualidad, se estudian los elementos fundamentales de la visualización científica y en particular la visualización distribuida, se desarrolla un modelo de arquitectura de centro de ciencia basada en cluster de computadoras donde se integra la visualización científica con los centros de ciencia.

Abstract

The emerging of science centers, where huge data volumes of peta bytes order are stored and manipulated, collate the development of new technologies, techniques and sceneries to manipulate these data. The Scientific Visualization has been developed as an important area of the Computer Graphics, which simplifies the analysis, understanding and communication of models, data and concepts in science and engineering. Different visualization techniques give to the scientist the adequate visual representation for show the internal correlation of the data and it constitute an efficacious alternative for the visual analysis of huge data volumes. The collect of data of a specific domain in a science center allow the access to them and the massive manipulation by the scientist of the entire world, increasing the possibilities of information and results interchange. In this work is presented the main features of the more relevant science centers at present, the principal elements of scientific visualization and distributed visualization were studied, we develop an overall architecture for the creation of a cluster-based science center that allows the integration of scientific visualization with the science centers.

Índice

Introducción

La Visualización Científica (VC) ha sido un área de investigación de interés creciente en los últimos años, motivado fundamentalmente por el incremento constante de los volúmenes de datos, generados en muchos campos de aplicación.

El desarrollo de las redes de comunicación y el surgimiento de Internet han ampliado la colaboración entre los científicos del mundo en determinadas áreas de la ciencia, gracias al desarrollo del hardware unido a la acumulación del conocimiento, se están generando a diario volúmenes de datos tan grandes y complejos, que no pueden ser analizados suficientemente en forma numérica por simples computadoras personales. Estos volúmenes de datos del orden de los peta bytes requieren de un nuevo estilo de trabajo.

Han surgido los centros de ciencia que brindan acceso tanto a los datos como a las aplicaciones que los analizan para algunos dominios científicos. En (Gray, 2005) se mencionan algunos ejemplos de estos centros de ciencia como: *SDSS (Sloan Digital Sky Survey)* en Fermilab (Laboratorio Nacional del Acelerador Fermi), *BaBar* en SLAC (Stanford Linear Accelerator Center), *BIRN (Biomedical Informatics research Network)* en SDSC (San Diego Supercomputer Center), *Entrez-PubMed-GenBank* en NCBI (National Center for Biotechnology Information).

Formulación del problema

Las aplicaciones científicas se están expandiendo por toda la comunidad internacional, cada vez son más grandes los conjuntos de datos que necesitan manipular los investigadores, por lo que se requieren de nuevas tecnologías y métodos para lograr una efectiva y eficiente manipulación de estos datos. Por otra parte la capacidad de cómputo requerida para analizar estos grandes conjuntos de datos, supone que se utilicen arquitecturas paralelas para solucionar estos problemas, específicamente GRID o Cluster de computadoras.

Muchos de los centro de ciencia en el mundo están basados en tecnología GRID, esta tecnología es cara, necesita redes de alto ancho de banda para la comunicación entre los

distintos servicios distribuidos que se precisan para un centro de ciencia, y se plantea la necesidad de desarrollar los servicios que brindan los centros de ciencia utilizando una arquitectura paralela más barata y al alcance de empresas pequeñas que quieran obtener resultados similares con buenos rendimientos, este tipo de arquitectura pudieran ser los cluster de computadoras.

Preguntas de investigación

¿Cuáles aspectos del diseño de un Centro de Ciencia pueden ser soportados sobre la tecnología de cluster de computadoras?

¿Cómo se podrían implementar las funcionalidades de los centros de ciencia en un cluster de computadoras?

¿Cuál sería el modelo de una arquitectura para los servicios y aplicaciones de un Centro de Ciencia basado en cluster?

¿Cuales podrían ser las soluciones tecnológicas para satisfacer los requerimientos de un Centro de Ciencias basado en cluster?

Objetivo General

Determinar la factibilidad de implementar un centro de ciencias de un área específica en un cluster de computadoras.

Objetivos Específicos

- Determinar las tendencias actuales en el desarrollo de centros de ciencia de áreas específicas, teniendo en cuenta aspectos tales como tecnologías, arquitectura y servicios brindados.
- Precisar qué aspectos del diseño de un Centro de Ciencia pueden ser soportados sobre la tecnología de cluster de computadoras.

- Definir como implementar las funcionalidades de los centros de ciencia en un cluster de computadoras.

- Diseñar un modelo de arquitectura para los servicios y aplicaciones brindados por un Centro de Ciencia.

- Proponer un conjunto de soluciones tecnológicas que satisfagan los requerimientos del Centro de Ciencias.

Justificación de la investigación

Como se comentó anteriormente la gran mayoría de los centro de ciencia en el mundo están basados en tecnología GRID, esta tecnología es bastante cara, necesita redes de alto ancho de banda para la comunicación entre los distintos servicios distribuidos que se necesitan para un centro de ciencia. Los cluster de computadoras podrían ser una solución más barata para empresas pequeñas y medianas que quieran manipular grandes conjuntos de datos con buenos rendimientos.

En el centro de estudios de informática de la UCLV se cuenta con un cluster de computadoras, en el cual se pretende probar y validar modelos creados a partir de los grandes centros de ciencia del mundo, adaptados para ejecutarse en cualquier cluster de computadoras específicamente el de este centro.

Hipótesis de investigación:

- Es posible determinar e implementar las funcionalidades de los centros de ciencia en un cluster de computadoras, basado en las arquitecturas conocidas adaptándolas a los requerimientos necesarios y disponibilidad de recursos que se tienen en el cluster del centro de estudios de informática de la UCLV (CEI).

Estructura

Capítulo 1. **"Estudio de los aspectos fundamentales para la creación de un centro de ciencia para la manipulación y visualización de grandes volúmenes de datos"**

En este capítulo se tratan los principales elementos teóricos que intervienen en esta investigación, características de los centros de ciencia, las facilidades que brinda la visualización científica, y en particular la visualización distribuida. Con el objetivo de lograr una integración entre los CC y la visualización de grandes volúmenes de datos. Para esto es necesario estudiar las distintas tecnologías paralelas, específicamente GRID y cluster de computadoras, además de los distintos formatos de datos científicos y las BD paralelas.

Capítulo 2. **"Tendencias actuales de los centros de ciencia"**

En este capítulo se analizan algunos de los CC de mayor relevancia en la actualidad (SDSS, BaBar, Birn, GenBank), con el propósito de conocer el funcionamiento, los servicios que brindan, y elementos del diseño que nos faciliten la creación de un modelo de CC basado en cluster de computadoras.

Capítulo 3. **"Diseño del modelo general de un centro de ciencia sobre un cluster de computadoras"**

Finalmente en este capítulo se parte de un conjunto de requisitos que se deben tener en cuanta para la creación del CC, después se analiza que aspectos de los centros estudiados son soportados sobre la tecnología cluster de computadoras. Luego se crea un diseño lógico basado en capas y finalmente se brinda un conjunto de soluciones tecnológicas para la implementación futura del centro.

1. "Estudio de los aspectos fundamentales para la creación de un centro de ciencia orientado a la visualización de grandes volúmenes de datos"

El surgimiento de centros de ciencia, donde se almacenan y manipulan grandes volúmenes de datos del orden de los peta bytes, compulsa el desarrollo de nuevas tecnologías, técnicas y escenarios para manipular estos datos. La visualización científica se ha desarrollado en los últimos años como un área importante de la computación gráfica, que simplifica el análisis, la comprensión y la comunicación de modelos, datos y conceptos en la ciencia y la ingeniería. Las diferentes técnicas de visualización proporcionan a los científicos representaciones visuales adecuadas para mostrar las correlaciones internas de los datos y constituyen una alternativa eficaz para el análisis visual de grandes volúmenes de datos. La concentración de datos de dominios específicos en centros de ciencia permite el acceso a ellos y su manipulación de forma masiva por científicos de todo el mundo, incrementando las posibilidades de colaboración e intercambio de información y resultados. En este capítulo se estudian las características generales de los centros de ciencia así como los requisitos de los datos almacenados en estos para poder ser visualizados eficientemente. Se hace una descripción general de las principales tecnologías paralelas utilizadas para la creación de los centros de ciencia.

Se presenta una clasificación de métodos y diversos algorítmicos que pueden usarse para visualizar y renderizar estos grandes conjuntos de datos. Esta clasificación provee la construcción de bloques para una solución que se basa en técnicas utilizadas en el diseño de los algoritmos paralelos. Estas técnicas enfocan su atención en el nivel de los sistemas y se basan en la descomposición paralela de los datos y / o las tareas requeridas. Para caracterizar esta definición de datos de gran escala, se consideran sólo situaciones en las cuales los conjuntos de datos no pueden ser procesados por una sola computadora. Además según (Hansen and Johnson, 2005, Heijmans, 2002) se trata un conjunto de técnicas usadas para la visualización, específicamente las cuatro técnicas más usadas para solucionar el

problema de visualización de datos de gran escala: Flujo de datos, Paralelismo de tareas, Paralelismo de tubería y Paralelismo de datos.

También se estudian los principales formatos de datos científicos, (HDF, NetCDF, FITS...), teniendo en cuenta sus facultades para procesamiento paralelo así como las base de datos paralelas como otra alternativa para el almacenamiento y procesamiento de datos científicos.

1.1 Centros de ciencia

La demanda de herramientas y recursos computacionales para realizar análisis de datos científicos crece continuamente, incluso más rápido que los propios volúmenes de datos. Lo cual es consecuencia de 3 fenómenos:

1. Los nuevos y sofisticados algoritmos consumen cada vez más instrucciones para analizar cada byte de dato.

2. Muchos algoritmos de análisis son súper lineales, a menudo de orden O(N2) o O(N3).

3. El ancho de banda de entrada-salida no ha mantenido el mismo ritmo de avance que la capacidad de almacenamiento. En la última década mientras la capacidad ha crecido más de 100 veces, el ancho de banda solo lo ha hecho 10 veces.

Estos 3 fenómenos provocan que el análisis de los datos necesite mucho más tiempo. Para mejorar estos problemas los científicos necesitarán mejores algoritmos de análisis, que puedan manejar inmensos conjuntos de datos con algoritmos aproximados (unos con tiempo de ejecución cercano al lineal), y según (Gray, 2005) serán necesarios algoritmos paralelos, que puedan utilizar muchos procesadores y discos para equiparar las demandas de densidad de CPU con el ancho de banda.

Los centros de ciencia o centros de datos científicos, que están surgiendo como estaciones de servicio para algunos dominios científicos y suministran acceso tanto a los datos como a

las aplicaciones que los analizan, se presentan como una vía para solucionar estas crecientes demandas de recursos computacionales.

Conjuntos de datos del orden de los peta bytes ocupan de 1000 a 10 000 discos, y requieren de miles de unidades de procesamiento. Un centro de ciencia es capaz de almacenar uno o más de estos grandes conjuntos de datos y ofertar aplicaciones que acceden a ellos y los analizan. Posee además, un equipo de personas que entiende los datos y está constantemente mejorándolos y agregando mas información.

El nuevo estilo de trabajo en estos dominios científicos es enviar solicitudes a las aplicaciones que están corriendo en el centro de datos, y recibir de vuelta las respuestas. Esto es mejor que hacer una copia de los datos hacia el servidor local para su posterior análisis con recursos propios.

Una tendencia actual que soportan muchos centros de ciencia es ofertar la posibilidad de almacenar espacios de trabajos personales y guardar allí las respuestas a las consultas realizadas. Esto minimiza el movimiento de los datos, y permite la colaboración entre grupos de científicos que hacen análisis sobre temas similares. Estos espacios de trabajo son también una vía para la colaboración entre grupos de analizadores de datos. Según (Gray, 2005) a largo plazo los espacios de trabajos personales de un centro de datos pueden convertirse en una forma para la publicación de datos, mostrando tanto los resultados científicos de un experimento o investigación, como los programas usados para generarlos.

Los centros de ciencia proporcionan herramientas y soporte para permitir la colaboración con otros centros de ciencia. Cuando un científico desea consultar datos de varios centros de ciencia, no quedaría otra opción que mover partes de los datos de un lugar a otro. Si esto fuera algo común, los centros de datos involucrados deben colaborar para suministrar la salva de los datos de manera mutua ya que el tráfico entre ellos justifica disponer de las copias.

Muchos científicos prefieren hacer gran parte de su análisis en los centros de datos, ya que les ahorrará tener que administrar datos locales y granjas de computadoras. Algunos

científicos pueden llevar extractos de datos a sus casas para hacer procesamiento, análisis y visualización de manera local. Pero esto será posible hacerlo en el centro de datos utilizando el espacio de trabajo personal. Como se mencionó anteriormente los conjuntos de datos del orden de los peta bytes requerirán de 1 000 a 10 000 discos y miles de nodos. En cualquier momento algunos de los discos o nodos pudieran colapsar. Estos sistemas tienen que tener un mecanismo para evitar la pérdida de datos, y proporcionar buena disponibilidad incluso con menos de la configuración completa por lo que requieren un sistema de auto recuperación. Por lo que se hace necesario disponer de un sistema de replicación tratado en (Özsu, 2005) para replicar los datos del centro de ciencia en distintos lugares geográficos.

Para el análisis científico va a ser de gran importancia la utilización de 3 técnicas avanzadas: (1) Uso de grandes metadatos y metadatos estándares para facilitar el descubrimiento de los datos que existen, que les sea cómodo tanto para las personas como para los programas entender los datos y se pueda saber fácilmente todos los procesos por los que han pasado. (2) Buenas herramientas de análisis que les permitan a los científicos de forma fácil hacer las consultas, entender y visualizar las respuestas. (3) Permitir el acceso paralelo a los datos soportados por los nuevos esquemas indexados y los nuevos algoritmos que nos permitan interactuar y explorar conjuntos de datos del orden de los peta bytes.

Como consecuencia, los centros de ciencia se mantendrán como la principal vía para la comunicación de datos e información entre las asociaciones mundiales y suministraran tanto los datos como la infraestructura para manipular, analizar, procesar, visualizar y publicar estos archivos de gran escala así como los algoritmos y herramientas.

1.1.1 Aspectos generales de los centros de ciencia.

Según (Moore, 2006), hay que tener en cuenta un conjunto de aspectos que son fundamentales para el funcionamiento de los centros de ciencia, entre estos requisitos se presentan:

Disponer de tecnologías de punta para el almacenamiento de los datos, está tecnología debe permitir aumentar la capacidad de almacenamiento en cualquier momento que se necesite. El tamaño de los datos actualmente es del orden de los tera bytes, y el número de objetos digitales se pueden medir en decenas de millones de millones de archivos, pero estos datos van aumentando de forma considerable cada año.

Los datos científicos son típicamente distribuidos a través de múltiples sitios, ya sea durante el proceso de generación, o durante el proceso de análisis. Se necesita una infraestructura independiente de designación de espacios de nombres lógicos para identificar los archivos. Cuando un archivo se mueve entre los sitios, su nombre lógico nunca cambia.

Los términos y conceptos utilizados por una disciplina científica no describen las características de los datos creados por otra disciplina académica. Cada disciplina diseña su propio metadato descriptivo. Los mecanismos de administración de la latencia son necesarios para reducir al mínimo el número de mensajes enviados a través de la red, y minimizar la sobrecarga asociada con la transmisión de datos.

Hasta que los datos no están calibrados y verificados, se limita el acceso a la colección a los miembros del equipo. El proceso de publicación es una afirmación por un equipo científico de que los datos son una representación exacta del mundo real. La publicación de los datos puede estar sujeta a la misma revisión por colegas como los procesos utilizados para la publicación de artículos científicos. Cada comunidad científica normalmente aplica una codificación estándar diferente para optimizar la capacidad de manipular sus estructuras de datos.

En el capítulo 2 se describe como los principales centros de ciencia del mundo manejan estas características para lograr un alto aprovechamiento de todos los recursos que ellos disponen y facilidades que brindan a los usuarios.

1.2 Tecnologías paralelas utilizadas en la gestión de grandes volúmenes de datos

La capacidad de cómputo requerida por los centros de ciencia para analizar estos grandes conjuntos de datos supone que se utilicen arquitecturas paralelas para solucionar estos problemas, específicamente GRID o Cluster de computadoras, a continuación se tratan las principales características de ambas tecnologías, ventajas y desventajas .

1.2.1 Computación Grid

Se llama GRID según (Boghosian, 2005) al sistema de computación distribuido que permite compartir recursos no centrados geográficamente para resolver problemas de gran escala. Los recursos compartidos pueden ser ordenadores (PCs, estaciones de trabajo, supercomputadoras, PDA, portátiles, móviles, etc), software, datos e información, instrumentos especiales (radios, telescopios, satélites etc), personas o colaboradores…

La computación Grid ofrece muchas ventajas frente a otras tecnologías alternativas. La potencia que ofrece una multitud de computadoras conectadas en red usando tecnología Grid es prácticamente ilimitada, además de que ofrece una perfecta integración de sistemas y dispositivos heterogéneos, por lo que las conexiones entre diferentes máquinas no generarán ningún problema. Se trata de una solución altamente escalable, potente y flexible, ya que evitarán problemas de falta de recursos (cuellos de botella) y nunca queda obsoleta, debido a la posibilidad de modificar el número y características de sus componentes. En (Moore and Jagatheesan, 2004, Paul Watson and Paton, 2003, Rajasekar et al., 2003) se puede profundizar sobre estos aspectos.

Los datos deben compartirse entre miles de usuarios con intereses distintos. Se deben enlazar los principales centros de súper computación de todo el mundo. Es necesario asegurar que los datos sean accesibles en cualquier lugar y en cualquier momento. Los sistemas GRID deben armonizar las distintas políticas de gestión de muchos centros diferentes, así como proporcionar la seguridad de los datos y aplicaciones que analizan los mismos.

Los principales inconvenientes tratados en (Boghosian, 2005) provienen de la dificultad para sincronizar los procesos de todos estos equipos, monitorizar los recursos, asignar la carga de trabajo y establecer políticas de seguridad fiables, por lo que el costo sería inalcanzable para pequeñas y medianas empresas.

1.2.2 Cluster de computadoras

Un cluster de computadoras es un grupo de equipos independientes conectados mediante una red local, que ejecutan una serie de aplicaciones de forma conjunta y aparecen ante los clientes y aplicaciones como un solo sistema, tienen memoria RAM, discos que pueden ser compartidos y procesadores similares, estrechamente acopladas, que pueden ser vistas como una sola unidad.

Los componentes del cluster están comúnmente conectados entre ellos por redes locales de alta velocidad. Su funcionamiento según (Özsu, 2005, Shankar, 2006) puede ser similar al de una supercomputadora pero es mucho más asequible, porque los procesadores no tienen que ser tan potentes, la fuerza está dada por la paralelización y el trabajo compartido entre las diferentes unidades de cómputo.

Los cluster de altos rendimientos de cómputo son implementados principalmente para aumentar la eficiencia de procesamiento, así como disminuir el tiempo de ejecución, repartiendo una tarea computacional a través de los diferentes nodos que componen el cluster, esta tecnología tiene gran aplicación en el área de la visualización científica y es una alternativa factible para las empresas medianas o pequeñas.

La comunicación mediante el paso de mensajes como paradigma de programación tratada en (Alonso, 1997), requiere que el paralelismo se exprese de forma explicita por el programador, responsable de analizar su algoritmo o aplicaciones, para identificar vías por las cuales puede repartir la carga de trabajo, entre los múltiples procesadores del cluster de la forma más eficiente posible, además de poder extraer concurrencia. Como resultado la programación mediante el paradigma de paso de mensajes, suele ser una actividad dura,

consumidora de tiempo, y de alta demanda intelectual. De otro lado, las aplicaciones diseñadas por paso de mensajes tienden a conseguir buenas prestaciones, y mantenerlas cuando el sistema aumenta a números grandes de tareas y/o procesadores.

Entre otras, algunas de las interfaces de los sistemas basados en paso de mensajes más utilizados son: PVM, MPI, BSP, Parmacs, P4, y Express, todas en pleno desarrollo. (Alonso, 1997)

1.3 Visualización científica de grandes volúmenes de datos

La visualización científica (VC) según (Hansen and Johnson, 2005, Morell and Pérez, 2006) significa encontrar una representación visual apropiada para un conjunto de datos, que permita mayor efectividad en el análisis y evaluación de los mismos. Simplifica el análisis, comprensión y la comunicación de modelos, conceptos y datos en la ciencia y la ingeniería.

Los softwares de visualización científica de propósito general, permiten a los científicos hacer programas visuales mediante la unión de diversos módulos en una red empleando un paradigma de programación visual. Los centros de ciencia también brindan estos servicios de forma tal que mediante un conjunto de componentes y relaciones entre ellos se pueden definir procesos complejos en un simple bloque de construcción, a este tipo de construcción de procesos complejos a partir de componentes y servicios interrelacionados se les conoce según (Shankar, 2006) como WorkFlow.

La Visualización Científica ofrece grandes ventajas sobre otros métodos de análisis de datos, permitiendo representar datos de varias dimensiones o variables, logrando visualizar cuatro o más variables al mismo tiempo.

Las técnicas y tecnologías actuales permiten representar sin mucha dificultad datos no homogéneos o que no se conozca detalladamente su estructura. La exploración visual es intuitiva, no requiere de complicados conocimientos matemáticos, estadísticos o de otra

índole. Otra gran ventaja de la VC es la gran cantidad de conocimiento que puede ser rápidamente extraída y consecuentemente con esto es necesario poseer gran capacidad de cómputo para obtener los resultados en un tiempo asequible, además de tener los datos bien organizados y ubicados, por lo que los centros de ciencia surgen como una solución viable, disminuyendo en gran medida el tráfico en las redes. Todo el análisis puede ser realizado en los centros sin necesidad de mover los datos a espacios personales, con la publicación de los resultados es posible eliminar la redundancia.

Debido al tamaño de los conjuntos de datos, conviene usar la computación paralela en el centro de ciencia para realizar los cálculos de visualización y transferir las imágenes más pequeñas en vez de los datos crudos a las computadoras personales de los científicos. Se trata el tema de la visualización distribuida, principalmente la visualización orientada a red. La estructura de los sistemas de visualización, las distintas funcionalidades de los sistemas por parte del cliente y del servidor, también se profundiza en el renderizado paralelo.

1.4 Visualización distribuida

Comenzando con el tratamiento de los principales sistemas de visualización distribuidos existentes tratados en (Heijmans, 2002). Analizándolos por categorías, a partir de sus principales características, además de crear una clasificación general de sistemas de visualización distribuidos.

¿Qué es la visualización distribuida? Primero se menciona una simple definición operacional de visualización: "Usar una computadora para darle al usuario una imagen de un conjunto de datos". Ahora se puede entender una definición de visualización distribuida: "Utilizando una o más computadoras para darle a uno o más usuarios una imagen o representación visual de un conjunto de datos"(Heijmans, 2002). Hay varias formas de realizar este mecanismo, una visión general de los distintos métodos se puede observar en la figura1.1, donde se usa el termino CPU cuando se refiere a una computadora.

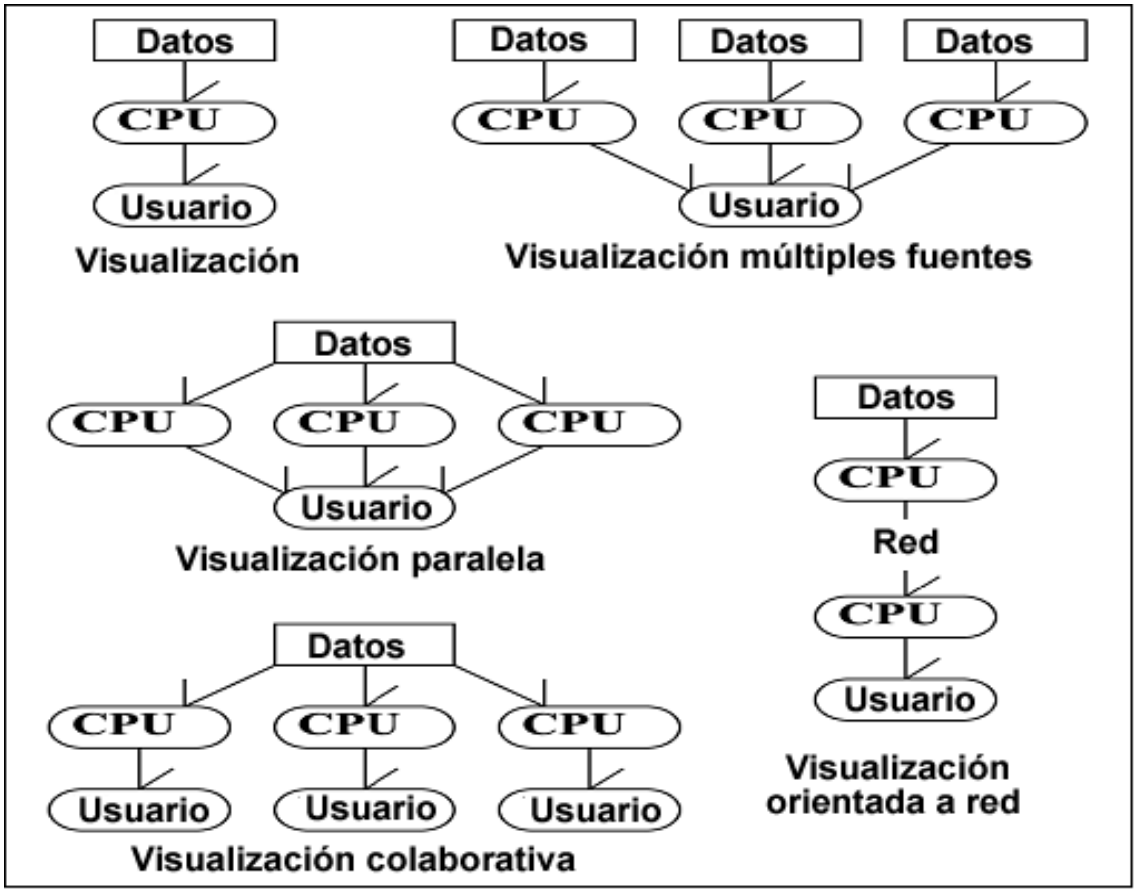


Figure 1.1. Vision general de los tipos de visualización distribuida.

La visualización paralela a menudo requiere un alto costo computacional, principalmente dado por el alto tiempo necesario para realizar una tarea, pero gracias al avance de las tecnologías se puede disminuir este tiempo utilizando múltiples CPU's conectadas en red.

La visualización colaborativa puede ser muy útil para facilitar el trabajo de un equipo de personas simultáneamente colaborando entre ellos, donde cada miembro del equipo tiene a la vista la misma visualización y ésta es dirigida por uno de ellos. Esta forma de trabajo elimina considerablemente la redundancia y aumenta la calidad de los resultados.

Con la visualización a partir de múltiples fuentes se que puede combinar resultados de múltiples conjuntos de datos en una misma imagen.

La visualización orientada a la red permite al usuario crear una visualización de datos que residen en cualquier lugar fuera de su computadora.

En este trabajo se centra la atención en la visualización orientada a la red y la visualización paralela. La visualización orientada a red no es más que un sistema de visualización en el cual los datos y la imagen residen en computadoras diferentes, conectadas por una red, ver en (Heijmans, 2002). Los datos y una parte del sistema de visualización se encuentran en el lado del servidor, mientras que la imagen y el resto del sistema de visualización están en el lado del cliente. La idea fundamental es que los resultados de la visualización siempre son enviados del servidor al cliente, la mayor parte de la visualización se realiza en el lado del servidor.

La estructura de un sistema de visualización puede ser comparada fácilmente con la estructura general de un sistema de información. La base consta de datos y metadatos descriptivos. El usuario tiene una interfaz mediante la cuál puede observar los metadatos y los datos, y con la cuál pueden controlar las rutinas de fondo que actúan sobre los datos. En un sistema de visualización al motor de visualización se le llama rutina de fondo. También conviene dividir la interfaz en dos partes (interfaz grafica del usuario (GUI) como indican sus siglas en ingles) y (observador). Esta estructura se puede entender mejor en la Figura1.2.

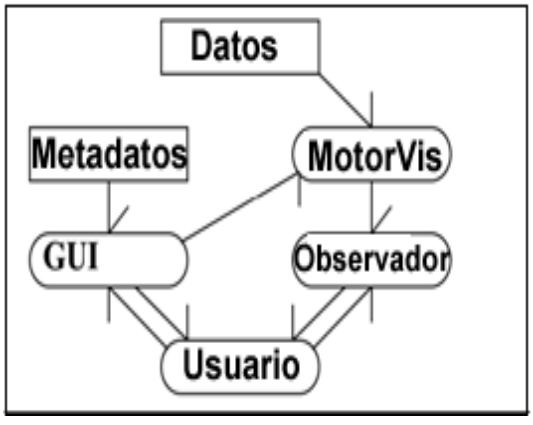


Figura 1.2 Control de flujo y visualización de datos.

El usuario interactúa con la GUI para mirar los metadatos y seleccionar los parámetros del motor de visualización. Los datos son transformados por el motor de visualización y mostrados a la vista del usuario por el observador.

A partir de ahora cuando se hable de visualización siempre se refiere a visualización distribuida y para abreviar se decidió llamar al motor de visualización del lado del servidor como (SVE) y al motor de visualización del lado del cliente como (CVE). La división no tiene que ser simétrica. Tal vez el SVE sólo envía los datos, mientras el CVE realiza la visualización, o el SVE hace todo el trabajo y el CVE sólo muestra las imágenes de información geométrica o 2D.

En la figura1.3 se muestra la estructura de un sistema de visualización distribuido ilustrando el flujo de metadatos a través de él. El servidor le envía los metadatos al cliente, y estos son mostrados al usuario por la GUI.

Los metadatos tienen gran importancia, porque permiten al usuario escoger correctamente los datos y determinar los parámetros que se deben tener en cuanta para la visualización. Sin embargo muchos sistemas de visualización no tienen esto en cuenta o no especifican que metadatos están disponibles.

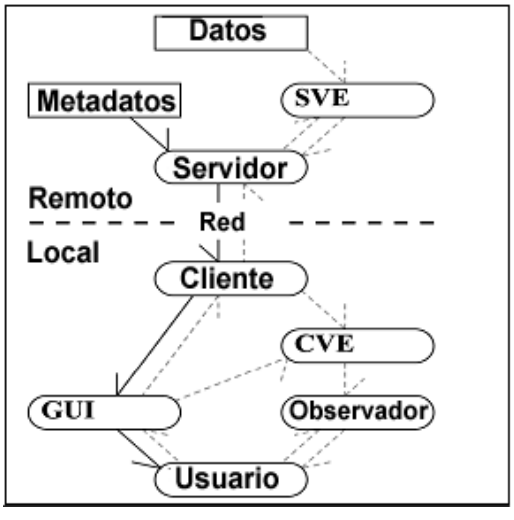


Figure 1.3 Flujo de metadatos

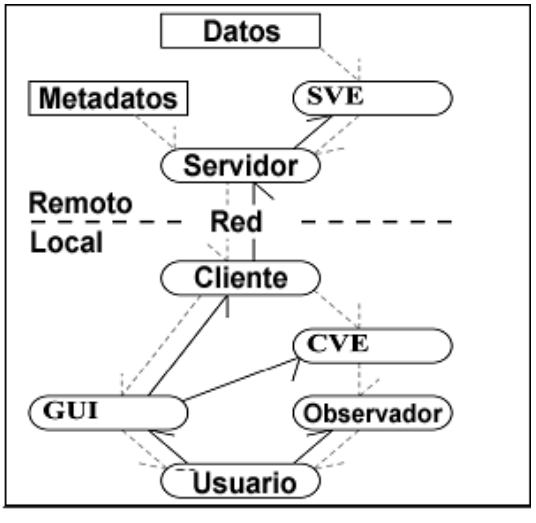
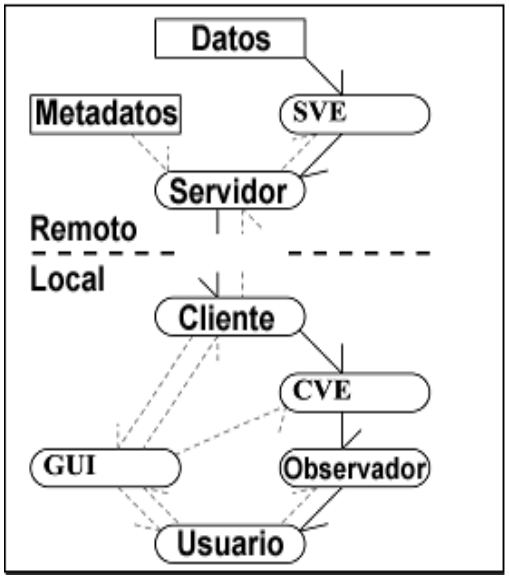


Figura 1.4 Control de flujo

En la figura1.4 se describe el control de flujo de un sistema de visualización. El usuario controla el observador, al CVE y al SVE a través del GUI. El usuario puede girar y escalar los objetos en el monitor y puede establecer los parámetros de visualización para el CVE y el SVE.



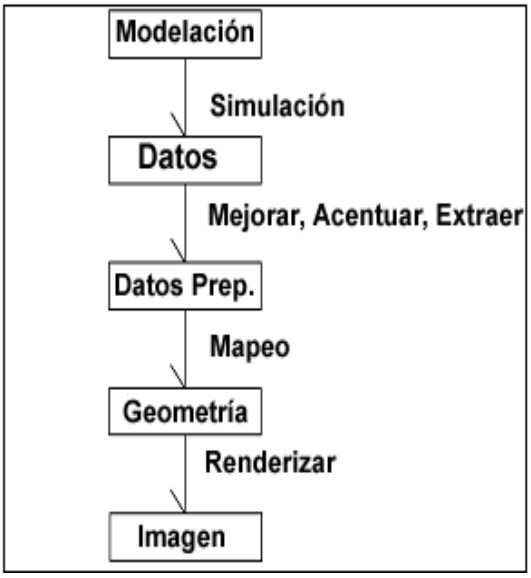


Figura 1.5. Flujo de datos **Figure 1.6. Tubería de Visualización.**

En la figura1.5 se muestra el flujo de datos a través del sistema. Los datos son procesados por el SVE, enviados del servidor para el cliente, procesados por el CVE y exhibidos por el observador para el usuario.

También se puede ilustrar el flujo de datos mediante la tubería de visualización en la figura1.6, para observar y entender mejor se puede observar lo que entra y sale de las diferentes partes del sistema. La entrada del SVE son los datos, la salida del CVE es la geometría 2D/3D, lo cual es también la entrada del observador. La salida del observador es una imagen 2D o una forma especial de imagen 3D siempre que esta sea soportada por el hardware.

La elección recae sobre la salida del SVE, como esta es enviada sin alteración sobre la red, también es la entrada del CVE. Es fácil concluir que para la tubería de visualización hay

tres opciones: que envíe; los datos, los datos preparados o la geometría 2D/3D. Esto también define la distribución de funcionalidad entre el servidor y el cliente. Si todos los datos son enviados por la red, el SVE no hace nada, mientras el CVE realiza la visualización. Si la geometría 2D/3D es enviada, el CVE no hace nada, la visualización es realizada por el observador.

Entonces, se presentan dos criterios importantes para definir la estructura de un sistema de visualización distribuido:

¿Qué es enviado por la red, o qué se hace por el servidor y qué por el cliente?

¿Qué control tiene el usuario sobre el proceso de visualización?

Cuando los datos están en un formato que el SVE no puede comprender directamente es necesario la conversión de los mismos para poder realizar la visualización, por lo cual se debe disponer de una utilidad de conversión o se debe usar un modulo especial para importar los datos en el motor de visualización, por razones de brevedad se decide no incluir los pasos de conversión en los diagramas, Conceptualmente cabe colocar la conversión de datos dentro de la fase de preparación de los datos. En ese caso no hay necesidad de distinguir conversión de datos de otros pasos de la visualización.

Hay tres niveles básicos de control: el usuario controla al observador, éste es el nivel más bajo de interacción. El usuario puede controlar sólo su punto de vista, pero no tiene el control de la visualización. Por consiguiente la visualización puede llamarse apenas interactiva; el usuario controla la visualización en el lado del cliente, si se envían todos los datos al cliente, la interacción se realiza según las necesidades del usuario, él puede cambiar la tubería de visualización y los parámetros de la visualización y por ultimo el usuario controla la visualización en el lado del servidor, si no se envían los datos al cliente, sino una selección o incluso el resultado de una visualización, el usuario debe poder controlar la visualización del servidor mediante un formulario ofrecido para que seleccione los parámetros según su interés.

1.5 Tubería de Renderizado Paralelo

Un proceso paralelo típico de renderizar un volumen consta de cuatro pasos. El paso de lectura de los datos en disco y la distribución para los procesadores. Cada procesador recibe un subconjunto del volumen de los datos. En el siguiente paso de renderizar, cada procesador renderiza el subvolumen asignado en una imagen 2D parcial que es totalmente independiente de los demás procesadores. Después, un paso de combinación, que generalmente requiere comunicación entre los procesadores, donde se mezclan las imágenes 2D parciales para formar la imagen 2D final. El paso final le muestra la imagen al usuario o esta es almacena para su análisis posterior.

Dado un volumen genérico que debe ser renderizado en paralelo y una máquina de P-Procesadores, según (Hansen and Johnson, 2005), hay tres formas posibles de administrar los procesadores para renderizar los volúmenes de datos, teniendo en cuenta el tiempo necesario para obtener la imagen final. El primer método simplemente es correr el volumen que se va a renderizar como una secuencia de subconjuntos uno tras otro. En cualquier punto de tiempo, la máquina entera de P-Procesadores está dedicada al renderizado de un volumen particular. El paralelismo solo se asocia con el renderizado de un solo volumen de datos, en esencia solo es explotado el paralelismo intravolumen.

El segundo método toma la estrategia opuesta exacta, renderizado de P volúmenes de datos simultáneamente, cada uno en un procesador. Este modus operandi sólo explota paralelismo de intervolumen, y está limitado por el espacio principal de memoria de cada procesador.

Para lograr un renderizado óptimo, e deben balancear dos factores en el funcionamiento, el uso eficiente de los recursos y la paralelización; Esto sugiere explotar ambos tipos de paralelismo, el paralelismo de intravolumen y el paralelismo de intervolumen. En lugar de usar todos los procesadores para el renderizado colectivamente de un volumen a la vez, se forma un proceso de renderizado de tubería dividiendo los procesadores en grupos de renderizado de múltiples volúmenes concurrentemente. De este modo según (Hansen and Johnson, 2005), el tiempo global de renderizado puede ser minimizado grandemente,

porque las tareas de la tubería de renderizado son superpuestas con la E / S requerida para cargar cada volumen en un grupo de procesadores.

El tercer método es así un híbrido, en el cual los nodos de los P-procesadores están subdivididos en L-grupos, L esta entre 1 y P, cada uno con un volumen de renderizado. La elección óptima de L generalmente depende del tipo y la escala de la máquina paralela así como del tamaño del conjunto de datos. La estrategia óptima de partición de discos para minimizar el tiempo global de renderizar puede ser caracterizada con el modelo de funcionamiento y revelada con un estudio experimental, lo cual demuestra que el tercer método ciertamente es el mejor de los tres.

Ahora se tratan los distintos métodos y diversos algorítmicos que pueden usarse para visualizar y rederizar estos grandes conjuntos de datos

1.5.1 Flujo de Datos

Flujo de Datos (FD) tratado en (Hansen and Johnson, 2005), es el método más usado para procesos independientes de subconjuntos de datos de gran escala. Un subconjunto en un instante de tiempo. Éste a menudo es el único método factible en situaciones donde el tamaño de un conjunto de datos excede la capacidad de los recursos disponibles en cuanto a memoria y espacio de intercambio. Por ejemplo, no es raro para conjuntos de datos científicos, especialmente series de tiempo, que fácilmente exceda la cantidad de memoria disponible. La ventaja crucial de este método es que todos los conjuntos de datos de cualquier tamaño pueden ser tratados exitosamente. El inconveniente de esta técnica es que a menudo requiere largo tiempo de ejecución y no tiene prevista la exploración interactiva de los datos. El uso de un subsistema de E/S de alto rendimiento específico es ventajoso para mejorar el funcionamiento global de los algoritmos de flujo.

1.5.2 Paralelismo de tareas

Con Paralelismo de Tarea, módulos independientes de una aplicación se ejecutan paralelamente. En la figura 1.7, esto es logrado haciendo que los módulos A, B, y C se

ejecuten al mismo tiempo. Requiere un algoritmo que resuelva tareas independientes y tener disponible múltiples recursos computacionales. La ventaja crucial de esta técnica es que permite ejecutar tareas de visualización paralelamente a múltiples porciones de datos. La principal desventaja de esta técnica es que el número de tareas independientes que pueden ser tratadas tiene que ser igual al número de CPU's disponibles. Además, puede ser difícil por el balance de carga de las tareas, y por eso a menudo puede ser muy desafiante para aprovechar los recursos disponibles.

El paralelismo de tarea es usado eficazmente en la industria cinematográfica, donde varios marcos en una producción animada son renderizados en paralelo. Las elecciones específicas del hardware para mejorar el funcionamiento del paralelismo de tarea depende de los detalles de las tareas requeridas, analizadas en (Hansen and Johnson, 2005).

1.5.3 Paralelismo de Tubería

El Paralelismo de Tubería ocurre cuando un número de módulos en una aplicación se ejecutan paralelamente pero en subconjuntos independientes de datos. En la figura 1.7, esto ocurriría cuando los módulos A , D, y E todos le hacen una operación a porciones independientes de datos. Este método es más conveniente para situaciones donde hay múltiples tareas heterogéneas. La ventaja de esta estrategia es que permite uso paralelo de los recursos de cómputo. Por ejemplo, un proceso puede estar leyendo de disco, otro proceso computando resultados utilizando al CPU, y un tercer proceso usando una tarjeta aceleradora de gráficos. La desventaja principal es que puede ser difícil balancear el tiempo de ejecución requerido por las distintas etapas; En una tubería desnivelada, la etapa más lenta directamente afecta la función global. Además, la longitud de la tubería directamente limita la cantidad de paralelismo que puede ser logrado, y se deben tener tantos procesadores disponibles como etapas halla en la tubería. Para minimizar el tiempo de ejecución, hay que mover rápidamente los datos de una etapa de la tubería a la siguiente. Esta estrategia requiere el uso de arquitecturas de memoria compartida y una red de interconexión de alta velocidad entre los procesadores.

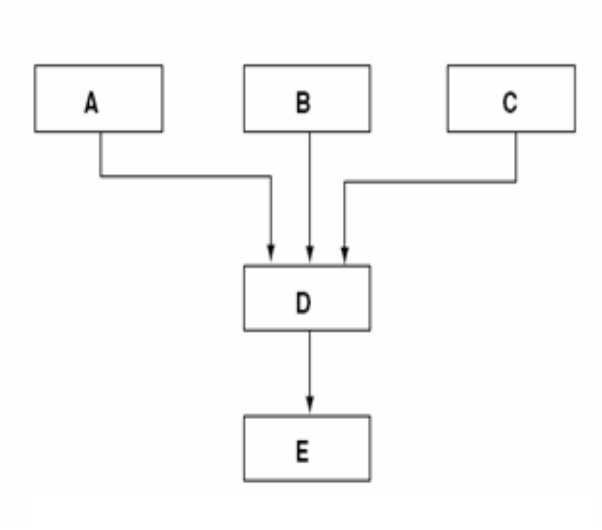


Figura 1.7. Módulos de ejecución.

1.5.4 Paralelismo de datos

Mediante Paralelismo de Datos, el código dentro de cada módulo de una aplicación se ejecuta paralelamente. En Lo Referente a la figura 1.7, esto ocurre cuando el código dentro del módulo A corre en paralelo. Esto requiere que el conjunto de datos sea subdividido en múltiples procesos que corren el mismo algoritmo en las partes resultantes concurrentemente. El paralelismo de datos puede ser implementado como una extensión de la técnica de descomposición de datos descrita en la sección de flujo de datos. En este caso, los datos se subdividen de la misma forma, pero se tiene un paso extra de asignar un procesador para cada una de las partes resultantes. Este método es comúnmente llamado modelo de programa simple de múltiples datos, porque cada proceso ejecuta el mismo programa en subconjuntos diferentes de datos. La principal ventaja de este método es que se puede lograr un alto grado de paralelismo; Las soluciones tienden a escalarse adecuadamente aumentando el número de procesadores. Cuando hay un gran número de procesadores disponibles, este método es a menudo uno de los mejores para lograr alto rendimiento. Un inconveniente posible es que la dimensionalidad puede estar limitada por el costo de la comunicación entre los procesos. Para lograr el mejor funcionamiento posible con paralelismo de datos, es a menudo importante considerar los costos de comunicación y

localidades de datos entre los procesadores. En las mejores situaciones posibles, no hay dependencia entre los procesadores, y si pasa lo peor cada procesador está obligado a compartir información con otro. Afortunadamente, muchos algoritmos de visualización tienen pocas dependencias de comunicación entre procesadores, y por eso el paralelismo de datos es a menudo una de las técnicas más efectivas para lograr alto funcionamiento, como se explica en (Hansen and Johnson, 2005).

1.5.5 Resumen

Las secciones precedentes han esbozado una clasificación de técnicas basadas en la descomposición de tareas y datos. La tabla mostrada en la figura 1.8, presenta un mapeo de las características del problema y la solución soportada. La columna de tamaño del conjunto de datos describe la cantidad de datos a visualizar. Un conjunto de datos es considerado grande cuando excede los recursos de una sola máquina. La columna de tareas describe el tipo de trabajo. Las tareas homogéneas es un mismo tipo de trabajo que se le aplica a diferentes datos. Las tareas independientes pueden ser corridas al mismo tiempo en paralelo. Las tareas secuenciales deben correr una tras otra en un orden fijo. La columna de recursos describe el número de recursos disponibles, y la columna de solución identifica la técnica que debe usarse en cada caso. Por ejemplo, si usted tiene un conjunto de datos de gran escala y sólo una CPU, flujo de datos quizás sea la única solución posible por la cual se pueden explorar la totalidad de los datos.

Tamaño de Datos	Tareas	Recursos	Solución
Grandes	Cualquiera	Una CPU	Flujo de Datos
Grandes	Homogénea	Múltiples CPUs	Paralelismo de datos
Cualquiera	Independiente	Múltiples CPUs	Paralelismo de tareas
Cualquiera	Secuencial	Múltiples CPUs	Paralelismo de Tubería

Figura 1.9 Resumen de las técnicas de solución.

1.6 Tecnologías para el almacenamiento de datos científicos

Los archivos secuenciales sin lugar a dudas resultan eficientes en el almacenamiento de datos científicos y permiten procesar y almacenar estos datos de forma relativamente sencilla.

Los científicos demandaban para su actividad un método eficiente e independiente para la lectura y escritura de datos científicos que permitiera desarrollar bibliotecas específicas para el almacenamiento de los datos científicos, una de las implementaciones más usadas son los archivos binarios con formatos estándares. Según (Gray, 2005), existe una tendencia de lo formatos de datos científicos (HDF, NetCDF, FITS,..) a centrarse principalmente en aspectos de los metadatos e intercambio de datos, y los sistemas tradicionales de gestión de datos (SQL y otros) a concentrarse en la gestión y el análisis de grandes conjuntos de datos almacenados principalmente en bases de datos paralelas. Los sistemas de bases de datos tradicionales tienen las virtudes de paralelismo automático, indexación, y de acceso no procedural, pero tienen que adoptar los tipos de datos de la comunidad científica y necesitan coexistir con los datos en el sistema de archivos.

Estos formatos incluyen estructuras para soportar grandes cantidades de datos, arreglos multidimensionales, una amplia variedad de tipos de datos multivariados, multievaluados y métodos para la gestión de metadatos.

En este epígrafe se presentan algunas características de los principales formatos de datos científicos y las bases de datos paralelas.

Entre los formatos estándares antes mencionados ha sido de más utilidad para esta investigación el HDF (Hierarchical Data Format) principalmente por la posibilidad del trabajo en paralelo a partir de la versión HDF5 y el NetCDF por su gran uso en los diferentes centros de ciencia. En este epígrafe se tratan de forma específica estos formatos de datos científicos, profundizando en su estructura y las funcionalidades de su biblioteca de funciones.

1.6.1 HDF

HDF (Hierarchical Data Format) es una biblioteca y un formato de fichero multiobjeto para la transferencia de datos gráficos y numéricos entre máquinas.

HDF se encuentra libremente disponible. La distribución consiste en la biblioteca HDF, las utilidades de línea de comando HDF permiten convertir de un formato a otro (ejemplo JPEG a HDF y viceversa), analizar y ver los ficheros hdf así como manipularlos.

Las interfaces de programación de aplicaciones incluyen conjunto de rutinas para el almacenamiento y acceso a tipos de datos específicos. A pesar de que cada interfaz requiere el uso de un (API), también necesita programación, todos lo detalles de bajo nivel pueden ser ignorados. Estas bibliotecas se encuentran disponibles en C y Fortran. Los tipos de estructura de datos que soporta HDF son Conjuntos de datos científicos, Imágenes de puntos y paletas de colores, en (Cheng et al., 2003) se pueden encontrar detalles sobre el manejo de estas herramientas.

HDF 4.0 (y versiones posteriores) soportan una interfaz de compresión de bajo nivel, la cual permite que cualquier objeto de dato sea comprimido utilizando una variedad de algoritmos.

Los principales rasgos que caracterizan los ficheros HDF son:

- **Es versátil**. HDF suporta muchos tipos diferentes de modelos de datos. Cada modelo de dato define un conjunto específico de tipo de dato y provee una API para lectura, escritura y organización de datos y metadatos del tipo correspondiente. Los modelos de datos soportados incluyen arreglos multidimensionales, imágenes de puntos, y tablas.

- **Es auto-descriptivo**. Permitiendo una aplicación para interpretar la estructura y contenidos de un fichero sin ninguna información proveniente del exterior.

- **Es flexible**. Con HDF, se pueden mezclar y asociar objetos relacionados agrupados en un fichero y acceder a ellos como un grupo o como objetos individuales. Los usuarios pueden además crear sus propias estructuras de agrupamiento utilizando un rasgo de HDF llamado vgroups.

- **Es extensible**. Puede acomodar fácilmente nuevos modelos de datos, sin tener en cuenta si fueron adicionados por el equipo de desarrolladores de HDF o por los usuarios de HDF.

- **Es portable**. Los ficheros HDF pueden ser compartidos a través de la mayoría de las plataformas comunes, incluyendo muchas estaciones de trabajo y computadoras de alto desempeño. Un fichero HDF creado en una computadora puede ser leído en un sistema diferente sin modificación alguna.

Además de las principales características del HDF estándar, para utilizar HDF paralelo (Parallel HDF), hay que tener en cuenta varios requisitos:

Los archivos paralelos del HDF tienen que ser compatibles con archivos HDF5 y compartidos entre diferentes plataformas paralelas.

La característica fundamental de esta tecnología es el trabajo con plataformas paralelas y una interfaz de Entrada-Salida paralela portátil para diferentes plataformas. HDF5 Paralelo tiene como meta inicial el soporte de la programación MPI pero no la programación de memoria compartida pues tuvo que ser diseñado sobre la idea del trabajo con un archivo único accedido por todos los procesos causando un procesamiento costoso, ampliado (Cheng et al., 2003).

HDF5 Paralelo permite que una aplicación lea de un directorio lleno de archivos en lugar de sólo un archivo. Los archivos de HDF5 son tratados de manera muy similar a los directorios. Es necesario antes de crear un archivo, asegurarse de que el directorio sobre el cual se va a trabajar esté creado, de no ser así es necesario hacerlo facilitando el trabajo

con dicho archivo desde cualquier lugar del sistema sin perdida de información. Los lenguajes de programación comunes usados para implementar las aplicaciones de HDF5 Paralelo son Fortran, C y C++.

1.6.2 NetCDF

NetCDF es una interfaz a una biblioteca de funciones de acceso a datos diseñada para el almacenamiento y recuperación de de datos en forma de arreglos, donde cada arreglo es una estructura rectangular n-dimensional (con $n = 0,1,2\cdots$) en la que todos sus elementos son del mismo tipo de dato (ejemplo: caracteres de 8-bit, entero de 32-bit) y un escalar (un solo valor) es un arreglo 0-dimensional.

La biblioteca NetCDF implementa un tipo de datos abstracto, por lo que todas las operaciones para acceder y manipular datos, en un conjunto de datos de netCDF, debe usar sólo las funciones que provee la interfaz. La representación de los datos está oculta para aplicaciones que usen la interfaz, así que la forma en la cual estos se encuentran almacenados puede ser cambiada sin afectar los programas existentes. La representación física de los datos de formato netCDF esta designada para ser independiente de la computadora en la cual sean guardados.

Una de las metas de netCDF es soportar acceso eficiente a pequeños subconjuntos de grandes conjuntos de datos. Para lograr esto, NetCDF utiliza accesos directos en lugar de accesos secuenciales. Esto puede ser mucho más eficiente cuando el orden en que los datos son leídos es diferente del orden en que fueron escritos, o cuando debe ser leído en diferentes órdenes para diferentes aplicaciones.

La cantidad de costo para la representación externa portable depende de muchos factores, incluyendo el tipo de dato, el tipo de computadora, la granularidad del acceso a datos, y de cuan bien se ha ajustado la implementación a la computadora en la que está corriendo. Este costo es típicamente pequeño en comparación con todo el conjunto de recursos utilizados por una aplicación. De cualquier forma, el costo de la capa de representación externa es usualmente un precio razonable a pagar por acceso de datos portables.

A pesar de que la eficiencia en el acceso a datos ha sido un asunto importante en el diseño e implementación de NetCDF, todavía es posible utilizar la interfaz NetCDF para acceder a datos de formas ineficientes: por ejemplo, mediante la petición de una porción de datos que requiere un solo valor de cada registro.

1.7 Bases de Datos Paralelas

Una de las implementaciones más eficientes para las Bases de Datos Científicas debido a su exitoso desempeño basado en su gran capacidad de cómputo son las Bases de Datos Paralelas. Para comprender mejor este concepto es necesario en primer lugar conocer las bases de datos distribuidas. Según (DeWitt2 et al., 1992) una Base de Datos Distribuida es una colección múltiple de datos lógicamente relacionada a través de una red de computadoras. Un sistema de administración de bases de datos (DBMS) distribuido se define como un sistema de software que permite la fragmentación y la distribución de los datos de forma transparente para los usuarios, la forma en que se implementa esta distribución resulta muy importante para esta investigación, pues resulta un factor determinante a la hora de gestionar los datos en esta alternativa de implementación. Una de las características principales de las Bases de Datos paralelas esta determinada por la arquitectura sobre la cual basan su funcionamiento, que puede variar de acuerdo a la capacidad económica de la institución que la implementa.

El avanzado desarrollo de la arquitectura de computadoras ha permitido establecer una nueva generación de computadoras personales, que utilicen más (y mejores) componentes integrados en un solo sistema. Este es el caso de aquellos computadores que soportan el trabajo de más de un procesador simultáneamente; proceso que es denominado actualmente procesamiento paralelo.

El soporte de varios procesadores en una misma mainboard ha agiliza el procesamiento de operaciones por parte de los computadores, produciendo un aumento de la eficiencia en los sistemas informáticos. La paralelización de operaciones consiste en separar una operación a procesar, en Sub-operaciones; cada una de las cuales es ejecutada de forma paralela en cada procesador para luego reunir los resultados en una sola respuesta.

Esta característica de las nuevas computadoras ha permitido que las bases de datos puedan ser implantadas en sistemas de esta naturaleza, tomando el nombre de Bases de Datos Paralelas, (DeWitt2 et al., 1992).

Estas Bases de Datos no solo son implantadas en una máquina centralizada, también pueden utilizar una red de alta velocidad que enlaza varios computadores (cluster de computadoras); todos ellos trabajando conjunta y simultáneamente (procesadores, memoria principal y discos de almacenamiento) para construir una base de datos paralela.

Los Sistemas de Bases de Datos Paralelas se comenzaron a desarrollar en los años 80 sobre un sistema de procesadores de propósito general funcionando en paralelo. Sobre este sistema desarrollado se basan los sistemas actuales de IMB, Oracle, Infomix entre otros. Los Sistemas Paralelos de Bases de Datos están constituidos por un conjunto de procesadores y discos de almacenamiento, comunicados por una red de alta velocidad, que trabajan de forma conjunta ejecutando consultas y transacciones de los clientes. Esta forma de ejecutar las transacciones mejora considerablemente la productividad, ejecutando sub-transacciones de forma simultánea. La arquitectura es sin lugar a dudas un factor determinante para la implementación de una base de datos, pues existen varios modelos para el procesamiento paralelo:

1. **Memoria compartida:** Todos los procesadores comparten una memoria común, Es muy eficiente en cuanto a comunicación entre procesadores.

2. **Disco Compartido:** Los procesadores comparten un conjunto de discos de almacenamiento pero no es posible trabajar con un conjunto grande de nodos, esta arquitectura permite una mayor escalabilidad entre los nodos.

3. **Sin compartimiento:** Cuando los procesadores no comparten ni memoria, ni discos. Cada nodo contiene un procesador, memoria y sus propios discos. Debido a que los datos se transportan desde los discos hasta la memoria dentro de un mismo

nodo se minimiza las colisiones en la red. Este tipo de arquitectura es escalable a miles de nodos.

4. **Jerárquico:** Cuando los procesadores comparten memoria principal y los discos de almacenamiento.

Es importante también señalar los tipos de paralelismo que se pueden encontrar en estas arquitecturas:

Paralelismo de Entrada salida: Permite la reducción del tiempo necesario para la recuperación de datos dividiendo la base de datos entre varios discos a través de una división horizontal.

1. **Paralelismo entre consultas:** Permite la ejecución de consultas o transacciones en paralelo.
2. **Paralelismo en consultas:** Permite que una única consulta se ejecute de forma paralela en diferentes procesadores y discos.
3. **Paralelismo entre operaciones.**

    a) **Paralelismo de Entrecruzamiento:** Permite ejecutar varias operaciones sin necesidad de acceder al disco para almacenar datos intermedios

    b) **Paralelismo Independiente:** Cuando se ejecutan varias operaciones independientes que pertenecen a la misma consulta.

La arquitectura de sistema de base de datos paralela es basada fundamentalmente en un hardware sin compartimiento donde los procesadores se comunican entre si enviando mensajes mediante la vía de una red de interconexión. En dichos sistemas, las tuplas de cada relación son particionadas para permitir que múltiples procesadores escaneen grandes relaciones paralelas sin necesitar dispositivos de E/S. Dicho diseño es utilizado actualmente

por Teradata, Tandem, NCR, Oracle-nCUBE, y por algunos otros productos de bajo desarrollo. La comunidad de investigación también ha abarcado esta arquitectura de flujo de datos sin compartimiento en sistemas como Arbre, Bubba, y Gamma, mencionados en (DeWitt2 et al., 1992).

Las bases de datos paralelas presentan importantes ventajas con respecto a otros sistemas de bases de datos pues permiten un aumento en la velocidad, necesitando menor tiempo de ejecución para cada transacción. También se logra la ampliación al poder procesar tareas más largas en el mismo tiempo. A pesar de estas posibilidades que brindan las bases de datos paralelas el costo de inicio es alto a la hora de la instalación y puesta a punto del sistema.

A pesar de parecer similares, los sistemas de bases de datos paralelas se diferencian de las distribuidas ya que estas se encuentran generalmente en diferentes puntos geográficos y se administran de forma separada y poseen una interconexión más lenta. También en los sistemas distribuidos se dan dos tipos de transacciones: locales y globales. Las transacciones locales son aquellas que se ejecutan solamente en el nodo donde se inicia la transacción, mientras que las globales involucran intercambio de datos entre los distintos nodos que conforman el sistema, mientras que en un sistema paralelo todas las transacciones tienen el mismo nivel, ya que todas son atendidas por el sistema en conjunto (independientemente de ser paralelizadas o no). Las bases de datos paralelas trabajan con arquitectura de computadoras multiproceso posibilitando un alto rendimiento y alta disponibilidad para servidores de bases de datos con menor costo(DeWitt2 et al., 1992).

De forma general se puede decir que las Bases de Datos Paralelas agilizan el proceso de consultas a un SBD a través de la ejecución de forma simultánea de las transacciones, sin embargo el costo de la implementación es mucho mayor que otros tipos de Bases de Datos. Los sistemas de bases de datos paralelas están teniendo una gran aceptación comercial, pero solo en organizaciones que tienen complejos sistemas de bases de datos.

1.8 Conclusiones Parciales

A modo de conclusión en este capítulo se tratan las principales características de los centros de ciencia, así como las dos tecnologías paralelas que pueden soportar los requerimientos de este nuevo estilo de trabajo. Como el objetivo principal de este modelo es brindar servicios de visualización, se analizaron las características de la visualización científica y las ventajas de las técnicas de visualización distribuida. Finalmente se centró la atención en los recursos de almacenamiento de los datos, tanto formatos de datos científicos (HDF, NetCDF…) como las bases de datos paralelas, ambos muy importantes para determinar cual utilizar en el modelo de centro de ciencia, que se pretende crear a partir del estudio de las características de algunos de los centros mas importantes del mundo, que se discuten en la próxima sección.

2. "Tendencias actuales de los Centros de Ciencia"

Los centros de ciencia o centros de datos científicos, que están surgiendo como estaciones de servicio para algunos dominios científicos y suministran acceso tanto a los datos como a las aplicaciones que los analizan, se presentan como una vía para solucionar las crecientes demandas de recursos computacionales. Según (Gray, 2005) un centro de ciencia es capaz de almacenar uno o más conjuntos de datos y ofertar aplicaciones que acceden a ellos y los analizan. Posee además, un equipo de personas que entiende los datos y está constantemente mejorándolos y agregando más información.

En este capítulo se analizan profundamente las principales características de cuatro de los centros de ciencia más importantes del mundo, escogidos por sus funcionalidades y servicios que brindan. Estos son de interés para la creación de un modelo de centro de ciencia, donde se pretende integrar la visualización científica con los CC como nueva alternativa para el tratamiento de grandes volúmenes de datos.

Se hace un análisis de una herramienta solucionadora de recursos de almacenamiento (SRB) analizada en (Rajasekar et al., 2003), que utilizan casi todos los centros de datos, para manejar sus recursos de almacenamiento y procesamiento, con el objetivo utilizarla por ser de código abierto o definir una similar que facilite integrar los recursos disponibles con los servicios que se pretenden brindar a los usuarios.

Es valido señalar que la documentación sobre el diseño e implementación de estos centros de ciencia es muy escasa, los principales artículos no están publicados en Internet o hay que pagarlos, las principales fuentes que se utilizaron para esta investigación es la interacción directa con estos centros de ciencia que si son de libre acceso para todo el mundo y la ayuda de científicos interesados en esta investigación que nos han facilitado bibliografía. Lo cual hace más importantes esta investigación, para ayudar al desarrollo de pequeñas instituciones que no disponen de suficientes recursos para brindar servicios similares con un costo asequible.

2.1 Solucionador de Recursos de Almacenamiento (SRB)

En el centro de supercomputadoras de San Diego, se ha desarrollado un sistema de gestión de datos que satisface las necesidades de la mayoría de las comunidades científicas. La tecnología se denomina "solucionador de recursos de almacenamiento" (Storage Resource Broker), o SRB. El sistema actualmente en funcionamiento en SDSC maneja más de 50 tera bytes de datos, incluyendo alrededor de 9 millones de archivos, almacenados en repositorios en SDSC y otros sitios. Existen otros sistemas SRB para el manejo de datos del orden de los tera bytes, pero sin duda el más importante es éste. Según (Peter Cao et al., 2005, Rajasekar et al., 2003), el sistema SRB es utilizado por proyectos patrocinados por La Fundación Nacional de Ciencias, La Administración Nacional de Aeronáutica y de La Agencia del Espacio, el Departamento de Energía, el Consejo Nacional de Archivos y Registros de Administración, el Instituto Nacional de Salud, y el Instituto Nacional de Publicaciones Históricas y La Comisión de Registros.

El SRB es un middleware basado en tecnología cliente-servidor que proporciona un servicio de construcción de colecciones de datos, así como su administración, brinda servicios de consulta y de acceso, y la preservación de los datos en un marco de red de datos distribuidos. En pocas palabras el SRB ofrece las siguientes capacidades.

- Identificadores globales persistentes para la asignación de nombres de archivos.
- Soporte para metadatos para describir la ubicación y la propiedad de los archivos.
- Soporte para metadatos descriptivos para apoyar el descubrimiento a través de los mecanismos de consulta de bibliotecas digitales.
- Norma mecanismos de acceso a través de navegadores Web, los comandos de shell de Unix, navegadores de Windows, scripts Python, Java, las llamadas biblioteca de C, Linux I / O redirección, WSDL, etc.
- Depósito abstracto de almacenamiento para interactuar con varios tipos de sistemas de almacenamiento.

- Sistema de autenticación para el acceso seguro a sistemas de almacenamiento remoto.
- El apoyo a la replicación de archivos entre sitios.
- Soporte para copias de los archivos de caché en un sistema de almacenamiento local y el apoyo para acceder a varios archivos en uno mismo como si estuvieran en el mismo lugar.
- Apoyo para la agregación de los archivos en los contenedores.
- Apoyo para la inscripción de nuevos ficheros en el sistema.

El SRB ha estado en producción y en uso durante los últimos cuatro años. Existen varios centros de distintas áreas de investigación que utilizan SRB como: Astronomía, Sistemas de la Tierra y Ciencias del Medio Ambiente, Ciencias Médicas, Ciencias Moleculares, Neurociencias, Física y Química, Archivos y Bibliotecas Digitales (Rajasekar et al., 2003).

En los epígrafes siguientes se muestran las principales características de los centros de ciencia SDSS, BaBar, BIRN y GebBank, escogidos por su prestigio, aporte al desarrollo de la ciencia y principalmente porque sus características de diseño y servicios que brindan a la comunidad que ayudaran a crear un modelo de centro de ciencia.

2.2 Sloan Digital Sky Survey (SDSS)

SDSS SkyServer, es un proyecto para confeccionar un mapa de gran parte del universo, ha revolucionado la astronomía por las facilidades que brinda a la comunidad de científicos que investigan en esta rama de la ciencia tan importante. Con la disponibilidad de acceso a los datos los astrónomos realizan consultas complicadas y obtienen respuestas en pocos segundos, y quizás minutos en caso de que la consulta requiera buscar en toda la base de datos.

El SDSS brinda conjuntos de datos del orden de los tera bytes, recolectados a partir de 1998 hasta la actualidad, estos datos cada día son más dóciles debido al crecimiento exponencial de la velocidad de procesamiento y el incremento de la capacidad de almacenamiento. El uso de la visualización distribuida, la réplica de los datos y las herramientas de programación paralela han logrado una equivalencia entre el nivel de almacenamiento y la capacidad de cómputo necesaria para poder analizar estos grandes volúmenes de datos. (sdss, 2008)

Los datos astronómicos obtenidos por los diferentes medios pasan por distintas etapas antes de ser publicados a la comunidad de científicos. Esta información primeramente es almacenada en bases de datos temporales en el Laboratorio Nacional de Fermi, que solo están disponibles para los especialistas encargados de demostrar que estos datos son una representación real de alguna parte del universo. Después del procesamiento y reexaminación de la calidad son exportados para el Centro de Ciencia, donde son replegados en las distintas instituciones pertenecientes a SDSS, con el archivo maestro residiendo en Fermilab.

Este centro de ciencia almacena los datos en bases de datos de SQL, organizados de forma jerárquica en multi-capas. Este repositorio es un conjunto de datos distribuido y esta compuesto por miles de bases de datos. Cada base de datos es un archivo en disco y la colocación de los objetos en contenedores es configurable por el administrador de la base de datos.

El repositorio de datos puede ser movido a través de plataformas heterogéneas sin pérdida de compatibilidad. Esto facilita la implementación y el mantenimiento de archivos distribuidos. Las herramientas administrativas brindan servicios de seguridad y gerencia de los datos. La replicación de los datos permite hacerle copia además de propagar las actualizaciones automáticamente en los sitios espejos, estas principales características posibilitan correr consultas complicadas para todo tipo de búsqueda en cualquier punto geográfico.

SDSS posee una herramienta online llamada CasJobs que permite el acceso a la gran base de datos de este centro de ciencia a través de los catálogos científicos. Esta herramienta está diseñada para emular y mejorar el acceso libre a las consultas en un entorno web, además permite a los usuarios abrir múltiples sesiones con servidores diferentes simultáneamente y enviar consultas a servidores en paralelo, las consultas propuestas pueden dirigir su salida de regreso a la interfaz gráfica del usuario, también pueden enviar la salida directamente a un archivo o a otra herramienta de análisis. La capacidad binaria de salida permite el flujo de los datos en forma compacta para agilizar el tráfico en la red. (CasJobs, 2008)

Algunas funciones de esta aplicación son:

- Permite ejecutar consultas de forma sincrónica y asincrónica, en forma de tareas rápidas o demoradas.
- Guarda un historial que registra las consultas y su estado.
- Posee una base de datos personal llamada "MyDB" en el lado del servidor que permite la creación de tablas, funciones y procedimientos. En esta base de datos personal se almacenan partes de los grandes volúmenes de datos del centro de ciencia que el usuario está investigando.
- Permite el compartimiento de datos entre los usuarios, a través de un mecanismo de "grupos de usuarios".
- Permite descargar en diferentes formatos los datos que están en la base de datos personal MyDB.
- Presenta varias opciones de interfaz, incluyendo un navegador cliente, así como una herramienta de línea de comandos.

Las consultas presentadas al sistema son tramitadas por un Agente de búsqueda del SDSS, que es transparente al usuario, lo cual es un servidor inteligente de búsqueda que:

- Maneja las sesiones del usuario.

- Analiza gramaticalmente, optimiza y ejecuta las búsquedas del usuario.
- Extrae atributos de objetos individuales conforme los solicite el usuario.
- Las salidas son puestas en las rutas especificadas por el usuario.

El primer paso del Agente de Búsqueda del Centro de Ciencia es analizar cada consulta propuesta para proveer una "estimación del costo de la tarea" en curso. Este costo es estimado a partir de los subconjuntos de la base de datos que debe ser registrada y mostrado al usuario como el tiempo mínimo obligado para completar la búsqueda, por lo que el usuario basado en el alcance y el costo de la consulta puede decidir abortarla o no.

El lenguaje de búsqueda del SDSS es SXQL; es un SQL simple que implementa el subconjunto básico de cláusulas y funciones necesarias para formular consultas para una base de datos objeto relacional. Reconoce las cláusulas SELECT-FROM WHERE del SQL estándar, además incluye declaraciones SELECT anidadas. Más allá permite especificación de enlaces de asociación en el SELECT, FROM y subcláusulas WHERE. Las asociaciones son enlaces para otros objetos, y pueden ser uno o varios enlaces. Además reconoce una sintaxis de búsqueda de proximidad, lo cual deja al usuario ir en busca de todos los objetos que están próximos a un objeto dado en el espacio. Además contiene un número de macros específicos de la astronomía y tiene soporte para funciones matemáticas.

Otra característica importante que no se debe dejar de mencionar es la capacidad de salida en formato XML, simplemente especificando FOR XML en la cláusula WHERE de la consulta, el usuario puede optar por recibir el resultado de la búsqueda en XML en vez de texto ASCII. Ésta es una característica especialmente importante para el Observatorio Virtual, en el cuál muchas aplicaciones de análisis de datos que los astrónomos necesitan, probablemente estarán disponibles como servicios web.

Este centro de ciencia posee además otro conjunto de herramientas dentro de las que se encuentran, la herramienta Famous Places que presenta una galería de imágenes de varias

galaxias y algunas de las estrellas mas distantes que se han descubierto. La herramienta Get Images que permite descargar imágenes individuales con varios niveles de resolución. Otras herramientas no menos importantes son Scrolling Sky, Search, Object Crossid, Help y Download. Además SDSS presenta varias herramientas de visualización (Visual Tools) como son Finding Chart que devuelve una imagen jpg con varios parámetros aplicados a los datos, Navigate que permite la navegación interactiva por lo datos del universo y otras como Image List , Quick Look, y Explore, este ultimo permite explorar interactivamente varias propiedades de un objeto individual.

El Sloan Digital Sky Survey identificará la ubicación de más de 100 millones de galaxias. Además mediante la combinación de los datos que se encuentran en los catálogos o el reanálisis de cada uno de los píxeles de cada imagen se pueden obtener nuevos objetos o generar estadísticas en forma de galaxias. Este último enfoque exige el flujo de imágenes desde su ubicación de almacenamiento a través de una plataforma de procesamiento. Para facilitar el análisis. Las colecciones han sido replicadas en NSF Teragrid que proporciona tanto los recursos de cómputo como de almacenamiento.

Los metadatos utilizados para describir los datos se basan en descriptores uniformes de contenido para los catálogos de entradas. La astronomía ha creado una comunidad estándar de formato de datos para las imágenes, llamado FITS. Este formato proporciona una manera de encapsular metadatos descriptivos acerca de la ubicación, la resolución, y el alcance de cada imagen, junto con los píxeles que componen la imagen.

2.3 BaBar

El Centro de ciencia BaBar en el Stanford Linear Accelerator Center, se encarga de almacenar grandes volúmenes de datos del orden de los peta bytes obtenidos de experimentos de física, así como brindar servicios a la comunidad de científicos para analizar estos datos. Al ocuparse del acceso concurrente de múltiples clientes para repositorios de datos de escala petabyte: el alto funcionamiento, la tolerancia a fallos, la robustez, y la dimensionalidad son cuatro aspectos muy importantes que deben estar presentes en la arquitectura del centro. (Babar, 2007)

La comunidad de Física de Altas Energías, tradicionalmente, ha venido acumulando experiencias en el almacenamiento y análisis de grandes volúmenes de datos, teniendo como uno de sus principales exponentes el centro de ciencia BaBar. Sólo la comunidad del experimento BaBar aglutina a más de 600 físicos e ingenieros de 75 instituciones a escala mundial y la tendencia es que sigan aumentado cada año, además han logrado consolidar una de las bases de datos orientadas a objetos más grandes del mundo que tenia en noviembre del 2004 casi 900 tera bytes. BaBar está orientado a la exploración de los principios físicos de la energía, la materia y la antimateria, utiliza recursos computacionales en modelo de GRID para aumentar la velocidad y la calidad en las investigaciones mundiales en este tema.

El LHC (Large Hadron Collider) Computing Grid Project (LCG) ha adoptado las herramientas computacionales y ha creado grupos de evaluación y prueba conjunta con el Proyecto Europeo DataGrid. La organización para la operación del LCG es jerárquica y está compuesta por Tiers, donde la raíz de esta jerarquía está en el CERN (Tier 0), luego nodos regionales (Tier 1) en IN2P3 Lyon-Fancia, PIC Barcelona-España, CNAF en Boloña-Italia, PPARC en Rutherford-Inglaterra, por mencionar los principales nodos europeos; luego seguirían nodos menores de servicios. Cada uno de estos nodos deberá realizar labores de limpieza de datos y proveer capacidades de cómputo para el análisis de esos resultados.

El experimento BaBar en Stanford Linear Accelerator Center produce una cantidad enorme de datos para ser accedidos por un número alto de trabajos de análisis. Por esta razón requiere un sistema confiable y escalable de acceso a datos. En 2002, el comité decidió migrar el sistema de almacenamiento de datos de Objectivity / DB para un sistema de archivo plano basado en flujo de objetos. El nuevo sistema de almacenamiento se basa en un mecanismo de persistencia, desarrollado en CERN 3, que es capaz de fluir un objeto en un archivo binario similarmente como el Java framework.

BaBar usa (SRB) como herramienta para suministrar una interfaz uniforme de acceso a sistemas de almacenamientos heterogéneos como (discos, cintas, bases de datos) que están

distribuidos en varios sitios, además como herramienta de soporte para compartir archivos de forma colaborativa, permite el control de replicas por lo que es muy fácil duplicar una BD de un sitio para otro, y proporciona búsquedas por atributos de los datos y búsquedas por metadatos.

La implementación de técnicas de gran capacidad de procesamiento, así como sistemas de acceso a datos capaces de maniobrar millones de archivos distribuidos y tolerantes a fallos, son objetivos fundamentales en las que se basa la arquitectura de este centro, algunas características que la distingue son:

- Los múltiples servidores cooperan entre si con el fin de maniobrar cantidades enormes de datos distribuidos y redundantes si es necesario, sin forzar al cliente a saber cuál servidor contactar para acceder a un conjunto particular.

- El servidor encubre las aplicaciones del cliente que están debajo del tipo del sistema de archivo, aun si maneja una o más unidades.

- Presenta un mecanismo de balanceo de carga, para distribuir eficazmente la carga entre grupos de servidores.

- Presenta un alto grado de tolerancia a fallos, para minimizar el número de Trabajos o aplicaciones que tienen que ser vueltos a arrancar después de algún problema que detecte el servidor o cualquier clase de fallo de la red.

La arquitectura diseñada por los especialistas para cumplir con los requisitos antes mencionados, permite la construcción de sencillos sitios de acceso a datos del servidor, cargar ambientes balanceados y estructura de implementaciones iguales. Esta estructura en cualquier caso presente es una interfaz definida como un protocolo de comunicación, que define las posibles interacciones y funcionalidades dadas a los clientes.

El protocolo de comunicación define una interfaz capaz de: Pedir acceso al sistema a través de los medios de autentificación, interrogar un sistema en busca de un recurso, obtener acceso al recurso pedido en el lugar donde puede ser accedido (o sea los servidores dándole acceso a los datos locales o permitiendo interoperabilidad remota de sitios), políticas sofisticadas de comunicación del lado del cliente y capaz de manejar cualquier clase de errores de comunicación. Las peticiones que fallen son intentadas de nuevo hasta que se encuentre otro servidor de trabajo, el mismo servidor se hace disponible otra vez o hasta que se alcance un número máximo especificado de nuevos intentos.

Este sistema por el lado del servidor está compuesto por cuatro capas: Red y capa de administración de hilos, capa de protocolos, capa de sistema de archivos y capa de almacenamiento. Mientras que por el lado del cliente por tres capas: La capa de interfaz, la capa de comunicación de alto nivel y la capa de comunicación de bajo nivel.

2.4 Biomedical Informatics Research Network (BIRN)

BIRN es otro centro de ciencia que está compuesto por gran cantidad de sitios, principalmente de Estados Unidos y Gran Bretaña, residiendo en la universidad de California, San Diego. Unidos todos con el objetivo de diseñar e implementar una arquitectura distribuida de recursos compartidos que sean usados por los investigadores de la biomedicina para avanzar en el diagnostico y tratamiento de enfermedades. (BIRN, 2008)

El proyecto BIRN es una iniciativa encaminada a la creación de un laboratorio para los investigadores biomédicos mediante el acceso y análisis de datos ubicado en diversos sitios de varios países. Cuestiones relacionadas con el manejo de los datos en la red, autentificación de usuarios, integridad de los datos, seguridad y la propiedad de los datos se requieren como parte del proyecto BIRN. Son necesarias dos redes para el intercambio de datos: una para bibliotecas digitales para publicar los datos, y se necesita otra para la persistencia de los archivos para conservar los datos. Un catálogo central fue creado en SDSC para la gestión de datos que lógicamente fue organizado en una colección. Los datos

originales residen en los sitios participantes. Mientras que los archivos están replicados en SDSC.

Una parte importante del proyecto BIRN es administrado centralmente para desplegar el hardware personalizado que es extensible tanto en la capacidad como la ubicación. Un sistema basado en linux, llamado BIRN RACK, se desplegó en cada sitio para crear memorias caché de datos distribuidos. Esta memoria en un componente esencial de la red necesario para la gestión de acceso a recursos de datos estrictamente controlados en una amplia área de red. El BIRN RACK puede personalizarse para ejecutar el Storage Resource Broker (SRB), al igual que Babar. El Centro Coordinador de SRB UCSD administra la red de datos. Los datos se replican entre sitios bajo el control de SRB. El SRB también apoya el acceso a los datos de visualización de los programas que están actualmente en uso por los diferentes asociados a BIRN.

En cierto sentido, los mecanismos de acceso uniforme de SRB hacen posible que se apliquen diferentes técnicas a los mismos datos para ver las ventajas. El objetivo principal del proyecto BIRN es superar los retos y problemas en el acceso a grandes conjuntos de datos en todos los sitios al mismo tiempo, atender el cumplimiento estricto que necesita el intercambio de datos médicos.

Un requisito particular del proyecto BIRN es la capacidad de aplicar controles de acceso tanto a las imágenes como los metadatos descriptivos y administrativos registrados en un espacio de nombres lógicos. Todos los controles de acceso se aplican directamente al nombre lógico. Esto significa que cuando una imagen se mueve, automáticamente los controles de acceso se aplican en la nueva ubicación. En el ambiente SRB, todos los archivos se almacenan bajo el control de sistema de manejo de datos SRB, asignando un identificador de UNIX. El sistema SRB utiliza los controles de acceso para decidir si se permite el acceso a los usuarios, mediante un certificado de llave pública.

Los controles de acceso a los metadatos son más complejos. Las restricciones de acceso permiten que los usuarios solo puedan ver los metadatos descriptivos de los archivos que se le permite el acceso. Todos los archivos para los que no tiene permiso no pueden ser vistos

mediante consultas. También son necesarias restricciones de acceso a los metadatos administrativos, de manera que los usuarios no pueden ver ningún valor de los atributos seleccionados. Así, los administradores pueden utilizar los valores de los atributos para administrar el repositorio sin ser visto por los usuarios.

Uno de los objetivos del proyecto BIRN fue la necesidad de desarrollar sistemas de bajo costo para almacenar las grandes colecciones de datos. Los investigadores prefieren mantener los datos en discos giratorios para reducir al mínimo las latencias de acceso. Con el desarrollo de los discos y los procesadores, ahora es posible tener un costo en cuanto a disco de almacenamiento alrededor de $ 3000 por tera bytes, incluyendo el acceso a la red, estas redes Grid son sistemas modulares que combinan una CPU de 1,7 Ghz, con un giga bytes de memoria, un tera bytes de disco, una controladora RAID, y una conexión de red que funciona bajo el sistema operativo Linux. Para ampliar la capacidad de procesamiento se añaden nuevos nodos al sistema y se pueden añadir más discos para aumentar la capacidad de almacenamiento.

La colaboración entre los diversos sitios que componen BIRN suministra al resto de la comunidad científica los siguientes aspectos:

- Intercambio de datos: La colaboración entre los múltiples sitios permiten desarrollar infraestructuras para compartir los datos y definir políticas sobre los conjuntos de datos disponibles para la comunidad científica.

- Interoperabilidad: La colaboración entre los diversos sitios puede motivar la interoperabilidad de las herramientas de análisis desarrolladas en los diferentes laboratorios.

- Estándares de comparación: Esta colaboración puede llevar al desarrollo de estándares de los nuevos formatos de datos y protocolos de análisis. Estas ayudas mueven el campo hacia delante estableciendo las marcas del mismo campo de trabajo.

- Sinergia: La colaboración puede fomentar la sinergia que significa poder producir más con la unión de varios sitios que lo que pudiera producirse en cada uno por separado. La combinación de fuerzas de cada sitio puede suministrar recursos más robustos que los que están disponibles en cada sitio.

BIRN posee una gran cantidad de herramientas para:

- Adquisición, calibración y control de calidad.
- Almacenamiento y administración de datos.
- Análisis y procesamiento.
- Visualización y construcción de mapas cerebrales, etc.
- Construcción de estándares, terminologías y ontologías.
- Intercambio de datos y colaboración.

Las comunidades de investigación biomédicas trabajan con grandes conjuntos de datos y a menudo altamente heterogéneos. BIRN provee una manera para integrar estos datos, permitiendo analizar los datos, distribuir la información para realizar nuevos descubrimientos. El Virtual Data Grid (BVDG) almacena los datos a través de un ambiente distribuido, donde estos son continuamente administrados y manipulados como un simple sistema virtual de archivos. Actualmente, dentro del BVDG existen más de tres millones de ficheros de datos, activos, que son accesibles directamente a través de un conjunto de interfaces estandarizadas.

A través de BIRN se puede compartir o publicar tanto datos como resultados de las investigaciones usando el BIRN Data Repository (BDR), el cual se rige por un conjunto de normas que tienen aspectos de publicación, garantía, seguridad y redistribución.

Los científicos pueden acceder desde cualquier ordenador que tenga acceso a Internet a través de un simple portal web que provee un conjunto de herramientas integradas, infraestructura, y servicios necesarios para realizar estudios de gran magnitud y en colaboración.

Además este centro tiene la capacidad de autentificar todas las personas que acceden a las imágenes, la gestión de los controles de acceso a todos los archivos de datos independientemente del lugar donde los datos están almacenados, la gestión de los controles de acceso a los metadatos, para proporcionar pistas de auditoría, todos los accesos a los datos de forma independiente de la ubicación de almacenamiento, y el apoyo de cifrado de datos. En la práctica, la auditoría se utiliza para demostrar la cantidad de intercambio de datos entre los sitios de colaboración. La cantidad de datos remotos visitados por un investigador puede ser cuantificada, así como la cantidad de datos que el investigador distribuye a otros sitios.

2.5 EntrezPubMedGenBank

El Centro Nacional para la Información Biotecnológica o National Center for Biotechnology Information (NCBI) es parte de La Biblioteca Nacional de Medicina de Estados Unidos y una rama de los Institutos Nacionales de Salud (National Institutes of Health o NIH) (ncbi, 2008). Está localizado en Bethesda, Maryland y fue fundado el 4 de noviembre de 1988 con la misión de ser una fuente importante de información de biología molecular. Almacena y actualiza constantemente varias bases de datos biológicas de acceso público. Entre las más conocidas y populares se encuentran las bases de datos de publicaciones científicas (PubMed), de secuencias de proteínas y ADN (GenBank), de estructuras tridimensionales de proteínas; y algunas otras no tan populares como OMIM (Online Mendelian Inheritance in Man). (GenBank, 2008)

El NCBI desarrolló Entrez como una herramienta para permitir a los usuarios interaccionar con estas bases de datos. Desde el punto de vista informático, Entrez es una 'interfaz de usuario', es decir, constituye el nexo entre el usuario y las bases de datos subyacentes.

Como interfaz, Entrez permite al usuario realizar consultas simples y obtener resultados, aun desconociendo la arquitectura de las bases de datos. Todas las bases de datos del NCBI están disponibles en línea de manera gratuita, y pueden ser accedidas usando el buscador Entrez, este centro es conocido como EntrezPubMedGenBank en lo siguiente se le llama GenBank.

El sistema de búsqueda PubMed es otro proyecto desarrollado por (NCBI) en el National Library of Medicine (NLM). Permite el acceso a las base de datos bibliográficas de NLM: MEDLINE, PreMEDLINE (citas enviadas por los editores antes de su publicación) y AIDS. Se pueden establecer alertas de correo electrónico automático para los nuevos artículos agregados a PubMed. Además tiene la opción de guardar las búsquedas, y una vez que ha guardado la búsqueda, puede escoger repetir la búsqueda en el futuro o bien que el sistema ejecute la búsqueda automáticamente y le envíe los resultados por correo electrónico.

Con el objetivo de escalabilidad en mente, este centro fue diseñado para usar una arquitectura distribuida. El sistema consta de una base de datos central de autentificación o de entrada y múltiples bases de datos distribuidas por categorías. Principalmente 8 categorías disponibles. La base de datos central de autentificación provee el URL de cada una de estas ocho bases de datos. Las ocho instancias de la base de datos comparten un esquema idéntico de base de datos, dejando una implementación genérica para el servicio Web. Las bases de datos de categorías adicionales pueden fácilmente ser incorporadas en el sistema existente con una adición mínima en la base de datos central de autentificación, no requiriendo cambio en la capa de servicio Web.

La base de datos del GenBank crece de manera exponencial, este crecimiento es debido a la forma misma en que la base se actualiza. Son los mismos autores quienes se encargan de mantener la base al día, pero además de remisiones de autores, el GenBank se nutre también de las otras bases de datos existentes actualizando interactivamente sus ficheros.

Existen muchas interfaces de acceso al GenBank, pero sin duda alguna la más efectiva es Entrez. Esta es una interfase que brinda acceso a muchas de las bases de datos mantenidas por el NCBI:

http://www.ncbi.nlm.nih.gov/Entrez /. Desde esta página se puede ir a:

1. División de publicaciones médicas (PubMed): interfase al servicio de citas bibliográficas del Medline.

2. Secuencias de nucleótidos, colección de archivos del GenBank

3. Base de datos proteica: esta base de datos combina la información de muchas fuentes con secuencias derivadas de la traducción de secuencias GenBank.

4. Base de datos de estructuras tridimensionales: información estructural de proteínas derivada de cristalografía de rayos X y resonancia magnética nuclear.

5. Bancos de Genomas: Compilación de mapas genéticos y físicos de una gran variedad de especies.

6. Taxonomía: Se usa la misma clasificación filogenético que en el GenBank, es principalmente un recurso de navegación.

Haciendo una búsqueda a través del Entrez es posible llegar a múltiples fuentes de información acerca del mismo tema, por ejemplo es factible para una secuencia encontrar su listado de citaciones bibliográficas contenidas en el Medline.

Un servicio de NCBI (http://www.ncbi.nlm.nih.gov) que vale la pena resaltar: el mapa genético del el genoma humano, este es un compendio de aproximadamente 30,261 secuencias.

A modo de resumen EntrezPubMedGenBank es un centro de ciencia que se dedica fundamentalmente a la bioinformática, medicina y biotecnología también esta compuesto por un grupo de sitios diferentes pero con objetivos similares. Presenta un conjunto de

servicios GRID que pueden ser accedidos a través de un cliente Desktop conocido como Taverna de ese GRID (Taverna, 2008). Este cliente cuando se ejecuta, automáticamente reconoce todos los servicios que hay en los diferentes sitios del centro de ciencia y permite el uso de estos servicios mediante diagramas de workflow. Los mensajes en XML son usados para la comunicación entre la capa del cliente y la capa de servicios web. La comunicación entre el cliente y servidor se realiza mediante un protocolo simple de acceso a datos sobre HTTPS. La capa de servicio Web es la interfaz entre el cliente y las funcionalidades que brinda el servidor. Esta arquitectura permite realizar cambios en las funcionalidades internas y localizaciones de los recursos sin que afecte la interfaz del cliente. GenBank provee servicios de búsqueda y de análisis en conjuntos no comprimidos de estos datos, requiriendo una cierta cantidad de 20 giga bytes de espacio en disco para datos, índices y las actualizaciones, así como también poderosos procesadores para poder analizarlos.

3. "Diseño del modelo general de un centro de ciencia sobre un cluster de computadoras"

Después de realizar el análisis de las características de los centros estudiados, teniendo en cuenta los elementos más importantes, es objetivo de la investigación, diseñar el modelo general donde se traten las principales funcionalidades que puede tener un centro de ciencia, creado sobre la tecnología de cluster de computadoras. Diseño que permitirá en tareas futuras la implementación del centro directamente en un cluster. En este capítulo se parte de una serie de requisitos que se asumen que deben tener los centros de ciencia para su estabilidad y buen funcionamiento, luego se analizan los aspectos que pueden ser soportados sobre la tecnología de cluster de computadoras. Se crea un diseño lógico de la arquitectura del centro, donde se analiza las distintas variantes que se presentan para el manejo de los datos. Posteriormente se tratan aspectos del módulo de visualización, para finalmente obtener una solución tecnológica donde se especifica el tratamiento tanto de los datos como los metadatos y los recursos de visualización.

3.1 Aspectos del diseño de un CC soportados sobre la tecnología de cluster de computadoras

Primeramente es necesario establecer un conjunto de requisitos que se asumen para la construcción del diseño de un centro de ciencia.

3.1.1 Requisitos asumidos para el diseño

El servicio más importante que debe brindar el centro es la visualización científica. Se debe tener en cuenta que un CC para la visualización tiene más accesos de lectura de los datos que de escritura en los discos, algo que es común en los sistemas de visualización. El tiempo de respuesta de este servicio debe ser lo más rápido posible, debido a que cuando se visualiza un conjunto de datos es con el objetivo de interpretar la imagen, ajustar algunos parámetros y volver a visualizar, por lo que es posible que un usuario tenga que realizar este proceso varias veces hasta obtener el resultado deseado, si no se logra un buen tiempo de respuesta esto podría causar que el usuario desista de utilizar el sistema.

El diseño debe contar con elevada capacidad de procesamiento y de almacenamiento, en un cluster de computadoras con discos duros se pueden cumplir perfectamente. Además la escalabilidad del cluster permite adicionar más nodos y disco duros en caso que se necesite aumentar la capacidad de cómputo o de almacenamiento en cualquier momento.

Se asume que el cluster de computadora en que se implemente este diseño debe tener nodos con características similares, igual sistema operativo, el mismo sistema de archivos y en el caso de tratarse de un cluster de BD, el mismo gestor de base de datos en para facilitar la implementación, aunque bien pudieran utilizarse varios gestores de BD.

Un aspecto fundamental es la conservación de la integridad de los datos almacenados en el CC, por lo cual el sistema debe ser tolerante a fallos. En caso que ocurra algún problema de perdida de información, se debe tener mecanismos de auto recuperación, que permita mantener la integridad de los datos. El uso de la replicación es una alternativa; con los archivos replicados al ocurrir cualquier fallo en el sistema, ya sea la perdida de datos por cualquier motivo como la rotura de un disco duro, el sistema debe detectar estos problemas y si es posible solucionarlo de forma automática, reestableciendo los datos a partir de un disco espejo o en caso que no se pueda, informarle al administrador el problema detectado.

El acceso concurrente a datos es una característica importante pero no hay que preocuparse por este especto en el diseño ya que todos los sistemas de archivos o sistemas de gestión de bases de datos se ocupan de esta tarea, aunque no siempre de manera distribuida.

Como se mencionó en el epígrafe 1.1, un requisito importante de los CC es brindar la posibilidad a los usuarios de tener espacios personales de trabajo, donde puedan almacenar los datos que con más frecuencia utilizan, y donde puedan guardar los resultados de sus investigaciones. A través de estos espacios de trabajos personales un usuario del centro puede compartir sus datos con otros usuarios. Por lo que el espacio de trabajo personal será un aspecto clave del diseño.

Manipular distintos tipos de metadatos es otro requisito importante. El manejo de catálogos de metadatos en sistemas distribuidos, facilita mantener la ubicación de cada uno de los

archivos en los diferentes nodos del cluster, así como otras informaciones importantes de los datos como fecha, el propietario y cualquier característica que sea de interés, sin entrar en detalles de los datos en sí. Los metadatos descriptivos de los datos son necesarios para comprender los datos, localizar conjuntos de datos dentro de un archivo y saber por los distintos procesos que estos han pasado.

Después de analizar aspectos y características imprescindibles a la hora de realizar el diseño, se pueden analizar los aspectos del diseño de los centros de ciencia estudiados, para ver cuales se pueden utilizar en un cluster de computadoras, para satisfacer estos requerimientos, así como ver las diferentes variantes y resaltar algunos aspectos que no se pueden utilizar.

3.1.2 Aspectos del diseño de un CC soportados sobre la tecnología de cluster de computadoras

En un cluster de computadoras se pueden implementar prácticamente todas las funcionalidades de los centros de ciencia, pero es necesario excluir algunas condiciones de heterogeneidad que poseen los CC para ajustarnos a las condiciones de un cluster de computadoras con similares configuraciones. Por ejemplo limitar las funcionalidades de un midleware para que en lugar de manipular diferentes sistemas de archivos y sistemas operativos, solo se dedique a una configuración con condiciones homogéneas.

Con el objetivo de visualizar grandes volúmenes de datos de manera rápida y eficiente, es fundamental explotar la potencia de procesamiento del cluster, entiéndase gran volumen de datos, como un conjunto que no puede ser visualizado de manera eficiente por un solo ordenador. Es por eso que dentro del modulo de visualización se deben implementar técnicas de visualización distribuida, visualización paralela y renderizado paralelo, las principales características de estas técnicas fueron discutidas en el epígrafe 1.3. Se Debe señalar que esta es una actividad bastante difícil y requiere un análisis profundo y que se puede contar con softwares de ayuda en la construcción de clusters para renderizado. Dentro de ellos se tienen a las herramientas Chromium, ParaView, VisIt, OpenDX todas de

código abierto y el HDF5 paralelo que como su nombre indica permite la visualización en paralelo de archivos HDF5.

En cuanto a la forma de almacenamiento se presentan dos variantes fundamentales, sonde se incluye el tratamiento de los datos, los metadatos y los espacios personales de trabajo:

1 Utilizar BD solo para almacenar los Catálogos de Metadatos y los archivos en formato de datos científicos (HDF, CDF, NetCDF, FITS) almacenarlos en los discos duros del cluster.

2 Tener Estructuras para almacenar objetos de formatos de datos científicos (HDF, CDF, NetCDF, FITS) en un Cluster de BD (Bases de datos objeto relacional, bases de datos orientadas a objetos, etc).

Según la primera variante, se tiene una base de datos para almacenar los catálogos de metadatos, que una vez que un usuario se autentifica le permite ubicar los datos que están replicados en los distintitos discos del cluster. Con esta variante la BD de catálogos de metadatos tiene el control de todos los datos del sistema, en caso que algún dato sea movido de lugar, se debe actualizar el catálogo.

Además con la facilidad de brindar espacios de trabajos personales, cuando un usuario se autentifica, el catálogo debe tener el control sobre el cliente, es decir permitir que el usuario tenga la localización de sus datos. Mientras que los grandes volúmenes de datos están almacenados en los discos duros del cluster, replicados entre varios nodos según la disponibilidad.

Los metadatos descriptivos estarían dentro de cada archivo. Con esta variante se pueden utilizar las facilidades que brindan algunos formatos de datos de trabajar en paralelo, tratado en el capítulo1, como el HDF paralelo, disponible a partir de la versión HDF5.

Mientras que en la segunda variante se presentan estructuras para almacenar objetos de formatos de datos científicos como (CDF, NetCDF, FITS), en un cluster de base de datos,

BD que pueden ser objeto relacional o orientadas a objetos, utilizando el gestor de BD que más nos convenga o varios gestores en caso que sea necesario.

En ésta variante el tratamiento de los catálogos de metadatos es similar a la primera. Una BD de catálogo de metadatos que puede estar replicada. Mientras que el tratamiento de los metadatos descriptivos es un poco más complicado, porque no se puede explotar las facilidades de los formatos de datos mencionados, donde los metadatos descriptivos se encuentran dentro de los archivos, una variante es tener un objeto en forma de tabla asociado a cada archivo, también en forma de tabla, que contenga sus metadatos descriptivos.

Si el trabajo con los metadatos descriptivos es un inconveniente para esta variante de almacenamiento, la posibilidad de brindar espacios de trabajo personales o BD personales es muy fácil, simplemente creando para cada usuario una BD con estructura similar a las que almacenan los grandes conjuntos de datos, que seria su BD personal, donde puede almacenar los datos de mayor interés para él.

Las BD personales de los clientes son específicas para su investigación y metas. Una vez definida, la instanciación de la BD de trabajo ha terminado con datos viniendo de las fuentes preexistentes. Esta base de datos de trabajo puede sufrir actualizaciones a partir de las fuentes de datos que son parte del esquema integrado, así como de conocimiento nuevo dado a saber por los usuarios. Los datos en la base de datos de trabajo se guardan en el formato que esté disponible o quizás el usuario pueda definir un formato deseado. La definición de esquema de la base de datos de trabajo y sus actualizaciones, son soportados por los dos esquemas de almacenamiento tratados.

En un cluster al igual que en los centros que utilizan tecnología GRID se pueden implementar buenos mecanismos de control de fallos para evitar la perdida de información y mantener la integridad de los datos. Estableciendo políticas de seguridad se evita cualquier problema que pueda causar un usuario ajeno al sistema, como la introducción de códigos malignos, por eso implementando un buen sistema de autentificación se tiene el control de los usuarios que acceden a los datos del CC. Utilizando la replicación de los

datos, se mantiene la integridad de los mismos. Haciendo una suma de chequeo de cada conjunto se puede determinar si ocurrió algún cambio o modificación que requiera la recuperación o actualización y se recuperaría copiando desde una fuente similar. Vale destacar que se debe tener un mecanismo de actualización, que sea capaz de propagar las actualizaciones de los datos a todos sus conjuntos replicados, así como a las bases de datos personales.

El hecho de contar con un cluster de computadoras implica buena capacidad de procesamiento, pero para explotar al máximo la sinergia en el cluster se necesita una buena programación de paso de mensajes, utilizando principalmente MPI, para lograr algoritmos capaces de repartir las tareas entre varios procesadores, así como balancear la carga a medida que los nodos se hacen disponibles después de terminar su tarea asignada. El uso de las BD paralelas puede ser un buen mecanismo para aumentar la capacidad de cómputo.

Una característica importante de los centros de ciencia es la colaboración entre varios de ellos. El cluster de computadoras no interviene directamente en esta tarea, el midleware es el que debe encargarse de tener enlaces con otros centros existentes. Por el momento no es objetivo de esta investigación desarrollar la colaboración con otros centros, esto puede ser una tarea a tener en cuenta en un futuro.

El diseño de la arquitectura de un centro de ciencia en un modelo lógico esta formado por tres capas fundamentales, como se muestra en la figura 3.1: En la parte superior se tiene la capa cliente, en un nivel más bajo la capa manejadora de los datos y por ultimo la capa de almacenamiento. Para facilitar la comprensión en este modelo se decidió no incluir el modulo de visualización que si se tendrá en cuenta posteriormente.

Los usuarios del centro de datos desde su estación de trabajo o computadora personal se autentifican en el sistema mediante un servicio. Localizan los datos en el catálogo de metadatos y acceden a los servicios que brinda el centro. Una vez que están dentro del sistema, pueden acceder a los datos (nivel más bajo) a través de una capa intermedia donde están las bibliotecas de manejo de datos que permiten trabajar con los distintos formatos de datos científicos (HDF, NetCDF, FITS). En esta capa intermedia también se

encuentran los sistemas manejadores de bases de datos, para trabajar con los datos que se encuentran almacenados en BD en el mismo nivel de almacenamiento. Ambos tanto bibliotecas de formatos de datos como sistemas manejadores de BD también son necesarios para acceder a los espacios de trabajo personales, que como se puede observar en la figura3.1, están el mismo nivel de los datos.

3.2 Diseño del modelo de arquitectura para los servicios y aplicaciones brindados por un Centro de Ciencia

Se realizó un diseño lógico de centro de ciencia basado en tres capas: Capa cliente, capa manejadora de datos y la capa de almacenamiento, teniendo en cuenta aspectos tales como catálogo de metadatos, servicios y espacios personales de trabajo. En la siguiente figura se muestra en que nivel interviene cada componente del CC.

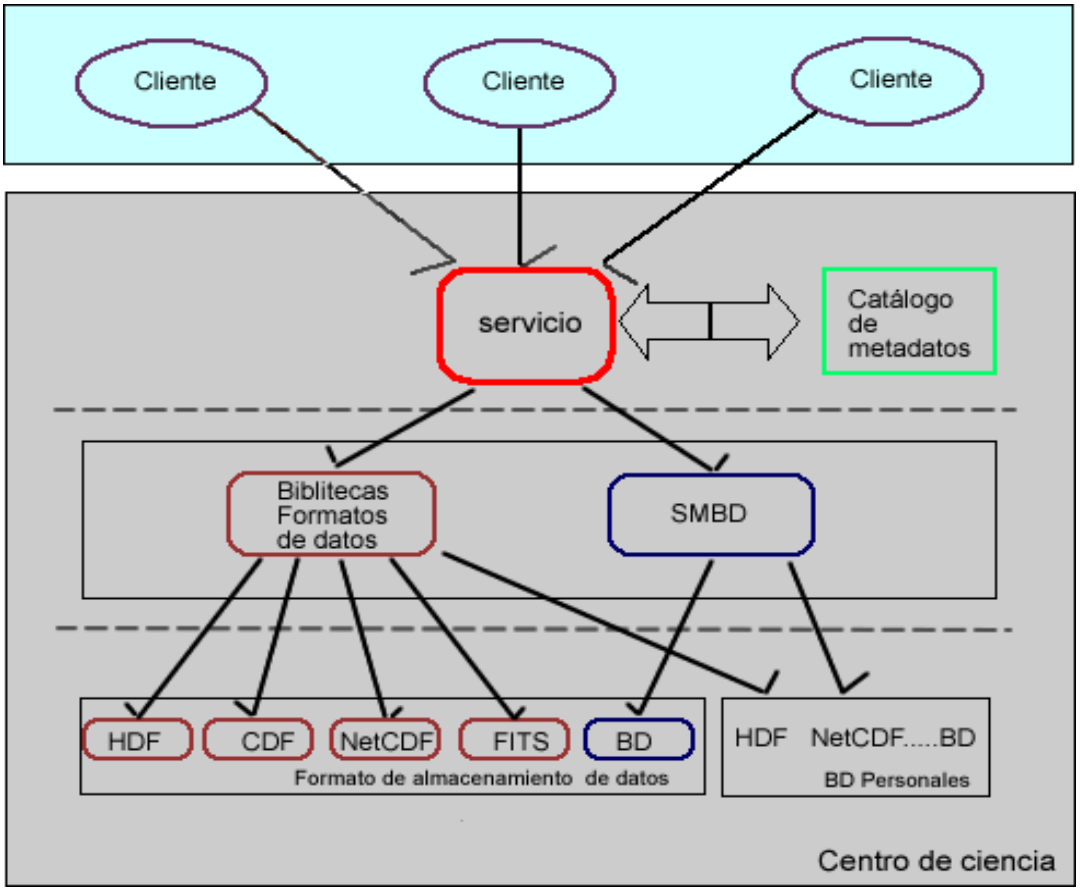


Figura 3.1 Modelo lógico del diseño del Centro de Ciencia.

3.2.1 Capa Cliente

La capa superior o capa cliente, permite que múltiples usuarios se conecten y trabajen simultáneamente en el sistema, accediendo a través de un servicio que les permite autentificarse mediante una interfaz común para todos. Una vez que un usuario se autentifica puede ver los servicios que están disponibles en ese momento o que sus privilegios le permitan acceder:

Entre los principales servicios para los cuales estará diseñado el centro de ciencia se encuentra la visualización. Una vez que un usuario se autentifica puede visualizar datos directamente desde el repositorio maestro o desde su espacio personal, las herramientas de visualización también las brindaría el centro. Además el centro ofrece la posibilidad de realizar consultas complicadas a los datos, mediante el uso del cluster y de la replicación, este trabajo puede ser muy eficiente y reducir considerablemente el tiempo con respecto a un servidor de base de datos normal, principalmente por las ventajas que ofrecen las bases de datos paralelas.

En el sistema hay que manejar distintos grupos de usuarios: Grupo de administradores, grupo de usuarios normales y grupo de usuarios temporales.

Los administradores del sistema tienen el control total de Centro de Ciencia, ellos puedes acceder directamente para publicar, mover, actualizar o eliminar datos. Publicarían nuevos datos después de analizarlos minuciosamente y saber que son reales además de que son útiles para alguna investigación, moverían los datos en caso de que necesiten una mejor organización o para aprovechar el espacio de almacenamiento, actualizarían datos que se vuelvan a obtener por nuevos mecanismos, algoritmos más sofisticados o cualquier otra fuente confiable y se eliminarían datos que ya no sean de interés para los usuarios del sistema. Todos estos procesos de trabajo con los datos serían invisibles para los demás grupos de usuarios del sistema.

Los usuarios normales serían investigadores de algún área científica que necesiten acceder al centro para aprovechar las facilidades de los servicios que este brinde. Ellos se

registrarían al sistema para consultar los datos que ya están almacenados en el centro. En caso que lo deseen pueden tener un espacio de trabajo personal, donde guardarían los datos de mayor interés o que usan con más frecuencia, en su espacio pueden guardar los resultados de su investigación y hacerlos públicos para el resto de los usuarios. También dentro de este tipo de usuarios puede que un grupo de ellos que investiguen el mismo tema quieran tener un espacio de trabajo común para colaborar entre ellos y evitar la redundancia. En caso que estos usuarios necesiten publicar nuevos datos que no se encuentren en el sistema, deben ser autorizados por los administradores del centro, que se encargarían de examinar esos datos. Los usuarios temporales serian usuarios que necesitan utilizar el sistema ocasionalmente, para realizar alguna búsqueda u obtener alguna imagen de un grupo de datos. Estos usuarios no tendrían espacios de trabajo personal, y su principal objetivo sería con fines educativos.

Otro elemento importante de este nivel es la BD de catálogo de metadatos. Como su nombre indica es una base de datos que almacena las direcciones de todos los recursos que están almacenados en el sistema, localización de los datos, espacios personales, y otras facilidades que ayude al cliente a encontrar lo mismo datos que recursos sin ser un experto en el funcionamiento del sistema, ni conocer los detalles de los datos. Cuando un usuario se autentifica, casi siempre lo primero que realiza es consultar esta BD.

A modo de conclusión se debe señalar que esta capa es muy importante en el sistema, ya que aquí es donde los usuarios interactuarían con el centro mediante una interfaz. Las funcionalidades de los demás niveles serian invisibles a los usuarios. Para la implementación requiere alto nivel de programación, para las otras capas hay herramientas que tienen las funcionalidades que necesarias. Esta capa esta enlazada directamente con la capa manejadora de datos que se trata a continuación.

3.2.2 Capa Manejadora de Datos

En la figura3.1 se observa la capa intermedia, que brinda las herramientas necesarias para poder manejar los distintos tipos de datos. Esta capa consta de dos partes bien definidas,

una donde están las bibliotecas para el manejo de los distintos formatos de datos (HDF, NetCDF….), y la otra parte formada por los sistemas manejadores de bases de datos. Ambos recursos permiten el trabajo en paralelo con los datos y el acceso de múltiples clientes simultáneamente al repositorio de datos.

En la figura3.2 se muestra detalladamente el modelo lógico de trabajo con los formatos de datos en paralelos tratado por. Aquí se observa como las aplicaciones de los usuarios que pueden ser paralelas o no, son tratadas por el sistema, en este caso se refiere a aplicaciones paralelas porque las aplicaciones secuenciales se realizan de igual forma y el principal objetivo de los centros de ciencia es explotar la fuerza de múltiples computadoras trabajando simultáneamente en una misma tarea.

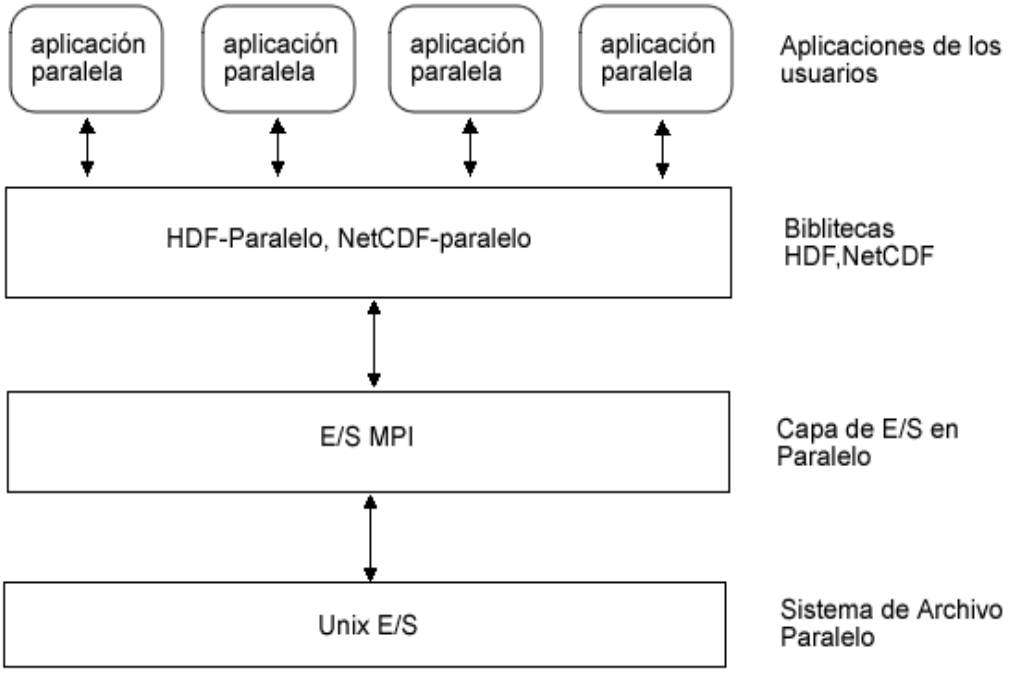


Figura 3.2 Capa de implementación de trabajo con formatos de datos en paralelo.

Las bibliotecas para el manejo de los datos (HDF, NetCDF….), tienen el soporte que permite el trabajo en paralelo con los distintos archivos almacenados en el centro de

ciencia. En un nivel inferior se tienen las aplicaciones de paso de mensajes MPI para poder distribuir las tareas entre los nodos de procesamiento del cluster y establece el orden en que van a ser procesadas, en la parte más baja el sistema de archivo, en este caso solo Linux pero para Windows el tratamiento es similar.

Estas bibliotecas de trabajo con los distintos formatos de datos permiten realizar las siguientes operaciones con los datos:

- Crear, abrir y cerrar un archivo.
- Crear, abrir y cerrar un conjunto de datos.
- Extender o reducir un conjunto de datos.
- Escribir o leer desde un conjunto de datos.

El trabajo con los datos puede de forma colectiva o independientemente. Una vez que un archivo es abierto por los procesos de comunicación:

- Todas las partes del archivo son accesibles por todos los procesos.
- Todos los objetos del archivo son accesibles por todos los procesos.
- Múltiples procesos pueden escribir en el mismo conjunto de datos.
- Cada proceso puede escribir en un conjunto de datos individual.

La comunicación mediante MPI (Message Passing Interface), permite la comunicación de un grupo de procesos entre si, esta comunicación tiene que realizarse en el orden correcto. Primeramente se abre un archivo paralelo con un comunicador, después obtiene el manipulador del archivo que va a ser usado para los accesos posteriores.

El sistema de archivo paralelo es un sistema de ficheros para cluster que proporciona alto rendimiento y acceso a la información por múltiples computadores simultáneamente. Los

conjuntos de datos están distribuidos en múltiples discos, permitiendo realizar lecturas y escrituras de los datos en paralelo. Como otros sistemas de archivos para cluster proporciona acceso de alta velocidad concurrente a aplicaciones ejecutándose en múltiples nodos del clusters. Proporciona un sistema de almacenamiento de ficheros, pone a disposición herramientas para la gestión y la administración de clusters y permite accesos compartidos al sistema de ficheros desde otros sistemas paralelos de archivos remotos. Ofrece otras funcionalidades como alta disponibilidad, el soporte para clusters heterogéneos, recuperación en caso de fallos y seguridad.

En éste mismo nivel se encuentran los sistemas manejadores de bases de datos, su propósito es el mismo de las bibliotecas de trabajo con los formatos de datos científicos. Estos SMBD tienen las herramientas necesarias para acceder y consultar los datos que se encuentran almacenados en BD en el nivel inferior.

En el caso de los Sistemas Manejadores de Bases de Datos (SMBD) el trabajo es mucho más sencillo, porque estos son mas robustos que las bibliotecas de los formatos de datos y la mayoría de ellos incorporan una gran cantidad de tareas que realizan sin que el usuario se tenga que preocupar. El usuario se tendría que concentrar en aprender el lenguaje de comunicación, para la formulación de las consultas.

Como se ha visto en esta capa, a pesar de ser invisible para el usuario, juega un papel fundamental en la interacción entre el cliente y los datos. Se debe señalar que hay muchas herramientas disponibles de código abierto, las de trabajo con bibliotecas de formatos de datos son más crudas y requiere mayor conocimiento, pero permiten explotar un alto grado de paralelismo en las tareas, mientras que los sistemas manejadores de BD son más robustos y por ende mucho más fáciles de usar, ambos sistemas manejadores de datos se encuentran en pleno desarrollo.

El SRB tratado en el capítulo 2 es un midleware que como su nombre indica (Solucionador de Recursos de Almacenamiento), controla gran parte del funcionamiento interno del sistema, del trabajo con los archivos, permite la autentificación y el manejo de los

diferentes tipos de metadatos. Es una herramienta muy robusta que utilizan casi todos de los grandes centros de ciencia del mundo, prácticamente es un estándar. Esta herramienta se logró descargar de Internet, pero no se pudo instalar porque Cuba está en la lista de los países que no les ofertan las contraseñas de desencriptación de los archivos principales de instalación. Existen otras herramientas que integradas con el SRB pudieran ser de mucho interés tenerlas en cuenta para este propósito, estas utilidades son el HDF-SRB, HDFView y otras que están en pleno desarrollo. HDFView es una interfaz muy cómoda que permite visualizar archivos HDF e interactuar con los diferentes conjuntos de datos de los archivos HDF. El HDF-SRB es una herramienta que integra SRB con el formato de datos HDF, es decir utiliza las facultades de SRB para el tratamiento de archivos HDF específicamente.

Actualmente se está desarrollando otra herramienta solucionadora de rcursos de almacenamiento, llamada iRODS que si es totalmente libre y de código abierto. Esta herramienta también tiene su base de datos de catálogos de metadatos llamada iCAT, mientras que SRB posee MCAT. Actualmente esta en construcción la herramienta HDF-iRODS similar a HDF-SRB, esta herramienta también será libre, de esta forma todos estos serán productos asequibles para todos los investigadores del mundo, HDF-iRODS se espera que sea liberada en agosto. En este diseño se cuentan con estas herramientas ya sea para utilizarlas o de otra forma se tendría que implementar una similar en el centro, tratando de cumplir con las mismas normas de los estándares.

3.2.3 Capa de Almacenamiento

La capa de almacenamiento es la encargada de soportar los formatos de datos físicos y las bases de datos en sí. Como se puede observar en la figura3.1 esta capa esta dividida en 2 secciones la de la izquierda significa que se tendrán formatos de datos científicos o BD para almacenar los grandes conjuntos de datos. La sección de la derecha significa que cada base de datos personal pudiera en principios tener las mismas funcionalidades para los almacenar formatos de datos o BD pero en menor escala.

Los datos se almacenarían en los discos duros de los nodos del cluster, en caso que estén en formatos de datos científicos, se utilizaría una herramienta de replicación, en este caso el iRODS o SRB apoyan esta función. Si los datos se encuentran en bases de datos, los mismos gestores de BD paralelas permiten realizar la replicación.

Las principales características que debe tener esta capa son: Gran capacidad de almacenamiento, que los discos sean rápidos para agilizar la comunicación entre los nodos y escalabilidad para poder aumentar la capacidad.

3.3 Módulo de visualización

Con el surgimiento de los centros de ciencia, se presentan nuevas oportunidades para el desarrollo de nuevos modelos de visualización. Estos modelos deberán considerar diferentes escenarios para distintos tipos de clientes con mayor o menor capacidad de cómputo. En cualquier caso, los datos a visualizar se concentrarán en el centro de ciencia, y los clientes interactuarán solo con la imagen que se les presenta. Este tipo de interacción, desarrollado para la visualización de fluidos bidimensionales se denomina interacción semántica, tratado en (Morell and Pérez, 2006). En los próximos años, con ayuda del soporte que brindan los centros de ciencia, este tipo de interacción debe extenderse para la visualización de otros tipos de datos.

El sistema permitirá definir una tubería de visualización (o diagramas de workflow) al cliente, utilizando los recursos que brinda el centro de ciencia. Además de explotar las facilidades que brinda la visualización distribuida, discutidas en la sección 1.3.

Cuando se trabaja con grandes volúmenes de datos, no es raro para algoritmos de visualización producir gran cantidad de geometrías y gráficos, lo cual puede sobrepasar la habilidad de un solo CPU y los aceleradores gráficos para renderizar estas imágenes. El uso de los métodos de renderizado paralelo de datos es requerido para lograr un funcionamiento adecuado. Las clasificaciones más usadas para los algoritmos de renderizado paralelo se basan en las localizaciones de la tubería de renderizado, así como el

espacio que ocupa el objeto en la pantalla. Estas clasificaciones son: sort-first, sort-middle, y sort-last, se puede encontrar una especificación detallada de cada una de ellas en el anexo 1 o en los capítulos 27 y 28 de (Hansen and Johnson, 2005).

Para abordar la estructura del modulo de visualización de forma óptima utilizando las ventajas de un cluster de computadoras, es necesario mencionar algunas de las herramientas de renderizado en paralelo con las que se pudiera contar en el diseño, como Chromium, Paraview, VisIt y OpenDx.

Chromium es una biblioteca libre y de código abierto que permite desarrollar aplicaciones de renderizado en paralelo en clusters de computadoras, posee una arquitectura completamente extensible. Soporta renderizado en paralelo de tipo sort-first, sort-last y un hibrido entre ambos. Esta biblioteca coordina renderizado con OpenGL en clusters de computadoras.

ParaView es una herramienta de código abierto y multiplataforma, diseñado para visualizar conjuntos de datos de tamaño variable, en un cluster de computadoras. Está montado sobre VTK (Visualization Toolkit).

VisIt es otra herramienta de código abierto y multiplataforma para renderizado en paralelo. Utiliza renderizado de tipo sort-first y esta diseñado para visualizar datos de grandes simulaciones.

OpenDX es una aplicación y un paquete de desarrollo de software de código abierto para la visualización de datos científicos, especialmente los adquiridos a partir de observaciones y simulaciones. Es un sistema multiplataforma de visualización de propósito general, que permite leer datos de distintas fuentes de una manera flexible y amigable. Aún cuando la organización de los datos puede ser muy diferente para cada fuente, OpenDX ofrece una manera de leer casi cualquier tipo de datos a través su herramienta conocida como Data Prompter. Ofrece una interfaz gráfica que permite crear programas de manera visual, mediante la interconexión de bloques o módulos (workfow), llamados redes. Por otro lado,

OpenDX contiene un lenguaje script para crear programas de visualización. OpenDX fue diseñado para trabajar en un ambiente cliente servidor. También fue diseñado para renderizar en paralelo en un cluster de computadoras utilizando sort-first.

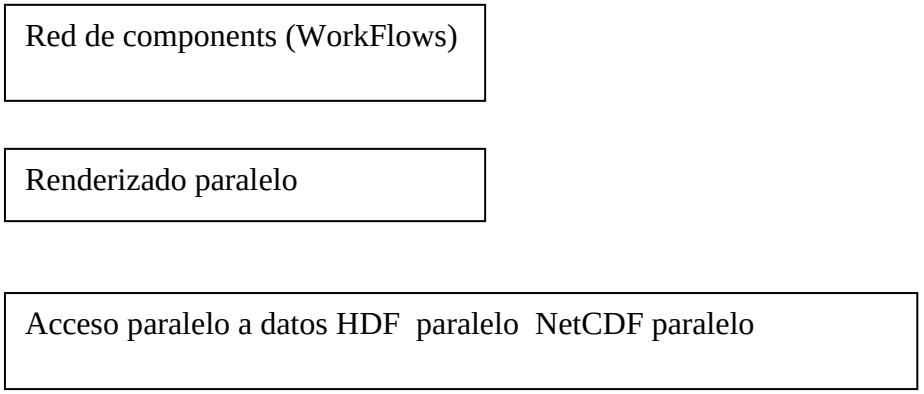


Figura 3.3 Diagrama de Visualización.

La figura anterior muestra los componentes más importantes que intervienen en el módulo de visualización, de abajo hacia arriba la primera capa: Acceso paralelo a datos HDF, NetCDF, analizados en el modelo lógico del diseño del centro de ciencia.

En la capa intermedia denominada "Renderizado paralelo", intervienen las herramientas Chromium, VisIt, ParaView y OpenDX, utilizadas con el objetivo de aprovechar la capacidad del cluster para renderizar en paralelo.

La última capa es la que está más cerca de los usuarios. La capa llamada Red de componentes (WorkFlow), es la capa que sirve de interfaz entre los usuarios y el centro de ciencia. La idea es que los usuarios puedan interactuar con el CC formulando complejas redes o flujos de trabajos (WorkFlows), especificando las tareas que se deseen realizar en el cluster. OpenDX interviene en las 2 capas superiores del diagrama anterior debido a que permite el renderizado paralelo y posee un conjunto de funciones para creación de workflows, siendo la principal herramienta a tener en cuenta en el diseño. Como es un paquete de código abierto, es posible modificarlo y adaptarlo para satisfacer los requerimientos del servicio de visualización.

Una herramienta que se puede utilizar para visualizar archivos HDF aunque no permite el renderizado paralelo es HDFVIEW. Herramienta que se pudiera utilizar para visualizar archivos HDF almacenados en el CC, e incluso en las estaciones locales. HDFVIEW permite acceso a datos HDF en un SRB a través del modulo HDF-SRB y puede ser modificado para acceder a datos HDF del iRODS a través del modulo HDF-iRODS.

3.4 Soluciones tecnológicas que satisfagan los requerimientos del CC en un cluster de computadoras

Este diseño tiene como principal requisito alto funcionamiento de E/S, la administración, el almacenamiento, el acceso eficiente y el análisis de estos datos que es una tarea sumamente desafiante. Se encontraron dos alternativas que sirven para dar solución a este problema. Almacenar los datos en archivos de formatos de datos científicos, que pueden proveer alto funcionamiento pero es una tarea tediosa para el manejo de grandes archivos y conjuntos de datos complicados. Las BD pueden ser convenientes, flexibles y poderosas pero su funcionamiento para aplicaciones paralelas es inferior y son menos eficientes para satisfacer los requerimientos de las aplicaciones científicas de gran escala. Para dar solución a este diseño se decidió combinar las buenas características de ambos sistemas, formatos de datos científicos y bases de datos.

Finalmente se decidió usar un sistema paralelo de archivos para almacenar los datos y una BD para almacenar los catálogos de metadatos. Así se aprovechan las buenas características de E/S que brinda MPI. Como consecuencia, los usuarios pueden escribir y pueden recuperar datos con alto funcionamiento paralelo, sin tener que perder tiempo en conocer los detalles de los archivos.

La definición de la interfaz de MPI-IO es parte del MPI estándar, API portable para las aplicaciones paralelas de E/S. Para las implementaciones de alto rendimiento de MPI-IO, el proveedor y las implementaciones son de dominio público y están disponibles para la mayoría de las plataformas. MPI-IO está diseñado específicamente para permitir alto funcionamiento de E/S en paralelo.

El usuario puede especificar los datos con una descripción de alto nivel, conjuntamente con anotaciones y el uso de un API para la recuperación de los mismos. Este modelo traduce internamente las peticiones de los usuarios en una llamada apropiada de MPI-IO.

El sistema permite al usuario especificar nombres y otros atributos que pueden asociar con los conjuntos de datos. Internamente selecciona un nombre de archivo en el cual se guardarán los datos. El mapeo que se establece entre los datos y los nombres de los archivos es almacenado en la BD de catálogos. El usuario puede recuperar un conjunto de datos especificando atributos de los datos deseados.

Como solucionador de recursos de almacenamiento (midleware), se propone utilizar iRODS, escogido por las facilidades que brinda para el manejo de datos, por ser de código abierto y se ajusta perfectamente a las necesidades del diseño. El iRODS instala la BD iCAT, que se utilizará para almacenar los catálogos de metadatos en un servidor postgresql. Es necesario destacar que se le pudieran realizar modificaciones al iRODS para quitarle algunas funcionalidades que ocupan recursos de cómputo y no se necesitarían en un cluster de computadoras.

El formato de datos que más se ajusta a las necesidades del diseño es el HDF, la extensión HDF-iRODS mencionada anteriormente es un modulo que se le puede agregar al servidor iRODS. Por otra parte debido a la potencia del trabajo en paralelo que permite la versión HDF5 paralelo se propone la instalación de esta herramienta.

Para las tareas de visualización se propone instalar del lado de servidor el paquete de visualización OpenDX, y en el lado del cliente una interfaz web o escritorio que permita al usuario de forma remota, explotar las facilidades que proporciona este paquete, para el renderizado en paralelo. Además se puede instalar en las estaciones de trabajo el HDFVIEW, para visualizar datos que se encuentran en el centro o en las estaciones locales de los usuarios.

Conclusiones

El nuevo estilo de trabajo que se presenta con el surgimiento de los CC permitirá acelerar las investigaciones en las distintas áreas de investigación, sin dejar de tener en cuenta la alta demanda de recursos computacionales necesarios para su creación. El objetivo de esta investigación fue propiciar un conjunto de herramientas y consideraciones que permitiera extender esta tecnología a instituciones que tengan limitaciones en cuanto a la disponibilidad de recursos.

1. Se determinaron las tendencias actuales en el desarrollo de centros de ciencia de áreas específicas, teniendo en cuenta aspectos tales como tecnologías, arquitectura y servicios brindados, a partir del estudio de algunos de los grandes CC de mayor auge en los últimos años.

2. Se precisaron qué aspectos del diseño pueden ser soportados sobre la tecnología de cluster de computadoras.

3. Se definieron como implementar las funcionalidades de los CC sobre un cluster de computadoras.

4. Se diseñó un modelo de arquitectura para los servicios y aplicaciones brindados por un CC que permita la creación del mismo en un futuro y se proponen un conjunto de soluciones tecnológicas que satisfagan los requerimientos de los CC.

Recomendaciones

1 Dada la variedad de posibles soluciones y la disponibilidad de herramientas libres, el estudio realizado no agota este campo de investigación. Por lo que se propone estudiar con mayor profundidad otras soluciones y las posibles transformaciones que se le puedan realizar a cada una de las herramientas.

2 Profundizar en las soluciones tecnológicas propuestas, principalmente las relacionadas con el modulo de visualización.

3 Extender el modelo para permitir la colaboración entre los distintos centros de ciencia basados en cluster de computadoras.

4 Realizar una comparación entre la solución propuesta en cuanto al almacenamiento y el uso de cluster de BD.

Referencias Bibliográficas

ALONSO, J. M. (1997) Programación de aplicaciones paralelas con MPI (Message Passing Interface).

BABAR (2007) BaBar Public Web Home Page.

BIRN (2008) Biomedical Informatics Research Network.

BOGHOSIAN, B. Y. P. V. C. (2005) Scientific Applications Of Grid Computing.

CASJOBS (2008) SDSS CasJobs site.

CHENG, A., KOZIOL, Q. & WENDLING, B. (2003) Flexible Parallel HDF5.

DEWITT2, D. J., GRAY, J., DEPARTMENT, C. S. & CENTER, S. F. S. (1992) Parallel Database Systems:

The Future of High Performance Database Processing1.

GENBANK (2008) GenBank Overview.

GRAY, J. E. A. (2005) Scientific Data Management in the Coming Decade.

HANSEN, C. D. & JOHNSON, C. R. (Eds.) (2005) *The Visualization Handbook*.

HEIJMANS, J. (2002) An Introduction to Distributed Visualization.

MOORE, R. W. (2006) Digital Libraries and Data Intensive Computing.

MOORE, R. W. & JAGATHEESAN, A. (2004) DATA GRID MANAGEMENT SYSTEMS.

MORELL, A. & PÉREZ, C. (2006) Biblioteca de módulos de visualización de fluidos para OpenDX.

NCBI (2008) National Center for Biotechnology Information.

ÖZSU, M. E. A. (2005) Preventive Replication in a Database Cluster.

PAUL WATSON & PATON, N. (2003) Grid Data Management Systems & Services.

PETER CAO, MIKE FOLK, MIKE WAN & REAGAN MOORE (2005) Integration of HDF5 and the SRB for Object-level Data Access.

RAJASEKAR, A., WAN, M., MOORE, R., SCHROEDER, W., KREMENEK, G., JAGATHEESAN, A., COWART, C., ZHU, B., CHEN, S.-Y. & OLSCHANOWSKY, R. (2003) Storage Resource Broker – Managing Distributed Data in a Grid.

SDSS (2008) What is the Sloan Digital Sky Survey.

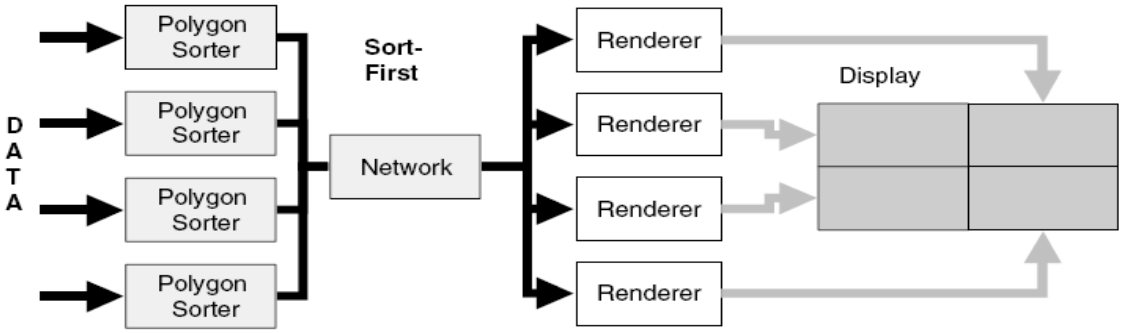
SHANKAR, S., ET AL (2006) Data Driven Workflow Planning in Cluster management Systems.

TAVERNA (2008) Taverna project website.

Anexos

Anexo 1. Sort-first, Sort-middle, Sort-Last.

Sort-First.



Los algoritmos Sort-first empiezan por distribuir primitivas de gráficos en el principio de la tubería de renderizado. Asignando las primitivas a los procesadores, subdividiendo la imagen de salida y asignando un procesador para maniobrar cada región resultante. Una vez que las primitivas han sido asignadas, cada proceso completa la tubería de gráficos para producir la subimagen final. La asignación inicial de primitivas para procesadores es el paso crucial en los algoritmos sort-first, requiere la transformación de las primitivas en coordenadas del espacio de escritorio e introduce unos gastos fijos computacionales al algoritmo. La asignación inicial para los procesadores se realiza generalmente de forma arbitraria, excepto después de que cada procesador completa la etapa de transformación de la tubería que reasigna primitivas para los procesadores correctos. La consideración adicional debe ser dada a las primitivas que traslapan las regiones subdivididas de la imagen final.

Sort-first es ventajoso porque los procesadores implementan la tubería completa de renderizado. Permitiendo el uso de una tubería acelerada por hardware, y / o buen comportamiento del sistema de memoria cache. Además, los requisitos de comunicación entre los procesadores pueden ser bajos, resultantes de una menor sobrecarga y mejor funcionamiento. La desventaja principal es que la distribución inicial de las primitivas entre los procesadores fácilmente puede conducir a un desequilibrio de carga de trabajo. Tratar

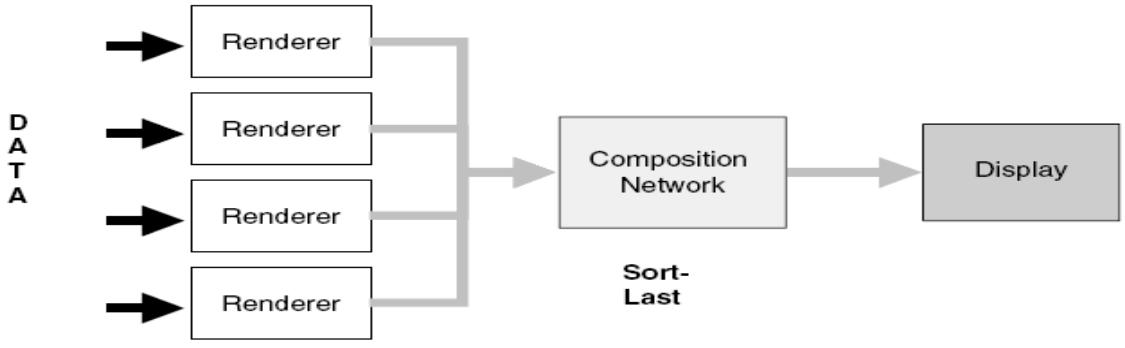
de redistribuir las primitivas durante el proceso de renderizar también puede conducir a una pobre escalabilidad debido a las cantidades de tráfico requerido por el envío de mensajes.

Sort-Middle.

En los algoritmos sort-middle, la redistribución de datos ocurre entre el procesamiento de la geometría y los estados de conversión y exploración de la tubería de renderizado. En este caso, es común describir los pasos como dos conjuntos de operaciones. La primera parte maniobra la porción de la geometría de la tubería (las transformaciones, y el alumbrado, etc.), y las primitivas están inicialmente asignadas en una forma arbitraria. El segundo grupo de operaciones es asignado a regiones continuas de la imagen final de salida, y estas operaciones son responsables de convertir una imagen vectorial en una foto computarizada en el espacio de pantalla, que las primitivas produjeron por el primer conjunto de procesadores. Note que cabe dedicarle procesadores a cada uno de estos conjuntos separadamente o dejar todos los procesadores realizar ambas tareas. Si los procesadores se usan de forma separada, cabe crear un conjunto de tubería en paralelo.

La división en la tubería de renderizado es la desventaja más notable de sort-middle, dificulta aplacar el acelerador de hardware de renderizado. Además, el costo de comunicación entre las etapas de la tubería escala con el número de primitivas renderizadas. Finalmente, el algoritmo puede padecer de desequilibrio de carga cuando las primitivas no son uniformemente distribuidas a través de la imagen de salida.

Sort-Last.



El método sort-last atrasa la ordenación de las primitivas hasta las etapas finales de la tubería de renderizado. Las primitivas pueden estar inicialmente asignadas en forma arbitraria, y cada procesador obtiene una porción de la imagen final. Todas estas imágenes luego son compuestas para formar la imagen completa final. Al igual que con sort-middle, cabe usar dos conjuntos de procesadores. La primera parte responsable de completar la tubería de renderizado, y el segundo conjunto es responsable de crear la imagen final. Esta asignación también tiene prevista la creación de una tubería paralela, que puede usarse para mejorar el funcionamiento global. Desde que las etapas de renderización de sort-last crean imágenes de resolución completa, la red de interconexión entre procesadores debe tener muy alto ancho de banda si se quiere renderizar interactivamente. Otras técnicas pueden usarse para reducir la cantidad de tráfico requerido para completar la etapa de composición.

Sort-last incluye la habilidad de implementar la tubería de renderizado completa. Es más fácil distribuir las primitivas uniformemente para los procesadores y así evitar asuntos de balanceo de carga. La principal desventaja es el alto costo de comunicación que introduce el envío de grandes cantidades de datos entre los procesadores. Las técnicas para reducir el costo de este tráfico de mensaje son todavía un área de investigación activa.