



UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES
PROYECTO DE GRADO**

Tema:

**“DISEÑO DE ALTA DISPONIBILIDAD EN SERVIDORES
SOLARIS 10 PARA APLICACIÓN EN EL CENTRO DE
COMPUTO DE LA U.C.S.G”**

Realizado por:

**ANDRÉS ANÍBAL LEÓN FEIJOÓ
ELIZABETH DEL ROCÍO GÓMEZ GUANGA**

Previo a la obtención del título de:

INGENIERO EN SISTEMAS COMPUTACIONALES

**GUAYAQUIL – ECUADOR
2010**

TRABAJO DE GRADO

Tema:

**“DISEÑO DE ALTA DISPONIBILIDAD EN SERVIDORES SOLARIS 10 PARA
APLICACIÓN EN EL CENTRO DE COMPUTO DE LA UCSG.”**

Aprobado por:

Dr. Ing. Walter Mera
DECANO

Ing. Vicente Gallardo
DIRECTOR DE CARRERA

Mgs. Ing. Cesar Salazar
DIRECTOR DEL TRABAJO DE GRADO

Ing. Edison Toala Quimí
VOCAL 1

Ing. Fernando Castro Aguilar
VOCAL 2

Andrés León Feijoó
AUTOR

Elizabeth Gómez Guanga
AUTOR

AGRADECIMIENTO

Agradecemos a Dios por darnos las fuerzas necesarias en los momentos que más lo necesitamos y bendecirnos con la oportunidad de terminar este trabajo, además damos las gracias con especial dedicación a nuestros padres por su apoyo incondicional y desinteresado a lo largo de esta carrera. De igual manera a nuestros profesores que han aportado con su experiencia y conocimientos. A nuestros amigos que son las personas con las que compartimos momentos gratos y tristes a la vez. También a todas y aquellas personas que de una u otra manera, colaboraron o participaron en la realización de este trabajo.

Andrés, Elizabeth

ÍNDICE DE CONTENIDO

PÁGINA

AGRADECIMIENTO	II
PRÓLOGO.....	1
INTRODUCCIÓN	2
OBJETIVOS.....	3
<i>General</i>	3
<i>Específicos</i>	3
ALCANCE.....	4
CAPITULO 1 : ALTA DISPONIBILIDAD	5
1.1. CONCEPTOS GENERALES	5
1.1.1. ALTA DISPONIBILIDAD (HIGH AVAILABILITY)	5
1.1.2. DISPONIBILIDAD	5
1.1.3. TIEMPO DE INACTIVIDAD PLANIFICADO.....	5
1.1.4. TIEMPOS DE INACTIVIDAD NO PLANIFICADO	5
1.2. CLÚSTER DE ALTA DISPONIBILIDAD.....	6
1.2.1. DEFINICIÓN	7
1.2.2. CLÚSTER DE ALTA DISPONIBILIDAD.....	8
1.2.3. ALTA DISPONIBILIDAD DE INFRAESTRUCTURA	8
1.2.4. ALTA DISPONIBILIDAD DE APLICACIÓN	8
1.3. COMPONENTES DE UN CLÚSTER	9
1.3.1. NODOS.....	10
1.3.1.1. TIPOS DE NODOS	11
1.3.1.1.1. NODOS DEDICADOS	11
1.3.1.1.2. NODOS NO DEDICADOS	11
1.3.1.2. CONFIGURACIONES DEL NODO	11
1.3.1.3. TÉCNICAS O CARACTERÍSTICAS DE LOS NODOS.....	12
1.3.2. SISTEMA OPERATIVO	13
1.3.3. CONEXIONES DE RED.....	14
1.3.4. ALMACENAMIENTO.....	16
1.3.5. MIDDLEWARE	18
1.3.6. AMBIENTES DE PROGRAMACIÓN PARALELA	19
1.3.7. DISEÑO DE APLICACIONES REQUISITOS	19
1.4. BALANCEO DE CARGA	21
1.4.1. DEFINICIÓN	21
1.4.2. CLÚSTER DE BALANCEO DE CARGA	21
1.4.3. SOLUCIÓN DE BALANCEO DE CARGA.....	21
CAPITULO 2 : SOLUCIÓN DE SUN MICROSYSTEM	23
2.1. LOS INICIOS DE ORACLE.....	23
2.2. HISTORIA DE SUN MICROSYSTEMS.....	25
2.3. LA COMPRA DE SUN POR PARTE DE ORACLE	27
2.4. ORACLE SUN CLÚSTER	28
2.4.1. INTRODUCCIÓN	28
2.4.2. SUN CLÚSTER 3.....	28
2.4.3. COMPONENTES DE SUN CLÚSTER.	29

2.4.3.1. NOMBRE GLOBAL DEL CLÚSTER.....	30
2.4.3.2. NOMBRES DE HOST PRIVADO	30
2.4.3.3. RED PUBLICA	31
2.4.3.4. RED PRIVADA	32
2.4.3.5. INTERCONEXIÓN DEL CLÚSTER	34
2.4.3.6. CONMUTADOR DE TRANSPORTE	35
2.4.3.7. ADAPTADORES DE TRANSPORTE.....	35
2.4.3.8. FENCING GLOBAL.....	35
2.4.3.9. GLOBAL DEVICES	36
2.4.3.10. DISPOSITIVO DE QUÓRUM.	36
2.4.3.11. RESERVACIÓN DE SCSI	39
2.4.3.11.1. INTERNET SCSI (ISCSI)	39
2.4.3.11.2. VOTACIÓN DEL QUÓRUM	39
2.4.3.12. CONFIGURACIONES DE QUÓRUM.....	40
2.4.3.12.1. DOS NODOS	40
2.4.3.12.2. MÁS DE DOS NODOS	41
2.4.4. SOLARIS CLÚSTER EDICIÓN GEOGRÁFICA.....	42
2.4.5. AGENTES	43
CAPITULO 3 : METODOLOGÍA	45
CAPITULO 4 : DISEÑO DE ALTA DISPONIBILIDAD	46
4.1. SITUACIÓN ACTUAL.....	46
4.2. EQUIPAMIENTO.....	46
4.3. SERVICIOS	47
4.4. ESTRUCTURA.....	48
4.5. DISEÑOS DE ALTA DISPONIBILIDAD PROPUESTOS.....	49
4.5.1. DISEÑO 1.....	49
4.5.1.1. DESCRIPCIÓN.....	49
4.5.2. DISEÑO 2.....	50
4.5.2.1. DESCRIPCIÓN.....	50
4.5.3. DISEÑO 3.....	52
4.5.3.1. DESCRIPCIÓN.....	52
4.6. SELECCIÓN DEL DISEÑO DE ALTA DISPONIBILIDAD.....	53
4.6.1. DESCRIPCIÓN.....	53
CAPITULO 5 : PROTOTIPO	54
5.1. DISEÑO.....	54
5.2. ESCENARIO 1:	54
5.3. ESCENARIO 2	59
CAPITULO 6 : CONCLUSIONES Y RECOMENDACIONES.....	64
6.1. CONCLUSIONES	64
6.2. RECOMENDACIONES.....	66
CAPITULO 7 : BIBLIOGRAFÍA	67
CAPITULO 8: ANEXOS	70
8.1. ANEXO 1.....	70
8.2. ANEXO 2.....	73

PRÓLOGO

El siguiente trabajo propone diseñar una alternativa de solución de alta disponibilidad con muy bajo costo, frente a soluciones propietarias con costos elevados para diseñar un esquema de alta disponibilidad utilizando la herramienta Sun Clúster en el sistema operativo Solaris 10 en Arquitectura Intel, ya que no se cuenta con los recursos que tiene la universidad, pero también puede ser aplicado para equipos con arquitectura Sparc que es con lo que cuenta la UCSG. Logrando con esto una alta disponibilidad para el servicio de Application Server del Centro de Cómputo de la UCSG, además ser guía de referencia para una futura implementación en esta institución.

INTRODUCCIÓN

Los sistemas de alta disponibilidad (High Availability), están diseñados con el propósito de minimizar el tiempo de recuperación de un sistema ante interrupciones de cualquier tipo (software o hardware).

El objetivo de esta tesis es el diseño de un sistema de alta disponibilidad en la arquitectura Intel con el sistema operativo SOLARIS 10, para que sirva como referencia en el futuro con la implementación de la alta disponibilidad en los Application server de la UCSG.

Solaris Clúster (a veces llamado Sun Clúster o SunCluster), es un clúster de alta disponibilidad del producto de software para el sistema operativo Solaris, creado por Sun Microsystems que ahora pertenece a Oracle. Se utiliza para mejorar la disponibilidad de servicios de software tales como bases de datos, compartir archivos en una red, sitios web de comercio electrónico u otras aplicaciones. Sun Clúster trabaja con equipos redundantes o nodos en los que uno o más ordenadores continuarán proporcionando el servicio si otra llegase a fallar. Los nodos pueden estar situados en el mismo centro de datos o en diferentes lugares.

El trabajo a realizar incluye un estudio preliminar y una comparación de las prestaciones de herramientas y sistemas aplicables al problema. En una segunda fase se elaborará un prototipo sobre el que se diseñará un plan de pruebas automáticas que validen su funcionamiento. Posiblemente, será necesario crear algunos programas o scripts para cubrir necesidades específicas.

OBJETIVOS

GENERAL

Lograr la alta disponibilidad para el servicio de Application Server del Centro de Cómputo de la UCSG.

ESPECÍFICOS

➤ Determinar la situación actual del servidor de aplicaciones de la UCSG, así como los problemas de disponibilidad del servicio.

➤ Diseñar un esquema de alta disponibilidad utilizando la herramienta Sun Clúster, desarrollada por Sun Microsystems para arquitecturas Intel o Sparc en el sistema operativo Solaris 10.

➤ Ser una guía de referencia para una futura implementación en esta institución.

ALCANCE

➤ Obtener un diseño de alta disponibilidad que este optimizado para el sistema de servicios administrativos, que se ejecuta bajo una arquitectura Sparc con servidores Sun Fire V490, en plataforma tecnológica Oracle Application Server 10gR2.

➤ El diseño de un esquema de alta disponibilidad, estará basado en la herramienta proporcionada por Oracle Sun Microsystems, Sun Clúster 3.2, compatible con la arquitectura Intel y Sparc. Se considerará la infraestructura que tiene actualmente el Centro de Cómputo con los recursos ya existentes.

➤ En base al diseño propuesto se elaborará un prototipo para la validación del mismo.

CAPITULO 1 : ALTA DISPONIBILIDAD

1.1. CONCEPTOS GENERALES

1.1.1. Alta disponibilidad (High Availability)

La alta disponibilidad consiste en una serie de medidas que garantizan la disponibilidad del servicio, es decir, asegurará que un servicio funcione las 24 horas.

1.1.2. Disponibilidad

Se refiere a la interconexión de dos o más computadoras con la probabilidad de que un servicio funcione en cualquier momento. Si un usuario no puede acceder al sistema se dice que está no disponible. Si el sistema no está disponible se utiliza generalmente el término tiempo de inactividad (downtime).

1.1.3. Tiempo de inactividad planificado

Es el resultante de un mantenimiento que es perjudicial para la operación del sistema y usualmente no puede ser evitado con la configuración del sistema actualmente instalado. Eventos que generan tiempos de inactividad planificados, quizás incluyen parches al software del sistema que requieran un reinicio o cambios en la configuración del sistema que toman efecto después de reiniciar. En general el tiempo de inactividad planificado es comúnmente el resultado de un evento lógico o de gestión iniciado.

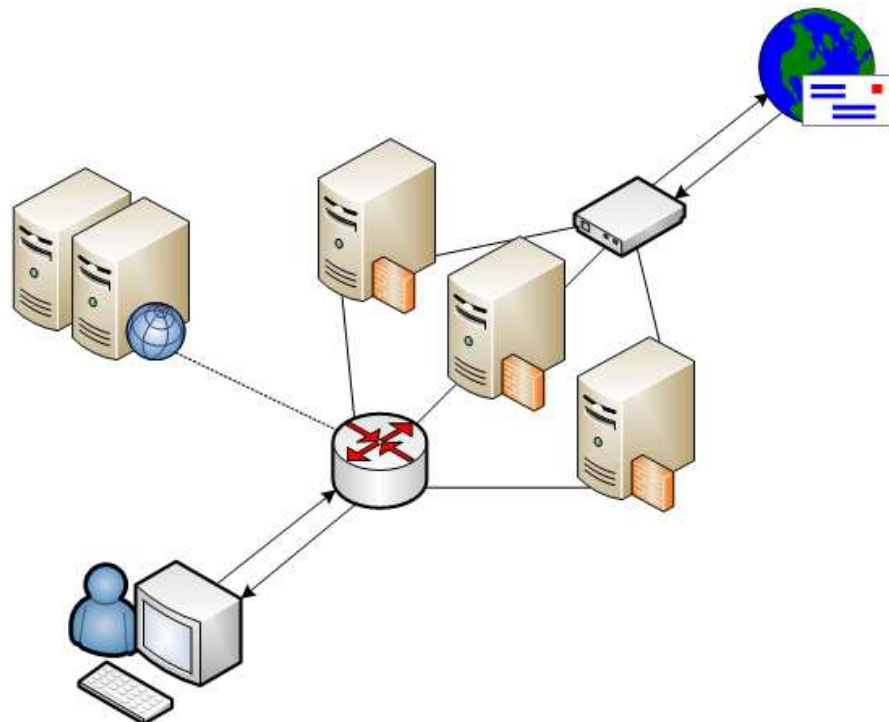
1.1.4. Tiempos de inactividad no planificado

Surgen de algún evento físico tales como fallos en el hardware o anomalías ambientales. Ejemplos de eventos con tiempos de inactividad no planificados incluyen fallos de potencia, fallos en los componentes de CPU o memoria RAM, una caída por recalentamiento, una ruptura lógica o física en las conexiones de red, rupturas de seguridad catastróficas o fallos en el sistema operativo, aplicaciones y capa media.

Muchos puestos computacionales excluyen el tiempo de inactividad planificado de los cálculos de disponibilidad, asumiendo, correcta o incorrectamente, que el tiempo de actividad no planificado tiene poco o ningún impacto sobre la comunidad de usuarios computacionales. Excluyendo tiempo de inactividad planificado, muchos sistemas pueden reclamar tener alta disponibilidad fenomenal, la cual da la ilusión de disponibilidad continua. Sistemas que exhiben verdadera disponibilidad continua son comparativamente raros y caros, y ellos tienen diseños cuidadosamente implementados que eliminan cualquier punto de fallo y permiten que el hardware, la red, el sistema operativo, middleware, actualización de aplicaciones, parches y reemplazos se hagan en línea.

1.2. CLÚSTER DE ALTA DISPONIBILIDAD

Gráfico 1. Clúster.



Elaborado por: Autores

1.2.1. Definición

Clúster es un grupo de múltiples ordenadores unidos mediante una red y es visto como uno solo, más potente que los de escritorio; estos se emplean para mejorar la disponibilidad y rendimiento. Los clúster han evolucionado mucho desde su primera aparición, ahora se pueden crear distintos tipos de clúster, en función a lo que se necesite como puede ser:

- Unión de Hardware
- Clúster de Software
- Alto rendimiento de bases de datos

De un sistema de este tipo se espera que presente combinaciones de los siguientes servicios:

- Alto rendimiento
- Alta disponibilidad
- Equilibrio de carga
- Escalabilidad

Para que un sistema clúster funcione como tal, no basta solo con conectar entre sí los computadores, sino que es necesario proveer un sistema de administración del clúster, el cual se encargue de interactuar con el usuario y los procesos que corren en él para optimizar su funcionamiento. Además no es necesario que todas las máquinas dispongan del mismo Hardware y sistema operativo (clúster heterogéneo). Este tipo de sistemas debe de disponer de un interfaz de manejo de clúster, el cual se encargue de interactuar con el usuario y los procesos, repartiendo la carga entre las diferentes máquinas del clúster.

1.2.2. Clúster de alta disponibilidad

Es un conjunto de dos o más servidores que se caracterizan por mantener una serie de servicios compartidos y por estar constantemente monitorizándose entre sí. Si se produce un fallo del hardware o de los servicios de alguno de las máquinas que forman el clúster, el software de alta disponibilidad es capaz de reiniciar automáticamente los servicios que han fallado en cualquiera de los otros equipos del clúster, y cuando el servidor que ha fallado se recupera, los servicios se migran de nuevo a la máquina original. El clúster de alta disponibilidad se lo puede dividir en dos clases:

1.2.3. Alta disponibilidad de infraestructura

Si se produce un fallo de hardware en alguna de las máquinas del clúster, el software de alta disponibilidad es capaz de iniciar automáticamente los servicios en cualquiera de las otras máquinas del clúster (failover), y cuando la máquina que ha fallado se recupera, los servicios son nuevamente migrados a la máquina original (failback). Esta capacidad de recuperación automática de servicios garantiza la alta disponibilidad de los servicios ofrecidos por el clúster, minimizando así la percepción del fallo por parte de los usuarios.

1.2.4. Alta disponibilidad de aplicación

Si se produce un fallo del hardware o de las aplicaciones de alguna de las máquinas del clúster, el software de alta disponibilidad es capaz de iniciar automáticamente los servicios que han fallado en cualquiera de las otras máquinas del clúster. Y cuando la máquina que ha fallado se recupera los servicios son nuevamente migrados a la máquina original. Esta capacidad de recuperación automática de servicios garantiza la integridad de la información, ya que no hay pérdida de datos, evitando molestias a los usuarios, que no tienen por qué notar que se ha producido un problema.

Los clústeres de alta disponibilidad conocido también como High Availability Clúster o Clúster de conmutación por error, son grupos de computadoras que se implementan con el propósito principal de ofrecer una alta disponibilidad de los servicios que ofrece el grupo. Operan por tener redundantes ordenadores o nodos que se utilizan para proporcionar el servicio cuando los componentes del sistema fallan. High availability remedia esta situación mediante la detección de hardware y errores de software e inmediatamente reiniciar la aplicación en otro sistema, sin necesidad de una intervención administrativa o de un proceso conocido como conmutación por error. Como parte de este proceso, el software de clustering puede configurar el nodo antes de comenzar el uso en él.

Un clúster de alta disponibilidad es utilizado a menudo para entornos críticos de bases de datos, compartición de archivos en una red, aplicaciones empresariales y servicios como el comercio electrónico. Los clúster suelen utilizar una conexión de red privada, también llamado latidos del corazón que se utiliza para controlar la salud y el estado de cada nodo del clúster.

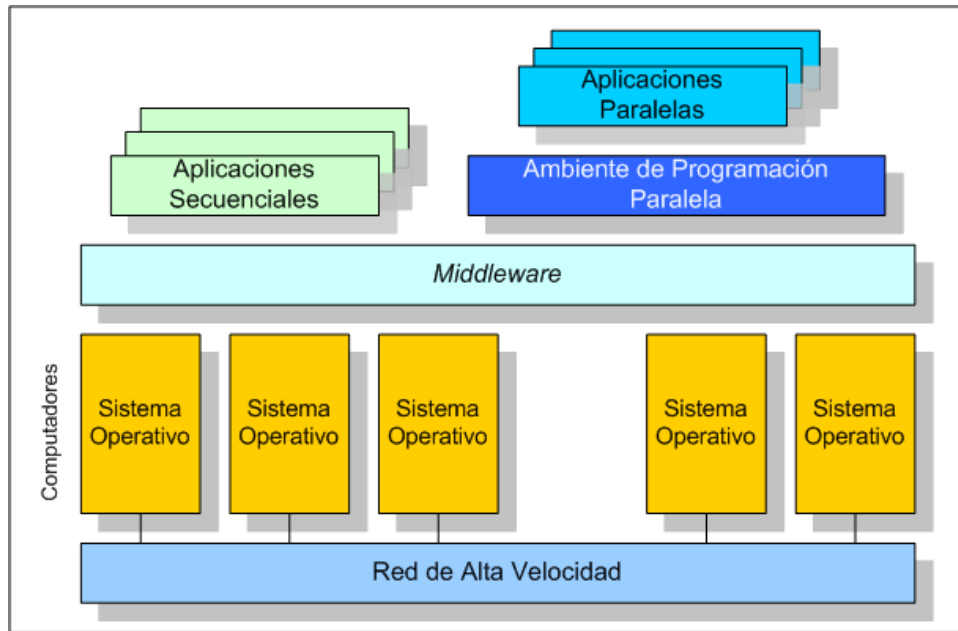
Uno de los principales problemas que se presenta al momento de implementar un clúster es el Split-cerebral, este se presenta cuando todos los enlaces privados bajan al mismo tiempo, pero los nodos del clúster se siguen publicando. Si eso sucede, cada nodo del clúster por error puede decidir que todos los demás nodos estén fuera de servicio y tratar de empezar a prestar servicios que otros nodos estaban publicando. Después de haber duplicado los servicios pueden causar daños en los datos en el almacenamiento compartido ya que cada uno de los nodos está escribiendo en dicho medio y así existiría inconsistencia.

1.3. COMPONENTES DE UN CLÚSTER

Son varios los componentes de hardware y software que necesita un clúster para su funcionamiento. Entre los principales se mencionan los siguientes:

- Nodos.
- Sistema operativo.
- Conexión de Red.
- Almacenamiento.
- Middleware.
- Ambientes de Programación Paralela.
- Aplicaciones.

Gráfico 2. Componentes de un clúster.



Fuente: Clúster

Elaborado por: Advanced Networks & Support

1.3.1. Nodos

Nodo es un punto de intersección o unión de varios elementos que confluyen en el mismo lugar. Estos pueden ser ordenadores de escritorio o servidores, de hecho se puede establecer un clúster con cualquier tipo de máquina.

Ahora bien, dentro de la informática la palabra nodo puede referirse a conceptos diferentes según el ámbito en el que se encuentre:

➤ En las redes de computadoras cada una de las máquinas es un nodo, y si la red es Internet, cada servidor constituye también un nodo.

➤ En estructuras de datos dinámicas un nodo es un registro que contiene un dato de interés, y al menos un puntero para referenciar (apuntar) a otro nodo. Si la estructura tiene sólo un puntero, la única estructura que se puede construir con él es una lista y si el nodo tiene más de un puntero ya se pueden construir estructuras más complejas como árboles o grafos.

1.3.1.1. Tipos de nodos

El clúster puede estar conformado por nodos dedicados o por nodos no dedicados.

1.3.1.1.1. Nodos dedicados

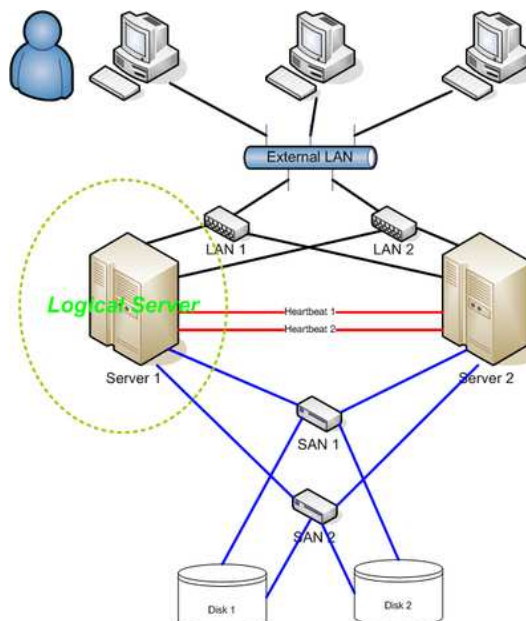
Estos nodos no disponen de teclado, ratón ni monitor y su uso está exclusivamente dedicado a realizar tareas relacionadas con el clúster.

1.3.1.1.2. Nodos no dedicados

Los nodos disponen de los dispositivos anteriormente excluidos y su uso no está exclusivamente dedicado a realizar tareas relacionadas con el clúster. El clúster hace uso de los ciclos de reloj que el usuario del computador no está utilizando para realizar sus tareas.

1.3.1.2. Configuraciones del nodo

Gráfico 3. Nodo de clúster de alta disponibilidad, diagrama de red.



Fuente: High-Availability cluster
Elaborado por: William George Herbert

El equipo lógico o host lógico de clúster se utiliza para describir la dirección de red, para acceder a uno o varios de los servicios prestados por el clúster. Esta identidad no está vinculada a un nodo de clúster único. En realidad, es una dirección de red/host que está vinculado con el servicio proporcionado por la agrupación. Si un nodo de clúster con una base de datos ejecutando baja, la base de datos se reiniciará en otro nodo de clúster y la dirección de red que los usuarios utilizaban para acceder se mantiene.

1.3.1.3. Técnicas o características de los nodos

Los clúster de alta disponibilidad suelen utilizar todas las técnicas disponibles para realizar los distintos sistemas, y la infraestructura para compartir lo más fiable posible. Estos incluyen:

- Disco reflejo de lo que el fallo de los discos internos no da lugar a fallos del sistema.

- Redundante red conexiones para que un solo cable, switch, o interfaz fallas en la red no dan lugar a interrupciones de la red.

- Redundantes de almacenamiento de red de área o conexiones de datos S.A.N de manera que un solo cable, switch, o fallos de interfaz no conduzcan a la pérdida de conectividad de almacenamiento.

- Redundantes de energía eléctrica en los circuitos de entradas diferentes, por lo general ambas o todas las protegidas por sistema de alimentación ininterrumpida de unidades, y redundantes de alimentación de unidades, de modo que, de alimentación eléctrica por cable, UPS, o la fuente de alimentación las fallas simples no conducen a una pérdida de potencia al sistema.

Estas características ayudan a reducir al mínimo las posibilidades, de que la conmutación por error entre los sistemas de agrupamiento sea necesaria. En tal conmutación por error, el servicio prestado no está disponible durante un mínimo tiempo, así que las medidas para evitar la conmutación por error son las preferidas.

Cabe aclarar que a la hora de diseñar un clúster, los nodos deben tener características similares, es decir, deben guardar cierta similitud de arquitectura y sistemas operativos, ya que si se conforma un clúster con nodos totalmente heterogéneos será ineficiente debido a que el middleware delegará o asignará todos los procesos al nodo de mayor capacidad de cómputo y solo distribuirá cuando este se encuentre saturado de procesos; por eso es recomendable construir un clúster cuyos nodos sean lo más similares posible.

1.3.2. Sistema operativo

Un sistema operativo debe ser multiproceso y multiusuario, cuanto más fácil sea el manejo del sistema menores problemas se tendrá. Un sistema operativo es un programa o conjunto de programas de computadora destinado a permitir una gestión eficaz de sus recursos.

Esté comienza a trabajar cuando se enciende el computador, y gestiona el hardware de la máquina desde los niveles más básicos, permitiendo también la interacción con el usuario, y se puede encontrar normalmente en la mayoría de los aparatos electrónicos que utilicen microprocesadores para funcionar, ya que gracias a estos se puede entender la máquina y que ésta cumpla con sus funciones (teléfonos móviles, reproductores de DVD, radios, computadoras, etc.).

Ejemplos de sistemas operativos multiproceso y multiusuario:

➤ **GNU/Linux**

ABC GNU/Linux2, OpenMosix, Rocks3, Cóndor, Sun Grid

➤ **Unix**

Solaris, HP-UX, Aix, FreeBSD

➤ **Windows**

NT, 2000 Server, 2003 Server, 2008 Server

➤ **Mac OS X**

Xgrid

1.3.3. Conexiones de Red

Las conexiones utilizadas en este tipo de sistema pueden ser muy variadas, se pueden utilizar desde simples conexiones Ethernet con placas de red comunes (adaptadores de red o N.I.C's Network Interface Card), o utilizarse sistemas de alta velocidad como Fast Ethernet, Gigabit Ethernet, Myrinet, Infiniband, S.C.I., etc.

➤ **Ethernet**

Son las redes utilizadas en la actualidad, debido a su relativo bajo costo. No obstante, su tecnología limita el tamaño de paquete, realizan excesivas comprobaciones de error y sus protocolos no son eficientes, y sus velocidades de transmisión pueden limitar el rendimiento de los clúster., para aplicaciones con paralelismo de grano grueso puede suponer una solución acertada.

La opción más utilizada en la actualidad es Gigabit Ethernet (1 Gbit/s), siendo emergente la solución 10 Gigabit Ethernet (10 Gbit/s). La latencia de estas tecnologías está en torno a los 30-100 μ s, dependiendo del protocolo de comunicación empleado.

En todo caso, es la red de administración por excelencia, así que aunque no sea la solución de red de altas prestaciones para las comunicaciones, es la red dedicada a las tareas administrativas.

➤ **Myrinet (Myrinet 2000 y Myri-10G)**

Su latencia es de 1,3/10 μ s, y su ancho de Banda de 2/10Gbps, respectivamente para Myrinet 2000 y Myri-10G.

Es la red de baja latencia más utilizada en la actualidad, tanto en clúster como en M.P.P.s estando presente en más de la mitad de los sistemas del top500. Tiene dos bibliotecas de comunicación a bajo nivel (GM y MX). Sobre estas bibliotecas están implementadas MPICH-GM, MPICH-MX, Sockets-GM y Sockets MX, para aprovechar las excelentes características de Myrinet. También existen emulaciones IP sobre TCP/IP, IPoGM e IPoMX.

➤ **InfiniBand**

Es una red surgida de un estándar desarrollado específicamente para realizar la comunicación en clúster. Una de sus mayores ventajas es que mediante la agregación de canales (x1, x4 y x12), permite obtener anchos de banda muy elevados. La conexión básica es de 2Gbps efectivos y con 'quad connection' x12 alcanza los 96Gbps. No obstante, los startups no son muy altos, se sitúan en torno a los 10 μ s.

Define una conexión entre un nodo de computación y un nodo de I/O. La conexión va desde un Host Channel Adapter (H.C.A.) hasta un Target Channel Adapter (T.C.A.). Se está usando principalmente para acceder a arreglos de discos S.A.S.

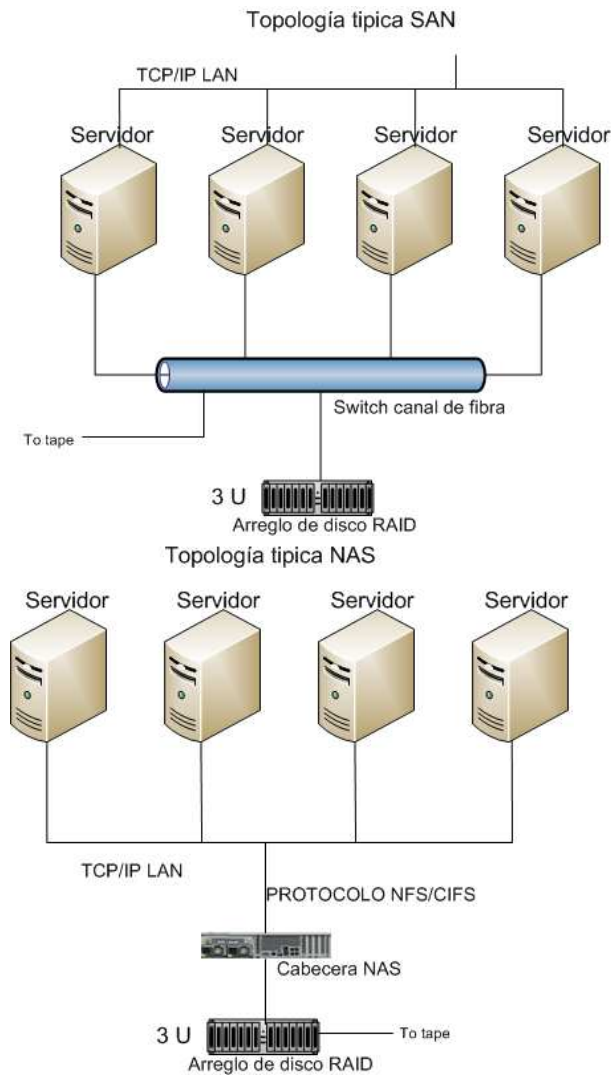
➤ **S.C.I. (Scalable Coherent Interface) I.E.E.E. estándar 1596-1992**

Su latencia teórica es de 1.43 μ s y su ancho de banda de 5333 Mbps bidireccional. Al poder configurarse con topologías de anillo (1D), toro (2D) e hipercubo (3D) sin necesidad de switch, se tiene una red adecuada para clústeres de pequeño y mediano tamaño.

Al ser una red de extremadamente baja latencia, presenta ventajas frente a Myrinet en clústeres de pequeño tamaño al tener una topología punto a punto, y no ser necesaria la adquisición de un conmutador. El software sobre S.C.I. está menos desarrollado que sobre Myrinet, pero los rendimientos obtenidos son superiores, destacando S.C.I. Sockets (que obtiene startups de 3 microsegundos) y ScaMPI, una biblioteca M.P.I. de elevadas prestaciones.

1.3.4. Almacenamiento

Gráfico 4. Típica topología S.A.N y N.A.S.



Elaborado por: Autores

El almacenamiento puede consistir en una N.A.S, una S.A.N, o almacenamiento interno en el servidor. El protocolo más comúnmente utilizado es N.F.S (Network File System), sistema de ficheros compartido entre servidor y los nodos. Sin embargo existen sistemas de ficheros específicos para clúster como Lustre (C.F.S) y P.V.F.S.2.

Tecnologías en el soporte del almacenamiento en discos duros:

- **I.D.E. o A.T.A.:** Velocidades de 33, 66, 100, 133 y 166 MB/s

- **S.A.T.A.:** Velocidades de 150, 300 y 600 MB/s

- **S.C.S.I.:** Velocidades de 160, 320, 640 MB/s. Proporciona altos rendimientos.

- **S.A.S.:** Aún a S.A.T.A.-II y S.C.S.I. Velocidades de 300 y 600 MB/s

- **D.L.T.:** Las unidades de cinta son utilizadas para copias de seguridad por su bajo coste.

- **N.A.S.:** (Network Attached Storage), es un dispositivo específico dedicado al almacenamiento a través de red (normalmente T.C.P./I.P.) que hace uso de un sistema operativo optimizado para dar acceso a través de protocolos; C.I.F.S., N.F.S., F.T.P. o T.F.T.P.

- **D.A.S.:** (Direct Attached Storage) consiste en conectar unidades externas de almacenamiento S.C.S.I. o a una S.A.N. (Storage Area Network) a través de un canal de fibra. Estas conexiones son dedicadas.

Mientras N.A.S. permite compartir el almacenamiento, utilizar la red, y tiene una gestión más sencilla, D.A.S. proporciona mayor rendimiento y mayor fiabilidad al no compartir el recurso.

1.3.5. Middleware

El middleware es el software que generalmente actúa entre el sistema operativo y las aplicaciones y que brinda al usuario la experiencia de estar utilizando una única súper máquina. Este software provee a un clúster lo siguiente:

- Una única interfaz de acceso al sistema, denominada S.S.I (Single System Image), la cual genera la sensación al usuario de que utiliza un único ordenador muy potente.

- Optimiza el sistema y provee herramientas de mantenimiento, como: migración de procesos, balanceo de carga, tolerancia de fallos, checkpoint-restart (congelar uno o varios procesos, mudarlos de servidor y continuar su funcionamiento en el nuevo host).

- Este sistema también se encarga de la escalabilidad del clúster, debe poder detectar nuevos servidores conectados al clúster para proceder a su utilización.

Existen diversos tipos de middleware, como por ejemplo: MOSIX, OpenMOSIX, Cóndor, OpenSSI, etc.

El middleware recibe los trabajos entrantes al clúster y los redistribuye de manera que el proceso se ejecute más rápido y el sistema no sufra sobrecargas en un servidor. Esto se realiza mediante políticas definidas en el sistema (automáticamente o por un administrador) que le indican dónde y cómo debe distribuir los procesos, por un sistema de monitorización, el cual controla la carga de cada CPU y la cantidad de procesos en él.

El middleware también debe poder migrar procesos entre servidores con distintas finalidades:

➤ Balancear la carga: si un servidor está muy cargado de procesos y otro está ocioso, pueden transferirse procesos a este último para liberar de carga al primero y optimizar el funcionamiento.

➤ Mantenimiento de servidores: si hay procesos corriendo en un servidor que necesita mantenimiento o una actualización, es posible migrar los procesos a otro servidor y proceder a desconectar del clúster al primero.

➤ Priorización de trabajos: en caso de tener varios procesos corriendo en el clúster, pero uno de ellos de mayor importancia que los demás, puede migrarse este proceso a los servidores que posean más o mejores recursos para acelerar su procesamiento.

1.3.6. Ambientes de Programación Paralela

Los ambientes de programación paralela permiten implementar algoritmos que hagan uso de recursos compartidos: CPU (Central Processing Unit), memoria, datos y servicios.

1.3.7. Diseño de aplicaciones Requisitos

No todas las aplicaciones se pueden ejecutar en un entorno de clúster de alta disponibilidad, y las decisiones de diseño necesarias deben hacerse en la fase de diseño de software, con el fin de ejecutar en un entorno de clúster de alta disponibilidad y una aplicación que debe satisfacer al menos los siguientes requisitos técnicos:

➤ Tiene que haber una forma relativamente fácil de iniciar, detener y comprobar el estado de la solicitud. En términos prácticos, esto significa que la aplicación debe tener una interfaz de línea de comandos o secuencias de comandos para controlar la aplicación, incluyendo soporte para múltiples instancias de la solicitud.

➤ La solicitud debe ser capaz de utilizar el almacenamiento compartido (N.A.S. / S.A.N.).

➤ Lo más importante, la aplicación debe almacenar la mayor cantidad de su estado en almacenamiento compartido y no volátil de ser posible. De igual manera lo relevante es la capacidad de reiniciar en otro nodo en el último estado, antes de la falla con el estado guardado del almacenamiento compartido.

➤ La aplicación no debe alterar los datos si se bloquea o se reinicia desde el estado guardado.

Los dos últimos criterios son esenciales para la funcionalidad fiable en un clúster, y son los más difíciles de satisfacer plenamente. Por último, el cumplimiento de licencias deben ser observadas.

1.4. BALANCEO DE CARGA

1.4.1. Definición

Es la técnica usada para compartir el trabajo que se realizará entre varios procesos, ordenadores entre otros recursos. Este se mantiene gracias a un algoritmo que divide de la manera más equitativa posible el trabajo, para así evitar los cuellos de botella.

1.4.2. Clúster de balanceo de carga

Está compuesto por uno o más ordenadores estos llamados nodos, que actúan como frontend del clúster, y que se ocupan de repartir las peticiones de servicio que recibe el clúster, a otros ordenadores del clúster que forman el back-end de éste. Un tipo concreto de clúster cuya función es repartir la carga de proceso entre los nodos en lugar de los servicios es el clúster openMosix. Las características más destacadas de este tipo de clúster son:

- Ampliar su capacidad fácilmente añadiendo más ordenadores al clúster.

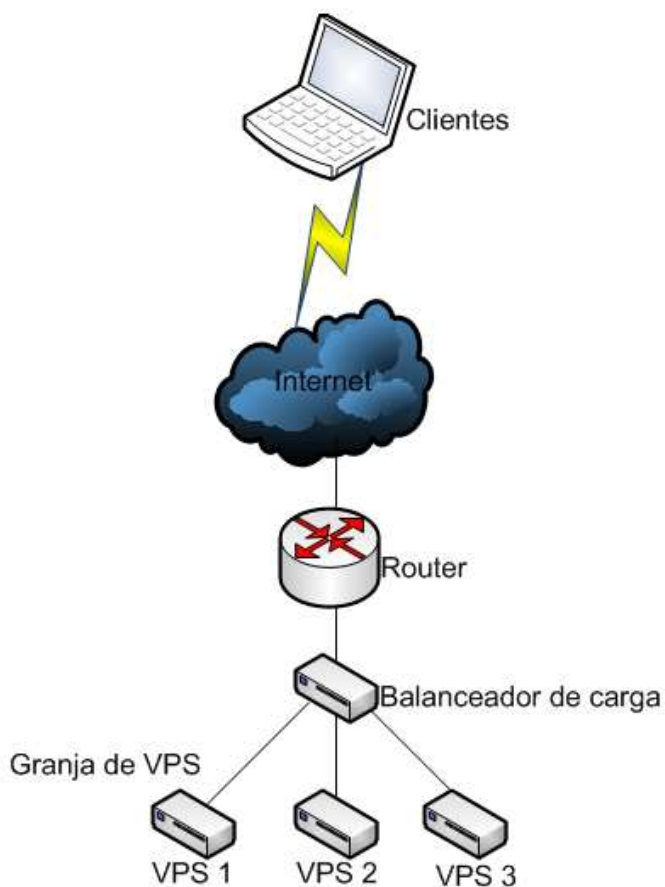
- Robustez. Ante la caída de alguno de los ordenadores del clúster el servicio se puede ver mermado, pero mientras haya ordenadores en funcionamiento, éstos seguirán dando servicio.

1.4.3. Solución de Balanceo de Carga

La Solución de balanceo de carga permite dividir las tareas que tendría que soportar una única máquina, con el fin de maximizar las capacidades de proceso de datos, así como de ejecución de tareas.

Esta Solución permite que ningún equipo sea parte vital del servicio que se quiere ofrecer. De esta forma se evita sufrir una interrupción del servicio debido a una parada de una de las máquinas.

Gráfico 5. Balanceo de carga.



Fuente: Solución de balanceo de carga

Elaborado por: Gravitynet E-Solutions S.L.U.

Además de esto, se debe destacar dos características importantes:

➤ Evita la saturación de una máquina. De esta forma, se puede evitar que los distintos procesos que deseen acceder a las máquinas afecten al normal funcionamiento del aplicativo.

➤ Gestiona de los recursos de manera inteligente. Permite gestionar y optimizar todos los recursos disponibles dando como resultado un acceso más rápido y estable a los aplicativos

CAPITULO 2 : SOLUCIÓN DE SUN MICROSYSTEM

2.1. LOS INICIOS DE ORACLE



Hace poco más de tres décadas Larry Ellison encontró un documento que describía un prototipo de un sistema administrador de bases de datos relacional y lo mostró a sus colegas Bob Miner y Ed Oates.

En ese momento, Este sistema pareció una oportunidad de negocios espectacular y fundaron la compañía Relational Software Inc., que luego llegaría a llamarse Oracle, pero nunca sabrían la importancia que tendría la base de datos que desarrollaron en aquel momento. En 1977 los sistemas computacionales de las grandes empresas eran enormes, arcaicos e ineficientes y con esta tecnología ellos lograron cambiar estos sistemas hasta lo que en la actualidad se conoce como la tecnología de bases de datos que utilizan todas las empresas, desde las grandes multinacionales hasta negocios.

La compañía Oracle se enfocó en los que sus clientes demandaban en un momento en que la era de la informática apenas comenzaba a iniciar, ofreciendo innovación y seguridad al momento de administrar la información y la cual se volvió el activo máspreciado de una compañía.

La empresa volvió a cambiar su nombre a Oracle Systems y se convirtió en una de las empresas norteamericanas con un crecimiento más alto, tras desarrollar estrategias básicas como la de simplificar el manejo de datos, crear soluciones para las plataformas de computación emergentes y aumentar la interoperabilidad en los sistemas para facilitar la sincronización o migración de datos. Al final de los años 80 la compañía había evolucionado teniendo 35 empleados hasta alcanzar el grado de una empresa multinacional, con más de US\$500 millones de ingresos.

Tras una década de mucho crecimiento, Oracle estaba en posición de invertir en innovaciones. Con cada nueva versión de sus productos, la empresa expandió los horizontes en todos los aspectos, como el poder de cálculo gracias a la introducción del lenguaje de programación orientado a base de datos PL/SQL, manejabilidad y rendimiento.

En los primeros años del nuevo milenio, Oracle ha continuado entregando innovación y resultados con productos como: Oracle E-Business Suite, Oracle Grid Computing, Oracle Fusion, Oracle Application Server, Oracle Virtual Manager y soporte para sistemas GNU/Linux a nivel empresarial, marcando así la tendencia de la compañía hacia futuro.

2.2. HISTORIA DE SUN MICROSYSTEMS



Sun Microsystems fue fundada el 24 de febrero de 1982 por Andreas Bechtolsheim, Vinod Khosla, Scott McNealy y Bill Joy como la concreción de un proyecto de Bechtolsheim que se había originado en 1981. Su intención era crear una estación de trabajo en red que fuera más funcional y más barata posible que las existentes en aquel momento. Este proyecto es, curiosamente, el que le da el nombre a la compañía: Stanford University Network (SUN).

La importante colaboración entre los cuatro socios provocó que en el transcurso de cuatro años la empresa, que hasta entonces se dedicaba a producir hardware más poderoso a partir de partes ya disponibles en el mercado, alcanzara un alto renombre mundial entre las compañías de desarrollo tecnológico.

Sin embargo, llegó el momento en que la visión de Sun Microsystems debió adaptarse a las necesidades de sus clientes quienes demandaban software más eficiente y dejó por un lado la fabricación de hardware para enfocarse en hacer accesible la tecnología de procesamiento y funciones de redes a las empresas medianas, cuando hasta este momento había sido algo exclusivo de las grandes corporaciones.

Esto lo logró basándose en estándares existentes, como UNIX. La estrategia implementada provocó que las ganancias empezaran a llegar en solo un par de meses, gracias a los bajos costos y elevadas ventas de sus productos.

Los años 80 fueron muy importantes para Sun, puesto que en 1984 se introdujo el sistema N.F.S (Network File System o sistema de archivos de red), permitiendo que los usuarios conectados a una red puedan trabajar con archivos ubicados en otros ordenadores y esto fuera casi de manera transparente para el usuario final. Además, empezó a cotizar en bolsa en 1986 e introdujo al primero de la familia de procesadores SPARC el año siguiente.

En los años 90 a Sun le fue mucho mejor, puesto que el desarrollo del sistema operativo Solaris 2.0, le permitió a Sun convertirse en un fabricante tanto de software como de hardware y crecer muy rápidamente, siguiendo de cerca a grandes compañías como IBM y Apple. Eventualmente hicieron su transición a software por la ardua competencia en el campo del hardware y desarrollaron el lenguaje de programación Java, el cual estaba orientado a objetos, con un modelo más simple que C y C++.

Sin embargo, el impacto que tuvo Java no se tradujo en dinero y la compañía seguía dependiendo de la línea de fabricación de servidores que corrían bajo el sistema operativo Solaris. Además el surgimiento de GNU/Linux provocó que algunos clientes se alejaran de Sun, para recurrir a opciones de otros fabricantes con costos más baratos. A partir de entonces Sun Microsystems, no ha tenido un crecimiento significativo como sus dos décadas anteriores y se ha estancado al no saber qué postura tomar frente al auge del código abierto u open source. Ante dicha situación, la compañía ha respondido con servicios como:

OpenOffice.org, OpenSolaris y GlassFish. Después de todo, el software distribuido libre y gratuitamente es el futuro o mejor dicho, el presente tanto para empresas como Sun Microsystems como la informática en general.

2.3. LA COMPRA DE SUN POR PARTE DE ORACLE

En el mes de marzo del año 2009 uno de los primeros interesados en comprar Sun fue IBM, que estuvo dispuesto a pagar 6.500 millones de dólares por Sun Microsystem pero luego de que un centenar de abogados estuvieron revisando por encargo de IBM los posibles problemas que una transacción como ésta podría suscitar con las autoridades reguladoras de Estados Unidos, así como otras dificultades de integración.

Después de aquel proceso, IBM rebajó su oferta de compra la cual no cumplía con las expectativas del Consejo de Administración de Sun Microsystem. Los directivos dejaron en claro que no seguirían negociando en exclusividad con IBM y desistieron finalmente el 6 de abril del 2009. Luego de esto Oracle finalmente adquirió a Sun Microsystems el 27 de Enero del 2010, en un monto total de 7,400 millones de dólares.

Larry Ellison consejero delegado de Oracle dijo: “La compra de Sun transforma el sector de las tecnologías de la información. Oracle será la única compañía que podrá diseñar un sistema integrado donde todas las piezas funcionan juntas, de forma que los clientes no tendrán que unirlos ellos mismos”.

El presidente de Sun, Scott McNealy declaró: “Oracle y Sun han sido empresas pioneras del sector y socio durante más de veinte años, esta unión entre ambas firmas es resultado de una evolución natural de su relación y supondrá un “hito” en el sector”.

Por su parte, el presidente de Oracle, Safra Catz, adelantó que el acuerdo podría aportar a los beneficios de su empresa al menos 15 centavos por acción. Un año después de que se haya cerrado la operación, lo que convertiría esta compra en una transacción más rentable que las de BEA Systems, PeopleSoft y Siebel juntas.

Con esta adquisición, Oracle se hace de los dos productos más importantes de Sun:

La tecnología Java y los servidores Solaris. A lo largo de los años, Oracle ha participado en el negocio de las bases de datos paralelamente con el de los servidores de Sun. La unión Oracle-Sun se convertiría en un gigante capaz de ofrecer software y hardware a sus clientes y soluciones certificadas en estas plataformas.

2.4. ORACLE SUN CLÚSTER

2.4.1. Introducción

Hoy en día las empresas son conducidas por las exigencias de una economía global 24x7, lo que significa que la optimización de los niveles de la alta disponibilidad de todos los tipos de servicios de aplicaciones se está convirtiendo en un foco primario para muchos administradores de TI. Con la finalidad de satisfacer esta creciente demanda de disponibilidad continua, las organizaciones necesitan soluciones de menor costo con el que predecir capacidades y características de alta disponibilidad de sistemas tolerantes a fallos costosos.

Es cierto que en la industria alrededor del 20 por ciento de los fallos del sistema son en realidad debido a fallas del hardware, software, o una combinación de ambos. El otro 80 por ciento es el resultado de fallas de proceso o errores humanos. Ante la demanda de servicios de aplicaciones disponibles de manera continua que minimicen el impacto de los fallos causados por productos, personas y procesos, los clientes recurren a las soluciones de clustering. Mediante la vinculación de múltiples servidores, dispositivos de almacenamiento y software, soluciones de clúster son capaces de tolerar el fallo de cualquier componente del entorno informático. Con el fin de ayudar a asegurar los más altos niveles de disponibilidad. El software del clúster debe ser diseñado, probado y desplegado con los más altos niveles de calidad y confiabilidad.

Calidad, tal como se define en el ámbito de software y disponibilidad de las aplicaciones, es la capacidad de un componente para consistentemente estar de acuerdo con su especificación funcional sin desviación o error inesperado.

2.4.2. Sun Clúster 3

La solución Solaris clúster también llamado Sun Cluster o SunCluster es un producto de software para la alta disponibilidad, diseñado para ejecutarse en sistemas operativos Solaris creado por SUN Microsystems ahora perteneciente a Oracle Corporation el cual esta soportado por una arquitectura Sparc o X86, entiéndase esto como procesadores Intel y AMD.

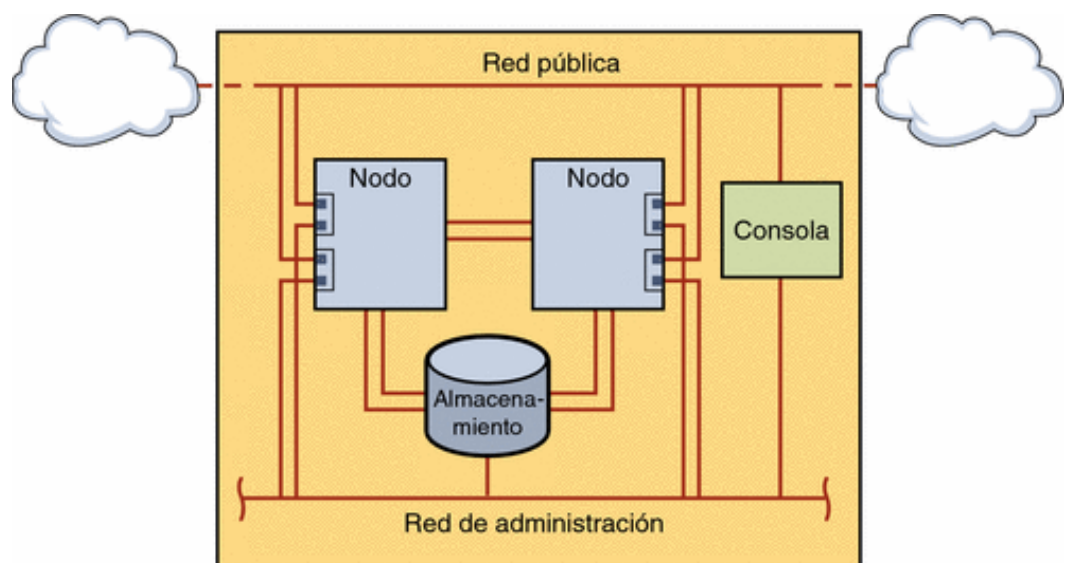
Esta solución ayuda a mejorar la previsibilidad y la capacidad de recuperación de las aplicaciones críticas de negocio. Está diseñado para proporcionar servicios altamente disponibles a nivel plataforma de gestión que permite a las empresas desplegar aplicaciones y proporcionar una alta disponibilidad, rendimiento y la escalabilidad de estos servicios. Sun Clúster es el único que se integra con el software Sun Java Enterprise System, que ofrece aplicaciones Web, redes de identidad, disponibilidad, servicios de seguridad, Portal de la Comunicación y Colaboración.

Este software es usado para proveer servicios de software de alta disponibilidad a diferentes sistemas como: Base de datos, Servicios de compartición de archivos, sitios de comercio electrónico y otros sistemas. Sun Clúster opera en equipos redundantes o nodos, donde una o más computadoras continua proveyendo el servicio si otra falla. Estos nodos pueden estar físicamente en el mismo centro de cómputo o en diferente continente.

2.4.3. Componentes de Sun Clúster.

En esta sección se detallará cada uno de los elementos que conforman un Sun clúster, así el lector tendrá una comprensión más clara y profunda.

Gráfico 6. Topología del clúster y conexiones de los cables.



Fuente: Guía de inicio rápido de Sun Clúster para el SO Solaris
Elaborado por: Oracle

2.4.3.1. Nombre global del clúster

Se debe especificar un nombre para el clúster durante configuración. El nombre del clúster debe ser único en todo el dominio.

2.4.3.2. Nombres de host privado

El nombre de equipo privado es el nombre que se utiliza para la comunicación entre las interfaces de red privada. Estos nombres privados se crean automáticamente durante la configuración de un clúster global o un clúster de la zona.

Estos nombres de equipo privados utilizan la nomenclatura siguiente:

nodeid-priv, donde **nodeid** es el número del identificador de nodo interno.

Durante la configuración de Sun Clúster, el número de ID de nodo se asigna automáticamente a cada voto de nodo cuando el nodo se convierte en un miembro del clúster. Un voto de un nodo del clúster global y un nodo perteneciente a un clúster de zona ambos tienen el mismo nombre de equipo privado, pero cada nombre de host se resuelve en una red de direcciones IP privadas diferentes.

Después de que un clúster global es configurado, puede renombrar el equipo privado mediante el comando **clsetup**. En la actualidad, no se puede renombrar un equipo privado de un nodo que este en un clúster de zona.

Para el sistema operativo Solaris 10, la creación de un nombre de equipo privado para una zona no global es opcional.

2.4.3.3. Red publica

Se debe Considerar los siguientes puntos cuando se configuren la red pública del clúster:

➤ **Separación de la red pública y privada.-** Las redes públicas y la red privada (interconexión del clúster) deben utilizar adaptadores por separado, o bien debe configurar etiquetado VLAN. Un adaptador con capacidad de etiquetado VLAN y un conmutador VLAN para utilizar el mismo adaptador de red para la interconexión privada y la red pública.

➤ **Mínimo.-** Todos los nodos del clúster deben estar conectado al menos a una red pública. Conexiones de red pública pueden utilizar subredes diferentes para los nodos diferentes.

➤ **Máximo.-** Usted puede tener tantas conexiones de red publicas adicionales como la configuración de hardware lo permita.

➤ **Servicios escalables.-** Todos los nodos que ejecutan un servicio escalable debe utilizar la misma subred o conjunto de subredes o utilizar diferentes subredes que se encaminen entre sí.

➤ **IPv4.-** El software Sun Clúster admite las direcciones IPv4 de la red pública.

➤ **IPv6.-** El software Sun Clúster admite direcciones IPv6 en la red pública, pero no es compatible si la interconexión privada utiliza adaptadores S.C.I.

➤ **IPMP grupos.-** Cada adaptador de red pública que se utiliza para el servicio de tráfico de datos deben pertenecer a una red IP multi-rruta (IPMP) del grupo. Si un adaptador de red público no se utiliza para el servicio de tráfico de datos, no tiene que configurar en un grupo IPMP.

El comando **scinstall** configura automáticamente un adaptador de grupo IPMP múltiples para cada conjunto de adaptadores de red pública del clúster que utiliza la misma subred y hace caso omiso de los adaptadores que ya están configurados en un grupo IPMP

➤ **Soporte local de dirección M.A.C.-** Todos los adaptadores de red públicos deben usar las tarjetas de interfaz de red (N.I.C) que soporten localmente la asignación de dirección M.A.C esto es un requisito de IPMP.

➤ **Establecer la variable local-mac-address.-** Debe utilizar el valor predeterminado igual a verdadero para los adaptadores de Ethernet. El software Sun Clúster no es compatible con un **local-mac-address** cuyo valor sea igual a falso. Este requisito es un cambio de Sun Clúster 3.0, que exigía un **local-mac-address=falso**.

2.4.3.4. Red Privada

El software Sun Clúster utiliza la red privada para la comunicación interna entre los nodos y entre zonas no globales, que son administrados por el software Sun Clúster.

Una configuración de Sun Clúster requiere al menos dos conexiones a la interconexión del clúster de la red privada. Al configurar el software Sun Clúster en el primer nodo del clúster, se especifica la dirección de red privada y máscara de red en base a una de las siguientes maneras:

➤ Aceptar la dirección de red privada predeterminada (172.16.0.0) y máscara de red por defecto.

▪ En el sistema operativo Solaris 10, la máscara de red por defecto es 255.255.240.0. Este rango de direcciones IP admite un máximo combinado de 64 nodos de votación y en zonas no globales, un máximo de 12 clústeres de zona, y un máximo de 10 redes privadas.

▪ En el sistema operativo Solaris 9, la máscara de red por defecto es 255.255.248.0. Este rango de direcciones IP admite un máximo combinado de 64 nodos y un máximo de 10 redes privadas.

- Se debe especificar una dirección de red privada diferente y aceptar la máscara de red por defecto.

- Aceptar la dirección de red privada predeterminada y especificar una máscara de red diferente.

- Especificar tanto una dirección de red privada diferente y una máscara de red diferente.

Si se desea especificar una máscara de red diferente, el comando **scinstall** solicitará que ingrese el número de nodos y el número de redes privadas que desea que el rango de direcciones IP soporte. En el sistema operativo Solaris 10, el comando también solicitará el número de clústeres de zona que se desea soportar.

Dicho comando calculará la máscara de red para el rango de direcciones IP mínimo que soportaran el número de nodos, los clústeres de zona, y redes privadas que ha especificado. La máscara de red calculada podría apoyar más que el número proporcionado de nodos, incluyendo las zonas no globales, las agrupaciones zona, y redes privadas.

El comando le preguntará qué máscara de red quiere elegir. Puede especificar cualquiera de las máscaras de red o una diferente. La máscara de red que se especifique como mínimo debe ser compatible con el número de nodos y redes privadas que se ha especificado al comando.

Si se especifica una dirección de red privada que no sea el predeterminado, la dirección debe cumplir los siguientes requisitos:

➤ Direcciones de red y de máscara.- La dirección de red privada no puede ser menor que la máscara de red. Por ejemplo, puede utilizar una dirección de red privada de 172.16.10.0 con una máscara de red 255.255.255.0. Pero no se puede utilizar una dirección de red privada de 172.16.10.0 con una máscara de red 255.255.0.0.

➤ **Direcciones aceptables.**- La dirección de red debe estar dentro del bloque de direcciones que la normativa **RFC 1918** reserva para utilizarla en redes privadas.

➤ **IPv6.**- El software Sun Clúster no es compatible con direcciones IPv6 para la interconexión privada.

2.4.3.5. Interconexión del clúster

Las interconexiones de un clúster proporcionan las vías de hardware para que los nodos del clúster se puedan comunicar por medio de redes privadas. Cada interconexión consiste en un cable que se conecta de una de las siguientes maneras:

- Entre dos adaptadores de transporte.
- Entre un adaptador de transporte y un conmutador de transporte

Durante la configuración de Sun Clúster, se especifica la información de configuración para una o dos interconexiones del clúster de la siguiente manera:

- Si el número de puertos de adaptador disponible es limitado, puede utilizar VLAN etiquetadas para compartir el mismo adaptador con la red pública y privada

➤ Puede configurar desde uno hasta seis interconexiones en un clúster. Mientras que una sola interconexión de clúster reduce el número de puertos del adaptador que se utilizan para la interconexión privada, y no proporciona redundancia y menos disponibilidad.

2.4.3.6. Conmutador de transporte

Si utiliza conmutadores de transporte, tales como un conmutador de red, usted debe especificar un nombre para cada conmutador de transporte de la interconexión. Se puede utilizar el nombre predeterminado switch N, donde N es un número que se asigna automáticamente durante la configuración, o puede establecer otro nombre.

También debe especificar el nombre de puerto del conmutador o aceptar el nombre predeterminado. El nombre del puerto por defecto es el mismo que el número interno de identificación del nodo. Sin embargo, no se puede utilizar el nombre del puerto por defecto para ciertos tipos de adaptadores, como los SCI-PCI.

2.4.3.7. Adaptadores de transporte

Los adaptadores de transporte, son puertos de las interfaces de red, especifican los nombres de adaptador de transporte y el tipo de transporte. Si la configuración es para un clúster de dos nodos, se puede elegir la interconexión punto a punto (adaptador al adaptador) o utilizar un conmutador de transporte como un switch. Para un clúster de más de dos nodos es obligatorio el uso de un conmutador de transporte.

2.4.3.8. Fencing global

El Fencing es un mecanismo utilizado por el clúster para proteger la integridad de los datos de un disco compartido en situaciones de cerebro dividido. De forma predeterminada, en el comando `scinstall` habilita el fencing global.

El comando **scinstall** le pregunta si quiere desactivar el fencing global. Para la mayoría de situaciones, responda No para mantenerlo habilitado. Sin embargo, puede deshabilitarlo en apoyo de las siguientes situaciones:

- El almacenamiento compartido no admite reservas SCSI.
- Desea habilitar sistemas que están fuera del clúster para acceder al almacenamiento que se adjunta al clúster.

2.4.3.9. Global devices

Sun Clúster utiliza los Global devices o también llamados dispositivos globales para ofrecer en todo el clúster, el acceso de alta disponibilidad a cualquier dispositivo de un clúster, desde cualquier nodo. En general, si un nodo falla mientras que proporciona acceso a un dispositivo global, el software Sun Clúster cambia a otra ruta de acceso al dispositivo y redirige el acceso a esa ruta.

Esta redirección es fácil con los dispositivos globales debido a que el mismo nombre se utiliza para el dispositivo, independientemente de la ruta. El acceso a un dispositivo remoto se realiza en la misma forma que en un dispositivo local que utiliza el mismo nombre.

Además, la API para acceder a un dispositivo global en un clúster es la misma que la API que se utiliza para acceder a un dispositivo local.

2.4.3.10. Dispositivo de Quórum.

En Sun Clúster, el mecanismo que determina la participación del nodo se conoce como quórum. Cada nodo del clúster es tomado en cuenta para la votación del quórum. Para que el clúster sea operacional, los votos requeridos son $N/2 + 1$, donde **N** es número de nodos en el clúster.

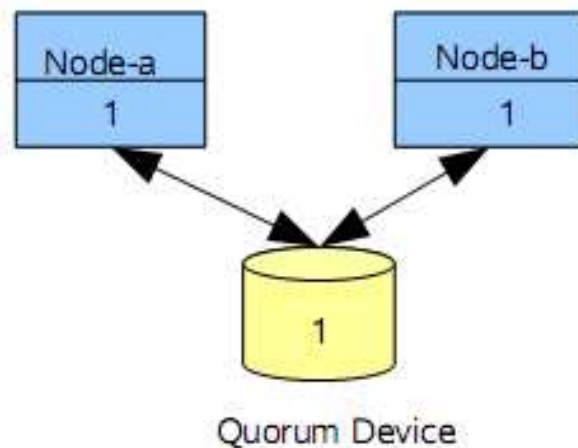
En un clúster de dos nodos, para que el clúster se pueda establecer, se requiere al menos dos votos. En el caso de que cualquiera de estos dos nodos del clúster este caído, todo el clúster se baja. Con el fin de evitar que esto suceda se necesita de un tercer elemento en el clúster como se

muestra en el grafico 7, el cual es un dispositivo de almacenamiento compartido, Este es identificado y configurado como un dispositivo de quórum.

El dispositivo debe ser un disco compartido que puede ser accedido por ambos nodos del clúster.

Sun Clúster hace uso de dispositivos de quórum en su configuración para mantener la integridad de los datos y los recursos, mediante la prevención de la amnesia y los problemas de cerebro dividido. Estos problemas serán explicados de manera amplia a continuación:

Gráfico 7. Alta Disponibilidad entre nodos remotos.



Elaborado por: Autores

Debido a que los nodos del clúster comparten datos y recursos, es importante que nunca un clúster se divida en dos particiones independientes que estén activas al mismo tiempo.

Hay dos tipos de problemas que surgen con particiones de clúster: Cerebro dividido y amnesia.

Cerebro dividido se produce cuando la interconexión del clúster entre los nodos se pierde y el clúster se divide en subgrupos, cada uno de ellos cree que es la única partición. Esto ocurre debido a problemas de comunicación entre los nodos del clúster.

Amnesia se produce cuando se reinicia el clúster después de un cierre con los datos del clúster más antiguo que en el momento de la parada. Esto puede suceder si las versiones múltiples de los datos se almacenan en el arreglo de disco y una nueva encarnación del grupo se inicia cuando la última versión no está disponible.

Cerebro dividido y amnesia puede evitarse dando a cada nodo un voto y ordenando una mayoría de votos para un clúster operativo. Una partición con la mayoría de votos tiene quórum y se le permite operar. Este mecanismo de voto de la mayoría funciona bien siempre y cuando haya más de dos nodos en el clúster. El voto externo es necesario para cualquier partición para obtener quórum. Este voto externo lo proporciona un dispositivo de quórum. Este dispositivo puede ser cualquier disco que se comparte entre los dos nodos.

Los requisitos para un dispositivo de quórum son:

- Un clúster de dos nodos debe tener un dispositivo de quórum. Para otras topologías, es opcional.
- Configurar impar el número de dispositivos de quórum. Para asegurar los dispositivos de quórum para tener completamente independientes las vías de fallos.
- No hay un requisito de tamaño para el dispositivo de quórum. También es posible asignar el dispositivo que podría contener datos de usuario como el dispositivo de quórum.

2.4.3.11. Reservación de SCSI

Sun Clúster utiliza el mecanismo de reserva SCSI en relación con el principio de quórum para la protección e integridad de los datos. Cuando se forma el clúster, un nodo se hace responsable del dispositivo de quórum y se etiqueta como el propietario.

Los otros nodos se etiquetan como capaces de convertirse en propietarios. Las claves se escriben en el dispositivo de quórum.

Si un nodo abandona el clúster y el otro nodo borra las claves que pertenecen al nodo que dejó el clúster. El nodo no podrá hacer operaciones de entrada/salida y básicamente estará aislado. Cuando ese nodo vuelve a unirse al grupo, las claves se vuelven a registrar y es habilitado el acceso. Sun Clúster emulado P.G.R (reserva de grupo persistente) para dispositivos SCSI-2. Para un clúster con más de dos nodos, el dispositivo de almacenamiento compartido debe ser SCSI-3.

2.4.3.11.1. Internet SCSI (iSCSI)

iSCSI(Internet Small Computer System Interface) es un protocolo que permite el transporte de bloques de datos sobre la red IP. Este protocolo no requiere una infraestructura de red especial que normalmente se requiere para el acceso de dispositivos de bloque como los basados en canal de fibra.

2.4.3.11.2. Votación del Quórum

De forma predeterminada, los nodos del clúster deben adquirir un recuento de votos de un quórum, cuando inicia y se convierten en miembros de la agrupación. Los nodos también pueden tener un conteo de los votos de cero, por ejemplo, cuando el nodo se está instalando, o cuando el administrador ha colocado un nodo en el estado de mantenimiento.

Dispositivos de adquisición de quórum la votación de quórum hacen el recuento basado en el número de conexiones de nodo en el dispositivo y cuando un dispositivo de quórum se establece, adquiere un recuento de votos máxima de **N-1**, donde **N** es el número de conexiones hacia los nodos, usted puede configurar dispositivos de quórum durante la instalación del clúster, o más adelante llamando al comando **ssetup**.

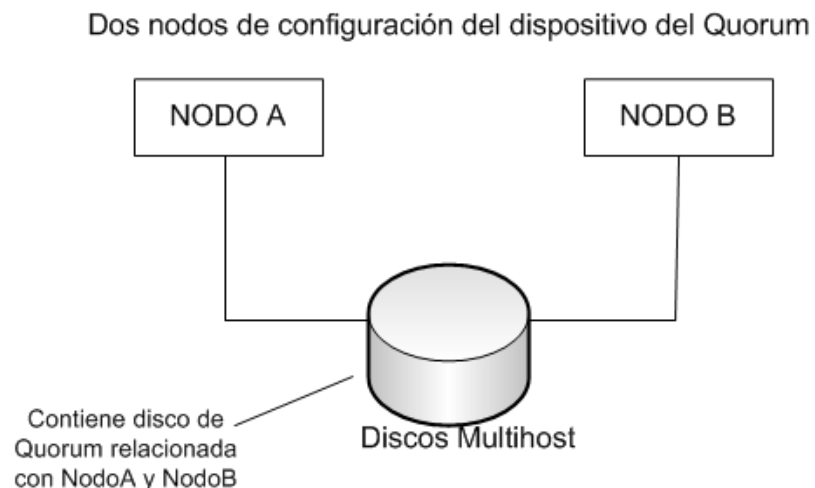
2.4.3.12. Configuraciones de quórum

Las configuraciones del Quórum en función del número de nodos en el clúster:

2.4.3.12.1. Dos Nodos

Dos votos del Quórum se requieren para un nodo de clúster-dos a la forma. Estos dos votos pueden ser los dos nodos del clúster, o de un solo nodo y un dispositivo de quórum. Sin embargo, un dispositivo de quórum debe estar configurado en un nodo de clúster-dos para asegurarse de que un único nodo puede continuar si el otro nodo falla.

Gráfico 8. Ejemplo de configuración de un dispositivo de quórum en dos nodos.



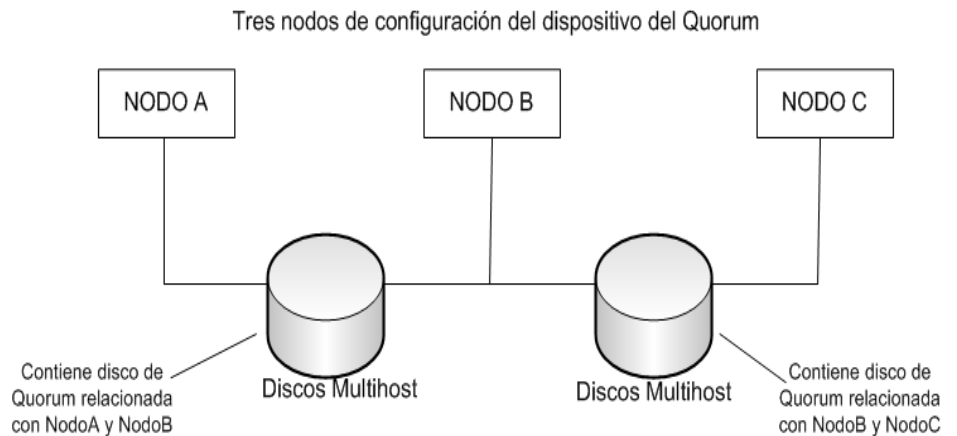
Elaborado por: Autores

2.4.3.12.2. Más de dos Nodos

Usted debe especificar un dispositivo de quórum entre cada par de nodos que las cuotas de acceso a un clúster de almacenamiento en disco. Por ejemplo, imagínese que tiene tres nodos similares en el clúster como se muestra en el gráfico 9. Se puede observar al Nodo A, compartir el acceso al mismo disco con el NodoB, y este mismo nodo (NodoB) compartiendo el acceso a otro disco con el NodoC.

Habría un total de cinco votos de quórum, tres votos pertenecientes a cada uno de los nodos y dos votos de los dispositivos de quórum compartido entre los nodos. Si aplicamos la regla mencionada anteriormente en el capítulo 2.4.3.4 referente a dispositivos de quórum que dice que el mínimo de votos necesarios para el quórum es $N/2-1$, aplicándola en este caso sería $((5\text{Votos}/2) - 1)=3.5$ votos, es decir 3 votos como mínimo en el clúster para formar quórum

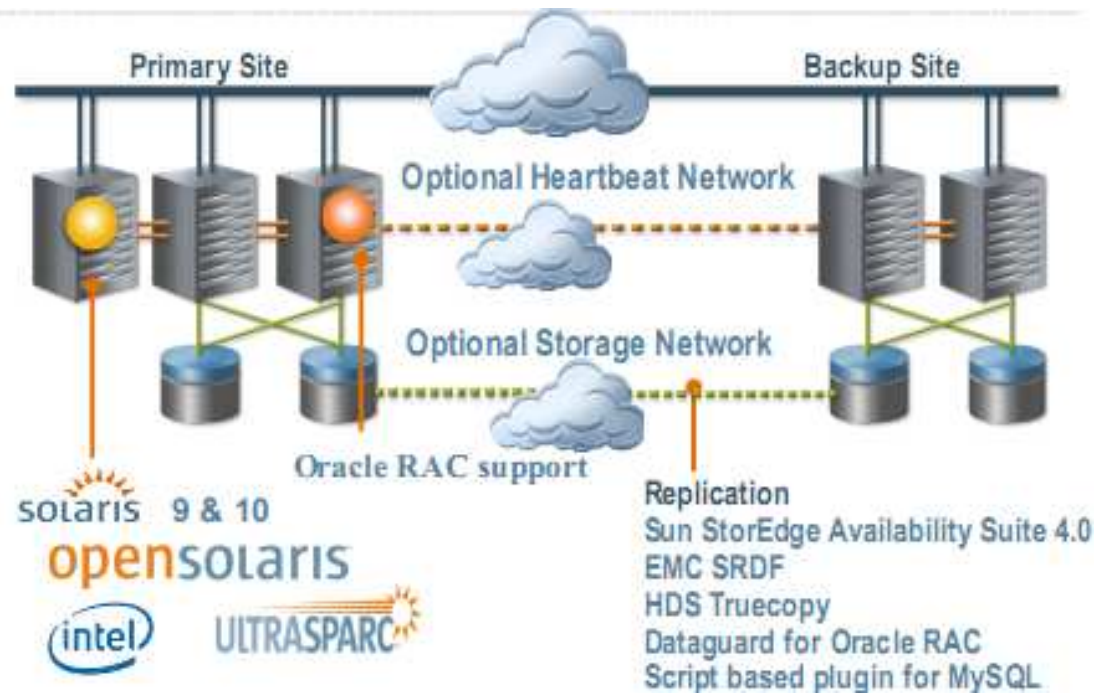
Gráfico 9. Ejemplo de configuración de dispositivos de quórum en tres nodos.



Elaborado por: Autores

2.4.4. Solaris Clúster Edición Geográfica

Gráfico 10. Alta Disponibilidad con Sun Clúster edición geográfica.



Fuente: Open HA Cluster

Elaborado: Solaris Cluster Geographic Edition

SCGE (Solaris Clúster Geographic Edición) fue introducido en Agosto del año 2005, y es un framework de administración que permite que dos instalaciones de Sun Clúster puedan ser administradas como una sola, junto con uno o más productos de replicación de datos, para proporcionar la recuperación de desastres en una instalación de cualquiera de los equipos del clúster. Por este motivo estos datos son actualizados continuamente y replicados a un sitio remoto en tiempo real, estos sitios pueden rápidamente tomar a cargo la provisión de un servicio en el evento que el sitio primario que se encuentre caído como un resultado de un desastre. Este es una clave para minimizar el objetivo de punto de recuperación (RPO) y objetivo de tiempo de recuperación (RTO) para el servicio.

2.4.5. Agentes

Solaris Clúster usa un conjunto de componentes de software llamados agentes o servicios de datos, estos componentes monitorean una o varias aplicaciones para detectar cuando estas no estén funcionando correctamente y toma una acción si un problema es detectado.

Los Agentes para aplicaciones comunes son incluidos en el software de Sun Clúster entre los cuales constan los siguientes servicios de datos:

- Servicio de datos de Sun Cluster para Apache.
- Servicio de datos de Sun Cluster para Apache Tomcat.
- Servicio de datos de Sun Cluster para DHCP.
- Servicio de datos de Sun Cluster para el Servicio de nombres de dominio (DNS).
- Servicio de datos de Sun Cluster para Kerberos.
- Servicio de datos de Sun Cluster para MySQL.
- Servicio de datos de Sun Cluster para N1 Grid Service Provisioning Server.
- Servicio de datos de Sun Cluster para Oracle.
- Servicio de datos de Sun Cluster para Oracle Application Server.
- Sun Cluster HA para PostgreSQL.
- Servicio de datos de Sun Cluster para Samba.
- Servicio de datos de Sun Cluster para Sun Java System Application Server.
- Servicio de datos de Sun Cluster para Sun Java System Message Queue Server.
- Servicio de datos de Sun Cluster para Sun Java System Web Server.

El software de un servicio de datos proporciona implementaciones de los métodos de Sun Management Cluster que realizar las siguientes operaciones sobre la aplicación:

- Inicio de la aplicación.
- Detener la aplicación.
- Control de fallas en la aplicación y la recuperación de estas fallas.

También incluyen un asistente el cual permite al operador del clúster crear agentes para otras aplicaciones.

Para lograr la alta disponibilidad en el servidor de aplicaciones de la U.C.S.G se utilizará el siguiente agente: **Servicio de datos de Sun Cluster para Oracle Application Server**

CAPITULO 3 : METODOLOGÍA

- Análisis de los servicios y procesos a diseñar en la alta disponibilidad.

- Desarrollo de los esquemas de alta disponibilidad, con sus ventajas y desventajas de cada diseño.

- Elaboración de prototipo de alta disponibilidad con el software Sun Cluster 3.2.

- Configuración de servicios en los servidores de alta disponibilidad

- Pruebas de simulación de desastres en servidores con alta disponibilidad.

- Documentación de todas las fases para alta disponibilidad.

CAPITULO 4 : DISEÑO DE ALTA DISPONIBILIDAD

4.1. SITUACIÓN ACTUAL.

Este proyecto de tesis está ideado para resolver la problemática actual del Centro de Cómputo de la U.C.S.G, el cual cuenta con pocos recursos tecnológicos y tiene la misión de que los servicios que ofrecen deben estar disponibles el mayor tiempo posible, ya que todas las facultades desde tesorería hasta empresariales cuentan con estos servicios.

Para que sus procesos internos no se vean afectados y así seguir con el ciclo de vida universitario, esto conlleva desde la matriculación de sus alumnos pasando por la selección de sus materias hasta finalización del semestre con todos los otros procesos administrativos internos que esto involucra, este ciclo se da en cada una de las facultades. Es por todo esto que hoy más que nunca estos servicios deben de estar altamente disponibles.

4.2. EQUIPAMIENTO

Se realizó un estudio de los recursos tecnológicos con los que cuenta el Centro de Cómputo de la UCSG. Cabe aclarar que estos recursos van directamente relacionados con los servicios que proveen.

Estos recursos son:

Dos Servidores SunFire V490 cada uno con 8Gb de RAM y un procesador dual 2.1 GHz UltraSPARC IV. El primer equipo es usado como Servidor de Aplicaciones y otro es el Servidor de Base de Datos. También cuenta con un Arreglo de discos Storeedge 3200 SCSI el mismo que solo es usado como repositorio de la base de datos.

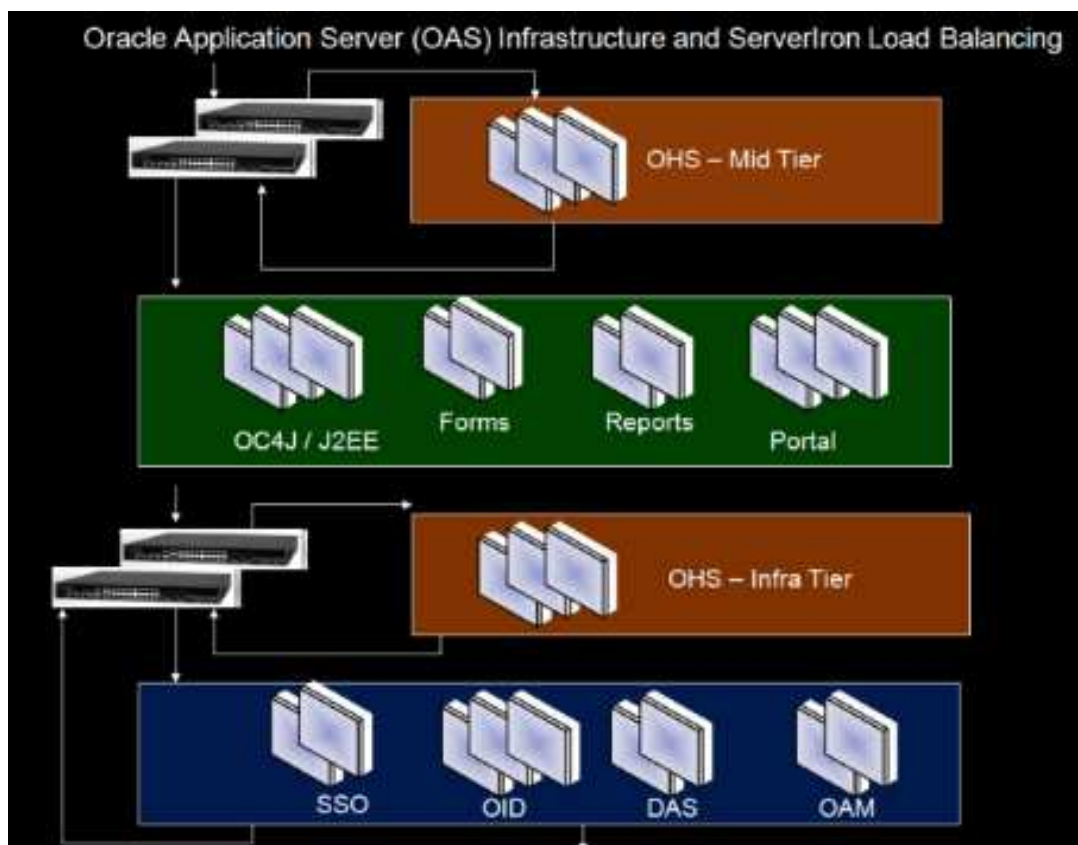


4.3. SERVICIOS

Entre los servicios que brinda el centro de computo de la UCSG, uno de los más críticos es el sistema S.I.U (Sistema Integrado Universitario), ya que contiene los subsistemas académicos, financiero, recursos humanos y administración, que es utilizado por docentes, empleados y directores de cada una de las facultades. Este sistema es de hecho indispensable para el correcto funcionamiento de la universidad.

Este sistema esta soportado en el Servidor de Aplicaciones de Oracle 10gR2 que tiene la siguiente arquitectura:

Gráfico 11. Arquitectura del servidor de aplicaciones de Oracle.



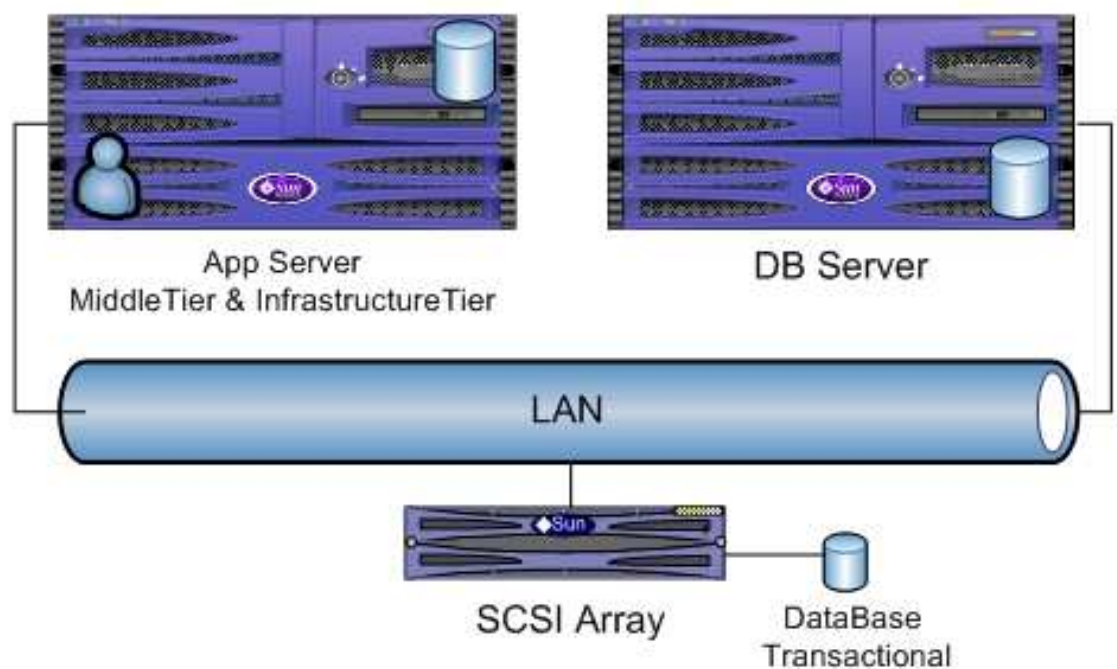
Fuente: Servidores de aplicaciones

Elaborado por: Brocade

4.4. ESTRUCTURA

El siguiente gráfico muestra la infraestructura tecnológica con la que cuenta actualmente el Centro de Cómputo de la UCSG, sobre el que se ejecuta el sistema S.I.U.

Gráfico 12. Infraestructura del centro de cómputo de la UCSG.



Elaborado por: Autores

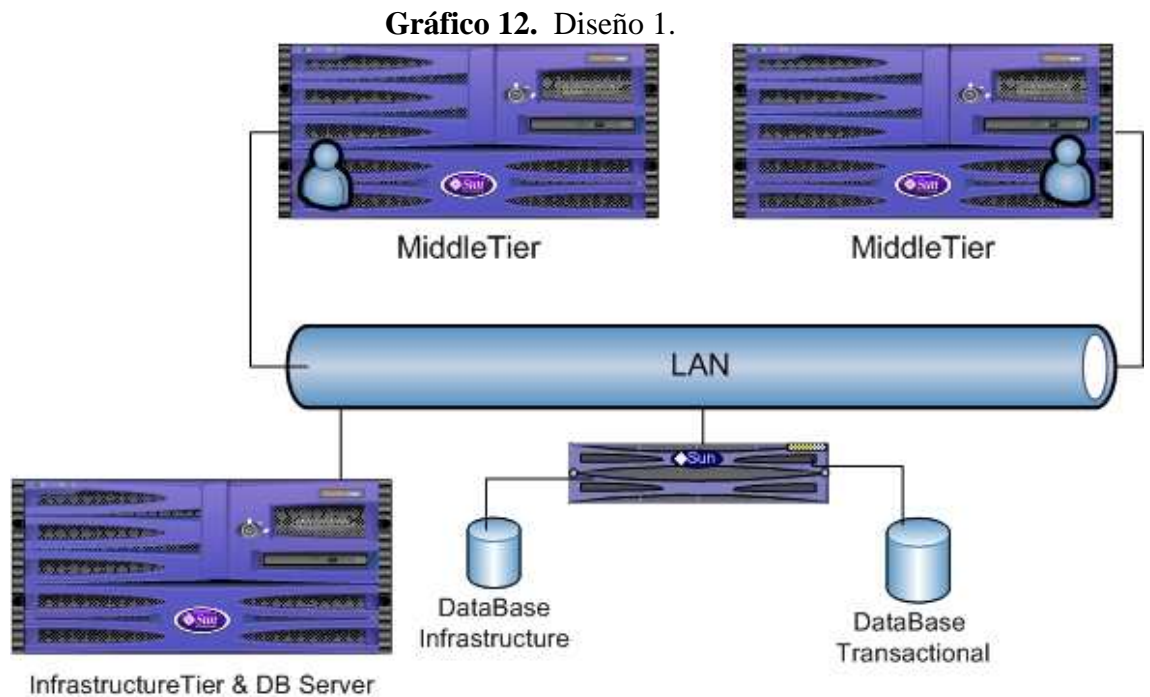
En uno de los servidores SunFire V490 se encuentra instalado el Servidor de aplicaciones de oracle 10gR2 y dentro del mismo las dos capas: Capa media y capa de infraestructura.

El sistema S.I.U fue creado con tecnología Oracle forms y reports, estos están contenidos dentro de Servidor de aplicaciones y son publicados mediante el portal de oracle.

El segundo servidor SunFire V490 es usado solo para base de datos ya que el sistema S.I.U tiene tablas con una alta transaccionalidad, para mantener la integridad de estos datos se tiene un arreglo de discos Storedge 3200.

4.5. DISEÑOS DE ALTA DISPONIBILIDAD PROPUESTOS

4.5.1. Diseño 1



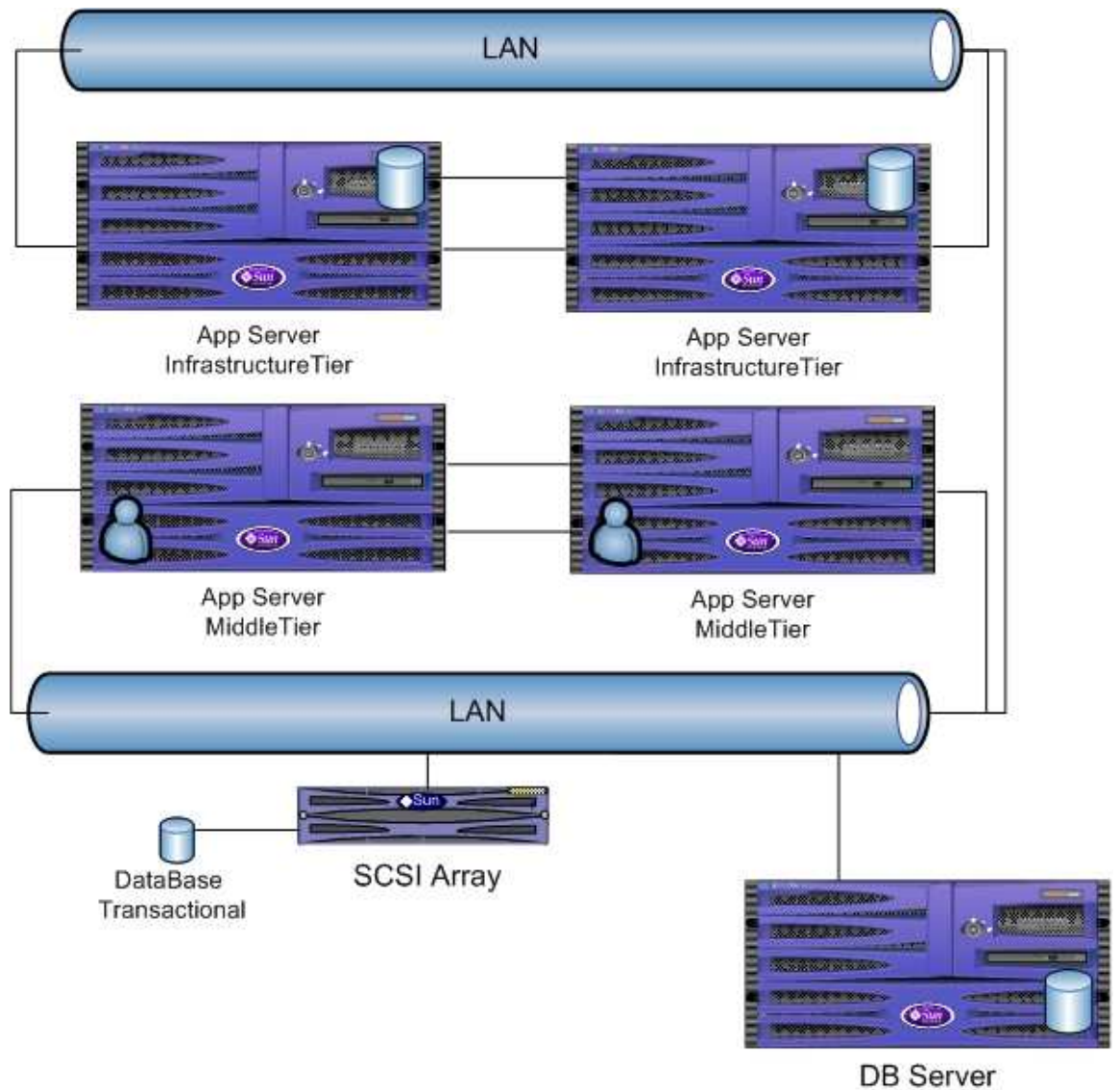
Elaborado por: Autores

4.5.1.1. Descripción

En este diseño se dividió en dos, la capa media del servidor de aplicaciones, asignándole dos servidores para proveerlo de alta disponibilidad, la capa de infraestructura iría instalada en el servidor destinado para base de datos, tanto los datos de la base transaccional como la base de infraestructura estarán en el arreglo de discos Storeedge 3200.

4.5.2. Diseño 2

Gráfico 13. Diseño 2.



Elaborado por: Autores

4.5.2.1. Descripción

Esta solución se basa en el diseño anterior pero agregando alta disponibilidad a la capa de infraestructura, para implementar esta capa en cada uno de los servidores se puede elegir tener la base de datos de la infraestructura de manera local, es decir, en el mismo servidor o en un

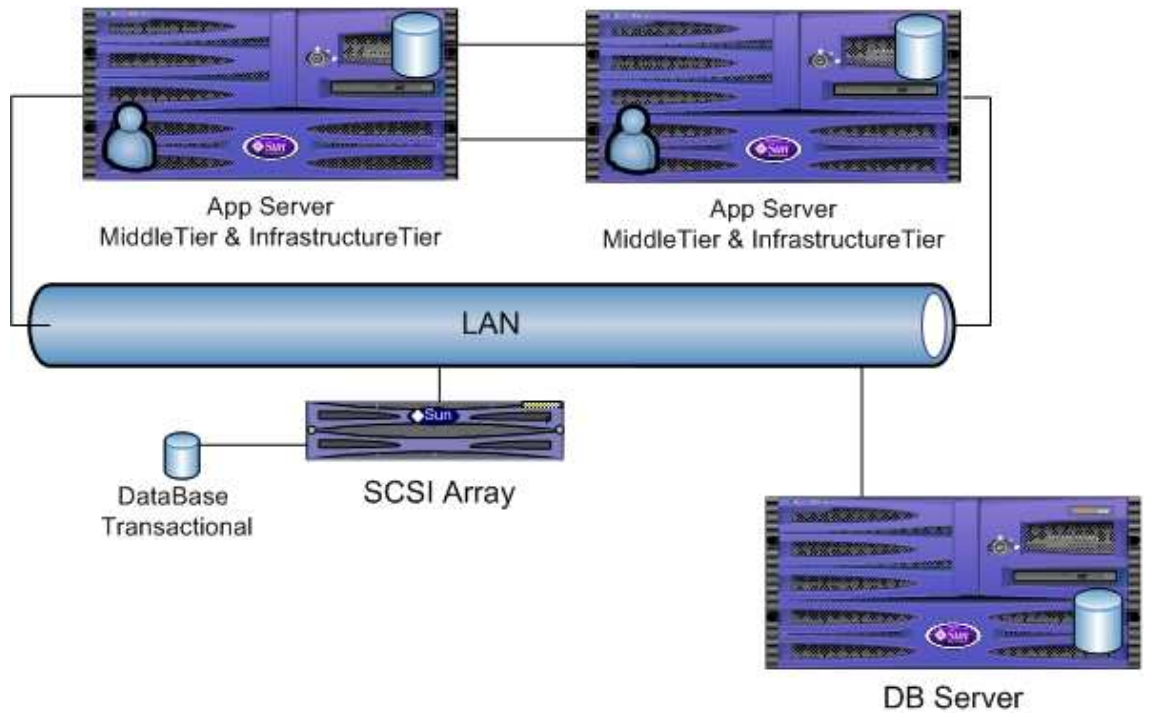
arreglo de discos, escoger una de las dos opciones depende mucho de las velocidades y disponibilidad que ofrezcan estos medios.

El tercer servidor es para uso exclusivo de la base de datos, esto es necesario debido a la alta transaccionalidad de las tablas como ya se menciono anteriormente.

La implementación de este diseño es más complicado que el anterior diseño que se presenta en esta tesis, ya que se necesita bastante inversión en tecnología de hardware y considerar en el refrigeramiento de estos equipos, el consumo de energía que cada uno utilizaría, espacio físico, etc. todo esto para poder brindar de alta disponibilidad al Centro de Cómputo de la U.C.S.G.

4.5.3. Diseño 3

Gráfico 14. Diseño 3.

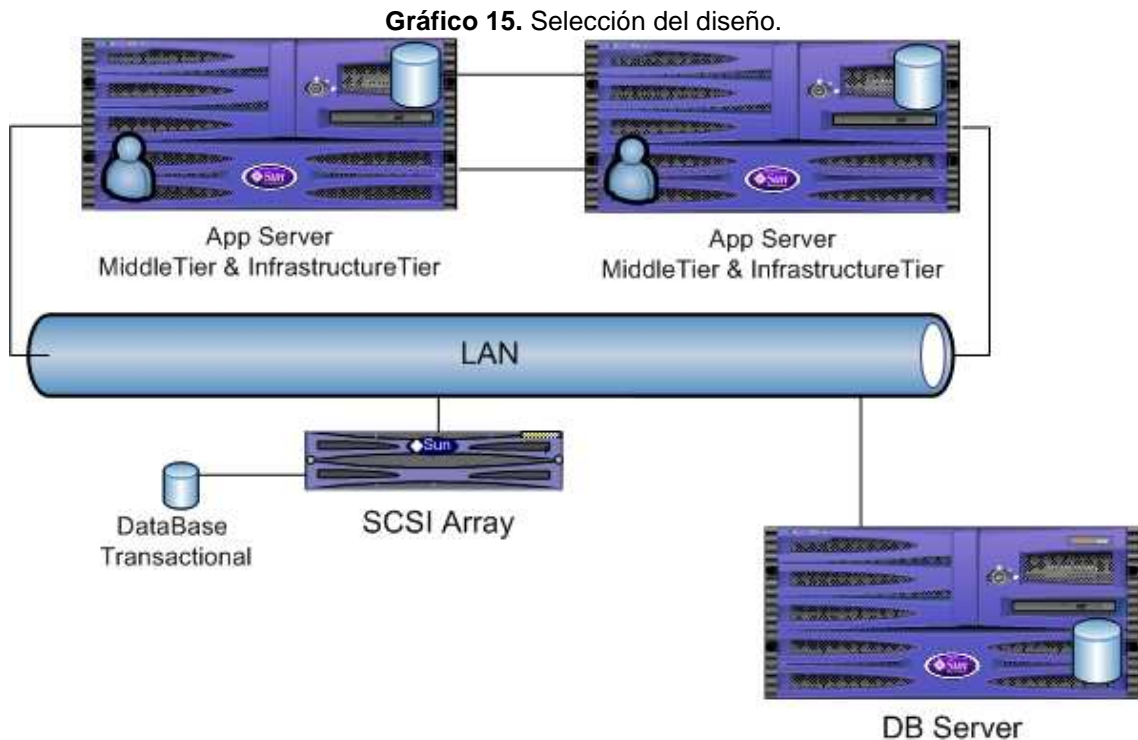


Elaborado por: Autores

4.5.3.1. Descripción

Este es un diseño óptimo el cual no incurre en inversión significativa de tecnología y permite conseguir alta disponibilidad. Las dos capas que conforman el servidor de aplicaciones están contenidas en cada uno de los servidores SunFire V490 y se mantienen sincronizados dentro de otras redes internas. Se utiliza un tercer servidor SunFire V490 que contiene la base de datos instalada y la data en el arreglo de discos Storeedge 3200.

4.6. SELECCIÓN DEL DISEÑO DE ALTA DISPONIBILIDAD



Elaborado por: Autores

4.6.1. Descripción

Se seleccionó el diseño 3 que brinda alta disponibilidad al Centro de Cómputo de la U.C.S.G. El recurso al que no se le puede dar el privilegio de estar fuera de servicio es el sistema S.I.U (Sistema Integral Universitario), el cual es utilizado por docentes, empleados y directores de cada una de las facultades. Este sistema se ejecuta en el servidor de aplicaciones de oracle 10gR2 el cual está contenido dentro de uno de los servidores SunFire V490.

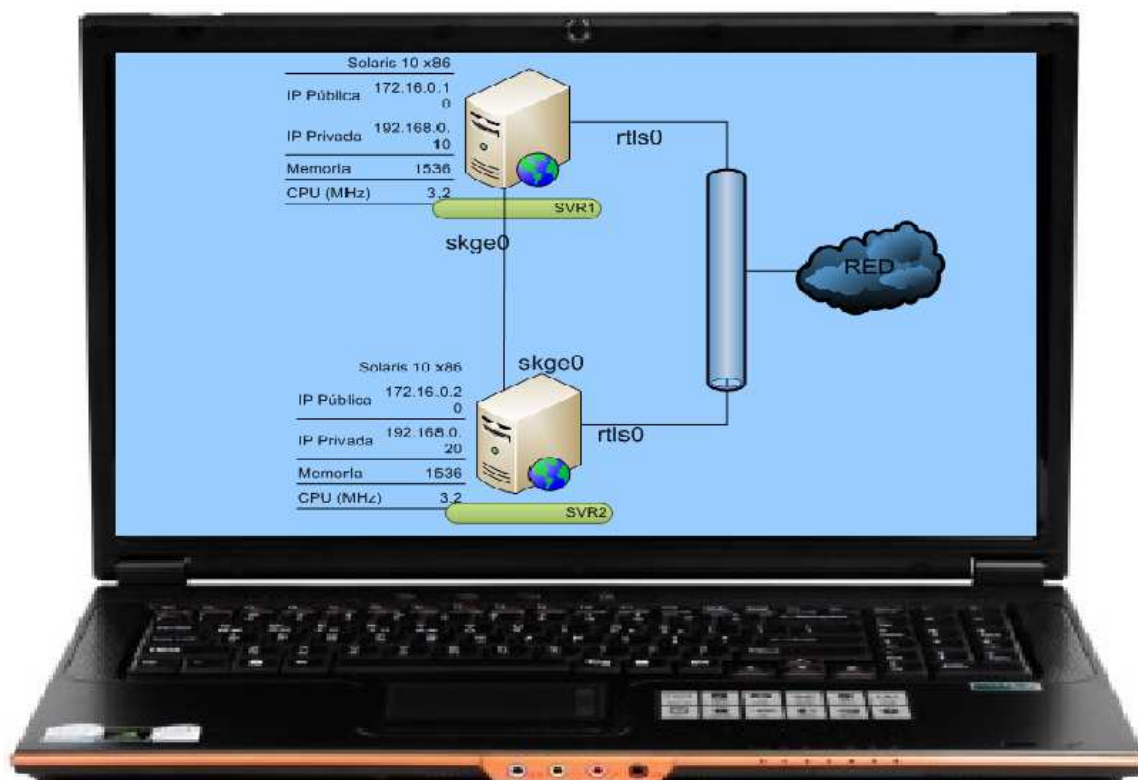
Para la validación de este diseño se ha evaluado crear un prototipo, pero lamentablemente el Centro de Cómputo de la U.S.C.G no cuenta con servidores disponibles, es por esta razón que se plantea el mismo diseño pero llevado a una arquitectura diferente y más accesible como Intel.

La implementación del prototipo solo contempla al servidor de aplicaciones, mas no al servidor de base de datos.

CAPITULO 5 : PROTOTIPO

5.1. DISEÑO

Gráfico 16. Diseño propuesto para alta disponibilidad en Arq. Intel.



Elaborado por: Autores

5.2. ESCENARIO 1:

Al principio de la elaboración de esta tesis se pensó que se tendría que dar mayor importancia a la instalación del Servidor de Aplicaciones ya que pertenece a Oracle y en ese entonces Noviembre del 2009 dicha compañía no había adquirido a Sun y en la mayoría de sus productos están certificados bajo GNU/Linux como RedHat u Oracle Linux.

Al empezar la instalación de Servidor de Aplicaciones se presentaron algunos inconvenientes; como la falta de documentación bajo el sistema operativo Solaris 10, la necesidad de un servidor de directorio.

Luego de superar todos los inconvenientes se pensó que la instalación y configuración de Sun Clúster 3.2 en ese entonces ya adquirido por Oracle no supondría ningún problema ya que dicho software fue desarrollado por Sun y por lo tanto compatible con su sistema operativo Solaris 10.

En la documentación oficial de Sun se especifica que el software Sun Clúster se puede ejecutar tanto en arquitectura Sparc e Intel. Con esta premisa se pensó que la instalación de dicho software no daría mayor problema como el Servidor de Aplicaciones.

En este escenario se pretende implementar el diseño propuesto como alta disponibilidad para el Centro de Cómputo de la U.C.S.G, en un ambiente virtual con las siguientes características:

Para virtualizar se utilizó el software Oracle Maquina Virtual VirtualBox 3.2.8 ya que esta solución fue desarrollada por Sun y por lo tanto se supuso que se “entendería” mejor con el sistema operativo anfitrión Solaris 10.

VirtualBox se lo instalo dentro del sistema operativo Windows 7 bajo la arquitectura Intel con un procesador i3 de 64bits.

Como se muestra en el grafico 17 se crearon dos maquinas virtuales asignándole a cada una lo siguiente:

1.5Gb de memoria, 30Gb de disco duro virtual, 4 tarjetas de red virtual Intel PRO/1000 MT Server (Bridged Adapter, Marvell Yukon 88E8040 PCI-E FastEthernet Controller).

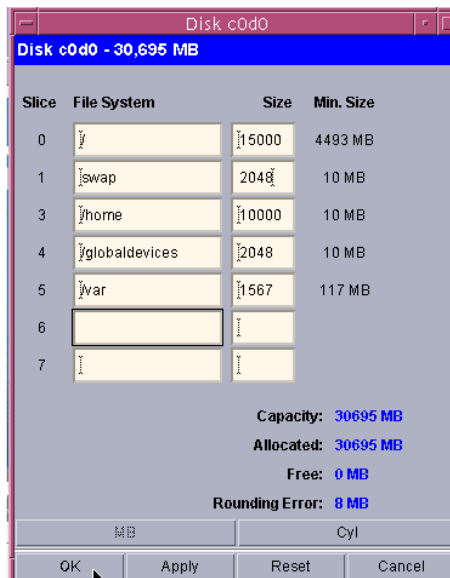
Gráfico 17. Descripción de maquinas virtuales.

OS Type:	Solaris
System	
Base Memory:	1536 MB
Processor(s):	1
Boot Order:	CD/DVD-ROM, Hard Disk
Display	
Video Memory:	12 MB
3D Acceleration:	Disabled
2D Video Acceleration:	Disabled
Remote Display Server:	Disabled
Storage	
IDE Controller	
IDE Primary Master:	SolarisNodo1.vdi (Normal, 30.00 GB)
IDE Secondary Master (CD/DVD):	Host Drive 'D:'
SCSI Controller	
Audio	
Host Driver:	Windows DirectSound
Controller:	ICH AC97
Network	
Adapter 1:	Intel PRO/1000 MT Server (Bridged adapter, Marvell Yukon 88E8040 PCI-E Fast Ethernet Controller)
Adapter 2:	Intel PRO/1000 MT Server (Bridged adapter, Marvell Yukon 88E8040 PCI-E Fast Ethernet Controller)
Adapter 3:	Intel PRO/1000 MT Server (Bridged adapter, Marvell Yukon 88E8040 PCI-E Fast Ethernet Controller)
Adapter 4:	Intel PRO/1000 MT Server (Bridged adapter, Marvell Yukon 88E8040 PCI-E Fast Ethernet Controller)

Elaborado por: Autores

En cada máquina virtual se le instalo Solaris 10 x86, asignándole las particiones para: (/ , /home, /globaldevices, /var, /swap) como se muestra en la siguiente imagen.

Gráfico 18. Partición disco.



Elaborado por: Autores

Luego de esto con el DVD oficial de Oracle Sun Solaris 10 se procedió a instalar en las dos maquinas los siguientes parches:

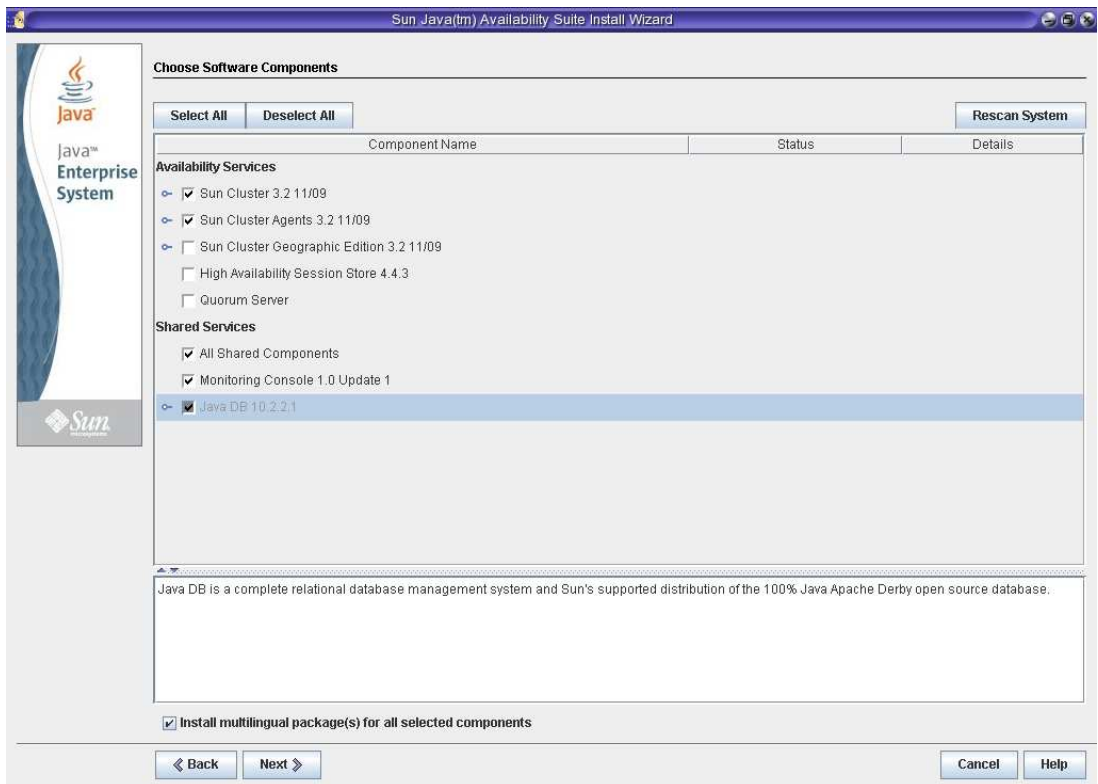
119255-70 140084-02 118919-21 140797-01 140900-01
122035-05 120720-02 123840-04 140861-01 121119-16
12785-01 126869-04 127756-01 138218-01 118855-36
140172-04

Al finalizar la instalación antes mencionada se procedió a descargar las actualizaciones recomendadas por el software Update Manager del Solaris, ya que con estas configuraciones básicas se tiene el ambiente requerido para implementar un prototipo en base al diseño planteado anteriormente.

Se descargo del sitio oficial de Oracle Sun el software de alta disponibilidad Oracle Sun Clúster 3.2 u3 el cual estaba en formato ISO, el mismo que tiene como ultima fecha de actualización el mes de noviembre del año 2009.

Las maquinas virtuales fueron configuradas para iniciar desde ese archivo de imagen, Solaris 10 monto dicho archivo como un unidad de DVD el cual se uso para instalar satisfactoriamente el clúster, escogiendo de entre sus productos los siguientes: Sun Cluster 3.2 y Sun Cluster Agents 3.2 además los servicios compartidos: Todos los componentes compartidos y Consola de monitoreo 1.0 actualización 1, como se puede apreciar en el grafico 19.

Gráfico 19. Instalación de Sun Clúster.



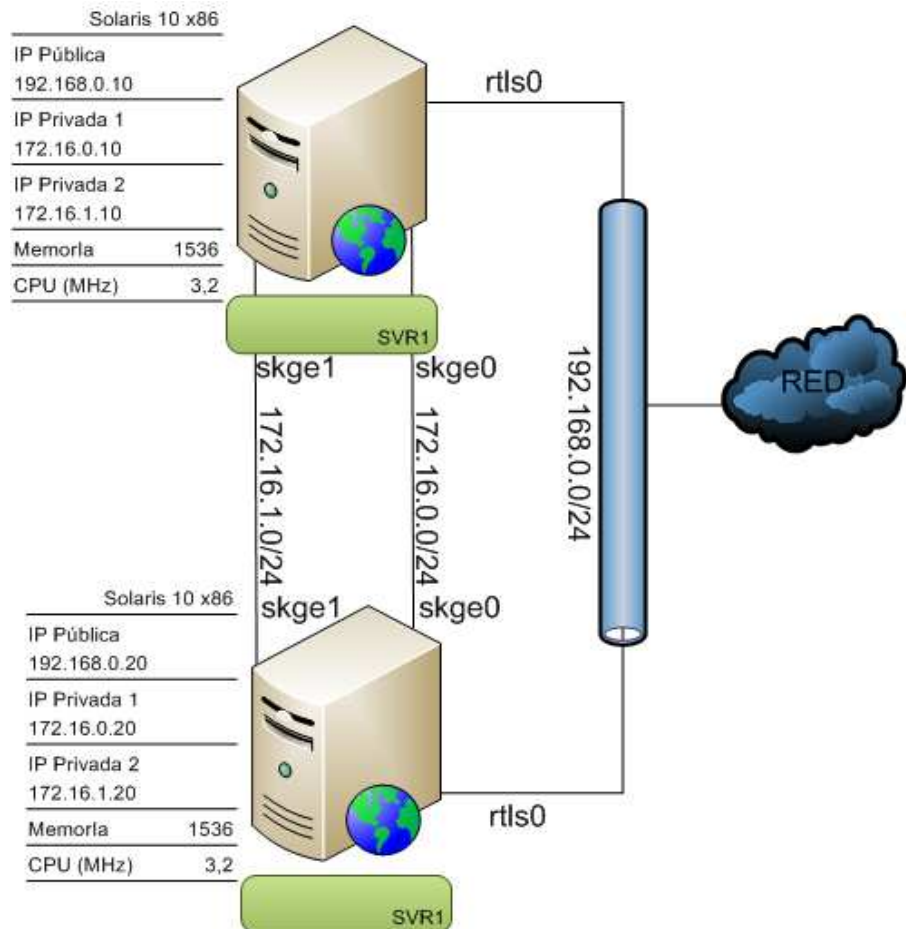
Elaborado por: Autores

Sun Cluster 3.2, Sun Cluster Agents 3.2 y los servicios compartidos fueron instalados exitosamente en los ordenadores, luego al primer servidor (SVR1) se lo configuró como clúster y se lo estableció como primer nodo y al servidor (SVR2) se lo añadió como segundo nodo perteneciente al clúster. Esta configuración envió dichos servidores a reiniciarse, pero cuando iniciaban sesión la configuración no persistía en ninguno de los dos servidores.

Por ese motivo se realizaron dos intentos más para establecer el clúster, el cual resulto sin éxito alguno, y por lo tanto se decidió que el ambiente virtual no era el más adecuado para la implementación del prototipo ya que Oracle Sun Clúster solo está certificado bajo arquitectura Sparc e Intel y por mucho que se quiera una maquina virtual nunca será lo mismo como una maquina real.

5.3. ESCENARIO 2

Gráfico 20. Diseño del prototipo en arquitectura Intel.



Elaborado por: Autores

Para la aplicación del prototipo se siguió la guía oficial de Oracle #819-0420 con el nombre de **Sun Cluster Software Installation Guide for Solaris OS** pagina 97, véase el anexo 1. El cual provee de una matriz de instalación personalizada del software Sun clúster 3.2.

En este escenario se pretende lo mismo que en el escenario anterior, pero en un ambiente real con dos equipos físicos cada uno con las siguientes características:

- Memoria 1.5 Gb DDR
- Procesador Intel
- Una Interface de red integrada en el mainboard.
- Dos interfaces de red PCI DLink DGE-530T de 10/100/1000 mbps.
- Disco Duro SATA de 80 Gb.

Se instalo en cada una de los equipos el sistema operativo Solaris 10 x86, asignándole las particiones para: (/ , /home, /globaldevices, /var, /swap, /tmp) similares a las de un sistema operativo G.N.U/Linux a excepción de la partición globaldevices.

Luego de esto se procedió a parchar el sistema operativo de forma similar como en el escenario anterior de maquinas virtuales, lo cual resulto sin novedad alguna.

Se instalo una sola vez el driver de red para de las tarjetas externas de marca DLink, lo cual fue un poco complicado ya que estos drivers o también llamados paquetes no se los puede descargar del sitio oficial de Solaris. Por motivo de que esta tarjeta (DLink 530T) no está soportada por la lista de compatibilidad de Hardware (H.C.L) publicada en el sitio oficial de Oracle Sun.

Solaris reconocerá el nuevo hardware de las tarjetas de red PCI instaladas en el mainboard, pero no les agregará una instancia de controlador hasta no instalar el driver, con el siguiente comando se puede observar las propiedades del hardware:

```
# prtconf -v  
  
pci1186,4b01 (driver not attached)  
  
Hardware properties:  
name='model' type=string items=1  
value='Ethernet controller'
```

La primera línea de las propiedades de una de las tarjetas de red, proporciona información sobre el identificador del fabricante y el identificador del producto de dicho hardware, separados por una coma respectivamente.

Se instala el controlador de red como se detalla a continuación; Este controlador servirá para todas las tarjetas PCI del mismo modelo y marca, para agregar el controlador de red o paquete en Solaris se lo consigue con el siguiente comando:

```
# pkgadd -d . SKGEsol
```

Luego de esto aparecerá un asistente por línea de comando que le pedirá que por cada interfaz de red le asigne una dirección IP y una máscara de red.

Con los identificadores antes mencionados se escribe la siguiente línea donde se actualiza el controlador de la tarjeta de red y se le adhiere una instancia del controlador.

```
# /usr/sbin/update_drv -a -i "pci1186,4b01" skge
```

Con el siguiente comando se pueden observar las propiedades del hardware de la tarjeta de red PCI y se filtra por los identificadores antes mencionados.

```
# prtconf -D | grep 1186,4b01
pci1186,4b01, instance #0 (driver name: skge)

pci1186,4b01, instance #1 (driver name: skge)
```

Para finalizar esta configuración se recomienda que, por cada interfaz de red instalada en el equipo ya sean estas destinadas para la red pública o privada se cree archivos en el directorio /etc con el nombre **hostname.skge#intance** como por ejemplo:

```
/etc/hostname.skge0
```

```
/etc/hostname.skge1
```

Y dentro de estos archivos se les agregue un nombre para luego resolverlo dentro del archivo `/etc/hosts` como por ejemplo:

```
Skge0 172.16.0.10
```

En el archivo `/etc/netmask` se agregara la dirección de red con su respectiva máscara como por ejemplo:

```
172.16.0.0      255.255.255.0
```

Además se instaló el software Oracle Sun Clúster 3.2 en cada uno de los nodos, escogiendo los siguientes componentes dentro del instalador: Sun Clúster 3.2 y Sun Clúster Agents 3.2 además los servicios compartidos similar al escenario anterior. Luego de esto se procedió a instalar el software de Quórum server para el nodo reservado como tal.

Una vez instalado satisfactoriamente el software de Quórum se debe editar un archivo de configuración ubicado en `/etc/scqsd/scqsd.conf`

Con el siguiente formato:

```
/usr/cluster/lib/sc/scqsd [-d quorumdirectory] [-i instancename] -p port
```

Entonces dicho archivo quedaría con la siguiente línea:

```
/usr/cluster/lib/sc/scqsd -d /var/scqsd -i QS1 -p 9000
```

Y para iniciar el servicio de Quórum Server se utiliza el siguiente comando:

```
/usr/cluster/bin/clquorumserver start QS1
```

Al registrar cada uno de los nodos en el quórum Server, los nodos se reinician y no se puede establecer el clúster, por motivo de que es necesario de un dispositivo de quórum o también llamado Quórum Device, esto es generalmente un arreglo de

discos SCSI con el cual no se cuenta, además de esto se necesitarían interfaces PCI SCSI para poder añadir este arreglo a cada uno de los nodos.

Estos problemas no se darían en una arquitectura Sparc por lo que se pueden tener un máximo de 64 nodos por cada clúster. Intel está limitado a tener un máximo de dos nodos por clúster como se describe en la documentación oficial de Oracle #819-2969 con el nombre de **Sun Cluster Concepts Guide for Solaris OS pagina 29, véase el anexo 2**, el cual no es suficiente para establecer un clúster ya que en el caso de estudio se necesitan de la mitad de votos más uno, es decir, tres votos.

El servidor1 asigna un voto, el servidor2 asigna otro voto, pero no se cuenta con el tercer voto (un arreglo de discos SCSI). Por esta limitante se decidió no continuar con la implementación del prototipo en arquitectura Intel y con ende las ultimas 3 etapas del la metodología planteada en el capítulo 4 de esta tesis.

CAPITULO 6 : CONCLUSIONES Y RECOMENDACIONES

6.1. CONCLUSIONES

- En el proyecto se ha podido determinar que la situación actual del servidor de aplicaciones de la U.C.S.G, que debería mejorar con respecto a la alta disponibilidad que al momento no cuenta.

- El diseño propuesto para este trabajo, consta de dos servidores con iguales o similares características, y un repositorio de la base de datos, el mismo que estará conectado a la red, para lograr la alta disponibilidad del servicio.

- El desarrollo de un prototipo con esta tecnología de clúster en ambientes virtuales no es posible simular.

- Solaris oficialmente no soporta dispositivos de hardware que no estén en H.L.C. (Hardware List Compatibility).

- Para el implementar clúster con Oracle Sun Clúster 3.2 es indispensable el uso de un dispositivo de quórum como un arreglo de discos.

- Para empresas medianas o pequeñas, cuyos servidores no sean de arquitectura Sparc el uso de esta tecnología implica inversión en hardware.

- Esta tecnología en arquitectura Intel tiene una limitante de solo 2 nodos por cada clúster.

- Para establecer el quórum sin importar la arquitectura es necesario de la mitad de votos más uno.

- El dispositivo de quórum tiene $N - 1$ votos, donde N es el número de conexiones hacia los nodos.

- Sun Clúster no admite dispositivos replicados como dispositivos de quórum.

- Existen dos problemas comunes en un clúster cerebro dividido y amnesia.

- Se puede configurar hasta seis interconexiones de clúster mediante el comando `clsetup`.

- Las interconexiones de un clúster pueden ser con tecnología Ethernet o Infiniband.

- Si se especifica un conmutador de transporte, puede añadir más fácilmente otro nodo al clúster, posteriormente.

- Para un clúster de más de dos nodos es obligatorio el uso de un conmutador de transporte.

- Para ofrecer la mayor disponibilidad del quórum posible, asegúrese de que el número total de votos con los que contribuyen los dispositivos del quórum es menor que el número de votos con el que contribuyen los nodos. En caso contrario, los nodos no pueden formar un clúster si todos los discos no están disponibles, incluso si todos los nodos están funcionando.

6.2. RECOMENDACIONES

➤ Después del análisis y evaluaciones de este proyecto se recomienda que el Centro de Cómputo de la U.C.S.G debería contar con un servidor respaldo de iguales o menores características al actual existente en producción, para una posible implementación de alta disponibilidad.

➤ Si se desea implementar Oracle Sun Cluster 3 en el Centro de Cómputo de la U.C.S.G se debería continuar con uso de servidores en arquitectura Sparc, ya que dicha arquitectura está soportada por dicho software.

➤ Se debe tomar en cuenta que el diseño propuesto podría ser implementado con los equipos de pruebas que tiene el Centro de Cómputo de la UCSG.

➤ Hacer uso de esta tecnología solo si se tiene servidores con arquitectura Sparc, ya que se pueden usar hasta 64 nodos por cada clúster, y existe más documentación oficial en esta arquitectura.

➤ Es indispensable verificar la existencia de uno o varios agentes para monitorear los servicios de software a los que se les quieran dar alta disponibilidad.

➤ Si desactiva el fencing global, sus datos podrían ser vulnerables a la corrupción durante la conmutación por error del servicio al que se le brinda alta disponibilidad.

➤ Siempre que sea posible, instalar dos o más interconexiones de nodos para proporcionar redundancia y escalabilidad y por lo tanto mayor disponibilidad.

➤ Para ciertos tipos de adaptadores, como el SCI-PCI no se debe utilizar el nombre del puerto predeterminado

CAPITULO 7 : BIBLIOGRAFÍA

- ❑ UOC, 2008, Administracion de sistemas operativos en red. Barcelona – España

- ❑ Sun Cluster, 2006. Guía de conceptos para el SO Solaris. Extraído el 14 de Agosto del 2010 desde:
<http://dlc.sun.com/pdf/819-2059/819-2059.pdf>

- ❑ Diego E. Aguirre, 2008. Configurando Sun Cluster 3.2 con Oracle 10g RAC y ASM. Extraído el 14 de Agosto del 2010 desde:
http://www.sun.com/bigadmin/content/submitted/es/span_cluster_rac.jsp

- ❑ Amy Rich, junio 2005. Instalación y configuración del software Sun Cluster 3.1 09/04 para aplicaciones de alta disponibilidad. Extraído el 14 de Agosto del 2010 desde:
http://www.sun.com/bigadmin/hubs/multilingual/spanish/content/install_cluster.jsp

- ❑ Sun Microsystems, Inc., 2004, Guía de instalación para el sistema operativo Solaris Extraído el 15 de Agosto del 2010 desde:
<http://docs.sun.com/app/docs/doc/817-6374/6mlp2dqdp?l=es&a=view>

- Paulo Clavijo, 2010, Clusters de Alta Disponibilidad (HA). Extraído el 15 de Agosto del 2010 desde:
<http://www.lintips.com/?q=node/119>

- Total Publishing Network S.A. 2008, Funcionamiento de virtualizacion y para que sirve. Extraído el 19 de Septiembre del 2010 desde:
<http://muypymes.com/tecnologia/software/4605-como-funciona-la-virtualizacion-y-para-que-sirve.html>

- Oracle Corporation, 2010, Alta disponibilidad real de las aplicaciones. Extraído el 19 de Septiembre del 2010 desde:
<http://docs.sun.com/app/docs/doc/817-6910/6mm91j7sv?l=es&a=view>

- Oracle Corporation, 2010, Estructura de alta disponibilidad. Extraído el 19 de Septiembre del 2010 desde:
<http://docs.sun.com/app/docs/doc/819-2059/6n4cgq17g?l=es&a=view>

- Oracle Corporation, 2010, Guía de instalación para SO Solaris. Extraído el 20 de Septiembre del 2010 desde:
<http://docs.sun.com/app/docs/doc/820-3928/geyni?l=es&a=view>

- Advanced Networks & Support, 2010, Componentes de un cluster. Extraído el 20 de Septiembre del 2010 desde:
<http://clusterfie.epn.edu.ec/clusters/Definiciones/definiciones.html>

- Scribd, 2010, Que es clúster. Extraído el 20 de Septiembre del 2010 desde:
<http://www.scribd.com/doc/33611705/Que-Es-Un-Cluster>

- Zerowait Corp, 2008, Diferencia entre NAS y SAN. Extraído el 20 de Septiembre del 2010 desde:
<http://www.nas-san.com/differ.html>

CAPITULO 8: ANEXOS

8.1. ANEXO 1

Custom Mode - If you will use Custom mode and customize the configuration data, complete the following worksheet.

Component	Description/Example	Answer
Software Patch Installation	Do you want <code>scinstall</code> to install patches for you?	Yes No
	If yes, what is the name of the patch directory?	
	Do you want to use a patch list?	Yes No
Sponsoring Node	What is the name of the sponsoring node? Choose any node that is active in the cluster.	
Cluster Name	What is the name of the cluster that you want the node to join?	
Check	Do you want to run the <code>sccheck</code> validation utility?	Yes No
Autodiscovery of Cluster Transport	Do you want to use autodiscovery to	Yes No

Component	Description/Example	Answer	
	configure the cluster transport? If no, supply the following additional information:		
Point-to-Point Cables	Does the node that you are adding to the cluster make this a two-node cluster?	Yes No	
	Does the cluster use transport junctions?	Yes No	
Cluster-Transport Junctions	If used, what are the names of the two transport junctions? Defaults: <code>switch1</code> and <code>switch2</code>	First	Second
Cluster-Transport Adapters and Cables	What are the names of the two transport adapters?	First	Second
	Where does each transport adapter connect to (a transport junction or another adapter)? Junction defaults: <code>switch1</code> and <code>switch2</code>		

Component	Description/Example	Answer	
	For transport junctions, do you want to use the default port name?	Yes No	Yes No
	If no, what is the name of the port that you want to use?		
Global-Devices File System	What is the name of the global-devices file system? Default: <code>/globaldevices</code>		
Automatic Reboot	Do you want <code>scinstall</code> to automatically reboot the node after installation?	Yes No	

8.2. ANEXO 2

x86: Sun Cluster Topologies for x86

A topology is the connection scheme that connects the cluster nodes to the storage platforms that are used in the cluster. Sun Cluster supports any topology that adheres to the following guidelines.

- ✓ ☐ *Sun Cluster that is composed of x86 based systems supports two nodes in a cluster.*
- ✓ ☐ *Shared storage devices must connect to both nodes.*

Sun Cluster does not require you to configure a cluster by using specific topologies. The following clustered pair topology, which is the only topology for clusters that are composed of x86 based nodes, is described to provide the vocabulary to discuss a cluster's connection scheme. This topology is a typical connection scheme.

The following section includes a sample diagram of the topology.

x86: Clustered Pair Topology for x86

A clustered pair topology is two nodes that operate under a single cluster administrative framework. In this configuration, failover occurs only between a pair. However, all nodes are connected by the cluster interconnect and operate under Sun Cluster software control. You might use this topology to run a parallel database or a failover or scalable application on the pair. The following figure illustrates a clustered pair configuration.

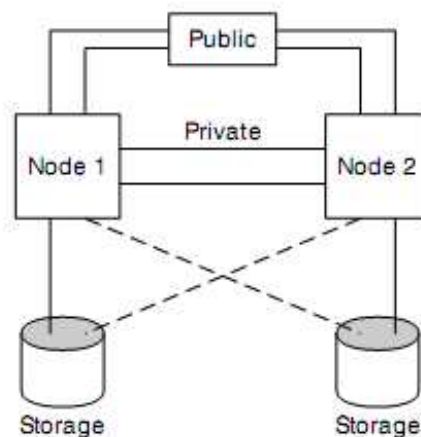


FIGURE 2-6 x86: Clustered Pair Topology