



SISTEMAS DE ALTA DISPONIBILIDAD.

Blanco Portillo, Marianela
Mblanco@ ITTACA.com
Universidad Dr. Rafael Belloso Chacín. Venezuela

RESUMEN

El propósito de esta investigación fue elaborar una propuesta para el diseño de Sistemas de Alta Disponibilidad para Petróleos de Venezuela, S.A. El estudio ameritó diagnosticar la plataforma tecnológica con la que actualmente cuenta la organización. El estudio realizado permitió determinar las debilidades y fortalezas de la empresa, utilizándose estos resultados como base para la propuesta. Este estudio se guió por una metodología de trabajo donde se definieron y analizaron las dimensiones e indicadores de la variable objeto de estudio, Alta Disponibilidad. Se utilizó una población de seis (6) expertos del área de Servidores. Se aplicó una encuesta que le permitió al investigador establecer el diagnóstico de las cuatro dimensiones consideradas en este estudio, Operabilidad Continua, Fiabilidad, Capacidad de Recuperación y Negocios. La propuesta consistió principalmente en la adecuación de la infraestructura actual para mantener operables las aplicaciones críticas de la organización. Se logró definir claramente las actividades a realizar por el personal de la empresa que presta sus servicios a PDVSA y las adquisiciones de hardware y software indispensables para iniciar con un proyecto de este tipo.

Palabras Claves: Inoperabilidad, Operabilidad, Fiabilidad, Capacidad de Recuperación, SAN.

ABSTRACT

The purpose of this investigation was to establish a proposal for the design of High Availability Programs for Petróleos de Venezuela S.A. The study required a diagnostic of the actual technological platform of the organization. The study allowed determinated of the weaknesses and strengths of the company and the results obtained were used as the base of the outlined proposal. This study was guided by a work methodology in which the dimensions and variable indicators studied, High Availability, were defined and analyzed. If used a population of six (6) experts was used in the Servers area. A survey was applied which allowed the researcher to establish a diagnostic for the four dimensions considered in this study: Continuous Operation, Trustworthiness, Recuperation Capability and Business. The proposal consisted mainly in the accommodation of the present infrastructure



to maintain in operation the critical applications of the organization. The activities that were to be carried out by the company personnel, which renders its service to PDVSA were clearly defined, as was the acquisition of the indispensable hardware and software needed to initiate such a project.

Key words: Inoperative, Operative, Trustworthiness, Recuperation Capability, SAN.

INTRODUCCION

En la actualidad un factor clave para el  xito de las empresas es el contar siempre con las aplicaciones que apoyan sus operaciones principales, pues la falta de disponibilidad de las mismas puede repercutir en costos, tiempos, esfuerzos y por supuesto en la confianza e insatisfacci n de sus clientes.

En este sentido se ofrece una propuesta de dise o para sistemas de alta disponibilidad para la empresa petrolera, en base al estudio de las diferentes y principales alternativas tecnol gicas existentes en el mercado.

SISTEMAS DE ALTA DISPONIBILIDAD

Martin (1999, p.107) define la disponibilidad como "la cantidad o porcentaje de tiempo que un sistema se encuentra disponible para la realizaci n de una determinada misi n". Sin embargo algunos autores consideran que no es solo un problema de aplicaci n sino del entorno que se afecta al no contar con el mismo, por ejemplo para Farley, Stearns y Hsu (1999, p.215) Los sistemas de Alta disponibilidad se deben entender como "sistemas que permiten a las aplicaciones seguir operando a pesar que el hardware o software falle. Estos sistemas deben proteger a los usuarios de fallos de software as  como de fallas presentes en las unidades de procesamiento, disco o en los componentes de red". Por lo que es razonable pensar que de ser posible eliminar los principales puntos o causas de falla en cualquiera de estos aspectos (software, hardware, red) mayores ser n las probabilidades de contar con la disponibilidad de los sistemas cr ticos de la organizaci n, m s aun si se implantan componentes redundantes o tolerantes a fallos quienes junto a una alternativa de software puedan entrar en funcionamiento y asumir la responsabilidad en el caso que un componente falle.

Ahora bien, si bien es cierto que la redundancia asegura un alto grado de la disponibilidad de los sistemas, tambi n es cierto que debido a los altos costos en los que se incurrir a por la duplicaci n de cada una de las partes



sería conveniente primeramente definir los requerimientos de disponibilidad, el tiempo tolerante a fallos ó los períodos de inoperabilidad aceptados.

Una vez identificadas realmente la criticidad de las aplicaciones y conocidos los niveles de tolerancia permitido, ya entonces se puede pensar en la corrección de puntos de fallo y en la redundancia de componentes

En este sentido, es necesario mencionar todos aquellos puntos de fallos en los cuales se debe centrar la atención

Falla de las Unidades de sistemas de Procesamiento (SPU): El SPU consiste en un grupo de elementos, cada uno de los cuales es propenso a fallar provocando la pérdida del servicio. Entre los más importantes de este grupo cabe la pena mencionar:

La unidad central de procesamiento, Los controladores de entrada y salida y la Tarjeta Principal.

Para eliminar los downtime asociados a la falla de estos componentes, se puede optar por el uso de una tecnología cluster, la cual permite reparar o reemplazar los mismos sin tener pérdida en el sistema. En lo que respecta particularmente a los fallos en la unidad central de procesamiento se podría quizás pensar en un esquema de multiprocesamiento simétrico,

Como otro punto de falla sobre el cual hay que centrar la atención es la falla en la red de área local (LAN). Las fallas en la red pueden ocurrir ya sea por el tipo de conexión utilizada, por las tarjetas de interfaz, por los routers, concentradores o incluso por las velocidades de transferencias utilizadas en la misma. Este tipo de problemas puede eliminarse bien sea proporcionando una redundancia total en las conexiones o configurando el switching de las tarjetas de interfaz.

Otro punto de fallo de interés radica en los subsistemas de almacenamiento, en este caso podría pensarse en la técnica de espejos, donde hay dos unidades independientes para cada disco de datos, siendo la unidad principal la encargada de manejar las peticiones de los usuarios y utilizándose la segunda unidad como un disco imagen de la primera, es decir, en este esquema los dos componentes realizan exactamente el mismo trabajo de forma entrelazada, teniéndose de esta manera dos sistemas equivalentes, de forma tal que si uno de ellos falla el otro continua trabajando.



Como otra alternativa podría ser la duplexación de discos o duplicación como también se le conoce, quien virtualmente opera de la misma manera que la técnica de espejo, excepto que añade un nuevo controlador, de allí que el rendimiento se incremente debido a la menor contención que ocurre en el bus de entrada y salida.

En lo que respecta a fallos por los subsistemas de almacenamiento, hoy día se cuenta con una serie de sistemas o metodologías de combinación de discos, conocidas como Arreglos de Discos Redundantes (RAID) que son mas que "un grupo de discos en los cuales se almacenan además de la información normal una serie de información redundante que permite, que en el caso que alguno de los discos falle, la información pueda regenerarse de forma automática". Palmer (2000, p. 289)

Este esquema puede ser implementado en diferentes formas dependiendo de la distribución que se haga en el disco, del tipo de rendimiento y del coste.

El método RAID nivel 0 fue el primero que apareció. En realidad este nivel RAID no incorpora tolerancia a fallos, por lo que un fallo en alguno de sus dispositivos provoca la perdida de la información almacenada en todas las unidades. Este esquema a pesar de no ofrecer tolerancia es considerado, por ser el primer método que apareció para la conjunción de varios discos en una única partición.

Un RAID de nivel 1 son discos en espejos, de esta forma se tiene duplicada la información en dos discos distintos, de manera que si uno de ellos falla, se continúa trabajando con el otro, sin por ello perder ni eficiencia en el servidor ni por supuesto la información. Este esquema consigue una fiabilidad en el sistema muy mejorada ya que aunque fallaran ambos discos simultáneamente (situación poco probable) se ahorraría tiempo de recuperación, adicionalmente las operaciones de lectura de un disco que forma parte de un espejo son más rápidas, ya que la lectura puede realizarse de forma simultanea en ambos discos. Sin embargo, en términos de almacenamiento solo se cuenta con el 50% de la capacidad total del disco, en operaciones de escritura se afecta el rendimiento ya que la información debe guardarse en dos sitios y en términos monetarios los costos de este esquema son sumamente elevados.

El RAID 5, método de tolerancia a fallos más utilizado en la actualidad, tiene la capacidad de acceder a tantos discos como necesite al mismo tiempo pero en peticiones independientes, por lo que su tasa de entrada y salida es mucho mayor si se compara con los esquemas predecesores. En



este esquema la información se reparte a nivel de bloque e incluso de registro. Aquí desaparece la necesidad de un disco de paridad, sin embargo su implementación requiere el manejo de un mínimo de 3 discos. En este esquema uno de los discos conocido como Spare es usado como backup de todos los volúmenes raid que han sido configurados en el arreglo, de forma que si cualquiera de los volúmenes falla, el spare es utilizado para restablecer la información.

Entre las principales ventajas que este esquema ofrece se pueden mencionar: Mejor utilización del disco, buena velocidad de lectura y disminución significativa de los costos. Sin embargo y a pesar de ser un muy buen esquema para el manejo de tolerancia a fallos, este esquema tiene como punto en contra, el bajo rendimiento que presenta para las operaciones de escritura.

Adicional a todos los esquemas raid ya mencionados, también se cuenta con los arreglos de discos EMC o arreglos simétricos que permiten conexiones desde el mismo disco de datos a múltiples nodos del clusters a través de diferentes buses.

El RAID, originó una nueva técnica conocida como intercambio en caliente, que no es mas que la capacidad de sacar o introducir discos, mientras este continua funcionando y suministrando entradas y salidas al sistema al que está conectado. Por supuesto, esto requiere un diseño que permita manejar las sobrecargas de energía y las interrupciones que se producen en el bus interno de entrada y salida como resultado de las extracciones o inserciones realizadas. En este ámbito se cuenta con una técnica en particular conocida como Disk Storage Enclosure, quien permite realizar un intercambio en caliente de los discos espejos, sin perder el servicio, mientras el sistema operativo continua corriendo y el dispositivo funcionando.

Como se ve existe una extensa variedad de alternativas de hardware para ofrecer alta disponibilidad de las aplicaciones, cada una de las cuales se adapta a un escenario en particular. Sin embargo como un esquema general, que actualmente asegura la alta disponibilidad, donde actualmente se enmarcan la mayoría de las soluciones y que quizás es el esquema mas implementado es la tecnología clusters, definida por Weygant (2000 p.73) como un "grupo de servidores que poseen suficiente redundancia de software y hardware de forma que la falla de alguno de sus componentes no impida la disponibilidad del servicio ".Consiguiéndose así potenciar las principales características de las aplicaciones de misión crítica.



Esta tecnología puede ser implementada de tres (3) maneras básicamente.

- Configuración Activa/Standby:
- Configuración Activa/Activa
- Configuración de Base de Datos Paralela

Como última opción no puede dejar de mencionarse la tecnología de Red de Area de Almacenamiento (Storage Area Network o SAN) pues es allí donde se dirigen las tendencias del mercado y donde se están invirtiendo grandes esfuerzos.

SAN es una red dedicada en exclusiva a funciones de almacenamiento, la cual se construye sobre una red de fibra óptica o Fibre Channel a la que se conectan todos los servidores que utilizan el almacenamiento, ofreciendo un ancho de banda potente. SAN es una tecnología fiable, económica y que ha desarrollado una potencia hasta hace años desconocida. Básicamente presenta dos opciones de implantación: Bucle Arbitrado y Fabric Switched.

En ambos casos se pueden integrar soluciones que eviten utilizar el ancho de banda de la red de área local para la realización de Backups, estas soluciones se integran en la SAN a través de controladoras específicas permitiendo así un altísimo rendimiento así como una operación desatendida.

Entre las ventajas más importantes que puede ofrecer la red de área de almacenamiento están:

- Superior ancho de banda.
- Administración centralizada de todo el almacenamiento.
- Estandarización en recursos de almacenamiento y configuración de servidores.
 - Mayor flexibilidad en la utilización del espacio disponible.
 - Mayor tolerancia de fallas en el sistema.
 - Al tratarse de dispositivos de ámbito corporativo, el rendimiento de los sistemas de almacenamiento puede llegar a ofrecer decenas de miles de I/O (operaciones de lectura/esritura) por segundo.
 - Fácil integración de soluciones de backup de ámbito corporativo de simple administración, alto rendimiento y fiabilidad.
 - Simplificación del proceso de adquisición de ampliaciones al sistema de almacenamiento, pues no estará basado en modelos ni marcas de servidores.
 - Mejora de la facilidad de planificación de las necesidades globales.



Ahora bien, la alta disponibilidad puede conseguirse tanto con alternativas de hardware como de software, En el mercado tecnológico actual, son muchas las empresas que se han dedicado, al desarrollo de herramientas de software orientadas a la alta disponibilidad.

En primer punto se puede mencionar un esquema propuesto por la empresa HP, conocido como Journaled File System (JFS). Esta alternativa orientada al sistema operativo Unix, utiliza una conexión especial que le permite registrar información acerca de los cambios hechos en el file system de metadatos. Dicha conexión permite mejorar la disponibilidad reduciendo a solo unos segundos el tiempo necesario para reiniciar un file system luego que el mismo ha fallado.

Como una alternativa opcional para expandir las funciones de JFS se puede considerar OnlineJFS, quien elimina los downtime planificados asociados con las actividades de mantenimiento de los file system pues actividades como desfragmentación, reorganización y expansión de los file system pueden ser ejecutadas mientras las aplicaciones acceden la data de los mismos. En lo que respecta a las actividades de backup las mismas se ejecutan utilizando la filosofía snapshot, lo que permite que las aplicaciones accedan la data mas reciente mientras que el proceso de backup accede la copia.

Otro esquema de software que igualmente apoya la disponibilidad de los sistemas es la herramienta Transaction Processing Monitors (TPM). Este esquema asegura la disponibilidad en segundos cuando es usado en conjunto con la tecnología clusters, pues se reenvían las transacciones a otro nodo cuando el nodo original presenta alguna falla.

Adicionalmente existen otras alternativas de software, quizás vistas como herramientas para el manejo de sistemas y redes que permiten manejar automatizadamente una red compuesta por diversos tipos de servidores y estaciones trabajo, pero que igualmente están orientadas a la alta disponibilidad.

ARQUITECTURA TECNOLÓGICA PARA LOS SISTEMAS DE ALTA DISPONIBILIDAD.

La definición de la propuesta comienza con el diagnóstico de la situación actual de la empresa objeto de estudio, donde se evalúan entre otros aspectos, las arquitecturas tecnológicas existentes que brindan soluciones de alta disponibilidad, así como los esquemas de implementación de las mismas.



Las debilidades detectadas en el proceso investigativo sirven de base para la propuesta de diseño para sistemas de alta disponibilidad, el cual persigue ofrecer alternativas que aseguren la continuidad operacional de las aplicaciones críticas, o en su defecto minimizar los períodos de inoperabilidad de las mismas.

En particular para este estudio, la organización está equipada con una excelente tecnología, sin embargo existen ciertos problemas en cuanto a la disponibilidad de las aplicaciones críticas, lo cual conlleva a definir reestructuraciones que permitan ofrecer a cabalidad la alta disponibilidad esperada.

En este sentido se presenta la propuesta de solución tecnológica sugerida para el logro de tal fin.

La propuesta planteada es una solución genérica que permite al personal de tecnología de la organización objeto de estudio, la selección de las alternativas en función del costo/beneficio que este a su alcance, sin limitarle a una solución particular de un proveedor determinado.

Para poder ofrecer un esquema alternativo que contemplara estos aspectos, era necesario optar por una tecnología que además permita aprovechar los recursos actuales y por supuesto asegurar la escalabilidad de la nueva arquitectura.

En tal sentido, la propuesta se basa en la implementación de una aplicación de tecnología de Fibre Channel para crear una red propia dedicada exclusivamente a dispositivos de almacenamiento, denominada SAN. (Storage .Area Network).

La arquitectura SAN es una infraestructura dedicada que permite conectar diferentes unidades al sistema de almacenamiento donde se pueden compartir datos, archivos, clips, y otros formatos de información sin necesidad de duplicarlos o moverlos a otros puntos de la red. Creando una unidad autosuficiente que no depende del funcionamiento del servidor.

SAN está orientada específicamente a funciones de almacenamiento, y control del volumen de acceso.

Para tal fin se requiere la adecuación de la infraestructura actual de la empresa objeto de estudio, de la siguiente manera:



- Conectar ambos servidores HPCLUST1 y HPCLUST2 a 1 Gbits a la red de la empresa.
- Conectar ambos servidores E10000 y Sun6500 a 1 Gbits a la red de la empresa.
- Establecer un enlace doble de fibra óptica entre los servidores que pertenecen al mismo cluster a la velocidad de 100 Mbits.
- Adquirir un Switch de Fibre Channel de 32 puertos como parte de esquema de redundancia planteado en la investigación.
- Establecer enlaces dobles entre el Switch de Fibre Channel con el subsistema de almacenamiento centralizado a la velocidad de 100 Mbits.
- Establecer enlaces dobles de 100 Mbits entre el Switch de Fibre Channel con cada uno de los servidores pertenecientes a los cluster evaluados.
- Adquirir una licencia de Parallel Server de Oracle e implementar el esquema de paralelismo para la base de datos de las aplicaciones críticas.
- Migrar todas las bases de datos Oracle que pertenecen a las aplicaciones críticas del negocio de Oracle versión 7.3 a Oracle 8i o Oracle 9i así como el sistema operativo de HP UX 10.20 a HP UX 11i y el sistema operativo Solaris 2.5 a 2.8. La selección del manejador dependerá del tiempo que transcurra entre esta propuesta y el tiempo considerado en la implementación de la misma o también de las mismas necesidades de las aplicaciones críticas.
- Adquirir e implementar la alternativa de solución del proveedor o proveedores que mejor se adapta a los criterios económicos y tecnológicos del personal de tecnología de la organización objeto de estudio, considerando las pautas establecidas para tal fin.



REFERENCIAS BIBLIOGRAFICAS

Compaq. "Compaq anuncia soluciones empresariales de almacenamiento y SAN con capacidad de actualización con tecnología líder en la industria y tolerancia de desastre" .[En línea].

<http://www.clac.compaq.com/news/040699b.html> [30-Mayo-2000]

Computer Associates. "Computer Associates anuncia una solución que confiere una alta disponibilidad a los servidores Windows NT para aplicaciones de misión crítica" .[En línea]

<http://www.cwordl.cl/archivos/ed179/reportaje.html> [18-Mayo-2000]

IBM. "Soluciones SAN" .[En línea]

<http://www.ibm.com./ar/news/almacenamiento.html> [19-Mayo-2001]

SMDATA. "Que es SAN". .[En línea].

<http://www.smdata.com.definiciones.html> [07-Mayo-2001]

Farley, M., &Stearns, T., & Hsu, J., Seguridad e Integridad de datos. España. Osborne/McGraw-Hill.

Hernandez R, Fernandez C, Baptista P, Metodología de la investigación. Mexico. Mc Graw Hill.

HP, Oracle, Cisco (1999). High Availability & Load Balancing Testing using Oracle Application, Cisco Local Director and HPWebQos.

Huidobro, José Manuel. Todo sobre comunicaciones. España. Paraninfo.

Martin, L (1999) Tolerancia a fallos en Windows NT. Windows NT Actual , 1, 70-73

Martin, L (1999). Servidores Nonstop Himalaya para alta disponibilidad. Windows NT Actual, 6, 46-49

Palmer, Michael. Redes informáticas. Seguridad e Integridad de datos. España. Paraninfo.

Weygant, Peter. Clusters for High Availability. Estados Unidos. Prentice Hall.