

IMPLEMENTACIÓN DE UN SERVIDOR CLÚSTER DE ALTA DISPONIBILIDAD BAJO HERRAMIENTAS DE CÓDIGO ABIERTO EN LA UJAP

IMPLEMENTATION OF A SERVANT CLÚSTER OF HIGH AVAILABILITY UNDER TOOLS OF CODE OPENED IN THE UJAP

Ari J. Morales J.

arimoralesjordan@gmail.com

Universidad José Antonio Páez

Marck F. Schmiedeler G.

Universidad José Antonio Páez

Resumen: En el presente trabajo de grado, se desarrollará una solución para combatir los problemas asociados a alta disponibilidad usando servidores en configuración de clúster bajo software libre y así prestar servicios de educación sin distancia, de manera tal que los usuarios, que en este caso serán los alumnos y docentes de la Universidad José Antonio Páez (UJAP), no presenten inconvenientes al momento de usar el sistema y hacer que la navegación en el portal sea lo más grato posible. Adicionalmente este tipo de soluciones se amoldará a estándares que permiten que su uso sea lo más económico posible y de esta forma ajustarse a los presupuestos de dicha casa de estudio. Debido al contexto general que comprende este proyecto, el mismo se enmarca dentro de la modalidad de proyecto factible, el cual servirá para el desarrollo de una propuesta basada en la implementación servidores redundantes o de alta disponibilidad en la UJAP.

Descriptor: Educación sin Distancia, Servidores tipo Clúster, Infraestructura de Tecnologías de Información.

Abstract: In the present work of degree, a solution will be developed to fight the associated problems to high availability being used servants in configuration of clúster under free software and thus to serve of education without distance, of way like the users, that in this case they will be the educational students and of the University Jose Antonio Páez (UJAP), they do not present/display disadvantages at the time of using the system and causing that navigation in the vestibule is most pleasing possible. Additionally, this type of solutions will be conformed to standards that allow that their use is economic possible and of this form to adjust to the budgets of this house of study. Due to the general context that includes/understands east project, the same is framed within the modality of feasible project, which will be used for the development as a proposal based on the implementation superfluous servants or as high availability in the UJAP

Key words: Education without Distance, Servants Clúster type, Infrastructure of Technologies of Information

PLANTEAMIENTO DEL PROBLEMA

Según una publicación de la revista electrónica de la Universidad Central de Venezuela (UCV) editada Fernández L. (1999). "La Educación a distancia ha mostrado mayor efectividad rompiendo las barreras del tiempo y espacio, al ofrecer métodos, técnicas y recursos que hacen más efectivo y flexible el proceso enseñanza-aprendizaje, esto mediante el uso de tecnologías como la radio, la televisión, el vídeo, el audio cinta, los sistemas de informática y el software interactivo. Esta nueva modalidad educativa da validez así a los principios de educación para todos, aprender a aprender, la enseñanza-aprendizaje personalizada, la educación para toda la vida, es decir, la educación permanente o continua."

El fenómeno de la educación sin distancia comienza a cobrar un desarrollo importante desde la utilización del vídeo y muy recientemente la telemática como medios pedagógicos. A medida que las comunicaciones avanzan a escala global, la información adquiere cada vez un papel más importante, dentro del ámbito, tanto empresarial como público.

Así mismo, el acceso a bases de datos vía Web, ha abierto una nueva puerta para el desarrollo de sistemas que se basan en la metodología cliente/servidor. Este tipo de plataforma requiere un hardware con características especiales para que este tipo de sistema sea confiable y de rápido acceso. Para prestar estos servicios no es suficiente un servidor común y corriente, donde éste tiene las características de una computadora común de hogar, puesto que por algún error o falla de cualquier índole, puede dejar de funcionar haciendo que el sistema cuando estén todos los usuarios conectados, interrumpa las tareas en ejecución.

Sucede pues, que una vez puesta en marcha una plataforma de educación sin distancia basada en la web 2.0, debe continuar trabajando durante un período académico determinado. Es decir, una plataforma de educación sin distancia necesariamente debe permanecer en funcionamiento durante las 24 horas del día, los 7 días de la semana y los 365 días del año, lo cual significa que debe prestarse una serie de servicios que soporten la diversidad multimedia y web, de manera continua e ininterrumpible. En otro orden de ideas, es notable el constante crecimiento de los archivos multimedia que se manejan en una plataforma de educación sin distancia, este crecimiento tiende a ser infinito puesto que la población estudiantil también persigue esta tendencia. Inclusive con el pasar del tiempo las aplicaciones se tornan más complejas y de gran tamaño en bytes. Es así como la capacidad de la plataforma de hardware debe ser ampliada constantemente junto a una implantación de políticas a la par del crecimiento, debido a su elevado costo

Sin embargo, tal período puede verse afectado por problemas de mal funcionamiento de hardware en cualquiera de los elementos requeridos para prestar el servicio, los cuales engloban una serie de equipos tanto servidores como dispositivos de red que, en conjunto, forman una plataforma interdependiente. Si por ejemplo, el servidor surge con un inconveniente en su fuente de voltaje eléctrico, inmediatamente se detienen los servicios prestados, y ocurrido el suceso hasta que el administrador de sistemas esté al tanto de la situación y pueda dar respuesta a la falla, definitivamente transcurre el tiempo suficiente como para interrumpir una clase online o alumnos que están subiendo al servidor archivos con asignaciones y trabajos; este error de ejemplo y su posible solución, sea instantánea o prolongada, es totalmente transparente para el usuario de la plataforma, por lo tanto sin saber los motivos de la falla, simplemente dejan de confiar en el sistema.

Evidentemente, este tipo de problemas los presentan los servidores que no cumplen con los requerimientos mínimos para este tipo de aplicación. Al momento en que ocurra una falla de hardware o software los usuarios que se encuentren conectados al sistema perderían su conexión y por consiguiente el trabajo que no ha sido guardado, hasta que el error o falla sea corregida.

Es así como la capacidad de la plataforma de hardware debe ser ampliada constantemente junto a una implantación de políticas a la par del crecimiento, debido a su elevado costo. Sucede pues, que cuando llega el momento de realizar estas tareas de ampliación, es necesario efectuarlas sin que ninguno de los dispositivos del conjunto de la plataforma deje de prestar los servicios. En caso de que la plataforma alcance su máxima capacidad, el sistema de educación sin distancia estaría sujeto a detenerse para realizar una actualización o mejora, perdiendo así la continuidad del servicio y poniendo en riesgo la confiabilidad de la plataforma.

Visto de esta forma, referente al costo, implementar soluciones de alta fiabilidad de cómputo implica una gran inversión inicial, la cual para el caso de la educación sin distancia no representa un retorno

de inversión considerable a corto plazo y por tanto los inversionistas de capital difieren al momento de invertir en un proyecto de educación sin distancia en el caso de que se posea el capital. Las grandes empresas de computación paralela y distribuida, proporcionan soluciones de servidores pre configurado. Sin embargo, independientemente a los problemas de costo ya mencionados, se encuentra una desventaja muy importante y es que al adquirir uno de estos servidores bajo un renombre, con el pasar del tiempo existe el riesgo de que el fabricante deje de elaborar repuestos y mejoras para el modelo adquirido, colocando una barrera limitante de alto costo para el mantenimiento y actualización de la plataforma.

En la universidad José Antonio Páez (UJAP), se cuenta actualmente, con una plataforma académica la cual se utiliza constantemente en el quehacer educativo. La misma presenta algunos inconvenientes que afectan al usuario, lo cual genera malestar e interrumpe el buen desarrollo de las actividades pedagógicas, lo que crea la necesidad de evaluar sus requerimientos a fin de instalar y/o mejorar la plataforma para el uso de sistemas de educación sin distancia que satisfaga las características de interacción demandadas por el usuario.

FORMULACIÓN DEL PROBLEMA

¿Haciendo una evaluación de los requerimientos necesarios para la plataforma, se podrán solventar los inconvenientes asociados a la necesidad de alta disponibilidad de servicios de educación sin distancia de la UJAP?

OBJETIVOS GENERAL

- Implementar un servidor tipo clúster, bajo el uso de herramientas de software libre en la UJAP, para solucionar los problemas asociados a la disponibilidad de los servicios que ofrece la educación sin distancia y fiabilidad en el uso de la plataforma educativa Específicos
- Diagnosticar los problemas que surgen en una plataforma de educación sin distancia a través de cuestionarios aplicados a los usuarios
- Determinar los requerimientos de hardware, software y conectividad a la plataforma de servidor para un aproximado número de estudiantes y docentes que se conectan a la plataforma de forma concurrente.
- Proponer una solución de servidor clúster bajo software libre para el establecimiento de la plataforma de educación sin distancia.
- Evaluar la propuesta de solución en función de distintas pruebas bajo situaciones de riesgo y la factibilidad de la misma.

JUSTIFICACIÓN DE LA INVESTIGACIÓN

Debido al avance de las aplicaciones vía Web y a la llegada de la Web 2.0, la interacción entre el usuario y el sistema vía Internet obligan a los servidores a aumentar su procesamiento de datos, disponibilidad y almacenamiento, haciéndolos mejores equipos que los servidores de algunos años atrás. La Universidad José Antonio Páez, (UJAP) ha avanzado conjuntamente con esta tecnología y pretende seguir haciéndolo, puesto que está creando nuevas extensiones de estudio, lo que genera la necesidad de incrementar los recursos y mejorar la plataforma que la mencionada institución posee actualmente, con el fin de optimizar su uso, a través de un sistema que pueda ser constantemente actualizado para la llegada de nuevas tecnologías y extensiones que la UJAP así lo requiera.

Cabe destacar que equipos con estas capacidades disponibles en el mercado, diseñadas e implementadas por las grandes casas de hardware, tienen un costo exorbitante y que solo pueden ser costeados por entes gubernamentales o por empresas con un gran capital. La UJAP con la ambición

del proyecto de educación sin distancia, no dispone de un capital de tan elevado costo para estos fines. Es por eso que con esta investigación de servidores clúster se pueden obtener resultados que cumplan con las expectativas propuestas, a un costo notablemente por debajo del sugerido por las grandes empresas de hardware, y con equipos convencionales de cómputo, pudiéndose utilizar de manera opcional equipos usados. Esto le permitirá a la universidad ser más competitiva en el mercado educativo, ya que podrá ofrecer mejores servicios, con tecnología de punta.

DISEÑO Y TIPO DE INVESTIGACIÓN

De acuerdo a los objetivos de este estudio, el mismo se enmarca dentro de la modalidad de proyecto factible. Según la Universidad Pedagógica Experimental Libertador (UPEL, 2003) un proyecto factible “consiste en la elaboración de una propuesta o requerimiento de solución de una necesidad para una organización o grupo social”. En este sentido, este estudio sigue los criterios de un proyecto factible, pues el trabajo está orientado a responder las necesidades del problema detectado en los grupos sociales, al implementar un clúster que sea viable, a fin de evitar que el sistema de educación sin distancia de la Universidad José Antonio Páez, de un mal servicio. En atención a esta modalidad de investigación, el estudio se desarrolló en cuatro fases. Se partió de un diseño experimental, fundamentado en una investigación cuasiexperimental-documental, lo cual está sustentada en la definición de acuerdo con Sierra A. (2008), como “aquella donde se toman grupos que ya están integrados por lo que las unidades de análisis no se asignan al azar ni por apareamiento aleatorio. La validez interna se ve afectada por el fenómeno de selección, la regresión estadística y el proceso de maduración”.

Por otra parte, Suárez N. (2007), señala que la investigación documental “es una concentración de informaciones para realizar con ella rápidamente, una difusión segura. Permite hacer provisión de materiales y de elementos intelectuales a quienquiera que ejerza una labor creadora, aprovechándose de la experiencia del pasado y de los últimos datos del progreso”.

POBLACIÓN Y MUESTRA

De acuerdo con Balestrini (1997), desde el punto de vista Estadístico, la población o universo puede estar referido a cualquier conjunto de elementos de los cuales se pretenden indagar y conocer sus características, o una de ellas, y para el cual serán válidas las conclusiones obtenidas en la investigación. Es el conjunto finito o infinito de personas, casos o elementos que presentan características comunes.

Considerándose el eje central de la investigación referido como la implementación de este servidor para evitar la sobrecarga en el sistema y mejorar el servicio de educación sin distancia, la población en esta investigación está representada por los estudiantes de la escuela de Ingeniería en Computación de la UJAP que utilizan o podrían utilizar el sistema de educación sin distancia.

Con el propósito de determinar cuáles son los sujetos que integraron la muestra, se planteó lo dicho por Sabino (1997), “en función de que la población total puede disgregarse en sub-conjuntos menores homogéneos internamente, pero heterogéneos entre sí, es decir, fragmentar la población en categorías de unidades, diferenciándolas de acuerdo a alguna variable de interés para la investigación”, por lo cual se tiene una muestra representativa, seleccionando así a los estudiantes del cuarto semestre en adelante de la carrera de Ingeniería en Computación. Basado en lo anterior, se realizó una selección de sujeción aleatoria, por ser un campo muy amplio, escogiendo un número representativo de personas implicadas en este problema, a los cuales se le aplicó un cuestionario, a fin de conocer los requerimientos del usuario.

FASES METODOLÓGICAS

Con el fin de dar respuesta a los objetivos de la presente investigación, la metodología usada para la elaboración del proyecto, fue una creación híbrida de los autores entre las metodologías RUP y XP unidas, y adaptada a los objetivos de la investigación. De allí que este trabajo de investigación se estructuró de la siguiente manera:

- **Fase I. Conceptualización:** Durante esta fase se realizó el levantamiento de la información de datos para los requerimientos del proyecto, la visión de adonde se quiere llegar, se hizo una propuesta de un modelo preliminar para solventar el problema y un plan de trabajo para el desarrollo del proyecto.
- **Fase II. Diagnóstico:** En esta fase se procedió a hacer un previo estudio de las posibles soluciones que se podrán implantar para evitar los problemas de alta disponibilidad que presentaría la Universidad al momento de instalar un sistema de educación sin distancia.
- **Fase III. Propuesta:** Durante esta fase se establecen los métodos para la instalación y configuración de la plataforma, siguiendo las recomendaciones que propone la filosofía de trabajo ITIL, la cual propone hacer un trabajo bajo ciertas reglas para lograr un proyecto bien esquematizado con toda práctica documentada y así dar base a toda la experiencia adquirida con el tiempo en determinada actividad. Por otra parte, son soportadas bajo esquemas organizacionales complejos bien definidos que se apoyan en herramientas de evaluación e implementación.
- **Fase IV. Evaluación:** Ya en esta fase se realizaron las pruebas funcionales de la puesta en marcha del servidor, para verificar de esta forma la alta disponibilidad del mismo para los sistemas de educación sin distancia, detectando al momento de la prueba la existencia de fallos o errores característicos que no deben ocurrir en los clústeres de alta disponibilidad. Así mismo la ejecución de escenarios típicos de fallas de hardware controladas.

RESULTADOS

FASE I. CONCEPTUALIZACIÓN. LEVANTAMIENTO DE INFORMACIÓN.

La obtención de la información necesaria para determinar los recursos físicos y conceptuales refieren al uso parcial del instrumento de encuesta, bajo la herramienta de código abierto y de uso gratuito LimeSurvey la cual arroja una serie de resultados cuantitativos, relacionados entre sí; en cuanto a la usabilidad de las plataformas de educación sin distancia e inclusive de cuanto los usuarios conocen de estas herramientas, así como también de la repercusión que tiene la disponibilidad de la plataforma en el desempeño y confianza de los usuarios en la misma.

Por otra parte, se hizo una entrevista no estructurada que genera notablemente una gran cantidad de ideas, aportes y lineamientos para el correcto entendimiento del problema y de cómo aportar soluciones que se encuentren dentro de las posibilidades descritas en las limitaciones del proyecto.

Entonces resulta que, luego de tomar en consideración los aspectos relevantes del problema gracias a los recursos anteriores, se continúa con la investigación vía web; cabe destacar que de una u otra forma las tecnologías de alta disponibilidad son relativamente nuevas y las bibliografías impresas debido al proceso de traducción, están desfasadas en cuanto a las aplicaciones para RIA y la web 2.0.

REQUERIMIENTOS BÁSICOS.

Existen unos requerimientos básicos que se necesita para la implementación del servidor clúster de alta disponibilidad, además, basándose en los alcances y limitaciones, dichos requerimientos se denotan de la siguiente manera:

1. El tipo de servidor para la implementación debe ser un Clúster de Alta disponibilidad.

2. Las herramientas usadas para la implementación del servidor clúster de alta disponibilidad deben ser herramientas con licencia GPL o software libre.

VISIÓN DEL PROYECTO.

La implementación de este proyecto permanece dentro de cierto criterio, relacionado directamente con el problema y los requerimientos planteados. En respuesta a estos requerimientos, se encuentra como candidato un clúster de alta disponibilidad, dentro de la cual subyace el concepto genérico de disponibilidad acuñado por Omar Crespo "Disponibilidad de un componente o servicio es hacer que este componente se encuentre operando permanentemente durante un periodo de tiempo deseado. La disponibilidad puede ser medida en un porcentaje de operación, por ejemplo, los "cinco nueves" (99.999%).", se espera entonces que este proyecto de paso al despliegue de una plataforma que a través de todos y cada uno de sus componentes soporte el lema de los "cinco nueves".

En síntesis, es necesario señalar algunos requerimientos mínimos para garantizar que el sistema visto como un todo, entregue a lo posible su máxima capacidad de disponibilidad. Esto se puede llevar a cabo haciendo que cada uno de los componentes, tanto de software como de hardware, aporte un granito de arena que a lo sumo represente el 100% de la alta disponibilidad del sistema. Desde el punto de vista del software, este se encarga de proveer fiabilidad, detectando fallas imprevistas en el sistema y permitiendo que el mismo se recupere de estas. Por otra parte, con el hardware se trata de evitar un único punto de fallo.

En este orden de ideas se debe entonces considerar los siguientes requerimientos mínimos:

- **Redundancia en el suministro eléctrico:** es necesaria la colocación de dispositivos de alimentación energética ininterrumpida (UPS) con al menos una duración de respaldo de 1 hora para el total de consumo en vatios por hora, bien sea de cada nodo o del total de los mismos. Por otra parte, se debe también proveer a cada nodo de fuentes de poder redundante e intercambiable en caliente, de manera que cuando una de las fuentes de poder del equipo falle, sea por horas de uso o defecto de componente; no represente una ruptura de disponibilidad a nivel de hardware.
- **Almacenamiento redundante:** Cada nodo debe contar con sistemas de almacenamiento redundante para asegurar la confrontación de fallas mediante arreglos de disco, que en este caso lo que haría falta es: Una tarjeta controladora RAID que soporte los esquemas RAID-0 para expansión y RAID-1 para el espejo de datos. Un mínimo de dos discos duros SCSI de 1TB cada uno; uno como principal y el otro como espejo.
- **Dispositivos de red:** El ruteo de paquetes de red hacia y desde el servidor clúster debe contar con dispositivos de red conectados a una fuente energética ininterrumpida, con el fin de evitar interrupciones por fallas eléctricas, ya sean momentáneas o permanentes; esto para darle integridad a la idea de que los servidores ya poseen suplidores UPS. Adicional al respaldo energético también es necesario asegurar las fallas internas de un dispositivo de ruteo, esto es posible al utilizar soluciones stateful switchover (SSO) las cuales permiten que un procesador pasivo esté en espera para tomar el control de una ruta fallida en el procesador activo mientras mantiene la conectividad.

MODELO PRELIMINAR.

Basándose en el resultado, se aprovechan las bondades que brinda un motor de búsqueda web, referente a debates y planteamientos de las distintas aplicaciones disponibles, y de cómo conjugar herramientas para dar respuesta a problemas de alta disponibilidad en general, se converge con definiciones y despliegues de modalidades de trabajo para brindar servicios de alta disponibilidad, tales definiciones pueden ser agrupadas de la manera expuesta a continuación:

SISTEMAS OPERATIVOS.

Existe una innumerable cantidad de variables a considerar para la instalación de un sistema operativo, sin embargo, para efectos de esta investigación se consideran los de mayor impacto en cuanto a los servicios a soportar y el personal de IT que manipularán la plataforma, a nivel de sistema operativo, hardware y middleware, claro está. Tomando en cuenta estas consideraciones, se enfoca la selección principalmente en tres aspectos clave: Manipulado, mantenido y soportado por una gran comunidad. Según DesktopLinux.com Enterprise Technology News and Reviews, soportado por eweek.com. Empresa que realizó la encuesta: 2007 Desktop Linux Market survey, que en español: Encuesta de mercado de escritorios usados en Linux 2007, y uno de sus resultados fue que la distribución Ubuntu alcanzó una votación del 30% en relación a las demás. Tomando en cuenta que el tamaño de la población de los encuestados es de 38.500 usuarios, queda explícitamente denotado que posee una gran diferencia de uso con respecto a su siguiente oponente OpenSuse la cual es de 10,4%. Definitivamente el hecho de que este sistema operativo permita la opción de seleccionar el sistema de ventanas también lo impulsa como buen candidato para ser utilizado en el presente proyecto de investigación para solucionar los problemas.

HERRAMIENTAS INCORPORADAS EN EL KERNEL PARA SOPORTE DE ALTA DISPONIBILIDAD.

Es de suprema relevancia, el hecho de que un sistema operativo posea incorporado en su kernel ciertos paquetes opcionales, bien sea de propósito back-end o de propósito front-end; debido a que si se requiere utilizar un paquete que no se encuentre incorporado será necesaria la instalación de tal paquete para luego compilar el kernel con la nueva aplicación instalada; esta tarea puede tornarse engorrosa y un tanto delicada puesto a que se encuentra sujeto a algunos colapsos de seguridad o integridad, si el administrador IT que la ejecuta no es un experto en el área. Además, es importante que para aplicaciones de alta disponibilidad el sistema operativo seleccionado, debe tener las cualidades de diseño, fiabilidad y escalabilidad para que soporte de manera natural tales aplicaciones.

Ahora bien, según Didio L. de Yankee Group en su análisis de la encuesta fiabilidad de SO de servidor 2008 donde menciona "Ubuntu, que aparece en encuestas de fiabilidad global del Yankee Group, por primera vez este año, también ha entrado muy fuerte teniendo en cuenta que es un sistema operativo de código abierto. Asimismo, Aproximadamente el 22% de los encuestados ejecutan al menos un servidor de Ubuntu en sus sitios, y ha demostrado ser altamente confiable, con 1.1 horas por servidor de tiempo de caída al año".

REQUERIMIENTOS MÍNIMOS DE UBUNTU SERVER:

Ubuntu Server está pensado para funcionar en cualquier procesador Intel o AMD x86, AMD_64, EM_64T. Se requiere un mínimo de 192 MB de RAM y 1 GB de espacio en disco. El análisis anterior nos proporciona un buen panorama para entender que Ubuntu server es el candidato número uno para implementar la solución.

ARQUITECTURA DE RED.

La arquitectura de red es descrita por los manuales y configuraciones de las tecnologías usadas para la interconectividad por Robertson A. (2008) de cada uno de los nodos y para establecer conexión con las puertas de enlace por donde llegan las peticiones a servicios del clúster, re-direccionando las peticiones al nodo que se encuentre marcado como activo en ese instante de cualquier petición. Es importante denotar, que cada nodo posee una conexión directa (física) a la topología tipo estrella de la subred donde se encuentra el clúster para el acceso a las puertas de enlace. Sabiendo que cada nodo debe contar con al menos dos interfaces de conexión NIC, una para la conexión a la estrella de la

subred comentada y la otra para establecer la conexión mediante un cable UTP con conectores RJ-45 y de una distancia no mayor a 100 metros, bajo el estándar TIA/EIA-568-B donde un extremo debe poseer un pin out de la norma T568B y el otro extremo T568A. Esto con el objetivo de formar una topología punto a punto para habilitar al gestor de recursos la correcta administración y monitoreo de cada nodo y al sistema de archivos la correcta sincronización de datos.

SISTEMA DE ARCHIVOS.

Luego de explorar una serie de ejemplares, sobresale notablemente una tecnología de sistema de archivos que posee un nivel de complejidad bajo y un rendimiento y versatilidad elevado llamado DRBD se refiere a software tanto en el (módulo del kernel y herramientas asociadas espacio de usuario), y también para dispositivos de bloque lógico específico gestionado por el software.

GESTOR DE BASES DE DATOS.

Ahora bien, al realizar una extensa investigación de los motores de bases de bases de datos para el uso de sistema de educación sin distancia, resaltó el motor de Mysql debido a que según un artículo por Schumacher R. et al (2008) señala que es un sistema de gestión de base de datos relacional, multihilo y multiusuario con más de seis millones de instalaciones; desde enero de 2008, es una subsidiaria de Sun 24

Microsystems y ésta a su vez de Oracle Corporation, desde abril de 2009 desarrolla Mysql como software libre en un esquema de licenciamiento dual.

Por un lado, se ofrece bajo una licencia GPL para cualquier uso compatible con esta licencia, pero para aquellas empresas que quieran incorporarlo en productos privativos deben comprar a la empresa una licencia específica que les permita este uso. Está desarrollado en su mayor parte en ANSI C.

Asimismo, MySQL es muy utilizado en aplicaciones web, como Drupal o phpBB, en plataformas (Linux/Windows-Apache-MySQL-PHP/Perl/Python), y por herramientas de seguimiento de errores como Bugzilla. También, su popularidad como aplicación web está muy ligada a PHP, que a menudo aparece en combinación con MySQL. Además, MySQL es una base de datos muy rápida en la lectura cuando utiliza el motor no transaccional MyISAM. Ahora, en aplicaciones web donde hay baja concurrencia en la modificación de datos y en cambio el entorno es intensivo en lectura de datos, lo que hace a MySQL ideal para este tipo de aplicaciones. Sea cual sea el entorno en el que va a utilizar MySQL, es importante adelantar el monitoreo sobre el desempeño para detectar y corregir errores tanto de SQL como de programación.

Gracias a las cualidades antes mencionadas de Mysql, es posible modificar sus funcionamientos y para este caso su estructura funcional, dando pie a la modificación de su código de configuración y de las direcciones donde sus datos son almacenados. Por otra parte, este manejador de bases de datos posee la cualidad de expandir su configuración a una plataforma distribuida, dejando una gran puerta abierta para la escalabilidad del clúster. Por consiguiente, este motor de bases de datos es ideal para este proyecto y por supuesto es libre.

SERVICIOS WEB.

Siguiendo en búsqueda de los servicios que se necesitan para poder implantar un servidor web y apoyando la alta calidad e integridad que nos dan los proyectos GNU y gracias a un artículo en Hartill R. (2004), se encontró el servidor web Apache el 25 cual es un servidor web HTTP de código abierto multiplataforma, que implementa el protocolo HTTP/1.11 y la noción de sitio virtual. Asimismo, Este servidor web es redistribuido como parte de varios paquetes propietarios de software, incluyendo la

base de datos Oracle y el IBM WebSphere application server. También, es soportado de alguna manera por Borland en las herramientas de desarrollo Kylix y Delphi. Adicionalmente, apache es incluido con Novell NetWare 6.5, donde es el servidor web por defecto, y en muchas distribuciones Linux. Ahora bien, Apache es usado para muchas otras tareas donde el contenido necesita ser puesto a disposición en una forma segura y confiable. Un ejemplo es al momento de compartir archivos desde una computadora personal hacia Internet, o bien para este proyecto la facilidad de modificar su estructura interna y cambiar su configuración a la más óptima deseada. En la imagen extraída de Netcraft.com (2010) se grafica la preferencia del uso del servidor apache sobre los demás servidores con un 51%, el cual ha tenido más actividad en la internet en los últimos años Gestor de recursos.

En búsqueda de un manejador de recursos que coloque los servicios deseados en disposición de una programación deseada y cumpliendo con el objetivo de utilizar software con licencias GPL, sobresalió como mejor opción el sistema de alta disponibilidad Heartbeat el cual según un artículo en la web por Alonso J. (2009) Señala que, es una aplicación de libre distribución disponible para los sistemas operativos de tipo Linux, FreeBSD y Solaris, que permite configurar sistemas de Alta Disponibilidad ofreciendo procesos de comunicación y monitorización de los nodos que conforman el clúster.

Asimismo, Heartbeat se encuentra dentro del proyecto Linux-HA, el cual mantiene un conjunto de bloques de construcción para sistemas de clúster de alta disponibilidad, incluyendo un grupo de mensajería de capa, un gran número de 26 agentes de recursos para una variedad de aplicaciones, y una colección de fontanería y kit de herramientas de informes de error.

Fase II. Diagnóstico

Seguidamente se muestran y analizan los resultados que arroja el estudio realizado a los estudiantes de cuarto semestre en adelante de ingeniería en computación de la UJAP, en relación a tres aspectos principales: la identificación del encuestado, el reconocimiento del conocimiento respecto al tema del trabajo de grado, y la experiencia de los encuestados con las plataformas utilizadas. Asimismo, analizando los resultados del levantamiento de información, se llegaron a las siguientes conclusiones según las secciones de la encuesta realizada que se mostrarán a continuación:

Con respecto a la sección de Identificación del Encuestado, se resaltó notablemente que un 33% de los encuestados tiene 25 o más años de edad, además un 74% son Bachilleres de los cuales un 23% tienen un título de técnico superior en informática, computación o sistemas, lo cual es un porcentaje significativo para brindar datos más acertados debido a su nivel académico. Continuamente, y observando que una moderada cantidad de encuestados están viendo materias de entre el 6to y 7mo semestre de la carrera de Ing. en Computación, implicando que cursan o han cursado materias como; Comunicación de datos, sistemas de bases de datos, y sistemas de operación. Brindando mejores respuestas y datos confiables en la recolección de datos.

Conjuntamente, la sección de Reconocimiento del Conocimiento, proporcionó datos importantes y positivos para que el diagnóstico de las posibles soluciones planteadas, sean lo más exactas posibles al momento de implantar alguna solución de servicios de alta disponibilidad, dichos datos arrojaron que entre un 40% a un 66% de los alumnos de la escuela de Ing. en Computación saben lo que se refiere a los sistema de educación sin distancia, lo cual es ideal ya que se tiene claro que dichos sistemas deben tener una alta disponibilidad debido a su característica metodología de trabajo, asimismo, se obtuvo que entre un 55% a un 77% de los encuestados saben a lo que se refiere el termino Clúster, aportando datos de mayor confianza a la encuesta

Ahora bien, los resultados de la sección de Experiencia del Encuestado, arrojaron que un 83% de los encuestados han usado la plataforma Acrópolis de la UJAP y además, que un 39% de estos la usan con poca frecuencia, al lado de ello, resalta que un 50% de los encuestados han presentado fallas con

la plataforma, si bien es cierto, entonces los alumnos no tomarían los cursos que se dictan por dicha plataforma. De esta manera se comprueba la hipótesis de que los alumnos no están motivados a ingresar al sistema y completar correctamente el curso, además que los usuarios usan con poca frecuencia, debido a que un 71% de los encuestados señalan que el sistema está fuera de línea con regularidad. De esta manera un 58% de los encuestados señalan que les gustaría ingresar al sistema entre las 6pm y 12m, debido a que la mayoría trabaja de 7am a 6 pm. Gracias a estos resultados se presenta un modelo de servidor tipo clúster el cual esté disponible las 24 horas del día, los 7 días de la semana y los 365 días del año para cubrir la demanda de usuarios que accederían al sistema de la UJAP

FASE III. PROPUESTA

Ahora se presentarán a continuación los pasos que se siguieron para configurar y poner en marcha los servidores desde el inicio. Asimismo, cumpliendo con las pautas que propone el Servicio de Soporte de Manejo de Configuraciones, básicamente el objetivo de este servicio es proveer con información real y actualizada lo que se tiene que configurar e instalado en cada nodo.

Entonces, se logró crear una plantilla de Solicitud de soporte de Configuración parecido al que postula la filosofía de trabajo ITIL, a diferencia que este fue creado para ajustarse a las necesidades del proyecto, como por ejemplo la plantilla llenada cuando se configuro Heartbeat y Apache Solicitud de la Configuración

Continuando con la fase de la propuesta, cada servidor o nodo debe tener ciertas configuraciones para que el servidor sea un clúster de alta disponibilidad y de esta forma preste sus servicios de tal manera que el usuario final no note la diferencia entre un servidor normal a un servidor tipo clúster. Para ello se han especificado ciertos puntos críticos para la instalación y configuración optima de los nodos dentro del clúster, los cuales se mencionarán a continuación.

INSTALACIÓN DEL S.O. UBUNTU SERVER.

Como se ha mencionado anteriormente el sistema operativo base instalado fue Ubuntu Server 9.04 para reducir el costo total por licencia con el fin, de que el clúster sea robusto y sea lo más estable posible, brindando la alta disponibilidad que es lo que a fin de cuentas es uno de los objetivos de la investigación. Sucede pues que, se realizaron los siguientes pasos:

- 1.- Descargamos el S.O. desde la página oficial de Ubuntu www.ubuntu.com, descargaremos una imagen de un disco con un peso de 600 Mb, con la extensión .iso.
- 2.-Luego de lo primero se grabará la imagen en un CD utilizando el programa K3b que nos facilita la gestión de dicha imagen.
- 3.-Continuando con la idea anterior procedemos configurar el BIOS del servidor, para que comience a leer el CD primero antes de la unidad de disco duro.
- 4.-Seguidamente se procede a colocar el CD en la bandeja de la unidad de CD-Rom, e iniciamos el instalador del sistema operativo.

Configuración del S.O. Ubuntu Server.

Una vez instalado el sistema operativo, se hace necesaria la instalación de algunas herramientas que nos ayudaran a configurar fácilmente los nodos, como por ejemplo un escritorio mínimo para facilitar el

manejo de archivos y exploración de las aplicaciones web, para ello necesitamos escribir en la consola lo siguiente:

1.-Para instalar Gnome escribimos lo siguiente: `sudo apt-get install x-window-system-core gnome-core`, pedirá escribir la contraseña del administrador que escribimos al instalar el S.O; luego de instalarlo lo ejecutamos con `startx`. Todo lo que incluye el menú es: "Accesorios" (con editor de texto y terminal) e "Internet" (con Firefox) y nada más.

2.-Ya que es una instalación mínima, queda todo en inglés. Para instalar los paquetes de idiomas en español escribimos: `sudo apt-get install language-pack-es language-pack-es-base language-pack-gnome-es language-pack-gnome-es-base language-selector language-support-es`.

3.-Luego necesitamos una Herramientas de red, para lo que escribimos `sudo apt-get install gnome-system-tools gnome-nettool`.

Con estos pasos ya tenemos instalado y configurado un sistema operativo estable y con el mínimo de espacio y recursos ocupado, donde tenemos un escritorio que solo se mostrará si ejecutamos `startx`, sino al iniciar solo nos mostrará la terminal, Ahora se procederá a instalar y configurar los servicios para configurar los nodos en paralelo.

Instalación de Tarjetas de Red Adicionales.

Como anteriormente se mencionó, se instalaron tarjetas de red Ethernet que utilizan conectores RJ-45 (10/100) los cuales utilizaremos para conectar directamente a los nodos entre sí con un cable UTP Crossover, el cual se registrara en el S.O. como Eth1 y la tarjeta integrada se registró como Eth0, esto para saber que tarjeta es la que usaremos para la replicación de datos y para prestar servicio de alta disponibilidad.

CONFIGURACIÓN DE LAS DIRECCIONES IP.

Continuando con la configuración, ahora se establecen las 3 direcciones ip que se usarán para que los nodos funcionen en paralelo entre ellos respectivamente, para nuestro clúster utilizaremos las siguientes direcciones ip:

- Para el nodo 1 utilizaremos la dirección ip para Eth0 192.168.1.10 y para Eth1 192.168.1.11.
- Para el nodo 2 utilizaremos la dirección ip para Eth0 192.168.1.12 y para Eth1 192.168.1.13.
- Y para el servicio de heartbeat utilizaremos la dirección ip 192.168.1.14.

Si es el nodo 2 solo se cambia la dirección ip en la línea que comienza con `address` por las correspondiente al nodo 2 además, luego reiniciamos el servicio de red para establecer las direcciones ip con el comando `sudo service networking restart`.

INSTALACIÓN DE DRBD.

Continuando con la configuración de los nodos y como ya se sabe para que se utilice el servicio de DRBD, Ahora vamos a instalarlo, para ello nada más necesitamos escribir lo siguiente en la consola: `sudo apt-get install drbd8-utils`

Con esto ya tendremos instalado el servicio de DRBD instalado en los servidores, ahora procederemos a configurar el servicio.

CONFIGURACIÓN DE DRBD.

Para la configuración necesitamos realizar los siguientes pasos:

1. Editar las rutas `/etc/drbd.conf` deben ser iguales para los nodos
2. Asegurarse de que las particiones no tengan formato. Las particiones deben ser del mismo tamaño en ambos servidores.
3. Sabiendo que `r0` es el nombre del recurso en `drbd.conf`, debemos crear el recurso y para esto, se escribe en consola lo siguiente: `sudo drbdadm create-md r0`
4. Aplicarle formato a la partición de drbd, para ello se escribe en la consola: `mkfs.ext3 /dev/drbd0 41`
5. Ahora se crea una carpeta en `/media/*` donde será montada la unidad.
6. Luego para probar hacemos un montaje opcional. para ello se escribe en consola: `sudo drbdadm -- --overwrite-data-of-peer primary r0 mount -t ext3 /dev/drbd0 /media/*` donde* representa la carpeta creada en paso 5.

Una vez montada comienza el proceso de sincronización que puede verificarse escribiendo en consola: `cat /proc/drbd`

INSTALACIÓN DE MYSQL V5.

Continuando con la instalación, ahora debemos instalar un sistema de bases de datos para poder tener sistemas dinámicos web, para este clúster HA se instalará Mysql versión 5 el cual nos permite tener un rendimiento óptimo en el clúster, para instalarlo necesitamos escribir en consola lo siguiente: `sudo apt-get install mysql-server mysql-client` Configuración de Mysql v5.

Ahora es necesario configurar MySQL para que escriba datos en la partición drbd y así la base de datos se replique en los dos nodos, para ello necesitamos realizar ciertos pasos que se mencionan a continuación:

1. Se tiene que editar archivo ubicado en `/etc/mysql/my.cnf` y buscar la línea `datadir=/var/lib/mysql` y sustituirla por `datadir=/media/*/mysql`.
2. Luego se tiene que editar el archivo ubicado en `/etc/apparmor.d/usr.sbin.mysql` y buscar las líneas que poseen `/var/lib/mysql/ **` y sustituirla por `/media/*/mysql`.
3. Asimismo, se tiene que modificar archivo ubicado en `/etc/mysql/debian.cnf` y cambiar la línea de contraseña en ambos nodos por una que sea igual ya que son distintas.
4. A continuación, se debe entrar a la base de datos y correr los siguientes scripts:

INSTALACIÓN DE APACHE2.

Continuando con los pasos a seguir para la configuración del clúster, ahora necesitamos un servicio páginas web para poder presentar los sistemas vía internet o intranet, para ello utilizaremos el servidor web apache y un compilador para poder hacer conexiones con bases de datos php5, el cual se instala escribiendo en consola: `sudo aptitude install apache2 php5`

CONFIGURACIÓN DE APACHE2.

Ahora hay que especificarle a apache, la ubicación del directorio donde tendrá la carpeta raíz, Instalación de Heartbeat.

Ahora como ya se sabe, se tiene que instalar el servicio que pondrá a los servidores en alta disponibilidad, para ello se escribe en la consola: `sudo apt-get install heartbeat` Configuración de Heartbeat.

La configuración de este servicio es crítica ya que si no se entiende la arquitectura o la funcionalidad del clúster este mismo no prestaría el servicio de alta disponibilidad de forma transparente al usuario final. Para una óptima configuración se han diseñado una serie de pasos que se mencionaran a continuación:

1. Crear y editar los siguientes archivos, esto en `/etc/ha.d`:
 - a) `auhkeys`, donde se especificará el cifrado de la conexión entre los nodos, este archivo debe tener un permiso de 600.
 - b) `ha.cf` en el cual establecemos la configuración de heartbeat.
 - c) `haresource`, donde especificamos los servicios que heartbeat administrara.
2. El archivo `auhkeys` debe ser igual en los dos nodos,
3. Asimismo, el archivo `ha.cf` debe contener:

```
# Archivo donde se guardan los mensajes del depurador
# Archivos donde se guardan otros logs
# El tiempo se especifica en milisegundo, o sea
# Especificamos que servidores están dentro del clúster
```
4. El archivo `haresource` Aquí se le dice a heartbeat quien va hacer el nodo principal, además de especificar los servicios que va arrancar y las particiones que va a montar para que haya un servidor de alta redundancia.
5. Luego de esto debemos reiniciar los servicios que configuramos

Ya para resumir, los pasos que se necesitan para la configuración del cluster de alta disponibilidad son:

1. Instalar Ubuntu Server 9.04
2. Instalar y configurar las tarjetas de red de cada nodo.
3. Instalar y configurar DRBD. Creando la partición de replicación de datos.
4. Instalar y configurar Apache2 de tal forma de que su directorio de acceso público se encuentre dentro de la partición creada en el paso 3.
5. Instalar y configurar Mysql 5.0 y de la misma manera que Apache2 direccionar la carpeta de datos hacia la partición creada en el paso 3.
6. Instalar y configurar heartbeat para que administre los servicios de DRBD, Apache2 y Mysql 5.0.
7. Hacer las pruebas respectivas para esta configuración y en caso de que surja algún inconveniente referirse al manual rápido de recuperación del clúster

FASE IV EVALUACIÓN.

PRUEBA DE TRANSFERENCIA DE NODO.

La prueba de transferencia de nodo se refiere a la comprobación del intercambio de modos entre los nodos, en cuanto uno o varios servicios del nodo principal (activo) se detengan y todos los servicios incluyendo puntos de montaje en disco del sistema de archivos, son transferidos al nodo que para ese instante de tiempo se encontraba en modo secundario (pasivo). Se muestra a continuación una serie de pasos para verificar el comportamiento de una transferencia provocada manualmente:

Se introduce el comando que permite verificar directamente el estado del nodo, se observa que es el terminal del nodo 1. Se observa que `ro`: Primario/Secundario por tanto a este punto es el nodo 1 el que está activo Luego se procede a detener los servicios manualmente ejecutándose en el nodo 1. Se

verifica de igual forma que el nodo 1 haya pasado a ser pasivo. Se tiene que ro: Secundario/Primario, efectivamente cambio de modo el nodo 1, ahora se verifica el nodo 2, que haya pasado de pasivo a activo. Y el resultado es afirmativo, el nodo 2 está en modo primario. Para cerciorarse de que los servicios han sido transferidos satisfactoriamente entre nodos se pedirá un estado de servicios en el nodo 1 donde uno de los servicios que interesa (Apache2) debe estar inactivo

PRUEBA DE REPLICACIÓN DE ARCHIVOS.

Para la realización de esta prueba se ha utilizado una transferencia de archivos simple desde la red LAN, tomando como referencia un archivo de gran tamaño para observar el desempeño de transferencia (espejo) desde el nodo principal (activo) hacia el nodo secundario (pasivo).

Debe señalarse que cada nodo puede cambiar de modalidad en cualquier momento, es por ello que debe separarse en dos secciones la prueba, con la convención de que la primera (a) es donde el nodo 1 permanece como activo y el nodo 2 permanece como pasivo, y la segunda (b) es donde el nodo 2 es el que permanece como activo y el nodo 1 se encuentra en modo pasivo.

Así se ha verificado el comportamiento de DRBD y puede resumirse en las conclusiones presentadas a continuación:

a) Entonces se hace un llamado al monitor de batch con parámetro `cat /proc/drbd` y para comprobar el estado cada 3 segundos:

PRUEBA DE SERVICIOS APACHE, MYSQL Y REPLICACIÓN DE BASES DE DATOS.

Esta prueba consiste básicamente en verificar si las tablas y las consultas de la base de datos son reflejadas en el nodo pasivo. También verificar la integridad de la información luego de que el nodo 2 cambie su estado a activo, evento que se realizará manualmente. Para realizar la prueba será utilizada la herramienta phpMyAdmin la cual necesita de manera obligatoria que estén en funcionamiento los servicios de apache2 y Mysql para así crear bases de datos, tablas y consultas de manera satisfactoria, esto implica que si alguno de estos dos servicios falla, la herramienta no podrá ejecutarse, para ello es necesario provocar que el servicio de Heartbeat se detenga y así arrancará los servicios de Heartbeat el nodo 2 y se podrá verificar si los servicios funcionan correctamente.

Ahora bien, nos cercioramos que los servicios en nodo1 funcionen correctamente, probaremos apache y Mysql realizando unas pruebas en la base de datos. Se prosigue a las pruebas cuyas pantallas se muestran a continuación: Se observa que es el nodo 1 el que se encuentra en modo activo.

PRUEBAS SISTEMÁTICAS.

PRUEBAS DE VARIABLES DE SESIÓN:

En esta prueba se realizó una desconexión manual de los servicios en el nodo 1, que para el caso de estudio es el nodo activo, entonces se inició sesión desde una estación cliente en un LMS llamado Moodle, el cual residía para el momento en el servidor, efectivamente configurado, a continuación, se muestra la pantalla de inicio de sesión del LMS desde el cliente

Según los resultados anteriores puede afirmarse que la integridad del servidor al intercambiar roles es bastante sólida y considerando al intercambio manual como un incidente no esperado el trabajo del usuario sobre el LMS se ve despreciablemente afectado ya que en teoría si existe una concurrencia con un evento disparador del usuario (al hacer clic en el envió), y en ese mismo instante de tiempo ocurre un intercambio de rolles ya no será transparente para el usuario dicho intercambio. No obstante, unos segundos después podrá continuar con sus labores en la plataforma. Considerándose

despreciable el tiempo perdido y entiéndase que ningún dato de formulario será desperdiciado, naturalmente reenviará los datos y se guardarán en el nodo activo una vez realizado el desconocido intercambio para el usuario cliente.

PRUEBAS DE DESCONEXIÓN FUENTE DE ALIMENTACIÓN ELÉCTRICA:

Esta prueba consiste en la desconexión repentina del cable de alimentación AC para simular una falla en uno de los nodos, ya sea una falla de suministro ó una falla de algún componente en el circuito de la fuente de poder del equipo. Para efectos de esta prueba se desconectará el cable de alimentación AC del nodo 1 (Activo). Seguidamente se muestra el estado de ambos sabiendo que el nodo 1 se encuentra en modo activo y el nodo 2 se encuentra en modo pasivo. También luego de ver el estado del nodo se muestra respectivamente la fuente de alimentación
Pruebas de interrupción de comunicación del cable de estado y búfer entre nodos.

La comunicación entre los nodos a través de este cable es de suma relevancia, ya que, mediante este, se transfiere la replicación del sistema de archivos y se verifica el estado de cada nodo. Razón por la cual se realiza esta prueba desconectando repentinamente el cable de comunicación para evaluar el comportamiento del nodo1 (activo) el cual debería permanecer brindando los servicios y manteniendo su rol. Se muestra a continuación el estado del cable y el estado de cada nodo respectivamente

Una vez realizada la desconexión, observe el resultado de ambos nodos, los cuales no cambian sus roles, pero desconocen mutuamente el estado del otro, esto implica que la replicación de datos no se está llevando a cabo y a pesar de que el nodo 1 sigue prestando sus servicios; el clúster posee un número menor de puntos de fallas, es decir, durante el tiempo de transcurso de un incidente de este tipo el clúster no soportaría una falla adicional a esta. Sin embargo, los servicios no han sido afectados bajo la condición de falla del cable de estado y búfer

EVALUACIÓN DE FIABILIDAD DE LA PLATAFORMA.

Es importante que el lector entienda la diferencia clave entre disponibilidad y fiabilidad, debido a que se persigue que los usuarios de la plataforma de educación sin distancia posean fiabilidad en la misma, esto no implica que sea posible medir con precisión o bien llegar a medir en si la fiabilidad. Desde el punto de vista sistemático sino no como la satisfacción del usuario al utilizar la plataforma. Entonces, obsérvese la diferencia acunada por (Sommerville, 2005, p46).

La disponibilidad y fiabilidad de un sistema son propiedades que están estrechamente relacionadas y que pueden expresarse como probabilidades numéricas. La fiabilidad de un sistema es la probabilidad de que un sistema funcione correctamente tal y como se ha especificado. La disponibilidad de un sistema es la probabilidad de que un sistema esté en disposición de funcionar para proporcionar los servicios a los usuarios que lo soliciten.

Si bien estas dos propiedades guardan una estrecha relación, no se puede deducir que los sistemas fiables estarán siempre disponibles y viceversa. Por ejemplo, algunos sistemas pueden tener como requisito una disponibilidad alta, pero una fiabilidad mucho más baja. Si los usuarios esperan un servicio continuo, entonces los requerimientos de disponibilidad son altos. Sin embargo, si las consecuencias de un fallo de funcionamiento son mínimas y el sistema puede recuperarse rápidamente de dichos fallos, entonces el mismo sistema puede tener requerimientos de fiabilidad bajos. El clúster que se objeta para el presente proyecto, requiere una combinación en alta disponibilidad de sus componentes aun cuando estos posean una confiabilidad no mesurable o baja.

García J. (2005) acuñó No todas las soluciones que se pueden utilizar para un servidor Web (sic) distribuido son igualmente fiables. La fiabilidad depende por una parte de la estructura de la arquitectura seleccionada, ya que la propia topología de la arquitectura puede incrementar o reducir la probabilidad de fallo. Por otra parte, la fiabilidad también se ve afectada por la estrategia de asignación de las réplicas a los nodos servidores.

La métrica utilizada en este caso es:

Fiabilidad Mide la probabilidad de que una petición de un elemento pueda completarse con éxito. Si la probabilidad de que un sistema falle es, su fiabilidad viene dada por la expresión: $F = 1 - P$

Esta probabilidad infiere una gran cantidad de variables inmensurables a simple vista. Es por ello que la determinación de fiabilidad F de la plataforma como un todo, resulta complejo y puede conllevar errores de cálculo. Por el contrario, es posible calcular una métrica relacionada con la disponibilidad, que se mide en “nueves” según (technet.microsoft.com, 2005) y refiere a:

$100 \times \frac{\text{tiempoActivo}}{\text{tiempoActivo} + \text{tiempoInactivo}} = \text{Porcentaje de disponibilidad}$

Aplicando la ecuación anterior, en base al tiempo que se probó el clúster el cual fue de 30 días continuos aproximadamente, se tiene que: $100 \times \frac{30(720)}{30(720) + 0} = 100\%$

Observe que el 100% representa una medida ideal, generalmente representaría las horas de disponibilidad anual del clúster, y como se utilizó un intervalo de tiempo relativamente pequeño no sucedió ninguna caída, por ejemplo, al suceder alguna caída durante el año la cifra sería arrojada en “nueves”.

CONCLUSIONES

Luego de un año y cuatro meses de investigación, de comparaciones, configuraciones, pruebas de ensayo y error, documentación, para sintetizar el presente trabajo de grado finaliza sus actividades planteadas según el cronograma de actividades correlacionadas a la fase de conceptualización de las fases metodológicas implantando y configurando un servidor de alta disponibilidad de prueba experimental, como se muestra en los anexos (ver anexos A al D), conjuntamente, de cumplir con la realización de un tutorial donde se muestra de manera interactiva el funcionamiento del clúster de alta disponibilidad, por consiguiente, allí se da a entender la infraestructura del servidor y todas sus configuraciones con una interfaz amigable y dinámica, donde se juega con los papeles de los nodos activos y pasivo. Ahora bien, se realizó la documentación pertinente a la elaboración del trabajo de grado, en el cual se dieron a conocer aspectos importantes y relevantes de las características del clúster, también se muestra opiniones y en otras ocasiones tomas de decisiones de las diferentes integraciones que se lograron para que el mismo se alineara a un 99,999% de disponibilidad.

También, se muestran las pruebas realizadas al servidor para comprobar que dicha infraestructura soporta un sistema de educación sin distancia, abriendo paso a la nueva era de los sistemas educativos 100% disponibles a toda la comunidad estudiantil y administrativa de la UJAP. Para así generar una certera confiabilidad de los usuarios hacia la plataforma y promover su uso.

En adición con el fin de obtener el título de ingeniero en computación damos a conocer nuestra obra final en nuestra carrera, donde paseamos por casi todas las diversas y distintas ramas de la computación, es aquí donde converge la integración de herramientas extraordinarias para realizar un proyecto y un servidor cuyo propósito es brindar sus servicios las 24 horas del día, los 7 días de la semana y los 365 días del año, se espera haber sido lo más explícito posible para que en futuros trabajos de grados sea de gran ayuda esta obra final que con la ayuda de Dios y amor a nuestra carrera se ha concretado y se ha vuelto un producto real y que ya no es un producto de la imaginación.

Gracias por acompañarnos en este viaje lleno de nuevas tecnologías y futuros servicios de educación sin distancia a la mejor disponibilidad

REFERENCIAS

- Alonso J. (2009) Alta Disponibilidad con Heartbeat [En Línea] <http://redes-privadas-virtuales.blogspot.com/2009/03/alta-disponibilidad-con-heartbeat.html> [febrero 2009]
- Álvarez F, Padilla A, (2002). Instrumentación del Proyecto de Educación a Distancia para migrantes mexicanos en Georgia Estados Unidos. [Libro en línea] <http://bibliotecadigital.conevyt.org.mx/colecciones/documentos/somece/54.pdf> [noviembre de 2008].
- Avilez J. (2008) Recolección de datos. [Artículo en Línea] <http://www.monografias.com/trabajos12/recoldat/recoldat.shtml#entrev> [febrero 2009].
- Balestrini, R. (1997). Técnica de la Investigación. Editorial Mc Graw Hill. [Libro en línea] <http://www.megaupload.com/?d=XKSWQV9M> [noviembre de 2008].
- Bustos A, (2007). Configuración de un clúster de alta disponibilidad y balanceo de carga en Linux para satisfacer gran demanda web y servicios de resolución de nombres [Libro en línea] <http://bieec.epn.edu.ec:8180/dspace/handle/123456789/914>. [noviembre de 2008].
- Bytecoders (2007) Servidores de alta disponibilidad: Heartbeat y DRBD [En Línea] <http://bytecoders.homelinux.com/content/servidores-de-alta-disponibilidad-heartbeat.html> [diciembre 2008].
- Canet J. Cubero S. Alta disponibilidad gracias a las tecnologías de vitalización y redes [libro en línea]. <http://www.rediris.es/difusion/publicaciones/boletin/82-83/ponencia1.4B.pdf> [septiembre 2009] 77
- Canonical Ltd. Ubuntu. Especificaciones Técnicas [Ficha Técnica en línea] <http://www.ubuntu.com/products/whatisubuntu/serveredition/techspecs/9.04>. [enero 2010]
- Cañas J, (2003). Computación paralela [Libro en línea] http://portal.inf.utfsm.cl/extension/noticias.php?nro_noticia=390 [noviembre de 2008].
- Crespo O. El mito de los clústers y la alta disponibilidad [Artículo en línea]. http://www.cwv.com.ve/index.php?option=com_content&task=view&id=100&Itemid=28 [febrero 2010].
- Desktoplinux.com. 2007 Desktop Linux Market survey (octubre 2007) [Artículo en línea]. <http://www.desktoplinux.com/cgi-bin/survey/survey.cgi?view=archive&id=0813200712407> [diciembre 2008].
- Didio L. 2008 Server OS Reliability Survey (febrero 2008) [Artículo en línea]. <http://www.iaps.com/2008-server-reliability-survey.html> [julio 2009].
- Estrategia Magazine. (junio 2002). Aprender sin Distancias - Educación a distancia y tecnologías. [Revista en línea]. <http://www.estrategiamagazine.com/tecnologia/aprendersin-distancias-elearning/> [agosto 2008].
- Fernández L. (1999). Proyectos de educación a distancia en Venezuela [En línea] <http://neutron.ing.ucv.ve/revista-e/No5/MQuevedo.htm> [febrero 2009]
- Gonzales B, Gómez A. Implementación de un entorno de alta disponibilidad en un Sistema de Información Clínica [Libro en línea]. http://revista.hospitalitaliano.org.ar/archivos/servicios_attachs/2641.pdf [enero 2010]
- Gonzales O. Clúster de Alta Disponibilidad y Alto Desempeño para Servidores Web (ADAD-SW) [libro en línea]. <http://www.fpune.edu.py/docs/revista2008.pdf#page=35> [marzo 2009].
- Hartill R. (2004) Servidor HTTP Apache [En Línea]. <http://wiki.apache.org/general/> [diciembre 2008].
- Hernández (1996). Síntesis de la Investigación. Ediciones Eneva. [Libro en línea] <http://www.megaupload.com/?d=GG3KLHXN> [noviembre de 2008]. 78
- Hernández C, (2009). Metodología ITIL [Libro en línea] <http://www.monografias.com/trabajos31/metodologia-til/metodologia-til.shtml>. [diciembre 2009].
- Hoeger H, Díaz G y Núñez L, (2008) Adaptación de clúster de Linux para servicios de redes [En línea] <http://www.saber.ula.ve/handle/123456789/15976> [noviembre de 2008].
- IT. Pergamino virtual (2009) [diccionario en línea]. <http://www.pergaminovirtual.com.ar/definicion/IT.html> [noviembre 2009].
- Jaimes D. (2010) Documentación [Artículo en Línea] <http://es.wikipedia.org/wiki/Documentacion> [marzo 2010].
- LINBIT HA-Solutions GmbH. Gestión de sistema de archives [manuales en línea]. <http://www.drbd.org/> [diciembre 2008].
- Linux-HA (2010) High Availability Cluster Systems [En Línea] http://www.linux-ha.org/wiki/Main_Page [Febrero 2010].
- Netcraft.com (2010). Totals for Active Servers Across All Domains. [En Línea]. http://news.netcraft.com/archives/2010/02/22/february_2010_web_server_survey.htm l. [febrero 2010].

- Oracle. Agrupamiento en clúster y alta disponibilidad [Artículo en línea]. <http://docs.sun.com/app/docs/doc/821-0058/ghqpo?l=es&a=view>. [noviembre 2009]
- O'Reilly T, (2006) Qué es Web 2.0. Patrones del diseño y modelos del negocio para la siguiente generación del software. [En línea] http://sociedadinformacion.fundacion.telefonica.com/DYC/SHI/seccion=1188&idioma=es_ES&id=2009100116300061&activo=4.do?elem=2146 [febrero 2009].
- Perera H, Álvarez M, (2002). Proyecto de Educación a Distancia de la Facultad de Ciencias Sociales Universidad de La República de Uruguay [Libro en línea]. <http://www.fcs.edu.uy/uae/Educacion%20a%20distancia.pdf> [noviembre de 2008]. 79
- Pfister G. (1998) Computing in the Classroom: Topics, Guidelines, and Experiences [Libro en línea] www.buyya.com/papers/CC-Edu.pdf [noviembre de 2008].
- Rich Internet Applications. (Enero 2010). Wikipedia [En línea] http://es.wikipedia.org/w/index.php?title=Rich_Internet_Applications&oldid=34552661. [enero de 2010].
- Robertson A. Linux-HA tutorial Release 2(julio 2008) [Libro en línea] http://www.linux-ha.org/_cache/HeartbeatTutorials__LWCE08-ha-tutorial.pdf [noviembre 2008].
- Rocks Core Development. Entendimiento de clusters de alto rendimiento. <http://www.rocksclusters.org/wordpress/> [manuales en línea]
- Rodríguez J. Tolerancia a fallos, en clusters de computadores geográficamente distribuidos, basada en replicación de datos [libro en línea]. http://www.tesisexarxa.net/TESIS_UAB/AVAILABLE/TDX-1013106-133133//jrs1de1.pdf [febrero 2010]
- Sabino C. (1997). El proceso de Investigación. Editorial Panapo. [Libro en línea] <http://www.megaupload.com/?d=JKMH9C4E> [noviembre de 2008].