**DESARROLLO DEL INFORME DE INVESTIGACION  
(ENFOQUE CUALITATIVO)**

PAGINAS PRELIMINARES

Página del jurado.

Dedicatoria

Agradecimiento

Declaratoria de autenticidad

Presentación

Índice

RESUMEN Incluyen palabras claves

ABSTRACT They include keywords

1. INTRODUCCIÓN

Las Bases de Datos hoy en día es una tecnología que se ha desarrollado a pasos agigantados en los últimos años debido al hecho de que las instituciones y empresas consideran de vital importancia a la información, convirtiendo a la misma en el activo principal de las organizaciones, como lo es el caso de la Universidad Particular de Chiclayo, en la cual se hace uso de esta tecnología para almacenar todo tipo de información referente a los estudiantes, como notas, asistencias, pagos, entre otro tipo de datos que se utiliza para poder tomar decisiones o planear estrategias administrativas.

Debido a la importancia que tiene la información que guarda las empresas en sus bases de datos se realizan técnicas de seguridad como medida de prevención ante desastres, la técnica más utilizada es la de realizar backups cada cierto tiempo, sin embargo, existe métodos más óptimos y eficientes para garantizar la disponibilidad y seguridad de la información como lo es la replicación de bases de datos sobre la cual se centra este proyecto de investigación.

En la actualidad los procesos de sincronización suelen realizarse de manera manual, lo que origina errores que pueden perjudicar el desempeño de la base de datos y, por ende, el de los sistemas que dependen de dicha base de datos. Los escenarios más frecuentes en donde se lleva a cabo este proceso son cuando se realizan pases a producción que traen consigo cambios en la estructura de la base de datos y ambientes en donde se utilizan bases de datos distribuidas para el procesamiento y almacenamiento de información en distintas sucursales.

El presente proyecto tiene como objetivo la implementación de servidores de bases de datos en las principales sedes de la Universidad Particular de Chiclayo, luego mediante la técnica de replicación identificar las diferencias y similitudes entre distintos objetos dentro de las bases de datos y a su vez, permita sincronizar la estructura de dichos objetos a una hora determinada.

1. PROBLEMA DE INVESTIGACION
   1. Aproximación temática: Observaciones y Estudios relacionados, preguntas orientadoras.

Actualmente existen investigaciones y estudios relacionados con el tema de comparador y sincronizador de bases de datos los cuales serán mencionados a continuación con un breve resumen de la investigación y las personas que realizaron dicho estudio.

**Autor :** Alejandra Zavala Mendoza

**Fecha :** Agosto del 2010

**Nombre del proyecto :** “Replicación y fragmentación de bases de datos distribuidas”

**País :** México

**Lugar :** Chihuahua, Coatzacoalcos

**Universidad :** Universidad Veracruzana

**Resumen**

El uso y la necesidad de contar con bases de datos va creciendo día con día, ya que en cualquier empresa necesita guardar información ya sea sobre sus ventas, compras, inventarios, clientes, empleados, información financiera, etc. Y al no contar con esta información será muy difícil llevar un buen funcionamiento de cualquier organización.

Las bases de datos facilitan la búsqueda y el almacenamiento de toda la información, garantizando la fiabilidad de la información que esta contiene.

Las bases de datos distribuidas tienen múltiples ventajas. En primer lugar los datos son localizados en lugar más cercano, por tanto, el acceso es más rápido, el procesamiento es rápido debido a que varios nodos intervienen en el procesamiento de una carga de trabajo, nuevos nodos se pueden agregar fácil y rápidamente. La comunicación entre nodos se mejora, los costos de operación se reducen, son amigables al usuario, la probabilidad de que una falla en un solo nodo afecte al sistema es baja y existe una autonomía e independencia entre los nodos.

Los almacenamientos de datos son un elemento fundamental para las empresas por ello ha llamado la atención de estas ya que en la actualidad cualquier compañía por pequeña que sea tiene que manejar grandes cantidades de información, para poder tener control sobre todas las áreas de su empresa, es por ello que surge la necesidad de emplear bases de datos dentro de las organizaciones, no importando el giro de las empresas.

Por lo cual, se le denomina “Base de Datos” a la colección de datos aparentes usados por el sistema de aplicaciones de una determinada empresa.

La mayoría de las empresas necesitan contar con bases de datos ya que las bases de datos facilitan la realización de sus actividades, garantiza la seguridad, la confiabilidad de la información, consistencia de los datos, mejora en la disponibilidad de información, guarda información sobres sus clientes, proveedores, información financiera, inventarios, compras ventas etc. Si contamos con un buen control de la información se podrá llevar una empresa de mejor manejar y más organizadamente. La gran cantidad de información que producen grandes empresas geográficamente distribuidas, ha generado una gran dispersión de los datos asociados a sus sistemas de información. Para compartir dichos datos se requiere de tecnología capaz de integrarlos para permitir el acceso concurrente de múltiples usuarios. Una de las tecnologías que se propone alcanzar este tipo de objetivos son las bases de datos distribuidas

Las bases de datos distribuidas mejoran la disponibilidad de los datos pues permite que el sistema pueda seguir operando y los usuarios desde cualquier sitio puedan acceder a la información, exactamente igual como si los datos estuvieran en el mismo sitio del usuario.

Las bases de datos, hoy en día, ocupan un lugar determinante en cualquier área ya que se emplea en diversas empresas no importando el giro de esta, es por ello que deben de tener los conocimientos necesarios para poder utilizar las bases de datos.

Las bases de datos distribuidas se conectan por medio de redes de computadoras, estas pueden tener cientos de usuarios conectados concurrentemente, por lo cual es un problema en muchas empresas que compiten en los mercados mundiales. Debido al alto grado de conectividad que provee Internet a las empresas, se debe tener un buen diseño de replicación y fragmentación.

Este problema consiste en tratar de determinar la mejor forma de dividir las relaciones de la base de datos en fragmentos. La fragmentación mejora la disponibilidad de la información notablemente por que el sistema puede seguir operando mientras, por lo menos, uno de los sitios está activo, con la replicación también se mejora el rendimiento de la recuperación en consultas globales, porque el resultado de semejante consulta se puede obtener localmente en cualquier sitio, esto ayuda a contar con la información necesaria de una forma más fácil y rápida y hoy en día para las empresas contra con la información en un corto tiempo es un factor clave para la toma de decisiones oportunas.

Para contar con una buena base de datos distribuida esta debe de ser replicada y fragmentada. La replicación y fragmentación permite tener disponible la información, reduce costos, mejora el rendimiento de la operación leer, aumenta la flexibilidad, se obtiene la información de una manera más fácil y rápida teniendo un mejor procesamiento de las consultas.

Mediante la replicación de información, las bases de datos distribuidas pueden presentar cierto grado de tolerancia a fallos haciendo que el funcionamiento del sistema no dependa de un solo lugar como en el caso de las bases de datos centralizadas.

**Autor :** Milán Franco, Jesús Manuel

**Fecha :** 30 de junio 2008

**Nombre del proyecto : “**Replicación Autonómica de Bases de Datos”

**País :** España

**Lugar :** Madrid, C/ Ramiro de Maeztu **universidad : U**niversidad Politécnica de Madrid

**Resumen**

En los últimos años, la evolución de los sistemas informáticos ha creado sistemas cada vez más complejos que constan de multitud de elementos interconectados a través de redes de comunicaciones rápidas. La configuración óptima de cada uno de estos elementos para obtener el máximo rendimiento del sistema depende de un gran número de factores como son la carga del sistema, el tipo de carga y el hardware sobre el que se ejecuta, entre otros. Esta diversidad de parámetros de configuración hace que la administración de dichos sistemas sea cada vez más complicada y que requiera administradores de sistemas con mucha experiencia y dedicación a estas tareas. Dichos sistemas, además, están sujetos a cambios de diversa naturaleza como son variaciones en la carga del sistema, en el tipo de carga, en los recursos disponibles; fallos y recuperaciones de los elementos del sistema, etc. Por esta razón ha surgido la necesidad de crear sistemas que sean capaces de adaptarse de forma automática a los cambios en el entorno que les rodea sin necesidad de intervención humana y que al mismo tiempo exhiban un alto rendimiento y calidad de servicio. Estos sistemas adaptables dinámicamente interaccionan con su entorno para detectar los cambios en el mismo, analizan la información obtenida para generar configuraciones óptimas a las condiciones actuales del entorno y modifican su configuración para adaptarse a estos cambios. El objetivo de estos sistemas es maximizar el rendimiento del sistema de acuerdo a las métricas de rendimiento de interés. En estos sistemas distribuidos, la replicación de datos juega un papel muy importante como forma de aumentar el rendimiento. Hasta ahora la replicación se implementaba dentro de la propia base de datos, lo que obligada a realizar modificaciones en la base de datos y estos no siempre eran posibles. Una línea de investigación nueva iniciada en el LSD y que ha sido adoptada por gran parte de los investigadores es la implementar la lógica de la replicación fuera de la base de datos a nivel de middleware. Este trabajo de investigación se enmarca dentro de esta línea de investigación y en él se estudia la adaptabilidad dinámica en el contexto de las bases de datos replicadas basadas en middleware. La adaptabilidad del sistema replicado afecta principalmente a dos áreas de trabajo: La adaptación local (el control del nivel de concurrencia) y la adaptación global (el equilibrado de carga). El objetivo de la adaptación local es conseguir el máximo rendimiento de cada réplica individualmente sin llegar a sobrecargarlas, mientras que el objetivo de la adaptación global es conseguir que la carga del sistema esté equilibrada, evitando de esta forma que algunas réplicas estén saturadas, mientras que otras estén ociosas. El trabajo realizado en esta tesis desarrolla un sistema de replicación adaptable a nivel de middleware que englobe ambas áreas de trabajo desarrollando protocolos y algoritmos para el control de la concurrencia a nivel local y el equilibrado de carga a nivel global y que maximizan de forma automática el rendimiento del sistema bajo distintos tipos de carga, fallos y recuperación de réplicas.

**Autor :** David Freire V

**Fecha :** 2010

**Nombre del proyecto :** “Replicación Síncrona master to master de bases de datos mysql en red hat Enterprise Linux o compatibles”

**País :** Ecuador

**Lugar :** Guayaquil, Guayas

**Universidad :** Universidad Pública de Navarra  
**Resumen**

El objetivo es realizar una replicación de Base de Datos MySQL, que optimice los recursos de la empresa al usar completamente ambas bases de datos, ya que, normalmente en los esquemas de replicación Amo a Esclavo (Master to Slave) se tiene una funcionando y la otra sirve de repositorio de información y de contingencia en caso de que la base principal deje de ser funcional, y no se logra explotar al máximo esta base. En el esquema actual, Amo a Amo (Master to Master), ambas bases trabajan de igual forma y se logra aprovechar todos los recursos de las misma, no dejando de contener el nivel de contingencia que posee la replicación tradicional Amo a Esclavo (Master to Slave). También se busca mantener el criterio de Alta disponibilidad, al realizar una replicación Síncrona, es decir, todos los cambios a nivel de inserción, actualización y eliminación de registros que se realicen a cualquiera del base de datos se replicaran automáticamente en la otra base de datos y viceversa.

**Autor :** Giancarlo Roberto Calderón Garay

**Fecha :** Abril del 2009

**Nombre del proyecto :** “Análisis, Diseño e implementación de un comparador y sincronizador de bases de datos relacionales de distintos manejadores”

**País :** Perú

**Lugar :** Lima, San Miguel.

**Universidad :** Pontificia Universidad Católica Del Perú

**Resumen**

El presente proyecto consiste en el análisis, diseño e implementación de un sincronizador de bases de datos relacionales de distintos manejadores, cuya finalidad es realizar la comparación de objetos entre dos bases de datos y sincronizar dichos objetos de acuerdo a las diferencias encontradas. Para la gestión del proyecto se ha seguido las prácticas recomendadas por PMI y se han considerado aquellos procesos de gestión necesarios para el desarrollo del proyecto. Para el desarrollo de la herramienta se optado por utilizar la metodología AUP (cuyas siglas en inglés significan Agile Unified Process), cuyas fases y disciplinas se han adaptado mejor al desarrollo de la aplicación. La arquitectura seleccionada ha permitido que la aplicación pueda trabajar con distintos manejadores de bases de datos relacionales. Para llevar a cabo este objetivo se ha implementado un componente que se encarga de realizar la abstracción de los manejadores e interactúa con las demás capas de manera transparente. La implementación de esta componente se ha realizado a través de archivos XML que, con una estructura definida, permiten que la aplicación consulte la metadata de la base de datos y construya las sentencias SQL para la sincronización de objetos. De esta manera, si se desea incorporar otra base de datos sólo se necesita definir el contenido de la plantilla XML y los parámetros de conexión que van definidos en un archivo de configuración. El proceso de sincronización implementado se puede realizar de dos maneras: mediante una comparación previa de los objetos a sincronizar o mediante un asistente de sincronización, el cual permite al usuario seleccionar de manera más personalizada qué objetos desea sincronizar sin importar qué diferencias existen con los objetos de la base de datos destino. En ambos escenarios, se genera un archivo de bitácora del proceso de sincronización donde se puede verificar los resultados del proceso.

* 1. Formulación del problema de investigación

¿Qué tecnologías nos permitirá el diseño e implementación del comparador sincronizador de bases de datos relacionales en servidores de la Universidad Particular de Chiclayo para que haya una alta disponibilidad de datos de los estudiantes universitarios?

* 1. Justificación

Debido a que actualmente la Universidad Particular de Chiclayo cuenta con una base de datos centralizada y a la lentitud de la red, cada vez que se quiere hacer la consulta de las notas y/o asistencias de los estudiantes desde una sucursal distinta, muchas veces la información solicitada no es mostrada o llega con bastante tiempo de retraso por ello es necesario la implementación de un comparador sincronizador de bases de datos relacionales como una solución ante este problema.

Al momento de llevar a cabo esta investigación se tuvo que evaluar el tipo de sincronización que se tendrá que realizar y la forma en que se realizará, se llegó a la conclusión de que el tipo de replicación más conveniente es la de Mater-to-Master para que las bases de datos trabajen al 100%, debido a que es necesario que se pueda leer y modificar la información en todas las sucursales, por otro lado la forma en que se llevará a cabo la sincronización es asíncronamente ya que es necesario que se realice el intercambio de información entre las bases de datos a una hora del día en que la red sea más rápida con el fin de evitar problemas en la escritura de datos.

* 1. Relevancia

Esta nueva alternativa, no convencional, es muy útil también para mantener segura la información de los estudiantes ya que ante cualquier desastre que ocurra en una de las bases de datos su información no será perdida debido a que también se encontrará disponible en las bases de datos de las otras sedes.

Es importante mencionar que las herramientas que se utilizarán para realizar la implementación del comparador sincronizador, tales como el sistema operativo, el gestor de bases de datos, todos ellos son open-source por lo que no es necesario pagar licencia por alguno de ellos, con lo que se demostrará que no es necesario realizar grandes inversiones económicas para obtener una alta disponibilidad de la información.

* 1. Contribución
  2. Objetivos

Los objetivos de este trabajo de investigación se pueden dividir en dos grupos: objetivos generales y en objetivos específicos.

* + 1. Objetivo general

Realizar el diseño e implementación de un comparador y sincronizador de bases de datos relacionales que permita tener una alta disponibilidad de las asistencias y notas de los estudiantes en todas las sedes de la Universidad Particular de Chiclayo.

* + 1. Objetivos específicos
* Instalación del sistema operativo CentOS 6 en los servidores que se utilizarán en las sedes de la Universidad Particular de Chiclayo.
* Analizar e Identificar como está estructurada y configurada la red en los locales de la Universidad Particular de Chiclayo.
* Diseñar la forma en que los servidores serán ubicados en la red local.
* Realizar configuración básica los de servidores web Apache y de bases de datos PostgreSQL 8.0.
* Instalación y configuración de la herramienta de replicación Bucardo en uno de los servidores Master.
* Determinar las fluctuaciones del ancho de banda de la red e identificar una hora del día en la que sea menor.
* Programar la herramienta de replicación para que realice la replicación en una hora asignada utilizando las herramientas cron y crontab propias en sistemas operativos Linux.
* Realizar las pruebas necesarias para confirmar que se realice correctamente la sincronización y que se garantice la disponibilidad de la información de las bases de datos.
* Crear un manual de instalación y configuración de los servidores web, base de datos, y configuración de la herramienta de replicación bucardo.
  1. Hipótesis
* La lentitud de la red afecta el tiempo de respuesta de una consulta hecha a una base de datos.
* El mayor problema de lentitud de respuesta de la base de datos se encuentra en la Ciudad Universitaria de la Universidad Particular de Chiclayo ya que se encuentra a gran distancia de la base de datos central ubicada en el local de Miraflores.
* Es posible que un comparador sincronizador, tenga la capacidad de evitar que haya tiempo de demora de respuesta al realizar una consulta, y mantener una alta disponibilidad de las asistencias y/notas de los estudiantes en todas las sedes principales de la Universidad Particular de Chiclayo.

1. MARCO METODOLOGICO
   1. Metodología
      1. Tipo de estudio

**Por su Método de Investigación Documental y Campo (Teórica y Práctica):**

Cuyo método de investigación se concentró en la recopilación de información en forma documental de replicación de bases de datos, respectivamente la recopilación de información se realizó enmarcado dentro del ambiente de redes LAN, uso del sistema operativo CentOS 6 (Linux/fedora), configuración de servidores web, servidores de bases de datos, el funcionamiento, configuración de servidores de replicación de bases de datos, topología de la red en las sedes de la Universidad de Particular de Chiclayo, y el uso de la herramienta de replicación BUCARDO, etc.., en el que se presentó un fenómeno o hecho que lo estudiamos los cuales son la utilización del proceso de replicación para copiar y mantener actualizados los datos en varios nodos o servidores de bases de datos en las diferentes sedes de la UDCH

* + 1. Diseño
  1. Escenario de estudio
  2. Caracterización de sujetos.

Los principales participantes involucrados en este proyecto de investigación son todos aquellos que hacen uso directa o indirectamente de los sistema de información de la universidad particular de Chiclayo tales como:

* Trabajadores administrativos
* Directores de escuela
* Decanos
* Estudiantes
* Padres de familia de los estudiantes
  1. Trayectoria metodológica
  2. Técnicas e instrumentos de recolección de datos.

Siguiendo los pasos de una investigación científica aplicada, precisamos encontrar los medios de recolección de información para luego procesarlas, analizarlas y posteriormente extraer las conclusiones sobre la base de la información recolectada, aplicando las técnicas e instrumentos pertinentes como la observación, entrevistas con los usuarios y administradores de las bases de datos, y revisión documental.

**Observación**

Previa la observación participativa detectamos y obtuvimos información sobre la estructura de la red LAN de los locales de la UDCH, como en el hardware de conectividad, hardware de interfaz, software de comunicaciones, fallas de los usuarios.

**Entrevista**

Con el propósito de conocer la opinión de los administradores de redes, usuarios, sobre la posibilidad de implementar uno o más servidores de replicación de base de datos relacionales para solucionar los problemas que ocurren.

**Técnica documental**

A fin de establecer el sustento teórico de la presente investigación se procedió a desarrollar la revisión bibliográfica de la literatura relacionada con el problema de estudio, como es la implementación de servidores locales para la replicación de base de datos relacionales.

* 1. Tratamiento de la información
  2. Mapeamiento
  3. Rigor Científico

Se puede hacer

Muestreo teórico }

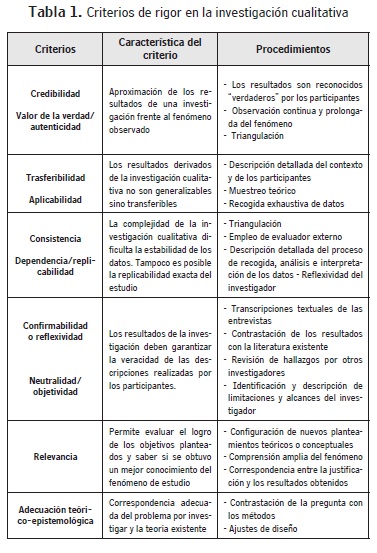
Recoger abundantes datos descriptivos }

Descripciones densas y minuciosas

} Solapamiento de métodos }

Pistas de revisión (cuadernos de campo, notas de las entrevistas, etc.) }

Investigador externo



1. RESULTADOS
   1. Descripción de resultados
   2. Teorización de unidades temáticas
2. DISCUSIÓN
3. CONCLUSIONES
4. RECOMENDACIONES
5. PROPUESTA ARQUITECTONICA
   1. Programación Arquitectónica
   2. Proyecto Arquitectónica
   3. Memoria Descriptiva
   4. Costo y Presupuesto
6. REFERENCIAS BIBLIOGRAFICAS
7. ANEXOS