UNIVERSIDAD PARTICULAR DE CHICLAYO

FAUIA-UDCH

FACULTAD DE ARQUITECTURA Y URBANISMO e INGENIERIAS

ESCUELA PROFESIONAL DE INGENIERIA INFORMATICA Y DE SISTEMAS

TESIS

“IMPLEMENTACION DE UNA ARQUITECTURA DE ALTA DISPONIBILIDAD DE BASE DE DATOS EN SERVIDORES LINUX PARA LA GERENCIA DE TRABAJO Y PROMOCION DEL EMPLEO LAMBAYEQUE ENERO – JULIO 2020”

PARA OPTAR EL TÍTULO PROFESIONAL DE

INGENIERO INFORMÁTICO Y DE SISTEMAS

PRESENTADO POR:

Bachiller Johan Manuel Rivas Rojas

ASESOR:

Ing. Oscar Castro Yoshida

Chiclayo, Julio de 2020

DEDICATORIA

A mis padres, por su amor, trabajo y sacrificio en todos estos años, gracias a ustedes he logrado llegar hasta aquí y convertirme en lo que soy. Ha sido un orgullo y un privilegio de ser su hijo. A mis hermanas por estar siempre presentes, acompañándome y por el apoyo moral, que me brindaron a lo largo de esta etapa de mi vida. A todas las personas que de alguna u otra forma me han apoyado y han hecho que el trabajo se realice con éxito en especial a aquellos que nos abrieron las puertas y compartieron sus conocimientos.

.

AGRADECIMIENTO

Agradezco a Dios por bendecirme la vida, por guiarme a lo largo de nuestra existencia, ser el apoyo y fortaleza en aquellos momentos de dificultad y de debilidad. Gracias a mis padres Noemí Rojas y Julio Rivas por ser los principales promotores de mis sueños, por confiar y creer en mis expectativas, por los consejos, valores y principios que me han inculcado. Agradezco a mis docentes de la Escuela de Ingeniería de Informática y Sistemas de la Universidad de Chiclayo, por haber compartido sus conocimientos a lo largo de la preparación de mi profesión, de manera especial, al Ingeniero Oscar Castro Yoshida, asesor de mi proyecto de investigación quien me ha guiado con su paciencia, y su rectitud como docente.

DECLARACIÓN DE AUTENTICIDAD

Yo, Johan Manuel Rivas Rojas, identificado con Documento Nacional de Identidad DNI 44130733, egresado de la Escuela de Ingeniería Informática y de Sistemas de la Facultad de Arquitectura y Urbanismo e Ingenierías, a efectos de cumplir con las disposiciones vigentes consideradas en el reglamento de Grados y Títulos de la Universidad Particular de Chiclayo, declaro bajo juramento que el informe del trabajo de suficiencia profesional denominada:

“IMPLEMENTACION DE UNA ARQUITECTURA DE ALTA DISPONIBILIDAD DE BASE DE DATOS EN SERVIDORES LINUX PARA LA GERENCIA DE TRABAJO Y PROMOCION DEL EMPLEO LAMBAYEQUE ENERO – JULIO 2020"

Que:

1. El informe desarrollado en el presente proyecto es de mi autoría.
2. He respetado y seguido los lineamientos de diseño, reglamento y parámetros establecidos para el desarrollo del informe.
3. El presente informe no ha sido plagiado, pues no ha sido objeto de publicaciones, ni presentada por algún tesista anteriormente para obtener un grado académico previo o título profesional.
4. La información presentada en el presente inform es el resultado de varias visitas realizadas en el campo, entrevistas y consultas al personal involucrado dentro del proyecto y éstos no han sido falseados, pues los resultados contribuirán a dar a conocer la realidad problemática con respecto a la investigación.

En tal sentido asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada por lo cual me someto a lo dispuesto en las normas académicas de la Universidad de Chiclayo.

Chiclayo, Julio del 2020

El autor

PRESENTACIÓN

Señores Miembros del Jurado:

De conformidad y cumpliendo lo estipulado en el Reglamento de Grados y Títulos de la Facultad de Arquitectura y Urbanismo, Ingenierías y Arte de la Universidad Particular de Chiclayo, para optar el Título Profesional de Ingeniero Informático y de Sistema, someto a vuestra consideración el presente informe Titulado:

“IMPLEMENTACION DE UNA ARQUITECTURA DE ALTA DISPONIBILIDAD DE BASE DE DATOS EN SERVIDORES LINUX PARA LA GERENCIA DE TRABAJO Y PROMOCION DEL EMPLEO LAMBAYEQUE ENERO – JULIO 2020".

Gracias

ÍNDICE GENERAL

Contenido

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

RESUMEN

El presente informe de tesis denominado “IMPLEMENTACION DE UNA ARQUITECTURA DE ALTA DISPONIBILIDAD DE BASE DE DATOS EN SERVIDORES LINUX PARA LA GERENCIA DE TRABAJO Y PROMOCION DEL EMPLEO LAMBAYEQUE ENERO – JULIO 2020", está orientado a garantizar el buen funcionamiento de los sistemas informáticos con los respaldos y réplicas de los datos de los datos de la Gerencia en mención.

Considerando la estructura brindada por la Escuela de Informática y de Sistemas de la Universidad de Chiclayo, para el desarrollo del informe de las tesis, es que presento el resumen a continuación de los capítulos:

En el Capítulo I se muestra el estudio y descripción de la situación actual o realidad problemática de la institución, y en base a esta situación se formula el problema para dar origen al planteamiento de la hipótesis y lograr obtener o desarrollar el problema definiendo los objetivos del presente proyecto de investigación.

En el Capítulo II, detallo las bases teóricas en el cual se mencionan los elementos principales del contenido del informe, donde desarrollamos los siguientes puntos como son los antecedentes de investigaciones ya desarrolladas que nos sirven de guía y ver su enfoque ante una realidad problemática específica, describimos en el marco teórico los principales temas y conceptos que se usaran en el desarrollo del proyecto detallando términos y sus definiciones.

En el Capítulo III mostraremos la explicación de los mecanismos utilizados para el análisis de nuestra problemática de investigación, este capítulo es conocido como el marco metodológico. Como nuestra investigación es de campo se deben de tener en cuenta algunos aspectos como son el diseño de la investigación, las variables a ser usadas, su operalizacion y sobre que muestra y población se va aplicar. También se define las técnicas de recolección de datos y que instrumentos se usan para obtener dichos datos, y así llegar a un análisis de dichos datos.

En el Capítulo IV y V se establecen los resultados obtenidos del capítulo anterior y se realiza la discusión de dichos resultados.

En el capítulo VI y VII se lograra definir a que conclusiones se llega y dar las recomendaciones necesarias para dar continuidad al presente proyecto en el tiempo.

En el capítulo VIII se describe cual es la propuesta tecnológica que se implantara en el informe y está conformada por las generalidades del proyecto, así como el análisis de la situación actual, diseño lógico de la propuesta tecnológica como los requerimientos y equipamiento a ser utilizados llegando a la implementación de la propuesta. También incluye el costo o presupuesto a ser usado en nuestro.

En el capítulo IX se lista principales referencias bibliográficas o fuentes que han sido revisadas para lograr desarrollar e implementar el presente proyecto desde el inicio y hasta el final.

Por último, en el capítulo X se considera algunas explicaciones adicionales respecto a la tecnología utilizada en el presente proyecto.

ABSTRACT

This thesis report called "IMPLEMENTATION OF AN ARCHITECTURE OF HIGH AVAILABILITY OF DATABASE IN LINUX SERVERS FOR THE LABOR MANAGEMENT AND PROMOTION OF EMPLOYMENT LAMBAYEQUE JANUARY - JULY 2020", is aimed at guaranteeing the proper functioning of computer systems with the backups and replicas of the data of the aforementioned Management data.

Considering the structure provided by the School of Informatics and Systems of the University of Chiclayo, for the development of the thesis report, I present the following summary of the chapters:

Chapter I shows the study and description of the current situation or problematic reality of the institution, and based on this situation the problem is formulated to give rise to the hypothesis statement and to obtain or develop the problem by defining the objectives of the present research project.

In Chapter II, I detail the theoretical bases in which the main elements of the content of the report are mentioned, where we develop the following points such as the background of investigations already developed that serve as a guide and see their approach to a specific problematic reality, We describe in the theoretical framework the main themes and concepts that will be used in the development of the project, detailing terms and their definitions.

In Chapter III we will show the explanation of the mechanisms used for the analysis of our research problems, this chapter is known as the methodological framework. As our research is in the field, some aspects must be taken into account such as the design of the research, the variables to be used, and their operation and on which sample and population it will be applied. It also defines the data collection techniques and what instruments are used to obtain said data, and thus arrive at an analysis of said data.

In Chapter IV and V the results obtained from the previous chapter are established and the results are discussed.

In chapter VI and VII it will be possible to define the conclusions to be reached and give the necessary recommendations to give continuity to this project over time.

Chapter VIII describes the technological proposal that will be implemented in the report and is made up of the generalities of the project, as well as the analysis of the current situation, logical design of the technological proposal as well as the requirements and equipment to be used arriving at the implementation of the proposal. It also includes the cost or budget to be used in our.

Chapter IX lists the main bibliographic references or sources that have been reviewed in order to develop and implement this project from the beginning to the end.

CAPÍTULO I

INTRODUCCIÓN

CAPITULO I: INTRODUCCIÓN

1.1. Realidad Problemática

Día a día los procesos se mejoran dentro de la Gerencia de Trabajo y desde año 2010 se ha aplicado una política de mejorar los procesos apoyados en la tecnología y desarrollo de sistemas informáticos, es por eso es que urge mantener la información actualmente procesada y digitalizada en las bases de datos a salvaguarda y siempre disponibles tolerante a fallas, ya que una demora en los tramites son propensos a sanciones por parte del personal que la tiene a su custodia. Los ciudadanos día a día vienen a realizar sus consultas y si esta no está disponible realizan quejas ya que algunos de ellos vienen de lugares lejanos a la ciudad, otros solicitan información de hace muchos años lo que ocasiona que la búsqueda se vuelva algo engorrosa y si a esto sumamos que los sistemas donde se encuentra digitalizados la información no se encuentre disponible las 24 horas nuestra atención será considerada pésima ante la ciudadanía. No contar con la disponibilidad de un servicio o datos críticos puede implicar un costo significativo para las empresas ocasionando grandes perdías en productividad, ingresos y ciudadanos insatisfechos y generando una mala imagen a nivel regional. Otro de los problemas detectados es que muchas veces existen perdidas de información o de datos por parte de que los equipos en donde se encentran almacenados las información por falla de los equipos que son muy antiguos donde se almacenan y los diversos cortes de energía eléctrica que hace que los equipos informáticos fallen en los discos duros perdiéndose la información de estas principalmente de los centro de datos, que aunque cuenten con equipos de UPS hay veces que son constantes y de larga duración.

La Gerencia debe de pensar en tener la base de datos en alta disponibilidad para asegurar que los datos estén siempre disponibles a los ciudadanos ante cualquier anomalía entendiéndose por este concepto a cualquier falla que se dé a consecuencia de factores de la naturaleza, fallas de hardware, error en el software, caídas en la conectividad de la red o por la intervención en muchos casos del usuario. Por esta razón, es muy necesario lograr

desarrollar un plan que garantice y logre la alta disponibilidad de la información en caso ocurra algún imprevisto o evento inesperado. En la actualidad existen en el mercado varias herramientas tecnológicas con las que podemos llegar a implementar este tipo de tecnologías para lograr respaldar y resguardar la información cola finalidad de disminuir los daños causados por alguna anomalía durante los procesos de la institución y así evitar daños irreparables de información a la Gerencia. El presente proyecto vamos a elaborar, diseñar y configurar una arquitectura que nos permita la tolerancia a fallos, que mantenga una tecnología denominada de alta disponibilidad para las Bases de Datos y así lograr el objetivo de no que no se detengan los procesos o transacciones del servicio de base de datos de la institucion, puesto que se contara de un grupo de servidores de base de datos en una configuración de servidores de la forma activo - pasivo o, es decir que si el servidor principal o primario o conocido como nodo activo sufre algún desperfecto o se necesita apagar para un mantenimiento preventivo el otro servidor secundario o conocido también como Pasivo pasa a modo Activo sin necesidad de tener grandes tiempos fuera de línea de los servicios informáticos.

Es por esto que se presenta el siguiente proyecto para mantener la información siempre presente para los ciudadanos y mejorar nuestro servicio de atención al público.

.

1.2. Formulación del Problema

¿Permitirá la implementación de una arquitectura de alta disponibilidad de base de datos en servidores Linux la continuidad de accesibilidad de la información en la Gerencia Regional de Trabajo y Promoción del Empleo?

1.3. Hipótesis

La implementación de una arquitectura de alta disponibilidad de base de datos en servidores Linux permitirá la continuidad de accesibilidad de la información en la Gerencia Regional de Trabajo y Promoción del Empleo.

1.4. Objetivos

1.4.1. Objetivo General

Implementar una arquitectura de alta disponibilidad de base de datos en servidores Linux para la Gerencia Regional de Trabajo y Promoción del Empleo para tener continuidad de accesibilidad de la información.

1.4.2. Objetivos Específicos

- Realizar un análisis de las diferentes bases de datos con las que cuenta la Gerencia Regional.
- Analizar las diferentes causas que han provocado las fallas y tiempo de inactividad de las base de datos.
- Determinar la arquitectura a implementar en alta disponibilidad y así lograr las réplicas de los datos.
- Implementar los servidores activo y pasivo de base de datos en Misal implementados en alta disponibilidad.
- Realizar pruebas de continuidad del servicio en caso de caídas y verificar el buen funcionamiento de los servidores en base a la arquitectura escogida.

Capítulo II:
BASES TEÓRICAS

CAPITULO II: BASES TEÓRICAS

2.1. Antecedentes de la Investigación

2.1.1. TRABAJO 01

Autores:

- Lester Juan De Dios Villar Zamora.

Fecha: 2014 – Cajamarca Perú

Nombre del proyecto:

"Clúster de servidores Linux para alta disponibilidad De la información"

Resumen:

"La investigación realizada tiene como objetivo implementar y diseñar un clúster de servidores en entorno Linux para brindar un servicio de alta disponibilidad de la información, para tal propósito se utilizaron herramientas como la base de datos postgreSQL, Pppool 11, Heartbeat. La investigación tiene como caso de estudios la empresa MINE SENSE SOLUTIONS que implementará un sistema de control de flotas pesadas para el proyecto minero Hudbay-Constancia en la región del Cusco. La empresa en cuestión tiene la necesidad de alta disponibilidad del servicio de información para el sistema que implementará por ser de vital importancia en las operaciones diarias. Para el diseño e implementación del clúster de alta disponibilidad se investigó las herramientas disponibles y que mejor se adaptan a las necesidades antes mencionadas.

La implementación comienza con la instalación del SO Ubuntu Server, luego se procede con la configuración de los IP estáticos necesarios para el tráfico de información y la configuración inicial de PostgreSQL para el acceso con el usuario root por defecto, se necesita acceso del tipo SSH entre los servidores para lo cual es necesario crear claves públicas para el acceso remoto desde los diferentes nodos que conforman el clúster".

"Luego de la configuración SSH se tendrá que configurar la replicación de información para lo cual se utilizó el mecanismo de espejo del motor

de Base de Datos PostgreSQL, Stream Replication". Al tener los nodos en espejo ya es posible manipular los roles de los nodos a través del administrador del clúster, el middleware pgpool-11 y después de la configuración del middleware es necesario configurar la alta disponibilidad del controlador del clúster a través de la herramienta por excelencia en Linux para alta disponibilidad, Heartbeat. los resultados obtenidos luego de la implementación del clúster son: alto rendimiento del servicio de datos, escalabilidad en la arquitectura del clúster, balanceo de carga entre los nodos del clúster los cuales se distribuyen el tráfico de información y las transacciones requeridas, haciendo así que se aminore el procesamiento en los nodos del clúster. Al tener un alto rendimiento respecto a la continuidad ante caída de nodos (failover) se pudo reducir el tiempo de inoperatividad estimado llegando a tener el 99.99 de disponibilidad de la información en un año.

2.1.2. TRABAJO 02

Autores:

- Efraín Ricardo Bautista Ubillús.

Fecha: 2014 – Lima Perú

Nombre del proyecto:

"Herramienta para el modelado de la replicación de mysql basada en la ingeniería dirigida por modelos"

Resumen:

"Para modelar la replicación de MySQL, los administradores de base de datos (DBA's) utilizan herramientas de diagramación, tales como Microsoft Visio. Sin embargo, este tipo de herramientas no permiten validar automáticamente si un modelo de replicación MySQL está libre de errores, lo que puede resultar tener documentación errónea de los modelos de replicación. Estas herramientas tampoco permiten generar automáticamente a partir del modelo, los comandos mysql replicate de configuración. La falta de estas funcionalidades conlleva a realizarlas de forma manual, la cual se torna en una tarea tediosa, propensa a

errores y que consume tiempo, más aún si el número de servidores MySQL del modelo es alto, con 15, 20, 25 o más servidores.

Este trabajo de investigación propone el desarrollo de la herramienta MySQL Replication Modeling que permita a los DBA's modelar la replicación de MySQL y validar automáticamente si el modelo es correcto, mostrando los errores en caso existan.

Además, una vez que el DBA ha corregido y validado el modelo, la herramienta es capaz de generar automáticamente los comandos mysqlreplicate de configuración. La herramienta se desarrolló siguiendo las fases del Proceso Unificado de Rational (RUP) y basada en la Ingeniería Dirigida por Modelos (MDE) bajo la plataforma Eclipse.

Los resultados demuestran que la herramienta propuesta MySQL Replication Modeling permite validar automáticamente un modelo de replicación de MySQL, y reduce en más del 87% el tiempo en identificar y corregir los errores de un modelo de replicación con 25 servidores comparado con el uso de la herramienta Microsoft Visio 2013. Los resultados también demuestran que con el uso de la herramienta propuesta se reduce en más del 99% el tiempo para generar los comandos mysqlreplicate de configuración comparado con el uso de la herramienta Microsoft Visio 2013".

2.1.3. TRABAJO 03

Autores:

- Beltrán Morales, Jefferson Tarsicio
- Alemán Fierro, Jonathan Sebastián

Fecha: 2016 – Quito Ecuador

Nombre del proyecto:

"Diseño de una metodología para clúster de base de datos Oracle MYSQL de alta disponibilidad, con un demo de aplicación en servidores Linux."

Resumen:

La implementación de MySQL Clúster asegura que: "la continuidad del

servicio de almacenamiento para cualquier tipo de aplicación transaccional y no transaccional, ya que, al ser un sistema de clustering, está diseñado para eliminar un punto único de falla. En el presente proyecto de investigación se muestran los principios teóricos para la implementación de un clúster de alta disponibilidad de base de datos, se establece una metodología para el dimensionamiento inicial del clúster en cuanto a características generales de hardware y software, y se demuestra el funcionamiento de esta tecnología con el uso de componentes libres (open-source): MySQL, Linux CentOS, MySQL Storage Engine NDB".

2.2. Marco Teórico

2.2.1. Base de Datos

2.2.1.1. Concepto

Marqués, M (2011). “Una base de datos es un conjunto de datos almacenados en memoria externa que están organizados mediante una estructura de datos. Cada base de datos ha sido diseñada para satisfacer los requisitos de información de una empresa u otro tipo de organización…”

Podemos decir también que una base de datos es un gran almacén de datos cuyo objetivo es almacenar grandes volúmenes de información de forma ordenada con diferentes propósitos y usos y tiene como principal propósito organizar y almacenar datos para su fácil manejo y acceso desde distintos puntos de acceso y con nueva tecnología desde distintos lugares geográficos con total transparencia y facilidad.

Durante años la forma de almacenar información era mediante papel pero cómo van los tiempos esto ya quedo atrás porque ya se necesita manejar grandes volúmenes de información y en tiempo de respuestas rápidas que nos dará la ventaja competitiva.

Una base de datos es una gran herramienta que es usada para lograr recopilar o almacenar grandes volúmenes de información y mantenerla organizada en forma rápida y sencilla. Las bases de datos pueden almacenar todo tipo de información o datos como por ejemplo de personas, productos, pedidos o cualquier dato de objetos que existen. Muchas de nuestras bases de datos sean iniciado con una simple lista en una hoja de papel o de cálculo o en algún aplicación de procesamiento de textos. A medida que nuestra lista va en aumento en su tamaño, empiezan a aparecer algunas redundancias e inconsistencias en los datos almacenados. Cada vez se hace más difícil lograr administrar volviéndose engorros el orden, y la extracción de datos en nuestra lista y ni

que hablar de la forma de respaldo de los datos en forma de. Una vez que estos problemas se detectan y comienzan a mostrarse, una solución es transferir los datos a una base de datos creada y administrada con un sistema de administración de bases de datos (DBMS).

Hay que tener en cuenta que no solo es guardar datos y almacenaros lo más seguro si no que nos dice Anguiano (2014) en su artículo “El almacenamiento de la información por sí sola no tiene un valor, pero si combinamos o relacionamos la información con diferentes departamentos nos puede dar valor.” Los datos hay procesarlos para darnos valor de utilidad sino siguen siendo simple bytes de información.

Ilustración 1: Base de datos

Fuente: https://www.ymant.com/blog/tipos-base-datos

2.2.1.2. Tipos de Base de Datos

Existen muchas empresas con diferentes giros y dependiendo del giro será el tipo de procesamiento que se le dará a la información, esto determinará el tipo de base de datos a utilizar. Existen diferentes tipos de bases de datos pero las más comunes son las OLTP y OLAP.

Las bases de datos de tipo OLTP (On Line Transaction Processing) también son llamadas bases de datos dinámicas lo que significa que la información se modifica en tiempo real, es decir, se insertan, se eliminan, se modifican y se consultan datos en línea durante la operación del sistema. Estas bases

de datos son las más usadas en los sistemas transaccionales. Las bases de datos de tipo OLAP (On Line Analytical Processing) también son llamadas bases de datos estáticas lo que significa que la información en tiempo real no es afectada, es decir, no se insertan, no se eliminan y tampoco se modifican datos; solo se realizan consultas sobre los datos ya existentes para el análisis y toma de decisiones. Este tipo de bases de datos son implementadas en Business Intelligence para mejorar el desempeño de las consultas con grandes volúmenes de información.

2.2.1.3. Modelos de Base de Datos

En este amplio mundo de la tecnologia existen diversas formas de ordenar y organizar la información para que sea accesible desde cualquier dispositivo y sea útil para una organización. Aunque de esta variedad de tipos de manejado es de datos no existe el sistema de base de datos que sea 100% perfecto es por eso es que hay que elegir aquel manejador de base de datos que mejor se adapte a nuestras necesidades. Los siguientes son los tipos más comunes:

- Las bases de datos en forma jerárquica.
- Las bases de datos en red que se derivan de las base de datos del tipo jerárquicas que logran mejorar la administración de los datos redundantes logrando un gran rendimiento al realizar consultas de datos.
- Las bases de datos transaccionales, son las más comunes y su función o finalidad esta derivada para el envío y recepción de datos las cuales se realizan a través de grandes velocidades y de forma continua. Su único objetivo es la recepción y envío de información. Este tipo de base de datos no ha sido diseñada para lograr la administración del almacenamiento o lograr la

redundancia, estas características están fuera de su propósito.

- Las bases de datos relacionales, son las más abundantes en las empresas, son las más utilizadas en el desarrollo de aplicaciones de procedimientos reales. La información que se genera se almacena siempre haciendo referencia o relación a otra por lo que se facilita la gestión, de ahí su nombre y es de fácil uso por cualquier usuario que no sea especialista o experto en manejadores de base de datos.
- Las bases de datos orientadas a objetos, son poco usadas y nacen de la necesidad o aparición de los sistemas de programación orientada a objetos. Los datos son almacenados en paquetes de objetos.
- Las bases de datos documentales u orientadas a documentos están especializadas en el almacenamiento de textos completos y manejan grandes volúmenes de datos, dentro de él puede contener diferentes tipos de datos.

2.2.1.4. Sistema de gestión de bases de datos relacionales (RDBMS)

Ilustración 2RDBMS

Fuente: *https://onlinetutorialhub.blogspot.com/2017/12/important-factors-of-rdbms.html*

Marqués, M (2011). “El sistema de gestión de la base de datos (en adelante SGBD) es una aplicación que permite a los usuarios definir, crear y mantener la base de datos, además de proporcionar un acceso controlado a la misma. Se denomina sistema de bases de datos al conjunto formado por la base de datos, el SGBD y los programas de aplicación que dan servicio a la empresa u organización”.

Como nos dice Marques en la cita anterior un sistema de gestión de bases de datos relacionales (RDBMS) es un software o programa que nos va a permitir realizar acciones sobre la base de datos como son los de crear, actualizar, eliminar es decir nos da la posibilidad de lograr administrar una base de datos del tipo relacional. La mayoría de los RDBMS comerciales utilizan el lenguaje de consultas estructuradas (SQL) para acceder a la base de datos.

Los principales productos en el mercado de los RDBMS son Oracle, DB2 de IBM y Microsoft SQL Server

Ahora porque usar una base de datos y un RDBMS, esto nos dice Date, C(2001) en su libro que una base de tos nos permite lograr la Compactación, es decir ya no existe la necesidad de tener grandes volúmenes de papeles almacenados en grandes cuartos y que con el tiempo se degeneren, otra es la velocidad de acceso y respuesta yaqué se usa tecnología que nos permite realizar esto mucho más rápido ya que se encuentra almacenado en orden, la velocidad es dada por la máquina que puede recuperar y actualizar datos más rápidamente que un humano, elimina ya el trabajo de llevar los documentos en la mano y trasladarlos de un lugar geográfico a otro.

A pesar de la gran variedad de modelos y soluciones comerciales, según Mora, A (2014) podemos enumerar una serie de funciones comunes a un gran número de SGBD:

• Nos permite lograr recuperar (leer) y grabar nueva

información o modificar la información de la base de datos de forma rápida y transparente en su funcionamiento para los usuarios.

• Usando un SGBS logramos garantizar la integridad de los datos, disminuyendo al mínima todo tipo de inconsistencias.

• Para lograr obtener los datos almacenados se puede hacer uso de un lenguaje de programación el que logra interaccionar entre el usuario y la base de datos para lograr obtener la información relevante para los procesos de la empresa.

• Logra proveer o almacenar las características lógicas de los datos conocido como metadatos o diccionario de datos.

• Permite gestionar los problemas derivados de la gran cantidad de accesos concurrentes a la información.

• Logra gestionar las transacciones, garantizando la unidad de varias instrucciones de escritura relacionadas entre sí.

• Todo SGBD incluye utilidades para realizar Backus y lograr una fácil restauración de las base de datos.

• Los SGBD proporcionan mecanismos de seguridad para lograr evitar los accesos no autorizados y logren manipular los datos en forma indebida.

A continuación mencionaremos algunos RDBMS mejor posicionados a julio del 2020 según el ranking mundial dado por DB-Enginas. (Julio 2020).

358 systems in ranking, July 2020

Rank Jul 2020	Rank Jun 2020	Rank Jul 2019	DBMS	Database Model	Score Jul 2020	Score Jun 2020	Score Jul 2019
1.	1.	1.	Oracle	Relational, Multi-model	1340.26	-3.33	+19.00
2.	2.	2.	MySQL	Relational, Multi-model	1268.51	-9.38	+38.99
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	1059.72	-7.59	-31.11
4.	4.	4.	PostgreSQL	Relational, Multi-model	527.00	+4.02	+43.73
5.	5.	5.	MongoDB	Document, Multi-model	443.48	+6.40	+33.55
6.	6.	6.	IBM Db2	Relational, Multi-model	163.17	+1.36	-10.97
7.	7.	7.	Elasticsearch	Search engine, Multi-model	151.59	+1.90	+2.77
8.	8.	8.	Redis	Key-value, Multi-model	150.05	+4.40	+5.78
9.	9.	↑ 11.	SQLite	Relational	127.45	+2.64	+2.82
10.	10.	10.	Cassandra	Wide column	121.09	+2.08	-5.91
11.	11.	↓ 9.	Microsoft Access	Relational	116.54	-0.64	-20.77
12.	12.	↑ 13.	MariaDB	Relational, Multi-model	91.13	+1.34	+6.69

Ilustración 3Ranking mundial a julio del 2020 de utilización de RDBMS

Fuente: https://db-engines.com/en/ranking

En la que se muestra en primer lugar al RDBMS Oracle y en

segundo lugar a Misal Server, siendo este el que vamos a usar por su popularidad y soporte a nivel mundial, adicionalmente por ser una RDBMS open source y también pertenece al grupo de Oracle.

Oracle.

"Oracle es un sistema de gestión de base de datos relacional (Relational Data Base Management System), desarrollado por la empresa Oracle Corporation". Se considera a este manejador de base de datos como uno de los sistemas de bases de datos mas completo y robusto que existe en el mercado, destacando como características principales:

- Gran soporte de transacciones
- Es muy estable.
- Tiene gran escalabilidad
- Funciona en múltiples plataformas.

Aunque su dominio en el mercado de los manejadores de base de datos implementados en servidores empresariales ha sido casi total en los últimos años, recientemente ha surgido una gran competencia por parte de Microsoft SQL Server de Microsoft y de la oferta de otros RDBMS que cuentan con licencia de software libre como PostgreSQL, MySQl o Firebird. Es por eso y siguiendo la corriente en el mercado que las últimas versiones de Oracle han sido certificadas para poder trabajar bajo GNU/Linux, pero la gran desventaja al igual que SQL Server, es el costo de la licencia del uso del software la cual desanima a muchos empresarios para su instalación.

Mysql.

"MySQL es un sistema de gestión de bases de datos relacionales de código abierto (RDBMS, por sus siglas en inglés) que usa el modelo conocido como cliente-servidor. RDBMS es un software o aplicación utilizado para crear y lograr

administrar bases de datos basadas en un modelo tipo relacional". Ahora, conozcamos algunos términos para entender mejor el concepto de Mysql:

"Código abierto es un termino que significa que eres libre de usarlo y modificarlo, esdecir que cualquier persona está autorizada para poder bajar e instalar el software. Nos permite aprender si tienes el conocimiento necesario para personalizar el código fuente y así adaptarlo mejor a tus necesidades. Sin embargo, la GPL (licencia pública de GNU) determina lo que puedes hacer según las condiciones. La versión con licencia comercial está disponible si necesitas una propiedad más flexible y un soporte avanzado.
Los equipos de cómputo que tienen instalado y ejecutan el software RDBMS se les conoce como clientes. Siempre que necesitan acceder a los datos, se conectan al servidor RDBMS". Esa es la parte "cliente-servidor". (Gustavo, 2019)

Microsoft SQL Server.

"Microsoft SQL Server es un sistema de gestión de base de datos relacional con gran aceptación en el mercado de los manejadores de base datos, fue desarrollado por la corporacion Microsoft. El lenguaje de desarrollo que es utilizado (por línea de comandos o mediante la interfaz gráfica de Management Studio) permite realizar la administración de la base de datos es llamado Transact-SQL (TSQL), una implementación del estándar ANSI del lenguaje SQL, que es utilizado para poder manipular y recuperar datos (DML), crear tablas y definir relaciones entre ellas (DDL)".
Entre sus características más importantes tenemos:

- Permite un gran soporte de transacciones.
- Pueden implementar procedimientos almacenados.
- Su instalación contiene un entorno gráfico que nos

permite la administración, logrando hacer el uso de comandos DML y DDL gráficamente.

- Trabaja en modo cliente-servidor, es decir la información y los datos se alojan en el servidor de base de datos y los puntos desde donde se acceden dicha información se les conoce como clientes.
- Además nos permite lograr administrar la información almacenada en otros servidores de datos logrando formar un cluster o granja de servidores.

Postgresql.

Es un sistema gestor de bases de datos relacionales, que está orientado a objetos, y una de sus principales características es es multiplataforma, es decir su código fuente funciona en múltiples tipos de sistemas operativos o tecnología.

Este gestor de base de datos fue desarrollado en el año 1996, gracias como en otros proyectos a una investigación militar de EEUU.

Postgresql maneja los elementos de las base de datos como si fueran objetos (orientado a objetos).

Se considera que un gestor de base de datos extensible, porque podemos ir agregando funcionalidades que no vienen instaladas por defecto en la instalación.

También lo definen con un software escalable y su potencialidad radica es que sus bases de datos pueden manejar grandes volúmenes de información, de más de 100 Terabytes y esto bajo una licencia de software libre, es decir que podemos usar este software para cualquier propósito sin ningún gasto o costo de licencia.

PostgreSQL esutiliado en diversos ámbitos, podemos daralgunos ejemplos donde se hace uso de este gestor de bases de datos:

- En tipo de almacenamiento conocidos como Dataware

house (DWH).

- Es usado en los servicios de grandes empresas como Amazon Web Services Redshift.
- Para realizar procesos de datos que están almacenados en la misma instancia y puedan conectarse con otras instancias.
- Es utilizado en sistemas complejos como los de información geográfica, para el servicio de mapas vía web.
- En bases de datos usados muy comúnmente para aplicaciones que se ejecutan bajo los entornos web.
- En CMS como Drupal o WordPress.

Algunos de los ejemplos más sobresalientes de empresas importantes a nivel mundial que lo utilizan como manejador de base de datos son por ejemplo Skype, que lo usan sus diferentes aplicativos, además de Telefónica, BBVA, Petrobras, el metro de Sao Paulo o el servicio de información meteorológica del Reino Unido. (Gil, 2018)

2.2.2. Sistema Operative GNU/Linux

Podemos tener en cuenta descripcion del Sistema operativo Linux tomando en cuenta la opinion de Barrios, Joel (2013) que nos lo describe como sigue:

> "GNU es un acrónimo recursivo que significa GNU No es Unix (GNU is Not Unix). Este proyecto fue iniciado por Richard Stallman y anunciado el 27 de septiembre de 1983, con el objetivo de crear un Sistema operativo completamente libre.
> GNU/Linux® es un poderoso y sumamente versátil sistema operativo con licencia libre y que implemente el estándar POSIX (acrónimo de Portable Operating System Interface, que se traduce como Interfaz de Sistema Operativo Portable). Fue creado en 1991 por Linus Torvalds, siendo entonces un estudiante de la Universidad de Helsinski, Finlandia. En 1992,

el núcleo Linux> fue combinado con el sistema GNU. El Sistema Operativo formado por esta combinación se conoce como GNU/Linux.

GNU/Linux es equipamiento lógico libre o Software Libre. Esto significa que el usuario tiene la libertad de redistribuir y modificar a de acuerdo a necesidades específicas, siempre que se incluya el código fuente, como lo indica la Licencia Pública General GNU (acrónimo de GNU is Not Unix), que es el modo que ha dispuesto la Free Software Foundation (Fundación de equipamiento lógico libre). Esto también incluye el derecho a poder instalar el núcleo de GNU/Linux® en cualquier número de ordenadores o equipos de cómputo que el usuario desee."

Distribuciones GNU/Linux para servidores

Red Hat

La distribucion Red Hat es considerado uno de los sistemas operativos más fiables usados por las empresas a nivel mundial en entornos de produccion corporativos, es una distribución de Linux de código abierto que fue desarrollada por la empresa Red Hat, para uso exclusivo comercial. Este Sistema operativo se basa en el Sistema operativo Fedora. La comunidad de este software desarrolla software y es probado en el Sistema operativo Fedora antes de ser lanzado en la distribucion RED HAT.

Si usted esta pensando en instalar un Sistema Operativo Red Hat usted instalara un Sistema operativo muy poderoso, estable y seguro para lograr administrar los centro de informacion o de datos modernos y que administran grandes volumenes de informacion. Red Hat es muy usado para cloud, IoT, Big Data, visualización y contenedores por su grans soporte y estabilidad.

Las distros de RHEL son compatibles con los distintos tipos de tecnologias como x86, x86-64, Itanium, PowerPC e IBM System Z, es decir existe una distro para cada uno de ellas. Pero una de las

debilidades o si quieren verlo como una ventaja es que si necesitamos alguna actualizacion o parche del Sistema operative debemos suscribirnos, esta suscripción a Red Hat nos permite una membresia para lograr obtener el último software para la empresa, amplias bases de conocimiento, nos dala garantia de la seguridad del producto y nos asigna el soporte técnico del personal de la empresa RedHat.

Red Hat Enterprise Linux esta probada paradarnos una garantia de tiempo de actividad de 99,999%, especializada para dar soporte a trabajados con cargas criticas. Al utilizar este Sistema nos ayuda a reasignar recursos para lograr mantener un estado óptimo de los procesos en nuestros servidores y asi poder hacer frente a nuevos desafíos.

Ilustración 4Logo Red Hat

Ubuntu Server

Ubuntu Server, uno de los sitemas operatvos mas comunes para los usuarios finales de linux, pero en su version de servidor este no cuenta con un entorno gráfico, es decir su administracion o configuracion se realiza totalmente desde consola ya sea en forma directa en el servidor pero via remota con software como elmuy conocido putty. El manejo, configuracion y administracion del Sistema operativo Ubuntu Server es muy similar al de cualquier otro Sistema Linux, pero claro como cualquier otra distro con sus particularidades.

Ubuntu Server nos permite lograr una gran escalabilidad en nuestro centro de datos especialmente en la parte económica por su costo de instalacion de cero. Por ejemplo si se desea implementar una nube OpenStack, un clúster de Kubernetes o una granja de unos 50,000 nodos, usted ha elejido bien, el sistema operativo Ubuntu Server ofrece el mejor rendimiento de escalabilidad horizontal disponible.

Ilustración 5Logo Ubuntu

Centos

El Sistema operativo CentOS Linux es una distribución que es mantenida por la gran comunidad de desarrolladores de linux y es derivada de los paquetes Fuentes liberados al público por la empresa Red Hat para el Sistema operative Red Hat Enterprise Linux (RHEL). CentOS Linux es por eso que este Sistema operativo es completamente compatible con RHEL. El Proyecto CentOS principalmente realiza algunos cambios en los paquetes que lo conforman para lograr eliminar todas las marcas comerciales y licencias propias de Red Hat. La distribución de CentOS Linux es de libre descarga, instalacion y actalizacion es decir no hay que pagarlo. Cada versión de CentOS que es lanzada al Mercado es mantenida por 10 años (por medios de actualizaciones de seguridad -- la duración que dan los creadores de esta ditribucion son en intervalos de mantenimiento que han ido variando a lo largo del tiempo con respeto a la relación de los paquetes fuentes liberados). Las versiones nuevas de CentOS es lanzada al Mercado en aproximadamente cada 2 años y cada versión de CentOS es actualizada periódicamente para incorporar

nuevos drivers de los Nuevo hardware que sin lanzados en el mercado. Entonces al instalar Centos en nuestra organizacion logramos obtener un entorno Linux seguro, de bajo costo mantenimiento, muy confiable, predecible y reproducible.

Ilustración 6: Logo Centos

2.2.3. Cluster

Un cluster es una unión de nodos independientes unidos por una infraestructura de red para procesar funciones en forma paralela trabajando como si fuese uno. A cada uno de estos servidores conectados al cluster se le denomina nodo. Otros componentes de un cluster son los Sistemas operativos que se encargan de las carcateristicas basicas de los procesos, debe ser robusto y fácil de usar, Conexión de Red: un componente muy importante para facilitar la transmisión sin retardos o latencia, hoy en día es mínimo a transmisiones de Ethernet Gigabit.

Que es Middleware: gracias a este software podemos lograr realizar operaciones de balanceo de carga, tolerancia de fallos, etc. el se ocupa de detector nuevos nodos que vayamos añadiendo al cluster de servidores, logrando asi obtener una gran escalabilidad de nuestro centrode datos.

Sistema de almacenamiento: cuando trabajamos con clusters podemos realizar o configurar sistemas de almacenamiento tanto interno como

externos, es decir por ejemplo interno en los servidores, utilizando los discos duros de manera similar a como lo hacemos como concualquier equipo de computo de esxritorio o bien recurrir a sistemas de almacenamiento más complejos de tipo externos o de red, que nospermiten proporcionar mayor eficiencia y disponibilidad de los datos, un ejemplo claro de este tipo de almacenamiento externos tenemos a los dispositivos NAS (Network Attaches Storage) y a las redes SAN (Storage Area Network).

2.2.3.1. Tipos de Cluster.

Ya que entendimos que es un cluster ahora veremos los tres tipos de cluster aconfigurar segun las necesidades de las empresas:

Clúster de alto rendimiento (HC Performance Clusters)

Son clústeres en los cuales se ejecutan tareas que requieren de una gran capacidad computacional o de procesador y de grandes cantidades de memoria, o ambos a la vez. El llevar a cabo estas tareas logran comprometer los recursos del clúster por largos periodos de tiempo.

Clústers de alta disponibilidad (HA o High Availability)

Son clústeres cuyo objetivo principal de diseño es de lograr proveer que los servidores sean confiables y esten disponibles ante cualquier falla ya sea de hardware o software. Estos clústeres logran brindar la máxima disponibilidad de los servicios que ofrecen. La confiabilidad que brindan estos cluster se da mediante software que permite detector fallos y poder recuperarse frente a los mismos, mientras que en hardware se evita tener un único punto de fallos.

Clúster de Alta eficiencia (HT o High Throughout)

Son clústeres cuyo objetivo de diseño es lograr ejecutar la mayor cantidad de procesos o tareas en el menor tiempo posible. El retardo entre los nodos del clúster no es considerado un gran problema.

2.2.4. Alta Disponibilidad

Los sistemas de alta disponibilidad garantizan los servicios de forma que si el sistema cae, sea reemplazado de forma automática.

La alta disponibilidad se basa en la duplicación del hardware, de esta forma, en caso de que algo falle en un sistema, tendremos otro servidor idéntico que lo sustituirá automáticamente, garantizando así la continuidad del servicio.

Imaginemos que tenemos un Servidor en el que se almacenan los datos del programa de gestión de una empresa. Tras un fallo en su funcionamiento, es necesario sustituir un componente de dicho sistema. El componente puede tardar mínimo 24h en recibirse del proveedor en caso de que la tenga en stock, por lo tanto el tiempo que pasa hasta que se soluciona el problema es muy alto. En este momento, la empresa queda parada y no puede continuar su actividad y las pérdidas económicas generadas por este parón pueden ser considerables.

Con alta disponibilidad: highavaiable

El servidor de almacenamiento dispone de una máquina de las mismas características que está totalmente sincronizada y simplemente el sistema sigue funcionando con normalidad. El tiempo de respuesta de dicho sistema son minutos.

Un protocolo de alta disponibilidad, conocido por sus siglas en ingles HA (High Availability) se aplica cuando queremos tener un plan de contingencia sobre cualquier componente que tenga alguna situación anómala para poder seguir dando el servicio del mismo.

Sistemas de alta disponibilidad

Actualmente los sistemas de alta disponibilidad cuentan con dos tecnologías:

Alta disponibilidad Activo - Pasivo: Este tipo de tecnología consta de dos servidores idénticos conectados y configurados en la misma red, donde uno de ellos activo es decir esta ofreciendo los servicios y existe

un segundo servidor idéntico en servicios, llamado pasivo, que está en contínua sincronización a la espera de una conmutación por algun error o por alguna tarea programada.

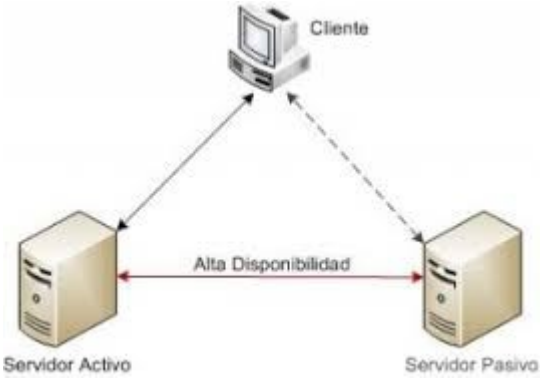


Ilustración 7: Alta disponibilidad activo-pasivo

Fuente: https://www.ecured.cu/Cluster_de_alta_disponibilidad

Alta disponibilidad Activo - Activo: Este tipo de tecnología consta de un servidor en el cual todos sus servicios está todo duplicado(existe otro servidor con los mismos servicios) y funcionando en forma simultanea quien toma la tarea elque esta mas libre de carga.. En caso de un problema de uno de los sistemas internos de uno de los servidores todo el control de la carga pasa uno de los servidores.

Ilustración 8: Alta disponibilidad activo-activo

Fuente: https://www.ecured.cu/Cluster_de_alta_disponibilidad

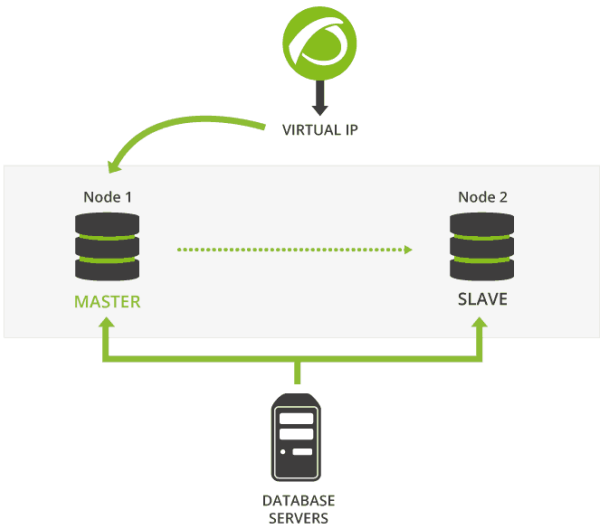


Ilustración 9: Modelo de base de datos activo-pasivo

Fuente: https://pandorafms.com/blog/es/base-de-datos-de-alta-disponibilidad/

2.2.5. Certificaciones de Centro de Datos

Podemos clasificar los data center, gracias a los estándares TIA-942 que incluyen información sobre los grados de disponibilidad (TIER).

Los Tiers se basan en información desarrollada por el Uptime Institute, un consorcio dedicado a promover las mejores prácticas para la planificación y gestión de centros de datos.

Para cada uno de los Tiers se detallan algunas recomendaciones para la infraestructura de seguridad, eléctrica y mecánica y telecomunicaciones. Cuando mayor Tier disponemos, mayor grado de disponibilidad.

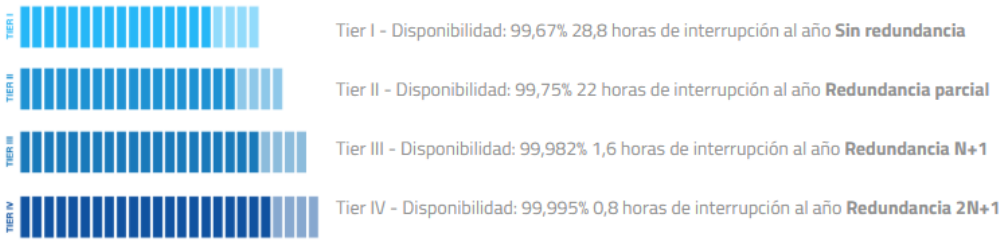


Ilustración 10: Niveles TIER

Fuente: https://bekkos.com/clientes-1/blog-bekkos/122-servidores-los-niveles-tier-3-y-tier-4

Tier I: Básico

El diseño de un Tier 1 permite admitir interrupciones planeadas y no planeadas. Se debe de disponer de sistemas de aire acondicionado y un sistema de distribución de energía, pero no suelen tener: suelo técnico, sistema de energía ininterrumpida o conocida como UPS o generados eléctricos.

Si implementamos este diseño, puede tener varios puntos de fallo, sobre todo cuando se llega a la carga máxima en situaciones críticas. Puede ocurrir también en este diseño errores de operación de los servidores o fallos en la infraestructura lo que provocaría que el centro de datos se interrumpiera en su funcionamiento causando un daño muy grande en la empresa en la cual esta instalada.

Además este diseño de infraestructura del data center prevee que este deberá en forma obligatoria estar fuera de servicio por lo menos una vez al año para su respectivo mantenimiento o reparación.

La tasa máxima de disponibilidad del CPD es 99.671% del tiempo.

Tier II: Componentes Redundantes

En este diseño los Data Centers usan componentes redundantes que permite asegurar ligeramente se vuelvan menos susceptibles a interrupciones, tanto las planeadas como las no planeadas. Estos CPDs deben de contener un suelo técnico, equipos UPS y generadores eléctricos que garanticen la continuidad del servicio, pero el problemas es que contempla que debe estar conectado a una sola línea o sistema de alimentación de distribución eléctrica. Su diseño es (N+1), que significa esto, es que existe al menos un duplicado de hardware y software de cada componente de la infraestructura. La carga máxima de los sistemas en situaciones críticas es del 100%. Y como se menciono anteriormente si se realiza un mantenimiento de la línea de la distribución eléctrica o en otros componentes de la infraestructura, pueden causar una gran interrupción del serviciode datos.

La tasa de disponibilidad máxima del CPD es 99.741% del tiempo.

Tier III: Mantenimiento Concurrente

Las capacidades de su Data Center Tier III, nos garantiza que podamos realizar cualquier actividad planeada sobre nuestro centro de datos, es decir sobre cualquier componente sin tener que detener algún componente o tener una interrupción en la operación.

Las actividades que se pueden planear incluyen:

- El Mantenimiento preventivodelos equipos.
- Raparación y reemplazamiento de componentes del data center.
- Agregar o eliminar los componentes en el data center.
- Realizar pruebas en sistemas o subsitemasen el data center.

Si queremos implementar este diseño de data center, debe de existir como punto principal una gran capacidad necesaria pararealizar las tareas y una doble línea de distribución para los componentes. Asi es posible realizar pruebas correctivas o preventivas mientras la otra línea atiende la totalidad de la carga.

Si siempre tiene el caso que se den actividades no planeadas como los errores de alguna operación en los componente o fallos espontáneos dentro de la infraestructura los que pueden causar daño en su centro de procesamiento de datos escoja este diseño. La carga máxima para situaciones críticas es de 90%.

Con la aplicación del Tier III, esta a un paso de certificar a Tier IV. La mayoría de los Data Centers Tier III son diseñadas para actualizarse a Tier IV, este cambio se produce si se logran aumentar las necesidades o requerimientos tecnológicos en la empresa.

La tasa máxima que se exige es de disponibilidad del CPD es 99.982% del tiempo ininterrumpido.

Tier IV: Tolerante a Fallos

Lograr un diseño en data center con este nivel deberá proveer la capacidad para realizar cualquier tipo de actividad como de mantenimiento correctivo o preventivo sin tener alguna interrupción en los servicios. Además se tiene una tolerancia a fallos que le permiten obtener que la infraestructura de su data center para continuar operando ante una actividad no planeada.

Para ello este sistema o diseño principalmente como requerimiento solicita que se logre instalar dos líneas de distribución en forma simultanea y que estén activas, típicamente en una configuración System+System. La carga máxima en esta situación crítica es del 90%. Persiste un nivel de exposición a fallos ya que es avisada por una alarma de incendio.

La tasa de disponibilidad máxima del CPD es 99.995% del tiempo

Capítulo III:
MARCO METODOLÓGICO

CAPÍTULO III: MARCO METODOLÓGICO

3.1. Variables.

Las variables de investigación que intervienen en el presente proyecto son:

a) **Variable Independiente (Causa):**

Implementacion de un Servidor Linux.

b) **Variable dependiente (efecto):**

Arquitectura de Base de datos en Alta disponibilidad.

c) **Solucion:**

Implementacion de una arquitectura de alta disponibilidad de base de datos en servidores Linux para la Gerencia Regional de Trabajo y Promocion del Empleo Lambayeque

3.2. Operacionalización de las variables:

Variable	Definición conceptual	Dimensiones	Indicadores	Tecnica / Instrumento
Implementacion de un Servidor Linux	Se instalara servidor con el sistema operativo Ubuntu Server 20.10 por su versatilidad y fácil configuración y soporte.	• Hardware • Software	• Costo • Facil mantenimiento • Soporte en linea	• Manual de instalacion • Instaladores (ISO) • Textos y foros en linea
Arquitectura de Base de datos en Alta disponibilidad	En una configuración de HA, se configuraran tres nodos un master en modo activo y dos esclavos en modo pasivo, cuando falle el master entran en acción no de los esclavos.	• Servicio de base de datos • Alta disponibilidad	• Calidad del Servicio • Disponibilidad del servicio • Tiempo de Respuesta ante caídas del sistema	• Encuesta a usuarios del servicio • Reporte de caídas del Servidor • Vitacoras de eventos de los servicios.

Tabla 1Operacionalización de las variables

Fuente: propia autoría

3.3. Metodología

3.3.1. Tipo de estudio:

Investigación aplicada

3.3.2. Diseño de investigación

- Definir las variables dependiente e independiente usadas en el presente proyecto.
- Formular el problema del presente proyecto de investigación.
- Lograr definir la hipótesis
- Plantear los objetivos
- Redactar el título del informe.
- Redactar el marco metodológico.

3.4. Población, muestra y muestreo.

a. Población:

Primera población: Personal que labora en las áreas de la Gerencia Regional.

Segunda población: Personal de la Oficina del Centro de Sistemas (solo cuenta con una persona encargada)

b. Muestra Satisfaccion del Servicio:

Para lograr determinar la muestra de la primera semana se aplica la siguiente fórmula:

n: Tamaño de la muestra

PQ: Varianza =0.25

N: Población

E: Margen de error 6%

K: Constante de corrección del error =2

$$n = \frac{PQN}{(N-1)\frac{(E)^2}{(K)^2} + PQ}$$

Para aplicar la formula se tomaran los siguientes datos:

N: Población (Trabajadores de la Gerencia) = 42

Lo cual aplicando la formulanos queda como sigue:

.n =10.5/0.2869

Donde n= 37 la muestra.

C. Muestreo

Muestreo no probalístico, se usa este tipo de muestras por que se va a elegir a los usaurios utilizando diferentes criterios relacionadas con las características de la investigación, no tienen la misma probabilidad de ser seleccionados.

D. Muestra: AplicacionTecnologica

Donde n= 01 persona

3.5. Técnicas e instrumentos de recolección de datos.

Para el análisis documental se emplearán las fichas textuales y fichas de resumen y se consideraran para lograr este proyecto como fuentes los informes y documentos.

3.6. **Encuesta**. Se empleará un cuestionario, que será aplicada a la cantidad de usuario como indica la muestra obtenida anteriormente para obtener la satisfacción del usuario en el uso de los sistemas en alta disponibilidad.

3.7. Métodos de análisis de datos

Se utilizó un cuestionario de 05 preguntas, nos permitio determinar o encontrar el conocimiento que tienen y la confianza qe tienen los trabajadores que son los usuarios de nuestro proyecto y otro de 05 preguntas al encargado del Centro de Sistemas de Informacion para ver funcionamiento de los sistemas antes y después de la implemenacion del proyecto.

3.8. Aspectos éticos

Se tuvo en cuenta la honestidad en los resultados de la encuesta, así como el desarrollo del proyecto considerando el contexto actual de la empresa.

Capítulo IV:

RESULTADOS

CAPITULO IV: RESULTADOS

A continuación se tabula los resultados de la encuesta realizada a la muestra del personal, esto referido a la calidad del servicio que ha tenido en el primer semestre sin la utilizacion del servicio de una base de datos de alta disponibilidad.

1. ¿Ha tenido usted problemas en el acceso a los datos de los sistemas informáticos que usted usa en su trabajo diario durante el último semestre?

RESPUESTAS	CANTIDAD	PORCENTAJE
SI	32	86%
NO	5	14%
TOTAL	37	100%

Tabla 2: Problemas por falla en los sitemas

Elaboración : Propia

2. Que tiempo en horas maximo por dia de trabajo ha tenido que suspender sus labores por falla en el acceso de los datos de los sistemas informáticos.

NRO TRABAJADORES	NRO HORAS PARALIZADAS	PORCENTAJE EN BASE A 08 HORAS
8	3	37.50 %
2	8	100.00 %
6	4	50.00 %
7	1	12.50 %
10	2	25.00 %
4	5	62.50 %
Total dia	23	

Tabla 3: Numero de horas por falla en los sistemas

Elaboración : Propia

3. Esta usted de acuerdo en la implementacion de tecnoligas que permitan la continuidad de los sistemas informáticos ante problemas tecnologicos naturales?

RESPUESTAS	CANTIDAD	PORCENTAJE
SI	36	97.00%

NO	1	3.00 %
TOTAL	37	100%

Tabla 4: Implementacion de nuevas tecnologias

Elaboración : Propia

4. Cuando se implemente una nueva tecnología en mejora de los sistemas de información a usted le gustaría participar en los cambios mediante capacitaciones o conocimiento básico de lo hecho o le gustaría sea transparente el uso para usted.

RESPUESTAS	CANTIDAD	PORCENTAJE
PARTICIPAR	2	5.00%
TRANSPARENTE	35	95.00 %
TOTAL	37	100%

Tabla 5: Participacion en la Implementacion de nuevas tecnologias

Elaboración : Propia

5. La calidad de los datos presentados por los sistemas después de las caídas de los sistemas son buenas en el último semestre.(datos no actualizados o inconsistentes)

RESPUESTAS	CANTIDAD	PORCENTAJE
SI	33	89.00%
NO	4	11.00 %
TOTAL	37	100.00 %

Tabla 6: Calidad de los datos despues de fallas

Elaboración : Propia

Encuesta al Personal del Centro de Sistemas.

1. Conoce usted de la existencia del software libre y su potencial en la implementación en los centros de datosy se siente capacitadao para el manejo de estas.

   RPTA: SI
2. Cual es el tiempo máximo de demora ante una falla de hardware que ha tenido usted para recuperar los sistemas hasta la adquisición de las piezas de repuesto.

RPTA: 01 semana en el peor de los casos, dos dias en promedio

3. Existen todas las condiciones en hardware para habilitar una arquitectura en alta disponibilidad.
   RPTA: SI
4. Usted esta dispuesto apoyar en la implementacion del sistema de alta disponibilidad y comprometerse en dar continuidad en el proyecto.
   RPTA: SI
5. Cree usted que la implementación de tecnologia en alta disponibilidad mejorara el funcionamiento de los sistemas y permitirá la mejora de respuesta en la recuperación y puesta en marcha de los servicios ante los inconvenientes tecnológicos o desastres.
   RPTA:SI

Capítulo V:
DISCUSIÓN

CAPITULO V: DISCUSIÓN

Ahora después de haber relizado las encuentas y haber realizado la tabulación de los datos procedemos a realizar las comparaciones finales para establecer la viabilidad delproyecto.

Para realizar esta discusión se tomo en cuenta las estaditicas del Inei para ver cortes de energía eléctrica promedio en el Peru por horas y del margen de error de los servidores por falla de la parte hardware.

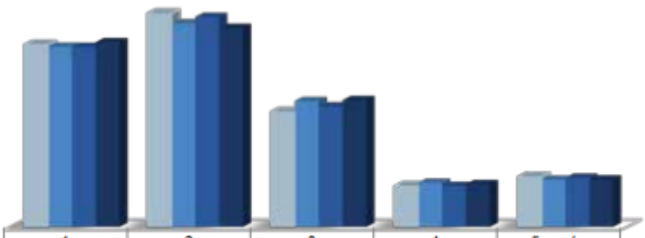
En los graficos siguientes se muestra el porcentaje de interrupciones por año, y tomaremos del año 2017 la que tiene menor variación porcentual para nuestro caso que es la de 04 interrupcciones (0.3) para nuestro proyecto.

Cuadro N° 1.4 Perú: Viviendas con servicio de energía eléctrica por red pública, por número de interrupciones o cortes en la energía eléctrica, 2014 - 2017
(Porcentaje)

Número de interrupciones o corte en la energía eléctrica	Año				Variación porcentual (2017 - 2016)
	2014	2015	2016	2017	
1	30,3	30,0	29,9	30,7	0,8
2	35,5	33,9	34,9	33,1	-1,8
3	19,1	20,9	20,1	21,1	1,0
4	6,8	7,4	6,8	7,1	0,3
5 a más	8,3	7,9	8,2	7,9	-0,3

Fuente: Instituto Nacional de Estadística e Informática – Encuesta Nacional de Programas Presupuestales 2014 - 2017.

Gráfico N° 1.4 Perú: Viviendas con servicio de energía eléctrica por red pública, por número de interrupciones o cortes en la energía eléctrica, 2014 - 2017
(Porcentaje)



	1	2	3	4	5 a más
2014	30,3	35,5	19,1	6,8	8,3
2015	30,0	33,9	20,9	7,4	7,9
2016	29,9	34,9	20,1	6,8	8,2
2017	30,7	33,1	21,1	7,1	7,9

Fuente: Instituto Nacional de Estadística e Informática – Encuesta Nacional de Programas Presupuestales 2014 - 2017.

Tabla 7: Numero de interrupciones electricas en Peru

Fuente:

https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1520/cap01.pdf

Ahora tenemos que en el año 2017, a nivel nacional, el tiempo promedio que duró la última interrupción o corte en el servicio de energía eléctrica por red pública en las viviendas, en el mes anterior a la encuesta, fue de 12 horas. Este dato lo tomaremos para realizar nuestro cálculo en nustra discusión.

Cuadro N° 1.5 Perú: Número de horas que duró la última interrupción o corte, según región natural, 2014 - 2017 (Porcentaje)

Región natural	Tiempo (Horas)				Variación porcentual (2017 - 2016)
	2014	2015	2016	2017	
Total	10	12	10	12	2
Costa	6	7	7	11	4
Sierra	15	15	12	14	2
Selva	10	19	9	9 a/	0

a/ Los resultados son considerados referenciales porque el número de casos en la muestra para este nivel no es suficiente y presentan un coeficiente de variación mayor al 15%.
Fuente: Instituto Nacional de Estadística e Informática – Encuesta Nacional de Programas Presupuestales 2014 - 2017.

Tabla 8: Promedio de duración de interrupción de energía electrica

Fuente:
https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1520/cap01.pdf

De los datos mostrados anteriormente podemos definir que el numero de horas por interrupción dela energía eléctrica se aproxima a :

Nro horas = 12 horas promedio * 04 interrupciones

Nro Horas = 48 horas de interrupciones

Ahora calcularemos el número de tiempo fuera de línea del servidor por falla del sistemas de base de datos esto sacado de la encuesta al encargado de la Oficina de Centro de Sistemas y datos del Inei.

Usaremos la formula:

$$Disponibilidad = \frac{(A - B)}{A} X\ 100\%$$

Donde:

$A = Horas\ comprometidas\ de\ disponibilidad$

$B = Número\ de\ horas\ \ fuera\ de\ línea$

Reemplazando los valores tenemos lo siguiente:

Horas por falla de hardware = 7 dias * 24 horas = 168 (peor escenario)

Horas por falla de hardware = 2 dias * 24 horas = 48 (promedio)

Horas de interrupción de energía = 48 horas

Recordar que la encuesta la tomamos por bimestre por eso se toma horas x semestre que es 4380 horas.

Peor Escenario:

$Disponibilidad$ = (4380 – 216) /4380 X 100 %

$Disponibilidad$ = 95,06849315 % (peor escenario)

Promedio de horas:

$Disponibilidad$ = (4380 – 96) /4380 X 100 %

$Disponibilidad$ = 97,80821917 % (peor escenario)

Ahora mostramos la tabla estándar para la medición de servidores en disponibilidad y tiempos permitidos fuera de línea.

Disponibilidad (%)	Tiempo offline/año	Tiempo offline/mes	Tiempo offline/día
90%	36.5 días	73 hrs	2.4 hrs
95%	18.3 días	36.5 hrs	1.2 hrs
98%	7.3 días	14.6 hrs	28.8 min
99%	3.7 días	7.3 hrs	14.4 min
99.5%	1.8 días	3.66 hrs	7.22 min
99.9%	8.8 hrs	43.8 min	1.46 min
99.95%	4.4 hrs	21.9 min	43.8 s
99.99%	52.6 min	4.4 min	8.6 s
99.999%	5.26 min	26.3 s	0.86 s
99.9999%	31.5 s	2.62 s	0.08 s

Tabla 9: Disponibilidad para un sistema 24×7 y tiempos de caída permitidos.

Fuente: Páginas eléctricas

Si ambos datos obtenidos lo comparamos con tabla de tiempos fuera de línea en ambos de los casos nos mantenemos en el rango del 95% no alcanzamos el superior del 98%, que es una situación critica para nuestro servicio brindado por la cantidad de horas de inactividad.

A continuación mostraremos los resultados aplicando nuestro proyecto que mejora en su totalidad la recuperación ante fallas.

Reemplazando los valores tenemos lo siguiente:

Horas por falla de hardware = 0 dias * 24 horas = 0 (peor escenario)

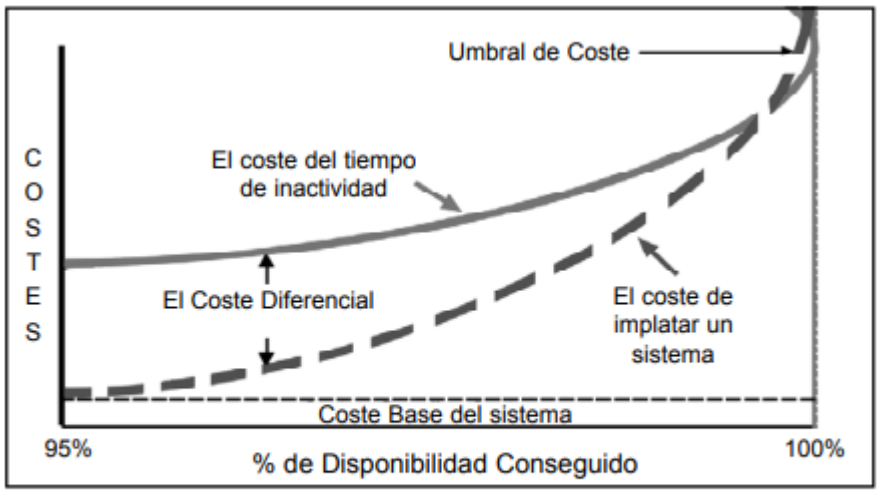
Horas de interrupción de energía = 48 horas

Cabe mencionar que estamos tomando 0 dias por que si falla uno de los equipos el otro entra a funcionar y la espera del repuesto no interrumpe el funcionamiento de todo el sistema.

$Disponibilidad$ = (4380 – 48) /4380 X 100 %

$Disponibilidad$ = **98,90410958 %**

Mostrando que ya subimos de nivel aplicando nuestro proyecto y estamos próximos a llegar al primer nivel de Certificacion Tier que seria Tier I, que nos solicita un 99.671 % de actividad del servidor.



El Coste/Beneficio de la Disponibilidad
Aquí se observa el diferencial entre el coste del tiempo de inactividad del sistema, y el coste para lograr ciertos niveles de disponibilidad del sistema.

Ilustración 11: Costo/Beneficio de la disponibilidad

Fuente: https://www.swgreenhouse.com/files/documents/continuidad-de-negocio/white-paper-los-fundamentos-de-la-disponibilidad-gestionada.pdf

Podemos tomar parte del texto de Akren, Ken. (2004) que encaja perfectamente con nuestros hallazgos:

> El gráfico del continuo costo/beneficio comienza con una disponibilidad presupuesta del 95%, y se acerca al 100% en la parte superior derecha. En general, el coste asociado con una

disponibilidad del 95% sólo difiere muy poco del coste total de referencia del sistema (hardware, software, servicios, etc.), porque es relativamente económico y más simple lograr una disponibilidad del 95%, la que equivale aproximadamente a 36 horas de tiempo de inactividad por mes. A medida que intentamos lograr niveles más altos de disponibilidad o de capacidad de reacción, pasando por tanto de prioridad de datos (reactiva a los requisitos empresariales) a prioridad de aplicativos (más proactiva a los requisitos empresariales), aumenta rápidamente el grado de sofisticación necesario para operar exitosamente a dichos niveles.

Ahora analizando las respuestas de las respuestas de las preguntas tenemos lo siguiente:

1. ¿Ha tenido usted problemas en el acceso a los datos de los sistemas informáticos que usted usa en su trabajo diario durante el último semestre?

   Podemos decir que de las 32 respuestas que dieron si, disminuirán en un gran alto valor porquelas fallas se limitaran solo por los cortes de energía eléctrica y ya no influyera las paradas por hardware.
2. Que tiempo en horas maximo por dia de trabajo ha tenido que suspender sus labores por falla en el acceso de los datos de los sistemas informáticos.

   Si sabemos que del total de horas paralizadas son 48 por corte de energía y máximo de 01 semana por hardware (168 horas),quiere decir que el porcentaje de corte de energía electrica corresponde al 22% del total y otros por hardware de 78%, si sacamos un porcentaje al cuadro tenemos

NRO TRABAJADORES	NRO HORAS PARALIZADAS	CON PROYECTO MINUTOS
8	3	39.6
2	8	105.6

6	4	52.8
7	1	13.2
10	2	26.4
4	5	66
Total dia	23	303.6

Ilustración 12: Numero de horas por falla en los sistemas con proyecto

Elaboración : Propia

El cual es una disminución muy justificatoria para nuestro proyecto.

3. Esta usted de acuerdo en la implementacion de tecnoligas que permitan la continuidad de los sistemas informáticos ante problemas tecnologicos naturales?

   Nos indica que el personal esta muy dispuesto a aceptar las mejoras tecnológicas en la institución.

4. Cuando se implemente una nueva tecnología en mejora de los sistemas de información a usted le gustaría participar en los cambios mediante capacitaciones o conocimiento básico de lo hecho o le gustaría sea transparente el uso para usted.

   De las respuestas anterior y y de esta pregunta podemos entender que el personal esta dispuesto a aceptar los cambios tecnológicos pero que no afecten su labor diaria, el cual lo estamos haciendo al realizar este proyecto ya que la solución del proyecto es toalmente transparente para ellos, solo se verá reflejado con las menores interrupciones en los accesos de los datos.

5. La calidad de los datos presentados por los sistemas después de las caídas de los sistemas son buenas en el último semestre.(datos no actualizados o inconsistentes)

   Esto se vera reflejado notablemente ya que se tendrán varias copias de los datos.

Encuesta al Personal del Centro de Sistemas.

De las respuestas dadas por el personal del Centro de Sistemas, el proyecto es viable ya que existe el conocimiento y la predisposición de este para implementar y dar continuidad al proyecto.

Por lo antes mencionado podemos decir que el proyecto es **viable**.

Capítulo VI:
CONCLUSIONES

CAPÍTULO VI: CONCLUSIONES

1. Se logro realizar un diagnostico de los sitemas con que cuenta la Gerencia, con sus Base de datos y tipo de datos.
2. Con las estadísticas y bitácoras de la sala de servidores se pudo determinar el número de fallas en los sistemas orígenes y causas.
3. Se logró implementar la base de datos Mysql Server en alta disponibilidad con sus requerimientos de harware y software.
4. Se realizó las validaciones de las pruebas de continuidad y buen funcionamiento del servidor.

Capítulo VII
RECOMENDACIONES

CAPÍTULO VII: RECOMENDACIONES

1. Capacitar al personal de Sistemas en cursos de certificacion en implementación de inafraestructuras de alta disponibilidad y redundancia,

2. Realizar pruebas de continuidad y contingencia bimestralmente en caso de caídas de los servidores por energía eléctrica, caída de discos y cambio de hardware, asi como caída total del servidor principal.

3. Realizar backups periodicos de las base de datos y lograr almacenarlas o resguardarlas en lugares que sean seguros y secretos en medios que nos garantice la calidad de los datos en el tiempo.

4. Realizar proyectos para lograr certificaciones TIER nivel I, asi mejorar la calidad de servicios brindados.

Capítulo VIII

PROPUESTA TECNOLÓGICA

CAPÍTULO VIII: PROPUESTA TECNOLÓGICA

8.1. Generalidades.

Actualmente, en la Gerencia no existe un plan de continuidad de los sistemas informáticos que respalde la caída de las base de datos lo cual paralizaría los servicios de atención a los ciudadanos siendo propenso a denuncias, motivo por el cual la implementación de servidores en alta disponibilidad de base de datos es una propuesta que dara solución a este tipo de problema, logrando obtener de esta manera una seguridad y confiabilidad en los sistemas informaticos para atender a los ciudadanos mediante herramientas de software libre, los mismos que estan incluidos dentro del nucleo de los Sistemas Operativos GNU/LINUX.

8.2. Análisis

8.2.1. Análisis de la situación actual de los servicios de la empresa

La Gerencia Regional cuenta con tres sistemas informáticos de uso diario mayormente para consultas para apoyo a los ciudadanos que son:

Sistemas	**Sistema informatico de Archivo**
Aplicación	Php
Base de Datos	Mysql
Area	Archivo Institucional
Personal	Solicitud de personas para realizar juicios laborales, lograr sus pensiones con tramite a ONP, pedido de expedientes por las áreas para ser remitidos a juzgados.

Sistemas	**Sistema informático de Multas**
Aplicación	Php
Base de Datos	Mysql

Area	Division de Administración
Personal	3
Descripcion	Solicitud por parte de las empresas para ver estado de sus multas, fraccionamientos y cálculos de cuotas.

Sistemas	**Sistema informático de Sindicatos**
Aplicación	Php
Base de Datos	Mysql
Área	Dirección de Prevención.
Personal	4
Descripcion	Por parte de los sindicatos y de la Policía Nacional para ver padrón de afilidos, se registra juntas directivas e inscripción de nuevos sindicatos.

Estos tres sistemas son cruciales para los ciudadanos, los cuales son de constantes accesos para consultas de los trámites de los ciudadanos. Cabe mencionar y como parte importante que estos sistemas son accedidos mas para lectura (Consultas y reportes) que ingreso lo que se tomara en cuenta para la decisión de la tecnología e infraestructura implementar.

8.2.2. Análisis actual de los servidores de la empresa

La institución cuenta con dos servidores en las cuales en uno de ellos corre el servicio web con sus aplicaciones locales conectadas a diversas base de datos, además de ser servidor de archivos para algunas áreas. No existe actualmente algún servidor que haga de controlador de dominio o haga de servidor para reutilizar equipos con bajos recursos de hardware. Se pudo encontrar el dato del 100% de equipos de cómputo el 30% de ellos son de bajos recursos de hardware que a pedido de los usuarios dejarían de funcionar lo cual acarrearía un gasto innecesario a la Gerencia Regional y afectaría su presupuesto.

Ilustración 13: Hp Proliant g6

Fuente: https://www.pccomponentes.com/hp-proliant-ml110-g6-x3430-5gb-2x250gb-kit

En la figuraanterior se muestra una imagen referencial delservidor usado como servidor de aplicaciones.

Ilustración 14 Hp Proliant G8

Fuente: https://kelsusit.com/hp-proliant-dl380p-g8-single-cpu-cto/

En la figura anterior se muestra una imagen referencial del servidor de archivos de la insttucion

8.3. Diseño

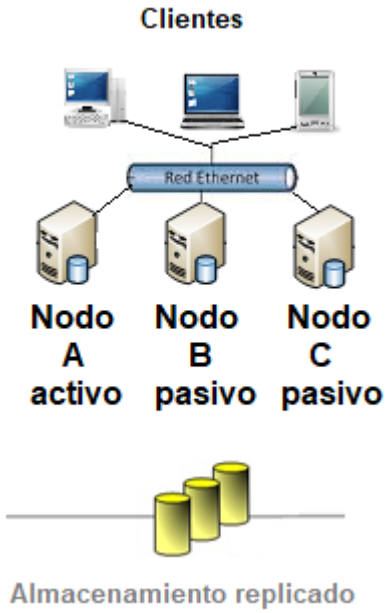


Ilustración 15Funcionamiento de una base de datos en alta disponibilidad

Fuente : Propia

Para la solución que se plantea en nuestro proyecto y luego de realizar el análisis mostrado anteriormente se plantea lograr un diseño de base de datos en alta disponibilidad implementando tres servidores uno principal y otros dos secundarios los cuales entraran en funcionamiento en caso de caída de servicio del principal, estos servidores deberán ser instalados y configurados haciendo uso del sistema operativos Linux. El diseño se muestra en la ilustración anterior.

Es cierto que existen dos formas de implementar uno es servidro activo-activo pero este es cuando existe un gran volumen de escritura por parte de los usuarios que deacuerdo al analisi anterior se descarta y se planteo en esquema Activo-pasivo.

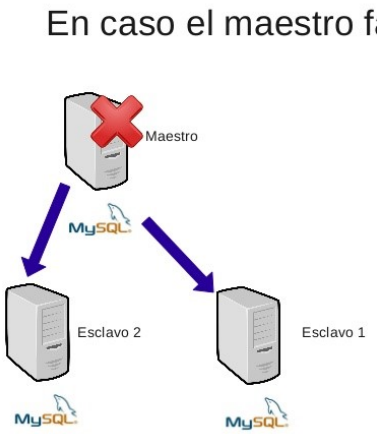


Ilustración 16Funcionamiento de una base de datos en alta disponibilidad en caso de fallo

Fuente : https://es.slideshare.net/denniscm20/alta-disponibilidad-con-mysql

8.4. Implementación

Para la implementación de nuestra solución utilizaremos el Sistema Operativo Ubuntu Server 20.10 e instalaremos la aplicación Mysql Server. En general utilizamos solo nodo del servidor de la base de datos para usos dentro de la LAN, pero si pensamos en aplicaciones que tienen miles de usuarios en línea a la vez, en esa situación, necesitamos una estructura que sea capaz de soportar esta carga y proporciona una alta disponibilidad con mayores gastos y ser una evaluación exhaustiva de las tecnologías en el mercado. Así que tenemos que agregar varios servidores de bases de datos interconectadas entre sí y mantenerlas sincronizada, para que en caso de que algún servidor se caiga, los otros servidores puede tomar su lugar y proporcionar los servicios a los usuarios.

- Un maestro, múltiples esclavos.
- En el maestro se escribe, en el esclavo se lee.

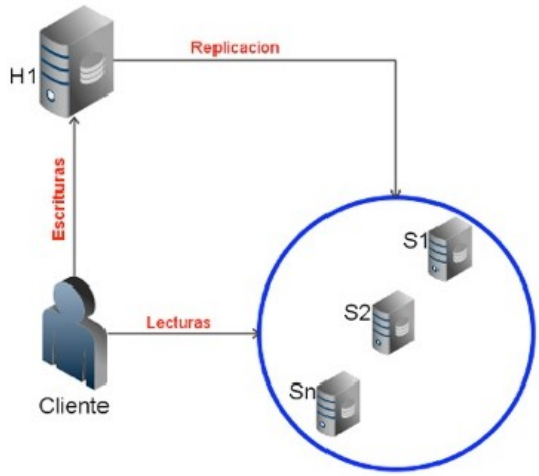


Ilustración 17 Implementación de arquitectura de replicación master-esclavo

Fuente: https://es.slideshare.net/miguelangelnieto/mysql-high-availavility-load-balacing-cluster

Como se menciono anteriormente se contara con tres servidores uno maestro y dos esclavos con la siguiente configuración:

Maestro
Ip : 192.168.0.5 /24
Nombre : nodoA

Esclavo 1
Ip : 192.168.0.6 /24
Nombre : nodoB

Esclavo 2
Ip : 192.168.0.7 /24
Nombre : nodoC

8.5. Instalación del equipamiento lógico necesario.

Para el proceso de instalación se seguirán los siguientes pasos:

- Instalación del sistema operativo Ubuntu Server 20 en los tres equipos que harán de servidores.

- Instalación de la base de datos Mysql Server y su configuración en Alta disponibilidad.
- Prueba de funcionamiento verificando replicas en las base de datos.

8.5.A. Instalación en Ubuntu Server.

- La primera pantalla nos pedirá el idioma, el cual seleccionaremos español

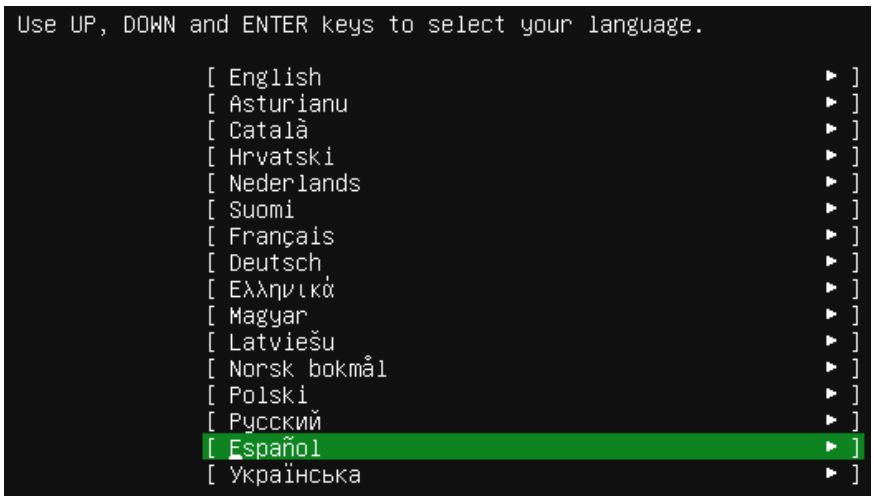


Ilustración 18: Seleccion de idioma

Fuente : Propia

- y después seleccionamos teclado latinoamericano
- luego configuramos la red de nuestro servidor

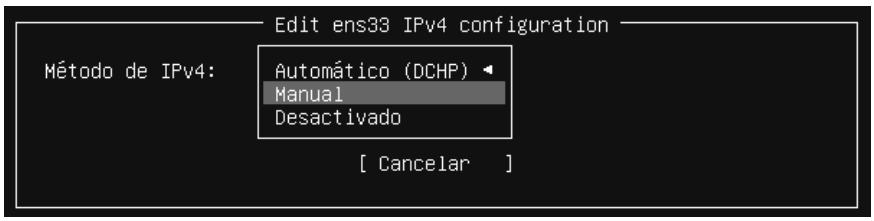


Ilustración 19: Configuracion de Red

- Ingresamos el IP para el servidor Master

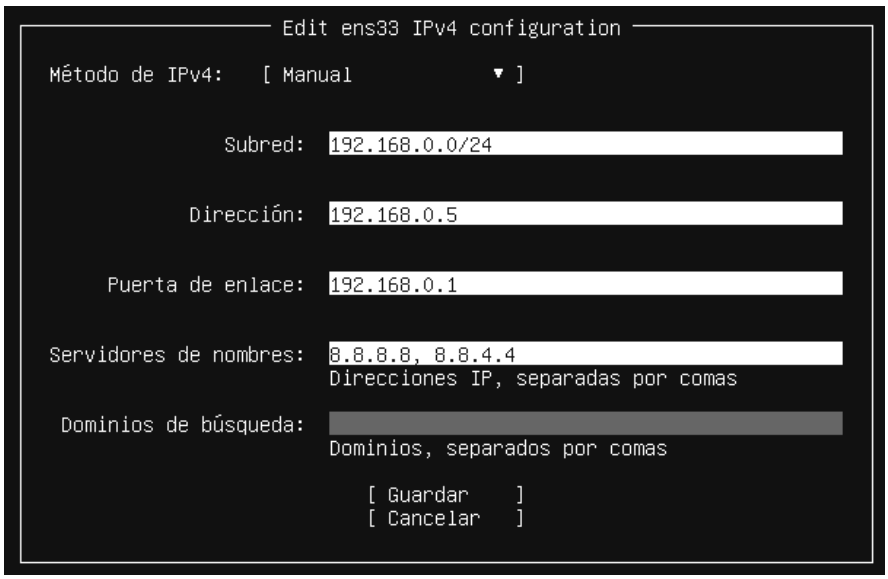


Ilustración 20: Datos Ip

Fuente : Propia

- Seleccionamos el particionamiento

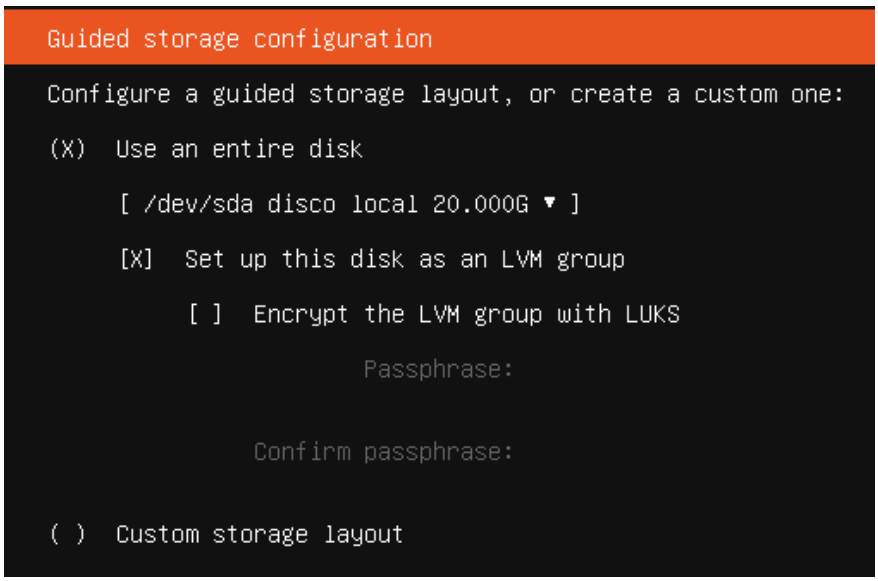


Ilustración 21: Particionamiento de disco

Fuente: Propia

- Seleccionamos perfil de usuario y servidor

Ilustración 22:Perfil de Usuario

Fuente:Propia

- Instalar ssh

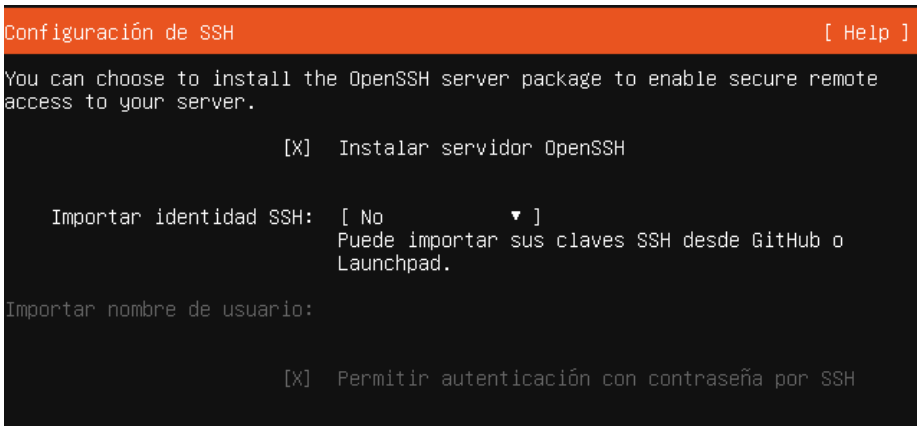


Ilustración 23:Instalacion ssh

Fuente : Propia

- Seleccionamos tipo de servicios, por el momento ninguno

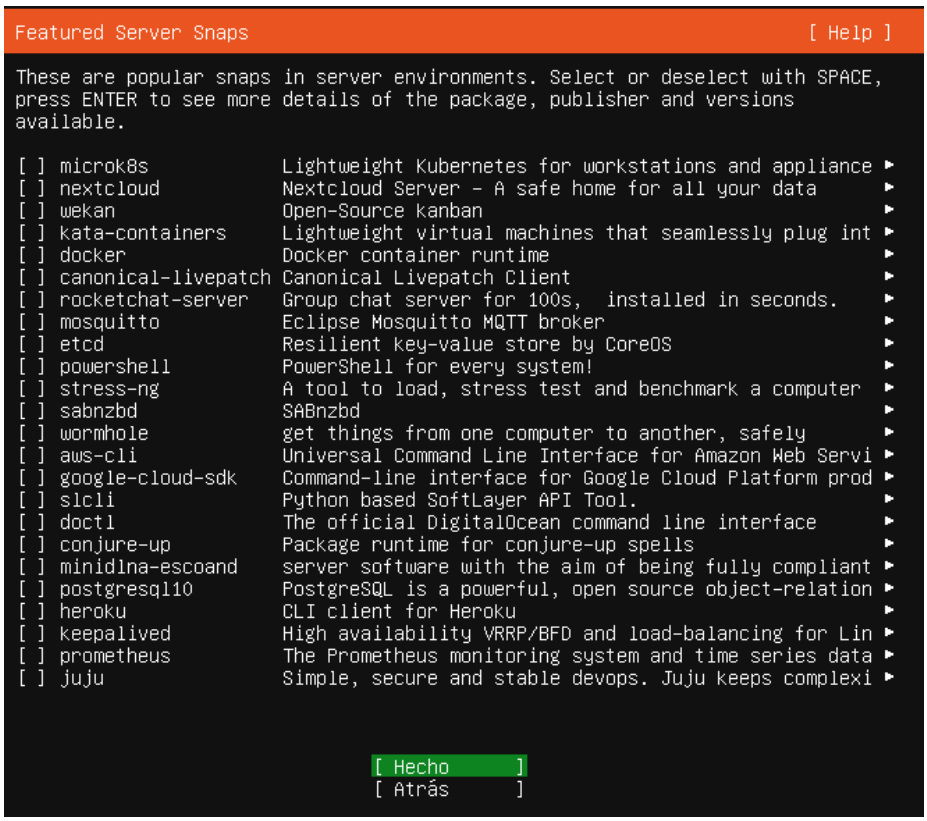


Ilustración 24: Lista de servicios

Fuente : Propia

Para el desarrollo del proyecto nos ayudaremos de la aplicación Galera el cual se será instalada en los tres servidores o nodos, los tres servidores del prsenete proyecto deberán tener la misma configuración y mismos componentes. Lo primero será instalar la clave para la clave de repositorio para la aplicación Galera.

#sudo apt-key adv --keyserver keyserver.ubuntu.com --recv BC19DDBA

Luego agregamos el archivo de repositorio para la aplicación de Galera dentro de la carpeta /etc/apt/sources.list.d/ en cada servidor:

#sudo nano /etc/apt/sources.list.d/galera.list

Agregar las siguientes lineas usando el comando

#nano /etc/apt/sources.list.d/galera.list

deb http://releases.galeracluster.com/mysql-wsrep-5.7/ubuntu bionic main

deb http://releases.galeracluster.com/galera-3/ubuntu bionic main

Cree otro nuevo archivo que lo llamaremos galera.pref ubicado dentro dentro del directorio /etc/apt/preferences.d/ de cada servidor:

#sudo nano /etc/apt/preferences.d/galera.pref

Ahora añada las siguientes líneas con su editor de texto:

/etc/apt/preferences.d/galera.pref

Prefer Codership repository

Package: *

Pin: origin releases.galeracluster.com

Pin-Priority: 1001

Luego se procede a actualizar

#sudo apt update

Ahora procedemos a Instalar MySQL en todos los servidores, ejecutamos el siguiente comando en los tres servidores para lograr instalar una instancia del manejador de base de datos.

#sudo apt install galera-3 mysql-wsrep-5.7

Desactive AppArmor ejecutando lo siguiente en cada servidor:

#sudo ln -s /etc/apparmor.d/usr.sbin.mysqld /etc/apparmor.d/disable/

#sudo apparmor_parser -R /etc/apparmor.d/usr.sbin.mysqld

Repita los siguientes pasos para sus otros dos servidores.

Ahora que ya se logro instalar MySQL server correctamente en cada uno de los tres servidores, se debe de continuar con el paso 02 que es realizar la configuración de la siguiente sección.

Configurar el primer nodo

Al llegar a este paso, se configurará el primer nodo. Cada nodo del nuestro clúster de servidores de base de datos tiene que tener una configuración casi idéntica. Se hará la configuración de la base de datos en su primera máquina o cluster y luego la copiará a los otros nodos.

De forma predeterminada, MySQL está configurado para leer las configuraciones en el directorio /etc/mysql/conf.d a fin de obtener configuración adicionales desde los archivos que terminan en la extension .cnf. ahora en su primer servidor, cree un archivo en este directorio con todas sus directivas específicas por clúster:

```
#sudo nano /etc/mysql/conf.d/galera.cnf
```

Añada la siguiente configuración al archivo

```
#nano /etc/mysql/conf.d/galera.cnf
[mysqld]
binlog_format=ROW
default-storage-engine=innodb
innodb_autoinc_lock_mode=2
bind-address=0.0.0.0

# Galera Provider Configuration
wsrep_on=ON
wsrep_provider=/usr/lib/galera/libgalera_smm.so
```

```
# Galera Cluster Configuration
wsrep_cluster_name="grtpe_cluster"
wsrep_cluster_address="gcomm://192.168.0.5,192.168.0.6,19
2.168.0.7 "

# Galera Synchronization Configuration
wsrep_sst_method=rsync

# Galera Node Configuration
wsrep_node_address="192.168.0.5"
wsrep_node_name="nodoa"
```

Ahora que configuró correctamente su primer nodo, puede configurar los nodos restantes en la siguiente sección.

Configurar los nodos restantes

Durante este paso, configurará los dos nodos restantes. En su segundo nodo, abra el archivo de configuración:

```
# nano /etc/mysql/conf.d/galera.cnf

[mysqld]
binlog_format=ROW
default-storage-engine=innodb
innodb_autoinc_lock_mode=2
bind-address=0.0.0.0

# Galera Provider Configuration
wsrep_on=ON
wsrep_provider=/usr/lib/galera/libgalera_smm.so

# Galera Cluster Configuration
wsrep_cluster_name="grtpe_cluster"
```

wsrep_cluster_address="gcomm://192.168.0.5,192.168.0.6,192.168.0.7 "

# Galera Synchronization Configuration
wsrep_sst_method=rsync

# Galera Node Configuration
wsrep_node_address="192.168.0.6"
wsrep_node_name="nodob"

Una vez que hallamos completado estos pasos, repítalos en el tercer nodo del cluster.

Abrir el firewall en todos los servidores y abra los puertos con los siguientes comandos:

sudo ufw allow 3306,4567,4568,4444/tcp
sudo ufw allow 4567/udp

Iniciar el clúster
Use la siguiente línea de comando en los tres servidores para lograr habilitar el servicio systemd de MySQL:
#sudo systemctl enable mysql

Despues de realizar en comando anterior usted verá el siguiente resultado, el que nos indica que el servicio se vinculó correctamente a la lista de servicios de arranque del servidor:

Output
Created symlink /etc/systemd/system/multi-user.target.wants/mysql.service → /lib/systemd/system/mysql.service.

Configure el primer nodo

Ahore ejecute el siguiente comando en el primer servidor:

#sudo mysqld_bootstrap

Con este comando no se mostrará un resultado tras una ejecución correcta. Cuando se realiza el comando con exito, el nodo se registra como parte del clúster y puede verlo con el siguiente comando:

#mysql -u root -p -e "SHOW STATUS LIKE 'wsrep_cluster_size'"

Después de ingresar los datos que nos solicita en nuestro caso la clave, verá el siguiente resultado, que indica que hay un nodo en el clúster:

Output

Variable_name	Value
wsrep_cluster_size	1

Tabla 10: Resultado de consulta numero nodos en nodo 1

En los otros dos nodos restantes, puede iniciar mysql de forma normal. Estos buscarán dentro de la red cualquier integrante de la lista de clústeres que esté en línea y, cuando lo encuentren, se unirán a el formando el cluster.

Activar el segundo nodo

Ahora, debe activar el segundo nodo. Inicie el servicio mysql:

#sudo systemctl start mysql

Ejecute el comando

#mysql -u root -p -e "SHOW STATUS LIKE 'wsrep_cluster_size'"

Verá el siguiente resultado después de realizar el comando anterior que indica que el segundo nodo se ha sumó al clúster y que hay dos nodos en total.

Output

Variable_name	Value
wsrep_cluster_size	2

Tabla 11: Resultado de consulta numero nodos en nodo 2

Active el tercer nodo

Ahora, es el momento de activar el tercer nodo. Inicie mysql:

#sudo systemctl start mysql

Ejecute el siguiente comando para encontrar el tamaño del clúster:

#mysql -u root -p -e "SHOW STATUS LIKE 'wsrep_cluster_size'"

Verá el siguiente resultado, que indica que el tercer nodo se sumó al clúster y que la cantidad total de nodos del clúster es tres.

Output

Variable_name	Value
wsrep_cluster_size	3

Tabla 12: Resultado de consulta numero nodos en nodo 3

Ahora en uno delos nodos usted puede realizar la creación de una base de datos:

#mysql -u root -p -e 'CREATE DATABASE testbd;

>CREATE TABLE testbd.usuario (id INT NOT NULL AUTO_INCREMENT, nombre varchar(50), apellido VARCHAR(25), PRIMARY KEY(id));

INSERT INTO testbd.usuario (nombre,apellido) VALUES ("Juan", "Suarez");'

Leer y escribir en el segundo y tercer nodo
A continuación, consulte el segundo nodo para verificar que la replicación funcione:

mysql -u root -p -e 'SELECT * FROM testbd.usuario;'
se vera el siguiente resultado en los demás nodos

Output

id	nombre	apellido
1	Juan	Suarez

Tabla 13: Resultado de consulta de la tabla usuario

De esta forma, habrá verificado correctamente funcionamiento del cluster de base de datos Mysql Server, y que los datos se están duplicando en los tres nodos.

8.6. Presupuesto

Recursos y presupuestos

- **Equipo de desarrollo**

Personal Tesista	TOTAL
Asesor	S/2,000.00
Responsable de Implementación	S/.1,000.00
Total	S/.3,000.00

Tabla 14 Costos equipo de desarrollo

- **Documentación**

DESCRIPCION	CANT.	TOTAL
FOTOCOPIAS	600	S/.240.00

IMPRESIONES	450	S/.380.00
ESPIRALADOS	5	S/. 80.00
Empastados	4	S/.100.00
Total		S/.800.00

Tabla 15 Costos Documentación

- **Otros**

DESCRIPCION	**CANT**	**PRECIO UNI.**	**TOTAL**
PASAJES	--	--	S/.450.00
PAQUETE DE DATOS	--	--	S/.220.00
TOTAL			S/.670.00

Tabla 16 Costos Otros

Resumen del Presupuesto:

- Equipo de desarrollo : S/. 3,000.00
- Documentación : S/. 800.00
- Otros : S/. 670.00

TOTAL : S/.4,470.00

Capítulo IX:
REFERENCIAS BIBLIOGRÁFICAS

CAPITULO IX: BIBLIOGRAFIA

Libros

- Barrios, Joel. (2013). Configuración De Servidores Con GNU/Linux (2013 edición). Mexico. Alcance Libre.

- Cabanes, Nacho. (2007). Introduccion a las Bases de Datos.
- Catherine, M.Ricardo.(2009). Bases de Datos (Primera edición). Mexico:McGRAW-HILL interamericana editores, s.a.

- Cortez, Stanley y Galarza, Monica. (2005). Cluster de Alta Disponibilidad para servidores web Linux (Informe para optar Titulo de Ingeniero). Cuenca Ecuador.

- Date,C. (2011). Introduccion a los sistemas de base de datos (séptima edición). Mexico: Pearson Educacion.
- Marqués, Mercedes. (2011). Bases de Datos. España: Castelló de la Plana: Publicacions de la Universitat Jaume. Servei de Comunicació i.
- Mora, Arturo. (2014). Bases de Datos: Diseño y Gestion. Madrid España: Editorial Sintesis.

Link

- DB-Engines. (July 2020). DB-Engines Ranking. Recuperado de https://db-engines.com/en/ranking

- Gil, J. G. (19 de Agosto de 2018). *openwebinars*. Recuperado el 1 de setiembre de 2019, de https://openwebinars.net/blog/caracteristicas-importantes-de-postgresql/
- Manterola,M y Curia, M. (26 de Noviembre de 2016). *Configuración y administración de un sistema GNU/Linux*. Recuperado de https://www.pdf-manual.es/redes/112-configuracion-y-administracion-de-un-sistema-gnulinux.html

- Soporte de Office. (2016). Conceptos básicos sobre bases de datos. EU.: Microsoft. Recuperado de https://support.microsoft.com/es-es/office/conceptos-b%C3%A1sicos-sobre-bases-de-datos-a849ac16-07c7-4a31-9948-3c8c94a7c204

- Wiki.Centos, M. (09 de Diciembre de 2019). *Centos*. Recuperado de https://wiki.centos.org/es

Articulo web

- Akren, Ken. (2004). Los Fundamentos de la Disponibilidad Gestionada. Recuperado de https://revistadigital.inesem.es/informatica-y-tics/los-gestores-de-bases-de-datos-mas-usados/
- Anguiano Morales, J. (03 de jumio, 2014). Características y tipos de bases de datos. Recuperado de https://developer.ibm.com/es/articles/tipos-bases-de-datos/

- Gustavo. (13 de Mayo de 2019). Recuperado el 01 de setiembre de 2019, de https://www.hostinger.es/tutoriales/que-es-mysql/

- Marin, Rafael. (16 de abril, 2019). Los gestores de bases de datos más usados en la actualidad. Recuperado de https://revistadigital.inesem.es/informatica-y-tics/los-gestores-de-bases-de-datos-mas-usados/

ANEXOS

Conexión de una aplicación en PHP al servidor nodoa

```
<?php
    // Conectando, seleccionando la base de datos
    $link = mysql_connect('172.16.40.125, ‘root', 'miclave')
                or die('No se pudo conectar: ' . mysql_error());
    echo 'conexión realizada!!!!´;
    mysql_select_db('testgrtpe') or die('No se pudo seleccionar la base de datos');

    // Realizar una consulta MySQL
    $query = 'SELECT * FROM usuario;
    $result = mysql_query($query) or die('Consulta fallida: ' . mysql_error());

    // Imprimir los resultados en HTML
    echo "<table>\n";
    while ($line = mysql_fetch_array($result, MYSQL_ASSOC)) {
        echo "\t<tr>\n";
        foreach ($line as $col_value) {
            echo "\t\t<td>$col_value</td>\n";
        }
        echo "\t</tr>\n";
    }
    echo "</table>\n";

    // Liberar resultados
    mysql_free_result($result);

    // Cerrar la conexión
    mysql_close($link);
?>
```

Conexión de una aplicación en PHP al servidor nodob

```
<?php
    // Conectando, seleccionando la base de datos
    $link = mysql_connect('172.16.40.126, ‘root', 'miclave')
                or die('No se pudo conectar: ' . mysql_error());
    echo 'conexión realizada!!!!´;
    mysql_select_db('testgrtpe') or die('No se pudo seleccionar la base de datos');

    // Realizar una consulta MySQL
    $query = 'SELECT * FROM usuario;
    $result = mysql_query($query) or die('Consulta fallida: ' . mysql_error());

    // Imprimir los resultados en HTML
    echo "<table>\n";
    while ($line = mysql_fetch_array($result, MYSQL_ASSOC)) {
        echo "\t<tr>\n";
        foreach ($line as $col_value) {
            echo "\t\t<td>$col_value</td>\n";
        }
        echo "\t</tr>\n";
    }
    echo "</table>\n";

    // Liberar resultados
    mysql_free_result($result);

    // Cerrar la conexión
    mysql_close($link);
?>
```

Conexión de una aplicación en PHP al servidor nodoc

```
<?php
    // Conectando, seleccionando la base de datos
    $link = mysql_connect('172.16.40.127, ‘root', 'miclave')
                or die('No se pudo conectar: ' . mysql_error());
    echo 'conexión realizada!!!!´;
    mysql_select_db('testgrtpe') or die('No se pudo seleccionar la base de datos');

    // Realizar una consulta MySQL
    $query = 'SELECT * FROM usuario;
    $result = mysql_query($query) or die('Consulta fallida: ' . mysql_error());

    // Imprimir los resultados en HTML
    echo "<table>\n";
    while ($line = mysql_fetch_array($result, MYSQL_ASSOC)) {
        echo "\t<tr>\n";
        foreach ($line as $col_value) {
            echo "\t\t<td>$col_value</td>\n";
        }
        echo "\t</tr>\n";
    }
    echo "</table>\n";

    // Liberar resultados
    mysql_free_result($result);

    // Cerrar la conexión
    mysql_close($link);
?>
```