



UNIVERSIDAD TÉCNICA DE BABAHOYO
FACULTAD DE ADMINISTRACIÓN, FINANZAS E INFORMÁTICA

PROCESO DE TITULACIÓN
NOVIEMBRE 2022- ABRIL 2023

EXAMEN COMPLEXIVO DE FIN DE GRADO O DE FIN DE CARRERA
PRUEBA PRÁCTICA

INGENIERÍA EN SISTEMAS
PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN SISTEMAS

TEMA:

ANÁLISIS COMPARATIVO ENTRE LOS SISTEMAS DE GESTIÓN DE BASE DE DATOS, CASSANDRA Y COUCHBASE, EN TÉRMINOS DE ESCALABILIDAD Y FLEXIBILIDAD DE DATOS.

EGRESADO:

CAMILO ANTONIO LUCIO MONTERO

TUTOR:

ING. ALCOSER CANTUÑA FABIAN EDUARDO

AÑO 2023

DEDICATORIA

Este trabajo se lo dedico con todo el amor del mundo a Dios por nunca dejarme y saberme guiar en todo momento por ser mi amigo y confidente. Por consiguiente, a mis padres por ser ellos el motor que me impulsa cada día a seguir a delante por darme fuerzas, por animarme, consolarme y darme todo su apoyo incondicional.

AGRADECIMIENTO

Agradezco inmensamente a DIOS por nunca dejarme y darme fuerzas para seguir adelante por guiarme y darme entendimiento en todo mi proceso académico

Agradezco a mi madre, padre, hermanos por ser mi fuerza constante para seguir adelante y no dejarme durante todo el proceso de convertirme en un profesional.

Agradezco inmensamente a la UNIVERSIDAD TECNICA DE BABAHOYO por otorgarme la oportunidad de haberme dado un cupo para poder entrar a la CARRERA DE INGENIERIA EN SISTEMAS y así poder cumplir uno de mis objetivos de convertirme en un profesional, gracias a cada docente que conocí en cada semestre gracias por impartir sus conocimientos.

RESUMEN

La presente investigación se centra en realizar un análisis comparativo entre dos sistemas de gestión de base de datos, Cassandra y Couchbase, para determinar cuál de ellos es más escalable y flexible en términos de manejo de datos. El objetivo principal es evaluar los pros y contras de ambos sistemas, y determinar cuál es el más adecuado según las necesidades específicas del proyecto o la organización.

En la investigación se realizará una revisión detallada de la arquitectura, características y funcionalidades de cada sistema, además de la realización de pruebas y experimentos para evaluar su capacidad de escalar horizontalmente y manejar grandes volúmenes de datos. También se analizará la flexibilidad de ambos sistemas para soportar diferentes tipos de datos y modelos de datos.

El presente estudio enfocará en comparar y contrastar los sistemas de gestión de base de datos Cassandra y Couchbase en términos de escalabilidad y flexibilidad de datos, con el objetivo de brindar una guía útil para los desarrolladores y organizaciones que necesitan seleccionar la mejor opción para sus proyectos o aplicaciones.

Sin embargo, se buscará identificar las fortalezas y debilidades de cada sistema de gestión de bases de datos, con el fin de ayudar a los usuarios a tomar una decisión informada sobre cuál es la mejor opción para sus necesidades específicas de escalabilidad y flexibilidad de datos.

Palabras claves: Cassandra, Couchbase, Sistemas de gestión de bases de datos, Escalabilidad, Flexibilidad.

SUMMARY

The present research focuses on performing a comparative analysis between two database management systems, Cassandra and Couchbase, to determine which of them is more scalable and flexible in terms of data management. The main objective is to evaluate the pros and cons of both systems, and to determine which one is the most suitable according to the specific needs of the project or organization.

The research will include a detailed review of the architecture, features and functionalities of each system, as well as tests and experiments to evaluate their ability to scale horizontally and handle large volumes of data. The flexibility of both systems to support different types of data and data models will also be analyzed.

This study will focus on comparing and contrasting the Cassandra and Couchbase database management systems in terms of scalability and data flexibility, with the objective of providing a useful guide for developers and organizations that need to select the best option for their projects or applications.

However, it will seek to identify the strengths and weaknesses of each database management system in order to help users make an informed decision on which is the best option for their specific data scalability and flexibility needs.

Keywords: Cassandra, Couchbase, Database management systems, Scalability, Flexibility.

INTRODUCCIÓN

La capacidad de manejar grandes volúmenes de datos de forma efectiva y eficiente se ha convertido en un desafío clave en el mundo de la tecnología de la información, debido a la creciente demanda de aplicaciones web y móviles escalables; por lo tanto, contar con sistemas de gestión de bases de datos capaces de satisfacer esta demanda se ha vuelto esencial.

Los sistemas de gestión de bases de datos han ganado popularidad gracias a su capacidad de manejar enormes cantidades de información tanto estructurados como no estructurados; además, su adaptabilidad a diferentes modelos de datos y requisitos ha sido otro factor importante que ha contribuido a su creciente uso; por ende, la flexibilidad de estos sistemas se refiere a su capacidad para manejar diferentes tipos y estructuras de datos, así como para permitir cambios en el modelo de datos sin necesidad de reestructurar completamente la base de datos, esto incluye la capacidad de agregar o modificar campos o atributos de los datos, manejar diferentes tipos de relaciones y permitir consultas complejas.

Cassandra y Couchbase son dos sistemas de gestión de bases de datos que se utilizan ampliamente en aplicaciones web y móviles de alta escala, ya que ambos sistemas se enfocan en la escalabilidad y la flexibilidad de datos, aunque abordan estos aspectos de manera diferente. La escalabilidad se refiere a la capacidad de los sistemas para llevar enormes demasía información sin comprometer el rendimiento aumentando la capacidad de procesamiento y almacenamiento a medida que la demanda de datos crece, agregar nodos de manera transparente y distribuir datos en múltiples servidores para mejorar la redundancia y la tolerancia a fallos.

El objetivo es identificar las fortalezas y debilidades de cada sistema en relación con los requisitos específicos de escalabilidad y flexibilidad de datos de una aplicación o proyecto llevando a cabo una revisión detallada de las características al compararlos para determinar cuál es más adecuado para diferentes aplicaciones y proyectos.

La metodología de investigación a utilizar en este proyecto es la investigación documental y bibliográfica aplicando un análisis comparativo entre los sistemas de gestión de base de datos, Cassandra y Couchbase, en términos de escalabilidad y flexibilidad de datos; por lo que, este análisis utilizará un enfoque cualitativo para recopilar y analizar la información para comparar ambos sistemas; además, se utilizó el método descriptivo para recolectar, resumir, organizar y presentar información que se ha podido recopilar en diferentes sitios web o fuentes bibliográficas. El fin de este método descriptivo es ir seleccionando información que existe sobre los sistemas de gestión de base de datos, Cassandra y Couchbase en comparación.

Cassandra es un sistema de gestión de bases de datos distribuido que permite el manejo de grandes volúmenes de datos estructurados; por lo que, su diseño está enfocado en la escalabilidad horizontal y la alta disponibilidad en entornos de varios centros de datos. Por otra parte, Couchbase es un sistema de gestión de bases de datos NoSQL basado en documentos que también se centra en la escalabilidad y la disponibilidad de datos porque ofrece flexibilidad para manejar varios modelos de datos, incluyendo documentos JSON, objetos y datos clave/valor; por lo tanto, es importante realizar un análisis comparativo detallado de ambos sistemas para evaluar sus fortalezas y debilidades en relación con los requisitos específicos de escalabilidad y flexibilidad de datos de una aplicación o proyecto.

La línea de investigación a utilizarse es Sistemas de la Información y Comunicación, Emprendimiento e Innovación, y la Sublínea de la investigación es Redes y Tecnologías Inteligentes de Software y Hardware para el desarrollo del análisis comparativo entre los sistemas de gestión de base de datos, Cassandra y Couchbase, en términos de escalabilidad y flexibilidad de datos.

DESARROLLO

En la actualidad, las empresas manejan grandes cantidades de datos que necesitan ser almacenados y procesados de manera eficiente; para lograr esto, es importante contar con un sistema de gestión de bases de datos que permita la escalabilidad y flexibilidad necesarias para manejar estos datos de manera efectiva. Entre los sistemas de gestión de bases de datos más utilizados se encuentran Cassandra y Couchbase, ambos diseñados para manejar grandes cantidades de datos distribuidos en múltiples servidores (SISTEMAS, 2020).

Cassandra es un sistema de gestión de bases de datos distribuido de código abierto desarrollado por Apache, que está diseñado para ser altamente escalable y tolerante a fallos; por ende, se basa en una arquitectura descentralizada y distribuida que permite el almacenamiento y procesamiento de grandes cantidades de datos en múltiples servidores. La escalabilidad de ella se logra a través de la adición de nodos adicionales al clúster, lo que permite aumentar la capacidad de almacenamiento y procesamiento. Además, Cassandra es altamente tolerante a fallos y puede recuperarse rápidamente de fallos en un servidor o en todo el clúster (Martín, 2021).

Couchbase, por otro lado, es un sistema de gestión de bases de datos distribuido de código abierto que combina la escalabilidad de una base de datos NoSQL con la funcionalidad de una base de datos SQL. Couchbase también se basa en una arquitectura distribuida y escalable que permite el almacenamiento y procesamiento de grandes cantidades de datos en múltiples servidores, además ofrece una gran flexibilidad en cuanto a los esquemas de datos y la capacidad de manejar diferentes tipos de datos, incluyendo datos estructurados, semiestructurados y no estructurados; por lo que, la escalabilidad de Couchbase se logra a través de la adición de nodos adicionales al clúster, lo que permite aumentar la capacidad de almacenamiento y procesamiento de datos (Fernández, 2021).

Couchbase fue desarrollado inicialmente por la empresa norteamericana Couchbase Inc., fundada en 2009; por lo tanto, la empresa tiene su sede en Santa Clara, California y

cuenta con oficinas en varios países del mundo; por lo que, Couchbase tiene varias patentes registradas en relación a su tecnología de base de datos NoSQL distribuida.

Cassandra es un tipo de DBMS (Sistema de Gestión de Bases de Datos) NoSQL (no relacional) distribuido, ya que fue desarrollado inicialmente por Facebook y ahora es un proyecto de código abierto mantenido por la Apache Software Foundation; por ende, está diseñado para manejar enormes volúmenes de datos en un entorno distribuido, lo que significa que puede escalar horizontalmente a través de múltiples servidores, este utiliza una arquitectura descentralizada sin puntos únicos de fallo, lo que lo hace altamente disponible y resistente a fallos; además, Cassandra se centra en la escalabilidad lineal, lo que significa que se puede agregar capacidad adicional de hardware sin que esto afecte el rendimiento de la base de datos y es particularmente adecuado para aplicaciones de alta disponibilidad, alta escalabilidad, como redes sociales, sistemas de mensajería y aplicaciones de IoT (Internet de las cosas).

Sistemas de gestión de base de datos

Los sistemas de gestión de bases de datos (SGBD) son programas de software que permiten la creación, el mantenimiento y la administración de bases de datos. Estos sistemas proporcionan una interfaz para que los usuarios puedan acceder y manipular la información almacenada en la base de datos de manera segura y eficiente. Los SGBD también se encargan de garantizar la integridad, la consistencia y la seguridad de la información almacenada.

Existen diferentes tipos de SGBD según el modelo de datos que utilizan y la funcionalidad que ofrecen. Algunos de los SGBD más utilizados son los SGBD relacionales, los SGBD NoSQL y los SGBD distribuidos. Los SGBD relacionales utilizan tablas para almacenar y relacionar la información, mientras que los SGBD NoSQL utilizan diferentes estructuras de datos como documentos, gráficos o claves-valor. Los SGBD distribuidos se utilizan para manejar grandes cantidades de datos que se encuentran en diferentes servidores y se utilizan para escalar horizontalmente la base de datos (Betania, 2023).

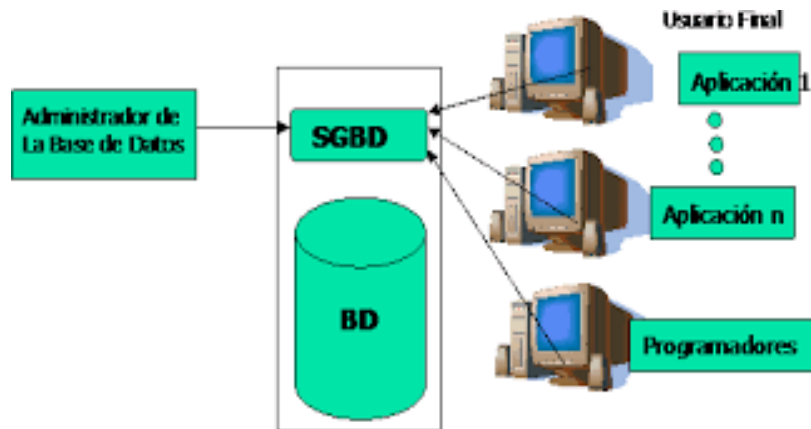


Ilustración 1 Sistemas de gestión de bases de datos

Fuente: <https://blog.alhambrait.com/redes-sd-wan-caracteristicas-y-beneficios>

Tipos de sistemas de gestión de bases de datos

Existen diversos tipos de sistemas de gestión de bases de datos (SGBD), cada uno con características específicas que lo hacen adecuado para diferentes situaciones; por lo que, los SGBD más comunes son:

- **Los SGBD relacionales**, que utilizan tablas para almacenar y relacionar la información.

- **Los SGBD NoSQL**, que utilizan estructuras de datos flexibles como documentos, grafos y claves-valor.

- **Los SGBD distribuidos**, que permiten el almacenamiento y acceso a la información en diferentes servidores.

- **También hay SGBD orientados a objetos**, basados en XML y de tiempo real, entre otros; por lo que, la elección del SGBD adecuado para un proyecto específico dependerá de factores como el tipo de datos que se manejen, la cantidad de información que se deba almacenar, el número de usuarios que accedan a la base de datos y los requerimientos de escalabilidad, rendimiento y seguridad (Bantu Group, 2020).

Qué permite y para qué sirve un Sistema Gestor de Bases de Datos

Los sistemas de gestión de bases de datos (SGBD) permiten el almacenamiento, acceso y manipulación de grandes cantidades de información de manera eficiente y segura, ya que los SGBD proporcionan una interfaz para que los usuarios puedan crear, modificar y consultar la información almacenada en la base de datos, además de garantizar la integridad, consistencia y seguridad de la información, lo que permiten también la realización de consultas complejas y la generación de informes y estadísticas de manera automatizada; por ende, gracias a su capacidad para manejar grandes cantidades de datos, los SGBD son una herramienta esencial en diferentes sectores empresariales, desde finanzas y manufactura hasta salud y educación, entre otros (Pérez, 2021).

Permite el almacenamiento, acceso y gestión de grandes cantidades de información de manera segura y eficiente porque los SGBD permiten la creación de bases de datos que pueden ser utilizadas por aplicaciones y sistemas para almacenar recuperando información, ya que los SGBD proporcionan una interfaz que permite a los usuarios definir la estructura de la base de datos, ingresar y manipular los datos, realizar consultas y generar informes de manera automatizada.

Además, los SGBD aseguran la integridad y la consistencia de los datos, lo que garantiza que la información almacenada sea precisa y confiable; por lo que, son utilizados en una amplia variedad de aplicaciones, desde sistemas de inventario y facturación hasta aplicaciones de salud y educación, lo que los convierte en una herramienta esencial en la gestión de la información empresarial (INGENIERÍA, 2022).

Las funciones principales de un Sistema Gestor de Bases de Datos

El gestor de bases de datos (DBMS) es un software que permite la creación, modificación, almacenamiento y recuperación de datos en una base de datos. Las funciones principales de un DBMS son las siguientes:

1. **Definición de la estructura de la base de datos:** El DBMS permite definir la estructura de la base de datos; es decir, la creación de tablas y la definición de campos y relaciones entre ellas.
2. **Manipulación de datos:** El DBMS permite la inserción, actualización y eliminación de datos en la base de datos, así como la consulta de información de manera rápida y eficiente.
3. **Control de acceso y seguridad:** El DBMS permite definir y controlar el acceso a la base de datos por parte de diferentes usuarios, así como garantizar la integridad y confidencialidad de la información almacenada.
4. **Gestión de transacciones:** El DBMS garantiza que las transacciones realizadas en la base de datos sean completas, consistentes y duraderas, incluso en caso de fallas del sistema.
5. **Gestión de la concurrencia:** El DBMS permite el acceso simultáneo a la base de datos por parte de múltiples usuarios y garantiza que no se produzcan conflictos en la actualización de datos (Sonix, 2021).

Objetivos de un Sistema Gestor de Bases de Datos

El objetivo principal de un sistema gestor de bases de datos (DBMS, por sus siglas en inglés) es permitir la gestión eficiente y efectiva de la información almacenada en una base de datos. Los objetivos específicos de un DBMS incluyen:

- **Acceso rápido a los datos:** El DBMS debe permitir un acceso rápido y eficiente a los datos almacenados en la base de datos; esto incluye la capacidad de realizar búsquedas, consultas y actualizaciones de datos de manera eficiente.

- **Protección de los datos:** El DBMS debe garantizar que los datos estén protegidos contra el acceso no autorizado o la pérdida; esto incluye la capacidad de establecer permisos de acceso y control de seguridad para proteger los datos.

- **Fiabilidad y disponibilidad:** El DBMS debe ser confiable y estar disponible en todo momento para garantizar que los datos estén disponibles cuando se necesiten, esto puede implicar la implementación de técnicas de copia de seguridad y recuperación de desastres.

- **Facilidad de uso:** El DBMS debe ser fácil de usar y comprender para los usuarios finales, ya que esto puede implicar la creación de interfaces de usuario intuitivas y la provisión de documentación clara.

- **Escalabilidad:** El DBMS debe ser capaz de manejar grandes volúmenes de datos y aumentar su capacidad de almacenamiento y procesamiento a medida que la cantidad de datos aumenta.

- **Integración:** El DBMS debe ser capaz de integrarse con otros sistemas y aplicaciones para permitir el intercambio de datos y la interoperabilidad; por lo que, esto puede implicar el soporte de estándares de interoperabilidad y la integración con otros sistemas mediante API o interfaces de programación (Felipe, 2021).

Características de un Sistema Gestor de Bases de Datos

- **Manejo de grandes cantidades de datos:** Un gestor de bases de datos debe ser capaz de manejar grandes cantidades de datos y proporcionar una forma eficiente de acceder a ellos.

- **Organización de datos:** El gestor de bases de datos debe ser capaz de organizar los datos en tablas, campos y registros, lo que facilita la gestión y el acceso a los datos.

- **Acceso a datos:** Los gestores de bases de datos deben proporcionar una interfaz para que los usuarios puedan acceder y manipular los datos.

- **Seguridad:** Es una característica fundamental de un gestor de bases de datos, ya que debe proporcionar medidas de seguridad para proteger los datos de accesos no autorizados.

- **Integridad de datos:** Un gestor de bases de datos debe garantizar la integridad de los datos almacenados en él; por lo que, esto se logra mediante la implementación de restricciones de integridad, como restricciones de clave primaria y foránea.

- **Transacciones:** El gestor de bases de datos debe proporcionar una forma de realizar transacciones, lo que garantiza que todas las operaciones realizadas en la base de datos se completen correctamente o se anulen en caso de error.

- **Escalabilidad:** El gestor de bases de datos debe ser capaz de manejar grandes cantidades de datos y ser escalable para adaptarse a las necesidades cambiantes de la organización.

- **Backup y recuperación:** El gestor de bases de datos debe proporcionar un mecanismo para realizar copias de seguridad y restaurar datos en caso de pérdida de datos o corrupción (DBA, 2020).

Componentes de un Sistema gestor de bases de datos

Motor de base de datos: Es el componente central del DBMS que proporciona la funcionalidad básica para administrar los datos, como la creación de tablas, la gestión de la integridad de los datos y la optimización de las consultas.

Lenguaje de consulta: El DBMS utiliza un lenguaje de consulta para que los usuarios puedan acceder y manipular los datos almacenados por ejemplo el SQL (Structured

Query Language) es uno de los lenguajes de consulta más populares y ampliamente utilizados en los DBMS.

- **Motor de almacenamiento:** Es responsable de la gestión del almacenamiento físico de los datos en el disco; por lo que, esto incluye la asignación de espacio en disco, la gestión del acceso a los datos y la optimización del rendimiento del acceso a los datos.

- **Controlador de transacciones:** Este componente asegura que las operaciones realizadas en la base de datos se completen correctamente o se anulen en caso de error, ya que se logra mediante el control de las transacciones y la garantía de que se cumplan los criterios de Atomicidad, Consistencia, Aislamiento y Durabilidad (ACID).

- **Gestor de seguridad:** El DBMS también debe proporcionar un sistema de seguridad que permita el acceso controlado a los datos, esto incluye la autenticación y autorización de usuarios, la gestión de permisos y la protección contra accesos no autorizados.

- **Herramientas de gestión:** Estas son herramientas que permiten a los usuarios administrar y configurar el DBMS, realizar tareas de mantenimiento, realizar copias de seguridad y recuperación de datos y monitorear el rendimiento del sistema (Cuestiones técnicas, 2020).

Lenguajes de datos

Un sistema gestor de bases de datos necesita contar con lenguajes para la manipulación de datos en diferentes niveles: para el acceso, modificación de datos, a nivel de seguridad y control; por lo que, estos lenguajes se dividen en tres categorías: el lenguaje de definición de datos (DDL), el lenguaje de control de datos (DCL) y el lenguaje de manipulación de datos (DML).

- **El DDL.** Permite crear, modificar y eliminar la estructura de la base de datos, incluyendo tablas, índices, relaciones y restricciones.

- **El DCL.** Permite establecer permisos sobre los objetos de la base de datos, la creación y eliminación de usuarios.

- **Por último, el DML.** Permite la inserción, actualización, eliminación y consulta de datos en las tablas de la base de datos (Princhi, 2022).



Ilustración 2 Gestor de bases de datos: Qué son y sus características
Fuente: <https://coworkings.co/gestor-de-bases-de-datos-ventajas-desventajas/>

Flexibilidad

La flexibilidad de un sistema gestor de base de datos (DBMS) se refiere a su capacidad para adaptarse a diferentes requisitos y necesidades de los usuarios y aplicaciones, ya que un sistema de gestión de base de datos flexible debe ser capaz de manejar diferentes tipos de datos, ofrecer diferentes modelos de datos y proporcionar diferentes niveles de seguridad y control de acceso a los datos.

Otro aspecto de la flexibilidad de un DBMS es su capacidad para integrarse con otros sistemas y aplicaciones en el entorno de una organización, esto significa que el DBMS debe ser capaz de interoperar con otros sistemas y aplicaciones, así como proporcionar una interfaz de programación de aplicaciones (API) que permita a los desarrolladores interactuar con la base de datos desde diferentes lenguajes y plataformas (Naeem, 2020).

Escalabilidad

La escalabilidad de un sistema de gestor de base de datos (DBMS) se refiere a su capacidad para manejar grandes cantidades de datos y un aumento en la demanda de procesamiento sin afectar significativamente su rendimiento y capacidad de respuesta; en otras palabras, un DBMS escalable es capaz de manejar una mayor carga de trabajo a medida que la cantidad de datos y usuarios aumenta.

Existen dos tipos principales de escalabilidad en un DBMS: la escalabilidad vertical y la escalabilidad horizontal. La escalabilidad vertical se refiere a la capacidad de un DBMS para manejar una mayor carga de trabajo aumentando la potencia de procesamiento y la capacidad de almacenamiento de hardware en una sola máquina; por otro lado, la escalabilidad horizontal se refiere a la capacidad de un DBMS para manejar una mayor carga de trabajo mediante la adición de más máquinas a un clúster de servidores.

Además de la escalabilidad vertical y horizontal, un DBMS escalable debe ser capaz de manejar diferentes tipos de datos y diferentes modelos de datos, ya que debe ser capaz de adaptarse a diferentes requisitos de seguridad y control de acceso a los datos proporcionando un alto nivel de disponibilidad y confiabilidad (Aguirre, 2020).

Ventajas de un Sistema Gestor de Base de Datos

- **Mayor eficiencia y velocidad:** Un DBMS está diseñado para manejar grandes cantidades de datos de manera eficiente y rápida, lo que permite realizar operaciones complejas de forma más rápida y precisa.

- **Mayor seguridad y control de acceso:** Un DBMS proporciona un alto nivel de seguridad y control de acceso a los datos, lo que significa que los usuarios solo pueden acceder a los datos para los que tienen permisos explícitos.

- **Mayor disponibilidad y confiabilidad:** Un DBMS está diseñado para proporcionar alta disponibilidad y confiabilidad, lo que significa que los datos están disponibles en todo momento y que se realizan copias de seguridad regulares para garantizar su integridad.

- **Mayor flexibilidad:** Un DBMS proporciona una mayor flexibilidad en la gestión de datos, lo que permite manejar diferentes tipos de datos y modelos de datos.

- **Mejora en la toma de decisiones:** Un DBMS permite la realización de consultas y análisis de datos complejos, lo que puede ayudar a los usuarios a tomar decisiones más informadas y basadas en datos.

- **Menor costo y tiempo de mantenimiento:** Un DBMS puede reducir el costo y tiempo de mantenimiento de los datos, ya que se elimina la necesidad de actualizar manualmente archivos de datos y tablas de Excel (Cutipa, 2021).

Desventajas de un Sistema Gestor de Base de Datos

- **Complejidad:** Los DBMS pueden ser complejos de implementar, configurar y pueden requerir personal especializado para administrarlos y mantenerlos.

- **Vulnerabilidades de seguridad:** Al igual que cualquier sistema de software, los DBMS pueden ser vulnerables a ataques de seguridad y piratería, lo que puede exponer datos sensibles y confidenciales.

- **Rendimiento limitado:** Los DBMS pueden presentar un rendimiento limitado si no se configuran correctamente, especialmente en entornos de alto tráfico o con grandes cantidades de datos.

- **Problemas de compatibilidad:** Al cambiar de un DBMS a otro, pueden surgir problemas de compatibilidad que dificulten la transferencia de datos o incluso la integridad de los datos (Bajonero, 2022).



Ilustración 3 Qué es un DBMS o sistema de administración de bases de datos
Fuente: <https://proximahost.es/blog/dbms-sistema-administracion-bases-datos/>

Tabla comparativa

Aspecto	Cassandra	Couchbase
Modelo de datos	Orientado a columnas, se enfoca en el almacenamiento de datos estructurados.	Orientado a documentos, se enfoca en la flexibilidad y manejo de datos no estructurados.
Escalabilidad horizontal	Altamente escalable horizontalmente, puede manejar grandes cantidades de datos y consultas en clústeres distribuidos.	Altamente escalable horizontalmente, puede manejar grandes cantidades de datos y consultas en clústeres distribuidos.
Escalabilidad vertical	No es tan escalable verticalmente como Couchbase, es mejor usarlo en un clúster distribuido.	Altamente escalable verticalmente, puede manejar grandes cargas de trabajo en una sola máquina.
Flexibilidad de datos	Buena, pero requiere una planificación cuidadosa para el esquema de la tabla y las consultas.	Excelente, ya que admite documentos JSON flexibles que se pueden agregar sin modificar la estructura.
Lenguaje de consultas	CQL (Cassandra Query Language)	N1QL (SQL for JSON)
Replicación de datos	Proporciona replicación de datos mediante su arquitectura descentralizada y esquemas de partición.	Proporciona replicación de datos mediante su arquitectura de clúster y políticas de replicación configurables.
Durabilidad de datos	Garantiza la durabilidad de datos mediante la replicación de datos y mecanismos de recuperación en caso de fallos.	Garantiza la durabilidad de datos mediante la replicación de datos y mecanismos de recuperación en caso de fallos.

Escalabilidad de lecturas y escrituras	Cassandra es capaz de manejar grandes cargas de escritura y lectura en paralelo mediante su arquitectura descentralizada.	Couchbase también es capaz de manejar grandes cargas de escritura y lectura en paralelo mediante su arquitectura de clúster y partición.
Soporte de transacciones	No soporta transacciones ACID. Se basa en el modelo eventual de consistencia.	Soporta transacciones ACID mediante su modelo de transacciones distribuidas.
Integración con Hadoop	Cassandra tiene una integración nativa con Hadoop y otros sistemas de análisis de datos.	Couchbase tiene una integración con Hadoop a través de su conector de Hadoop.
Compatibilidad con múltiples lenguajes	Cassandra es compatible con varios lenguajes de programación, incluyendo Java, Python, y C#.	Couchbase es compatible con varios lenguajes de programación, incluyendo Java, Python, y Ruby.
Facilidad de administración	Cassandra requiere un conocimiento avanzado para la configuración y administración.	Couchbase es más fácil de administrar y configurar, especialmente para usuarios nuevos.
Comunidad y soporte	Cassandra tiene una gran comunidad de usuarios y es respaldado por Apache.	Couchbase tiene una comunidad activa y soporte comercial de la compañía.

La tabla anterior presenta una comparación detallada entre dos sistemas de base de datos NoSQL, Cassandra y Couchbase, en términos de escalabilidad y flexibilidad de datos; por lo que, ambos sistemas son altamente escalables horizontalmente y tienen una buena capacidad de manejar grandes cantidades de datos en clústeres distribuidos; sin embargo, Couchbase también es altamente escalable verticalmente, lo que significa que puede manejar grandes cargas de trabajo en una sola máquina.

Cassandra se enfoca en el almacenamiento de datos estructurados y utiliza un modelo de datos orientado a columnas, mientras que Couchbase se enfoca en la flexibilidad y manejo de datos no estructurados utilizando un modelo de datos orientado a documentos. En cuanto a la facilidad de administración, Couchbase es considerado más fácil de administrar y configurar que Cassandra, especialmente para usuarios nuevos.

Ambos sistemas proporcionan replicación de datos, durabilidad de datos y escalabilidad de lecturas y escrituras, aunque Cassandra no soporta transacciones ACID y se basa en el modelo eventual de consistencia, mientras que Couchbase soporta transacciones ACID mediante su modelo de transacciones distribuidas. En resumen, Cassandra y Couchbase son opciones viables para aplicaciones de alta escalabilidad y rendimiento, y la elección dependerá de las necesidades específicas del caso.

Tabla comparativa de la flexibilidad de Cassandra y Couchbase como Sistemas de Gestión de Bases de Datos

Característica	Cassandra	Couchbase
Estructura de datos	Cassandra utiliza una estructura de datos basada en columnas, lo que significa que los datos se organizan en columnas en lugar de filas.	Couchbase utiliza una estructura de datos basada en documentos, lo que significa que los datos se almacenan como documentos JSON.
Flexibilidad de esquema	Cassandra proporciona una flexibilidad de esquema limitada, ya que los tipos de datos y las columnas deben definirse de antemano.	Couchbase proporciona una mayor flexibilidad de esquema, ya que los documentos JSON pueden contener cualquier número de campos y tipos de datos diferentes.

Tabla comparativa de la escalabilidad de Cassandra y Couchbase como Sistemas de Gestión de Bases de Datos

Característica	Cassandra	Couchbase
Escalabilidad	Cassandra está diseñado para escalar horizontalmente a través de múltiples servidores, lo que lo hace altamente escalable.	Couchbase también está diseñado para escalar horizontalmente a través de múltiples servidores, lo que lo hace altamente escalable.

Que tipos de DBMS son Cassandra y Couchbase

Cassandra y Couchbase son dos sistemas de gestión de bases de datos NoSQL (Not Only SQL) que comparten algunas características, pero también tienen algunas diferencias significativas; por lo que, Cassandra es una base de datos distribuida de alta escalabilidad que se basa en el modelo de datos de columnas, ya que Cassandra fue desarrollado originalmente por Facebook y luego se convirtió en un proyecto de código abierto de Apache.

Algunas de las ventajas de Cassandra son:

- **Alta escalabilidad horizontal:** Cassandra es capaz de manejar grandes cantidades de datos en un entorno distribuido sin comprometer el rendimiento.
- **Tolerancia a fallos:** Cassandra está diseñado para tolerar fallos en los nodos y recuperarse rápidamente sin perder datos.
- **Modelo de datos flexible:** el modelo de datos de columnas permite agregar nuevas columnas sin modificar el esquema de la base de datos.

- **Buen rendimiento de lectura y escritura:** Cassandra está optimizado para leer y escribir grandes cantidades de datos en un entorno distribuido.

Sin embargo, también tiene algunas desventajas, como:

- **Complejidad de configuración:** configurar un clúster de Cassandra puede ser complicado, especialmente si no se tiene experiencia en la administración de bases de datos distribuidas.
- **Falta de soporte transaccional:** Cassandra no es una base de datos transaccional, lo que significa que no soporta transacciones ACID (Atomicidad, Consistencia, Aislamiento, Durabilidad).
- **Escasa herramienta de análisis:** aunque Cassandra tiene algunas herramientas para analizar los datos, no son tan completas como las de otras bases de datos.

Couchbase, por otro lado, es una base de datos distribuida de alto rendimiento que combina las características de una base de datos de documentos y una base de datos clave-valor.

Algunas de las ventajas de Couchbase son:

- **Escalabilidad horizontal:** Couchbase se puede escalar fácilmente en un entorno distribuido para manejar grandes cantidades de datos.
- **Alta disponibilidad:** Couchbase tiene la capacidad de recuperarse rápidamente de fallos de nodo sin perder datos.
- **Modelo de datos flexible:** el modelo de documentos de Couchbase permite agregar nuevos campos sin modificar el esquema de la base de datos.

- **Soporte para transacciones ACID:** Couchbase soporta transacciones ACID, lo que significa que garantiza la atomicidad, la consistencia, el aislamiento y la durabilidad de las operaciones.

Sin embargo, también tiene algunas desventajas, como:

- **Complejidad de configuración:** configurar un clúster de Couchbase puede ser complicado, especialmente si no se tiene experiencia en la administración de bases de datos distribuidas.
- **Consumo de recursos:** Couchbase utiliza muchos recursos del sistema, lo que puede afectar el rendimiento en algunos entornos.
- **Precio:** aunque hay una versión de código abierto de Couchbase, la versión empresarial de Couchbase es costosa.

Tanto Cassandra como Couchbase son sistemas de gestión de bases de datos distribuidas, lo que significa que pueden escalar horizontal y verticalmente para manejar grandes cantidades de datos y altas cargas de trabajo; por lo que, son capaces de escalar horizontalmente y verticalmente para manejar grandes cantidades de datos y altas cargas de trabajo, ya que la escalabilidad horizontal se logra agregando más nodos al clúster, mientras que la escalabilidad vertical se logra agregando más recursos a cada nodo existente.

CONCLUSIONES

Una vez finalizado el presente compendio investigativo Después de comparar los sistemas de gestión de bases de datos Cassandra y Couchbase en términos de escalabilidad y flexibilidad de datos, podemos concluir lo siguiente:

- Ambas bases de datos son altamente escalables y ofrecen un alto rendimiento para lecturas y escrituras a través de clústeres de nodos; por lo que, Cassandra tiene un modelo de datos columnar, basado en clave, mientras que Couchbase tiene un modelo de datos de documentos JSON, lo que la hace más adecuada para casos de uso donde se requiere flexibilidad en los datos.

- Cassandra requiere una planificación cuidadosa para el esquema de la tabla y las consultas, mientras que Couchbase admite documentos JSON flexibles que se pueden agregar sin modificar la estructura, lo que ambas bases de datos permiten configurar la consistencia de los datos para adaptarse a diferentes requisitos de aplicación.

- Cassandra utiliza menos memoria que Couchbase, lo que la hace más adecuada para entornos con recursos limitados, ya que Cassandra cuenta con una comunidad de usuarios más grande y una amplia base de usuarios en comparación con Couchbase, aunque la comunidad de usuarios de Couchbase está en constante crecimiento. Por lo tanto, Cassandra es de código abierto y gratuito, mientras que Couchbase tiene una versión de comunidad gratuita y una versión empresarial con costo.

BIBLIOGRAFÍA

- Aguirre, M. F. (13 de Noviembre de 2020). *appvizer*. Obtenido de appvizer:
<https://www.appvizer.es/revista/organizacion-planificacion/gestion-proyectos/escalabilidad>
- Bajonero, A. M. (22 de Septiembre de 2022). *centrobanamex*. Obtenido de centrobanamex:
<https://www.centrobanamex.com.mx/cuales-son-las-desventajas-de-los-gestores-de-bases-de-datos>
- Bantu Group. (27 de Abril de 2020). *bantugroup*. Obtenido de bantugroup:
<https://www.bantugroup.com/blog/tipos-de-sistemas-de-gestion-de-base-de-datos>
- Betania, V. (24 de Enero de 2023). *Hostinger*. Obtenido de Hostinger:
<https://www.hostinger.es/tutoriales/sgbd>
- Cuestiones técnicas. (16 de Marzo de 2020). *ionos*. Obtenido de ionos:
[https://www.ionos.es/digitalguide/hosting/cuestiones-tecnicas/sistema-gestor-de-base-de-datos-
sgbd/#:~:text=Un%20sistema%20de%20gesti%C3%B3n%20de,lenguaje%20de%20manipulaci%C3%B3n%20de%20datos.](https://www.ionos.es/digitalguide/hosting/cuestiones-tecnicas/sistema-gestor-de-base-de-datos-
sgbd/#:~:text=Un%20sistema%20de%20gesti%C3%B3n%20de,lenguaje%20de%20manipulaci%C3%B3n%20de%20datos.)
- Cutipa, G. (26 de Agosto de 2021). *guidocutipa*. Obtenido de guidocutipa:
<https://guidocutipa.blog.bo/principales-ventajas-desventajas-bases-de-datos-relacionales-no-relacionales-nosql-vs-sql/>
- DBA. (16 de Junio de 2020). *Dbá Dixit*. Obtenido de Dbá Dixit: [http://dbadixit.com/caracteristicas-sistema-gestor-bases-datos-
sgbd-dbms/](http://dbadixit.com/caracteristicas-sistema-gestor-bases-datos-
sgbd-dbms/)
- Felipe. (30 de Junio de 2021). *Hostingplus*. Obtenido de Hostingplus:
<https://www.hostingplus.pe/blog/objetivos-de-las-bases-de-datos-y-sus-caracteristicas/>
- Fernández, P. (26 de Octubre de 2021). *Silicon*. Obtenido de Silicon:
<https://www.silicon.es/couchbase-anuncia-capella-la-modalidad-dbaas-para-su-base-de-datos-nosql-2447176>
- INGENIERÍA. (22 de Marzo de 2022). *Universidad Europea*. Obtenido de Universidad Europea:
<https://universidadeuropea.com/blog/para-que-sirve-gestor-base-datos/>
- Martín, S. (26 de Julio de 2021). *Pandorafms*. Obtenido de Pandorafms:
<https://pandorafms.com/blog/es/bases-de-datos-nosql/>
- Naeem, T. (1 de Enero de 2020). *Astera*. Obtenido de Astera:
<https://www.astera.com/es/type/blog/relational-database-management-system/>
- Pérez, S. (9 de Septiembre de 2021). *intelequia*. Obtenido de intelequia:
<https://intelequia.com/blog/post/gestor-de-base-de-datos-qu%C3%A9-es-funcionalidades-y-ejemplos>

- Princhi. (10 de Noviembre de 2022). *huawei*. Obtenido de huawei:
<https://forum.huawei.com/enterprise/es/%C2%BFcu%C3%A1les-son-los-3-lenguajes-de-sql/thread/1031524-100325>
- SISTEMAS. (28 de Abril de 2020). *Ilimit*. Obtenido de Ilimit: <https://www.ilight.com/blog/base-de-datos-sql-nosql/>
- Sonix. (18 de Enero de 2021). *Tech Krowd*. Obtenido de Tech Krowd:
<https://techkrowd.com/bases-de-datos/sistema-gestor-de-bases-de-datos-sgbd/>

ANEXOS

ENTREVISTA

Para 2 empleados:

- 1. ¿Cómo definirías la escalabilidad de Cassandra y Couchbase en términos de capacidad de almacenamiento y procesamiento de datos?**

En la respuesta del primer empleado menciona, tanto Cassandra como Couchbase son sistemas de bases de datos altamente escalables, ya que Cassandra es especialmente útil para grandes volúmenes de datos debido a su capacidad para escalar horizontalmente, lo que permite agregar más nodos al clúster para distribuir la carga de trabajo; además, Couchbase también es muy escalable y puede manejar grandes volúmenes de datos gracias a su modelo de documentos y su capacidad para escalar vertical y horizontalmente.

Desde la experiencia del segundo empleado, se puede decir que tanto Cassandra como Couchbase son sistemas de bases de datos distribuidas altamente escalables, ya que Cassandra es conocido por su capacidad para escalar horizontalmente, lo que permite manejar grandes cargas de trabajo y grandes volúmenes de datos. Por otro lado, Couchbase es muy escalable tanto en términos de almacenamiento como de procesamiento de datos, lo que lo hace ideal para empresas que necesitan manejar grandes cantidades de información en tiempo real. En general, creo que ambos sistemas son excelentes opciones para empresas que buscan una base de datos escalable y de alto rendimiento.

2. ¿Cómo pueden los sistemas de gestión de bases de datos Cassandra y Couchbase ayudar a las empresas en términos de escalabilidad y flexibilidad de datos?

En la respuesta del primer empleado, Cassandra y Couchbase son sistemas de gestión de bases de datos altamente escalables y flexibles. Cassandra es ideal para empresas que necesitan manejar grandes volúmenes de datos y necesitan una alta disponibilidad y tolerancia a fallos; por otro lado, Couchbase ofrece una gran flexibilidad gracias a su modelo de documentos y la capacidad de trabajar con múltiples modelos de datos en la misma base de datos.

Desde la experiencia del segundo empleado, Cassandra y Couchbase son excelentes opciones de bases de datos para empresas que buscan escalabilidad y flexibilidad. Cassandra es conocido por su capacidad para escalar horizontalmente, mientras que Couchbase ofrece una gran flexibilidad de datos gracias a su modelo de documentos y la capacidad de realizar consultas SQL y NoSQL en la misma base de datos. En general, ambas soluciones pueden ayudar a las empresas a manejar grandes cantidades de datos de manera efectiva.

3. ¿Cómo se comparan Cassandra y Couchbase en términos de flexibilidad de datos?

Desde la perspectiva del primer empleado, tanto Cassandra como Couchbase son muy flexibles en términos de datos, ya que Cassandra ofrece una estructura de datos basada en

columnas, mientras que Couchbase utiliza un modelo de documentos, lo que permite a las empresas trabajar con múltiples tipos de datos y realizar consultas SQL y NoSQL en la misma base de datos.

Desde la perspectiva del segundo empleado, Cassandra y Couchbase ofrecen una gran flexibilidad de datos porque Cassandra es especialmente útil para empresas que necesitan manejar grandes volúmenes de datos, mientras que Couchbase es una excelente opción para empresas que necesitan trabajar con múltiples modelos de datos y realizar consultas complejas en la misma base de datos.

4. ¿Cuáles son algunos casos de uso comunes para Cassandra y Couchbase en términos de escalabilidad y flexibilidad?

El primer empleado dice lo siguiente, Cassandra es comúnmente utilizado en casos de uso que requieren alta escalabilidad y tolerancia a fallos, como en aplicaciones de IoT o en la gestión de grandes volúmenes de datos en tiempo real; por otro lado, Couchbase es una buena opción para empresas que necesitan trabajar con múltiples tipos de datos y realizar consultas complejas en la misma base de datos, como en aplicaciones de comercio electrónico o en la gestión de datos de usuario.

El segundo empleado dice lo siguiente, Cassandra es ampliamente utilizado en casos de uso que requieren alta escalabilidad y disponibilidad, como en aplicaciones de juegos en línea o en la gestión de grandes volúmenes de datos para la toma de decisiones empresariales. Couchbase, por otro lado, es ideal para empresas que necesitan trabajar con múltiples modelos de datos y realizar consultas complejas, como en aplicaciones de seguimiento de inventario o en la gestión de datos de transacciones financieras.

Para 2 docentes:

1. ¿Cómo han visto que evolucionan Cassandra y Couchbase en términos de escalabilidad y flexibilidad de datos en los últimos años?

El primer docente menciona que, Cassandra y Couchbase han evolucionado significativamente en términos de escalabilidad y flexibilidad de datos en los últimos años, ya que ambas plataformas han mejorado su capacidad para manejar grandes volúmenes de datos y han añadido nuevas características para mejorar la flexibilidad de los datos y la capacidad de consulta.

El segundo docente define que, Cassandra y Couchbase han seguido evolucionando en los últimos años para satisfacer las necesidades cambiantes de las instituciones en términos de escalabilidad y flexibilidad de datos; por lo que, ambas plataformas han mejorado sus capacidades de escalabilidad horizontal y vertical, así como su capacidad para trabajar con múltiples modelos de datos y realizar consultas complejas. Además, han introducido nuevas características para mejorar la integración con otras herramientas de Big Data y la seguridad de los datos.

2. ¿Qué consideraciones de escalabilidad deben tenerse en cuenta al seleccionar una base de datos?

En la opinión del primer docente, al seleccionar una base de datos, es importante considerar la escalabilidad horizontal y vertical, así como la capacidad para manejar grandes volúmenes de datos y la capacidad de consulta, también es importante considerar la capacidad para integrarse con otras herramientas de Big Data y la seguridad de los datos.

En la opinión del segundo docente, al seleccionar una base de datos, es esencial considerar la escalabilidad, la flexibilidad y la capacidad de consulta, ya que es importante pensar en la escalabilidad horizontal y vertical, así como en la capacidad de manejar grandes volúmenes de datos. También es importante considerar la capacidad de trabajar con múltiples modelos de datos y realizar consultas complejas, así como la capacidad para integrarse con otras herramientas de Big Data y la seguridad de los datos.

3. ¿Cómo pueden los sistemas de gestión de bases de datos Cassandra y Couchbase ayudar a la Universidad a cumplir con los requisitos de regulación y cumplimiento en términos de escalabilidad y flexibilidad de datos?

El primer docente manifiesta, tanto Cassandra como Couchbase pueden ayudar a la Universidad a cumplir con los requisitos de regulación y cumplimiento en términos de escalabilidad y flexibilidad de datos. Ambas plataformas ofrecen características de seguridad robustas, así como opciones de replicación de datos y copias de seguridad automatizadas, lo que garantiza la integridad de los datos y la disponibilidad continua de los mismos.

El segundo docente manifiesta, los sistemas de gestión de bases de datos Cassandra y Couchbase pueden ser herramientas valiosas para ayudar a la Universidad a cumplir con los requisitos de regulación y cumplimiento en términos de escalabilidad y flexibilidad de datos, ya que ambas plataformas ofrecen capacidades de auditoría y seguimiento para garantizar la transparencia y la responsabilidad en el manejo de los datos, así como opciones para cifrar los datos en reposo y en tránsito para garantizar la seguridad de los mismos.

4. ¿Cómo afecta la escalabilidad y flexibilidad de datos a la toma de decisiones empresariales en el ámbito de la administración financiera e informática?

Primer docente menciona que, es la capacidad para manejar grandes volúmenes de datos y realizar consultas complejas de manera eficiente puede proporcionar información valiosa que puede ayudar a la Universidad a identificar tendencias y patrones y tomar decisiones informadas sobre la asignación de recursos y la inversión.

Segundo docente menciona que, la escalabilidad y flexibilidad de datos pueden ayudar a las empresas a implementar soluciones de tecnología de la información más eficientes y rentables, lo que puede mejorar la eficiencia y la productividad de la organización.