

# Couchbase: Multimodelo NoSQL.

## Introducción

Couchbase es una base de datos NoSQL distribuida y multimodelo, que soporta un modelo de datos JSON flexible y familiar acceso a datos relacionales y multimodelo. Ofrece un alto rendimiento en memoria, escalabilidad fácil, sincronización móvil, disponibilidad 24x365, seguridad avanzada y opciones de implementación en la nube asequibles. También se puede acceder a Couchbase como un servicio de base de datos completamente administrado llamado Couchbase Capella, y ofrece implementaciones de clústeres contenerizados gestionados por Kubernetes con su línea de productos de automatización de base de datos nativa de la nube. Como base de datos multimodelo, Couchbase admite múltiples métodos de acceso a datos dentro de una estructura dinámica de contención de datos en un formato de documento JSON flexible. Esto permite consolidar múltiples capas y motores de acceso a datos en una sola plataforma que de otra manera requeriría bases de datos de propósito único que trabajen juntas. Couchbase elimina la necesidad de administrar modelos de datos y coherencia entre múltiples sistemas, aprender diferentes lenguajes y APIs, y gestionar tecnologías independientes.

Las necesidades comerciales pragmáticas para requisitos técnicos más avanzados han impulsado a las bases de datos NoSQL multimodelo al frente. Las necesidades empresariales para un alto rendimiento, flexibilidad impulsada por la aplicación sobre la composición de sus datos, procesamiento y movilidad distribuidos, y la necesidad general de reducir los costos operativos y escapar del encierro del proveedor son impulsores clave por qué las organizaciones buscan sistemas NoSQL nativos de la nube. Estos requisitos modernos han impulsado el desarrollo de Couchbase desde su inicio, asegurando un alto rendimiento, flexibilidad en el modelo de datos y el acceso a los datos, soporte para redes de clústeres distribuidos y movilidad, y un valor increíble y bajo TCO (Total Cost of Ownership). Como resultado, Couchbase se ha convertido en una base de datos NoSQL multipropósito y moderna.

## Modelo de datos JSON y métodos de acceso.

En esta sección se describe el modelo de datos JSON que se utiliza en Couchbase. El modelo de datos JSON es compatible con tipos de datos básicos y complejos, como números, cadenas de caracteres, objetos anidados y matrices. Couchbase almacena datos como documentos individuales, cada uno con una clave y un valor. Cuando el valor es en formato JSON, Couchbase ofrece múltiples formas de acceso, incluyendo operaciones clave-valor básicas, consultas SQL++, búsqueda de texto completo, análisis en tiempo real, eventos en el lado del servidor y sincronización de aplicaciones móviles.

El modelo de documentos JSON en Couchbase ofrece una gran flexibilidad en la definición de esquemas, ya que permite documentos con esquemas variados y estructuras anidadas. Los desarrolladores pueden expresar relaciones muchos a muchos sin necesidad de una tabla de referencia o de unión. Además, los subcomponentes de los documentos pueden ser accedidos y actualizados directamente. Couchbase no impone uniformidad en la estructura de los documentos, lo que permite que las diferencias entre objetos se representen de manera eficiente y que el esquema de los documentos evolucione progresivamente. Esto significa que las aplicaciones pueden cambiar su comportamiento

sin tener que reorganizar todos los datos fuente o detener la aplicación para hacer cambios básicos.

Couchbase es una base de datos de documentos que permite que las aplicaciones administren datos JSON a través de diferentes métodos. El método clave-valor brinda acceso a los datos al proporcionar una ID de documento. El método de consulta y análisis permite que la sintaxis de consulta basada en SQL interactúe con datos JSON para devolver resultados JSON coincidentes, mientras que el método de búsqueda de texto completo utiliza analizadores de texto con tokenización para buscar una variedad de funciones de coincidencia booleana y de campo, devolviendo ID de documentos, puntuación de relevancia y datos de contexto opcionales. El método de eventos permite que se ejecuten funciones de Javascript personalizadas dentro de la base de datos a medida que cambian los datos o en función de los temporizadores, lo que admite el acceso y la actualización de datos, la escritura en un registro o la llamada a un sistema externo. Además, Couchbase Mobile es una pila de extremo a extremo que permite sincronizar datos con aplicaciones móviles.

Los documentos JSON en Couchbase consisten en claves, valores y subdocumentos. La clave es inmutable y debe ser una cadena UTF-8 sin espacios, con una longitud máxima de 250 bytes y debe ser única dentro de su depósito. El valor puede ser un tipo básico, como un número, string o booleano; o un tipo complejo, como un documento incrustado o una matriz, y el tamaño máximo de un valor es de 20 MB. Los subdocumentos son componentes internos de un documento JSON que utilizan una sintaxis de ruta para especificar atributos y posiciones de matriz para lectura/escritura, lo que hace innecesario transferir documentos completos a través de la red cuando solo se requieren modificaciones parciales.

Couchbase ofrece una estructura de organización y contención de datos multinivel flexible para organizar documentos, que consta de cuatro niveles: buckets, ámbitos, colecciones y documentos, que se asigna fácilmente a construcciones RDBMS familiares de bases de datos, esquemas, tablas y filas. Los depósitos son el contenedor superior en Couchbase, que es el equivalente lógico de una base de datos en los sistemas relacionales, y los vBuckets son la forma en que Couchbase crea y usa su mapa de ubicación de datos al segmentar un depósito en 1024 vBuckets, distribuidos entre nodos en un clúster y replicados para disponibilidad y redundancia.

## **Servicios de Couchbase.**

Couchbase es una plataforma que implementa métodos de acceso a datos a través de servicios dedicados, con el Servicio de Datos en el centro. Cada servicio tiene sus propias cuotas de recursos y, cuando corresponde, capacidades de indexación y comunicación inter-nodo. Esto proporciona varios métodos muy flexibles para escalar los servicios cuando sea necesario, no solo escalando a máquinas más grandes o a más nodos, sino también escalando servicios específicos de manera independiente entre sí. La arquitectura "shared-nothing" permite un control de aislamiento de carga de trabajo. La plataforma utiliza un núcleo de capacidad de datos que alimenta todos los demás servicios. Esto permite a los desarrolladores controlar la carga de trabajo. Las aplicaciones se comunican directamente con cada servicio a través de un SDK común que siempre está al tanto de la topología del clúster y cómo están configurados los servicios.

El Servicio de Datos es la base para almacenar datos en Couchbase Server. Es responsable de la memoria caché, persistencia y servir datos a aplicaciones y otros servicios dentro del clúster. El principal componente de la arquitectura del servicio

de datos es el sistema de gestión de clave-valor conocido simplemente como el Motor KV. Cada nodo que ejecuta el Servicio de Datos tiene su propio proceso de Motor KV y es responsable de persistir y cachear una porción del conjunto de datos general. Couchbase ha introducido un nuevo formato de motor de almacenamiento llamado Magma, que combina el rendimiento de los árboles de mezcla de registro con la capacidad de organización y la inmutabilidad de las tablas de cadena ordenadas, para proporcionar un alto rendimiento en un motor de baja latencia que se adapta a cargas de trabajo de búsquedas rápidas y de baja latencia. Esto minimiza el aumento de espacio en disco llamado "amplificación de almacenamiento" y la complejidad que se produce cuando los documentos se mutan sin ser reorganizados.

Cuando Couchbase recibe una solicitud para escribir un documento, ocurre lo siguiente:

1. Cada servidor en un clúster de Couchbase tiene su propia caché de objetos administrados. El cliente escribe un documento en la caché y el servidor envía una confirmación al cliente. Por defecto, el cliente no tiene que esperar a que el servidor persista y replique el documento, ya que ocurre de forma asíncrona.
2. El documento se agrega a la cola de replicación intra-cluster para ser replicado a otros servidores dentro del clúster.
3. El documento se agrega a la cola de escritura en disco para ser persistido de forma asíncrona en disco. El documento se persiste en disco después de que se haya vaciado la cola de escritura en disco.
4. Después de que el documento se persiste en disco, se replica a otros clústeres mediante XDCR y, eventualmente, se indexa.

Couchbase es una base de datos de documentos, pero en su núcleo hay un almacenamiento de clave-valor (KV) distribuida. Una tienda KV es un enfoque extremadamente simple y sin esquema para la gestión de datos que almacena un ID único (clave) junto con una pieza de información arbitraria (valor); se puede pensar en ella como un mapa hash o un diccionario. La tienda KV en sí puede aceptar cualquier tipo de dato, ya sea un blob binario o un documento JSON, y las características de Couchbase, como el servicio de consulta SQL++, hacen uso de la capacidad de la tienda KV para procesar documentos JSON.

El servicio de índices de Couchbase es responsable de mantener y administrar los índices secundarios globales (GSI), los cuales son esenciales para consultas eficientes. Couchbase soporta varios tipos de índices, como primarios, secundarios, funcionales, cubiertos, de matriz y compuestos, entre otros. El servicio de índices funciona separado del servicio de consultas, permitiendo que sus cargas de trabajo se aislen cuando sea necesario. Además, el Asesor de Índices es un comando de consulta integrado que sugiere qué índice o GSI usar o construir en función de las selecciones de objeto y predicado contenidas en la instrucción de consulta.

Couchbase proporciona varios niveles de control sobre la consistencia de la consulta para permitir intercambios entre la latencia de consulta y la consistencia de los datos. Estos niveles incluyen `no_delimitado`, `en_plus` y `solicitud_plus`, con cada nivel especificado por consulta para permitir un control aún más fino y dinámico de estos intercambios. En el fondo, los índices de Couchbase se actualizan de forma asíncrona después de que la aplicación haya modificado los datos. Esto permite una mayor capacidad de escritura, pero aumenta la posibilidad de inconsistencia entre los datos y sus índices. Para abordar esto, Couchbase utiliza índices optimizados en memoria (MOI), los cuales utilizan una estructura de lista de saltos en lugar de índices B-tree, optimizando el consumo de memoria y el procesamiento concurrente de actualizaciones y escaneos de índices.

Couchbase también incluye un servicio de búsqueda es un motor que realiza búsquedas de texto completo (FTS) en datos JSON almacenados dentro de un cubo o colección. Le permite crear, administrar y consultar índices invertidos para buscar texto libre en un documento. El servicio proporciona analizadores que realizan operaciones como tokenización multilingüe, truncamiento y puntuación de relevancia. Los nodos de búsqueda incorporan tanto un indexador como un procesador de consultas, con ambas cargas de trabajo ejecutándose en cada nodo de búsqueda. La creación de índices es altamente configurable a través de un archivo de definición de índice JSON, el SDK de Couchbase o una interfaz web gráfica como parte de la consola de administración.

Este también incluye servicio de analítica. El servicio de análisis proporciona una capacidad de consulta ad hoc sin la necesidad de índices, lo que permite un modelo híbrido de procesamiento operacional y analítico (HOAP) para análisis en tiempo real y operacionales sobre los datos JSON activos en Couchbase. Utiliza el mismo lenguaje SQL++ que el servicio de consulta y está diseñado para ejecutar eficientemente consultas complejas sobre un gran número de documentos, incluyendo características de consulta como joins ad hoc, operaciones de conjunto, agregación y agrupamiento. Debido a que el servicio admite el procesamiento de consultas paralelas eficientes y el manejo de datos masivos, y se ejecuta en nodos separados, a menudo es preferible para consultas costosas, incluso si las consultas están predeterminadas y podrían ser admitidas por un índice operacional. Además, Couchbase proporciona tanto la base de datos operacional como una base de datos de análisis escalable, todo en una plataforma NoSQL.