

Comparación de Despliegue de un Gestor de Base de Datos NoSQL Mediante Docker

MARKO ANTONIO RIVAS RIOS

JORGE LUIS MAMANI MAQUERA

ORLANDO ANTONIO ACOSTA ORTIZ

YOFER NAIN CATARI CABRERA

ORESTES RAMIREZ TICONA

ROBERTO ZEGARRA REYES

Universidad Privada de Tacna

Junio 22, 2019

Abstract

Docker is an open source project created in 2013 and which has been a revolution for the development and deployment of operations. Docker abstracts the host's hardware and operating system by running the applications in containers, isolated compartments that contain all the resources for an application or service. In this work we will see how to use Docker for the development of simple applications, learning how to deploy a NOSQL database with Docker.

Abstract

Docker es un proyecto open source creado en 2013 y que ha supuesto una revolución para el desarrollo y despliegue de operaciones. Docker abstrae el hardware y el sistema operativo del host ejecutando las aplicaciones en contenedores, compartimentos aislados que contienen todos los recursos para una aplicación o servicio.

En este trabajo veremos cómo usar Docker para el desarrollo de aplicaciones sencillas, aprendiendo a desplegar una base de datos NOSQL con Docker.

I. INTRODUCCIÓN

EL potente concepto de Microservicios está cambiando poco a poco la industria. Grandes servicios monolíticos están dando paso lentamente al enjambre microservicios pequeños y autónomos que trabajan en conjunto. El proceso va acompañado de otra tendencia del mercado: la contenerización. Juntos, ayudan a construir sistemas sin precedentes. La contenerización cambia no sólo la arquitectura de los servicios, sino también la estructura de ambientes utilizados para crearlos.

Ahora, cuando el software se distribuye en contenedores, los desarrolladores tienen plena libertad para decidir qué aplicaciones necesitan. Como resultado, incluso los entornos complejos, como los servidores de grandes bases de datos e infraestructura de análisis complejos pueden crear instancias en cuestión de segundos. El desarrollo de software se hace más fácil y más eficaz.

II. MATERIALES Y MÉTODOS

i. Creación de una base de datos NoSQL en docker

ii. Inserción de datos y Consulta de datos (en una base de datos NOSQL)

iii. Comparacion de distintos tipos de base de datos NoSQL

Tabla comparativa			
	MongoDB	Cassandra	DynamoDB
Descripción	Servicio de base de datos alojado y escalable de Amazon con los datos almacenados en la nube de Amazonas	Almacén de gran columna basado en ideas de BigTable y DynamoDB	Una de las tiendas de documentos más populares
Modelo de base de datos primaria	Document store: Las tiendas de documentos, también llamadas sistemas de bases de datos orientadas a documentos, se caracterizan por su organización de datos sin esquema. Key-value store: Las tiendas de valores clave son probablemente la forma más simple de sistemas de administración de bases de datos. Solo pueden almacenar pares de claves y valores, así como también recuperar valores cuando se conoce una clave.	Wide column store: Las tiendas de columnas anchas, también llamadas tiendas de discos extensibles, almacenan datos en registros con la capacidad de mantener un gran número de columnas dinámicas.	Document store: Las tiendas de documentos, también llamadas sistemas de bases de datos orientadas a documentos, se caracterizan por su organización de datos sin esquema.
Versión inicial	2012	2008	2009
Basado en la nube	sí	no	no
Lenguaje de implementación		Java	C ++
Esquema de datos	sin esquema	sin esquema	sin esquema
Índices secundarios	sí	restringido	sí
SQL	no	Sentencias DML y DDL similares a SQL (CQL)	no
API y otros métodos de acceso	API RESTful HTTP	Protocolo propietario Ahorro	protocolo propietario usando JSON
Lenguajes de programación admitidos	.Net Erlang Groovy Java JavaScript Perl PHP	C# C++ Clojure Erlang Go Haskell Java	C C# C++ Clojure info D info Erlang Go info

	Python Ruby	JavaScript info Perl PHP Python Ruby Scala	Groovy info Haskell Java JavaScript PHP PowerShell info Prolog info Python Ruby
Scripts del lado del servidor	no	no	JavaScript
Triggers	sí	sí	no
Métodos de particionamiento	Sharding	Sharding	Sharding
Métodos de replicación	sí	factor de replicación seleccionable	Replicación maestro-esclavo
Llaves Foraneas	no	no	no
Conceptos de transacción	no	no	no
Concurrencia	sí	sí	sí
Durabilidad	sí	sí	sí
Entornos de operación	hosting de amazon	BSD Linux OS X Windows	Linux OS X Solaris Windows
Ventajas	<p>Totalmente administrado</p> <p>Amazon DynamoDB es un servicio de base de datos NoSQL en la nube totalmente administrado; usted simplemente crea una tabla de base de datos, establece su rendimiento y deja que el servicio se encargue del resto.</p>	<p>Ningún punto único de falla garantiza el 100% de disponibilidad.</p> <p>Simplicidad operativa para el menor costo total de propiedad.</p> <p>La mejor escalabilidad de las plataformas NoSQL.</p>	<p>MongoDB mantiene las características más valiosas de las bases de datos relacionales: coherencia fuerte, lenguaje de consulta expresivo e índices secundarios. Como resultado, los desarrolladores pueden construir aplicaciones altamente funcionales más rápido que las bases de datos NoSQL.</p> <p>MongoDB proporciona flexibilidad de modelo de datos, escalabilidad elástica y alto rendimiento y disponibilidad de bases de datos</p>

			NoSQL. Como resultado, los ingenieros pueden mejorar continuamente las aplicaciones y entregarlas a escala casi ilimitada en hardware básico.
Desventajas	<p>- DynamoDB es una base de datos NoSQL. Eso significa que no puede hacer consultas complejas o de unión en dynamoDB. Además, no ACID, ya que no es un RDBMS</p> <p>- No se pueden escribir objetos grandes (BLOB) en dynamo DB. Definitivamente puede almacenar metadatos para BLOBS en dynamoDB, pero objetos reales que puede almacenar en S3. La recuperación del ajuste S3 es otra cosa que debe tener en cuenta para obtener un mejor rendimiento, como el uso de orden lexicográfico, el uso el uso de orden lexicográfico, el uso aleatorio y la paralelización de los GET.</p>	<p>La desventaja más grande para Cassandra es que no es compatible con los índices B-Tree y, por lo tanto, no admite consultas de rango.</p> <p>-Los CounterColumns no son un sustituto del autoincrement de las bases relacionales. Los contadores no son muy confiables. Si está haciendo decrementos, este puede no ser el resultado deseado.</p>	<p>No utiliza joins. Se diseña como si nunca hubieran sido una opción.</p> <p>Se escala bien en un rango estrecho, pero otras soluciones NoSQL son mejores para escalar.</p> <p>Sin transacciones, excepto en el nivel de registro. Si desea una transacción, debe ajustar toda la transacción en un solo registro.</p> <p>Difícil de asegurar correctamente sin contar con una licencia Enterprise.</p> <p>No hay parches, tiene que hacer actualizaciones completas y las actualizaciones completas y las actualizaciones completas se emiten varias veces al año. Entonces tienes que construir un programa de parches alrededor de ellos. No hay garantía de que su actualización funcione con su controlador dado, por lo que las pruebas deben programarse antes de la actualización.</p>

CouchDB:

CouchDB es catalogado muchas veces como una base de datos NoSQL, un término que se hizo cada vez más popular a finales de 2009 y principios de 2010. Si bien este término es una caracterización más bien genérica de una base de datos, o almacén de datos, sí define claramente un descanso de SQL tradicional bases de datos. Una base de datos CouchDB carece de un esquema o estructuras de datos pre-definidos rígidos tales como tablas.

Los datos almacenados en CouchDB es un documento (s) JSON. La estructura de los datos, o documento(s), puede cambiar dinámicamente para adaptarse a las necesidades cambiantes. CouchDB es una base de datos que abarca por completo la web. Almacene sus datos con documentos JSON. Tenga acceso a sus documentos y consultar sus índices con su navegador web, a través de HTTP. Índice, combinar y transformar sus documentos con JavaScript. CouchDB funciona bien con la web moderna y aplicaciones móviles. Usted puede incluso servir aplicaciones web directamente de CouchDB. Y usted puede distribuir sus datos o sus aplicaciones, de manera eficiente mediante la replicación incremental de los CouchDB. CouchDB soporta configuraciones maestro-maestro con detección automática de conflictos. CouchDB viene con una serie de características, como la transformación de documentos sobre la marcha y notificaciones de cambio en tiempo real, que hace que el desarrollo de aplicaciones web una brisa. Incluso viene con un fácil utilizar la consola de administración web.

Las principales características son las siguientes:

- **Almacenamiento de documentos:**

Almacena los datos como documentos esto es, uno o más pares campo/valor expresados en JSON. Los valores de los campos pueden ser datos simples como cadenas de caracteres, números o fechas. Pero también se pueden usar listas ordenadas y vectores asociativos. Todos los documentos en una base de datos CouchDB tienen un identificador único y no requieren un esquema determinado.

- **Vistas e índices Map/Reduce:**

Los datos almacenados se estructuran por medio de vistas. En CouchDB, cada vista se construye por medio de una función JavaScript que actúa como la mitad Map de una operación map/reduce. La función recibe un documento y lo transforma en un único valor, retornándolo. CouchDB puede indexar vistas y mantener actualizados esos índices a medida que se agregan, eliminan o actualizan documentos.

- **Arquitectura distribuida con replicación:**

Se diseñó con teniendo en mente la replicación bidireccional (o sincronización) y la operación off-line. Eso significa que múltiples réplicas pueden tener cada una sus propias copias de los mismos datos, modificarlas y luego sincronizar esos cambios en un momento posterior.

- **Interfaz REST:**

Todos los ítems tienen una URI única que queda expuesta vía HTTP. REST usa los métodos HTTP POST, GET, PUT y DELETE para las cuatro operaciones básicas CRUD (Create, Read, Update, Delete) con todos los recursos.

- **Consistencia Eventual:**

Garantiza consistencia eventual para poder ofrecer tanto disponibilidad como tolerancia a las particiones.

- **Hecha para operar offline:**

Puede replicar datos a dispositivos (como smartphones) que pueden quedar offline y manejar automáticamente la sincronización de los datos cuando el dispositivo vuelve a estar en línea.

Neo4j:

Es una base de datos orientada a grafos escrita en Java, es decir la información se almacena de forma relacionada formando un grafo dirigido entre los nodos y las relaciones entre ellos. Se integra perfectamente con múltiples lenguajes como Java, PHP, Ruby, .Net, Python, Node, Scala, etc. La base de datos está embebida en un servidor Jetty. Está especialmente indicada para modelar redes sociales y sistemas de recomendación.

Se distribuye en dos versiones: la community edition (open source) y la Enterprise edition. Para hacer pruebas de concepto nos basta con la community edition pero si quieres sacarle todo el partido a Neo4j la opción enterprise es la más recomendable ya que permite ponerla en cluster, monitorización, backups en caliente y un sistema de cache de alto rendimiento, además de soporte de sus creadores.

Otra de las ventajas que tiene Neo4j es que se pueden efectuar las consultas directamente a través de un API Rest lo que hace especialmente interesante su integración con aplicaciones web.

Principales características de neo4j:

- Alto desempeño y alta disponibilidad (Escalamiento de lectura) Soporte sólido y real para transacciones ACID.
- Escalable: 32 miles de millones de Nodos, 32 miles de millones de Relaciones, 64 miles de millones de Propiedades.
- Servidor con una API REST o usable como una biblioteca Java.

III. RESULTADOS

Comparaciones de 2 Bases de Datos NoSQL

i. Grafos

ii. Tabular (Column-Store)

iii. Documental

iv. Clave-Valor

IV. DISCUSIÓN

V. CONCLUSIONES

REFERENCES

1. <http://revistatelematica.cujae.edu.cu/index.php/tele/article/view/23/21>
2. <https://programarfacil.com/blog/que-es-un-orm/>
3. <https://www.beeva.com/beeva-view/tecnologia/mas-alla-de-la-virtualizacion-contenedores/>
4. <https://searchdatacenter.techtarget.com/es/definicion/virtualizacion-basada-en-contenedores-virtualizacion-a-nivel-de-sistema-operativo>
5. <https://www.incibe-cert.es/blog/asegurando-virtualizacion-tus-sistemas-control>
6. <http://www.datakeeper.es/?p=716>
7. <https://sigmodrecord.org/publications/sigmodRecord/1012/pdfs/04.surveys.cattell.pdf>
8. <https://ieeexplore.ieee.org/abstract/document/6625441>
9. <http://nosql-database.org/>