

Cassandra

Integrantes: Jeisson Andrés Carrillo Alvarado
Omar Ocampo Rivera



¿ Que es Cassandra ?



- Apache Cassandra es una base de datos NoSQL distribuida y basada en un modelo de almacenamiento de «clave-valor»
- Código abierto que está escrita en Java.
- Permite grandes volúmenes de datos en forma distribuida.

Arquitectura Distribuida



- Cassandra está basada en una serie de nodos iguales que se comunican con un protocolo P2P con lo que la redundancia es máxima.



Características



- Cassandra ofrece soporte robusto para múltiples centros de datos
- Replicación asincrónica sin necesidad de un servidor maestro
- Permiten operaciones de baja latencia para todos los clientes.
- También ofrece un gran rendimiento.
- **Descentralizado**
- **Escalabilidad**

Características

- Escalabilidad
- Tolerancia a fallos
- Consistencia



Usos en Estudios investigación



- En 2012, investigadores de la Universidad de Toronto que estudian los sistemas NoSQL concluyeron que "En términos de escalabilidad, hay un claro ganador a través de nuestros experimentos.
- Cassandra logra el más alto rendimiento para el número máximo de nodos en todos los experimentos", aunque "esto tiene como precio una alta latencia de escritura y lectura"

Herramientas de Cassandra



- cqlsh: the CQL shell
- Nodetool
- Usage
- Cassandra Stress

Administración y Monitoreo



- un sistema a base de Java que puede ser administrado y monitoreado con Java Management Extensions (JMX). Nodetool, por ejemplo, puede ser usado para administrar un clúster

Uso Académico



- BigTable
- Distributed database
- Distributed hash table (DHT)
- Dynamo (storage system) - Cassandra toma muchos elementos de Dynamo
- NoSQL

BigTable



- Es un sistema de gestión de base de datos creado por Google con las características de ser: distribuido, de alta eficiencia y propietario.
- **BigTable** almacena la información en tablas multidimensionales cuyas celdas están, en su mayoría, sin utilizar.

NO SQL



- Recordemos que SQL es un lenguaje de consulta estructurado, SQL nos permite acceder a las bases de datos y obtener cualquier información o realizar operaciones con los datos. Utiliza el álgebra y el cálculo relacional para obtener los datos de una manera sencilla.
- Es una base de datos relacional de código abierto que no usaba un lenguaje de consultas SQL

NOSQL VS Bases de datos relacionales

Feature	NoSQL Databases	Relational Databases
Performance	High	Low
Reliability	Poor	Good
Availability	Good	Good
Consistency	Poor	Good
Data Storage	Optimized for huge data	Medium sized to large
Scalability	High	High (but more expensive)

Versiones de Cassandra

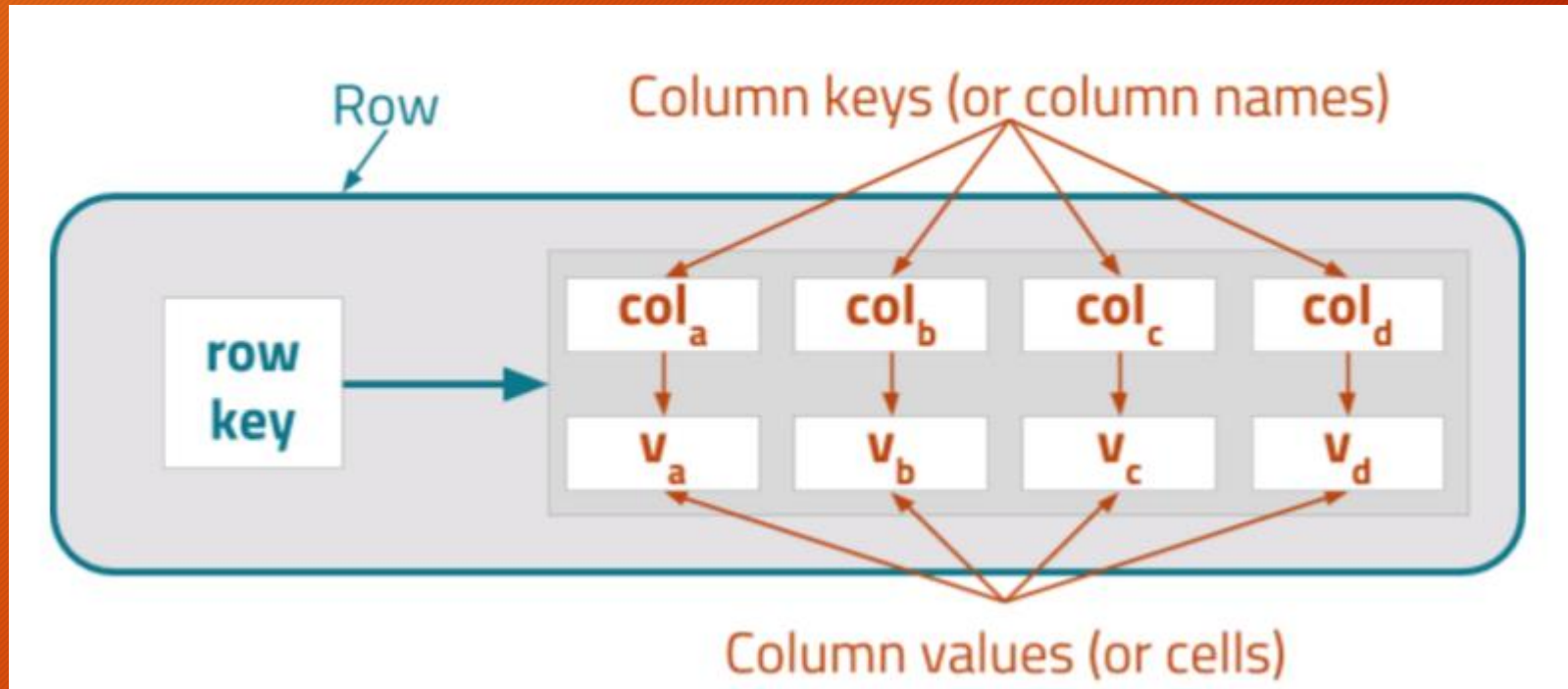


Version	Original release date	Latest version	Release date	Status ^[15]
0.6	2010-04-12	0.6.13	2011-04-18	No longer supported
0.7	2011-01-10	0.7.10	2011-10-31	No longer supported
0.8	2011-06-03	0.8.10	2012-02-13	No longer supported
1.0	2011-10-18	1.0.12	2012-10-04	No longer supported
1.1	2012-04-24	1.1.12	2013-05-27	No longer supported
1.2	2013-01-02	1.2.19	2014-09-18	No longer supported
2.0	2013-09-03	2.0.17	2015-09-21	No longer supported
2.1	2014-09-16	2.1.20	2018-02-16	Still supported, critical fixes only
2.2	2015-07-20	2.2.12	2018-02-16	Still supported
3.0	2015-11-09	3.0.16	2018-02-19	Still supported
3.11	2017-06-23	3.11.2	2018-02-10	Latest release
Legend: Old version Older version, still supported Latest version Latest preview version				

Modelo de datos



- También combina propiedades de una base de datos clave-valor y una orientada a columnas.



En qué se diferencia Cassandra de otras bases de datos NoSQL



- La alta disponibilidad es una de las características principales de esta base de datos, pero no la única. De hecho, es todo el conjunto de características de esta base de datos lo que la hace única.
- Cassandra es una base de datos distribuida, ¿Qué quiere decir esto? Pues bien, la información que se almacena se reparte a través de todos los nodos que componen el clúster, de manera que si algún nodo se cae o se degrada va a ofrecer alta disponibilidad, asegurando el servicio on-line de la base de datos.

En qué se diferencia Cassandra de otras bases de datos NoSQL



- Además, algo que puede marcar la diferencia con respecto a otras bases de datos, tanto tradicionales (RDBM) como NoSQL, es que el sistema se puede escalar de forma horizontal añadiendo nodos basados en *hardware commodity* de bajo coste

¿Cuándo Cassandra es la mejor solución?



- Si buscamos una base de datos que nos permita una escalación lineal con alto rendimiento y además alta disponibilidad.... bueno, ésta es una de sus grandes características. La elegiremos sin duda.
- Igualmente, su arquitectura distribuida y la manera en que se almacenan los datos va a permitir tener una alta disponibilidad de la base de datos, evitando caídas que pueden suponer pérdida de dinero o pérdida de datos.... cuando no, ambas a la vez.

Ejemplos



```
cqlsh> create keyspace sparktestcassandra with replication = {'class': 'SimpleStrategy', 'replication_factor': 1};
AlreadyExists: Keyspace 'sparktestcassandra' already exists
cqlsh> use sparktestcassandra ;
cqlsh:sparktestcassandra> create table emp (empid int primary key, emp_first varchar, emp_last varchar, emp_dept varchar);
AlreadyExists: Table 'sparktestcassandra.emp' already exists
cqlsh:sparktestcassandra> select * from emp;
```

empid	emp_dept	emp_first	emp_last
-------	----------	-----------	----------

(0 rows)

```
cqlsh:sparktestcassandra> insert into emp (empid, emp_first, emp_last, emp_dept) values (1, 'fred', 'smith', 'eng');
cqlsh:sparktestcassandra> insert into emp (empid, emp_first, emp_last, emp_dept) values (2, 'mathan', 'pillai', 'cogni');
cqlsh:sparktestcassandra> insert into emp (empid, emp_first, emp_last, emp_dept) values (3, 'giri', 'r', 'databricks');
cqlsh:sparktestcassandra> insert into emp (empid, emp_first, emp_last, emp_dept) values (4, 'parthi', 'pillai', 'cogni');
cqlsh:sparktestcassandra> insert into emp (empid, emp_first, emp_last, emp_dept) values (5, 'mathan', 'pillai', 'cogni');
cqlsh:sparktestcassandra> select * from emp;
```

empid	emp_dept	emp_first	emp_last
-------	----------	-----------	----------

5	cogni	mathan	pillai
1	eng	fred	smith
2	cogni	mathan	pillai
4	cogni	parthi	pillai
3	databricks	giri	r

(5 rows)

```
cqlsh:sparktestcassandra> █
```

Conclusión



- Cassandra es una solución brillante para muchos casos de uso que podemos encontrar en el mundo Big Data.
- Cassandra no es una base de datos cualquiera. Ofrece innumerables beneficios con respecto a otras
- Mejor Tolerancia a fallos.
- No requiere tiempos de espera durante mantenimientos.
- Puede ser muy escalable.
- Es gratuita.



· Muchas
! Gracias !