

Real Time Ingestion from Cassandra Ring

Neha.C.D

National Institute Of Engineering

ABSTRACT

Cassandra is fault tolerant,scalable Nosql database. It is ideal for supporting high throughput low latency writes. Our aim is to build a generic ingestion adaptor for cassandra that,ingests data in real time.

Keyword: Application, Nosql database

I. INTRODUCTION

Cassandra is a fault tolerant, preformant , decentralized and scalable NOSQL database. It is ideal when we need to support high throughput, low latency writes. Data in Cassandra is distributed over several nodes and these nodes together form what is called Cassandra ring. To build a system that can 'ingest' data from a Cassandra ring and transfer it our data late.

Database vendors usually provide a mechanism by which change events in the database are captured and emitted to external systems. This is common especially in database replication technologies.Ingestion can take advantage of this change data capture and transmit this to the data link.

II. OBJECTIVE

The objective is to build generic ingestion adapter for cassandra that, ingests change data in real time, which doesn't impact the performance of transactional systems that also access the database for reads and writes and constant performance and latency regardless of the amount of data in the database.

III. EXISTING SYSTEM

Database systems are mainly Relational Database, which needs to be atom compliant. They pose major problem in Scalability since they are only vertically scalable. Most of the data is unstructured these days, which cannot be handled efficiently by databases like RDBMS.

IV. PROPOSED SYSTEM

To overcome the problems of above, we use Nosql database where Cassandra is one such database. this support horizontal scaling and support all kind of data whether its structured, unstructured or semi-structured.

V. NOSQL DATABASE

Nosql database is "not only database" or "non relational database" database provides a system for storage and retrieval of data that is ideated in means other than sequential relations used in RDBMS.

Nosql databases are majorly used in big data applications as well as real time web applications. Nosql database supports SQL like query languages.

There is a simplicity of design in Nosql database ,it is unstructured database and more simpler than the SQL database. Nosql includes very less query languages. It stores in the favour of the

- Availability
- Partition tolerance
- Speed

Cassandra is one of the example for the Nosql database.

There is no any other properties to span the data in RDBMS.

Nosql database has write ahead logging to avoid data loss. Distributed data processing accross multiple data and data consistency is the biggest challenge for both the Nosql and relational database.

VI. ABOUT CASSANDRA

Cassandra is a free open source No sql database. Cassandra stores the data in the form of key value pairs. It uses the clustering form. Cassandra is highly robust.

- Cassandra has proved its constant use at Github, GoDaddy, instagram,Netflix uses large data sets.
- Cassandra is fault tolerant replication of data to multiple nodes.
- Cassandra performs Nosql alternatives in benchmarks and inreal application.
- Cassandra is decentralized and all the nodes present in the cluster is identical.
- Scalable
- Durable
- Elastic

VII. CONCLUSION

To build a system that can 'ingest' data from a Cassandra ring and transfer it our data late.

VIII. REFERENCE

- <https://apachecassandra.com>
- <https://nosqldatabase.com>
- <https://cassandradatabase.com>