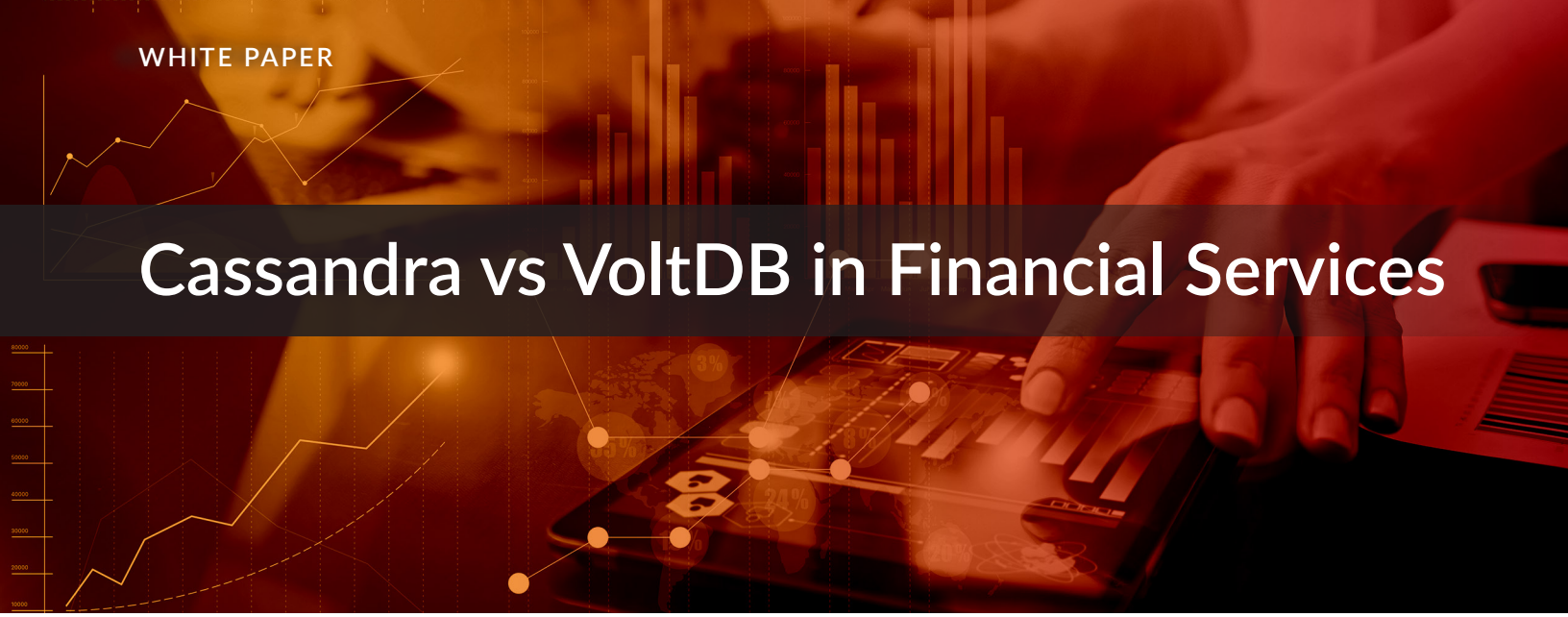
WHITE PAPER



Cassandra vs VoltDB in Financial Services

Cassandra has become a very popular NoSQL database. Cassandra, like other NoSQL databases, was originally designed to solve the scalability and availability problems in web-scale applications and it does that well. Using it in financial services should be given serious consideration due to some of its limitations:

1 ACID Guarantees

Most of the banking and financial applications require strict ACID guarantees. Having multiple copies of inconsistent data (i.e., different account balances, stale security prices) can cause significant problems in the financial transactions. Cassandra is an AP (Availability & Partition Tolerance) database that sacrifices consistency for availability. VoltDB is 100% ACID compliant relational database that delivers always consistent and correct data that can be safely used for critical financial transactions.

2 SQL vs CQL

Cassandra's CQL offers very limited features compared to the standard SQL despite the similar names. For example, there is no Join or Group By in CQL. CQL also provides limited support on aggregation and even simple data aggregation functions like MIN, MAX, SUM, AVG are very resource intensive in Cassandra because it is a key-value database. VoltDB uses ANSI Standard SQL-99 and it supports most of the commonly used aggregation and math SQL functions in the financial services industry.

3 Structured vs. Unstructured Data

Cassandra is a scalable key-value database that was originally developed to manage Facebook data. It works the best for large amount of unstructured data but not for the structured data like financial accounting, trade records, customer transactions, compliance and market data which are the most important data to the organizations in the financial services. VoltDB is a NewSQL high performance relational database that minimizes application complexity and maximizes usability of structured financial data.

4 Data Modeling

Data modeling in Cassandra is based on query patterns. If the query pattern changes, you need to change the data model and migrate data which can affect many applications and business processes such as research, trade, audit, compliance etc. This makes data modeling very challenging in Cassandra because it is difficult to predict the query patterns for future use cases. VoltDB is a relational database that consists of tables and columns, with constraints, indexes, and views. VoltDB uses standard SQL database definition language (DDL) statements to specify the database schema. Designing and maintaining the schema for a VoltDB database uses the same skills and knowledge as managing a database for Oracle, MySQL, or any other relational database.

Single Source of Truth

Single Source of Truth (SSOT) refers to an architectural practice of accessing and updating data from one place that stores the complete picture of the data objects as a whole. VoltDB is a fully ACID compliant and relational database designed to be a secure and safe source of the Single Source of Truth data store for financial transactions.

VoltDB can be used as a primary Operational Data Store (ODS) to create a SSOT by compiling, cleansing, enriching the data from multiple System of Records (SORs) in real-time. It is critical for the ODS to read, insert, update records in real-time as it arrives in order to achieve a successful SSOT. Cassandra performs poorly on inserting and updating large amount of records so using Cassandra as the primary ODS for heavy financial transaction data is not a good choice.

In an organization with SSOT, when a record changes the entire system in the organization should see the same change at the same instance. An eventual consistency database like Cassandra is not a good fit for SSOT for time sensitive financial transactions and operations. VoltDB guarantees 100% transactional consistency in real-time data processing and ensures the data is always accurate no matter how your application accesses it.

Near Real-Time (NRT) vs. Real Real-Time (RRT)

In financial services, analyzing of information and taking an appropriate action in real-time is more than just a nice idea – it's essential. Preventing fraudulent credit card transactions, identifying suspicious trade activities, monitoring risk exposures and incremental risks in portfolio management must to be done in real-time in order to make the most effective business decisions.

Cassandra is often coupled with Apache Spark Streaming to build a streaming analytics solution but one should note that Apache Spark Streaming is mini-batch, not true real-time. In mini-batch, records of streaming data are collected in a buffer and periodically executed as scheduled. In other words, Spark splits the stream into mini-batches for faster processing but it will never be Real Real-Time. That's why we call it Near Real-Time solution compared to Real Real-Time solution like VoltDB. Real Real-Time data processing requires continuous processing of data and immediate outputting of actionable business information as data arrives. VoltDB is a RRT solution to fuel real-time applications that can make data-driven decisions on each event as it enters the data pipeline.

There is a fundamental difference between NRT and RRT use cases in financial services. NRT is often used to build a control that detects and alerts suspicious activities such as fraud detection system. RRT is used to build highly time-sensitive and critical control such as fraud prevention system which is designed to keep fraud from occurring.