A Single Source of Truth —
is it the mythical creature of data management?

In the world of data management, a "single source of truth" is a fully trusted data source — the ultimate authority for the particular data it provides. It is important for the smooth operations of a company since applications often need an absolute authority for a given bit of data, say a customer's credit limit. "Single source of truth" is often synonymous with a "system of record" but, as we will see, that creates data management challenges.

In an organization with SSoT, when a record is changed, the entire organization should be able to see the change at the same instant. But legacy systems, disparate SaaS applications with divergent data formats and different user needs can be obstacles to establishing a single source of truth.

Several approaches have been proposed to create and maintain a Single Source of Truth data source. These include:

- **User Data Management** (UDM) — a unified data model which serves as a system or architecture that provides a 'bridge' between data sources, against which a user or application may execute queries.

- **Master Data Management** (MDM) — a method or model that enables an enterprise to link all critical data to one 'master' file, which provides a single point of reference to the master data. MDM promises to increase collaboration and reduce data silos.

- **System of Record** — a database. application or data store that is the authoritative source for a given data record or piece of data.

- **Enterprise Service Bus** — A precursor to message queues like Apache Kafka, ESBs enable multiple systems to be updated when a record is updated (published) in another, connected system.

VOLTDB

A common problem with each of these approaches is the difficulty in establishing a 'Golden Record' — the master source of a data record. A financial institution, for example, might have data about a customer's identity stored in a mobile banking app (name, account number(s), balance), a Customer Relationship Management system (name, account number, physical address, SSN and/or tax ID), and an email marketing system (name, account number, email address, communications preferences, activity). Each of these systems of record might collect and store data in different ways, resulting in multiple different formats for name fields for the same person, for example. Each attribute will be valid in its own silo/application, but the challenges of unifying them into a single, correct view of the data — a Single Source of Truth about the customer — is challenging from a technical, organizational and operational standpoint.

Single Source of Truth vs. System of Record

A System of Record (SoR) is an information system, an application, or a system or application backed by a database or data store that is the authoritative source for a given data element or piece of information. Thus, a SoR is a system, application or data store; a Single Source of Truth is an architectural approach or the practice of using only one source for any given data element or record.

As we have seen, an enterprise may have many SoRs, each governed by different business units or departments, and each conceivably using different data formats, taxonomies, and semantics to serve different use cases. At the data management layer, these SoRs are overwhelmed by daily lookup requests, with some writes thrown in, so one approach is to add a caching layer that can take the brunt of ad-hoc queries throughout the day, buffering content and reducing the load on the SoRs.

Single Source of Truth (SSOT) refers to an architectural practice of accessing and updating data from one place that stores a unified, complete picture of the data objects. If limited to reads, many products could be used to present this unified view to corporate business intelligence users and satisfy casual ad-hoc queries. But when an enterprise's data management team wants to have a caching layer take writes to correct or update records and you want that new information to be guaranteed to flow back to the systems of record, they will need to have strong ACID, particularly isolation, to make sure the records are updated consistently and logically based on the requests.

Why is having a Single Source of Truth important?

Enterprises with multiple Systems of Record face difficulties reconciling different sources of data to inform operations. Keeping customer data, for example, in silo'd applications using different data formats makes it impossible to trust a data record, which makes it difficult to act with confidence. Data quality may be suspect, since each data record has been created by different people, with different skill levels, using different applications. It can be daunting to link different technologies and applications to achieve a single view of a customer's data.

In an ideal world, an enterprise would have a 'single source of truth' — an information architecture in which each data element is stored only once. In the real world, with a proliferation of legacy applications and systems, an SSoT may not seem feasible. Yet many organizations seek to add a data management platform to front-end multiple Systems of Record (SoR) to create the "single source of truth" for their operational and analytical data.

VOLTDB

The benefits of having a Single Source of Truth include:

- **Omnichannel experience** — if a customer's information is consistent throughout touchpoints — for example in a mobile operator's billing systems, authentication systems, CRMs, mobile apps, web portal and marketing systems — the mobile operator will be able to provide each customer with a consistent experience of the brand.

- **Personalization** — if a customer's information is consistent throughout touch-points — for example in a mobile operator's billing systems, authentication systems, CRMs, mobile apps, web portal and marketing systems — the mobile operator will be able to personalize interactions with each customer, making appropriate and timely offers.

- **Improved customer experience**

- **Cross-sell, up-sell opportunities** — the mobile operator in the example above will have more opportunities to cross-sell and upsell customers about whom the operator has more consistent, perfect data.

Why VoltDB?

In-memory databases with strong ACID compliance are perfect solutions for establishing a SSoT, especially in the caching use case.

VoltDB is a fully ACID-compliant relational database designed to be a secure and safe source of the Single Source of Truth data store for transactions in many industries. In an SSoT system, VoltDB can be used as the primary ODS that stores the complete picture of data objects as a whole within the enterprise. While NoSQL systems are often used for this use case, because IT doesn't have to worry about defining schema when they ingest data from different sources especially for unstructured data, most offerings are eventually-consistent, and thus not a good fit for SSoT for time-sensitive transactions and operations. VoltDB guarantees 100% transactional consistency in real-time data processing and ensures data is always accurate no matter how an application accesses it. NoSQL and open source offerings also present IT and data management teams with issues with poor or non-standard query models, lack of immediate consistency, and no guarantees as to correctness of data.

VoltDB can be used as a primary Operational Data Store (ODS) to create a SSoT by compiling, cleansing, and enriching data from multiple Systems of Records (SoRs). It is critical for the ODS to read, insert, update, filter, and aggregate data in real-time as it arrives to achieve a successful SSoT.

VOLTDB

VoltDB is an ideal choice for SSoT because it offers:

- **ACID compliant transactions**. ACID compliance is the best way to ensure the data integrity in your DBMS and is most preferred for developers of modern mission-critical financial systems to achieve SSoT. Financial transactions such as credit/debit card transactions, trading executions, and wire transfers are very costly or often not possible to reverse once executed; therefore, having the most accurate data to make the right business decisions at the time of each transaction is critical.

- **In-memory technology for performance**: In order to satisfy the needs of real-time analytics on live streaming financial market data from multiple sources, you need a high performance in-memory database. An average cost of one megabyte of RAM was about $100 in 2000s. Each megabyte of RAM costs less than a penny today. In-memory technology has become a very cost-effective tool for operational data stores that require high throughput and predictable low latency.

- **Scalability**: Scalability is an important aspect of systems in banking and financial institutions. As of 2016, an average annual growth rate in data for investment banks was 40%. This rate is expected to continue or even accelerate in the foreseeable future. VoltDB scales horizontally by adding servers to a shared-nothing cluster. It scales linearly to millions of transactions per second to meet the growing needs of institutions' data processing requirements.

- **High Availability and Disaster Recovery**: Downtime is not an option for today's mission critical financial transaction systems. Customers expect 24/7 transactions, and global capital markets never sleep. With VoltDB there is no single point of failure. Built-in database cluster replication with XDCR (Cross Datacenter Replication) ensures your system will be up and running all the time.

- **Feeds to downstream systems**: Even with rapidly declining memory prices, you would not want to keep the data that you don't need for current operations and real-time analytics in main memory. You need the technology that can safely and efficiently move those historical data into your downstream systems. VoltDB can export data to HTTP, Kafka, HDFS, etc. with its default exporters, or build your own custom exporter.

- **Structured and unstructured data**: Most mission critical data for financial service organizations is structured data (i.e., market data, transactions, account data, trade records etc.), but some interaction data may be semi-structured or unstructured. VoltDB is a relational database that maximizes the usability of highly-structured financial data, but it also supports JSON column values. With VoltDB you can work with JSON data while maintaining the efficiency and transactional consistency of a relational database.

VOLTDB

For more information about how VoltDB can be used to create a Single Source of Truth for your institution, contact us at **info@voltdb.com**

Sources

https://medium.com/zetta-venture-partners/beyond-systems-of-record-7b9777695f38

https://www.linkedin.com/pulse/difference-between-system-record-source-truth-santosh-kudva

https://blogs.informatica.com/2015/05/08/golden-record/#fbid=ZtiftjBhFLj

https://www.cbinsight.com/the-benefit-a-single-source-of-truth-brings-to-customer-data.html