splice MACHINE

Take the ACID Test with Splice Machine Data Warehouse and RDBMS in One

Data Warehouse & RDBMS in One

Modern applications need to combine fast data ingestion, web-scale transactional and analytical workloads, and continuous machine learning. Splice Machine RDBMS is architected to support these Intelligent Applications, with a hybrid data platform that can process OLAP and OLTP concurrently in real time.

Take the ACID Test with Splice Machine

Splice Machine provides CRUD operations with fully ACID-compliant transactions, which scale both to high-frequency operational queries and to very large-scale analytic workloads. You can review this capability by taking the ACID test, an easy way to set up Splice Machine, along with data-sets for analytical and transactional workloads on an AWS cluster. To get started, go to:

www.splicemachine.com/acidtest/

After you fill out some basic information and provide your AWS credentials, a CloudFormation script will set up your EC2-cluster, Hadoop, Splice Machine and the data-sets. You can then run OLTP and OLAP workloads concurrently and experience how easy Splice Machine is to operate.

The hybrid architecture of Splice Machine makes it the perfect data platform for Intelligent Applications in industries such as Cyber Security, Financial Services, Healthcare, Manufacturing, Marketing Automation, Retail, Supply Chain. Now it is easy for you to see how it will work for your application!

4 Reasons Why You Need a Hybrid OLAP/OLTP Architecture

1. **Simplify Operational Complexity**
   Avoid having to manage separate systems, tuning each separately for performance and writing low-level code and batch programs to keep them in sync.

2. **Eliminate Need for Special Coding Skills**
   Developers can use a single industry-standard SQL and JDBC/ODBC interface to work with the system.

3. **Power Concurrent Applications**
   The ACID transaction implementation is designed for both analytical and operational workloads. This means that it supports high concurrency with even thousands of users or devices updating the system at the same time. Its MVCC, using snapshot isolation, can handle fine-grained updates without locking reads.

4. **Support Machine Learning**
   Modern applications adapt over time by continuously transforming operational data into aggregated features that train statistical machine learning models and deploy those models in real-time decision systems. Splice Machine enables the feature engineering, model selection, and deployment process to take place on one platform without significant data movement.

Create Real-Time, Scalable Applications with the Hybrid Splice Machine Data Platform

Splice Machine offers a better solution to the complexity of separate RDBMS's and Data Warehouses. With Splice Machine, you can now get all the benefits of these separate systems, but with a much simpler architecture.

For example, here's how a machine learning application can use Splice Machine:

- **Batch File Ingestion –** Imports of raw data files are directly inserted into sharded tables in parallel with indexes that are atomically updated with the data for fast access
- **Real-Time Stream Ingestion –** Stored procedures continuously ingest streams with standard SQL and auto-shards
- **Data Cleansing –** Use standard SQL, with constraints and triggers, to clean up small subsets of data as well as entire data sets efficiently, without big batch runs or file explosions
- **Feature Engineering and Extensive ETL –** Execute complex aggregations, joins, sorts, and groupings with efficient SQL that is automatically parallelized and optimized without writing code at the application level
- **Model Training –** Stored procedures execute analytics directly on the data, for example, using built-in functions like ResultSetToRDD that take SQL results and treat them as Spark RDDs or execute R and Python libraries directly on database result sets
- **Application Logic –** ACID semantics enable the architecture to power concurrent CRUD applications without additional moving parts
- **Model Execution –** Stored procedures and user-defined functions wrap models
- **Reporting and Data Visualization –** Use Tableau, Domo, MicroStrategy and other ODBC/JDBC compliant tools turnkey

Why Choose Splice Machine

Splice Machine is the open-source SQL RDBMS, powered by Apache Hadoop® and Apache Spark™.

The Splice Machine RDBMS provides:

- **ANSI SQL –** Splice Machine provides ANSI SQL-99 coverage, including full DDL and DML
- **ACID Transactions –** Splice Machine provides CRUD operations with fully ACID-compliant transactions, which scale both to high-frequency operational queries and very large-scale analytic workloads
- **Secondary Indexing –** Splice Machine supports true secondary indices on data, in both unique and non-unique forms
- **Referential Integrity –** Referential integrity, such as Primary and Foreign key constraints, can be enforced without requiring any behavior from the underlying application
- **Joins –** Splice Machine supports inner, outer, cross and natural joins using join algorithms such as broadcast, merge, merge sort, batch nested loop, and nested loop joins

By centralizing on a relational hybrid architecture on Splice Machine, teams can build machine learning applications very quickly, maintain them with standard operational personnel, and be able to tightly integrate machine learning into the application without extensive efforts to integrate disparate tehnologies.
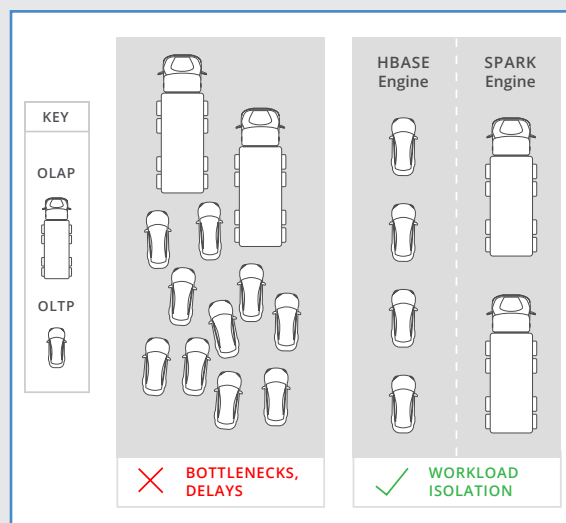
Workload Isolation – The cost-based optimizer chooses a dataflow engine based on the estimation of the query plan – OLTP runs on HBase and OLAP runs on Spark