Splice Machine & Talend Enable Customers to Streamline Data Workflows on Hadoop

A common struggle for data-driven enterprises is managing unnecessarily complicated data workflows with bloated ETL pipelines and a lack of native system integration. Designed to solve these big data integration issues, Splice Machine and Talend have partnered to automate data workflows and cohesively bring together IT architecture to power real-time operational applications that are overwhelmed by increased data growth and cost pressures.

Splice Machine leverages the proven ability of Hadoop and Apache Spark™ to scale seamlessly to petabytes of data with commodity servers. Splice Machine can parallelize queries in a shared-nothing architecture, enabling it to deliver compelling results versus traditional RDBMS like Oracle, Microsoft SQL Server and IBM DB2:

- 10–20x increase in query speeds
- 75% reduction in TCO
- 10–20x better price/performance

The Best of All Worlds

Splice Machine's partnership with Talend provides business the best of all worlds: a standard SQL database, the proven scale-out of Hadoop, and outstanding in-memory performance for OLAP queries with Spark.

Customers can benefit from two different approaches for managing their data integration needs with Splice Machine and Talend.

1. Customers migrating from legacy databases to Splice Machine can continue to use their existing Talend ETL process with minimal rewrites. Talend connects to Splice Machine via JDBC and all existing ETL workflows can be effortlessly migrated over to Splice Machine.

2. In order to fully maximize the power of Hadoop's parallel processing capabilities, customers can use Talend Open Studio for Big Data to move data from source systems into HDFS. They have the option of performing transformations on that data in Hadoop, via Open Studio, without having to do any coding. They can then invoke Splice Machine's bulk loader to ingest the data from HDFS into the Splice Machine database.

Splice Machine

OVERVIEW

The Splice Machine RDBMS is the first hybrid, in-memory RDBMS powered by Hadoop and Spark. Leveraging in-memory technology from Spark and scale-out capabilities from Hadoop, Splice Machine can replace Oracle® and MySQL™ databases, while increasing performance by 10–20x at one-fourth the cost. With an innovative, hybrid architecture and advanced resource isolation, the Splice Machine RDBMS provides exceptional performance for simultaneous OLAP and OLTP workloads, enabling companies to unlock the insights in their Big Data to make decisions in the moment.
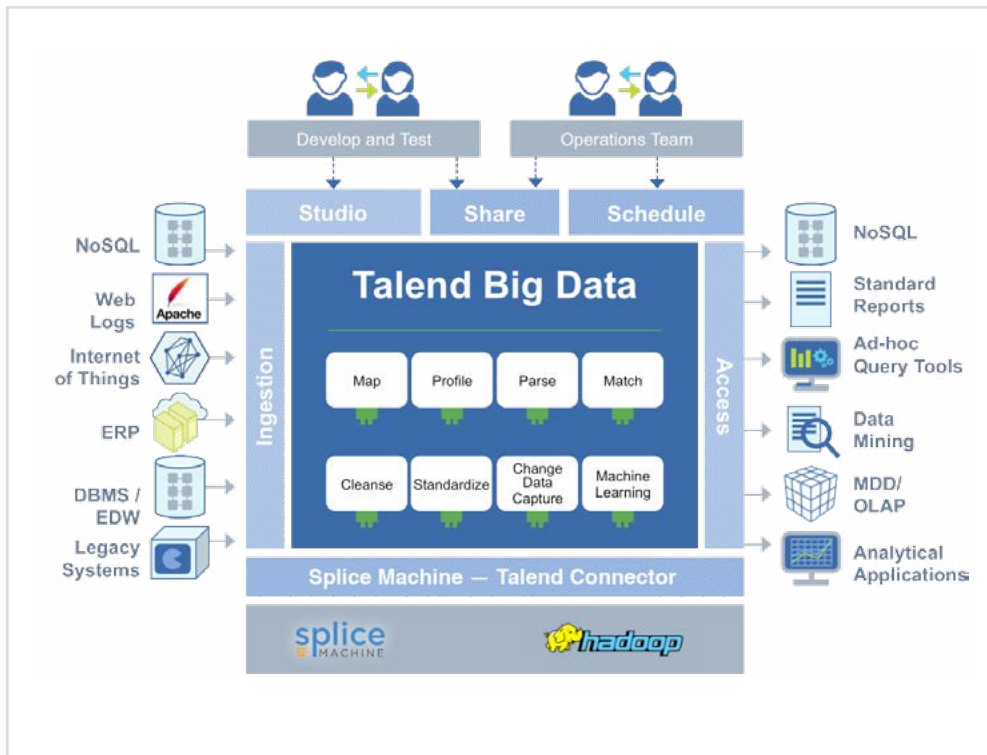
SOLUTION HIGHLIGHTS

- ACID compliant transactions
- ANSI standard SQL
- Complex joins
- Secondary indexes
- Aggregations
- Sub-queries
- Triggers
- User-defined functions (UDFs)
- Column-level security

SOLUTION BENEFITS

- Power both operational applications and reports that require data updated in real time
- Support existing applications with minimal migration effort with a true ANSI SQL-compliant database
- Query acceleration with advanced in-memory technology from Apache Spark
- Scale-out on commodity servers using the proven auto-sharding of HBase
- Reduce TCO by over 75% and increase by performance by 10–20x versus traditional scale-up databases
- State-of-the-art design allows very high throughput and avoids troublesome locks

The Eclipse-based Talend UI enables one to drag, drop and configure graphical components representing a wide array of Splice Machine related operations and the Talend software automatically generates optimized code for your target data platform. This easily created code can then be deployed as a stand-alone job, an executable, or a big data integration service.

Organizations planning to build operational data lakes with Splice Machine can augment Talend's data integration technology with its data quality capabilities. Talend data quality tools can profile data, identify anomalies, cleanse and monitor data quality on the Splice Machine database. It can also apply de-duplication, validation, standardization and enrichment processes.

About Splice Machine

The Splice Machine RDBMS is the first hybrid, in-memory RDBMS powered by Hadoop and Spark. Leveraging in-memory technology from Spark and scale-out capabilities from Hadoop, Splice Machine can replace Oracle® and MySQL™ databases, while increasing performance by 10–20x at one-fourth the cost. With an innovative, hybrid architecture and advanced resource isolation, the Splice Machine RDBMS provides exceptional performance for simultaneous OLAP and OLTP workloads, enabling companies to unlock the insights in their Big Data to make decisions in the moment.

www.splicemachine.com

The Benefits of Talend

INCREASE PRODUCTIVITY

- Provides rich, graphical tools and an easy to use, drag-and-drop UI
- Integrated solutions make testing and de-bugging simpler
- Talend Studio proven 30x more productive than hand-coding

LOWER TCO

- Leverage existing ETL and Java developers, eliminating need for specialized hires
- Subscription model is predictable with linear scaling

FUTURE-PROOF ARCHITECTURE

- 100x faster: optimized for Amazon EMR, Cloudera, Hortonworks and MapR
- Creates native Hadoop, Java, Spark & SQL, eliminating need to learn MapReduce
- Build once and run anywhere with in-cloud and on-premises deployments

About Talend

At Talend, it's our mission to connect the data-driven enterprise, so our customers can operate in real-time with new insight about their customers, markets and business. Founded in 2006, our global team of integration experts builds on open source innovation to create enterprise-ready solutions that help unlock business value more quickly. By design, Talend integration software simplifies the development process, reduces the learning curve, and decreases total cost of ownership with a unified, open, and predictable platform. Through native support of modern big data platforms, Talend takes the complexity out of integration efforts.