



Talend Big Data Integration with Hortonworks Data Platform



Partner Brief



"Companies that can make informed decisions faster win. Talend and Hortonworks offer the quickest way to get data in and out of Hadoop and deliver a real-time advantage when combined with Apache Spark and Storm."

Laurent Bride
CTO
Talend

To address the data demands of the growing digital universe, Hortonworks has partnered with Talend for easily connecting Apache Hadoop to hundreds of data sources without having to write code. Talend Big Data is a powerful and versatile, open source solution for big data integration that natively supports Apache Hadoop, including connectors for Hadoop Distributed File System (HDFS), MapReduce 2.0 (YARN), HCatalog, HBase, Pig, Sqoop, Oozie and Hive.

By leveraging Apache Hadoop's architecture for highly distributed data processing, Talend Big Data generates optimized Hadoop code – as native MapReduce, Spark or Storm – and runs data transformations directly inside Hadoop for maximum scalability. Its easy-to-use graphical development environment dramatically improves the efficiency of data integration job design.

Talend lets you spend more time analyzing your data instead of integrating data. Otherwise, up to 70% of your time spent doing data analysis involves data integration and resolving data quality issues. With Talend, you can quickly load, extract, transform, cleanse and process large and diverse data sets from disparate systems. Talend Big Data is the only solution to natively run data quality rules on Hadoop at infinite scale to parse, cleanse and match all of your data.

The Highest Performing Hadoop Code

Not all big data integration solutions perform equally. Talend generates optimized and high performing Hadoop code, so you can answer more questions in a shorter time period, and you do not need to do further performance tuning. TPC-H benchmark (www.tpc.org) tests show a 3 to 5X performance gain using Spark over MapReduce. Additional use cases highlight up to a 100X gain. Recent Talend benchmarks show a 53% performance increase in MapReduce code with Talend 5.6.



Develop Integration Jobs 10 Times as Fast

- Easily connect to 800+ data sources and applications including big data and NoSQL
- Graphical drag-and-drop tools, wizards, and team collaboration tools
- Visually optimize MapReduce jobs for faster development

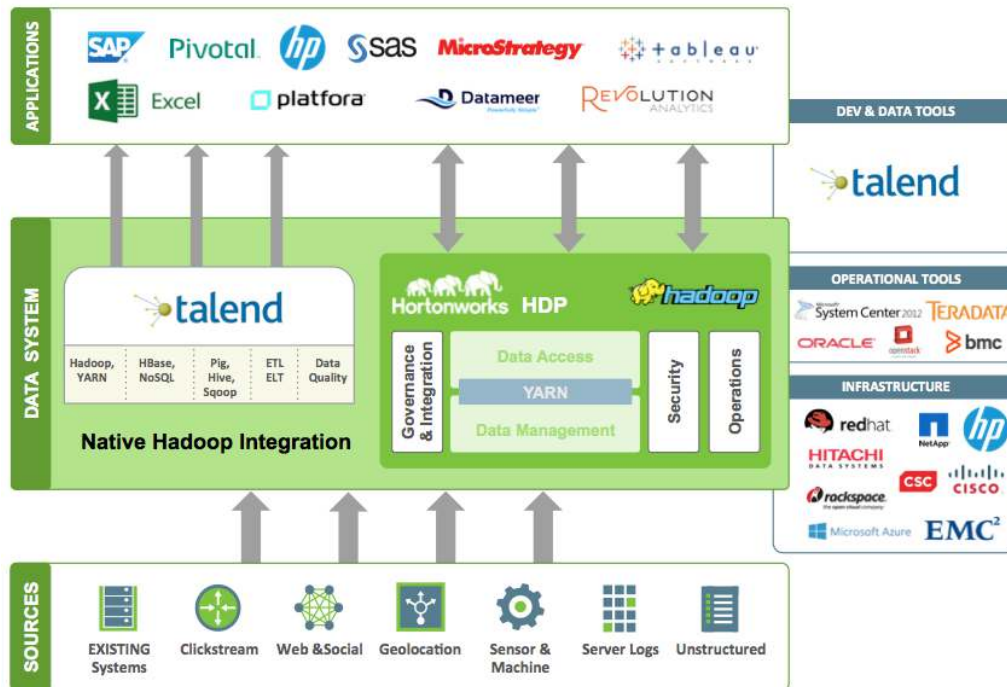
Run at the Speed and Scale of Hadoop

- YARN and Hadoop 2.0 support for better resource optimization
- Support for ETL and ELT, real-time and event driven delivery
- No need to learn MapReduce or tweak custom code

Protect your Investments with a Future-Proof Architecture

- The first integration vendor to implement MapReduce, YARN, Spark and Storm
- Software created through open standards and development processes

Talend in the Modern Data Architecture



Talend simplifies the movement of data between Apache Hadoop and enterprise data systems.

Data Warehouse Optimization

To reduce operational costs and improve data warehouse performance, firms are optimizing how they use and archive data, e.g. infrequently used data and aged data can be offloaded to less expensive Hadoop systems for storage and still be quickly retrieved for analysis.

With Talend and Hortonworks, the process of moving data between traditional data sources, Hadoop and enterprise data warehouses is greatly simplified, so firms can store more data while keeping costs flat.

Zero to Big Data in Less than 10 Minutes

The Talend Big Data Sandbox is a ready-to-run virtual environment that includes Talend Platform for Big Data, Hortonworks Data Platform and big data examples. Download your free sandbox at: <http://www.talend.com/talend-big-data-sandbox>

Talend is a Certified Technology Partner

The Hortonworks Certified Technology Program reviews and certifies technologies for architectural best practices, validated against a comprehensive suite of integration test cases, benchmarked for scale under varied workloads and comprehensively documented.



Hortonworks. We do Hadoop.

Hortonworks is a leading commercial vendor of Apache Hadoop, the open source platform for storing, managing and analyzing Big Data. Hortonworks Data Platform, our distribution of Apache Hadoop, provides an open and stable foundation for enterprises and a growing ecosystem to build and deploy Big Data solutions.

Hortonworks is the trusted source for information on Hadoop, and together with the Apache community, Hortonworks is making Hadoop an enterprise data platform. Hortonworks provides unmatched technical support, training and certification programs for enterprises, systems integrators and technology vendors.

For additional questions, contact:

- Talend
www.talend.com
(650) 539-3200
- Hortonworks
www.hortonworks.com
(855) 8-HORTON