



Talend

Industry

Data Management

Website

www.talend.com

Company Overview

Data-driven organizations use Talend's integration solutions to gain instant value from all their data. Natively supporting Hadoop and Spark technologies, Talend's open and unified solutions take the complexity out of any integration project and equip IT to be more responsive to the demands of the business, at a predictable cost.

Product Overview

Talend Big Data on Cloudera is a powerful and versatile data integration and data quality solution to extract, load, and process large and diverse data sets for better insights or process optimization..

Solution Highlights

- Easy and fast – Visually develop jobs that run 100% on Hadoop and Spark. A single solution for batch, streaming and real-time data integration and data quality.
- Native support – Unlike legacy integration solutions, Talend natively generates Hadoop and Spark code and requires zero-footprint on the Hadoop cluster.
- Connect your legacy IT landscape – Bridges and augments your legacy data assets through 900+ connectors and components for RDBMS, NoSQL, ERP/CRM, SaaS, and all types of flat files.

Optimizing your Enterprise Data Warehouse with Cloudera and Talend

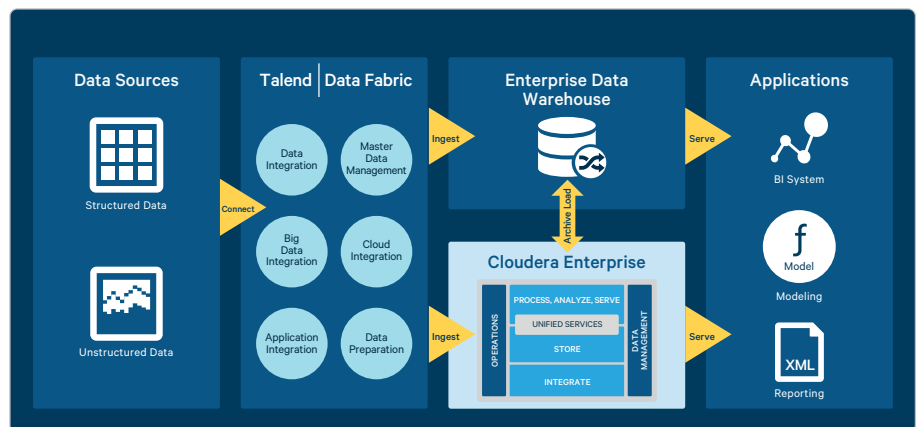
Enterprise data warehousing (EDW) has been a major IT pillar for many organizations over the past two decades. However, with the volume of data doubling every two years, and growing exponentially with the advent of sensor and smart devices, enterprise data warehouses are under distress. For example, load-processing windows are being maxed out, affecting service and threatening the delivery of critical business insights. Additionally, the need to store “just in case” data exacerbates this challenge by taking up expensive EDW capacity.

Organizations that use Hadoop to optimize their data warehouses can keep more data for longer periods of time, and can even mine and analyze the data for more advanced and impactful insights. They can also combine, enrich, and augment their structured data with weblogs, social media content, and other unstructured data to gain new analytical insights.

Now there is a better way to extend your EDW—thanks to the combination of Talend and Cloudera, which provides a new class of scalability, fast data processing, and management across vast amounts of structured and unstructured data. Furthermore, this combination allows data management projects to be completed faster, at a lower cost, and using existing skillsets.

Running Talend Big Data on Cloudera Enterprise enables you to create and execute Hadoop integration jobs, process and reconcile data, and implement data governance processes using an intuitive drag-and-drop interface. This robust, easy-to-use environment makes it possible for organizations to get up and running on Cloudera Enterprise faster, to scale to any level, and to support any data type (regardless of complexity), so they can focus on deriving insights instead of time-consuming data integration tasks.

Offload Your EDW to Hadoop Without Having to Code



Talend Real-Time Big Data combines components for Hadoop, Spark, MapReduce 2.0 (YARN), HBase, Hive, HCatalog, Oozie, Sqoop, and Pig into a unified environment so you can quickly load, extract, transform, and process large and diverse data sets from disparate systems.

Talend provides an easy-to-use graphical environment that allows developers to visually map data sources and targets without the need to learn and write code. Running 100% natively on Hadoop, Talend Big Data provides massive scalability. Once a big data connection is configured, the underlying code is automatically generated and can be deployed remotely as a zero-footprint job to process the data where it is on your cluster without mandating to move it to an external engine.

Cloudera Enterprise Benefits

Stores and Analyzes Any Type of Data

- Leverage the full power of your data to achieve pervasive analytics, increase business visibility, and reduce costs
- Bring diverse users and application workloads to a single, unified pool of data on common infrastructure; no data movement required

Enterprise Approach

- Compliance-ready perimeter security, authentication, granular authorization, and data protection through encryption and key management
- Enterprise-grade data auditing, data lineage, and data discovery

Industry-Leading Management and Support

- Best-in-class holistic interface that provides end-to-end system management and zero-downtime rolling upgrades
- Open platform ensures easy integration with existing systems
- Open source to achieve stability, continuous innovation, and portability

Talend Benefits

Fast time-to-value

- Talend dramatically lowers the integration and training efforts needed to bring Hadoop to the core of your IT landscape, accelerating the time-to-value

Improved data integrity

- Native Hadoop data quality helps you discover and cleanse data and reconcile it with your current processes, business entities, and applications

Predictable pricing

- Both Talend and Cloudera use an open core and subscription-based pricing model allowing to predictably scale your data management costs

Future-proof

- Based on an open source foundation, Talend provides an extensible solution that runs at the innovation pace of the Hadoop community and Cloudera

Talend provides data quality functions that take advantage of the massively parallel environment of Hadoop—allowing you to understand the completeness, accuracy, and integrity of data, as well as to remove duplicates. Hadoop data profiling allows you to collect information and statistics about big data to assess data quality, repurposing, and metadata. Additional functions include standardization, parsing, enrichment, matching, survivorship, and monitoring on-going data quality.

EDW Optimization Projects Simplified

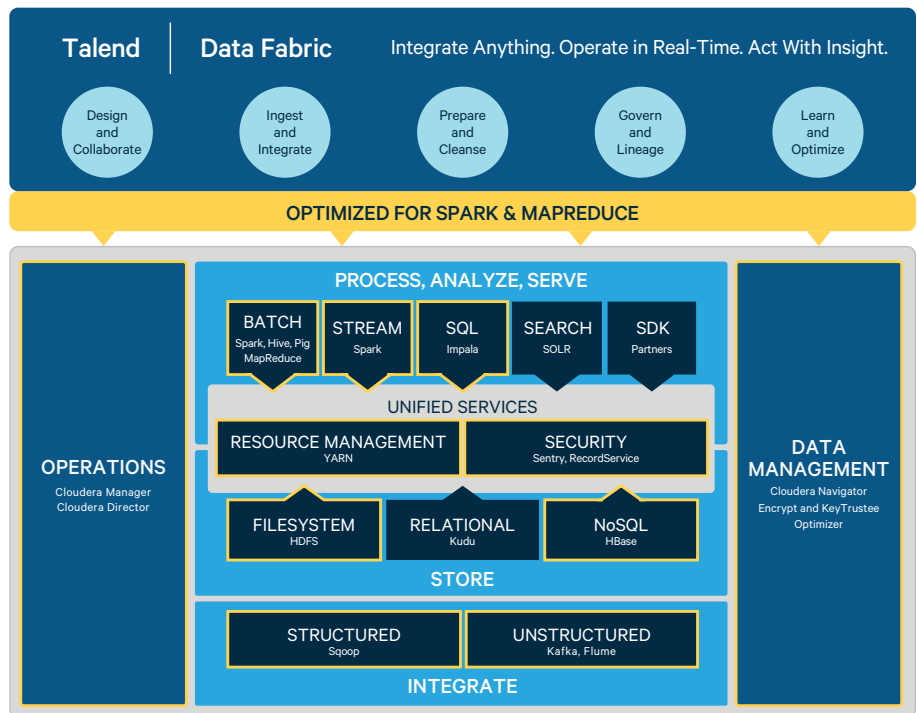
Together, Talend and Cloudera simplify the task of integrating diverse data structures, regardless of the volume or velocity. As the first enterprise data management solution that natively leverages the power and scalability of Cloudera Enterprise, the combined solution provides:

High Performance

Talend generates native MapReduce, Hive, Pig, and Spark code, provides native integration with Cloudera, leverages Spark Streaming in-memory processing, enables end-to end Internet of Things interoperability, and is the only data quality solution to run inside of Hadoop. This translates into faster performing applications.

Governed Data Management

Talend Real-Time Big Data is integrated with Cloudera Navigator. This lets users trace data lineage for MapReduce and Spark down to the level of the schema defined by the developer in a Talend data Job, which is crucial for both impact analysis and data lineage.



 Talend Integration

Low technology and financial risk

Your developers can benefit from a familiar, Eclipse-based environment using open standards. As Talend generates Spark, MapReduce/YARN, Pig, and Java, your ramp-up time is shortened.

Both Talend and Cloudera use a subscription-based pricing model, allowing you to start small and scale over time, while easily managing your up-front financial commitments.



Extensive connectivity

Talend provides connectivity to Cloudera from existing and legacy systems with more than 900 connectors and components for RDBMS, NoSQL, ERP/CRM, SaaS, and all types of flat files.

With Cloudera and Talend, you can transform and drive down the costs of your existing IT assets—such as ETL and batch process offloading, or enterprise data warehousing optimization.

About Talend

Talend's integration solutions allow data-driven organizations to gain instant value from all their data. Through native support of modern big data platforms, Talend takes the complexity out of integration efforts and equips IT departments to be more responsive to the demands of the business, at a predictable cost. Based on open source technologies, Talend's scalable, future-proof solutions address all existing and emerging integration requirements. Talend is privately-held and headquartered in Redwood City, CA. For more information, please visit www.talend.com and follow us on Twitter: @Talend.

About Cloudera

Cloudera delivers the modern data management and analytics platform built on Apache Hadoop and the latest open source technologies. The world's leading organizations trust Cloudera to help solve their most challenging business problems with Cloudera Enterprise—the fastest, easiest, and most secure data platform available for the modern world. Our customers efficiently capture, store, process, and analyze vast amounts of data—empowering them to use advanced analytics to drive business decisions quickly, flexibly, and at lower cost than has been possible before. To ensure our customers are successful, we offer comprehensive support, training, and professional services. Learn more at cloudera.com.

cloudera.com

1-888-789-1488 or 1-650-362-0488

Cloudera, Inc. 1001 Page Mill Road, Palo Alto, CA 94304, USA

© 2016 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries. All other trademarks are the property of their respective companies. Information is subject to change without notice.

Cloudera_Solution-Brief_Talend_Optimizing_Ent_Data_Warehouse_103