What it Does

Integration at Cluster Scale

Talend redefines the development skills needed for big data and facilitates the organization and orchestration required by these projects so that you can focus on the key question: “What use should we make of data, big and small, and how am I going to be the leader in using data to help my business?” Talend’s big data product combines big data components for MapReduce 2.0 (YARN), Hadoop, HBase, Hive, HCatalog, Oozie, Sqoop and Pig into a unified open source environment so you can quickly load, extract, transform and process large and diverse data sets from disparate systems.

How it Works

Big Data Without The Need To Write / Maintain Code

Ready to Use Big Data Connectors

Talend provides an easy-to-use graphical environment that allows developers to visually map big data sources and targets without the need to learn and write complicated 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 job that runs natively on your big data cluster - HDFS, Pig, HCatalog, HBase, Sqoop or Hive.

Big Data Distribution and Big Data Appliance Support

Talend’s big data components have been tested and certified to work with leading big data Hadoop distributions, including Amazon EMR, Cloudera, IBM PureData, Hortonworks, MapR, Pivotal Greenplum, Pivotal HD, and SAP HANA. Talend provides out-of-the-box support for big data platforms from the leading appliance vendors including Greenplum/Pivotal, Netezza, Teradata, and Vertica.

Open Source

Using the Apache software license means developers can use the Studio without restrictions. As Talend’s big data products rely on standard Hadoop APIs, users can easily migrate their data integration jobs between different Hadoop distributions without any concerns about underlying platform dependencies. Support for Apache Oozie is provided out-of-the-box, allowing operators to schedule their data jobs through open source software.

Pull Source Data from Anywhere Including NoSQL

With 450+ connectors, Talend integrates almost any data source so you can transform and integrate data in real-time or batch. Pre-built connectors for HBase, MongoDB, Cassandra, CouchDB, Couchbase, Neo4J and Riak speed development without requiring specific NoSQL knowledge. Talend big data components can be configured to bulk upload data to Hadoop or other big data appliance, either as a manual process, or an automatic schedule for incremental data updates.
**Big Data Products**

Talend Open Studio for Big Data is an Apache licensed, open source development tool. Talend Enterprise Big Data adds teamwork and management features. Talend Platform for Big Data adds data quality, clustering features with extended support services.

<table>
<thead>
<tr>
<th>Features</th>
<th>Talend Open Studio for Big Data</th>
<th>Talend Enterprise Big Data</th>
<th>Talend Platform for Big Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Designer</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>Components for HDFS, HBase, HCatalog, Hive, Pig, Sqoop</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>Hadoop Job Scheduler</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>NoSQL Support</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>Versioning and Centralized Metadata Management</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>Shared Repository</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>Reporting and Dashboards</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>Big Data Profiling, Parsing and Matching</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>Indemnification/Warranty and Talend Support</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>License</td>
<td>Apache</td>
<td>Subscription</td>
<td>Subscription</td>
</tr>
</tbody>
</table>

**Specifications**

**Big Data**

Talend Big Data supports the following third party components, products and operating systems. For detailed information, please reference the product installation document and release notes.

**SUPPORTED BIG DATA HADOOP DISTRIBUTIONS AND NOSQL**
- Amazon Redshift
- Amazon EMR
- Apache Hadoop (HBase, HDFS, Hive)
- Cassandra
- Couchbase
- CouchDB
- Cloudera
- Google BigQuery
- Greenplum/Pivotal HD
- Hortonworks Data Platform
- IBM PureData System for Hadoop
- MapR
- MongoDB
- Neo4J

**SUPPORTED DATABASE CONNECTIVITY**
- Riak
- SAP HANA
- Teradata
- Vertica
- Windows Azure Blob Storage

**BIG DATA FILE FORMAT SUPPORT**
- SEQ, JSON, RC, ORC and AVRO

**SUPPORTED DATABASE AND STORAGE CONNECTIVITY**
- Amazon RDS, Amazon Redshift, Amazon S3, AS400, DB2, Derby DB, Exasol, eXist-db, Firebird, Google Storage, Greenplum, H2, HIVE, HSQLDB, Informix, Ingres, InterBase, JavaDB, JDBC, MaxDB, Microsoft OLE-DB, Microsoft SQL Server, MySQL, Netezza, Oracle, ParAccel, PostgreSQL, PostgresPlus, SAS, SQLite, Sybase, Teradata, VectorWise, Vertica

**SUPPORTED SAAS AND 3RD PARTY APPLICATIONS**
- Alfresco
- Centric CRM
- Marketo
- Microsoft CRM and AX
- Open Bravo
- SAGE X3
- Salesforce.com
- SAP
- SugarCRM
- Vtiger CRM

**SUPPORTED OPERATING SYSTEMS**
- CentOS Linux
- OS X
- Redhat Enterprise Linux
- Solaris
- SUSE Linux
- Ubuntu Linux
- Microsoft Windows

---

**Get White Paper**

Get White Paper - How Big is Big Data Adoption?


**Watch Webinar**

Watch Webinar - Discover Big Data Tools