ORACLE DATA SHEET

Oracle Data Integrator For Big Data

ORACLE
FUSION MIDDLEWARE
DATA INTEGRATOR

Oracle Data Integrator For Big Data offers customers enterprise scale big data Integration. Oracle Data Integrator For Big Data extends big data heterogeneity to include multiple big data standards. Through its decoupling of logical design and physical implementation, Oracle Data Integrator lets customers choose between multiple underlying big data platforms that best suits the customer requirement. Customers can now future proof their Hadoop investment and increase Hadoop platform inter-operability. Oracle Data Integrator decreases big data projects' time to value by offering out of the box code templates which increases developer productivity, streamlines the development process and improves performance.

KEY BUSINESS BENEFITS

- Extreme scale and performance to big data integration
- Expanded big data heterogeneity
- Ability to seamlessly switch between underlying big data platforms at will
- Increased developer productivity and streamlined development process for big data integration projects
- Quick time to value for big data projects

KEY FEATURES

- Native code language generation and execution for Pig Latin and Apache Spark
- Native process scheduling choice through Oozie or ODI Agent
- Out of the Box support for Loads and Integration to Big Data Sources and Targets

Oracle Data Integrator for Big Data

Oracle Data Integrator for Big Data brings advanced data integration capabilities to customers who are looking to implement a seamless and responsive Big Data Management platform.

Oracle Data Integrator is a transparent and heterogeneous Big Data Integration technology based on an open and lightweight ELT architecture. It runs a diverse set of workloads, including Spark, Spark Streaming and Pig transformations, to enable customers solve their most complex and time sensitive data transformation and data movement challenges. It is a core component of Oracle Data Integration solutions, integrating seamlessly with the rest of Oracle's Data Integration and Business Application solutions.

Oracle Data Integrator for Big Data provides the following benefits to customers.

- It brings expanded connectivity to various Big Data source such as Apache Kafka or Cassandra
- It decreases time to value for Big Data projects
- It provides a future proof Big Data Integration technology investment
- It streamlines and shortens the Big Data development and implementation process

Future Proof Big Data Integration through Seamless Portability between Big Data Platforms

Oracle Data Integrator allows customers to define mappings through a logical design

ORACLE

Spark, and Pig as the generated transformation language. This allows users to pick the best implementation method based on the environment and use case; it is even possible to choose different implementations simultaneously using multiple physical designs making development for big data flexible and future-proof.

Efficient Big Data Processing and Scheduling Through Native Execution

Oracle Data Integrator provides efficient Big Data processing by offloading its workload into the Big Data platform. Since Oracle Data Integrator's ELT architecture does not require middleware, customers do not have to invest in additional processing hardware to handle increasing data loads. Oracle Data Integrator's ability to natively generate code translates into greater efficiency because it eliminates the need for external agents to be installed on the Hadoop cluster.

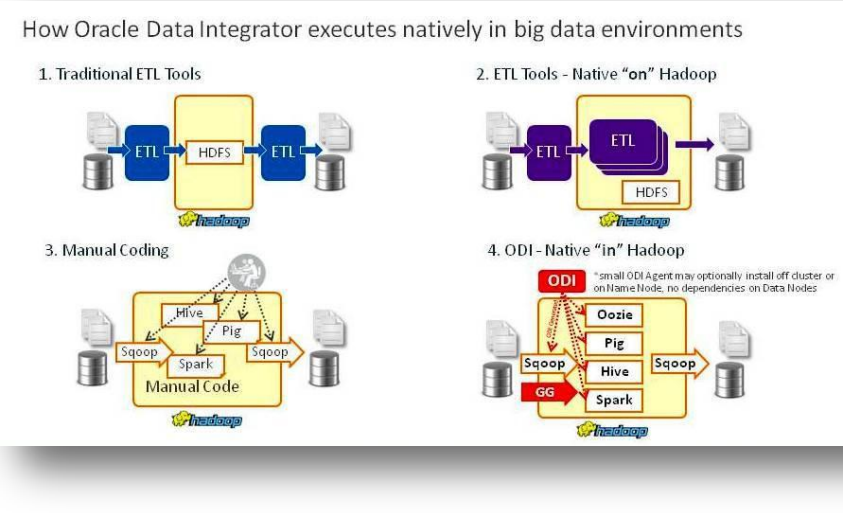

FIG1: ORACLE DATA INTEGRATOR PROCESSES BIG DATA NATIVELY TAKING ADVANTAGE OF UNDERLYING CLUSTER CAPABILITIES UNLIKE TRADITIONAL ETL TECHNOLOGIES THAT RUN ON TOP OF, OR OUTSIDE THE BIG DATA CLUSTER.

Currently ODI supports

Generation of Pig Latin transformations: Users can choose Pig Latin as their transformation language and execution engine for ODI mappings. Apache Pig is a platform for analyzing large data sets in Hadoop and uses the high-level language Pig Latin for expressing data analysis programs. Any Pig transformations can be executed either in local or map-reduce mode. Custom Pig code can be added through user-defined functions or the table function component.

Generation of Spark and Spark Streaming transformations: ODI mappings can also generate PySpark, which exposes the Spark programming model in the Python language. Apache Spark is a transformation engine for large-scale data processing. It provides fast in-memory processing of large data sets. Custom PySpark code can be added through user- defined functions or the table function component.

ARCHITECTURAL DIFFERENTIATORS

- Leverages cluster compute and query capabilities
- Eliminates middleware and additional processing investment
- Generation of native big data code

RELATED PRODUCTS

Oracle GoldenGate for Big Data

Oracle Data Integrator Enterprise Edition

Oracle GoldenGate

Oracle Enterprise Metadata Management

Orchestration of ODI Jobs using Oozie: Users have a choice between using the traditional ODI Agent or Apache Oozie as orchestration engines for jobs such as mappings, packages, scenarios, or procedures. Apache Oozie allows fully native execution on Hadoop infrastructures without installing an ODI agent for orchestration. Users can utilize Oozie tooling to schedule, manage, and monitor ODI jobs. ODI uses Oozie's native actions to execute Hadoop processes and conditional branching logic.

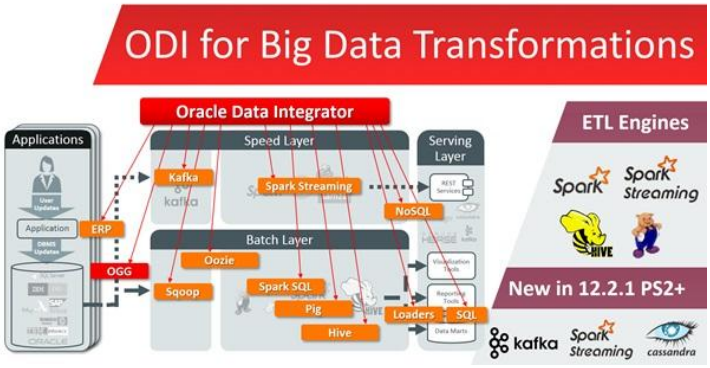

Fig 2: Oracle Data Integrator For Big Data Transformations

Shortened Time to Value On Big Data Projects Through Out Of The Box Big Data Knowledge Modules

Oracle Data Integrator provides out of the box transparent and customizable modules, called Knowledge Modules that shortens and streamlines development cycles, increases standardization and repeatability for Big Data projects.

Oracle Data Integrator includes the WebLogic Hive JDBC driver that provides a number of advantages compared to the Apache Hive driver, such as full JDBC compliance and improved performance. Knowledge Modules whose main purpose is to load from a source are now provided as Loading Knowledge Modules, enabling them to be combined in a single mapping with other Loading Knowledge Modules. A new class of "direct load" Loading Knowledge Modules also allows loading of targets without intermediate staging.

Oracle Data Integrator Brings Big Data To The Center Of Technology And Business Decisions

Oracle Data Integrator For Big Data brings speed, ease of use and trust to how enterprises capitalize on data. Big data management is essential to any organization that wants to make serious headway in their decision making culture. Data is being generated in all forms, from various traditional and nontraditional sources to provide competitive advantages. Oracle Data Integrator for Big Data addresses this growing need in the market by providing a future proof, powerful platform to build your enterprise around its Data Management framework.

CONTACT US

For more information about Oracle Data Integrator For Big Data, visit oracle.com/dataintegration or call +1.800.ORACLE1 to speak to an Oracle representative.

CONNECT WITH US

blogs.oracle.com/oracle

facebook.com/oracle

twitter.com/oracle

oracle.com

Hardware and Software, Engineered to Work Together