

Oracle GoldenGate for Big Data



ORACLE GOLDENGATE FOR BIG DATA

KEY FEATURES

- Non-invasive, real-time transactional data streaming
- Secured, reliable and fault-tolerant data delivery
- Simple to install, configure, and maintain
- Streams real-time changed data into big data systems, real-time messaging systems, NoSQL databases, object storages, cloud data warehouses and on-premises MPP appliances
- Easily extensible and flexible to stream changed data to other big data targets and message queues like JMS, Apache Kafka, Amazon Kinesis etc
- Process and analyze streaming data using an interactive user interface on a scalable and highly available clustered Spark environment

Oracle GoldenGate for Big Data 12c streams data into big data systems in real time, without impacting the performance of source systems. It streamlines real-time data delivery into the most popular big data solutions, including Apache Hadoop, Apache HBase, Apache Hive, Apache Flume, Apache Kafka, NoSQL Databases, Elasticsearch, JDBC, Amazon Web Services, Cloud Databases, and Data Warehouses to facilitate improved insight and timely action.

Oracle GoldenGate for Big Data Product Overview

Big data analytic solutions play a pivotal role in improving business insight and enhancing consumer experience. It is increasingly evident that organizations can realize the full potential value of their data assets by combining the structured transactional data with semi-structured and unstructured data. Businesses also notice that to be agile and react to situations in real time, access to transactional data with low latency is essential.

Low-latency transactional data brings additional value especially for dynamically changing operations that day-old data, structured or unstructured, cannot deliver. Oracle GoldenGate for Big Data offers a comprehensive platform that streams transactional data into big data solutions in real time, without degrading the performance of the source production systems. It lays the foundation for more efficient operations and an improved customer experience. With its easy-to-use solution, Oracle GoldenGate for Big Data enables IT organizations to quickly integrate into big data systems without extensive training and management resources.

Oracle GoldenGate is a time tested and proven product for real-time data integration and heterogeneous database replication. It supports organizations for real-time data integration for analytical systems, data warehouses, data marts, zero downtime migration and consolidation into cloud architectures, offloading reporting from transactional systems (on premises and cloud) for maximum performance, and active-active database replication for data distribution and continuous availability. In this new offering, Oracle GoldenGate for Big Data provides optimized and high performance delivery to Flume, HDFS, Hive and Kafka to support customers with their real-time big data analytics needs for any organization. Oracle GoldenGate for Big Data provides optimized and high performance capture and delivery to NoSQL databases such as HBase, MongoDB, and Cassandra to support customers with their real-time new data analytics and machine learning initiatives. Oracle GoldenGate for Big Data includes Oracle GoldenGate for Java and JDBC to enable you to easily integrate to additional big data systems, such as Amazon Web Services, Oracle

KEY BENEFITS

- Improve IT productivity in integrating with big data systems
- Use real-time data in big data analytics for more timely and reliable insight
- Improve operations and customer experience with enhanced business insight
- Minimize overhead on source systems to maintain high performance

ORACLE GOLDENGATE FOR BIG DATA

Oracle GoldenGate for Big Data helps organizations raise the quality and timeliness of business insights by streaming real-time data to big data environments. It lays the foundation for more efficient operations and an improved experience.

RELATED PRODUCTS

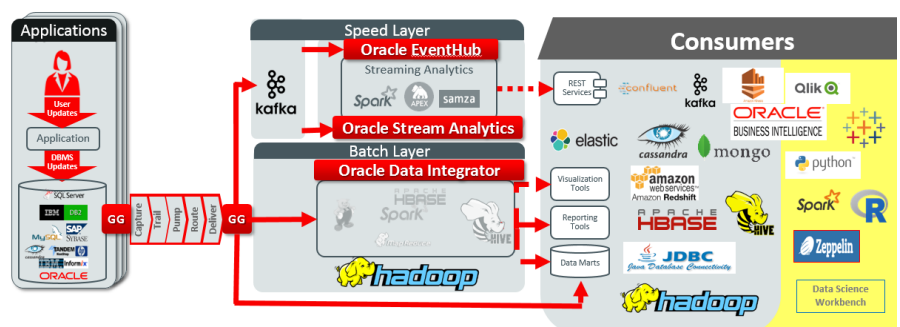
The following products support and complement Oracle GoldenGate for Big Data

- Oracle GoldenGate
- Oracle GoldenGate Studio
- Oracle GoldenGate Veridata
- Oracle Management Pack for Oracle GoldenGate
- Oracle Stream Analytics
- Oracle GoldenGate Cloud Service
- Oracle Data Integration Platform Cloud

NoSQL, Elasticsearch, SAP HANA, IBM PureData System for Analytics, Greenplum, and many others.

The Oracle GoldenGate Big Data real-time data streaming platform also allows you to keep your big data reservoirs, or big data lakes, up to date with your production systems. With this solution, you can perform ad-hoc discovery, organization, and enrichment of low-latency data before it traverses to more refined sets of analytics tools.

Oracle GoldenGate for Big Data is a key component of Oracle's big data integration offering along with Oracle Data Integrator and Oracle Stream Analytics. Oracle combines big data integration with Oracle's Big Data Architecture to give customers the ability to capture the performance and real time advantages of Hadoop, NoSQL and cloud technologies.



Oracle GoldenGate for Big Data supports real-time data capture and streaming to major big data targets. You can combine Oracle GoldenGate for Big Data with other Oracle products to have an end-to-end platform for big data machine learning and analytics needs.

Oracle GoldenGate Handler for HDFS and Hive

Hadoop Distributed File System (HDFS) is a Java-based file system that provides scalable and reliable data storage that is designed to span large clusters of commodity servers. HDFS, MapReduce, and YARN form the core of Apache Hadoop and also commercial vendors such as Microsoft Azure HDInsight, Cloudera Platform, Hortonworks Data Platform, and MapR Platform. Using Oracle GoldenGate for Big Data to stream changed data into HDFS enables the downstream applications to further process data natively on Hadoop for better business insights and for lower total cost of ownership. Hive is a data warehousing infrastructure provided over Hadoop for doing summarization and analysis. Oracle GoldenGate can provide real-time data to the Hive Data Store to enable scalable, agile, and cost effective real-time data warehousing on Hadoop.

Oracle GoldenGate Handler for HBase

Apache HBase is a non-relational (NoSQL) database that runs on top of the Hadoop Distributed File System (HDFS). It is columnar and provides fault-tolerant storage and quick access to large quantities of sparse data. It also adds transactional capabilities to Hadoop, allowing users to conduct updates, inserts and deletes. With Oracle GoldenGate for Big Data, inserts, updates, and deletes can be applied to HBase in real-time for deriving value from large data sets.

Oracle GoldenGate Handler for Apache Flume

Apache Flume is a distributed service for efficiently transporting high-volume, streaming data flows. It performs real-time analytics on data flows to support timely insight and action in today's fast-paced digital business environment. By using Oracle GoldenGate for Big Data to provide data into Flume in real-time, you can build real-time analytical applications.

Oracle GoldenGate Handler for Kafka

Apache Kafka is a high throughput publish-subscribe messaging system. It is distributed, partitioned, replicated, fault tolerant, and also has stronger ordering of message sequence when compared to the conventional messaging systems. Kafka is also a general purpose enterprise messaging system beyond Hadoop and GoldenGate can also integrate with Confluent Platform using Kafka Pub/Sub API, Kafka Connect API and Kafka REST Proxy API. By streaming real-time data into Kafka using Oracle GoldenGate for Big Data, you can build fault-tolerant, highly reliable, and extensible real-time analytical applications.

Oracle GoldenGate Handler for MongoDB

MongoDB is an open-source document database that provides high performance, high availability, and automatic scaling. It can provide fast read and publishing of data to consuming users to support timely insight and action in today's fast-paced digital business environment. By using Oracle GoldenGate for Big Data to provide data into MongoDB in real-time, organizations can build scalable and highly available real-time analytical applications.

Oracle GoldenGate Handler for Cassandra

Apache Cassandra is a NoSQL Database Management System designed to store large amounts of data scaling horizontal across multiple machines. It is one of the databases that can write data really fast in today's fast-paced digital big business environment where data is generated by devices and high speed computers. By using Oracle GoldenGate for Big Data to capture in real-time and provide data into Cassandra in real-time, organizations can build fast, scalable, and dynamic real-time based analytical systems.

Oracle GoldenGate Handler for JDBC

JDBC stands for Java Database Connectivity. JDBC is one of the most generic technology interfaces that is used by most of the spreadsheets, databases, and application providers for on-premises and on the cloud. By using Oracle GoldenGate for Big Data to provide data into JDBC interface in real-time, organizations will be able to replicate transaction data to a large set of long-tail target databases like Amazon Redshift, IBM Netezza, and Greenplum for building real-time analytical reporting systems.

Oracle GoldenGate Handler for Amazon Kinesis

Amazon Kinesis is fully managed Amazon Web Services application that helps to ingest, buffer, and process streaming data in real-time, to derive insights in seconds or minutes. Amazon Kinesis helps you to process and analyze data as it arrives and respond instantly instead of having to wait until all your data is collected before the processing can

begin. Kinesis streams can be used to stream data to other Amazon Cloud applications such as Amazon S3 and Amazon Redshift. Oracle GoldenGate for Big Data can stream data from any of the transactional systems or message bus to Amazon Kinesis in real-time to derive valuable insights in few seconds.

Oracle GoldenGate Handler for Elasticsearch

Elasticsearch is a highly scalable, open-source full-text search and analytics engine. Elasticsearch allows you to store, search, and analyze large volumes of data quickly and in near-real time. It is generally used as the underlying engine or technology that drives applications with complex search features. Oracle GoldenGate for Big Data feeds data into Elasticsearch in real-time so that organizations can build scalable and highly available real-time complex search capabilities into their analytical applications.

Oracle GoldenGate Handler for File Writer

Oracle GoldenGate for Flat File can write micro batches to disk to be consumed by tools that expect batch file input such as ETL, proprietary or legacy applications. The data is formatted to the specifications of the target application such as delimiter separated values, length delimited values, ORC or Parquet formats into HDFS or AWS S3. Near real-time feeds of events to these systems are accomplished by decreasing the time window for batch file rollover to minutes or even seconds.

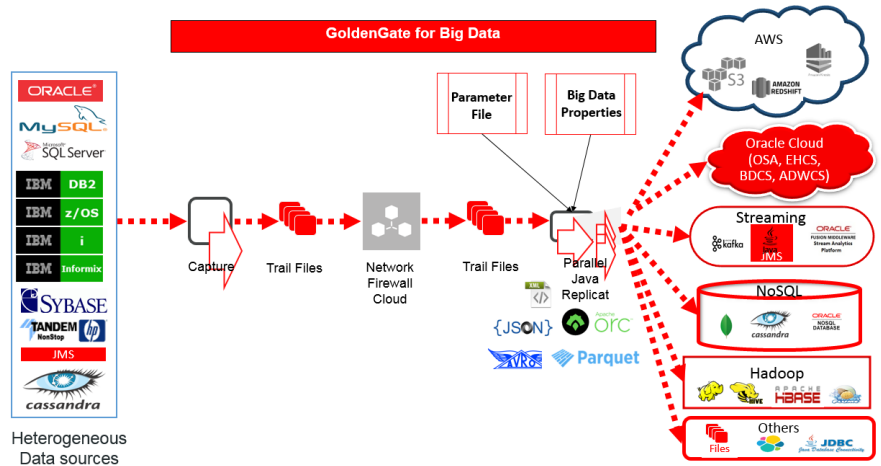
Oracle Stream Analytics

Oracle Stream Analytics allows users to process and analyze large scale real-time information by using sophisticated correlation patterns, enrichment, and machine learning. It offers real-time actionable business insight on streaming data and automates action to drive today's agile businesses. Oracle Stream Analytics is included in Oracle GoldenGate for Big Data to process events originating from the Oracle GoldenGate Handler for Kafka.

Key Technical Value Drivers

Oracle GoldenGate offers a modular architecture designed for extreme performance, fault tolerance, and flexibility. Oracle GoldenGate for Big Data is built on the same architecture to enable extensible solutions for customers' big data environments.

Oracle GoldenGate captures data from heterogeneous source systems including Java-based messaging systems, non-invasively and with negligible overhead. It stores database transactions in Trail Files and pumps them into the parallel Java Replicat framework. Oracle GoldenGate for Big Data includes handlers for different big data technologies such as HDFS, Hive, HBase, Flume, Kafka, NoSQL, JDBC, Amazon Kinesis, Amazon S3 and more. It also includes pluggable formatting technology that can be used to transform data to standard formats such as XML, JSON, Avro, ORC, Parquet and Delimited text. It is built using extensible Java architecture, which enables you to easily deliver to additional big data systems and support specific use cases to meet your business demands.



With Oracle GoldenGate real-time streaming platform, you can capture from the source system once and deliver all, or a portion of, the changed data to multiple targets on-premises, cloud and hybrid, including databases, messaging systems, and now big data environments.

With constantly expanding heterogeneous support, Oracle GoldenGate enables you to standardize on a single real-time data streaming platform for their enterprise. Oracle GoldenGate supports log-based capture from, and delivery to Oracle, DB2 for z/OS, i Series, and LUW (Linux, Unix, Windows), SQL Server, MySQL, Informix, Sybase ASE, SQL/MX, SQL/MP, Enscribe, flat files, and more. Oracle GoldenGate’s delivery capabilities also include Oracle Stream Analytics, Oracle TimesTen In Memory Database, Teradata, and PostgreSQL, in addition to Hadoop-based and non-Hadoop based big data systems.

Summary

Oracle GoldenGate for Big Data offers a high-performance, fault-tolerant, easy-to-use, and flexible real-time data streaming platform for big data environments. It easily extends your real-time data integration architectures to big data systems without impacting the performance of the source systems and enables timely business insight for better decision making.



CONTACT US

For more information about Oracle Management Pack for Oracle GoldenGate, visit oracle.com or call +1.800.ORACLE1 to speak to an Oracle representative.

CONNECT WITH US



blogs.oracle.com/dataintegration



facebook.com/oracledataintegration



twitter.com/OracleDI



oracle.com/goto/goldengate

Hardware and Software, Engineered to Work Together

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group. 0518

