SharePlex™ Connector for Hadoop®

Real-time and near real-time data integration from Oracle® to Hadoop® databases, cost-effectively empowering analytics

You need data analytics for a business intelligence advantage and essential decision-making insight. The Hadoop® framework gives you that, but integrating Oracle® data can be time-consuming, providing only snapshots that quickly become out of date. But now there's an affordable and time-saving solution for improved analytics and far more immediate data replication.

SharePlex is a highly advanced, log-based Oracle database replication product used by customers worldwide for a range of needs, including real-time data access, data archiving, high availability, operational reporting, migrations, regulatory compliance, and integration.

SharePlex Connector for Hadoop adds to those capabilities by enabling you to easily load and continuously replicate changes from an Oracle database to a Hadoop cluster. This gives you all the benefits of maintaining a real-time or near real-time copy of source tables, so your organization can efficiently and cost-effectively perform business-boosting big data analytics.

Best of all, it's an elegant solution that ensures business continuity while meeting your database operational goals. It provides a real-time copy in a Hadoop environment with no impact on the availability of your Oracle online transaction processing (OLTP) applications or the performance of your systems.

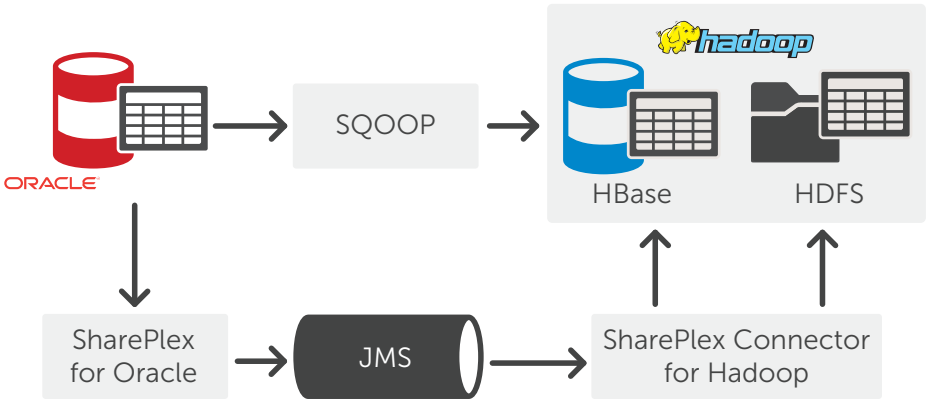


Figure 1. SharePlex Connector for Hadoop delivers real and near real-time replication of Oracle data to Hadoop HBase and HDFS environments.

Benefits:

- Improve analytics with fast and easy access to the most up-to-date information.
- Simplify real-time data integration with an automated solution that includes an easy-to-use GUI dashboard.
- Satisfy the growing demands of your users without impacting performance on production systems.
- Save time with an affordable solution that includes all the tools you need for effective and efficient Oracle-to-Hadoop data integration.
- Works with proven, enterprise Hadoop distributions: Apache®, IBM® InfoSphere® BigInsights®, Hortonworks® and is Cloudera® certified for CDH5.

System requirements

Software
SharePlex™ 8.5
Apache® ActiveMQ® 5.6
Java Development Kit® (64-Bit) 1.6 or later

Hadoop® environments
Apache 1.2.1 (64-Bit)
IBM® InfoSphere® BigInsights® (IBM BI 2.1.2)
Cloudera CDH 4.2.0 (64-Bit), 5.0.0 (64-Bit), 5.1.0 (64-Bit)
Hortonworks® (HDP 1.3)
Optional: Install and configure HBase® to replicate tables to HBase and Hive® over HBase.
Optional: Install and configure Hive to replicate tables to Hive (Hive over HDFS or Hive over HBase).

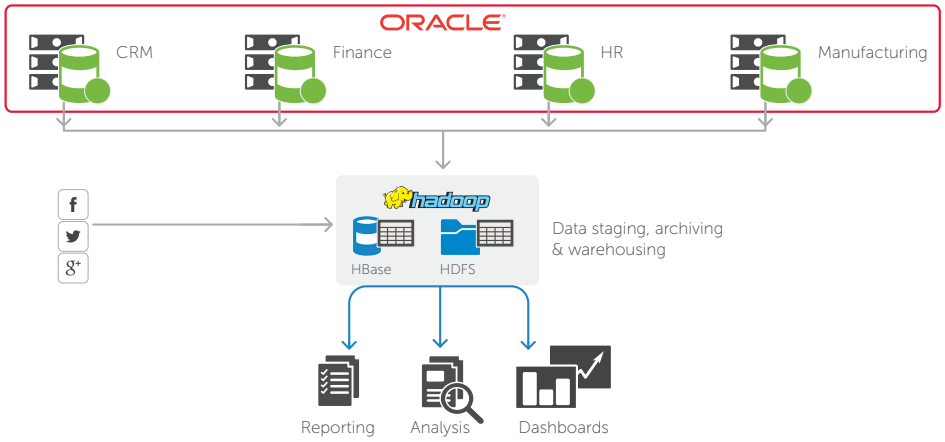


Figure 2. Gain greater insight by integrating Oracle data with Hadoop environments for better business intelligence, predictive analytics, data warehousing, data staging and archiving.

Features that deliver valuable business insight

- **Oracle data replication to Hadoop**—Replicates multiple copies of data to Hadoop Hive®, HDFS, and HBase® environments locally, remotely, or in-the-cloud for better business intelligence, predictive analytics, data warehousing, data staging, and archiving.
- **Hadoop partitioning**—Enhances analytic-processing performance with custom and range partitioning for your specific Hadoop environment.
- **Near real-time replication to Hadoop HDFS**—Writes data to HDFS files in text, sequence or Avro format. Rebuilds tables periodically by merging all the changes captured in a batch mode, avoiding time-consuming manual refreshes.
- **Real-time replication to Hadoop HBase**—Changes can be posted immediately with all the columns on the source table replicated to a single column family for fast data analytics.
- **Change data capture**—Enables audit capability with change data capture of insert, update and delete activity, optionally including pertinent metadata like SCN, ROWID, timestamp, user ID and more.

About Dell Software

Dell Software helps customers unlock greater potential through the power of technology—delivering scalable, affordable and simple-to-use solutions that simplify IT and mitigate risk. This software, when combined with Dell hardware and services, drives unmatched efficiency and productivity to accelerate business results. www.dellsoftware.com.

Dell Software
5 Polaris Way, Aliso Viejo, CA 92656 | www.dell.com
If you are located outside North America, you can find local office information on our Web site.

© 2014 Dell, Inc. ALL RIGHTS RESERVED. Dell, Dell Software, the Dell Software logo and products—as identified in this document—are registered trademarks of Dell, Inc. in the U.S.A. and/or other countries. All other trademarks and registered trademarks are property of their respective owners.
Datasheet-SharePlexConn4Hadoop-US-VG-25042