

# Real-Time Transactional Data Movement and Replication for Big Data Analytics

## Table of Contents

- 3 For Big Data Analytics, Real-Time Data Movement and Updates Are Critical
- 6 Core Functionality in SAP Replication Server
- 8 High-Performance, Real-Time Data Replication for Big Data Use Cases





Information and data-driven insight is what powers business today. But to [get the most from your enterprise data](#), you need a way to bring transactional, streaming, social media, and other data together – regardless of its format and whether it's structured or unstructured – and be able to analyze it. The challenge is moving, replicating, and centralizing a wide variety of data efficiently, cost-effectively, and quickly enough to meet business demands for active insight. SAP® Replication Server® can help.



SAP Replication Server enables the continuous movement of mission-critical application data.



# For Big Data Analytics, Real-Time Data Movement and Updates Are Critical

Today, most enterprise IT environments support a complex, ever-growing data architecture that often includes incompatible systems and data schemas. Enterprise data used to be generated primarily by internal data sources such as transactional systems, data logs, and e-mails.

But now, data comes from a variety of external data sources such as the Internet of Things, social media, Web sites, instant messaging systems, and third-party sources. And much of this data is unstructured, such as audio files, photos, and videos that don't fit neatly into traditional databases or play nicely with common data management and movements tools. These vast amounts of complex data are commonly referred to as Big Data.

The typical Big Data solution consolidates volumes of data into one central location – for example, a Hadoop data store – so people across your organization can easily access data and generate insights from it. You can even make certain data sets available to parts of your extended enterprise for self-service reporting or other purposes.

And of course, as all of this data is consolidated, you need to ensure that the data quality is high, even as it is transformed into different formats.

Whatever you have on the source side must be intact at the target. And there should be no loss of data integrity at any point.

But in most cases, moving and synchronizing data from one (or many) primary data sources to one (or many) data targets remains a challenge. Think about your own organization. Chances are, you need a way to move data from many sources at once – continuously, bidirectionally, and in real time to your analytics system, so that decision makers can generate insights based on the very latest data. You want your data to stay secure at all times, and governance must be enforced to prevent compliance and risk issues – particularly if you work in closely regulated industries such as finance and healthcare.

To meet all these requirements, you need a solution that supports:

- Secure data distribution
- Data consolidation
- Data assurance
- Log-based, nonintrusive operation
- Real-time, high-volume processing

SAP Replication Server can do all this and more.



SAP Replication Server is engineered to supply your target systems with the most current information from diverse data sources across the enterprise.



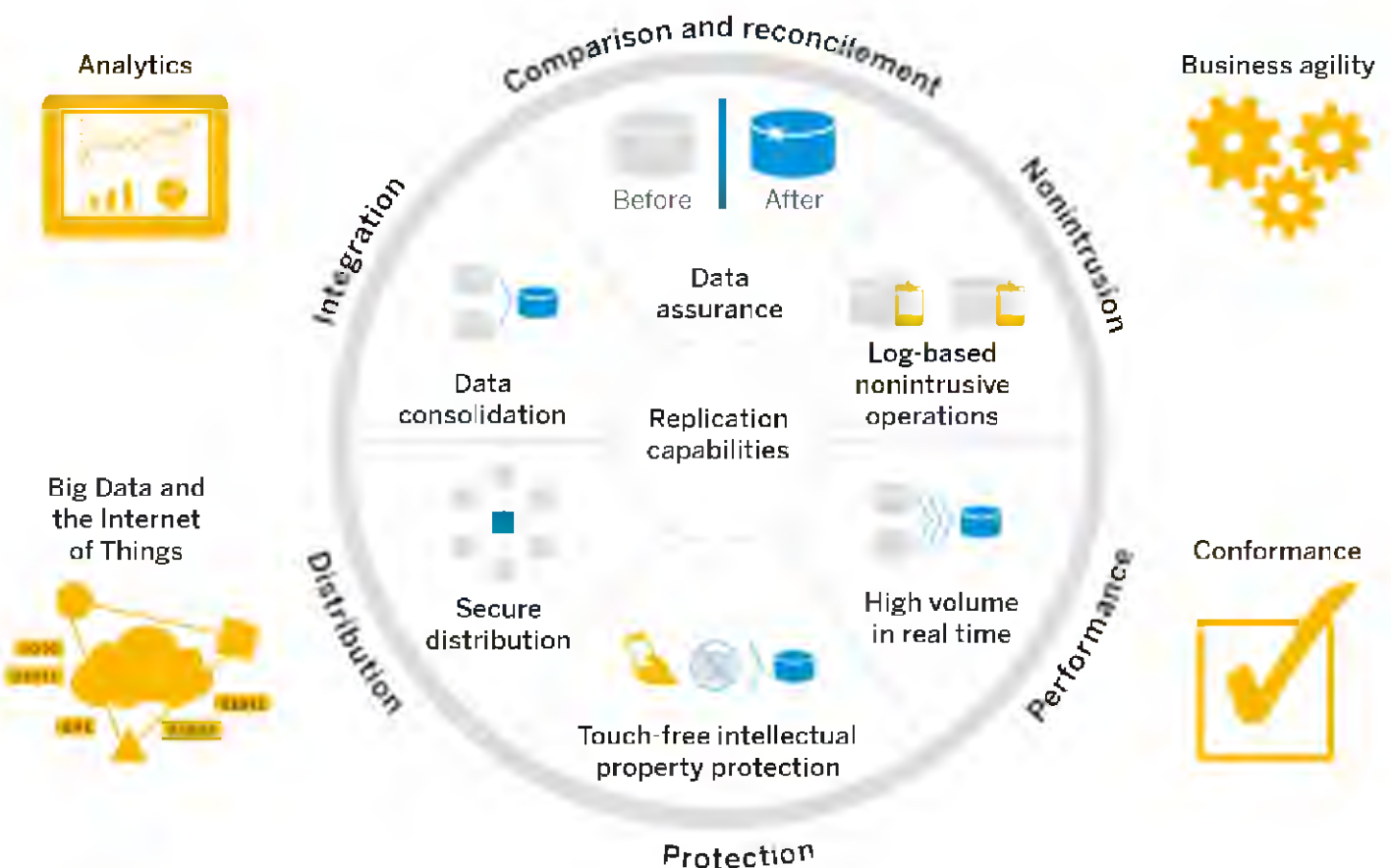
## SAP REPLICATION SERVER: OPTIMIZED FOR BIG DATA MOVES

SAP Replication Server was designed to meet today's business-critical data requirements. It supports your enterprise data management environment by replicating and synchronizing transactions originating in SAP HANA®, SAP Adaptive Server® Enterprise (SAP ASE), Oracle, IBM DB2 Universal Database (UDB), and Microsoft SQL

Server databases (see Figure 1). It provides high-performance, secure, reliable, and consistent delivery of data across the enterprise, addressing:

- Data movement, distribution, synchronization, and migration of both structured and unstructured data
- Data consolidation for real-time analytics and reporting

Figure 1: How SAP® Replication Server® Benefits Governance, Risk, and Compliance





With SAP Replication Server delivering consolidated, real-time data to analytics systems, people across every part of your business can make decisions based on the most current data. In addition, you can:

- Distribute and synchronize data across geographies to multiple systems
- Comply with industry regulations requiring strict transaction tracking
- Integrate acquired data assets for 360-degree access across the enterprise
- Migrate data to new databases, hardware, or OS platforms

#### An Industry-Proven Solution

Originally developed in conjunction with three Wall Street giants to quickly and securely distribute large volumes of sensitive financial data between geographic locations, SAP Replication Server was one of the first transactional replication tools on the market. Over the years, it has been put to the test and proven in some of the world's most demanding environments – including most Wall Street financial institutions, most U.S. government services in both civilian and defense agencies, and many large retail and logistics companies.

And now, new functionality available in the latest release makes it the optimal choice for companies seeking to move data to Hadoop and harness their Big Data for real-time, active insights.

#### Enabling SAP Solutions for the Internet of Things

SAP Replication Server is also foundational to supporting the larger SAP vision for Big Data and the Internet of Things (IoT). The new, Big Data-specific functionality built into the solution moves and replicates data to Hadoop – a low-cost storage solution for Big Data. In the future, any core IoT system will require data movement, replication, and storage capabilities that enable **both** a nearly infinite store of low-value data **and** high-performance storage for highly valuable real-time business data. In addition, IoT solutions will require access to new machine-generated data, as well as an ever-expanding array of other business information sources, in order to build robust applications that provide true business value.



With over 5,000 customers globally across all industries, SAP Replication Server remains a recognized industry leader.

---

SAP® Replication Server® has garnered recognition from Forrester and Gartner as a leader in the data integration market. For example, it's one of the SAP data integration offerings that earned a leadership position in the 2015 Gartner Magic Quadrant for data integration tools. And just in the last few years, it has received over 10 patents for its performance-boosting innovations.





# Core Functionality in SAP Replication Server

With SAP Replication Server, you benefit from:

- High-performance, 100% secure data movement
- A nonintrusive, log-based methodology that captures transactions without degrading the performance of the source database or the applications running on it
- Dependable, low-maintenance, source-to-target database movement that eliminate data loss
- Real-time, bidirectional synchronization across heterogeneous data sources
- Centralized modeling and administration of complex deployment architectures
- Flexible data translations
- Distributed architecture with reliable and consistent data delivery

Let's delve deeper into these key areas of functionality and understand how they work together to enable Big Data analytics scenarios with Hadoop.

## HIGH-PERFORMANCE DATA MOVEMENTS TO BIG DATA TARGETS

At its core, SAP Replication Server is a real-time data movement solution that captures, moves, and synchronizes data from one (or many) primary database sources to one (or many) database targets – including Big Data targets such as Apache Hive – in fractions of a second. Movements start with a transactional system and either an SAP Business Suite application or a non-SAP application running on any of the supported source databases. Since SAP Replication Server is a log-based replication tool, it reads directly from the database transaction log. This eliminates interference with the database itself. The data changes are read off the log and sent through the replication engine,

where subscription and publish rules and other processes are applied. Then the data is routed and distributed to its target or targets.

## NONINTRUSIVE TRANSACTION CAPTURE WITHOUT SOURCE DEGRADATION

All data movement and replication occurs with virtually no performance impact on the primary source database. The software supports a large variety of database sources and targets, including message bus targets and unstructured data. The transactional Web service and other applications running on the source database never slow down as data changes are captured from the database log and moved to target secondary sites, including Big Data targets such as Hive.

## DEPENDABLE MOVEMENT TO TARGET DATABASES FOR ZERO DATA LOSS

As data is routed and distributed to its target or targets, the software maintains its transactional integrity by using the same transaction rules and order defined in the source database transaction. In the end, sources and targets are fully synchronized.

---

SAP® Replication Server® runs in the following operating environments:

- HP-UX
- Microsoft Windows
- IBM AIX
- Red Hat Enterprise Linux
- Solaris
- SUSE Linux





## REAL-TIME SYNCHRONIZATION ACROSS HETEROGENEOUS DATA SOURCES

With its support for real-time data distribution and synchronization from heterogeneous systems, SAP Replication Server is engineered to supply your target systems with the most current information from diverse data sources across the enterprise. This level of agility helps ensure that the analyses and reports you receive provide a holistic, up-to-the-second view of your business. There's no need for complex manual updating of processes or multiple tools for different databases, as SAP Replication Server provides data movement and synchronization across heterogeneous data sources and targets. Sources include SAP HANA, SAP ASE, Oracle, Microsoft SQL Server, IBM DB2/UDB, and IBM DB2/OS390. Targets include SAP HANA, SAP ASE, SAP IQ software, SAP Event Stream Processor, SAP Data Services software, Oracle, Microsoft SQL Server, IBM DB2/UDB, IBM DB2/OS390, Hadoop, and message buses – specifically, IBM WebSphere MQ, TIBCO Rendezvous, and Java Message Service (JMS).

## CENTRAL MODELING AND ADMINISTRATION OF COMPLEX DEPLOYMENTS

SAP Replication Server also works with modeling and systems management tools from SAP. With SAP PowerDesigner® software, you can create and capture the metadata used to describe your replication architecture and make changes more rapidly and flexibly. SAP PowerDesigner also automatically generates many of the scripts needed to create the replication logic definitions. And its powerful yet easy-to-use console gives you one

convenient place to administer and manage your replication environment in an intuitive, secure, and low-cost way.

## FLEXIBLE DATA TRANSFORMATIONS

The software also supports data transformations, including those performed through:

- **Internal mapping** – These transformations are performed using an automatic mapping scheme; the process is hard coded so that incoming data will map it to this type on SAP HANA, for example.
- **Custom function strings** – These flexible data transformations allow you to write custom add-on code to automatically manipulate data (for example, to convert it to a new format).

For more-complex data transformation needs, you can access SAP Replication Server functionality within SAP Data Services software, which provides comprehensive transformation functionality.

## A DISTRIBUTED ARCHITECTURE WITH RELIABLE DATA DELIVERY

If your organization operates across multiple geographical locations, you can realize significant benefits by having a central, synchronized source of all critical data. The flexible topology of SAP Replication Server uses your existing network to replicate data from diverse geographical locations. You can set SAP Replication Server to load from a single database source into one or more instances of your target database simultaneously, or from multiple database sources into one or more instances. This greatly increases data visibility and access across the enterprise.



# High-Performance, Real-Time Data Replication for Big Data Use Cases

What makes this solution uniquely optimized for Big Data use cases is its high performance, its support for data replication in high data volume environments, and its support for mission-critical enterprise data movement. As data is added or changed in your data sources, SAP Replication Server immediately captures it and replicates what's changed to Hadoop, your Big Data target for analytics.

## HIGH PERFORMANCE REDEFINED

Performance is a defining characteristic of SAP Replication Server. During the past few years, SAP Replication Server has received over 10 patents for its performance-boosting technology. And with the most recent release, its legendary performance is up to 600% faster than its predecessor. This increase in performance is due to the following features and functionality:

- Continuous, real-time, and log-based transactional data capture
- Structured Query Language (SQL) statement replication, which replicates only the SQL statement itself to the target database
- Stored procedure replication, which reduces replication system latency
- Optimized transaction performance, which reduces latency between the source database and target database

This extraordinary performance enables what SAP refers to as “active insights.” With active insights, there are no delays. You can aggregate and consolidate data to your analytics systems in near-real time and get to insights – including predictive insights – incredibly quickly, so you can take informed action at just the right time.

## DATABASE REPLICATION GEARED FOR BIG DATA ENVIRONMENTS

The software's high performance allows it to handle data movement, replication, and synchronization of multiple, massive data sources and targets simultaneously. But the data replication process itself also makes it ideal for Big Data environments such as Hive, which can act as the interface to replicate data into Hadoop when you need to receive data from multiple sources. In these cases, the interface between SAP Replication Server and Hive can be extended and qualified to support other sources for replication into Hive (see Figure 2). You can also extend the replication canonical interface for SAP Replication Server by adding new Hive interfaces for applications with data that must be brought into Hive; these interfaces enable Hive to consume data actively, as data is created in the source apps. As a result, connected applications become more proactive and predictive, and they function as close to real time as possible.

Figure 2: Using Apache Hive with SAP Replication Server



DSI = Data Server Interface; HVAR = High-volume adaptive replication





For Big Data support, SAP Replication Server currently works with two sources – SAP ASE and Oracle – and one target – Hive. This list will expand in the near future. SAP has focused its initial release on Hive because it is the most widely used open-source solution for storing and processing Big Data in a distributed fashion on large clusters of commodity hardware.

For specific information on operating system and database support, please see the article "[Real-Time Transactional Replication for Big Data Analytics](#)" on SAP Community Network.

#### **FULL SUPPORT FOR MISSION-CRITICAL ENTERPRISE DATA MOVEMENTS**

SAP Replication Server enables the continuous movement of mission-critical application data. Once established, this environment can be automated to help ensure that information is replicated to meet changing business demands. Whatever the environment, no matter how complex or broadly distributed, and whatever the time constraints, SAP Replication Server can meet your organization's most demanding data movement requirements.

What makes this possible is the software's Big Data replication architecture, which includes key features such as:

- Table-level replication from Oracle and SAP ASE to Hive
  - Insert, update, or delete operations, with additional columns for operation type and time stamp
  - High volume adaptive replication (HVAR)
- Manual and direct load materialization from Oracle and SAP ASE to Hive
- Movement of a variety of structured, semistructured, and unstructured data from Oracle and SAP ASE to Hive, except LOB (large object)
- Anonymous and Lightweight Directory Access Protocol (LDAP) authentication when connecting from Oracle and SAP ASE to Hive
- Secure Sockets Layer (SSL) secure communication from SAP Replication Server to Hive
- Transaction recovery from SAP Replication Server

For more detailed information about these functions, please see the article "[Real-Time Transactional Replication for Big Data Analytics](#)."



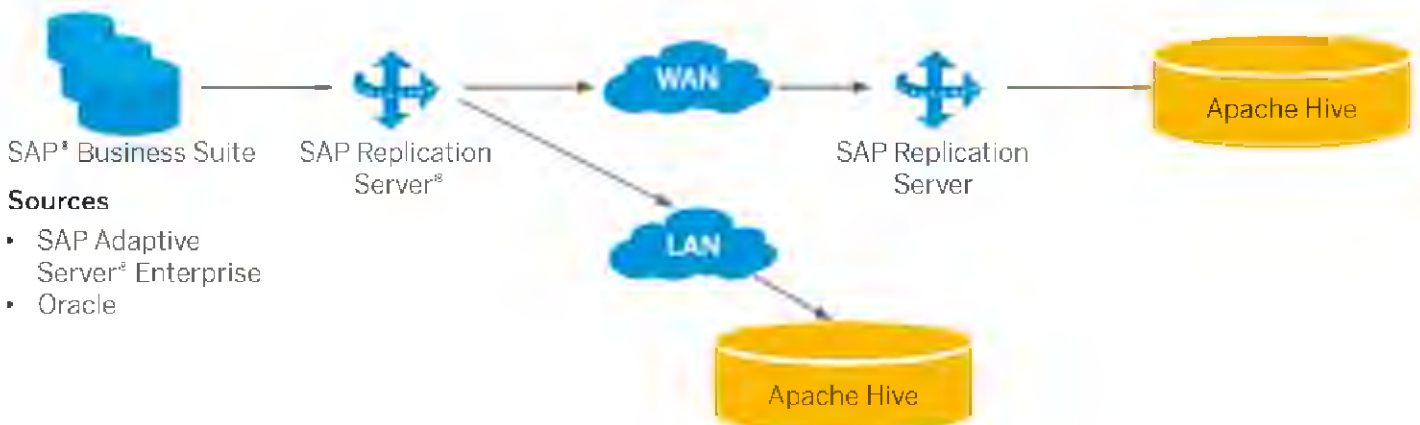


From a Big Data perspective, this architecture and functionality means you can consolidate data from disparate sources (both ASE and Oracle) into Hive tables for analysis (see Figure 3).

#### FIND OUT MORE

For more information about how SAP Replication Server can help meet your Big Data challenges, visit <http://scn.sap.com/community/sap-replication-server>.

Figure 3: Big Data Replication Architecture



SAP Replication Server provides high-performance, secure, reliable, and consistent delivery of data across the enterprise.

© 2016 SAP SE or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. Please see <http://www.sap.com/corporate-en/legal/copyright/index.epx#trademark> for additional trademark information and notices. Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors.

National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP SE or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP SE or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE's or its affiliated companies' strategy and possible future developments, products, and/or platform directions and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.



The Best-Run Businesses Run SAP®

