



# Data Replication for Hadoop



Partner Brief



"Our customers rely on Continuent Tungsten Replicator to perform the real-time data replication needed to run business-critical applications on cost-effective open source software. With the Hortonworks certification, our mutual clients can apply a high-performance, low-impact approach to transfer data from multiple upstream systems into Hadoop and rapidly extract useful information from the billions of daily transactions common in today's modern, data-driven businesses."

Robert Hodges  
CEO  
Continuent

Enterprise data has been traditionally stored in databases such as MySQL and Oracle. With the advent of Big Data, enterprises need to move away from database silos so that data can be shared and analyzed seamlessly across different systems. Traditional methods for loading data into Hadoop have relied on either full dump and load, or intermittent exports from RDBMS environments, which is slow and time-consuming.

Continuent, a leading provider of database clustering and replication offers the Tungsten Replicator solution that loads data into Hadoop at the same rate as the data is loaded and modified in the source RDBMS. Tungsten Replicator is an open source replication engine for MySQL and Oracle with high performance. Tungsten Replicator provides the cost effectiveness and flexibility of open source but coupled with enterprise grade replication features.

Tungsten Replicator supports a variety of databases and handles failures and planned maintenance. Tungsten works with open source technology to manage data requirements including replication, filtering and support for complex technologies. Tungsten helps enterprises handle terabytes of data every day using open source solutions.

Tungsten Replicator performs extraction on a MySQL server, reading the MySQL binary log containing a list of all the change events. For Oracle, information is extracted using the Oracle CDC mechanism, which provides a stream of the changes on chosen schemas within Oracle.

The partnership between Continuent and Hortonworks builds upon open source technology and provides enterprises with a cost effective solution to run business critical applications. Tungsten Replicator has been certified with the Hortonworks Data Platform.

Continuent has had many customer successes including Booking.com that is using Tungsten Replicator to move data effectively from their MySQL environment into Hadoop to perform analytics. This replaced a nightly Sqoop of the entire dataset with a live stream of the changes and enabled Booking.com to speed up the delivery and supply of information to their analysts.

## Features and Benefits

### Customizable Data Materialization

Schema, database and table definitions are loaded from the source and the changes replicated from MySQL and Oracle can be used to create carbon copy tables, combined with previously SqooPED data, translated into time series data or other transformations.

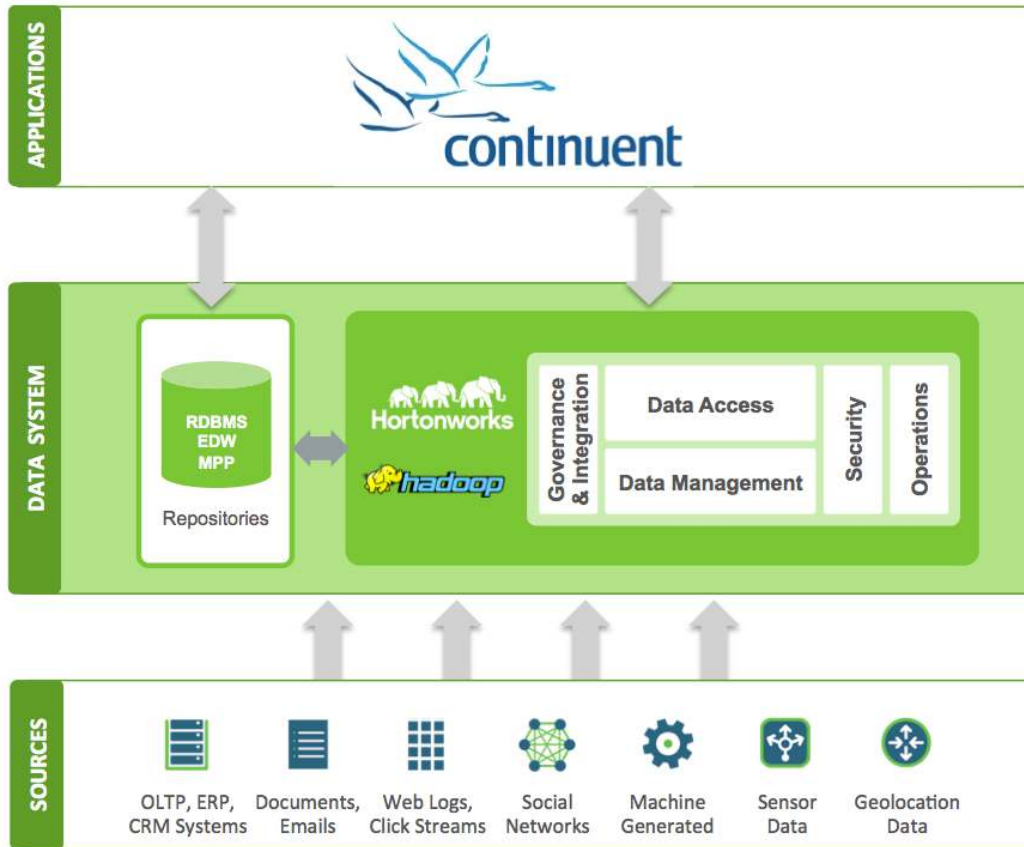
### Real-Time Data Loading

Tungsten Replicator provides a continuous stream of all the changes from the MySQL or Oracle database and writes into Hadoop in real-time providing a list of changes per schema and table.

### Low Impact Extraction and Replication

Tungsten Replicator reads and writes information with a very low impact on either the source database or the Hortonworks Data Platform. The data replicated by Tungsten matches on the source and target.

# Continuent in the Modern Data Architecture



## Features and Benefits of the Combined Solution

- Load data in real-time from MySQL or Oracle and sends data through to a replicator service within Hortonworks Data Platform (HDP)
- Generate carbon-copy of the table data to produce live tables of the replicated data
- Materialize at set intervals
- Migrate DDL and schema definitions
- Generate time series and change data

Continuent provides database replication for Hadoop environments

## Continuent is a Certified Technology Partner

The Hortonworks Certified Technology Program reviews and certifies technologies for architectural best practices, validated against a comprehensive suite of integration test cases, benchmarked for scale under varied workloads and comprehensively documented.



## Hortonworks. We do Hadoop.

Hortonworks is a leading commercial vendor of Apache Hadoop, the open source platform for storing, managing and analyzing Big Data. Hortonworks Data Platform, our distribution of Apache Hadoop, provides an open and stable foundation for enterprises and a growing ecosystem to build and deploy Big Data solutions.

Hortonworks is the trusted source for information on Hadoop, and together with the Apache community, Hortonworks is making Hadoop an enterprise data platform. Hortonworks provides unmatched technical support, training and certification programs for enterprises, systems integrators and technology vendors.

For additional questions, contact:

- Continuent  
[www.continuent.com](http://www.continuent.com)  
(866) 998-3642
- Hortonworks  
[www.hortonworks.com](http://www.hortonworks.com)  
(855) 8-HORTON