

WhereScape®

WHEREscape RED FOR BIG DATA

*Automating the integration of
enterprise and big data*

Highlights

- *Speeds up big data development*
- *Supports the integration of enterprise and big data*
- *Builds data lakes and big data platforms using standard open source projects, e.g., Hive, Sqoop*
- *Automates ingestion of data into Hive*
- *Builds standard database and big data components – no black box*
- *Automated document generation for data scientists, big data engineers and end users*
- *Big data platform and scheduler integration ensure ease of use and up-to-date information*
- *Open and integrated metadata across both big data and enterprise solutions*
- *Supports agile and iterative development best practices*

WhereScape RED is data warehouse and big data automation software for building, deploying and renovating your analytic solutions whatever the size of your data. WhereScape RED sets the standard for delivery speed using familiar industry standards, frameworks and best practices to dramatically accelerate time to value.

WhereScape RED customers will be able to fully manage their Apache Hive™ big data environments through the WhereScape RED data automation platform. This centralizes development of the entire decision support infrastructure into one integrated platform and toolset.

There is no need to license separate ETL, data integration, or data modeling tools and because WhereScape RED supports industry standard SQL, customers can leverage their existing resources and training rather than having to rely on tool or platform-specific expertise.

WhereScape®

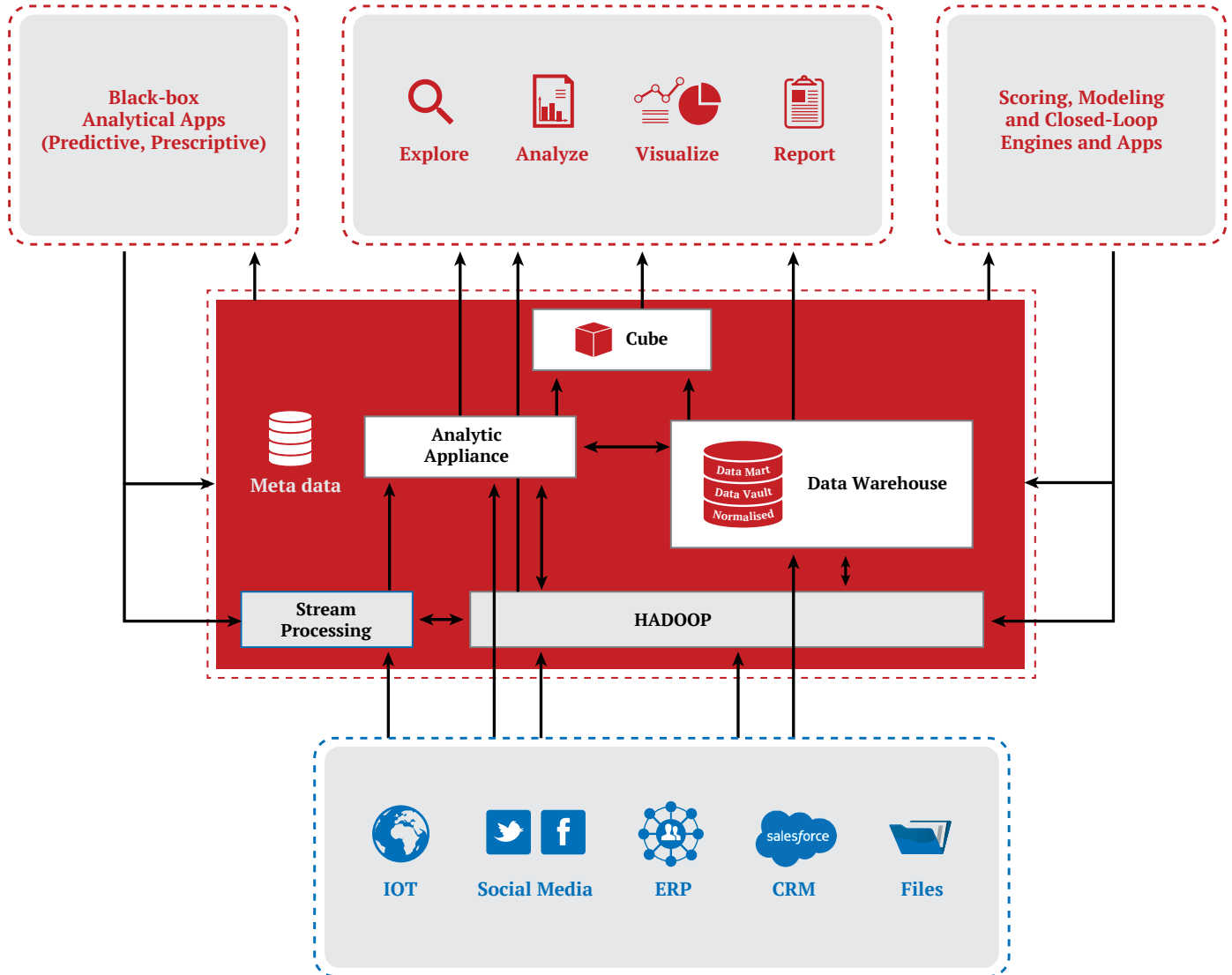
Features:

- Consistent Interface**
 - Developers use the familiar RED application interface to build both big data and enterprise solutions.
- RED Automation**
 - RED automation functionality eliminates hand-coding and tedious, repetitive tasks throughout the solution lifecycle. This provides consistent code across the entire platform and eases ongoing maintenance.
- Automatically collects and manages metadata**
 - RED repository collects and manages complete and transparent metadata for design and operational purposes. Data Lineage available to show how data spans the Hive and EDW environments, i.e., ability to see how data moves between environments.
- Automated and Integrated Scheduling and Workflow engine**
 - RED Scheduler will execute the defined processing routines on Hive, as part of our standard Job processing. This is incorporated in our broader Job processing routines, allowing Hive tasks to be managed in conjunction with the Data Warehouse processing.
- Bi-directional Data transfer between EDW and Hive**
 - Ability to transfer data in both directions, between Hive and Data Warehouse environments.
- Hive DDL Management**
 - RED will create and manage Hive DDL for Tables, Views and Indexes defined within the RED metadata repository. RED supports setting Hive DDL options, such as partitioning, clustering, table properties, skewing, row format, file format, etc.
- Hive ELT SQL Generation**
 - RED will generate native Hive SQL code and takes advantage of database-specific features where available.
- Hive Object Support**
 - Full support for Hive Managed tables (including Partitioned), External tables (including Partitioned), Views and Indexes.
- Detailed Logging**
 - All logs generated during Hive processing are captured and made available within RED.
- Documentation**
 - RED automatically manages documentation as you make changes including full data lineage, track back, track forward, and impact analysis across both enterprise and big data solutions managed by RED.

Requirement:

- RED Metadata Repository**
 - RED will utilize the existing Data Warehouse metadata repository, either on Oracle, Teradata or SQL Server.
- Edge-Node access**
 - RED server component will need to be installed on an edge-node.
- Java 1.7+**
 - Installed on edge-node. This is required for the RED server-component installed on the edge-node.
- HiveServer2**
 - HiveServer2 (HS2) component accessible. Used by both RED UI and edge component to issue DDL/DML commands during development and operations.
- Apache Sqoop™**
 - Sqoop (v1) is required if data is to be transferred between the Data Warehouse RDBMS and Hive. This must be accessible from the edge node where the RED server component is installed.

WhereScape®



Supporting the integration of enterprise and big data



About WhereScape

The pioneer in data warehouse automation software, WhereScape empowers organizations constrained by time, money or lack of resources, to deliver business value from their decision support infrastructure – including enterprise data warehouses, business facing data marts, and big data solutions. WhereScape has global operations in the USA, UK, Singapore, and New Zealand. www.wherescape.com