WhereScape®

RED Clouds Whitepaper

WhereScape®

RED Clouds

Contents

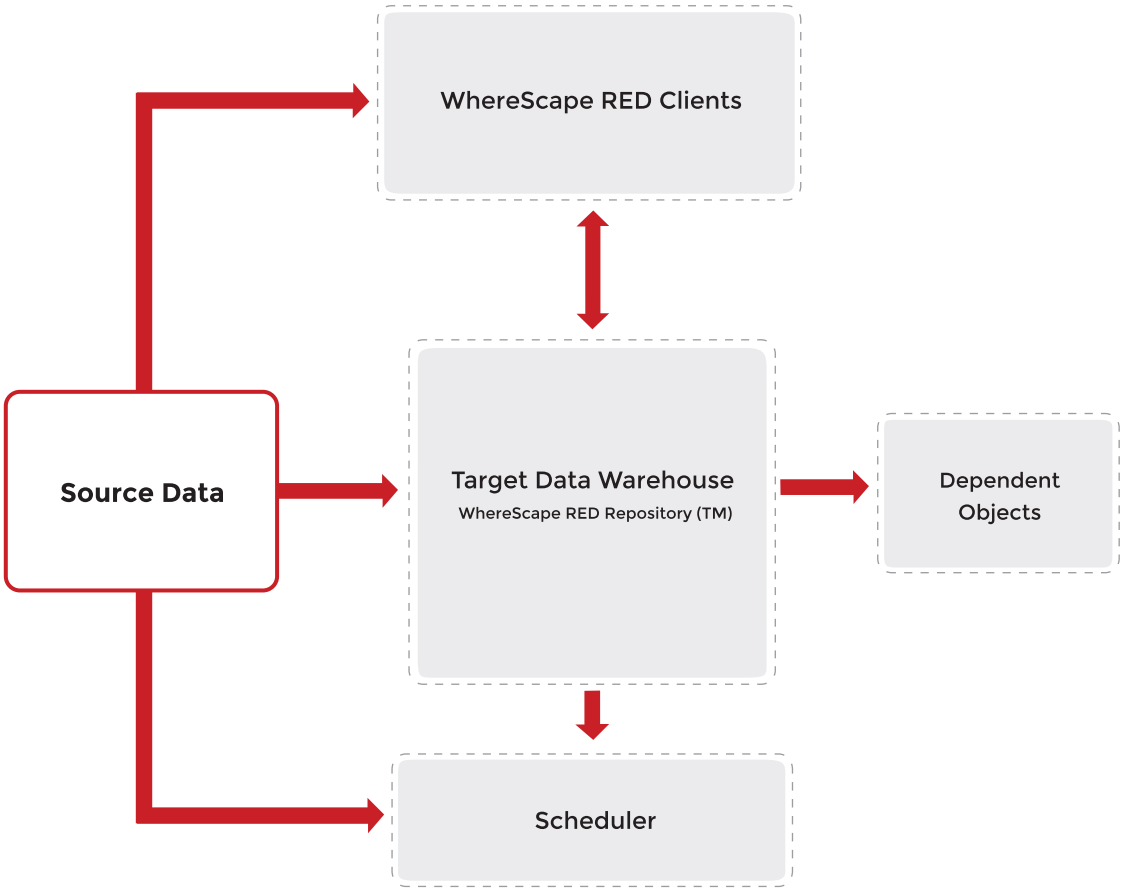
Introduction

At WhereScape we have started to be asked about data warehousing in the cloud. Currently the questions we are fielding are primarily relating to the cloud as infrastructure, rather than on models for enterprise data warehousing in the cloud. While the models for cloud based data warehousing are still being developed, it can still make sense today to share data and information in the cloud, which serves a data warehouse as a source or target.

For instance, common supplier information could be shared across the cloud and used as a source for a data warehouse. Other examples include small, specialized data marts that need to be shared across geographically distributed sales forces or departments. Using cloud services is particularly attractive for smaller and mid-sized organizations to build a data warehouse that can be accessed globally without the cost of building a local infrastructure.

This whitepaper discusses how WhereScape RED can be successfully integrated into the cloud using the cloud database services Microsoft SQL Azure and Amazon Cloud Services.

WhereScape RED Architecture



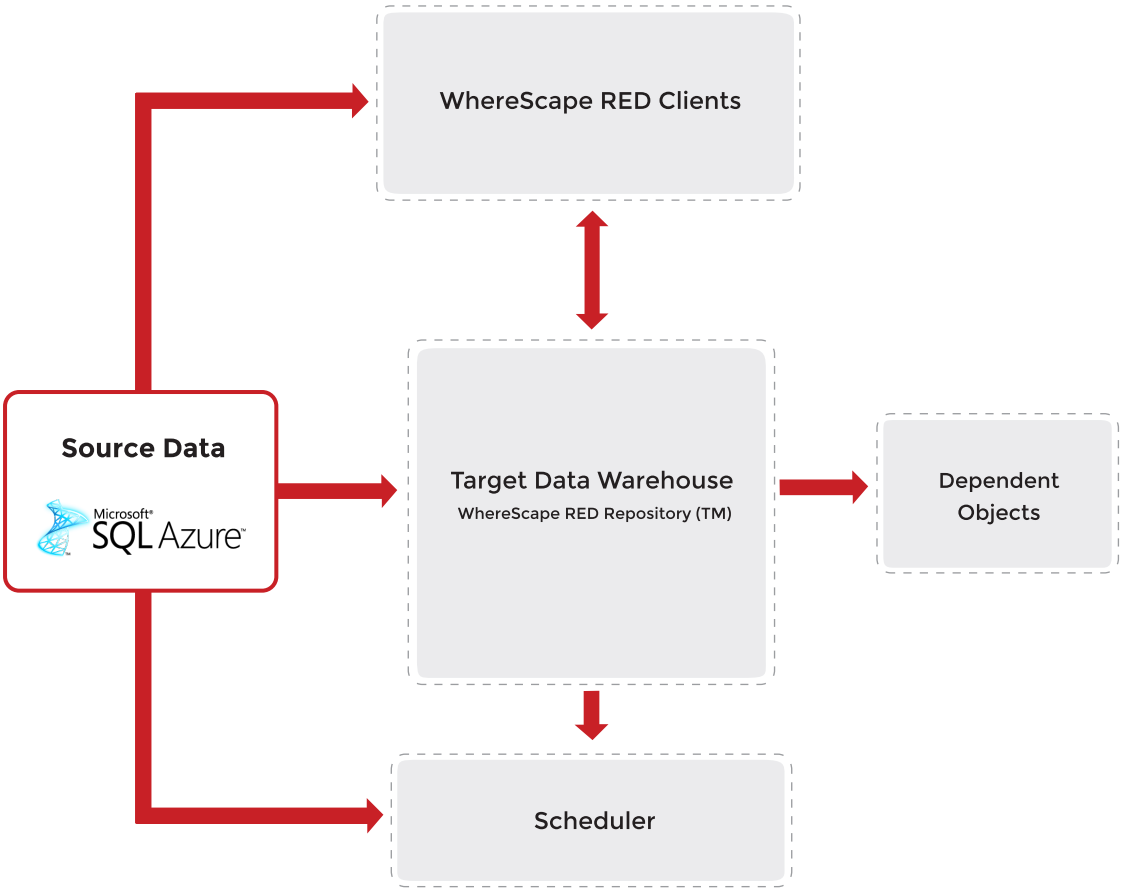
WhereScape RED is an integrated development environment (IDE) for developing and managing data warehouses. It is designed to help build data warehouses faster than using traditional methods.

WhereScape RED can be used to build SQL Server, Teradata, Oracle or DB2 data warehouses.

The core element is a single integrated metadata repository. The WhereScape RED Repository™ manages the metadata and objects associated with all of the elements of a coherent data warehouse implementation: source systems, ETL logic, schema design and client application integration.

The RED Repository™ is self-managing, self-documenting and fully versioned, so you can understand what you did when, and why, and you can restore previous versions at will. The RED Repository™ is transparent – everything is stored in the relational database of your choice, and accessible directly via SQL.

WhereScape RED and Microsoft SQL Azure

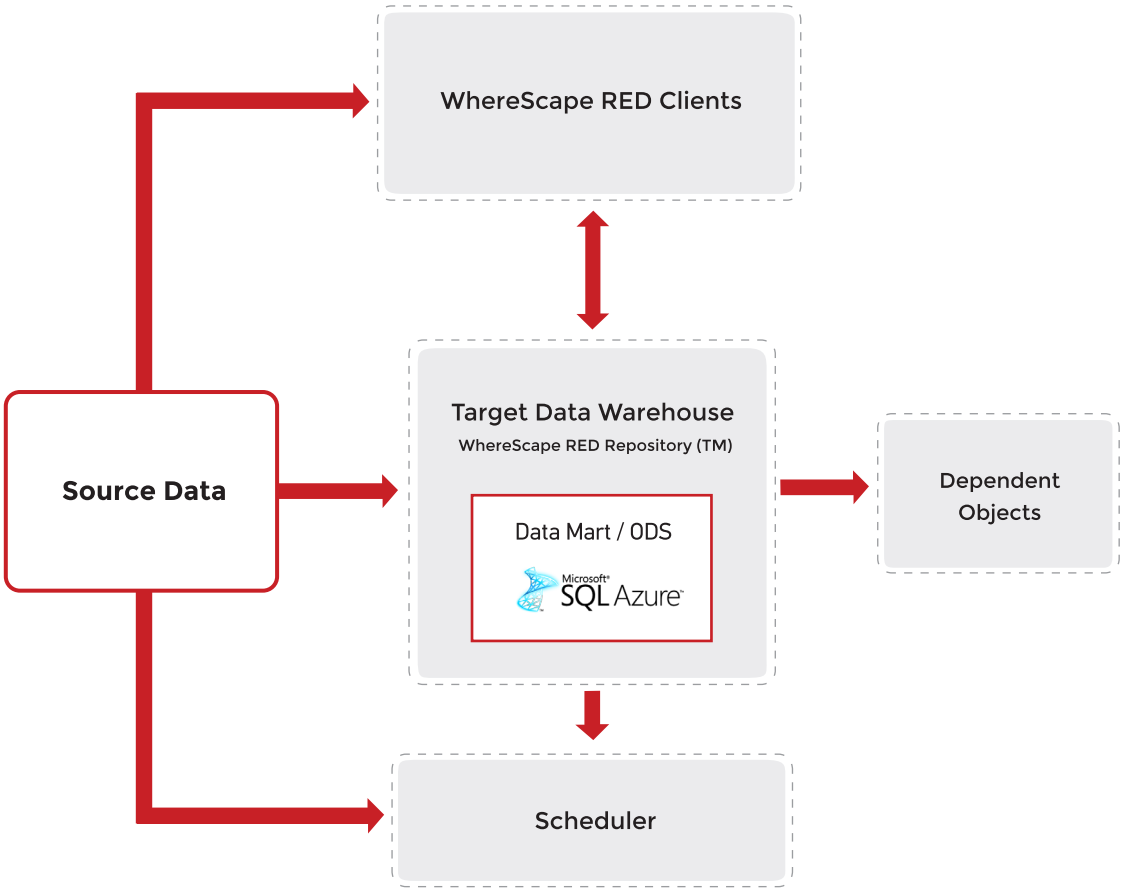


Windows Azure is Microsoft's cloud platform that provides Windows services in the cloud. SQL Azure is the Data as a Service (DaaS) component providing relational database services built on SQL Server technologies. SQL Azure contains a large subset of SQL server functionality allowing you to host your database applications in the cloud.

As WhereScape RED can be used with SQL Server as the underlying database platform, SQL Azure can be used to integrate parts of your data warehouse into the cloud. Because SQL Azure does not provide the full functionality of SQL Server it cannot host the RED Repository™ but can be used for following scenarios:

Data stored in SQL Azure can be easily loaded into a data warehouse by using SQL Azure as an ODBC source.

Moving data marts or data stores into the cloud:

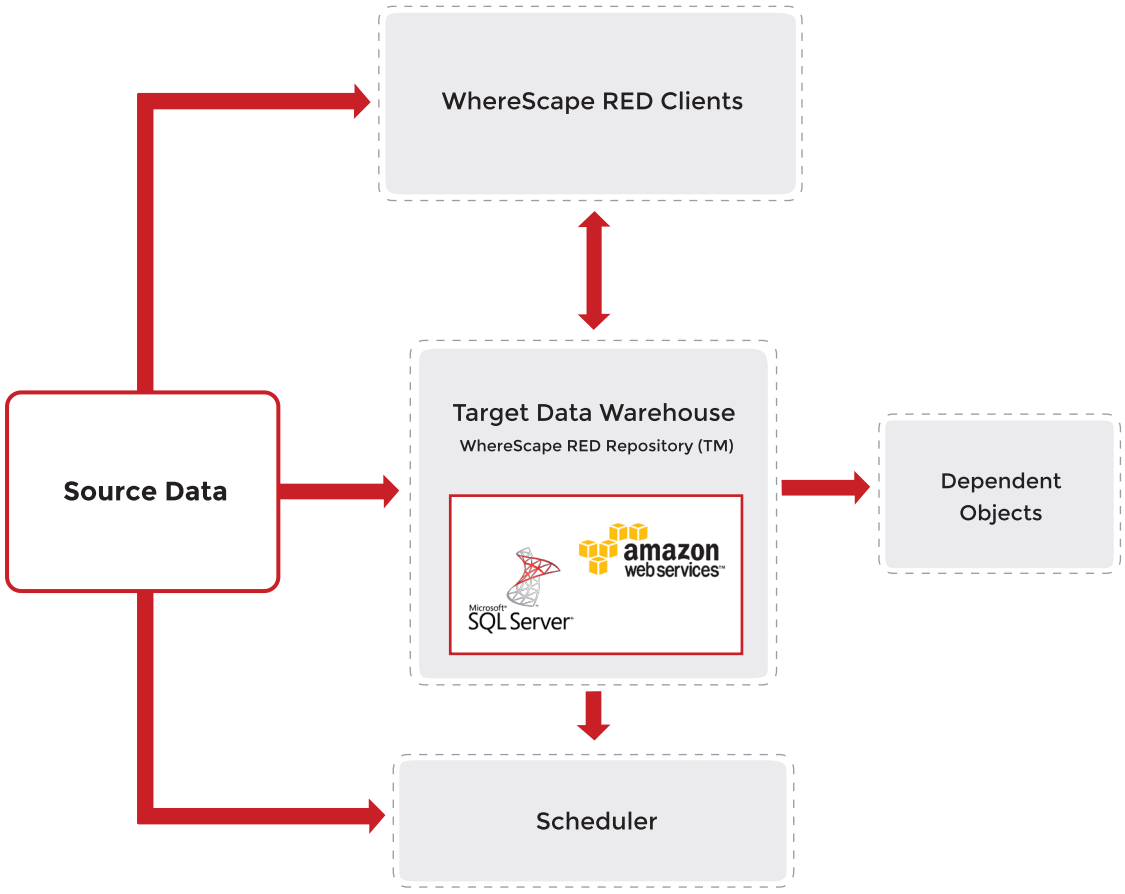


Functionality of the local SQL Server database hosting the WhereScape RED Repository™ allows you to move data marts and data stores (DS) into SQL Azure via linked server configuration and view creation in the repository database.

This method ensures that the tables in SQL Azure appear in WhereScape RED as local objects which you can process using the standard tools of WhereScape RED.

WhereScape RED and Amazon Relational Database Services

The cloud as a source of information:



Amazon Relational Database Service (Amazon RDS) allows you to set up and operate a relational database in the cloud including SQL Server. As Amazon RDS provides you with the full functionality of SQL Server, you can build a WhereScape RED Repository™ on this database and operate a fully functional data warehouse in the cloud.

Conclusion

Amazon relational database services and Microsoft SQL Azure allow you to run a data warehouse or data mart in the cloud without building the necessary infrastructure to operate and maintain. This option is particularly interesting when you have the requirement to share data across geographically distributed entities. WhereScape RED can help you to build this data warehouse fast and efficient in these early days of data warehousing in the cloud.

Windows Azure > SQL Database Notes

1. Many of the WhereScape RED repository tables do not by default have either a Primary Key, nor a Clustered Index. Tables in Windows Azure SQL Database require a Clustered Index in order to be populated with data, ie SQL Database does not allow heap tables

Note: By default a Clustered Index is created when a Primary Key is created.

It is best to check back to the following url for any changes to the requirements of needing a Primary Key or a Clustered Index on every table.

http://msdn.microsoft.com/en-us/library/windowsazure/hh974312

2. Many of the stored procedures that make up the WhereScape RED metadata repository are encrypted so they cannot be altered. Windows Azure SQL Database does not support the with **encryption option** within the **create procedure** statement.

It is best to check back to the following url for any changes to the **create procedure** statement:

http://msdn.microsoft.com/en-us/library/windowsazure/ee336289.aspx

3. The following url has the latest newsletters, that seem to correspond with Windows Azure Service Releases. Service Releases seem to be on a quarterly basis. The most recent one being around 07-Jun-2012.
Prior to that Mar-2012 and Dec-2011.

http://www.windowsazure.com/en-us/community/newsletter

4. The following url lists the known limitations of SQL Database:

http://msdn.microsoft.com/en-us/library/windowsazure/ff394102

5. The following url is a good general document on SQL Database:

http://social.technet.microsoft.com/wiki/contents/articles/1695.inside-windows-azure-sql-database.aspx

References

http://www.itbusinessedge.com/slideshows/show.aspx?c=87269&slide=9

https://www.wherescape.com/resource-library/white-papers/

http://www.windowsazure.com/en-us/home/features/what-is-windows-azure/

http://aws.amazon.com/rds/