

WhereScape: Automated DW Development

by Robin Bloor
Chief Analyst

The Immediacy of the Data Warehouse

The data warehouse (DW) is a fixture of enterprise computing, whose importance is increasing. Not so long ago, a data warehouse was designed to remain fairly static with a slow pace of change. Businesses reluctantly accepted the wait time associated with requests for new reports and specialized data marts. Data analysis was often performed on data that was days – if not weeks or months – old. Data warehouse maintenance and tuning were hand coded. In many cases, the business had no choice but to employ their own IT staff who were invested heavily in the design process.

Those days are over or rapidly coming to an end. Big Data has seen to that. The data warehouse is no longer a static creature. New data sources are added or created on a regular basis, sometimes daily. Business rules and requirements have become a prominent driver in how data gets modeled and delivered. All of this places a strain on the once-simple data warehouse design and makes maintenance far more difficult than it once was.

The Multitude of Data Sources

Big data comes in many forms: log files, streaming data, social network data, click stream data, etc. These are commonly from external sources, such as web services or cloud applications. Sometimes such data is unstructured, as is the case with text files, emails and social network data.

The data warehouse must keep pace with demand, but the established techniques and tools are not helping the cause. Typically, IT departments use ETL to connect new data sources to the data warehouse and make the data available, in a usable format, to business users. Some new data sources, particularly Hadoop, don't conform to standard ETL practices and so new skills are required to adopt and implement them.

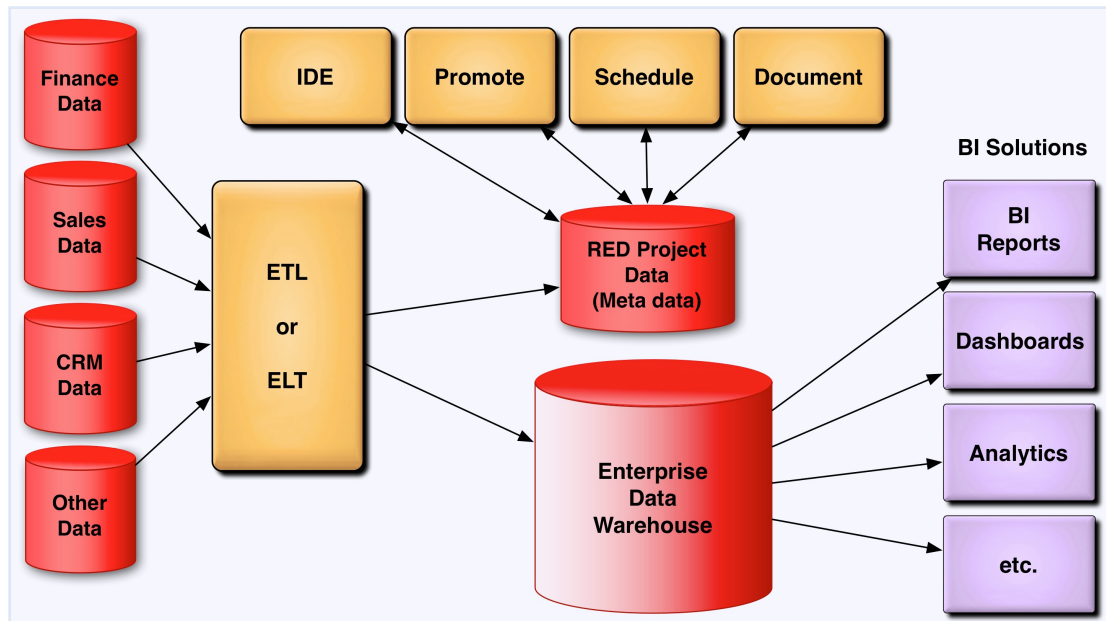
Therein lies the problem: today's data warehouse is being handicapped by the increasing number of data sources and data types – and in some situations it is being swamped. Managing the environment has evolved into an onerous chore, one that demands increasingly more time, more resources and more technology. At the same time, the business side of the house requires an increasing supply of data and at a faster pace than ever.

To get that data, many line of business users implement their own "shadow IT" – pain point solutions designed to mitigate the delay in getting the information they want. These rogue applications usually ignore important best practices, particularly in the area of data quality and documentation. However, best practices exist for good reasons.

What is needed to fix the situation is a method to ensure delivery of that data in business time engineered around a comprehensive, integrated map of all data used by the business. Fortunately, WhereScape has a product that delivers just that.

WhereScape: Automated DW Development

“WhereScape RED provides an extremely intuitive and integrated drag-and-drop interface that allows developers the type of visibility and connectivity to build models in a fraction of the time it would normally require.”



The Power of WhereScape RED

WhereScape RED is an end-to-end warehouse automation solution that provides a framework for the development and maintenance of the data warehouse. While this may seem like a simple concept, implementing it is no breeze. WhereScape RED combines ETL, ELT, metadata design and documentation, and automated iterative development in a tightly knit solution.

As a technique for fast and effective development, a rapid, iterative approach is more commonly associated with writing software and designing interfaces. The fact is, as WhereScape has demonstrated repeatedly in consultancy engagements, it is also well-suited to data warehouse design and maintenance. WhereScape RED has been constructed around this capability.

As a consequence the time taken to build, test and deploy a data warehouse is reduced dramatically. As rough rule of thumb, a brand new data warehouse can be developed and implemented in as little as 12 weeks. Similarly the time required to revise a data warehouse that has fallen into a dilapidated state can be a matter of weeks. Small tasks such as setting up new data marts are very swift – and everything is documented.

WhereScape RED delivers a collaborative environment for both IT and business users. It provides a highly intuitive integrated drag-and-drop interface for developers, giving them the type of visibility and connectivity to data sources that allows them to build models in a fraction of the usual time.

WhereScape RED stores complete audit trail and descriptive documentation for end users. Users of WhereScape RED, whether from the business side or the technical side, have the ability to access metadata views. If the source system has some descriptive properties, WhereScape RED pulls that information in directly to the data warehouse. Business users can also write in business definitions or requirements

“Perhaps the biggest differentiator between WhereScape RED and other data warehouse management tools is its pioneering automation capability.”

about the column contents.

The technical documentation allows users to look at code, data origination, target source, lineage and indexing for each object. The documentation view also exposes fact tables, hierarchies, dimensions and transformations. Everything within the documentation is hyperlinked, allowing for further navigation and drill downs.

Perhaps the biggest differentiator between WhereScape RED and other data warehouse management tools is its pioneering automation capability. WhereScape RED can automatically update or merge tables, detect changes to the data warehouse and process either batch or distinct data sets. WhereScape RED also automates table creation and population according to business rules and logic. Every automated event is immediately reflected in the metadata as well, resulting in complete documentation of both manual and automatic processes.

An aspect of WhereScape RED that is worth highlighting is its database independence. It can target multiple different databases (Oracle, SQL Server, IBM DB2, Teradata, Greenplum, etc.) and can even be used to migrate data from one to another.

WhereScape RED and Hadoop

What the world calls Big Data often runs into its own set of challenges when it comes to integrating with the data warehouse. Many businesses are now turning to Hadoop to act as a Big Data reservoir or landing area. Although it can be complex, this is similar to the data warehouse data staging areas of the past. WhereScape RED provides two methods for pulling in data from Hadoop.

The first approach lets users access the Hadoop Distributed File System (HDFS) directly, enabling movement of data to and from the data warehouse. The second approach WhereScape RED employs interacts with Hadoop via Hive. This provides the user with the ability to do much more with the data before moving it from the cluster if they so choose. Users can query, analyze and process data natively on Hadoop.

The Bottom Line

In today’s business and technology landscape, if something can be automated, it probably should be automated. We certainly believe this to be the case with the building and maintenance of the data warehouse. Pressure comes from two directions: the number of useful data sources is expanding swiftly and business users want faster access to that data. With these pressures, automation makes perfect sense.

In our view, WhereScape RED excels at delivering value in data at the speed that business needs it. Companies that suffer from data warehouse disconnect would do well to consider WhereScape RED.

Company: WhereScape
Location: Portland, OR

Product: WhereScape Red
DW Design and Maintenance