

WhereScape®

DELIVERING DATA
WAREHOUSES ON
MICROSOFT SQL
SERVER

White Paper

WhereScape®

| | |
|---|-----------|
| Delivering a Data Warehouse on the Microsoft Platform | 01 |
| • WhereScape RED and SSIS - Different approaches, different architectures | 01 |
| • Agile project delivery - A best practice approach | 01 |
| • The Data Warehouse lifecycle | 02 |
| • Comparing SSIS with WhereScape RED | 02 |
| • <i>Positioning and scope</i> | 03 |
| • <i>Data Warehouse development</i> | 04 |
| • <i>Data Warehouse management</i> | 05 |
| • <i>Data Warehouse meta data</i> | 05 |
| • WhereScape RED Benefits | 06 |
| • Summary | 06 |
| FAQ | 07 |

WhereScape®

DELIVERING DATA WAREHOUSES ON MICROSOFT SQL SERVER

Delivering a Data Warehouse on the Microsoft Platform

WhereScape RED and SSIS - Different approaches, different architectures

Microsoft SQL Server Integration Services (SSIS) is an ETL (Extract Transform and Load) tool that has been built to efficiently extract and transform data in memory on a row by row basis and then load them into a data warehouse database structure. Row level operations are very flexible, and because they are performed in memory can also be fast. SSIS provides many useful data transform tasks out of the box, however, support staff need technical knowledge of data warehousing best practice and SSIS behaviour to troubleshoot any technical issues as SSIS does not come with any frameworks or Data Warehousing best practices built in.

WhereScape RED generates code using best practice frameworks that is performed by the database itself in the form of set based code. Data is extracted and loaded very quickly and simply using bulk operations written into staging tables during intermediate steps of transformation and then published into permanent presentation tables (facts and dimensions). The set based nature of the transforms provides performance, and the use of staging tables between steps makes support and development much simpler as support staff just needs to troubleshoot SQL code and data stored in native database objects.

Both approaches and architectures described above can provide high performance. We have found that experienced data warehouse managers will always choose supportability, the ability to troubleshoot and enhance over a few percent

of processing time. They recognise that user experience is what makes a data warehouse successful, and the factors that make user experience better are:

- ▶ Fast response to change requests
- ▶ Reliable operation
- ▶ Confidence in the data

Agile project delivery – A best practice approach

Data warehousing suits an agile project approach. Agile practises are becoming common in data warehouse projects because there is recognition that users need to be involved during project development to drive requirements and prioritisation, and the project actually needs to deliver something with limited resources or time.

Agile practises in data warehousing boils down to;

- ▶ Collaboration with the business user community
- ▶ Deliver just enough to meet requirements
- ▶ Deliver early, often and be responsive to change

Building data warehouses in an agile way requires purpose built software that supports the methodology and process. Doing it any other way gets very hard very quickly. WhereScape RED has been designed specifically to address the challenges of building data warehouses in an agile way. RED supports processing, modelling, workflow, changes and documentation in a single toolset that covers the entire cycle of data warehouse development. WhereScape RED is a comprehensive data warehouse automation and lifecycle product that scales as large or as small as your data warehouse.

WhereScape®

Traditional non-agile data warehouse projects do not accept that users do not know what they want at the start of the project. The typical project phases are:

1. Gather detailed user requirements
2. Build a model based on requirements
3. Develop ETL to publish into model
4. Test
5. Deliver

This traditional process is lengthy, requires multiple toolsets and specification phases. It does not easily cope with changing requirements, this leads to the well publicised high failure rates of projects that adopt this approach. Typical reasons for project failure are as follows:

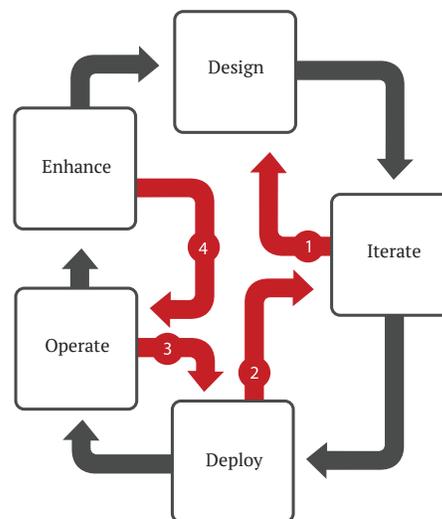
- ▶ Changing business objectives
- ▶ Inaccurate & changing business requirements
- ▶ Source data does not fit model
- ▶ Key assumptions based on static requirements and data

For traditionally run projects these problems are not exposed to the users until user acceptance testing which occurs close to the end of the project and the project budget. By comparison projects following best practice agile approaches supported by the right toolsets such as WhereScape RED reduce risk by rapidly closing the feedback loop with the business so the project can move forward with confidence.

The Data Warehouse lifecycle

The data warehouse needs to change, frequently. The data warehouse is a part of business intelligence solution that provides feedback on organisational strategy and operations. These change frequently as the organisation reacts to external or internal forces. The data warehouse lifecycle is made up of a series of cycles:

- ▶ Rapid design and agile prototyping phase
- ▶ One or several deployment phases
- ▶ Changes to solution due to operational requirements
- ▶ Changes to solution due to changing business requirements



The Data Warehouse Lifecycle process

Comparing SSIS with WhereScape RED

Microsoft SQL Server Integration Services (SSIS) and WhereScape RED are very different products however both can be used to build a data warehouse. It is difficult to compare purely SSIS and WhereScape RED, because SSIS would be part of a suite of products required to provide equivalent functionality to WhereScape RED. The areas that are covered by this document:

- ▶ Positioning and scope
- ▶ Data warehouse development and enhancement
- ▶ Data warehouse management and operation
- ▶ Scalability and performance

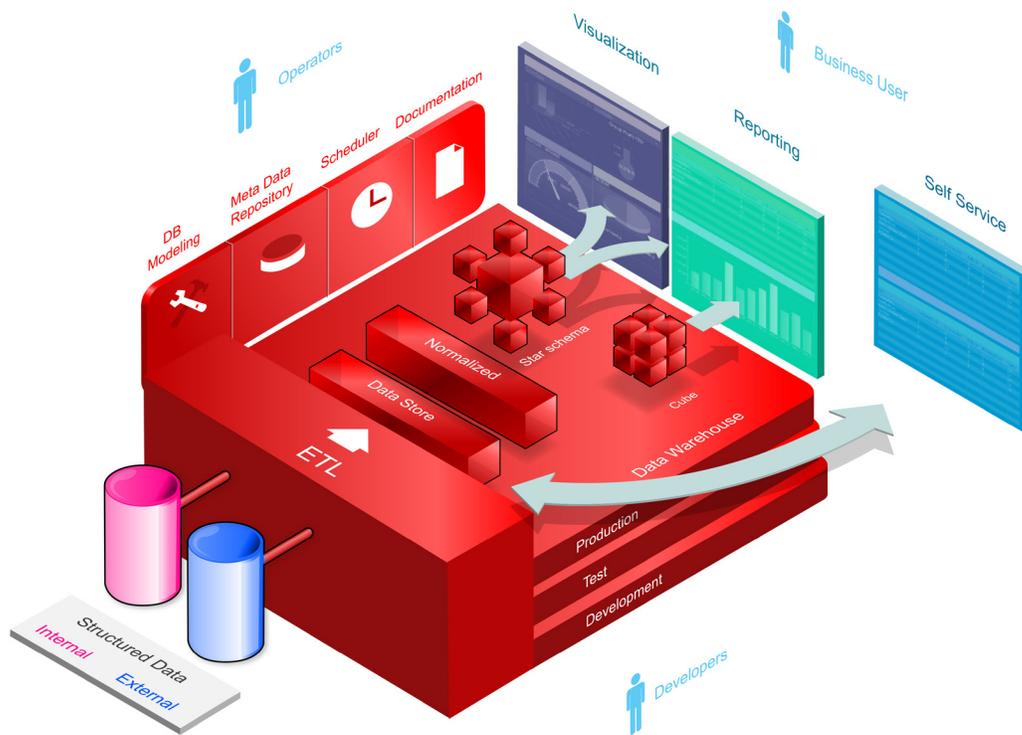
WhereScape RED has been used many times to take over the management of a data warehouse that has been built using generic ETL tools (including SSIS) because the cost of managing and enhancing the data warehouse using these tools grows over time.

WhereScape®

Positioning and scope

WhereScape RED is a complete data warehouse lifecycle product that incorporates data processing together with data modelling, workflow, metadata and documentation. It is used to build, change, migrate, operate and document a data warehouse, using one interface, one technology and one set of skills.

SSIS is a generic data processing platform that can be used to build data integration and workflow solutions, including extraction, transformation, and loading (ETL) operations for data warehousing.



Data Warehouse environment with WhereScape RED

In the diagram above the scope of SSIS is the “ETL” component, whereas WhereScape RED encompasses all functionality between source systems and BI Solution. The Microsoft software stack does not contain all of the features that WhereScape RED provides most notably in the area of Meta data and documentation. Meta data is crucially important in the data warehouse to maintain the relationships between objects and to provide the ability to describe and document the data warehouse. Without Meta data it is impossible to automate impact analysis estimates, code re-generation or automatic documentation.

WhereScape®

Data Warehouse development

WhereScape RED provides significant productivity benefits for the initial development of the data warehouse in the following main areas:

▶ **Automation** - WhereScape RED only builds data warehouses, so it can apply data warehousing best practice by default. SSIS is a generic tool so it will not apply best practice for data warehousing by default, WhereScape RED will automate a task that will take several steps in SSIS.

▶ **Scope** – WhereScape RED builds and manages all aspects of the data warehouse – SSIS does not. In order to develop a data warehouse object (e.g. a dimension) RED will automatically create the processing, workflow, table structure, indexes and documentation. SSIS can only manage the processing and workflow. This manifests itself in a siloed development environment where different tools and skills are required to manage the equivalent development in WhereScape RED introducing considerable productivity overheads.

▶ **Consistency** – SSIS is a very flexible platform but it offers no framework for providing consistency of layout of code or the application of standards. WhereScape RED provides an operational framework, standardized code templates, and configurable naming conventions. Consistency of implementation (and documentation) is crucial for team development or when development is outsourced to external parties to manage.

▶ **Testing** – continuous testing is an agile technique to measure development completeness and quality through continuously run unit and system testing. WhereScape RED provides a pre-built framework to build, run and report on tests as part of the development and deployment process. SSIS has no equivalent functionality.

▶ **Collaboration** – the user community (typically a user representative) needs to have constant input into the development process to ensure

that development effort is in alignment with business priorities. To facilitate user collaboration WhereScape RED provides the ability to quickly build and deploy discovery cubes to expose the data and structures to the user community with minimal developer effort. SSIS has no equivalent functionality.

▶ **Documentation** – during development WhereScape RED collects Meta data relating to data warehouse design, descriptive information, workflow, transformations and implementation information. This meta data is compiled into documentation and impact analysis functionality, SSIS has no equivalent functionality.

The recommended approach for building SSIS packages is to extract the data from source systems, transform the data in memory using pipeline processing tasks and then publish the data into presentation tables (dimensions and facts), minimising the number of writes to disk. This approach is very much in line with mainstream ETL tools and suits development against a source to target mapping design based on SDLC the requirements gathering approach.

Once developed, anyone wanting to examine or change a SSIS package uses the graphical interface or Package Explorer to try and track down the processes they are maintaining. The actual Data Warehouse design is managed separately using a modelling tool such as Management Studio, Visio or ERWin. SSIS's lack of integration between database design and process design slows both development and maintenance and increases the risk of error. Iterations take between 2 and 10 times longer using SSIS over WhereScape RED.

WhereScape RED enables short, fast iterations of working software during the design phase that allows business users and stakeholders to see and experience the actual Data Warehouse during development. WhereScape RED generates the Data Warehouse tables, indexes along with the load routines that populate these objects at the same time. They are managed as one object, which

WhereScape®

provides huge productivity gains when building and maintaining the Data Warehouse, and reduces technical risk.

The agile development style involves users early in the project, aligns expectations and reduces the risk that the data warehouse will not be useful. Consistency of naming standards, indexing strategy, object placement and code templates are crucial to reduce the costs of ongoing support and maintenance. WhereScape RED allows these to be defined up front and then applies them automatically to all database objects, load code, and documentation.

WhereScape RED's built in meta data supports a range of meta data reports across the entire data warehouse that are essential for larger projects – for example impact analysis reports, data lineage diagrams and workflow dependency diagrams.

Data Warehouse management

A data warehouse (when it has been developed well) is a significant organisational asset, it will change often and will become business critical. Not only that, but it is likely to have different developers building and maintaining different parts of it over its lifespan. An important function of any Data Warehouse tool should be to reduce the cost and technical risk of ongoing maintenance and enhancements.

Due to the free format of SSIS packages the only way to implement code consistency is to use external guidelines imposed on the developers. It is rare that these guidelines are prescriptive enough to save much time when troubleshooting or extending SSIS packages and, naturally, there is no guarantee that, over time, developers will follow guidelines consistently. Task precedence and flow is easy to follow in SSIS packages as they are shown graphically, however crucial object relationships are maintained and shown in other tools. For example if a support person wanted to find out which SSIS packages, and which tasks “touch” a table during data warehouse processing

they would need to open each package up and examine each task.

The WhereScape RED framework ensures that process flow is easy to implement and follow for troubleshooting or development in the data warehouse. Most, if not all, of the process logic that loads the Data Warehouse is automatically generated by WhereScape RED and is therefore very consistently implemented. Coding standards can be pre-defined in RED, so that all generated code conforms to the corporate standard. The entire data processing flow of the Data Warehouse, from source tables to presentation layer, can be viewed graphically, either interactively in WhereScape RED or in the automatically generated HTML documentation.

All WhereScape RED metadata, data warehouse objects and generated code are SQL Server tables, views, indexes and stored procedures. No hidden or binary structures are used. WhereScape RED can be un-installed and all of the data warehouse structures and processing would still be usable.

Data Warehouse Meta data

SSIS does not generate Meta data that is easy to integrate across the data warehouse. This makes documentation a manual process for the development team to hand over to support staff or end users. Any documentation that is produced is likely to be out of date when it is needed so the development team would need to check through the actual packages prior to making any changes.

WhereScape RED is a Meta data driven tool. All data warehouse objects (tables, views, database links, indexes etc.) are held in Meta data. This Meta data can be exported, deployed, backed up and documented using the tools supplied with WhereScape RED. Developers can run reports (many standard ones are built in) to perform essential tasks such as impact analysis, data lineage diagrams and workflow dependency diagrams.

WhereScape®

WhereScape RED can also generate HTML documentation right across the data warehouse from two different perspectives to support End Users and separately to support Technical support staff. WhereScape RED integrates business Meta data with technical Meta data and this can also be accessed from end user query tools.

WhereScape RED benefits

- ▶ **WhereScape RED makes your budgets go further** - Data warehouse enhancement and development costs are reduced, meaning that more money is available for reporting.
- ▶ **WhereScape RED accelerates project delivery** - New analysis areas can be built in 10 percent of the time, and Live Prototypes™, using real data, can be built in hours or days.
- ▶ **WhereScape RED makes your users advocates, not critics** - Live Prototyping™ means business users can see what their reports

will look like, with real data, sooner than ever before. This means that changes can be made earlier in the process and the final solution more targeted.

Summary

While SSIS is a useful tool for extracting and moving data into or out of Microsoft SQL Server, it is not specifically designed for the development and maintenance of Data Warehouses. Since much of the upfront cost of a data warehouse is in the development effort and the longer term costs in the effort required to operate and maintain the data warehouse, IT managers developing high value data warehouses should look carefully at the cost benefit of using WhereScape RED. WhereScape RED is a tool designed from the ground up to build high quality SQL Server data warehouses quickly and efficiently with low risk of failure and with a low TCO.

WhereScape®

Frequently Asked Questions

Q: What is it that WhereScape RED does?

A: WhereScape RED is an Integrated Development Environment (IDE) for data warehousing that prototypes, builds and enhances data warehouses ten to one hundred times faster than any other available software.

Q: Is WhereScape RED an ETL tool?

A: No – WhereScape RED does not have a proprietary engine and generates native SQL Server code, scripts and objects – it is a development and management environment not an ETL tool.

Q: What are the similarities and differences between WhereScape RED and an ETL tool?

A: Traditional ETL tools focus on data movement and finish at the stage or model layer. WhereScape RED starts with the loading, or once the data has been loaded into stage or model tables. It is focused on building sophisticated data warehouses, rather than transforming and loading data.

Q: Practically speaking how fast can you build with WhereScape RED?

A: Typically data warehousing development teams will take three to six months to bring a data warehouse from initial design through implementation, testing, and into production, even with the aid of traditional ETL tools. WhereScape RED allows small project teams – often no more than one or two people – to accomplish a task of similar magnitude in six to nine days. That’s less than 10% of the time. Assuming every project hour spent costs your organization, say, \$100 USD – and that’s a pretty conservative estimate in our experience – WhereScape RED can take \$100,000-200,000 in costs out of a project. This is easily enough to pay for itself in the first project.

Q: I have Microsoft SQL Server 2014. Can I use WhereScape RED? How?

A: Yes – WhereScape supports SQL Server 2014 including the new columnar storage (both clustered and non-clustered columnstore indexes). The tables, cubes, indexes and Transact SQL created by WhereScape RED are identical to what would be created manually by an experienced data warehouse practitioner, as each object is created metadata is maintained that can be used for lineage reporting, diagrams and documentation generation.

Q: I use SQL Server with an ETL tool. Can I use WhereScape RED after the data is loaded?

A: Yes. For instance, you can use WhereScape RED to build presentation layer objects such as summary tables, aggregate tables, join indexes and even Microsoft Analysis Services cubes from model or stage tables that have been created manually or via an ETL tool.

Q: Do I need to buy in to your way of doing things in order to use your product?

A: No, WhereScape RED is unique in its support for rapid prototyping, but you can use it with any methodology you choose including classic requirements-driven waterfall methodologies.

WhereScape®

Q: I've got an existing data warehouse. It works well. Users seem to like it. How does WhereScape RED add value for me?

A: Potentially in two distinct ways.

(1) If your warehouse is mature, and is serving all the user constituencies it's ever likely to serve, migrating your warehouse into the WhereScape RED Repository™ allows you to reduce your operations and management costs significantly and respond to incremental enhancement requests more rapidly.

(2) If your data warehouse's user constituencies are growing and their demands for breadth, depth or distribution are growing, migrating your warehouse into the WhereScape RED Repository™ allows you to build out your warehouse data sets more rapidly. With WhereScape RED you can develop and deploy a tier of data marts for specialized needs with no incremental cost and very little effort, while reducing your costs to operate and manage the data warehousing infrastructure as a whole.

Q: Do I need to start the project with WhereScape RED in order to get its benefits? Can I introduce WhereScape RED into a project that's already started?

A: Absolutely. You can slipstream WhereScape RED into a project at nearly any stage of its evolution.

Q: Can I get an evaluation copy of WhereScape RED?

A: Sure. An evaluation copy is freely downloadable from www.wherescape.com. Or, if you prefer, contact sales@wherescape.com to arrange a no obligation web demo.

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