

white paper

WhereScape®

Delivering Data Warehouses on Microsoft SQL Server

*WhereScape RED™ and SSIS -
Different approaches and architectures*

Introduction

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 can also be fast because they are performed in memory. SSIS provides useful data transformation tasks out of the box. However, support staff needs technical knowledge of data warehousing best practices and SSIS behavior to troubleshoot any technical issues. SSIS does not come with any frameworks or data warehousing best practices built-in.

WhereScape RED generates code that is performed by the database itself using best practice frameworks. Data is extracted and loaded very quickly. Simply using bulk operations written into staging tables during intermediate steps of transformation. Then, published into permanent presentation tables like Data Vault (Hubs, Links, and Satellites) or Dimensional Models (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 recognize that user experience is what makes data warehouse successful, and the factors that make the user experience better are:

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

Developing a Data Warehouse

Initial development of the data warehouse falls into the following main areas:

- **Automation** - WhereScape RED only builds data warehouses so it can apply data warehousing best practices by default. SSIS is a generic tool that will require several steps to implement a data warehouse. WhereScape RED will automate that task.
- **Consistency and Scope** - WhereScape RED provides an operation framework, standardized code templates, and configurable naming conventions. Whilst SSIS is a very flexible platform, it does not have this type of framework to support teams that rely upon documentation.
- **Testing** - For agile development environments, which go beyond pure ETL requirements, and that measure success and quality through continuous system testing, an additional benefit of WhereScape RED is its provision of a pre-built framework upon which to build, run and report on as part of the development and deployment process.

Whilst SQL Server Integration Services excels at loading data from different sources. Once the data is loaded, WhereScape RED can build all the downstream SQL Server objects such as facts, dimensions, and Microsoft Analysis Services cubes in the database 10 times faster than with traditional, hand-coding methods.

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). Minimizing 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.

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

WhereScape developers enjoy full technical and user documentation for change management and data lineage.

Once developed, anyone wanting to examine or change an SSIS package uses the graphical interface or Package Explorer to try and track down the process they are maintaining. The actual Data Warehouse design is managed separately using a modeling tool such as Management Studio, Visio, or ER/Studio. 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.

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 costs of ongoing support and maintenance. WhereScape RED allows these to be defined upfront and then applies them automatically to all database objects, load code, and documentation.

WhereScape RED's built-in metadata 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

Once data is loaded into a SQL Server data warehouse, WhereScape RED works within SQL Server's powerful database engine to transform data and build data warehouse objects within the data warehouse or project mart, in a fraction of the time that it typically takes with traditional methods. WhereScape RED will continue to support the native SQL Server load utilities it has used to import data into the database, such as BULK INSERT, linked servers, and ODBC.

A data warehouse (when it has been developed) is a significant organizational 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. These guidelines are rarely 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 are 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 support wanted to find out which SSIS packages, and which tasks "touch" a table during the data warehouse process. They would need to open each package 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 processing logic that loads the Data Warehouse, is automatically generated by WhereScape RED. Coding standards can be predefined 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 the presentation layer, can be viewed graphically, either interactively in WhereScape RED or in the automatically generated HTML documentation.

Data Warehouse Metadata

SSIS does not generate Metadata 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 before making any changes.

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

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 Metadata with technical Metadata and this can also be accessed from end-user query tools.

Agile project delivery - A WhereScape RED approach

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

Agile practices in data warehousing boil 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 require purpose-built software that supports the methodology and process. Doing it any other way becomes difficult real quick. WhereScape RED has been designed specifically to address the challenge of building data warehouses in an agile way. RED supports processing, modeling, 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 small as your data warehouse.

Traditional non-agile data warehouse projects do not accept users that do not know what they want at the start of a 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 well-publicized high failure rates of projects. Typical reasons for project failure are as follows:

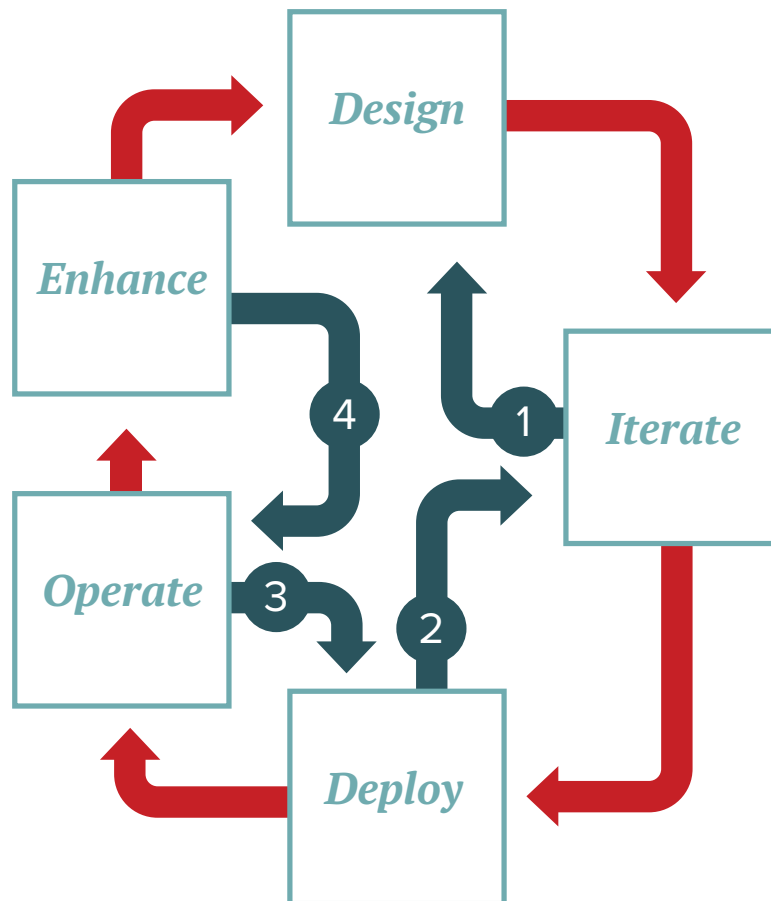
- Changing business objectives
- Inaccurate & changing business requirements
- Source data does not fit the model
- Key assumptions based on static requirements

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. By comparison, projects following the best agile approaches supported by 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 part of a business intelligence solution that provides feedback on organizational strategy and operations. These change frequently as the organization reacts to internal and external forces. The data warehouse lifecycle is made up of a series of cycles:

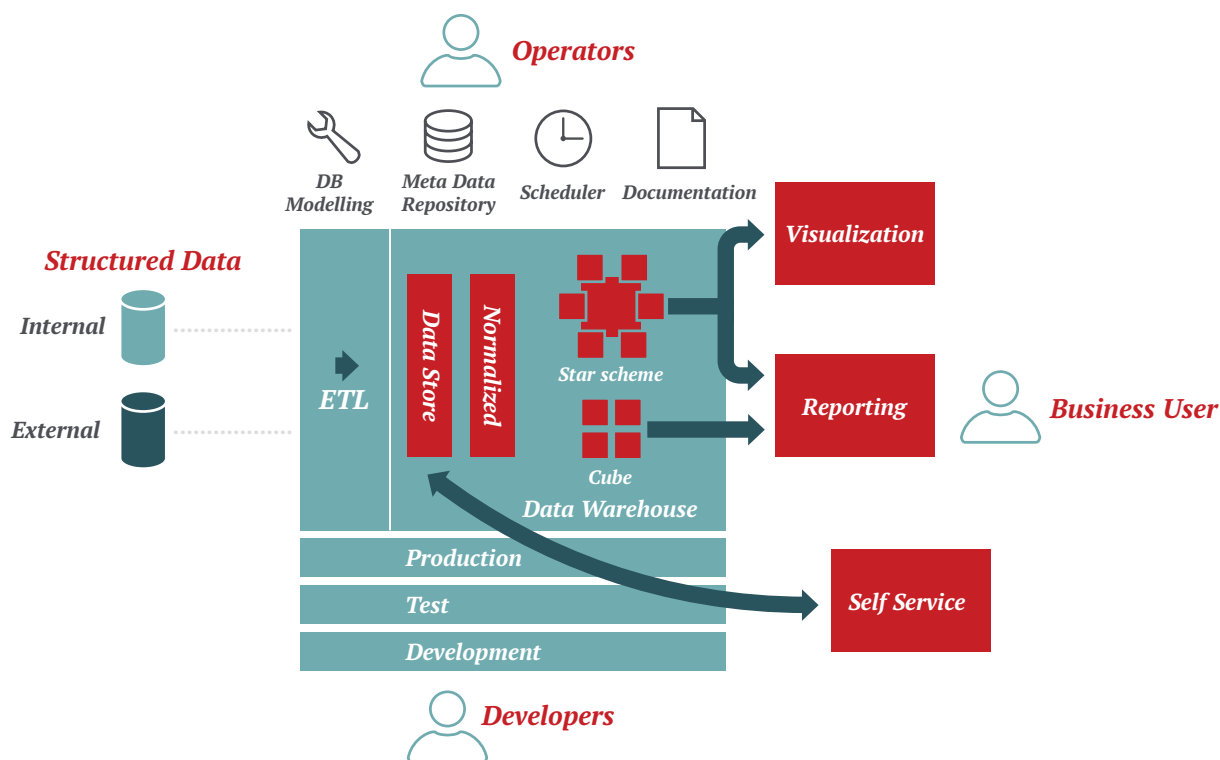
- Rapid design and agile prototyping
- One or several deployment phases
- Changes to solution due to operational requirements
- Changes to solution due to changing business requirements



Comparing ETL Tools 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 purely compare SSIS and WhereScape RED because SSIS would need to be part of a suite of products to adequately compare functionality.

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. WhereScape RED is a complete data warehouse lifecycle product that incorporates data processing together with data modeling, 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.



The diagram above shows the scope of SSIS as the “ETL” component - this provides WhereScape customers with another option to load data that is extremely fast, efficient, and flexible. It also shows WhereScape RED covering the entire journey between source systems and the target BI Solution including metadata and documentation. Metadata is crucial in maintaining the relationships between objects and to provide the ability to describe and document the data warehouse.

WhereScape RED benefits

The same WhereScape agile integrated development environment is used to build SQL Server data warehouses as well as Analysis Services cubes. The benefits are compelling – organizations only need to learn one development environment, leveraging end-to-end metadata from the source system through to the cube layer (and where the supported client tool is, as well). WhereScape delivers an integrated workflow and scheduling for the data warehouse and the cubes, automatically maintained lineage information and documentation from source system extract through to the data warehouse and the cubes.

- **Your budgets go further - Data warehouse enhancement and development costs are reduced, meaning that more money is available for reporting.**
- **Project delivery acceleration - 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 maintain the data warehouse. IT Managers 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 a lower risk of failure and low TCO.

Frequently Asked Questions

Q: What does WhereScape RED do?

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 begins with the loading, then once the data has been loaded into stage or model tables; It's focus shifts to helping the user build more sophisticated data warehouses rather than just 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 easy enough to pay for itself in the first project.

Q: I have Microsoft SQL Server 2019. Can I use WhereScape RED? How?

A: Yes – WhereScape supports SQL Server 2019 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 permanent presentation tables like Data Vault (Hubs, Links, and Satellites) or Dimensional Models (Facts, Summary, Aggregates, and Views), or stage tables that have been created manually or via an ETL tool.

Q: Do I need to buy into your way of doing things 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.

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. Migrating your warehouse into the WhereScape RED Repository™ allows you to reduce your operations and administrative costs significantly. Also the ability to 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 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. You can retrofit your existing objects by bringing them straight into WhereScape RED.

Q: Can I see a demonstration of WhereScape RED?

A: Sure. We run our virtual tours weekly at <https://www.wherescape.com/wherescape-virtual-tour>.

Feel free to sign up here.