

Accelerating Data Vault Success with Automation

A Modelers Guide to Data Vault Express™

Topics covered include:

- Comparing Data Vault with other modeling approaches
- How to evaluate which approach is best for you
- Ensuring success with WhereScape

Introduction

An overview of Data Vault 2.0 and the importance of data automation to ensure successful project delivery.

Venturing into Data Warehouse methodologies and automation may be an intimidating subject for some but it is necessary in today's fast-moving world. There are countless methodologies used in regards to data warehousing and there is no "wrong" answer. This guide will look at two of the big players in the data warehousing industry; Data Vault, and WhereScape.

Data Vault and WhereScape are a matchmade in data warehouse automation heaven. Data Vault is a methodology for data warehouses and WhereScape is there to assist Data Vault with its automation capabilities. This guide will detail the benefits of using WhereScape with the Data Vault system. Also, how it can help your company and team achieve their data goals.

Contents

Introduction	2
A background to Data Vault	3
How does Data Vault stack up against Dimensional Modeling?	5
What's better than Data? Faster Data!	6
How does WhereScape accelerate Data Vault development?	7
Demo in images	7
Successful Data Vault 2.0 Projects	10
Vodafone	10
Aptus Health	10
Further reading / Additional resources	11

A background to Data Vault

Data Vault is an agile system of business intelligence built to solve inadequacies in the Data Warehouse models that preceded it. While 3NF and Dimensional modeling address the data, they fail to address the people and the technology that impact data warehouses in 2020, especially at the enterprise level.

Today a data warehouse is not a single entity. It is a living fabric of data spread across Cloud and on-premises databases. It amalgamates sources from multiple geo-locations in a combination of batch, real-time and big data formats. You shouldn't be able to point to a single server and say "there lives my data warehouse", at least not at enterprise level. New data sources and the unpredictability of people and changing business needs must be catered for with Agile architectures and methodologies.

This is where Data Vault excels over legacy modeling techniques. "Collaboration" and "responding to change" are two of the four main principles of the Agile Manifesto, and these in turn enable the other two: working software and individuals over processes and tools.

Biography of Dan Lindstedt Inventor of Data Vault 2.0

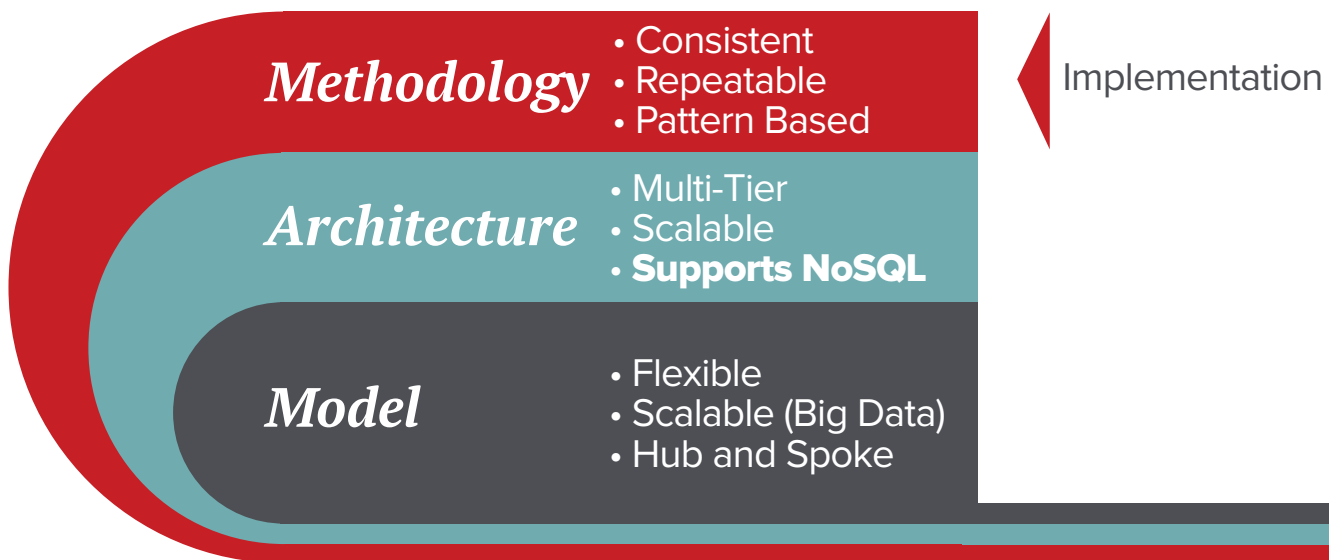
A 25-year IT Veteran, Dan Lindstedt has worked in data warehousing since 1990. Lindstedt has worked for numerous companies, doing everything from ETL to data warehousing build outs. He even solved some of the government integration problems and disparate data sets through the 90's. At the start of the millennium, Dan was convinced that if the data vault were to proceed successfully it must be done through a grassroots effort. Thus, began the creation of Data Vault 2.0. Data Vault 2.0 arrived in 2013 and brought to the table Big Data, NoSQL, unstructured, and semi-structured seamless integration, along with best practices for methodology, architecture, and implementation.



What is Data Vault 2.0?

Data Vault 2.0 is a complete system of business intelligence that contains the necessary components needed to accomplish the enterprise vision of Data Warehousing and Information Delivery. The Data Vault 2.0 approach simplifies data ingestion from multiple sources, makes it easy to add new data sources without disruption, and increases scalability and consistency. The Data Vault is designed for solving enterprise-level issues such as scalability, agility, flexibility, and consistency. The methodology standards help you drive teams in the same direction, at the same speed to meet your company's enterprise needs.

A System of Business Intelligence Comprises of



“We saw in Data Vault 2.0 a way to achieve our objectives of integration, speed, auditability, and security. It enables an agile development approach, which we were eager to use”

Michael Magalsky,
Enterprise Data Architect

How does Data Vault stack up against Dimensional Modeling in the Data Warehouse?

In scenarios where you know all your requirements and data sources before the project begins and nothing will change, Dimensional modeling and Star Schemas on SQL Servers work well. A smaller company whose business requirements remain relatively static and who do not depend upon real-time data ingestion, big data, and rapidly changing business rules, for example. But for international and enterprise companies dealing with macroeconomics, mergers, acquisitions and rapidly changing markets, etc. Dimensional modeling is too brittle.

In large companies, if you build a data warehouse using this approach for one department and another department wants one, and another, your costs spiral. If your team can't get approval to pay for building data warehouses for each department, they may use a different team to build one, and then you get data silos which make it hard to achieve a single version of the truth. Instead, you should create data marts for each department to avoid siloing.

Additional problems include:

- Uncoordinated and unintegrated data staging areas
- Challenge to change modeling styles
- Complex load patterns for conformed dimensions that make processes run slow
- Real-time data ingestion is not supported
- High cost of maintenance
- Inability to react to changes in business rules (while keeping historical data for audit)

If a business has one or more of these issues, its data warehouse is unsustainable, it could decide to rip it up and build a new one with their new requirements and sources – this is known as the greenfield approach. But these inherent problems will just emerge again and again. With Data Vault, you only have projects that add functionality. The older projects do not need to be fully tested after every change. Saving you time and costs when building.

A Data Vault enables companies to:

- Adapt to change without re-engineering
- Scale to petabyte levels
- Tackle Master Data Management

A problem can't be solved by using exactly the same plan that created it. What's required is a holistic framework that addresses technology and people as well as the data.

What's better than Data? Faster Data!

Data is all around us and tracked constantly from infinite sources. Unfortunately, bandwidth and speed are limited. Businesses want the data ASAP, but sometimes the ability to deliver is in the realm of weeks or months or years. Even after you gather the necessary data, how do you know it's valid? What's the confidence level? What if requirements change?

The Data Vault 2.0 system combined with WhereScape technologies can help with the authentication and automation of data analytic projects. This collaborative force is known as WhereScape Data Vault Express. This data automation suite consists of two products, WhereScape 3D and WhereScape RED.

3D is where all of the data modeling takes place. It provides complete project documentation, suitable for review, distribution and governance submission and makes planning, modeling, and designing data warehouses, vaults, lakes, and marts easy and effective.

RED is the customized development environment for rapidly building, deploying, managing, and updating your Data Warehouse. RED does this through a metadata-driven automation approach that generates code from templates and automates the creation for your Data Vault. This approach cuts project cost and reduces risk. With simplified and faster development, your data integration and infrastructure is at a lower risk and cost than other DIY methods.

“Our data warehouse environment was on premises, built for one particular legacy organization, and had virtually no documentation”
“While we bolted data from various systems onto the existing structure, we were challenged to fulfill the analytic needs of the business in a timely way”

Steve Gordon,
Director of Data Warehousing

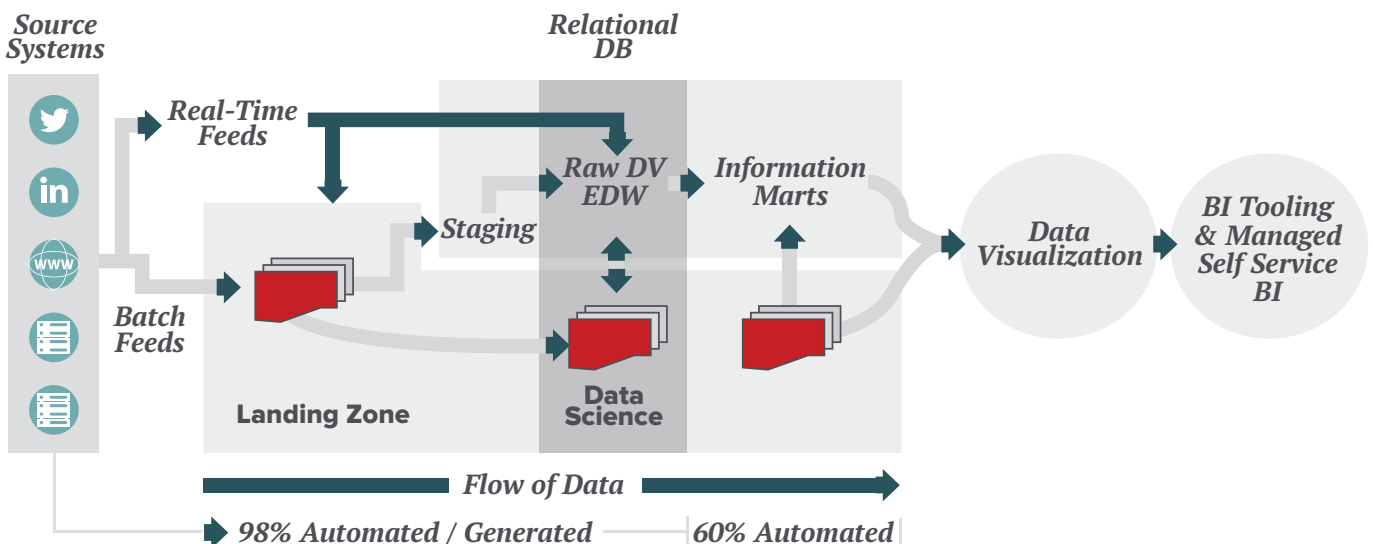
How does WhereScape accelerate Data Vault development?

To execute successfully in a Data Vault 2.0 environment you need skills, people, and processes, but you also need automation technology.

WhereScape Data Vault Express dramatically cuts the time to develop Data Vault-based analysis solutions with built-in automation, wizards, patterns, models and templates. This can lead to delivery of Data Vault-based solutions in hours and days as opposed to months and years.

Data Vault Express reduces the risk of failure by getting your project to production faster with higher quality and data consistency. WhereScape does this through automating the design, development, deployment and operation of enterprise data vaults. This includes building hubs, satellites, and links, in addition to automatically managing metadata attributes such as load date and record source. All while generating uniform and optimized code native to your target platform.

As well as creating the Data Vault tables, Data Vault Express also creates the supporting objects, which often grow to a very large number as you increase the functionality of the Data Vault. This is one of the many reasons this modeling approach cannot be effectively managed with manual processes. For Data Vault, automation tool is not a nice to have - as it was often perceived to be with dimensional modeling - it provides essential support both during implementation and day-to-day/ongoing management.



The evidence is tangible. Dan Linstedt talks about projects in which he had to ingest data from 125 source systems in six months with a team of three people and at WhereScape one customer, Micron, reported the processing of 3.2 trillion records from ten different geo-locations every 24 hours.

This is the reality of data warehousing today and projects like this just aren't possible without automation.

WhereScape automates the repetitive manual processes that once stifled Data Vault projects. There are phases of Data Vault development that must be automated as much as possible such as ELT from disparate sources to the database, and the design and creation of Data Vault structures. About 98% of this can be automated, and all developers need to do is kick the process off.

That 2% is all about metadata. Here, data acquisition teams profile the data and connect it to the business keys and business processes. WhereScape will document the metadata around it. We need to understand how the data set flows and understand at a conceptual level how the business deals with the data sets. This human element is taught in Data Vault boot camp sessions. Data Vault is not just a modeling style but a holistic system that comes with training in how to work in an augmented, Agile mindset.

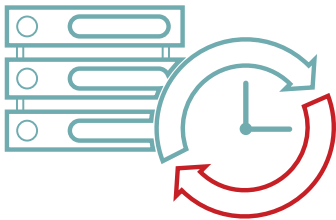
WhereScape uses templates that follow industry best practices, based on the successful completion of hundreds of Data Vault projects. Developers will want to use these templates where possible for maximum speed, but they can always customize model conversion rules, and templates before generating the physical model and code. Once these patterns are in place, you can generate thousands of times.



“Getting the opportunity to abandon an existing data warehouse and start over meant that we needed to show very rapid progress,” she says. “Automation was the only way to accelerate the delivery of business value with the same size team”

Alethea Granberg,
Senior Director of Product Development

However when it comes to modeling the data into information marts, there is no substitute for human intelligence and creativity, so with developers taking a more prominent role here as things get interesting, automation does around 60% of the work. Data teams can set up the design then use automation to accelerate team processes. You can define and WhereScape will support building information marts with all of the types of dimensions you need. It's a templated approach rather than spending precious time starting from scratch every time you need a new factoid dimension. If you can put the definition of the business rules and the right components into the metadata we can automate 60% of the downstream objects.



As well as completing development work in a fraction of the time it would take by hand, without human error, WhereScape ensures all actions are standardized and adheres to industry best practice: the methodology, the approach, the design, the naming conventions and so on. There is no room for individual flair here, as if mistakes are made at this point the Data Vault will not scale effectively. When it comes to laying foundations, uniformity and repeatability are key.

Conclusion

WhereScape is a multi-developer environment. It brings version control and enables teams around the world to leverage the same platform and follow the same methodology, the same sprint cycles and achieve the same outcomes in terms of cost, maintenance with almost zero errors put into production. If you don't have the right standards and rigor, your Data Vault, or any kind of data warehouse for that matter, will fail. We need to get everybody on the same platform and the same level of understanding, and then we can start to move faster.

The major advantage of a Data Automation tool like WhereScape sitting on top of your Data Vault, building and orchestrating all facets, is that every action you or the tool takes is fully documented in HTML. This means every action is recorded in a standardized, exportable language which reduces your reliance on individuals and means new employees can see how structures were produced. This also frees up development time, and makes sure the documentation is always produced right away, at the click of a button, rather than being done belatedly or, as can often be the case, not at all.

Who is benefitting from Data Vault with WhereScape?



vodafone

- Decommissioned an old system that had very poor data quality and plenty of bugs.
- Vodafone leveraged the power of Data Vault 2.0 and WhereScape to produce significant cost savings and favorable results.
- Processing of their run was 14-16 hours on the previous system. Now, a day's run is processed in 10% of that time.



- Replaced legacy, on-premises data warehouse and data silos with cloud-based data vault, centralizing data access and management.
- Create first data vault design in WhereScape Data Vault Express for Snowflake in 3 days.
- Delivered first production release of cloud-based data vault in three months.

Further reading / Additional resources

Videos

- Webcast “Data Vault Demystified”
- Video “Dan Linstedt explains why organizations should take a look at WhereScape”
- Webcast “TDWI on how to unlock the Data Vault with Automation”

Solution Briefs

- Is your organization ready for Data Vault?
- WhereScape Data Vault Express Solution Brief

www.wherescape.com/datavault

“We chose WhereScape Data Vault Express for its template-driven implementation, ease of use, ease of maintenance and automatic documentation”

Steve Gordon,
Director of Data Warehousing