Data Warehouse Design Patterns

Ready-to-use patterns to architect, implement and fully automate your data solution.

Workshop with Roelant Vos

Significantly improve your delivery

Research shows that most of the effort in a typical Business Intelligence project is still spent on data integration – surmounting to approximately 80% of a project's efforts. This is, as you can expect, easily costing many organisations millions - even with the semi-automation of some of this work.

As a result, our industry has identified the need to begin adopting a variety of automation techniques to deliver data solutions with better time-to-value.

However, this has not always prevented us from 're-inventing the wheel' – especially when new requirements, technologies, platforms or concepts are introduced.

This means that the question remains: how do we address the challenge of delivering data solutions quickly, consistently and reliably across our industry?

A new generation of data modelling techniques, including Data Vault and Anchor, answered by providing strong foundations to address these challenges. But data solutions are inherently complex – and complexity does not go away by using different modelling techniques, automation technology or platforms.

What is the solution?

A deep understanding of these complexities and how to address them using design patterns is the answer. Especially in order to truly leverage data, allow for automation of delivery

and the ability to look at data from different perspectives with no or limited lead time.

As a result, this training focuses on:

- The guidelines, concepts and patterns you need in your data solution architecture
- How to generate the supporting data integration processes
- How to automate the delivery using best-practice release management ('DevOps' for data)

The techniques you will learn will allow you to develop a data solution that not only maintains the pace of your business, but improves the ability to adjust direction where required.

Same data, many interpretations

Like a map that shows us partial information (or an image) of a landscape, data representing business objects or processes is not necessarily complete – and does not need to be.

Much like how you have maps for different purposes, some data models are better suited for some situations than others. You may need different ones to represent different issues, even within a single company. Effective data models are purpose driven and, although you can have a 'wrong' model, usually there is no single 'right' one.

This is where Data Modelling comes into effect, as it is the art of defining the structure by which the data is stored with a view to efficiently process it, whilst simultaneously ensuring the ability to draw the correct meaning. The data model for a data solution, such as a Data Warehouse, is purposefully designed to bring together multiple sets of data. Therefore, it needs be able to receive all incoming data from the source systems but also support the information demands of all consumers.

However, the more purposes you try to satisfy, the harder it can be to design (and especially also feed) your data model.

Data and what it represents

Data is essentially the representation of events within a business that have been recorded as evidence of business processes. Therefore, deriving an interpretation that suits a specified objective can be complex. Finding the 'right way' to represent data for any given context could require a number of iterations where business Subject Matter Experts (SME) and Data Professionals have to collaborate.

This is why a Data Warehouse model is not something you can always get right in one go. In fact, it can be tedious and time consuming for a Data Warehouse model to stabilise, and in the current fast-paced environments, this may never occur.

However, design patterns assist here by enabling you to manage the incongruity between the need to represent data through different, and sometimes conflicting, models - and the goal to pursue a single integrated model, or, the 'single version of the truth'. This approach allows users to (re)fit data in new or different perspectives without complex redevelopment.

What can this training do for me?

This Data Warehouse Design Patterns training is relevant for anyone, and everyone, seeking to understand how to leverage 'model-driven-design' and 'pattern-based code-generation' techniques to accelerate development.

By combining hybrid data modelling (e.g. Data Vault, Anchor and Ensemble Modelling approaches) with a Persistent Historical Data Store, and supporting this with code generation and process automation ('DevOps', 'DataOps') - we can reduce the repetitive aspects of data preparation whilst maintaining consistency in development.

It is, in a way, an evolution in Data Warehouse automation thinking.

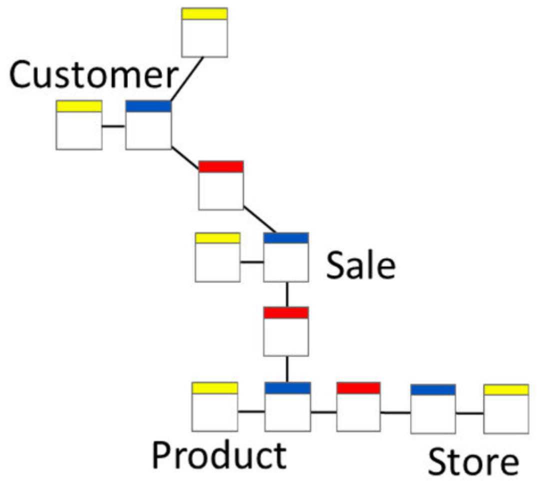
Regardless if you work on a Do-It-Yourself (DIY) solution, or have invested in any of the available Data Warehouse Automation (DWA) platforms, the concepts behind design patterns must be deeply understood to get the correct results. Success depends on correct modelling of the information, in combination with adequate application of the patterns – something software simply cannot replace yet.

Ultimately, leveraging code generation and automation techniques allows for a great degree of flexibility because you can quickly refactor and test different modelling approaches to understand which one is the best fit for you. This enables you to spend more time on higher value-adding work, such as improving the data models and delivery of your data.

As advanced modelling and implementation techniques are also covered, this training can be applied to a wide range of data professionals. The intent of the training is, after a brief introduction, to move to implementation and advanced techniques as quickly as possible.

Flexible design and implementation

Data Vault modelling has emerged as the leader of contemporary hybrid data modelling concepts for Data Warehousing. To facilitate the training, this approach will be referenced in various examples and pattern explanations.



Even though many data professionals are familiar with the basic concepts, the intricacies of implementing these into a maintainable, scalable and consistent manner are largely unknown.

Although Data Vault modelling is used for reference, the concepts and technology as covered by the training are applicable to a wider range of other Data Warehouse approaches.

The training will cover:

- The implementation of the various modelling concepts
- Complex edge-cases and advanced considerations
- The mechanisms to deliver information for consumption by business users (i.e. 'marts')
- How to produce the 'right' information by implementing business logic
- Managing multiple timelines for reporting
- This training also provides free software tools and sample data by which you can adopt in order to begin automating your own development. Alternatively, you can use these to simply understand the approaches used in commercial 'off-the-shelf' software in order to be able to fully utilise them
- A complimentary evening hands-on session (optional) is included to configure these examples, and try out code generation on your own device

In short, this training covers everything you need to implement a data solution from start to finish.

Training content and schedule

Day 1:

- Model Driven Design overview
- Solution Design & Architecture
- Solution pre-requisites and components.
- What needs to be in place? What concepts should be supported?
- Staging concepts, implementation and approaches
- Collaborative group modelling workshop
- Overview of loading patterns and their metadata requirements
- Core Business Concept pattern considerations and implementation approach (key distribution)
- Optional complimentary hands-on evening workshop

Day 2:

- Natural Business Relationship pattern considerations and implementation approach
- Context pattern considerations and implementation approach
- Handling time-variant data
- Technical considerations (indexing, partitioning, joining)
- Control framework - managing scheduling, workflows and parallelism
- Process automation and release management
- Exception handling

Day 3:

- Temporality concepts; joining time-variant data sets ('date math')
- Base tables and derived tables
- Handling late-arriving data
- Application of business logic
- Helper constructs (Point-in-Time calculations, snapshots and semi-constructs)
- Dimensions and facts

Prerequisites

- Understanding of Data Warehouse and ETL development (e.g. ETL, database and Business Intelligence platforms)
- Good understanding of SQL (e.g. joining tables, using window functions)
- Basic scripting / programming awareness (e.g. C#, Python)
- Familiarity with data modelling for Data Warehousing (e.g. CIF, Kimball / Dimensional, Ensemble techniques such as Data Vault, Anchor)