Data Solution
Design Patterns
by Roelant Vos

Course Outline

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

| Significantly improve your delivery

Research shows that most of the effort in a typical Business Intelligence (BI) 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 when requirements change?

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?

Finding a solution to this problem begins with accepting that complexity needs to be addressed somewhere, and that ultimately the data platform needs to be flexible to accommodate new requirements.

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.

This can be achieved by understanding these complexities and embed them into standard design patterns that are suitable for (data processing) code generation and automation.

That is the purpose of this training.

This is 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.

Similar to 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 in, as the art of defining the structure in which the data is stored so that the correct, unambiguous, meaning can be drawn from it.

The data model for a data solution, such as a Data Warehouse, is purposefully designed to bring together multiple sets of data. It needs be able to receive all incoming data from the operational systems that produce the original data, but also support the (changing) information demands of all consumers.

However, the more purposes you try to satisfy, the harder it can be to design your data model as well as populate it with data for consumption.

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 several iterations where business subject matter experts (SME) and data professionals must collaborate.

Therefore, a data solution model is not something you can always get right in one go. In fact, it can be tedious and time consuming for a model to stabilise, and in the current fast-paced environments, this may never be the case.

Design patterns can 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?

The Data Solution Design Patterns training is relevant for anyone, and everyone, seeking to understand how to leverage *model-driven design/engineering* and *pattern-based code-generation* techniques to accelerate development.

By combining 'hybrid' data modelling (Data Vault, Anchor and Ensemble Logical Modelling approaches) with a Persistent Staging Area (PSA), 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 products, the concepts behind design patterns must be thoroughly understood to get the correct results.

Success depends on correct modelling of the data, combined with adequate application of the patterns.

It is 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 extensively covered, this training can be applied to a wide range of data professionals.

| Flexible design and implementation

Data Vault modelling has emerged as the leader of contemporary hybrid data modelling concepts for Data Warehousing. 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.

To facilitate the training, the Data Vault approach will be referenced in various examples and pattern explanations.

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.

| Everything you need for a reliable data solution

The training will provide:

- Implementation patterns for the essential modeling concepts

- Real-world scenarios and 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

- Free software tools and sample data to begin automating your own development or apply to commercial 'off-the-shelf' software

- Time reserved for optional after-hours sessions, allowing for further discussion and development

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