



# Data Warehouse Modernization



quest for knowledge

W. [www.q4k.com](http://www.q4k.com)

E. [info@q4k.com](mailto:info@q4k.com)

P. +31 76 57 21 99

P. +32 2 808 99 46

P. +46 8 525 07 005

# COURSE DESCRIPTION



## OVERVIEW

Today, with most people connected to the Internet, the power of the customer is almost limitless. They can browse your competitors' web sites. They can compare prices, view sentiment about your business, and switch loyalty in a single click any-time, anywhere all from a mobile device. In addition, social media has given customers a voice to express opinion and sentiment about products and brands and to create social networks by attracting followers, and following others. For many CEO's, customer retention, loyalty, service and growth are top of their agenda. Therefore, they want access to new data to enrich what they already know about customers.

In addition, COO's are adding telemetry to capture new data to optimise operations. Yet at the same time, regulations like GDPR, KYC, MiFiD are everywhere, making governance and risk management also a priority.

Given these new requirements, many companies running traditional data warehouses and data marts are realising that just recording historical transaction activity is not enough. The pace of change is quickening, business is demanding lower latency data, the backlog of changes to data warehouses and data marts is growing rapidly while testing remains slow and complicated. Also with business unit autonomy, new technology available on the cloud, and pent up demand for machine learning everywhere, shadow IT is springing up in business units fracturing the analytical effort and building new analytical silos that are not integrated with data warehouses. With so much pressure to remain competitive, how then do you modernize your analytical setup, to improve governance and agility, bring in new data, re-use data assets, modernize your data warehouse to easily accommodating change, lower data latency and integrate with other analytical workloads to provide a new modern logical data warehouse for the digital enterprise?

This 2-day course looks at the business case as to why you need to do this, discusses the tools and techniques needed to capture new data types, establish a data pipeline to produce re-usable data assets, modernize your data warehouse and bring together the data and analytics needed to accelerate time to value, deliver new insights to foster growth, reduce costs, improve effectiveness and enable competitive advantage.

# COURSE DESCRIPTION



## WHY ATTEND

After completing this course, you will:

- Understand why data warehouse modernization is needed to help improve decision making and competitiveness
- Have the ingredients to know how to modernize your data warehouse to improve agility, reduce cost of ownership, facilitate easy maintenance
- Understand modern data modelling techniques and how to reduce the number of data stores in a data warehouse without losing information
- Understand how to exploit cloud computing at lower cost
- Understand how to reduce data latency
- Know how to migrate from a waterfall based data warehouse and data marts to a lean, modern logical data warehouse with virtual data marts that integrates easily with other analytical systems
- Know how to use data virtualisation to simplify access to a more comprehensive set of insights available on multiple analytical platforms running analytics on different types of data for precise evidence-based decision making
- Understand the role of a modern data warehouse in a data-driven enterprise.



## WHO SHOULD ATTEND

CDO's, CIO's, IT managers, CTO's, business analysts, data analysts, data scientists, BI managers, business intelligence and data warehousing professionals, enterprise architects, and data architects.

# COURSE OUTLINE

## 01 THE TRADITIONAL DATA WAREHOUSE AND WHY IT NEEDS MODERNIZED

---

For most organisations today, their data warehouse is based on a waterfall style architecture with data flowing from source systems into operational data stores, staging areas, then on to data warehouses under the management of batch ETL jobs. However, analytical landscape has changed. New data sources continue to grow with data now being collected in edge devices, cloud storage, cloud or on-premises NoSQL data stores, Hadoop systems as well as data warehouse staging. Hadoop, Spark, streaming data platforms and Graph databases are also now used in data analysis. Also many business units are using the cloud to quickly exploit these new analytical technologies at lower cost.

This module looks at these new activities and explains why data warehouses have to change not only to speed up development, improve agility, reduce costs but also to exploit new data, enable self-service data preparation, utilize advanced analytics and integrate with these other analytical platforms.

- The traditional data warehouse
- Multiple data stores, waterfall data architecture and data flows
- New data entering the enterprise
- The changing face of analytics – new analytical data stores and platforms
- Big Data analytics on Spark, cloud storage and Hadoop
- Real-time streaming data analytics
- Graph analysis in Graph Databases
- New challenges brought about by:
  - Data complexity
  - Data management siloes
  - Managing data in a distributed and hybrid computing environment
  - Self-service data prep vs ETL/DQ
- Problems with existing data warehouse architecture and development techniques
- The need to avoid silos, accommodate new data and integrate to deliver value

## 02 MODERN DATA WAREHOUSE REQUIREMENTS

---

This module looks at the key building blocks of modern data warehouses that need to be in place for flexibility and agility.

- Modern data modelling techniques
- Accelerating data preparation using data lake, automated data discovery, an information catalog and re-usable data assets
- Cloud based analytical DBMS
- External tables and in-database analytics
- Shortening development time using data warehouse automation
- Data Virtualisation for data independence, flexibility and to integrate new analytical data stores into a logical data warehouse
- Incorporating fast streaming data, prescriptive analytics, embedded and operational BI

# COURSE OUTLINE

## 03 MODERN DATA MODELLING TECHNIQUES FOR AGILE DATA WAREHOUSING

---

In order to improve agility, change friendly data modelling techniques have emerged and are becoming increasingly popular in designing modern data warehouses. In particular this includes Data Vault data modelling. This module looks at this data modelling approach and discusses the main components of a Data Vault model and why it is becoming popular for a modern data warehouse design but not for data mart design. It also looks at the disadvantages of such techniques and how you can overcome these.

In order to improve agility, change friendly data modelling techniques have emerged and are becoming increasingly popular in designing modern data warehouses. This module discusses which data warehouse modelling technique is best suited to handling change? Should you use Data Vault? Does Data Warehouse design need to change? Does data mart design need to change?

- Data warehouse modelling approaches - Inmon vs Kimball vs Data Vault
- The need to handle change easily
- What is Data Vault?
- Data vault modelling components - hubs, links and satellites
- Pros and cons of data modelling techniques
- Using data virtualisation to improve agility in data marts while reducing cost

## 04 MODERNIZING YOUR ETL PROCESSING

---

This module looks at the challenges posed by new data on ETL processing. Also, what options are available to modernize ETL processing, where should it run and what are the pros and cons of each option? How does this impact on your data architecture?

- New data and ETL processing - high volume data, semi-structured data, unstructured data, streaming data (e.g. IoT data)
- What are the implications and challenges of this new data on ETL processing?
- Should all this data go into a data warehouse or not?
- What options are available to modernize data warehouse ETL processing?
  - Offloading staging data to a data lake and use Spark or Hadoop for big data ETL processing
  - Using data warehouse automation software to generate ETL processing
- Pros and cons of these options
- Data architecture implications of modernizing ETL processing

# COURSE OUTLINE

## 05 ACCELERATING ETL PROCESSING USING A MULTI-PURPOSE DATA LAKE & DATA CATALOG

---

This module looks at how you can use a multi-purpose data lake to accelerate ETL processing and integration of data for your data warehouse.

- What is a data lake?
- How can it accelerate ETL processing and self-service data preparation?
- Ingesting and staging your data in a data lake
- Using an information catalog to automatically discover, profile, catalog and map data
- GDPR - Detecting sensitive data during automatic data discovery
- Creating an information supply chain to process data in a data lake
- Using Spark or Hadoop for scalable big data ETL processing
- Masking GDPR sensitive data during ingestion or ETL processing
- Is using ETL tools for processing unstructured data a good idea?
- ETL processing for streaming data in a real-time data warehouse
  - What is streaming data?
  - Types of streaming data - IoT data, OLTP system change data capture, weblogs...
  - Key technologies for processing streaming data - Kafka, streaming analytics and event stores
  - Turning OLTP change data capture into Kafka data streams
  - Linking Kafka and ETL tools to process data in real-time
  - Running ETL processing at the edge Vs on the cloud or the data centre
  - Future proofing streaming ETL processing using Apache Beam
  - Ingesting streaming data into your data lake
- Real-time data warehouse - Integrating your data warehouse with streaming data - external tables, data virtualisation and data lake
- Using ETL data pipelines to produce re-usable data assets for use in your data warehouse and other analytical data stores
- Publishing reusable data in a catalog ready for consumption
- Using data science to develop new analytical models to run in your data warehouse

# COURSE OUTLINE

## 06 RAPID DATA WAREHOUSE DEVELOPMENT USING DATA WAREHOUSE AUTOMATION

---

In order In addition to a data lake, this session looks at how you can use metadata driven data warehouse automation tools to rapidly build, change and extend modern cloud and on premises Data Warehouses and data marts. It looks at how these tools help you adopt new modern data modelling techniques quickly, how they generate schemas and data integration jobs and how they can help you migrate to your new data warehouse systems on the cloud.

- What is Data Warehouse Automation?
- Using Data Warehouse Automation Tools for rapid data warehouse and data mart development
- Generating Data Vault, E/R and Star Schema design
- ETL job generation
- Processing streaming data using Data Warehouse Automation
- Integrating big data with a data warehouse using Data Warehouse Automation
- Integrating cloud Data Warehouses with data lakes using Data Warehouse Automation
- Integrating business glossaries with Data Warehouse Automation Tools
- Using Data Warehouse Automation to migrate data warehouses
- Using Data Virtualisation to shield existing BI tools from changes in design
  - The Data Warehouse Automation Tools market e.g. WhereScape, Trivadis BIGenius, Attunity Compose, Balanced Insight, BML and more
  - Metadata driven data warehouse maintenance

## 07 BUILDING A MODERN DATA WAREHOUSE IN A CLOUD COMPUTING ENVIRONMENTG

---

A key question for many organisations is what do you do with your existing data warehouse? Should you try to change the existing set-up to make it more modern or re-develop it in the cloud? This module looks at the advantages of building modern data warehouses in a cloud computing environment using a cloud based analytical Relational DBMS.

- Why use Cloud Computing for your Data Warehouse?
- Pros and cons of deploying on the cloud?
- Cloud based data warehouse development – what are the options?
- Cloud based analytical relational DBMSs
- Amazon Redshift, Google BigQuery, Microsoft Azure SQL Warehouse, Snowflake, Teradata, Kinetica, IBM Db2 Warehouse on Cloud
  - Separating storage from compute for elasticity and scalability
  - The power of GPUs and In-memory caching in an analytical DBMS

# COURSE OUTLINE

- Managing and integrating cloud and on-premises data
- Using iPaaS software to integrate data in cloud ETL processing – Informatica IICS, Dell Boomi, SnapLogic, Talend, StreamSets.....
- Managing streaming data in the cloud
- Integrating big data analytics into a cloud based data warehouse
- Train and deploying machine learning in your analytical database for in-warehouse analytics
- Tools and techniques for migrating an existing data warehouse to the cloud
- Dealing with cloud DW migration issues
- Managing access to cloud based data warehouses
- Integrating cloud based BI systems with on-premise systems

## **08** CREATING VIRTUAL DATA MARTS AND A LOGICAL DATA WAREHOUSE ARCHITECTURE TO INTEGRATE BIG DATA WITH YOUR DATA WAREHOUSE

This module looks at how you can make use of data virtualisation software to modernize your data warehouse architecture, and simplify access to and integrate data in your data warehouse and big data underlying data stores and improve agility.

- What is data virtualisation?
- How does data virtualisation work?
- How can data virtualisation reduce cost of ownership, improve agility and modernize your data warehouse architecture?
- Simplifying your architecture by using data virtualisation to creating Virtual Data Marts
- Migrating your physical data marts to virtual data marts to reduce cost of ownership
- Layering virtual tables on top of virtual marts to simplify business user access
- Publishing virtual views and queries as services in a catalog for consumption
- Integrating your data warehouse with your data lake and low latency data using external tables and data virtualisation
- Enabling rapid change management using data virtualisation
- Creating a logical data warehouse architecture that Integrates data from big data platforms, graph databases, streaming data platforms and your data warehouse into a common access layer for easy access by BI Tools and applications
- Using a business glossary and data virtualisation to create a common semantic layer with consistent common understanding across all BI tools



# COURSE OUTLINE

## 09 GETTING STARTED WITH DATA WAREHOUSE MODERNIZATION

---

This final module looks at what you have to do to get started with a data warehouse modernization initiative. In particular it looks at:

- Data Warehouse Modernization options
- Change vs Rebuild?
  - What order do you do this in?
  - How do you minimise impact on the business while you modernize?
  - How to you deal with a backlog of change when you are also trying to modernize?
  - Pros and cons of build Vs automating data warehouse development
  - What new skills are needed?
  - Delivering new business value while you are in the progress of modernizing
  - How do you involve business professionals in the modernization effort?



## INSTRUCTOR

**Mike Ferguson** is Managing Director of Intelligent Business Strategies Limited. As an analyst and consultant he specialises in business intelligence and enterprise business integration. With over 35 years of IT experience, Mike has consulted for dozens of companies. He has spoken at events all over the world and written numerous articles. Mike is Chairman of Big Data LDN – the fastest growing Big Data conference in Europe, and chairman of the CDO Exchange. Formerly he was a principal and co-founder of Codd and Date Europe Limited – the inventors of the Relational Model, a Chief Architect at Teradata on the Teradata DBMS and European Managing Director of Database Associates. He teaches popular master classes in Analytics, Big Data, Data Governance & MDM, Data Warehouse Modernisation and Data Lake operations.



## COURSE DATES

---

10-11 JUNE 2021

VIRTUAL LIVE CET TIME

---

10-11 NOVEMBER 2021

VIRTUAL LIVE CET TIME

---



## PRICING

The fee for this 2-day course is EUR 1.450 (+VAT) per person. We offer the following discounts:

- 10% discount for groups of 2 or more students from the same company registering at the same time.
- 20% discount for groups of 4 or more students from the same company registering at the same time.

Note: Groups that register at a discounted rate must retain the minimum group size or the discount will be revoked. Discounts cannot be combined.