



Practical Guidelines for Implementing a Data Products in a Data Mesh



quest for knowledge

W. www.q4k.com

E. info@q4k.com

P. +31 76 57 21 99

P. +32 2 808 99 46

P. +46 8 525 07 005

COURSE DESCRIPTION



OVERVIEW

Most companies today are storing data and running applications in a hybrid multi-cloud environment. Analytical systems tend to be centralized and siloed like data warehouses and data marts for BI, Hadoop or cloud storage data lakes for data science and stand-alone streaming analytical systems for real-time analysis. These centralized systems rely on data engineers and data scientists working within each silo to ingest data from many different sources, clean and integrate it for use in a specific analytical system or machine learning models. There are many issues with this centralized, siloed approach including multiple tools to prepare and integrate data, reinvention of data integration pipelines in each silo and centralized data engineering with a poor understanding of source data unable to keep pace with business demands for new data. Also, master data is not well managed.

To address these issues, a new approach has emerged attempting to accelerate the creation of data for use in multiple analytical workloads. That approach is to create Data Products. This is a decentralized business domain-oriented approach to data ownership and data engineering to create a mesh of reusable data products that can be created once and shared across multiple analytical systems and workloads. Multiple data architecture options are available to create data products. These include using one or more cloud storage accounts on cloud storage, on an organized data lake, on a Lakehouse, on a data cloud, using Kafka or using data virtualization. Data products can then be consumed in other pipelines for use in streaming analytics, Data Warehouses or Lakehouse Gold Tables, for use in business intelligence, data science and other analytical workloads.

This 2-day course looks at:

- Data Products in detail and examines its strengths, and weaknesses
- The strengths and weaknesses of Data Product implementation options
- Which architecture is best to implement Data Products?
- How to co-ordinate multiple domain-oriented teams
- How to use of common data infrastructure software like Data Fabric to create high-quality, compliant, reusable, data products
- How to use a data marketplace to govern and share data products
- The objective to shorten the time to value while also ensuring that data is correctly governed and engineered in a decentralized environment
- Organizational implications of a democratized Data Product development
- How to create sharable data products for master data management AND for use in multi-dimensional analysis on a data warehouse, data science, graph analysis and real-time streaming analytics to drive business value
- Technologies like data catalogs, Data Fabric for collaborative development of data integration pipelines to create data products, DataOps to speed up the process, data orchestration automation, data marketplaces and data governance platforms

COURSE DESCRIPTION



WHY ATTEND

You will learn about:

- The problems caused in existing analytical systems by a hybrid, multi-cloud data landscape
- Strengths and weaknesses of centralized data architectures used in analytics
- What Data Products are and how they differ from other approaches?
- What benefits do Data Products offer and what are the implementation options?
- What are the principles, requirements, and challenges of implementing Data Products?
- How to organize to create data products in a decentralized environment so you avoid chaos
- The critical importance of a data catalog in understanding what data is available
- How business glossaries can help ensure data products are understood and semantically linked
- A best practice operating model for coordinating development of Data Products across different domains to succeed in implementation
- What software is required to build, operate and govern Data Products for use in a Data Lake, a Data Lakehouse or Data Warehouse?
- What is Data Fabric software, how does it integrate with data catalogs and connect to data in your data estate
- An implementation methodology to produce ready-made, trusted, reusable Data Products
- Collaborative domain-oriented development of modular and distributed DataOps pipelines to create Data Products
- How a data catalog, Generative AI and automation software can be used to generate DataOps pipelines to create Data Products
- Managing data quality, privacy, access security, versioning and the lifecycle of Data Products
- Publishing semantically linked Data Products in a data marketplace for others to consume and use
- Federated data architecture and Data Products - the emergence of lakehouses open tables as a way for multiple analytical workloads to access shared data products
- Persisting master Data Products in an MDM system
- Consuming and assembling Data Products in multiple analytical systems like data warehouses, lakehouses and graph databases to shorten time to value
- How to implement federated data governance



WHO SHOULD ATTEND

This course is intended for business data analysts, data architects, chief data officers, master data management professionals, data scientists, master data management professionals, business intelligence professionals, data warehouse professionals and data governance professionals.

COURSE DESCRIPTION



PREREQUISITES

This course assumes you understand basic data management principles and data architecture plus a reasonable understanding of data cleansing, data integration, data catalogs, data lakes and data governance.



INSTRUCTOR

Mike Ferguson is the Managing Director of Intelligent Business Strategies Limited. As an independent IT industry analyst and consultant, he specializes in BI/Analytics and data management. With over 40 years of IT experience, Mike has consulted for dozens of companies on BI/Analytics, data strategy, technology selection, data architecture, and data management.

Mike is also conference chairman of Big Data LDN, the fastest-growing data and analytics conference in Europe. He has spoken at events all over the world and written numerous articles.

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.

He teaches popular master classes in Data Warehouse Modernization, Big Data Architecture & Technology, Centralised Data Governance of a Distributed Data Landscape, Practical Guidelines for Implementing a Data Mesh (Data Catalog, Data Fabric, Data Products, Data Marketplace), Real-Time Analytics, Embedded Analytics, Intelligent Apps & AI Automation, Migrating your Data Warehouse to the Cloud, Modern Data Architecture and Data Virtualisation & the Logical Data Warehouse.

COURSE OUTLINE

01 WHAT ARE DATA PRODUCTS AND WHY ARE THEY NEEDED?

This module looks at the challenges facing companies trying to become data-driven and at the problems of siloed analytical systems and data engineering. It then looks at the emergence of Data Mesh and its introduction of Data Products as a potential way to address current problems. It explains how you can enable the creation of trusted, reusable Data Products using different data architecture options such as data lakes, data lakehouses, data virtualisation, message oriented middleware topics etc., for use in multiple analytical workloads. It also asks if combining multiple architecture approaches is advantageous or not.

- Data complexity in a hybrid, multi-cloud environment
- The growth in new data sources
- Siloed analytical systems and the IT data engineering bottleneck
- The need to industrialise data engineering to shorten time to value
- The emergence of Data Mesh and Data Products
- What is a data product?
- What types of data product can you build?
- Decentralised development of data products
- What are the challenges with this decentralised approach?
- Is data management software ready for Data Products?
- How will decentralised development of Data Products impact your current IT organization and data culture?
- Is federated data governance possible?
- What are the architectural options for implementing Data Product development and what are their strengths and weaknesses?
- Implementing Data Products on Cloud Storage Vs Lakehouse Vs Cloud Data Warehouse Vs Data Virtualisation Vs Kafka
- The promise of open table formats and a federated hybrid data architecture for building data products
- Implementation requirements to create data products?
 - Federated operating model
 - Common business vocabulary
 - Data producers and data consumers
 - Architecture independence
 - A unified data platform for building any pipeline to process any data
 - DataOps - component-based CI/CD pipeline development
 - Distributed pipeline execution
 - Reusable, semantically linked data products
 - Governance of a distributed data landscape
- Key technologies: Data Fabric, Data Catalogs, data classifiers, Generative AI in data management, Data Marketplace, Data Automation tools
- Vendor's offerings in the market - Alation, AWS, Ataccama, Boomi, Cambridge Semantics, Collibra, Denodo, Dremio, Global IDs, Google, IBM, Informatica, Microsoft, Oracle, One Data, Qlik (Talend), SAP, SAS, SnapLogic, Stratio, Starburst Data

COURSE OUTLINE

02 ORGANIZING AND STANDARDIZING YOUR ENVIRONMENT TO SUPPORT DEMOCRATISED DATA PRODUCT DEVELOPMENT

This module looks at how to standardize the setup in each business domain to optimize the development of Data Products.

- The importance of a program office
- Federated organizational structure
- Implementing Data Products on a single cloud vs a hybrid multi-cloud environment
- Implementing Data Products on a Data Lake or Lakehouse
- Standardizing the domain implementation process – ingest, process, persist, serve
- Creating zones in a domain cloud storage account, a Data Lake or Lakehouse to produce and persist Data Products
- Using Kafka as an option to persist Data Products
- Selecting Data Fabric software as a platform for domain-oriented teams to build Data Products
- Applying DataOps development practices to help standardize and version control Data Products development?

03 METHODOLOGIES FOR CREATING DATA PRODUCTS

This module looks at methodologies on how to produce business-ready, reusable Data Products for use by data consumers in multiple analytical use cases who need them to drive business value.

- A best practice step-by-step methodology for building reusable Data Products
- How does structured, semi-structured and unstructured data impact the methodology?
- Steps-by-step data product development?
 - Data concept model
 - Business glossary
 - Data source registration
 - Automated data discovery, data quality profiling, sensitive data detection, governance classification, lineage extraction and cataloguing
 - Data ingestion
 - Data Product pipeline development
 - Improving data pipeline development productivity using Generative AI
 - Standardising on best practice and taking complexity away from citizen data engineers
 - Data Product publishing for consumption
 - Global and domain policy creation for federated governance of classified data

COURSE OUTLINE

04 DEFINING AND DESIGNING DATA PRODUCTS USING A CATALOG BUSINESS GLOSSARY AND DATA MODELLING

This module looks at how you can create common data names and definitions for your data products in a business glossary so data consumers can understand the meaning of the data produced and available in a Data Products. It also looks at how business glossaries have become part of a data catalog.

- Why is a common vocabulary relevant?
- Data catalogs and the business glossary
- The Data Catalog market, e.g., Alation, Atlan, Amazon Glue, Ataccama ONE, BigID, Cambridge Semantics ANZO Data Catalog, Collibra Catalog, data.world, Denodo Data Catalog, Google Data Catalog, Hitachi Vantara Lumada, IBM Watson Knowledge Catalog, Informatica IDMC Data Governance & Catalog, Microsoft Purview Data Catalog, Oracle, Qlik (Talend) Catalog, SAP DataSphere, Top Quadrant TopBraid
- Roles, responsibilities and processes needed to manage a business glossary
- Jumpstarting a business glossary with a data concept model
- Defining Data Products using glossary terms
- Using a catalog and glossary to ensure data products are semantically linked
- Design options for Data Product data models
- Incrementally building an Enterprise Data Model while designing and building Data Products
- Assembling Data Product data model components to create a data warehouse

05 SOURCING, MAPPING AND DATA QUALITY PROFILING DATA FOR YOUR DATA PRODUCTS

This module looks at how you can use the capabilities of a data catalog to source data for your Data Products using automated data discovery. It also looks at how a data catalog can help you automate the mapping of automatically discovered raw data to the business terms of your Data Products defined in a business glossary. Finally, it looks at how data catalogs can also automatically profile the data quality of your data sources to quickly determine what data needs to be cleaned when building your Data Products.

- Sourcing data for Data Products using a data catalog for automated data discovery
- Mapping discovered physical data to Data Product business terms in your business glossary



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.



COURSE DATES

24-25 APRIL 2025

STOCKHOLM

03-04 NOVEMBER 2025

AMSTERDAM
