



Big Data Architecture and Technology for Analytics



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

This 1-day course is aimed to get you up to scratch on big data technologies such as Spark, Cloud Storage Data Lakes, Hadoop, Lakehouses (e.g., Databricks Lakehouse Platform, Apache Iceberg), Flink, Analytical SQL, NoSQL DBMSs and Multi-Platform Analytics. What is big data? How can you make use of it? How does it integrate with a traditional analytical environment? How do you re-define your architecture to create a stronger analytical foundation for your company? What skills do you need to develop for big data analytics? All of these questions are addressed in this knowledge packed course.



WHY ATTEND

You will learn:

- How big data creates several new types of analytical workloads
- Big data technology platforms beyond the data warehouse
- Big data analytical techniques and front-end tools
- Understand when to use what where - business use cases for different big data technologies
- How to create a stronger data and analytical architecture by integrating big data, data science, data warehouses and BI
- How to integrate real-time data into your data warehouse
- How to analyse un-modelled, multi-structured data using cloud storage, Spark and Hadoop
- How to leverage predictive analytics in BI reports & dashboards



WHO SHOULD ATTEND

IT directors, CIO's, CDO's, IT managers, BI managers, BI and data warehousing professionals, data scientists, enterprise architects, data architects and data engineers.

COURSE OUTLINE

01 AN INTRODUCTION TO BIG DATA

This module defines big data and looks at why business wants to use big data technology. It looks at big data use cases and the difference between big data, traditional BI and data warehousing.

- The demand for data
- Types of big data
- Why analyze big data?
- Industry use cases – popular big data analytic applications
- What is data science?
- Data warehousing and BI versus big data
- Popular patterns for big data technologies
- Types of big data analytical workloads
- Architecture options for an extended analytical ecosystem

02 BIG DATA TECHNOLOGY

This module looks at big data platforms and storage options and how all of them fit together in an end-to-end data architecture. The topics covered include:

- The new multi-platform analytical ecosystem
- Analytical RDBMSs and NoSQL options
- An introduction to the Hadoop stack
- Apache Spark framework
- The big data Hadoop marketplace
- The cloud analytics option – cloud storage versus Hadoop, Amazon (Data Lake Formation, Kinesis, Elastic MapReduce and Redshift), Google (Pub/Sub, Dataplex, Data Fusion, DataProc, Big Lake and BigQuery), Microsoft Azure (Event Hub, Stream Analytics, Data Lake Storage, HDInsight, Data Factory and Synapse Analytics, ML Service, Power BI), IBM (Streams, Analytics Engine, Db2 Warehouse on Cloud, Cloud Pak for Data), Oracle Autonomous Data Warehouse and Oracle Analytics Cloud, SAP Data Warehouse Cloud and SAP Analytics Cloud
- Accessing big data via SQL on cloud storage, SQL on Hadoop or External Tables in Cloud Data Warehouses
- The increasing power of analytical relational DBMSs
- Streaming and analyzing data on Kafka
- Analyzing big data – What's in the data scientist's toolkit
 - Streaming, natural language processing, classic machine learning at scale, deep learning, graph analytics

COURSE OUTLINE

03 INTEGRATING BIG DATA ANALYTICS INTO THE ENTERPRISE

This module looks at how new big data platforms can be integrated with traditional data warehouses and data marts to create a new data and analytics architecture for the data-driven enterprise. It looks at stream processing, cloud storage, Hadoop, NoSQL databases and data warehouse and shows how to put them together in an end-to-end architecture to maximize business value from big data.

- Beyond data warehouse – a new analytical architecture and ecosystem for the data-driven enterprise
- Integrated management of the analytical ecosystem
- Integrating stream processing, cloud storage data lakes, Hadoop, Data Warehouses and MDM
- Simplifying access to a multi-platform analytical ecosystem using data virtualization
- Multi-platform optimization – the final frontier

04 INGEST, PREPARE, ANALYZE AND GOVERN BIG DATA

This module will look at the challenge of integrating and governing big data and the unique issues it raises. How do you deal with very large data volumes and different varieties of data? How does loading data into cloud storage or Hadoop differ from loading data into analytical relational databases? What about NoSQL databases? How should low-latency data be handled? It also looks at tools and techniques available to data scientists, business analysts and traditional DW/BI professionals to analyze big data. Topics that will be covered include:

- Connecting to big data sources
- Data ingestion into cloud storage or Hadoop
 - Data ingestion options
 - Challenges of capturing different types of big data
 - Streaming data ingest
 - Parsing unstructured data
 - Change data capture – what's possible?
- ELT data preparation, transformation and integration at scale using Spark or parallel SQL on cloud data warehouses
- Managing data scientist and business analyst self-service data preparation – Alteryx (including Trifacta), Azure Data Factory, DataRobot, Google Cloud Data Fusion, IBM Cloud Pak for Data, Tamr, MicroStrategy, Salesforce Tableau Data Prep Builder and others
- Unified data delivery – a common data integration supply chain for the entire analytical ecosystem
- Multi-platform data and analytical pipelines from data lake to enterprise data marketplace
- Data governance in a big data environments
 - The importance of a data catalog
 - Organizing data in data lake storage or lakehouses
 - Governing data privacy
- Governing data in a data science environment
- Analyzing big data

COURSE OUTLINE

- Supervised and unsupervised machine learning
- Natural language processing & sentiment analysis
- Search, BI & big data
- Graph analytics
- Analyzing data in motion using streaming analytics
- Integrating it all with a self-service BI tool



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.



PRICING

The fee for this 1-day course is EUR 725 (+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

This course is only available as Customer Specific Training, whereby we can deliver private courses arranged at both a location (or virtual) and time to suit you, covering the right content to address your specific learning needs. Contact us by e-mail at info@q4k.com.