

# Introduction To Big Data

Course: **00027**

Filter: **Beginner**

Duration: **2 days**

Category: **Business Intelligence**

Price: **2800,00 €**

## About Course

What is Big Data? + Key Reasons to Learn Big Data Analytics starting with a vendor-agnostic approach: This Intro to Big Data is a unique approach to help you act on data for real business gain – not what a tool can do, but what you can do with the output from the tool. Big data as defined by Wiki is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. In this hands-on Introduction to Big Data Course, learn to leverage big data analysis tools and techniques to foster better business decision-making – before you get into specific products like Hadoop training (just to name one). Learn ways of storing data that allow for efficient processing and analysis, and gain the skills you need to store, manage, process, and analyze massive amounts of unstructured data to create an appropriate data lake.

## What you'll learn

- Store, manage, and analyze unstructured data
- Select the correct big data stores for disparate data sets
- Process large data sets using Hadoop to extract value
- Query large data sets in near real time with Pig and Hive
- Plan and implement a big data strategy for your organization

## Targeted audience

- Anyone needing to implement, enhance your big data environment and looking to advance their analytics career by ensuring foundational knowledge

- Typical job roles include: Project Managers and IT Managers, Database Administrators & Data Architects, Developers & SQL Developers, Data Scientists & Business Intelligence

## **Pre-requisites**

- Working knowledge of the Microsoft Windows platform and basic database concepts

## **Curriculum**

### **Module 1: Defining Big Data**

- The four dimensions of Big Data: volume, velocity, variety, veracity
- Introducing the Storage, MapReduce and Query Stack

### **Module 2: Delivering business benefit from Big Data**

- Establishing the business importance of Big Data
- Addressing the challenge of extracting useful data
- Integrating Big Data with traditional data
- Storing Big Data

### **Module 3: Analyzing your data characteristics**

- Selecting data sources for analysis
- Eliminating redundant data
- Establishing the role of NoSQL

### **Module 4: Overview of Big Data stores**

- Data models: key value, graph, document, column-family
- Hadoop Distributed File System
- HBase
- Hive

- Cassandra
- Hypertable
- Amazn S3
- BigTable
- DynamoDB
- MongoDB
- Redis
- Riak
- Neo4J

### **Module 5: Selecting Big Data stores**

- Choosing the correct data stores based on your data characteristics
- Moving code to data
- Implementing polyglot data store solutions
- Aligning business goals to the appropriate data store

### **Module 6: Integrating disparate data stores**

- Mapping data to the programming framework
- Connecting and extracting data from storage
- Transforming data for processing
- Subdividing data in preparation for Hadoop MapReduce

### **Module 7: Employing Hadoop MapReduce**

- Creating the components of Hadoop MapReduce jobs
- Distributing data processing across server farms
- Executing Hadoop MapReduce jobs
- Monitoring the progress of job flows

### **Module 8: The building blocks of Hadoop MapReduce**

- Distinguishing Hadoop daemons
- Investigating the Hadoop Distributed File System

- Selecting appropriate execution modes: local, pseudo–distributed and fully distributed

### **Module 9: Handling streaming data**

- Comparing real–time processing models
- Leveraging Storm to extract live events
- Lightning–fast processing with Spark and Shark

### **Module 10: Abstracting Hadoop MapReduce jobs with Pig**

- Communicating with Hadoop in Pig Latin
- Executing commands using the Grunt Shell
- Streamlining high–level processing

### **Module 11: Performing ad hoc Big Data querying with Hive**

- Persisting data in the Hive MegaStore
- Performing queries with HiveQL
- Investigating Hive file formats

### **Module 12: Creating business value from extracted data**

- Mining data with Mahout
- Visualizing processed results with reporting tools
- Querying in real time with Impala

### **Module 13: Defining a Big Data strategy for your organization**

- Establishing your Big Data needs
- Meeting business goals with timely data
- Evaluating commercial Big Data tools
- Managing organizational expectations

### **Module 14: Enabling analytic innovation**

- Focusing on business importance

- Framing the problem
- Selecting the correct tools
- Achieving timely results

### **Module 15: Implementing a Big Data Solution**

- Selecting suitable vendors and hosting options
- Balancing costs against business value
- Keeping ahead of the curve