Big Data for Decision Makers

BD4DM - Version: 1

3 days Course

Description:

This course is intended for decision makers, technical managers and team leaders, who are interested to learn how to design big data solutions. This course will introduce concepts, use-cases and leading products that are used to design a scalable solution for big data in the modern data landscape.
This course will introduce the different components a big data solution is comprised of, using pre-defined use cases as an example on how to plan a solution, from start to finish.

Intended audience:

Technical managers and Decision makers

Prerequisites:

Objectives:

Be familiar with data trends in recent years, the potential hidden in that data and the problems new data platforms need to handle with.
Be familiar with the different components needed to build a big data solution, such as data ingestion, analysis and transformation.
Be familiar with big data use cases
Be familiar with big data products

Topics:

The modern data trends

- The building blocks of big data solutions

- Ingesting, Querying, Indexing, Processing and Analyzing
- Hadoop and Hadoop eco-system

- The Lambda architecture, and other design patterns
  - Horizontal Scaling
  - Micro-Services

Ingesting Data

- How data is collected?
- Where to keep data?
- Kafka vs. Flume vs. Sqoop

Saving and Querying Data

- Distributed File Storage
  - Saving data over HDFS
  - Querying data using SQL over data files
  - Using Hive
- NoSQL Introduction
  - Key-Value Stores (Riak)
  - Columnar Databases (Cassandra)
  - Document Databases (Couchbase)
  - Graph Databases (Neo4j)

Processing Data

- Extract, Transform, Load
- ETL using Pig
- Batch Processing
- Hadoop vs. Spark
- Near-Real-Time Processing
- Storm vs. Spark Streaming

Search and Indexing data

- Full text search and beyond
- Elastic Search/Solr

Big Data solutions, End-to-End

- Public Cloud Deployments
- Workflow Management
- Security