

Course code: **BIGDATA/ARCH**

Course title: **Big Data system design for architects**

Days: 2

Description:

Course intended for

The course main audience are corporate architects designing both software and systems for processing Big Data. The material may be of interest for business analysts designing project requirements.

Course objective

The attendees will learn about new problems that arise during Big Data analysis from various sources. The demonstrated approach to solving Big Data problem is more high level and general to suite the needs of architects. Moreover, a number of up-to-date Big Data technologies will be presented during the course, both for computer cluster and cloud solutions, that suits the needs of different Big Data problems. In addition, integration of Big Data systems with existing software and systems will be discussed.

Course strengths

The course is conducted by people that have practical work experience with Big Data problems in their everyday practice. Hence, the material often goes beyond the common textbook information, that are often selective. Moreover, the content of the training is continuously updated following the modern advancements in the field. After the course the graduate will have a broad view of the Big Data and NoSQL ecosystem and will be able to choose a suitable solution for their specific business needs.

Requirements

The course does not have any strong technical requirements. Useful skills are: ability to design software and/or systems, familiarity with data analysis, business analysis, basic project management.

Course parameters

2 working days, 2*7 working hours, group 8-10 people. The course is mainly in a form of presentation and case-based exercises. Examples of demo systems can be arranged.

- Introduction to Big Data
 - Definition
 - What is Big Data?
 - History of Big Data
- Defining Big Data problems
 - Problem classification Big Data
 - Stakeholders in Big Data project
 - Do we have a Big-Data-type problem?
 - Requirements in Big Data project
- Big Data systems overview
 - Batch processing:
 - Hadoop family
 - Cascading
 - Spark
 - Stream processing:
 - Storm
 - Spark Streaming
- Data storage
 - Files:
 - HDFS
 - GlusterFS



- Databases:
 - Hbase
 - Cassandra
 - MongoDB
- Big Data distributions
 - Definition
 - Pros and cons
 - Example distributions
 - Cloudera
 - Hortonworks
 - MapR
 - Pivotal
- Cloud solutions
 - Cloud and Big Data
 - IaaS vs PaaS
 - Pros and cons
 - Solution examples
 - Amazon Web Services
 - Google Cloud
- Systems integration
 - We have Big Data system: what is next?
 - Data access: files and databases
 - Client libraries



- Database interfaces
- Business Intelligence systems
- Services
- Queues
- Full-text search (ElasticSearch, Solr) in Big Data

