

Course code: **HBASE**

Course title: **NoSQL databases - HBase**

Days: 2

Description:

Course intended for:

The training is aimed at programmers and databases administrator, who want to develop systems used for storage and/or analysis of large data sets, reaching the size of several terabytes, where use of a relational database is impossible or is not an optimum solution. The training is dedicated for beginners among the NoSQL database users, in particular, HBase, and for those, who already have some experience and would like to develop or strengthen their knowledge.

Course objective:

The users will get familiar with the theoretical aspects of the NoSQL database, they will find out, when such databases are used, and they will get familiar with the practical aspects of use of the Apache HBase database.

Course strengths:

The training strengths include its practical aspects – the participants will not only get familiar with the database architecture, but they will learn to use it effectively by writing and using MapReduce algorithms with API. In addition, during the first part of the training, the participants are introduced to the issues of NoSQL on a general level and get an opportunity to compare HBase with other databases, thanks to which they will be able to assess the best applications for individual solutions.

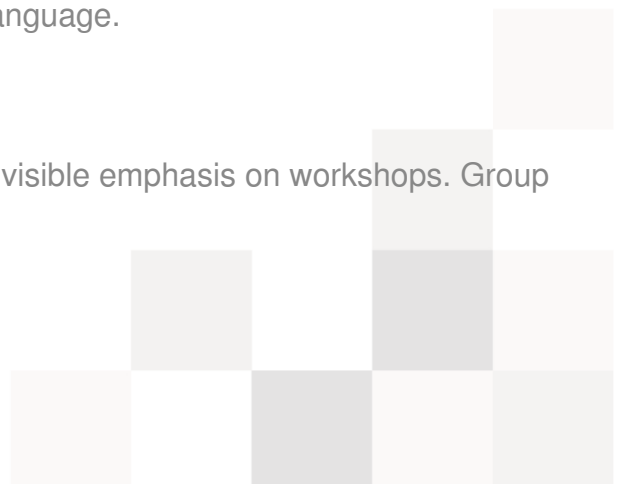
Requirements:

The participants are required to have the basic Java language programming skills, as well as basic knowledge of relational databases and SQL language.

Course parameters:

2*8 hours (2*7 net) of lectures and workshops, with visible emphasis on workshops. Group size: no more than 8-10 participants

Course curriculum:



- Introduction to BigData and NoSQL
 - History of NoSQL movement
 - Differences between NoSQL and relational databases
 - NoSQL database types
 - ACID theorem
 - CAP theorem
- HBase
 - Introduction to Hbase
 - What is it?
 - History
 - Architecture and basic components
 - Hosting machine parameters planning
 - Data model
 - Data scheme design
 - Key generation strategies
 - Use and communication
 - WWW console
 - Shell, Unix console
 - Api
 - JPA
 - JDO
 - Thrift, Avro



- Rest Api
- Asyncbase
- Integration with MapReduce
- Unit tests with Hbase
- Advanced concepts
 - Coprocessors
 - Data filtering
 - Optimization
 - Geographic operations
- Integration with other tools
 - Hive
 - PIG
 - Phoenix
 - Sqoop (integration with relational databases)
- Installation and administration
 - Installation
 - configuration
 - Basic administrative operations (launching and stopping, node adding, region management etc.)
 - Aggregation and division of regions
 - Replications
 - Backup copies
 - Work monitoring
 - Security



- Comparison with other NoSQL databases
 - MongoDB
 - Cassandra
 - Impala
 - Accumulo
 - Titan, Giraph, Neo4j

