Sages

Course code: **CASSANDRA**

Course title: **NoSQL databases - Cassandra**

Days: 3

Description:

Course intended for:

The training is intended for Web application programmers, who want to develop systems based on scalable databases, in particular, for persons, for whom performance of the solutions developed is a high priority. The training is also intended for those, who are currently engaged in development of relational databases, who would like to broaden their knowledge by learning about alternative technologies that are gradually replacing relational databases from various fields of application.

Course objective:

The training participants will acquire general knowledge on NoSQL databases, their functionalities, applications and limitations. This will allow them to select the appropriate database for their projects. The training, apart from a general introduction to non-relational databases, is focused on Cassandra database, originally developed by Facebook.

In particular, the participants are to:

- Get familiar with BigTable type data model and its implementation in Cassandra database,

- Find out what data is stored and how physical access to data is provided

- Get familiar with the mechanisms ensuring high availability and performance and the associated compromises,

- Get familiar with various clients that connect with Cassandra database, in particular, those using Java language (Pelops, Hector, Kundera) and the universal ones (Thrift, CQL).

Course strengths:

The curriculum includes a general introduction to non-relational databases and a comprehensive presentation of Cassandra database including the key database clients for

Java language.

The training is unique, as the topics discussed have not been sorted out in the available literature, and the knowledge on them is strongly dispersed.

The curriculum is constantly being updated due to fast development of the solutions analyzed.

Requirements:

The training participants are required to be able to program in Java language (the skill can be acquired during the J/JP course).

Course parameters:

3*8 hours (3*7 net hours) of lectures and workshops (with a visible emphasis on workshops).

During the workshops, apart from simple exercises, the participants will design a database and implement access to it.

Group size: no more than 8-10 participants.

Course curriculum:

1. Introduction to NoSQL databases

    - The causes of development of cloud databases

    - Consistency, Availability, Partition Tolerance

    - CAP theorem

    - Differences between NoSQL and relational databases

    - The basic parameters of NoSQL databases

    - Classification and review of NoSQL databases (Cassandra, Hbase, Mongo, Riak, CouchDB, Tokyo Cabinet, Voldemort, etc.)

    - Where it is good to use Cassandra database

    - Cassandra installation and launching

    - The in-built interface

2. The data structure

- Databases, tables, rows, columns

- Data types

- Row grouping: partitioning key and grouping key

- Collections

- Counter columns

- Help indexes

- Database development examples

- Examples of conversion of a relational database to a NoSQL database

3. Working with data and programming interfaces

- CQL3:

  - Setting of consistency level

  - Adding/ overwriting data

  - Working with collection

  - Multiple data writing operations

  - Atomic conditional write operations (small transactions)

  - Primary key inquiries

  - Index-based inquiries

  - Triggers

- An official Cassandra controller for Java language

  - Connection to cluster, sessions

  - Making inquiries and receiving results

  - Asynchronous requests
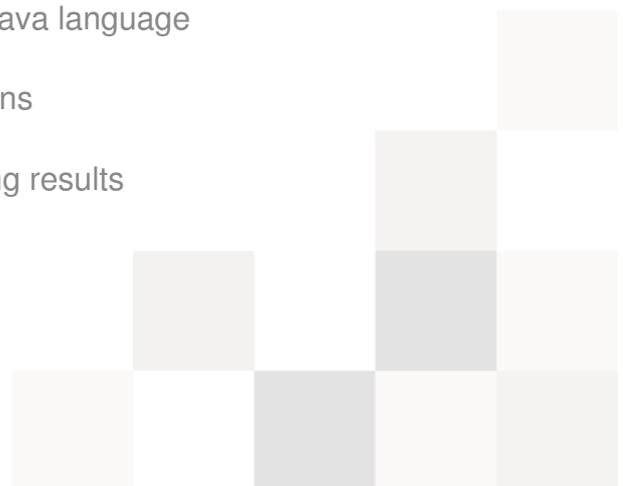
- Other interfaces: Thrift, JDBC, Pelops, Hector

- Cooperation with Hadoop infrastructure and Map-Reduce mechanism

4. Internal data structure and mechanisms

- Read-Repair mechanism

- Hinted-Handoff mechanism

- Commit-Log

- Flushing

- Compaction

- SSTables

- Compression

- Repair

5. Configuration and optimization

- Cassandra configuration

- Log and data separation

- Cluster configuration

- Partitioning

- Monitoring

- Authentication and authorization

- Backup

- Cache and memory configuration

6. Additional sources of knowledge

7. Demo and a review of capabilities of DataStax Enterprise.