Hive

Course Summary

Topics

- Big Data
- Hive DML (Data Manipulation)
- HiveQl: Queries
- HiveQL: View
- HiveQL: Indexes
- Tuning

Duration

One day

Hive

Course Outline

I. Big Data
A. MapReduce
B. Overview
C. **Example
D. Hive Overview
E. Cassandra Overview
F. Getting Access to Cassandra Data through Hive
G. Hive DDL (Data Definition)
H. Hive QL
I. Databases in Hive
J. Creating Tables
K. Dropping Tables
L. Altering Tables
M. Partitioned Managed Tables

II. Hive DML (Data Manipulation)
A. Loading Data into Tables
B. Inserting into Tables via Queries
C. Exporting Data

III. HiveQl: Queries
A. SELECT ... FROM
B. WHERE
C. GROUP BY
D. JOIN
E. ORDER BY and SORT BY
F. DISTRIBUTE BY and SORT BY
G. CLUSTER BY
H. Casting
I. Sampling Data
J. UNION ALL

IV. HiveQL: View
A. Reducing query complexity with views
B. Restricting data with views

V. HiveQL: Indexes
A. Creating/Rebuilding an index
B. Showing an Index
C. Dropping indexes

VI. Tuning
A. EXPLAIN
B. EXPLAIN EXTENDED
C. Optimized Join
D. Parallel Execution
E. Tuning # of Mappers and Reducers
F. Speculative Execution
G. JVM Reuse
H. Virtual Columns
I. Single MapReduce Multi GROUP BUY