HBase for Developers

Course Summary

Description
This course introduces HBase – a NoSQL store on top of Hadoop. The course is intended for developers who will be using HBase to develop applications, and administrators who will manage HBase clusters.

Objectives
By the end of this course, students will learn:

- HBase architecture and design
- HBase internals
- Using APIs to interact with HBase
- Data modeling on HBase

Topics
- Introduction to Big Data & NoSQL
- HBase Intro
- HBase Data model
- Accessing HBase using Java API
- HBase schema Design: Group session
- HBase Internals
- HBase installation and configuration
- HBase eco-system
- Monitoring and Best Practices

Audience
This course is designed for Developers and Administrators.

Prerequisites
Prior to taking this course, you should:

- Be comfortable with Java programming language
- Be comfortable in Java programming language (navigate Linux command line, edit files with vi / nano)
- Be familiar with a Java IDE like Eclipse or IntelliJ

Duration
Three days
HBase for Developers

Course Outline

I. Introduction to Big Data & NoSQL
   A. Big Data ecosystem
   B. NoSQL overview
   C. CAP theorem
   D. When is NoSQL appropriate
   E. Columnar storage
   F. HBase and NoSQL

II. HBase Intro
   A. Concepts and Design
   B. Architecture (HMaster and Region Server)
   C. Data integrity
   D. HBase ecosystem
   E. Lab : Exploring HBase

III. HBase Data model
   A. Namespaces, Tables and Regions
   B. Rows, columns, column families, versions
   C. HBase Shell and Admin commands
   D. Lab : HBase Shell

IV. Accessing HBase using Java API
   A. Introduction to Java API
   B. Read / Write path
   C. Time Series data
   D. Scans
   E. Map Reduce
   F. Filters
   G. Counters
   H. Co-processors
   I. Labs (multiple): Using HBase Java API to implement time series, Map Reduce, Filters and counters.

V. HBase schema Design : Group session
   A. Students are presented with real world use cases
   B. Students work in groups to come up with design solutions
   C. Discuss / critique and learn from multiple designs
   D. Labs : implement a scenario in HBase

VI. HBase Internals
   A. Understanding HBase under the hood
   B. Memfile / HFile / WAL
   C. HDFS storage
   D. Compactions
   E. Splits
   F. Bloom Filters
   G. Caches
   H. Diagnostics

VII. HBase installation and configuration
   A. Hardware selection
   B. Install methods
   C. Common configurations
   D. Lab : installing HBase

VIII. HBase eco-system
   A. Developing applications using HBase
   B. Interacting with other Hadoop stack (mapreduce, Pig, Hive)
   C. Frameworks around HBase
   D. Advanced concepts (co-processors)
   E. Labs : writing HBase applications

IX. Monitoring And Best Practices
   A. Monitoring tools and practices
   B. Optimizing HBase
   C. HBase in the cloud
   D. Real world use cases of HBase
   E. Labs : checking HBase vitals