

HBase For Developers

Course Outline:

Section 1: Introduction to Big Data & NoSQL

- Big Data ecosystem
- NoSQL overview
- CAP theorem
- When is NoSQL appropriate
- Columnar storage
- HBase and NoSQL

Section 2 : HBase Intro

- Concepts and Design
- Architecture (HMaster and Region Server)
- Data integrity
- HBase ecosystem
- Lab : Exploring HBase

Section 3 : HBase Data model

- Namespaces, Tables and Regions
- Rows, columns, column families, versions
- HBase Shell and Admin commands
- Lab : HBase Shell

Section 3 : Accessing HBase using Java API

- Introduction to Java API
- Read / Write path
- Time Series data
- Scans
- Map Reduce
- Filters
- Counters
- Co-processors
- Labs (multiple) : Using HBase Java API to implement time series , Map Reduce, Filters and counters.

Section 4 : HBase schema Design : Group session

- students are presented with real world use cases
- students work in groups to come up with design solutions
- discuss / critique and learn from multiple designs
- Labs : implement a scenario in HBase

Section 5 : HBase Internals

- Understanding HBase under the hood
- Memfile / HFile / WAL
- HDFS storage
- Compactions
- Splits
- Bloom Filters

- Caches
- Diagnostics

Section 6 : HBase installation and configuration

- hardware selection
- install methods
- common configurations
- Lab : installing HBase

Section 7 : HBase eco-system

- developing applications using HBase
- interacting with other Hadoop stack (MapReduce, Pig, Hive)
- frameworks around HBase
- advanced concepts (co-processors)
- Labs : writing HBase applications

Section 8 : Monitoring And Best Practices

- monitoring tools and practices
- optimizing HBase
- HBase in the cloud
- real world use cases of HBase
- Labs : checking HBase vitals