Hadoop, Part 4 of 4: HBase and MapReduce

page 1

Meet the expert: Kevin McCarty is a computer professional with over 30 years of experience in the industry as a programmer, project manager, database administrator, architect, and data scientist. He is a Microsoft Certified Trainer with over 25 individual certifications in programming and database technologies and serves as the chapter leader of the Boise SQL Server Users Group. A former Army officer and Eagle Scout, he holds a doctorate in Computer Science and a lifelong love of learning.

Prerequisites: You should have some programming background and some familiarity with a Unix-based operating system. No specific experience with Java programming language or Hadoop is required. As with any such course, the more experience you bring to the course, the more you’ll get out of it. This course moves quickly through a broad range of topics, but it does not require any prior experience with Hadoop. The course does assume that you are well familiarized with how to use the version of Windows that you are running. For example, the course might say simply “Open PuTTY” without explaining how to do that. You should also be able to navigate the folder hierarchy using Windows Explorer.

Runtime: 02:19:36

Course description: Big Data development is a growing field and understanding how technologies such as the MapReduce Combiner and HBase can make data analysis easier is important. Learn how to use the Combiner, look deeply at HBase, a NoSQL tool for managing the storage and retrieval of large and diverse data sets, and create a starter app to use HBase in Java.

Course outline:

MapReduce Combiner
- Introduction
- MapReduce Components
- Combiner
- Why Use a Combiner?
- How the Combiner Works
- Combiner, Reducer Differences
- Summary

MapReduce Combiner Demo
- Introduction
- Demo: Mapper Stub
- Demo: Reducer & Combiner Stubs
- Demo: Main
- Demo: Fill in Methods
- Demo: Run WordCountCombiner
- Summary

HBase Basics
- Introduction
- Relational vs. Non-Relational Databases
- What Is HBase?
- What Does HBase Do?
- Other HBase Features
- The HBase Shell
- Summary

HBase Shell Intro
- Introduction
- Demo: The HBase Shell
- Demo: Ambari
- Demo: Build a Table
- Demo: Add Data
- Summary

HBase Lifecycle Architecture
- Introduction
- HBase Architecture
- HBase Scalability
- HBase Durability
- HBase Consistency
- HBase Input/Output
- Working with Data
- Partitioning a Table
- Serving Regions
- Partitioning Advantage
- Region Server Failover/Load Balancing
- Summary

HBase Lifecycle Services
- Introduction
- The HBase Write Path
- HBase MemStore
- HBase Write-Ahead Log
- HBase on a Single Server
- Scaling HBase
- The Flush Operation
- HBase Cache
- Read Optimization
- Minimizing Disk Seeks
- HBase Compaction
- HBase Load Balancing
- HBase Split Operation
- HBase Scalability
- HBase Limitations
- Summary

Creating an Hbase App
- Introduction
- HBase Archetypes
- What Is a Good HBase Archetype?
- Effective Design
- Effective Design of Queries
- Row-Key Design
- Effective Application Design
- Entities in HBase
- Working with Simple Entities
- Simple Entity Guidelines
- Working with Linked Entities
- Creating a Linked Entity
- Linked Entity Guidelines
- Compound Key
- Compound Time Series
- HBase Techniques for Historical Data
- Considerations for Timestamps
- Summary

Creating an Hbase App Demo
- Introduction
- Demo: Setup
- Demo: HBase Methods
- Demo: Flesh Out Methods
- Demo: Main Method
- Demo: HBase Environment
- Demo: Run HBase App
- Summary