Hadoop, Part 1 of 4: Introduction and HDFS

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: This course assumes that students 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: 01:42:00

Course description: Join Hadoop expert Kevin McCarty as he takes a high level look at Hadoop beginning with its history. Next, McCarty examines a number of key components in the Hadoop ecosystem used for storage, processing, data ingest, and transformation. He will show how Hadoop addresses problems that plague large systems such as failover and redundant storage as well as explore how an organization might incorporate Hadoop into their existing IT framework.

Course outline:

Prerequisites
- Introduction
- Hadoop Sandbox
- PuTTY
- Linux
- Linux BASH
- Linux Commands
- WinSCP
- Notepad++
- Demo: Open Source Tools
- Demo: BASH
- Demo: Other Tools
- Summary

Introduction
- Introduction
- The Lure of Big Data
- What Is Big Data?
- Where Do We Get Big Data?
- Types of Big Data
- Managing Big Data
- The Goal of Big Data
- Companies Using Big Data Today
- The Challenge of Big Data
- How Do We Process Big Data?
- Summary

History
- Introduction
- The Motivation for Hadoop
- Enter Hadoop
- History of Hadoop
- What Hadoop Provides
- Major Users of Hadoop
- The Future
- Summary

Architecture
- Introduction
- Hadoop's Architecture
- HDFS - Name Node
- HDFS - DataNode
- Hadoop Architecture
- Job and Task Tracker
- Hadoop Architecture
- Summary

Ecosystems
- Introduction
- Hadoop Ecosystem
- Zookeeper - Motivation
- What Is ZooKeeper
- Data Ingest
- Apache Flume
- Flume Overview
- Apache Sqoop
- Sqoop Example
- Pig
- Example Pig Script
- Apache Hive
- Example Hive Script
- HBase
- HBase Is...
- Oozie - Job Workflow
- Mahout
- Mahout Use Cases
- Summary

Data Delivery in Hadoop
- Introduction
- Hadoop in the Enterprise
- Demo: Hadoop Sandbox
- Summary

HDFS Basics
- Introduction
- Motivation for HDFS
- Hadoop Distributed Architecture
- HDFS
- HDFS - Nodes
- HDFS NameNode
- HDFS - Secondary NameNode
- HDFS - Standby NameNode
- HDFS - DataNode
- Demo: HDFS
- Demo: Other Commands
- Summary