Hadoop, Part 3 of 4: YARN and NiFi

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: 02:12:29

Course description: Tackle Hadoop tools and services like NiFi, YARN, and Flume as well as the Spark shell, an alternative to MapReduce. Discover why Hadoop has such a large and growing following among sys admins and data scientists. Learning how Hadoop has something for just about everybody to gain and maintain competitive advantage.

Course outline:

YARN Basics
- Introduction
- Scalable Computing
- Scalable Computing - YARN
- YARN
- Hadoop = YARN + HDFS
- Managing Data and Processes
- Limitations of MapReduce v1
- YARN Processing
- Summary

YARN Services
- Introduction
- YARN Daemons
- Resource Manager
- Node Manager
- Job History Server
- Application Master
- Tasks in YARN
- YARN Architecture
- Demo: YARN Tools
- Demo: YARN Parameters
- Summary

Tez and Spark
- Introduction
- Hadoop Structure
- Tez
- What is Apache Spark?
- Benefits of Spark vs. MapReduce
- Spark Framework
- Spark Languages
- Summary

The Spark Shell
- Introduction
- Demo: The Spark Shell
- Demo: Using Scala
- Summary

Flume, Linux, and Nifi
- Introduction
- What Is Flume?
- What Does Flume Do?
- Flume Overview
- Linux ETL
- What is Nifi?
- Working with Nifi
- Nifi Capabilities
- Demo: Linux ETL
- Demo: More Linux Commands
- Summary

Installing Nifi
- Introduction
- Demo: Install Nifi
- Demo: install-nifi.sh
- Demo: Run Nifi
- Summary

Nifi Components
- Introduction
- Nifi Components
- The FlowFile
- The GetFile Component
- The UnpackContent Component
- The ControlRate Component
- The EvaluateXPath Component
- The SplitXML Component
- The UpdateAttribute Component
- The AttributesToJSON Component
- The MergeContent Component
- The PutFile Component
- The RouteOnAttribute Component
- Summary

Nifi Workflow
- Introduction
- Demo: Nifi Workflow
- Demo: Processors
- Demo: Source File
- Summary

Configuring Nifi Workflow
- Introduction
- Demo: GetFile, Unpack Content, and ControlRate
- Demo: Evaluate XPath and Split XML
- Demo: Update Attribute and Parse Records
- Demo: Route on Attribute
- Demo: MergeContent and PutFile
- Demo: Debugging
- Summary