From 0 to 1: The Oozie Orchestration Framework

Modality: On Demand

Duration: 4 Hours

About this course:

Oozie is formidable because it is entirely written in XML, which is hard to debug when things go wrong. However, once you've figured out how to work with it, it's like magic. Complex dependencies, managing a multitude of jobs at different time schedules, managing entire data pipelines are all made easy with Oozie. Oozie allows you to manage Hadoop jobs as well as Java programs, scripts and any other executable with the same basic set up. It manages your dependencies cleanly and logically. Knowing the right configurations parameters which gets the job done, that is the key to mastering Oozie

The average salary of Systems Administrator is **$72,762** *per year.*

Course Objectives:

After completing this course, students will be able to:

- Install and set up Oozie
- Configure time-triggered and data-triggered Workflows
- Configure Workflows to run jobs on Hadoop
- Configure data pipelines using Bundles

Audience:

Engineers, analysts and sysadmins who are interested in big data processing on Hadoop.

Prerequisites:

Students should have basic knowledge of the Hadoop eco-system and should be able to run MapReduce jobs on Hadoop.

Suggested prerequisites courses:

- [HDP Operations: Hadoop Administration 2 (HW HDP OpAdm2)](#)
- [HDP Overview Apache Hadoop Essentials (HW HDP ESS)](#)

Career path:

[Big Data…](#)

Course Outline:

Introduction

- You, This Course and Us

A Brief Overview Of Oozie

- What is Oozie?
- Resource Material
- Oozie architectural components

Oozie Install And Set Up

- Installing Oozie on your machine
- Resource Material

Workflows: A Directed Acyclic Graph Of Tasks

- Running MapReduce on the command line
- Resource Materials
- The lifecycle of a Workflow
- Resource Material
- Running our first Oozie Workflow MapReduce application
- The job.properties file
- The workflow.xml file
- A Shell action Workflow
- Control nodes, Action nodes and Global configurations within Workflows

Coordinators: Managing Workflows

- Running our first Coordinator application
- Resource Material
- A time-triggered Coordinator definition
- Coordinator control mechanisms
- Data availability triggers
- Running a Coordinator which waits for input data
- Coordinator configuration to use data triggers

Bundles: A Collection Of Coordinators For Data Pipelines

- Bundles and why we need them
- Resource Material
- The Bundle kick-off time

Installing Hadoop in a Local Environment

- Hadoop Install Modes
- Hadoop Install Step 1 : Standalone Mode
- Resource Material

- Hadoop Install Step 2 : Pseudo-Distributed Mode

Appendix

- [For Linux/Mac OS Shell Newbies] Path and other Environment Variables

https://www.quickstart.com/   Contact Us: (866) 991-3924