

Advanced Hadoop: Hadoop Troubleshooting

Duration: 24 hours

This course is specifically designed for advanced hadoop users, who already have experience with Hadoop. In this course we will be covering troubleshooting skills and look at production use cases with logs from production systems.

Note: This course will give you access to a real cluster for troubleshooting and root cause analysis.

Module 1: Track YARN job flow in logs.

This module is the core of the course, as it build the skill to troubleshoot. If we do not know, what is unusual in the execution of jobs, then we will never be able to diagnose them.

Understand State Machine for YARN

Trace the execution of stages of a job in logs and break down each step.

Module 2: Common Issues and Data points.

In this we will talk about most common errors and what to look for.

Build skills to quickly identify errors and find data points.

Module 3: Hadoop log Management and processing.

In this module we will look at how to manage cluster logs and arrange them for quick search.

Use Spark to process Hadoop logs and find errors.

Module 4: Troubleshooting Hive

MySql metadata issues and tuning.

Hive thrift server common issues.

Hive query resolution and troubleshooting slowness.

Hive log analysis.

Module 5: Spark Troubleshooting

Common Spark issues and how to look for clues.

Module 6: Troubleshooting Challenge

In this module, a real cluster will be giving to troubleshoot complex issues ranging from HDFS, Hive, YARN, Hbase, Spark, Oozie and Tez.

Cost: US \$700 or INR 42000 /-

Email: trainings@netxillion.com