Advanced Hadoop: Data Lake and Streaming

Duration: 24 hours

Module 1: Using Hadoop as an warehouse. − Data policies.

   Various ingestion and extraction methods.
   Archiving policies

Module 2: Flume Configuration. − Flume Installation and Configuration. − Flume channels and various formats. − Flume twitter use case.

Module 3: Data Ingestion using Sqoop and Hive − Sqoop details.

   MySql imports and exports
   Tuning Sqoop
   Hive details and integration with Sqoop
   Hbase integration

Module 4: Spark installation and Configuration. − Spark Architecture

   Spark standalone mode setup. − Spark in YARN mode. − Spark use cases and programs.

Module 5: Data Pipeline − Understand Kafka architecture and configuration. − Building a Kafka Data pipeline. − Example and common issues.

   Integrating Spark with Kafka.

Module 6: Storm Architecture. − Storm Cluster Setup. − Storm Use Cases. − Adding Storm to the Data Pipeline. − Storm performance tuning.

Module 7: Real Life Project.

Cost: US $350 or equivalent Rs 20,000 INR

Email: trainings@netxillon.com