

BIGDATA TESTING

TRACK -1 / BIGDATA BASIC / DURATION - 30HRS

MODULE 1 – 2 HRS

Introduction of Big Data Testing

- » Understanding what is Bigdata
- » Difference between ETL and big data testing
- » Types of testing in BIGDATA

MODULE 2 - 2 HRS

Introduction to ETL | Data warehouse VS Bigdata testing

- » Introduction to ETL and its importance for data warehousing/ Project Architecture
- » Learning all about OLAP, ETL and OLTP systems
- » Preparation of ETL test data/ different kinds of Sources
- » Writing ETL Testing scenarios & test cases
- » Real time scenarios/ ETL Validation Techniques

MODULE 3 – 3 HRS

Introduction to HDFS and file operations & Unix file system

Part A

- » UNIX Overview
- » Unix commands
- » Unix File operation AWK, SED, GREP, CUT, VI mode

Part B

- » Hadoop Distributed File System (HDFS) Concepts and its Importance
- » HDFS Deep Dive – Architecture, Data Replication, Name Node, Data Node, Data Flow
- » HDFS files system Hands on/ Real time scenarios

MODULE 4 – 5 HRS

Introduction of Hive and Testing Hive & Hadoop SETUP

Part A) SQL

- » Introduction To DDL/DML
- » Introduction To SQL Joins /Set Operators
- » Introduction To Subquery, Aggregate functions
- » Introduction To of query used in ETL validation

BIGDATA TESTING

Part B) Hive and Pig

- » Hive Schema and Data Storage
- » Comparing Hive to Traditional Databases
- » Hive vs. Pig
- » Hive Use Cases
- » Interacting with Hive
- » Hive Databases and Tables
- » HiveQL Syntax/Data Types
- » Joining Data Sets/ Common Built-in Functions
- » Hands-On Exercise: Running Hive Queries on the Shell, Scripts, and Hue
- » Understanding the test cases of Hive & Testing Data loaded in hive/ Real time scenarios

MODULE 5 – 4 HRS

Map Reduce Introduction and Testing Mapr/Spark JOB

- » Introduction OF MAP and REDUCE
- » Understanding MAP and Reduce Algo
- » How TO run MAP and REDUCE job
- » Validating and testing the MAP Reduce job

MODULE 6 – 2 HRS

Introduction of HBASE

- » Introduction to HBASE
- » Introduction to HBASE Architect
- » Introduction to HBASE CRUD operation
- » HBase commands

MODULE 7 – 2 HRS

SQOOP Testing

- » what is Sqoop?
- » why Sqoop?
- » Importing and exporting data from Sqoop to HDFS
- » Sqoop test case and Sqoop testing

BIGDATA TESTING

MODULE 9 – 10HRS

Hadoop Testing Life Cycle and project work | Hadoop Automation Scope

- » Understanding the Requirement
- » preparation of the Testing Strategy
- » Test Cases
- » Test Data
- » Test Execution
- » Defect Reporting, Defect Retest, Daily Status report delivery, Test completion.
- » ETL testing at every stage (HDFS, HIVE, HBASE) while loading the input (logs/files/records etc.) using Sqoop which includes but not limited to data verification, Reconciliation.
- » Big Data Pipeline Testing
- » Validation of Upstream Down Stream impacts
- » Data Quality Checks on Bigdata
- » scope of Hadoop Automation
- » shell scripts for file testing in Hadoop
- » Hand on session