BIGDATA HADOOP TESTING COURSE CONTENTS

Java Basics

➢ What is JAVA?
➢ How to set path?
➢ JDK, JRE and JVM
➢ Data Types and Operators
➢ Installing Java on UNIX
➢ Running Java Program in UNIX

Oops Concepts

➢ Objects and Classes
➢ Constructors
➢ Method Over-loading
➢ Method Overriding
➢ Static Keyword
➢ This Keyword
➢ Super and final Keyword

Oops Concepts -Part 2

➢ Abstract and Interface
➢ Package
➢ Abstract Class
➢ Encapsulation
➢ Polymorphism
➢ Overloading vs. Overriding

Collections API

➢ Topics
➢ List
➢ Set
➢ Map
➢ Need of collection API

UNIX

➢ What is Unix and Architecture
➢ Unix Creating file
➢ Listing Files in UNIX
➢ Hidden Files in UNIX
➢ Displaying Contents of file
➢ Copying Files
➢ Renaming file in UNIX
➢ Deleting Files in UNIX

Directories and File Permission

- ➤ Home Directory
- ➤ Creating Directory and parent Directory
- ➤ Listing Directory
- ➤ Removing Directory and Changing Directory
- ➤ CHMOD command and File Access Modes
- ➤ Changing Groups and Owners

Additional Commands

- ➤ grep and sort
- ➤ Consuming output of one command to another
- ➤ Starting a Process in foreground and background
- ➤ Listing Running Process
- ➤ Stopping a Process
- ➤ Top Command

Data-ware-housing

- ➤ Data Warehouse Architecture
- ➤ OLTP v OLAP
- ➤ What is Data Warehouse
- ➤ What is Enterprise Data Warehouse
- ➤ What is Data Marts

Component of Data-Ware-house

- ➤ Source System and target Systems
- ➤ Staging Area
- ➤ Drill up and Drill down
- ➤ Facts and Dimensions
- ➤ Slowly Changing Dimensions

Hadoop Testing

- ➤ Introduction to Big Data
- ➤ Data and Growth
- ➤ What is Big Data?
- ➤ What are three V's in Big Data
- ➤ Big Data Storage and Processing challenges
- ➤ Testers Role in Big Data Project?
- ➤ Pre-requisites for Hadoop Testers?

Introducing Hadoop

- ➤ What is Hadoop and Why Hadoop?
- ➤ Hadoop Eco-System, how solutions fit in?
- ➤ What Tester should know in Eco-System?
- ➤ What are Hadoop Core-Components?
- ➤ How to Start and Stop the hadoop dameons?
- ➤ Hadoop Versions, Flavor and what testers need to Know?
- ➤ Hadoop Cluster Modes

Detailing Hadoop HDFS

- HDFS Architecture
- What is Namenode , Data Node and Secondary Namenode
- How to Browse the HDFS file System?
- Validating the logs for the HDFS dameons?
- Generating Test Data in HDFS
- Basic operations like mkdir, chmod, cat in HDFS.
- Storing Data in Hadoop
- Testing and Validating the daemons

Detailing Hadoop Map-Reduce

- What is Job Tracker and Task Tracker?
- Explaining a Weather Data Set Program?
- Compiling and Verifying the Map-Reduce Program
- Difference between traditional RDBMS and Map Reduce

Hadoop Configuration Files and Properties

- Location of Configuration files?
- What are different configuration files in Hadoop
- Explaining the default parameters in the configuration files.
- How to overwrite the default parameters?
- Testing Impact of Configuration

Hadoop Data loading: SQOOP: Part 1

- What is Hadoop Data loading
- What is Sqoop
- Sqoop Architecture
- Sqoop 1 vs. Sqoop 2
- Display Sqoop Version
- Creating Tables in MySQL and adding Data
- Sqoop Commands
- List Database
- List tables
- Importing tables with map-reduce task
- Testing and Validating the output
- Importing tables with target directory
- Testing and Validating the output
- Importing tables with options file
- Testing and validating the output
- Importing all rows of a table but specific columns
- Testing and validating the output
- Import and filter rows using where clause
- Testing and Validating the output

Hadoop Data loading: SQOOP: Part 2

- Sqoop Import Architecture

- ➢ Code Generation
- ➢ Testing for Data Quality and Data Consistency for Import
- ➢ Text and Binary File Formats in Scoop
- ➢ Testing and Validating the output
- ➢ Importing Large objects like CLOB
- ➢ CLOB through SQL Insert
- ➢ CLOB through Prepared Statement
- ➢ CLOB through Set Stream Method
- ➢ Testing and Validating the large Objects

Hadoop Data loading: SQOOP: Part 3

- ➢ Sqoop Import into Hive
- ➢ Dealing with the delta records?
- ➢ last Modified and Append Modes
- ➢ Test and Validating for Incremental Load
- ➢ Export Architecture
- ➢ Exporting Data from Hive to Sqoop
- ➢ Export HDFS file to RDBMS
- ➢ Sqoop Commands Eval
- ➢ Sqoop Job

Hadoop Data loading: Flume Part 1

- ➢ What is Flume?
- ➢ Flume Architecture
- ➢ Understanding the Source and Sink
- ➢ Understanding the Memory Channel
- ➢ Flume first Example
- ➢ Working with flume Configuration file

Hadoop Data loading: Flume: Part 2

- ➢ Flume Second Example for HDFS sink
- ➢ Testing and Validating the Output of Flume Operations
- ➢ Overview Additional Source and Sink Support
- ➢ Validating the Configuration and Error from output
- ➢ Working with Twitter Source?
- ➢ Configuration for Twitter Source

Pig -Part 1

- ➢ Why Pig was created?
- ➢ Why Pig when Map-Reduce is there?
- ➢ Pig Components.
- ➢ Pig Execution Modes
- ➢ Pig Example: Analysis for Weather Data Set.
- ➢ Testing Pig Output with Sampling.

Pig -Part 2

- ➢ Pig for Data Cleaning

- Data Types and Models
- Data Type like Bag, Tuple ,Map
- Use Cases in Healthcare
- Execution Pig Programs
- Storing the Analysis Results from PiG
- Join Operations like Left , Right , Outer Join
- Testing the output with Sampling
- Map Side Joins for Pig
- Testing the Output with MapSide Joins
- Co-Group , Sort , Union and Cross Joins Examples in Pig
- Illustrate , Explain and Describe Commands

Pig -Part 3

- Relation Operator
- Advance Concepts like Parallelism and Streaming
- Functions is Pig
- AVG
- COUNT
- COUNT_STAR
- CONCAT
- DIFF
- ISEmpty
- MAX
- MIN
- PluckTuple
- SIZE
- SUBTRACT
- SUM
- TOKENIZE
- Load Functions
- PIGSTORAGE
- JSONLOADER
- JSONSTORAGE
- BINSTORAGE
- HANDLING COMPRESSION
- Map,Tuple and Bag
- TOTUPLE
- TOBAG
- TOMAP
- TOP
- Working with UDF in pig
- Testing and Validating the UDF in Pig

Hive - Part 1

- Hive Introduction
- Comparison with RDBMS Databases
- Working with Hive Shell
- Hive Data Types

- ➢ Hive Create Database and Drop Database
- ➢ Hive Create Table and Alter Table
- ➢ Hive Partitions

Hive -Part 2

- ➢ Hive Functions and Operators
- ➢ Creating View and Indexes in Hive
- ➢ Hive QL: Select, Group, and Joins, Where clauses?
- ➢ Analyzing Weather Data Set
- ➢ Testing the Hive Analyzed results against Sample Output

Hive - Part 3

- ➢ Hive and Metastore Derby
- ➢ Changing the Hive Metastore to MySQL
- ➢ Hive Managed vs. External Tables
- ➢ Hive Storage with JSON
- ➢ Hive Storage and SerDE
- ➢ Testing the Output from Hive
- ➢ 3 Critical Things Testers should be aware of?

Hive Part 4

- ➢ Hive UDF's
- ➢ Configuring and Testing the UDF's
- ➢ Hive Configurations
- ➢ Hive Web UI
- ➢ Hive Property Precedence
- ➢ Running Batch commands in Hive
- ➢ Why Tester need to be aware of Configurations?
- ➢ What Testers need to look out in Case Expected outcome is not met

Hive vs. Pig

- ➢ Hive and Pig Comparison
- ➢ Using Hive Schema for Pig Load
- ➢ What Tester need to know?

Testing Big Data

- ➢ Key Challenges in Testing Big Data
- ➢ Identifying the Testing Gates and Entry Points
- ➢ Finding Bad Data and doing data Quality Checks
- ➢ Functional and Regression Testing
- ➢ Big Data Testing Stages and Testing Task