Javahomecloud
Master Your Skills

Hadoop Course

Introduction to Scala:

1.Variable declaration
2.Working with Option
3.Scala data types
4.Type inference
5.Scala shell
6.Lazy variables
7.Control structures
8.Pattern matching in Scala
9.Defining functions in Scala
10.Collections in Scala

OOPS in Scala

1.Classes
  i. Working with mutator and accessor methods
  ii. Overriding default mutator and accessor methods
2.Singletons
3.Companion object
4.Access specifiers
5.Constructors
  i.Primary constructors
  ii.Auxiliary Constructors
6.Case classes
7.Inheritance
  i.Fields overriding
  ii.Methods overriding
  iii.Calling super class constructors from sub class Functional programming in Scala
1.Understanding functional programming concepts
2.Higher order function Scala environment setup
  1.Java setup
  2.Scala setup

Javahomecloud
Master Your Skills

Spark

Introduction to Spark

1. Evolution of Distributed systems & Challenges faced
2.Need of new generation
3.Hardware/software evolution in last decade 4.History of Spark
5.Features of Spark
6.Spark Architecture
7.Spark installation
8.Spark shell
9.Creating Spark context
10. Introduction to RDD
    i. Transformations in RDD
    ii. Actions in RDD
    iii.Element-wise operations
    iv.Working with Key/Value pairs Persistence(Caching)

Spark API Examples

1.Loading data to RDD
2.Item wise count
3.Find a specific item
4.Serialization
5.Discount
  i.Per sale
  ii.Sumbased
6.Error Handling
  i.Counting (Accumulators)
  ii.Malformed record saving
7.Joins
  i.Shuffle based
  ii.Broadcast based(Broadcast variable) Anatomy of RDD
  1.Partitions
  2.Manipulating Partitions
  3.Hash Partition
  4.Custom practitioner
  5.RDD dependencies
  6.Run Job API
  7.Cache Example
  8.Spark on YARN

Anatomy of job execution

Spark SQL

1 Introduction
2.SQL Context
3.Spark SQL vs Hive
4.Data Source
  i.Reading csv Data in RDD
  ii.CSV data
  iii.Json Data
  iv.Mix sources
5.Creating DataFrame
   i.Using case classes
   ii.Programmatic Schema
6. Querying with SQL dialect
    i.Basic Queries
    ii.Joins
7.Dataframe DSL
    i.Basic Queries
    ii.Joins
8. Interaction with databases
   i.MySQL
   ii.Mongodb
   iii.ParquetFile
9.Spark and Hive integration 10.Cache in Spark SQL

Spark Streaming

1.Batch vs Streaming
2.Architecture and Abstraction
3.DStreams, DStreams vs RDD
4.Transformations
5.Input Streams(Socket, HDFS, Twitter, Kafka)
6.Checkpointing, Persist and Caching
7.Batch and Window Sizes
8.Level of Parallelism

Javahomecloud
Master Your Skills

HADOOP Overview

HDFS

1.What is Hadoop
2.History of Hadoop
3.Problems with Traditional Large-Scale Systems and Need for Hadoop
4.Understanding Hadoop Architecture
5.Fundamental of HDFS (Blocks,Namenode, Datanode,Secondary Namenode)
6.Rack Awareness
7.Read/Writefrom HDFS
8.Drawbacks of HDFS1.x
10.HDFS commands

Map Reduce

1 Understanding Mapreduce
2 Job Tracker and Task Tracker
3 Architecture of Mapreduce
4 Data Flow of Mapreduce
5 Map Function & Reduce Function
6 How Mapreduce Works
7 Anatomy of Mapreduce Job
8 Understand Difference Between Block and Input Split
9 Map Side VS Reduce Side

Hive:

1 Introduction to Apache Hive
2 Architecture of Hive
3 Hive Metastore
4 Hive data types
5 Hive – Internal vs External tables
6 Hive – Data Partitioning
7 Buckets & Sampling
8 Indexes & Views
9 Developing hive scripts
10. Parameter Substitution

11. Creating UDF

Overview on liveproject:

Trainer will discuss live projects that cover architecture and low level implementation to give exposure on real-time systems.

Assignments on MR, HIVE, SCALA,Spark Core,SQL and Streaming

Others:

1.Helping in resume preparation

2.Sharing frequently asked interview question