Quandatics

Big Data, IoT, Smart Analytics

Big Data Analytics – Hadoop

4-Days Practical Training

Why Big Data and Hadoop?

The use of Big Data is becoming a crucial way for leading companies to outperform their peers. It enables us to tap into large information flow where data about products and services, buyers and suppliers, consumer preferences and intents can be captured and analyzed. Therefore nurturing Big Data Analytics expertise plays an essential role in developing new products or services.

Who should attend this Hadoop training?

This training is ideal for professionals who wish to implement Big Data Analytics using Hadoop framework.

What you will learn on Hadoop?

The course equips participants to work on the Hadoop environment with ease and learn vital components such as Apache Spark, Flume, Hive, Pig, MapReduce and other advanced concepts like HBase, ZooKeeper and Sqoop.

This course consists of:

4 days classroom training
20 hours of real-time industry-based projects
15 hours of lab exercises with proprietary VM
3 hours of doubt clarification session
2 projects with unique data sets included
Industry specific projects on the retail and telecom sectors
Java Essentials for Hadoop included
Hadoop installation procedure included
Hadoop deployment and maintenance tips
Packed with latest & advanced modules like Yarn, Spark, Flume, Oozie, Mahout, and Chukwa
Tips and techniques to clear CCDH certification exam

Quandatics

Big Data, IoT, Smart Analytics

www.quandatics.com

Course Outline:

Day 1

Introduction to Big Data

- What is Big Data?
- Examples of Big Data
- Use cases of Big Data
- What is Hadoop
- History of Hadoop
- Problems with Traditional Large-Scale Systems and Need for Hadoop
- Where Hadoop is being used
- Understanding distributed systems and Hadoop
- RDBMS and Hadoop

Software Installation

- Pre-requisites
- Understanding Hadoop configuration files
- Setup single node Hadoop cluster
- The Command-Line Interface

Hadoop Ecosystem

- HDFS
- MapReduce
- Hive-Introduction
- Sqoop-Introduction
- Pig-Introduction
- HBase-Introduction
- Flume- Introduction
- Spark-Introduction

Understanding Hadoop Distributed File System

- Understanding HDFS Architecture
- Hadoop Components- HDFS, MapReduce
- Overview of Hadoop Processes
- Overview of Hadoop Distributed File System
- Name Nodes and Data Nodes
- The building blocks of Hadoop
- Running Hadoop
- Web-based cluster UI-Name Node UI, Map Reduce UI
- Hands-On Exercise: Using HDFS commands

Exercise / Hands-On: HDFS

- HDFS commands

Understanding MapReduce

- How MapReduce Works
- Data flow in MapReduce
- Map operation
- Reduce operation
- MapReduce Driver Class
- Running your first program
- Split, Record Reader (RR), Sorter, Shuffler and Partitioner
- Combiner in-depth
- Distributed Cache
- Writing MapReduce Drivers, Mappers and Reducers in Java with eclipse
- Code Walkthrough
- Error handling in MapReduce
- MapReduce Job execution flow In-depth

Day 2

Exercise / Hands-On: MapReduce

- First MapReduce program with Basic Word count
- Calculate Sum for structured data
- Handling unstructured data
- Processing fixed-length values

Hive

- Introduction to Apache Hive
- Hive architecture
- Installing Hive
- Getting data into Hive
- Hive-HQL
- Query execution
- Working with WHERE Clause
- Partitions in Hive (Static and Dynamic)
- Performing JOIN in Hive (Map and Reducer Side joins)
- Compression in Hive (ORC)
- Executing Hive queries in real time
- User defined functions (UDFs)
- Programming practices and projects in Hive

Exercise / Hands-On: Hive

- Loading data, sample query with where and joins
- Partitions and UDFs

Quandatics

Big Data, IoT, Smart Analytics

www.quandatics.com

Sqoop

- Installing Sqoop
- Configure Sqoop
- Import RDBMS data to Hive using Sqoop
- Export from Hive to RDBMS using Sqoop
- Incremental load

Exercise / Hands-On: Sqoop

- Import data from RDBMS to HDFS/ Hive
- Export data from HDFS/Hive to RDBMS

Day 3

Pig

- Introduction to Apache Pig
- Installing Pig
- Pig architecture
- Pig Latin – Reading and writing data using Pig
- Parameter passing with Pig
- UDFs in Pig
- Managing multiple Pig scripts in real-time case
- Executing Pig scripts in real-time projects

Exercise / Hands-On: Pig

- Load data, execute data processing statements
- Programming with Pig UDFs

HBase

- What is HBase?
- Installing HBase
- HBase Architecture
- Command line interface Exercise
- Programming In HBase
- MapReduce Programs in HBase
- Filters in HBase
- Load and Managing large volumes data sets with HBase
- HBase -Hive integration
- Real time projects in HBase (use cases MapReduce)

Exercise / Hands-On: HBase

- HBase command line interface
- Loading data into HBase with MapReduce

Day 4

Spark

- What is Spark and why use Spark?
- Installing Spark
- Spark components
- RDD introduction
- Spark interface with Scala and Java
- Command line interface of Spark
- Programming in spark

Flume

- Introduction to Flume and why?
- Installing Flume
- Flume components (Agent, Source, Channel, Sink, Receiver)
- Flume configuration with Source to write date into file (Local and HDFS)
- Multiple sources with Flume
- Running Flume Agent
- Running Receivers and Test with sample data

For any enquiries, please contact:
Quandatics
e: contact@quandatics.com
m: +6 016 236 2220
(Jaden Teo)