What is hadoop?

Hadoop is a platform written in java where we can able to process large amount of data. Hadoop eco system has lots of tools which make processing the bigdata made easy.

Let's learn how to do that end to end..!!

Objective:

Over the past years, Hadoop & Spark has seen enormous industry adoption and facing lack of skills in the market. To help bridge the gap we have designed this course with industry expectations with real time examples. This is course will help you understand variety of big data application development options and let you develop your own and Performance tune the same.

This course is for,

- Professionals who wants to learn & develop Hadoop & Spark applications.
- Professionals who wants to do certification (Hortonworks : HDPCD, HDPCDSPARK)(Cloudera: CCA175, CCA159).
- And those are is interested to learn about latest technology for their career improvement.

Course Structure:

- This course is designed with 50% theory and 50% Hands on.
- You will be given real time POC to solve and learn.

Hadoop – Project (English) – Click Here

Hadoop - Intro Session (Tamil) – Click Here

SPARK - Intro Session (Tamil) – Click Here

After this class you will be able to,

- Have in-depth knowledge about hadoop.
- Have hands-on experience on hadoop.
- Complete a project on hadoop independently.
- Know how to switch career to hadoop from any other technology.
- Develop your own spark application.
- Understanding different components of spark.
- Performance tune a spark application.
- Prepare and complete Horton works spark developer certification (With min 1 month of practice)
- Build data pipeline using spark API's and Dataframes.
- Analyze Spark jobs using the UI's and logs.
- Create Streaming jobs and run on YARN cluster.

Course Overview:

- Introduction to Hadoop
- Hadoop Architecture In-depth travel.
- Map Reduce 1.0 & YARN
- Pig & Hive
- Sqoop & Flume
- Hbase, oozie & Zookeeper
- Welcome to Spark.
- Programming with RDD.
- SparkSQL & DataFrames.
- Spark Job Execution.
- Cluster Architecture for Spark.
- Introduction to Kafka.
- Introduction to Spark Streaming.

Module 1: Introduction to Hadoop World:

- ✓ Dataaaaaaa.....Bigdata..!
- ✓ What is bigdata? 3 + 1 V's.
- ✓ What is Hadoop , why hadoop & Its history.
- ✓ Hadoop Eco System an overview. (HDFS,MAPREDUCE,SQOOP,FLUME,PIG, HIVE,OOZIE,HBASE..etc)
- ✓ Current Requirements and Future possibilities in Hadoop.
- ✓ RDBMS vs Hadoop
- ✓ Wait..Finally what hadoop is not?
- ✓ Do we need java to learn hadoop?
- ✓ Hadoop installation

Module 2: Hadoop Architecture In-depth travel:

- ✓ HDFS - An introduction.
- ✓ How data is stored in hdfs? (Travel of a byte).
- ✓ Hadoop Daemons:
  - Name node.
  - Data node.
  - Job Tracker.
  - Task tracker.
- ✓ Fault tolerance in hadoop.
- ✓ HA mode in HDFS.
- ✓ How files are handled in projects (sample Project Scenario Execution)

Module 3: Map Reduce 1.0 & YARN.

- ✓ Mapreduce history.
- ✓ How Map Reduce is being used in Projects.
- ✓ Mapreduce architecture,Key-Value pair.
- ✓ YARN 2.0 architecture.
- ✓ Java Implementation of map reduce. (Sample POC)
- ✓ Mapper, Reducer, Combiner Different combination.

Module 4: Pig & Hive.

- ✓ Hive introduction.
- ✓ Hive data model.
- ✓ Hive implementation of sample project.
- ✓ Pig Introduction.
- ✓ Pig Data structure.
- ✓ Pig Implementation on sample project.
- ✓ How pig & hive is used in real time project?
- ✓ Module 4 assignment.

Module 5: Sqoop & Flume.

- ✓ Flume introduction.
- ✓ Flume configuration.
- ✓ Flume sample Project.
- ✓ Sqoop Introduction.
- ✓ Sqoop configuration.
- ✓ Sqoop Sample project.

Module 6: Hbase, oozie & Zookeeper

- ✓ oozie introduction.
- ✓ oozie Overview and configuration.
- ✓ zookeeper overview.
- ✓ HBASE Introduction.
- ✓ HBASE Overview.
- ✓ SPARK Over view

SPARK

Intro Session(Tamil) – Click Here

Module 1: Welcome to Spark:

- ✓ Welcome to the world of Spark.
- ✓ Bye Bye Hadoop? (Hadoop Vs Spark).
- ✓ Spark Components:
  - Spark Core
  - Spark SQL
  - Graphx
  - Mlib
- ✓ Spark Use cases in real time.

Hands on:

- Installing and configuring spark in your machine.
- Running a sample program in spark.
- Executing a spark use case.

Module 2: Programming with RDD:

- ✓ What is RDD?
- ✓ Why RDD?
- ✓ How RDD gets executed in a spark application.
- ✓ Transformations in RDD.
- ✓ Actions in RDD.
- ✓ RDD Programming API's.

Hands On:

- Creating RDD from a Data file.
- Applying transformations & actions in RDD.
- Interactive queries using RDD.

Module 3: Spark SQL/DataFrames.

- ✓ SparkSQL/Dataframe Uses.
- ✓ DataFrame / SQL API's
- ✓ Spark & Hive Integration.
- ✓ Catalyst query optimization.

Hands on:

- Create dataframe from a file.
- Create dataframe from a table.
- Caching and reusing dataframes.
- Query with dataframes API and SQL.

Module 4: Spark Execution & Optimization.

- ✓ Jobs Stages & tasks.
- ✓ Partitions and Shuffles.
- ✓ Data locality.
- ✓ Job Performance (tuning).

Hands on:

- Visualizing DAG execution.
- Measuring memory usage.
- Understanding performance.

Module 5: Introduction to Kafka.

- ✓ Introduction to Kafka.
- ✓ Kafka architecture.
- ✓ Producers,Consumers in Kafka.
- ✓ Working with kafka.

Hands on:

- Installing & configuring kafka.
- Producing and consuming messages.

Module 6: Spark Streaming.

- ✓ Introduction to Spark Streaming.
- ✓ DSTREAM API's and Stateful Streams.
- ✓ Realiablity and fault recovery.

Hands on:

- Creating DStream from source.
- Integration of Kafka and Spark streaming.
- Developing a kafka-spark application.
- Viewing Stream jobs in WebUI.

For More details :

Mail: Arumugam@tamilboomi.com,

arumugamsip@gmail.com

Whatsapp: +91 9619663272

Visit to: www.tamilboomi.com

For cloudera VM and Free Bigdata Startup kit: Startup kit link Click here.

Happy Learning..!