HADOOP ADMIN:

Hadoop is an open source Java-based programming framework that supports the processing of large data sets in a distributed computing environment. To process and store the data, it utilizes inexpensive, industry-standard servers. The key features of Hadoop are Cost effective system, Scalability, Parallel processing of distributed data, Data locality optimization, Automatic failover management and supports large clusters of nodes.

Hadoop admin is responsible for capacity planning and estimating the requirements for decreasing or increasing the capacity of the Hadoop cluster. Hadoop admin is also responsible for deciding the size of the Hadoop cluster based on the data to be stored in HDFS. The Hadoop frameworks comprised of two main core components i.e. HDFS and MapReduce framework.

Key Features:

- License Free
- Open Source
- Meant for Big Data Analytics
- Ecosystem Approach
- Shared Nothing Architecture
- Distributed File System
- Commodity Hardware
- Horizontal Scalability

- Distributors

- Cloudera

- Horton works

Course Content:

Hadoop Architecture:

Introduction to

- Parallel Computer vs. Distributed Computing

- How to install Hadoop on your system

- How to install Hadoop cluster on multiple

- Hadoop Daemons introduction: NameNode, DataNode, JobTracker, TaskTracker

- Exploring HDFS (Hadoop Distributed File System) Exploring the HDFS Apache Web UI

- NameNode architecture (EditLog, FsImage, location of replicas) Secondary NameNode architecture

- DataNode architecture

MapReduce Architecture:

- Exploring JobTracker/TaskTracker

- How a client submits a Map-Reduce job

- Exploring Mapper/Reducer/Combiner

- Shuffle: Sort & Partition

- Input/output formats

- Job Scheduling (FIFO, Fair Scheduler, Capacity Scheduler) Exploring the Apache

  MapReduce Web UI

Hadoop Developer Tasks:

- Writing a map-reduce programme

- Reading and writing data using

- Java Hadoop Eclipse integration

- Mapper in details

- Reducer in details

- Using Combiners

- Reducing Intermediate Data with Combiners

- Writing Partitioners for Better Load

- Balancing Sorting in HDFS

- Searching in HDFS

- Indexing in HDFS

- Hands-On Exercise

Hadoop Administrative Tasks:

- Routine Administrative Procedures

- Understanding dfsadmin and mradmin Block Scanner, Balancer

- Health Check & Safe mode

- DataNode commissioning/decommissioning

- Monitoring and Debugging on a production

- cluster NameNode Backup and Recovery

- ACL (Access control list) Upgrading Hadoop

HBase Architecture:

- Introduction to HBase

- HBase vs. RDBMS

- Exploring HBase Master & region server

- Column Families and Regions

- Basic HBase shell commands.

Hive Architecture:

- Introduction to Hive

- HBase vs Hive

- Installation of Hive

- HQL (Hive query language)

- Basic Hive commands

Pig Architecture:

- Introduction to Pig

- Installation of Pig on your system

- Basic Pig commands

- Hands-On Exercise

Sqoop Architecture:

- Introduction to Sqoop

- Installation of Sqoop on your system

- Import/Export data from RDBMS to HDFS

- Import/Export data from RDBMS to HBase

- Import/Export data from RDBMS to Hive

- Hands-On Exercise

Mini Project / POC (Proof of Concept):

- Facebook-Hive POC

- Usages of Hadoop/Hive @ Facebook

- Static & dynamic partitioning

- UDF (User defined functions)

Our learning methods include:

- Comprehensive course selection of Instructor-Led Training

- Logistical convenience and interactive classroom experience of Online Training

- Flexible pacing and instructor-guided support of Mentored Learning

- Self-paced convenience of Online ANYTIME

In addition:

- Interview preparation with mock interview drills

- Effective resume building

- Process of applying jobs at the right places

Reach us:

Call   : +1 720 738 4411

Email ID: info@procareer.io

Website: https://www.procareer.io/