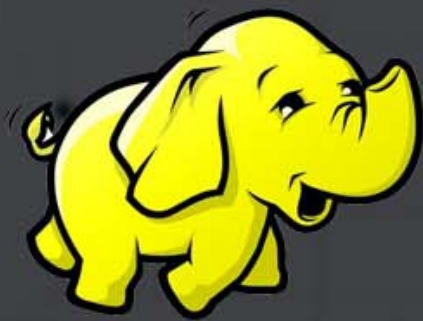# Hadoop Developer

BIG DATA
HADOOP DEVELOPER

SevenMentor
PVT.LTD

Introduction to Hadoop

1. RDBMS Vs Hadoop
2. Difference in between Mysql and Hadoop
3. Why Hadoop is better that Mysql??
4. 5 V's of big data

Introduction to Java

1. Basics of Java required for Hadoop
2. OOPS - Class, Object and Interface
3. Inheritance and types of inheritance
4. Method overriding and overloading
5. Exception Handling

Introduction to SQL

1. Basics of Sql required for Hadoop
2. DML,DDL statements

Introduction to HDFS (Storage) & Understanding cluster environment

1. NameNode and DataNodes
2. HDFS has a master/slave architecture
3. Overview of Hadoop Daemons
4. Hadoop FS and Processing Environment's UIs
5. Block Replication
6. How to read and write files
7. Hadoop FS shell commands
8. MR1.x vs 2.x

SevenMentor
PVT.LTD

Understanding Map-Reduce Basics

1. The introduction of MapReduce.
2. MapReduce Architecture
3. Data flow in MapReduce
4. How MapReduce Works
5. Writing and Executing the Basic MapReduce Program using Java

TOOLS

A. SQOOP:
   Sqoop architecture
   Sqoop commands
   Sqoop practical implementation
   Importing data to HDFS
   Importing data to Hive
   Exporting data to RDBMS
   Sqoop show tables, databases, eval
   Sqoop jobs

B. HIVE:
   Hive Architecture
   Hive Query Language (HQL)
   Managed and External Tables
   Partitioning & Bucketing
   UDF in hive
   Working with different file formats
   JDBC , ODBC connection to Hive
   Hands on Multiple Real Time datasets.

SevenMentor
PVT.LTD

C. PIG:

Pig Latin (Scripting language for Pig)
Schema and Schema-less data in Pig
Structured , Semi-Structure data processing in Pig
Built-in functions
UDF in pig

D. HBASE:

Introduction to HBASE
Basic Configurations of HBASE
Fundamentals of HBase
What is NoSQL?
HBase Data Model
Table and Row
Column Family and Column Qualifier
Cell and its Versioning
Get commands
Scan -Put commands
Namespace and drop tables
Hive table with hbase data

E. Oozie

Introduction to Oozie
Designing workflow jobs
Job scheduling using Oozie
Time based job scheduling
Oozie Conf files

F. Apache Flume:

Introduction to flume
Source,Sink and Channel
Fetching twitter data

G. Introduction to Spark:
Overview of Spark, Scala and its features