

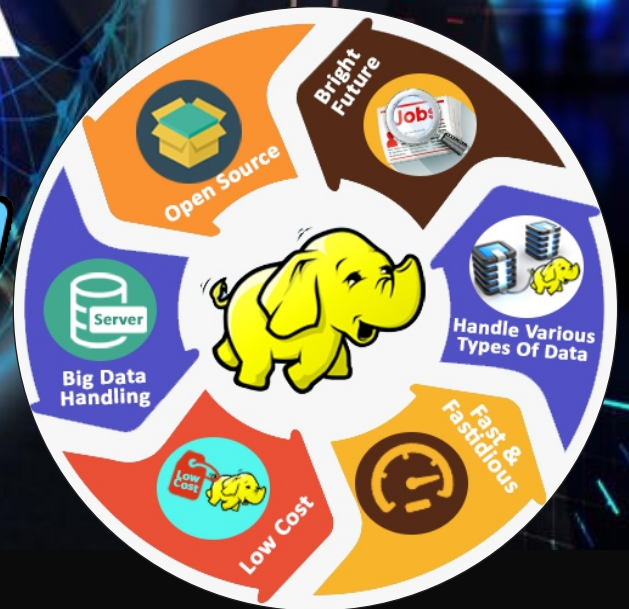


# SKY VISION COLLEGE OF DIGITAL TECHNOLOGY AND MANAGEMENT

Contact: +232 73797979/+232 32797979

**DURATION  
216  
Hours**

## BIG DATA



### WHY STUDENT CHOOSE US!

FACULTY WITH  
INDUSTRY  
EXPERIENCE

100% PRACTICAL

FREE WI-FI  
INTERNET FACILITY  
FOR STUDENTS

DIGITAL PROJECTORS  
FOR LIVE PRESENTATION  
& TRAINING

HIGHLY CONFIGURED  
MACHINE-ROOM FOR  
PRACTICAL

100 % PLACEMENT  
ASSISTANCE

RECORDED  
SESSION

INTERNATIONAL  
CERTIFICATION

- Get trained from Sky Vision College Certified Trainers having rich experience in Finishing School Programs Today!
- Experience state-of-the-art infrastructure at Sky Vision College Computer Labs.



Your Future, Our Vision

# Hadoop Programme Details

## 1. Java Fundamentals

- 1.1 Basic java concepts
- 1.2 Multi-threading
- 1.3 File I/O –Java. IO
- 1.4 Collections –Java.Util.\*, Java.Math, Java.Lang
- 1.5 Java Generics
- 1.6 Java Serialization
- 1.7 Java Database Connectivity –JDBC
- 1.8 Java Common Design Patterns
- 1.9 Java Open Source Frameworks (Spring, Apache Maven, Logging, etc...)
- 1.10 Java Apache Hadoop Frameworks (Hadoop Common, Map Reduce etc.)
- 1.11 Understand Web Servers & Application Servers - JBoss Application server, Apache Tomcat server
- 1.12 Java Unit testing Frameworks (JUnit / TestNG)
- 1.13 Eclipse IDE – Java Development.
- 1.14 Version Control – GIT, SVN, etc.
- 1.15 Java Continuous Integration frameworks – Hudson, Jenkins, etc.
- 1.16 Handling XML and XSD using Java frameworks
- 1.17 Java XML Parsers frameworks – DOM and SAX
- 1.18 Java Web services concepts – SOA, SOAP, XML, JAXB,
- 1.19 SOAP Web services
- 1.20 REST web services

## 2. Hadoop Fundamentals

- 2.1 What is Big Data? Why Big Data?
- 2.2 Hadoop Architecture & Components
- 2.3 Hadoop Storage & File Formats (ASCII, Avro, Parquet, RC4, JSON, EBCDIC etc.)
- 2.4 Hadoop Processing – Map Reduce, Spark Frameworks

## 3. HDFS

- 3.1 HDFS Basics
- 3.2 File Storage
- 3.3 Fault Tolerance

## 4. Map Reduce

- 4.1 What Is MapReduce?
- 4.2 Basic MapReduce Concepts
- 4.3 Concepts of Mappers, Reducers, Combiners and Partitioning
- 4.4 Inputs and Output formats to MR Program
- 4.5 Error Handling and creating UDFs for MR

## 5. Spark

- 5.1 What Is Spark?
- 5.2 Basic Spark Concepts
- 5.3 How Spark differs from Map Reduce?
- 5.4 Working with RDD's
- 5.5 Parallel Programming with Spark
- 5.6 Spark Streaming

## 6. Hive

- 6.1 What is Hive, why we need it and its importance in DWH?
- 6.2 How Hive is different from Traditional RDBMS
- 6.3 Modeling in Hive, creating Hive structures and data load process.
- 6.4 Concepts of Partitioning, Bucketing, Blocks, Hashing, External tables etc.
- 6.5 Concepts of serialization, deserialization
- 6.6 Different Hive data storage formats including ORC, RC, and Parquet.
- 6.7 Introduction to HiveQL and examples.
- 6.8 Hive as an ELT tool and difference between Pig and Hive
- 6.9 Performance tuning opportunities in Hive, learnings and Best Practices.
- 6.10 Writing and mastering Hive UDFs
- 6.11 Error Handling and scope of creating Hive UDFs.

## 7. Pig and Latin

- 7.1 Basics of Pig and Why Pig?
- 7.2 Grunt
- 7.3 Pig's Data Model
- 7.4 Writing Evaluation
- 7.5 Filter
- 7.6 Load & Store Functions
- 7.7 Benefits of Pig over SQL language
- 7.8 Input and Output formats to MR program.
- 7.9 Error Handling and scope of creating UDFs for Pig.

## 8. Hbase

- 8.1 HBase – Introduction
- 8.2 When to use Hbase
- 8.3 HBase Data Model
- 8.4 HBase Families & Components
- 8.5 Data Storage and Distribution
- 8.6 HBase Master



**Address: 5 Main Motor Road, Wilberforce,  
Freetown, Sierra Leone.**

**info@skyvisioncollege.com | www.skyvisioncollege.com**

**Phone : +232 73 797979 | 32 797979**

