Diploma in Data Analytics

1. INSTALLATION OF VMWARE

2. MYSQL DATABASE

3. CORE JAVA

1.1 Types of Variable

1.2 Types of Datatype

1.3 Types of Modifiers

1.4 Types of constructors

1.5 Introduction to OOPS concept

1.6 Types of OOPS concept

4. ADVANCE JAVA

1.1 Introduction to Java Server Pages

1.2 Introduction to Servlet

1.3 Introduction to Java Database Connectivity

1.4 How to create Login Page

1.5 How to create Register Page

5. BIGDATA

1.1 Introduction to Big Data

1.2 Characteristics of Big Data

1.3 Big data examples

6.HADOOP

i) BigData Inroduction,Hadoop Introduction and HDFS Introduction

1.1. Hadoop Architecture

1.2. Installing Ubuntu with Java on VM Workstation 11

1.3. Hadoop Versioning and Configuration

1.4. Single Node Hadoop installation on Ubuntu

1.5. Multi Node Hadoop installation on Ubuntu

1.6. Hadoop commands

1.7. Cluster architecture and block placement

1.8. Modes in Hadoop

- Local Mode

- Pseudo Distributed Mode

- Fully Distributed Mode

1.9. Hadoop components

- Master components(Name Node, Secondary Name Node, Job Tracker)

- Slave components(Job tracker, Task tracker)

1.10. Task Instance

1.11. Hadoop HDFS Commands

1.12. HDFS Access

-Java Approach

ii) MapReduce Introduction

1.1 Understanding Map Reduce Framework

1.2 What is MapReduceBase?

1.3 Mapper Class and its Methods

1.4 What is Partitioner and types

1.5 Relationship between Input Splits and HDFS Blocks

1.6 MapReduce: Combiner & Partitioner

1.7 Hadoop specific Data types

1.8 Working on Unstructured Data Analytics

1.9 Types of Mappers and Reducers

1.10 WordCount Example

1.11 Developing Map-Reduce Program using Eclipse

1.12 Analysing dataset using Map-Reduce

1.13 Running Map-Reduce in Local Mode.

1.14 MapReduce Internals -1 (In Detail) :

- How MapReduce Works

-Anatomy of MapReduce Job (MR-1)

- Submission & Initialization of MapReduce Job (What Happen ?)

- Assigning & Execution of Tasks

- Monitoring & Progress of MapReduce Job

- Completion of Job

- Handling of MapReduce Job

- Task Failure

- TaskTracker Failure

- JobTracker Failure

1.15 Advanced Topic for MapReduce (Performance and Optimization) :

- Job Sceduling

- In Depth Shuffle and Sorting

1.16 Speculative Execution

1.17 Output Committers

1.18 JVM Reuse in MR1

1.19 Configuration and Performance Tuning

1.20 Advanced MapReduce Algorithm :

1.21 File Based Data Structure

- Sequence File

- MapFile

1.22 Default Sorting In MapReduce

- Data Filtering (Map-only jobs)

- Partial Sorting

1.23 Data Lookup Stratgies

- In MapFiles

1.24 Sorting Algorithm

- Total Sort (Globally Sorted Data)

- InputSampler

- Secondary Sort

1.25 MapReduce DataTypes and Formats :

1.26 Serialization In Hadoop

1.27 Hadoop Writable and Comparable

1.28 Hadoop RawComparator and Custom Writable

1.29 MapReduce Types and Formats

1.30 Understand Difference Between Block and InputSplit

1.31 Role of RecordReader

1.32 FileInputFormat

1.33 ComineFileInputFormat and Processing whole file Single Mapper

1.34 Each input File as a record

1.35 Text/KeyValue/NLine InputFormat

1.36 BinaryInput processing

1.37 MultipleInputs Format

1.38 DatabaseInput and Output

1.39 Text/Biinary/Multiple/Lazy OutputFormat MapReduce Types

iii)TOOLS:

1.1 Apache Sqoop

- Sqoop Tutorial
- How does Sqoop Work
- Sqoop JDBCDriver and Connectors
- Sqoop Importing Data
- Various Options to Import Data
  - Table Import
  - Binary Data Import
  - SpeedUp the Import
  - Filtering Import
- Full DataBase Import Introduction to Sqoop

1.2 Apache Hive

- What is Hive ?
- Architecture of Hive
- Hive Services
- Hive Clients
- How Hive Differs from Traditional RDBMS
- Introduction to HiveQL
- Data Types and File Formats in Hive
- File Encoding
- Common problems while working with Hive
  - Introduction to HiveQL
  - Managed and External Tables
- Understand Storage Formats
  - Querying Data
- Sorting and Aggregation
- MapReduce In Query

- Joins, SubQueries and Views

-Writing User Defined Functions (UDFs)

-Data types and schemas

-Querying Data

-HiveODBC

-User-Defined Functions

1.3 Apache Pig :

- What is Pig ?

- Introduction to Pig Data Flow Engine

- Pig and MapReduce in Detail

- When should Pig Used ?

- Pig and Hadoop Cluster

- Pig Interpreter and MapReduce

- Pig Relations and Data Types

- PigLatin Example in Detail

- Debugging and Generating Example in Apache Pig

1.4 HBase:

- Fundamentals of HBase

- Usage Scenerio of HBase

- Use of HBase in Search Engine

- HBase DataModel

- Table and Row

- Column Family and Column Qualifier

- Cell and its Versioning

- Regions and Region Server

- HBase Designing Tables

- HBase Data Coordinates

- Versions and HBase Operation

- Get/Scan

- Put

- Delete

1.5 Apache Flume:

- Flume Architecture

-Installation of Flume

-Apache Flume Dataflow

- Apache Flume Environment

-Fetching Twitter Data

1.6 Apache Kafka:

-Introduction to Kafka

-Cluster Architecture

-Installation of kafka

-Work Flow

-Basic Operations

-Real time application(Twitter)

4)HADOOP ADMIN:

-Introduction to Big Data and Hadoop

-Types Of Data

-Characteristics Of Big Data

-Hadoop And Traditional Rdbms

-Hadoop Core Services

Hadoop single node Cluster Setup for 1X series :

-Hadoop single node cluster(HADOOP-1.2.1)

-Tools installation for hadoop1x.

-Sqoop,Hive,Pig,Hbase,Zookeeper.

-Analyze the cluster using

a)NameNode UI

b)JobTracker UI

-SettingUp Replication Factor

Hadoop Distributed File System:

-Introduction to Hadoop Distributed File System

-Goals of HDFS

-HDFS Architecture

-Design of HDFS

-Hadoop Storage Mechanism

-Measures of Capacity Execution

-HDFS Commands

The MapReduce Framework:

-Understanding MapReduce

-The Map and Reduce Phase

-WordCount in MapReduce

-Running MapReduce Job

Hadoop single node Cluster Setup :

-Hadoop single node cluster(HADOOP-2.7.3)

-Tools installation for hadoop2x

-Sqoop,Hive,Pig,Hbase,Zookeeper.

YARN

-Introduction to YARN

-Need for YARN

-YARN Architecture

-YARN Installation and Configuration

Hadoop Multinode cluster setup:

-hadoop multinode cluster

-Checking HDFS Status

-Breaking the cluster

-Copying Data Between Clusters

-Adding and Removing Cluster Nodes

-Name Node Metadata Backup

-Cluster Upgrading

Hadoop ecosystem:

-Sqoop

-Hive

-Pig

-HBase

-zookeeper

7) MONGODB

8) SCALA

1.1 Introduction to scala

1.2 Programming writing Modes

i.e. Interactive Mode,Script Mode

1.3 Types of Variable

1.4 Types of Datatype

1.5 Function Declaration

1.6 OOPS concepts

9) APACHE SPARK

1.1 Introduction to Spark

1.2 Spark Installation

1.3 Spark Architecture

1.4 Spark SQL

- Dataframes: RDDs + Tables

- Dataframes and Spark SQL

1.5 Spark Streaming

- Introduction to streaming

- Implement stream processing in Spark using Dstreams

-Stateful transformations using sliding windows

1.6 Introduction to Machine Learning

1.7 Introduction to Graphx

10) TABLEAU

11) DATAIKU

12) Product Based Web Application Demo based on java(Ecommerce Application)

13) Data deduplication Project

14) PYTHON

1.1 Introduction to Python

- What is Python and history of Python?

- Unique features of Python

- Python-2 and Python-3 differences

- Install Python and Environment Setup

- First Python Program

- Python Identifiers, Keywords and Indentation
- Comments and document interlude in Python
- Command line arguments
- Getting User Input
- Python Data Types
- What are variables?
- Python Core objects and Functions
- Number and Maths
- Week 1 Assignments

1.2 List, Ranges & Tuples in Python

-Introduction
- Lists in Python
- More About Lists
- Understanding Iterators
- Generators , Comprehensions and Lambda Expressions
- Introduction
- Generators and Yield
- Next and Ranges
- Understanding and using Ranges
- More About Ranges
- Ordered Sets with tuples

1.3 Python Dictionaries and Sets

-Introduction to the section
- Python Dictionaries
- More on Dictionaries
- Sets
- Python Sets Examples

1.4 Input and Output in Python

- Reading and writing text files

- writing Text Files

- Appending to Files and Challenge

- Writing Binary Files Manually

- Using Pickle to Write Binary Files

1.5 Python built in function

-Python user defined functions

-Python packages functions

-Defining and calling Function

-The anonymous Functions

- Loops and statement in Python

-Python Modules & Packages

1.6 Python Object Oriented

- Overview of OOP

- Creating Classes and Objects

- Accessing attributes

- Built-In Class Attributes

- Destroying Objects

1.7 Python Exceptions Handling

- What is Exception?

- Handling an exception

- try....except...else

- try-finally clause

- Argument of an Exception

- Python Standard Exceptions

- Raising an exceptions

- User-Defined Exceptions

1.8 Python Regular Expressions

- What are regular expressions?

- The match Function

- The search Function

- Matching vs searching

- Search and Replace

- Extended Regular Expressions

- Wildcard

1.9 Python Multithreaded Programming

- What is multithreading?

- Starting a New Thread

- The Threading Module

- Synchronizing Threads

- Multithreaded Priority Queue

-Python Spreadsheet Interfaces

-Python XML interfaces

1.10 Using Databases in Python

-Python MySQL Database Access

-Install the MySQLdb and other Packages

- Create Database Connection

- CREATE, INSERT, READ, UPDATE and DELETE Operation

- DML and DDL Oepration with Databases

- Performing Transactions

- Handling Database Errors

-Web Scraping in Python

1.11 Python For Data Analysis

- Numpy:
- Introduction to numpy
- Creating arrays
- Using arrays and Scalars
- Indexing Arrays
- Array Transposition
- Universal Array Function
- Array Processing
- Arrary Input and Output

1.12 Pandas:

- What is pandas?
- Where it is used?
- Series in pandas
- Index objects
- Reindex
- Drop Entry
- Selecting Entries
- Data Alignment
- Rank and Sort
- Summary Statics
- Missing Data
- Index Heirarchy

1.13 Matplotlib: Python For Data Visualization

1.14 Welcome to the Data Visualiztion Section

1.15 Introduction to Matplotlib

1.16 Django Web Framework in Python

1.17 Introduction to Django and Full Stack Web Development

15) R Programming

1.1 Introduction to R

1.2 Installation of R

1.3 Types of Datatype

1.4 Types of Variables

1.5 Types of Operators

1.6 Types of Loops

1.7 Function Declaration

1.8 R Data Interface

1.9 R Charts and Graphs

1.10 R statistics

16) Advance Tool for Analysis

1.1 git

1.2 nmpy

1.3 scipy

1.4 github

1.5 matplotlib

1.6 Pandas

1.7 PyQT

1.8Theano

1.9 Tkinter

1.10 Scikit-learn

1.11 NPL

17)Algorithm

1.naive bayes

2.Linear Regression

3.K-nn

4.C-nn