Hadoop Training Content by Mr. Sasidhar
Cloudera certified Hadoop professional
sasis937@gmail.com

--- Hadoop ---

Hardware Requirements: -- Systems must have at least 2 GB RAM.

Software Requirements: -- I will provide all software (Operating System also).

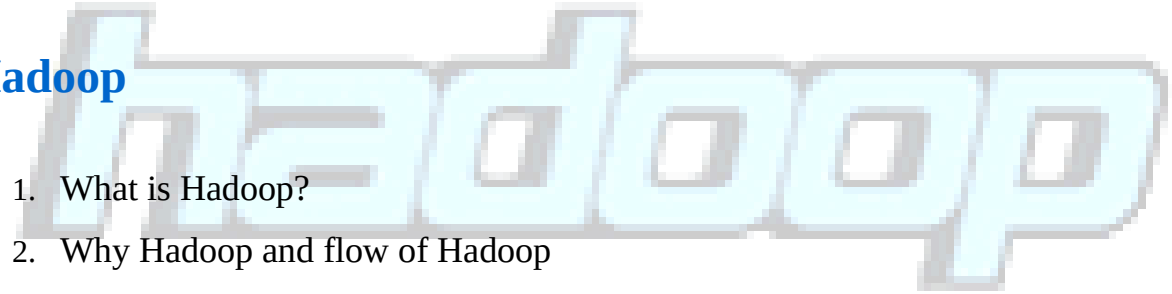
CONTENTS

Virtual box/VM Ware

1. Basics & Installations

Linux

1. Basics



Hadoop

1. What is Hadoop?
2. Why Hadoop and flow of Hadoop
3. Scaling
4. Distributed Framework
5. Hadoop v/s RDBMS
6. Brief history of Hadoop

Hadoop installation in pseudo mode

Hadoop installation in cluster mode

1. Adding and removing nodes (without down time)
2. Decommissioning nodes
3. Block size
4. Hadoop Processes (NN, SNN, JT, DN, TT)

Common errors when running Hadoop cluster, solutions

HDFS- Hadoop distributed File System

1. HDFS Design and Architecture
2. HDFS Concepts
3. Interacting HDFS using command line
4. Dataflow
5. Introduction about Blocks
6. Data Replication
7. Admin Commands
8. Hadoop archives

Hadoop Processes

1. Name node and its functionality
2. Secondary name node and its functionality
3. Job tracker and its functionality
4. Task tracker and its functionality
5. Data node and its functionality
6. Resource manager and its functionality

7. Node manager and its functionality

Map Reduce

1. Developing Map Reduce Application
2. Phases in Map Reduce Framework
3. Map Reduce Input and Output Formats
4. Advanced Concepts
5. Combiner
6. HAR
7. Partitioner, sorting, shuffling
8. Different phases of MapReduce programs
9. Data localization
10. Different unstructured data processing examples
11. Image processing by using MapReduce

Joining datasets in MapReduce jobs

1. Map-side join
2. Reduce-Side join

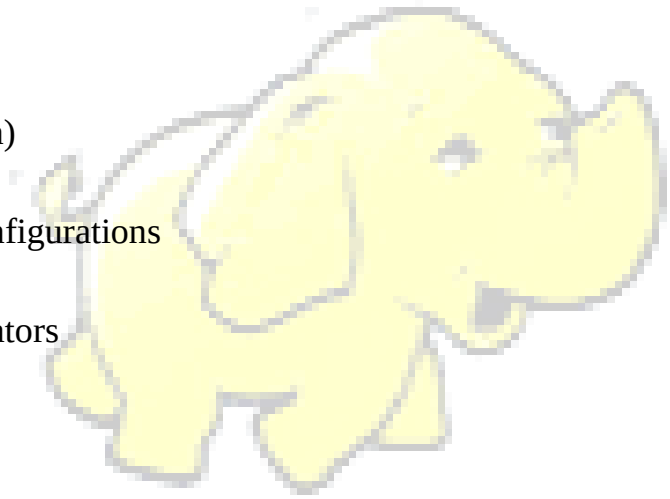
Hadoop Programming Languages:-

PIG

1. Introduction (Basics)
2. Installation and Configuration
3. Different datatypes in PIG

4. Interacting HDFS using PIG
5. Map Reduce Programs through PIG
6. PIG Commands
7. Execution mechanisms (grunt, script...)
8. Loading, Filtering, Grouping, joins….
9. Sample programs in PIG with Real time

Hive



a. Basics (Introduction)

b. Installation and Configurations

c. Datatypes and operators

d. HQL Commands

e. Interacting HDFS using Hive

f. MapReduce programs through Hive

g. Joins, groups, filter.....

h. Sample Programs in hive with real-time

I. Join vs Map Join

Impala

a. Basics

b. Commands

Sqoop

a. Introduction to sqoop

b. Installations & Configurations

c. Sqoop commands

d. Connect to relational database using sqoop and downloading lakhs of records to Hadoop (in single minute)

Flume

a. Basics (Introduction)

b. Installation and Configurations

NOSQL Databases Concepts

a. Hbase

I. Basics & Installations
II. commands
III. Interacting Hbase with HDF

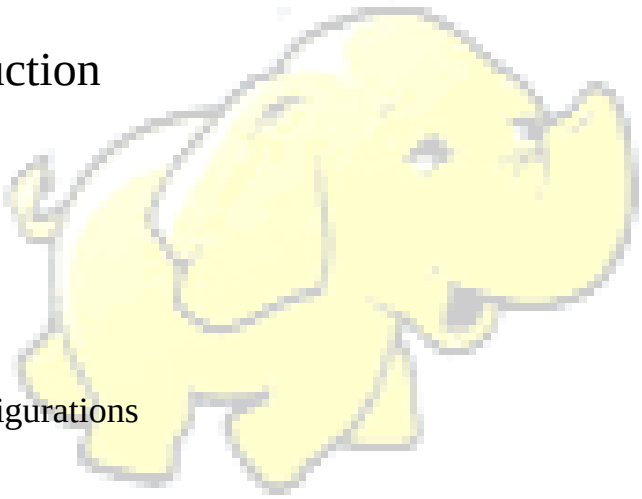
b. MongoDb

I. Basics & Installations

II. All queries for processing data

OOZIE Introduction

Zookeeper introduction



Apache Spark

a. Introduction

b. Installations and configurations

c. RDD , SC....

d. Scala Introduction

e. Interacting spark with HDFS

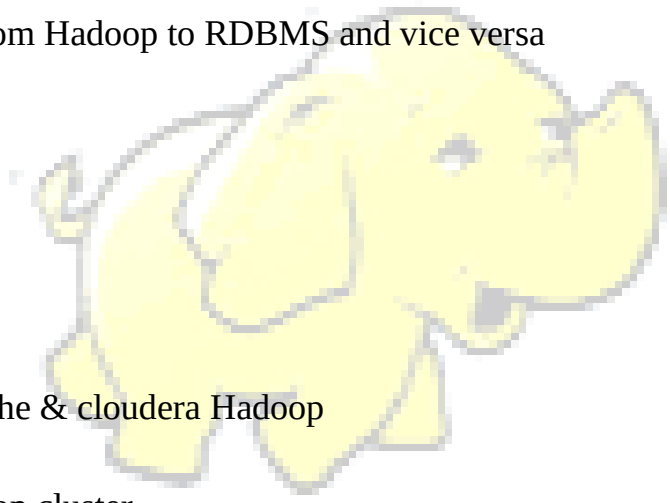
f. Programs in Spark through Scala

Specialties:--

ETL tool (Data Warehousing BI Tools):--

PDI

1. Introduction
2. Creating RDBMS database
3. Establishing Connection between PDI to RDMS database
4. Creating data in Hadoop
5. Establishing Connection between PDI to Hadoop data
6. Moving data from Hadoop to RDBMS and vice versa
7. Summarization



Highlights

a. Working with Apache & cloudera Hadoop

b. Practical's on Hadoop cluster

c. Real life use cases

d. Will cover old version of Hadoop and latest version of Hadoop

If you need anything more, please feel free to contact.

Contact – Mr. Sasidhar

sasis937@gmail.com

91 - 7795851053

Hadoop Training Content by Mr. Sasidhar
Cloudera certified Hadoop professional
sasis937@gmail.com