Course Syllabus

Installation

- Installation of bigdata - hadoop required softwares

How To Solve Big data Problem

- Introduction to Big Data
- Big Data Problem
- Solution to Big Data
-

Hadoop Framework, Cluster, Hadoop Introduction

- Breaking data into chunks
- Introduction to Hadoop Clusters
- Hadoop cluster concepts
- Why Hadoop cluster?
- Introduction to Hadoop
- Hadoop framework components
- Hadoop architecture
- How hadoop works

HDFS

- Introduction to Hadoop file distribution system(HDFS)
- HDFS read and write operation
- HDFS commands

Linux, Hadoop Installation, Hadoop Eco Sysytem

- Linux introduction
- Linux file system
- Linux basic commands
- Hadoop installation on single system
- Hadoop installation on distributed system
- Hadoop eco system

MapReduce

- Introduction to MapReduce
- Mapreduce lifecycle on Hadoop
- Examples of MapReduce - word count

Advanced MapReduce concepts with examples

- Counters
- Distributed cache
- Custom input format

Continuation of Advanced MapReduce concepts

- Sequence input format
- Reduce join

Project I

- Meentes can select project from predefined set of AcadGild projects
- We encourage all mentees to come up with own ideas for their projects

Project I

- Meentes can select project from predefined set of AcadGild projects
- We encourage all mentees to come up with own ideas for their projects

Apache Pig, Apache ETL Commands

- Introduction
- architecture
- Installation
- Apache pig environments
- Data types
- Detailed demostration of pig latin commands

Advanced Pig Concepts, Pig UDF

- Joins in Pig
- UDF introduction
- Why UDF ?
- UDF example

Introduction to Hive

- Inroduction
- Function
- Hive architechture
- Hive componets
- Limitations of Hive

Hive Data Model and Type System

- Data model and type system
- Data storage
- Introduction to HQL
- Hive query lifecycle on Hadoop
- Basic operations: create table and load data
- Partitioning and bucketing

Advanced Hive Operations

- Joins and union
- UDF introduction
- Why UDF ?
- UDF example
- Thrift server using Hive

HBase Introduction

- NoSQL databases
- HBase v/s RDBMS
- CAP theorem
- HBase: column family and HBase data model

HBase Architecture

- HMaster and slave
- HBase components
- Zookeeper
- Hbase work flow

Hands-On on HBase CRUD Operations, HBase Thrift Server

- Put and get
- Scan
- Filters in HBase
- Delete

- Data loading techniques
- What is HBase thrift server
- Integrating HBase with your application
- Sending request and response from Thrift server
- Example

Apache Sqoop

- Import and export of structured data on Hadoop
- Introduction to Apache Sqoop
- Sqoop Architecture
- Import and Export in Sqoop
- Sqoop commands
- Sqoop Installation

Apache Flume

- Injecting unstructured data into Hadoop
- Apache Flume introduction
- Flume Architecture
- Flume components
- Flume Installation
- Flume setup and schedule for Twitter social media data.

Introduction to Apache Oozie

- Overview + hands-on

Project II

- Meentes can select project from predefined set of AcadGild projects
- We encourage all mentees to come up with own ideas for their projects

Project II Cont...

- Meentes can select project from predefined set of AcadGild projects
- We encourage all mentees to come up with own ideas for their projects

Project II Cont...

- Meentes can select project from predefined set of AcadGild projects

- We encourage all mentees to come up with own ideas for their projects

Project II Cont...

- Meentes can select project from predefined set of AcadGild projects
- We encourage all mentees to come up with own ideas for their projects