ALTALUNE TECHNOLOGY

BIGDATA HADOOP

- Day 1
  - Python
    - What is Programming Language?
    - What is Python?
    - Data Types
    - Type Conversion
    - Operators
    - What are Algorithms?
    - Flow Charts
    - Control Statements
      - Conditional Statements
      - Loops
- Day 2
    - Functions
      - Predefined Functions
      - User Defined Functions
    - Closures
    - Decorators
- Day 3
    - Serialization & Deserialization
      - Json
      - Pickle
    - Socket Programming
    - Exception Handling
- Day 4
    - Application Programming Interface (API)
    - Database Programming
- Day 5
  - Basic Concepts of Linux
    - Installation
    - Basic Commands
    - Managing Users
    - Permissions & Special Permissions
    - Creating Sudo Users
    - Package Installation
      - RPM
      - YUM
    - Creating Partitions
    - Secure Shell (SSH)
    - Configuration of Database Server

- Day 6
  - **Introduction to Big Data**
    - Big Data Definition
    - Enterprise/Structured Data
    - Social/Unstructured Data
    - Unstructured data needs for analytics
    - What is Big Data?
    - Big Deal about Big Data
    - Big Data Sources
    - Industries Using Big Data
    - Big Data Challenges
- Day 7
  - **Cluster Installation**
- Day 8
  - **Introduction HDFS (Hadoop Distributed File System)**
    - History of Hadoop
    - The Ecosystem and Stack
    - Components of HDFS
    - Design of HDFS
    - Java Interfaces to HDFS
    - Architecture Overview
    - Development Environment
    - Hadoop Distribution and basic commands
- Day 9
  - **MapReduce**
    - Introduction to MapReduce
    - How MapReduce Works
    - Developing a Map Reduce Application
    - Hadoop 2.x MapReduce Architecture
    - Hadoop 2.x MapReduce Components
- Day 10
    - MapReduce components
      - Combiner
      - Partitioner
      - Reducer
    - Work Flow of YARN framework
    - Relation between Input Splits and HDFS Blocks
    - MapReduce Practical and Troubleshooting
- Day 11
  - **Hive**
    - About Hive
    - History of Hive
    - Use of Hive
    - Hive Use Case
    - Hive Vs Pig

- Hive Architecture and Components
- Metastore in Hive
- Limitations of Hive
- Traditional Database Vs Hive
- Hive Data Types and Data Models
- Hive Management
- Partitions and Buckets
  - Hive Tables(Managed Tables and External Tables)
  - Importing Data
  - Querying Data
  - Managing Outputs
  - Hive Script

❖ Day 12

- HiveQL
  - Joining Tables
  - Dynamic Partitioning
  - Custom Map/Reduce Scripts
  - Hive Indexes and views Hive query optimizers
  - Hive : User Defined Functions
- Hive Practical and Troubleshooting

❖ Day 13

- **Sqoop**
  - Introduction to Sqoop
  - History of Sqoop
  - Usage and Management of sqoop with RDBMS
  - Sqoop Architecture
  - Sqoop Commands
    - Command to get data from RDBMS form HDFS
    - Command to put data in RDBMS form HDFS
  - Importance of sqoop with HDFS and RDBMS
  - Sqoop Practical and Troubleshooting

❖ Day 14

- **Apache Spark**
  - Introduction to Apache Spark
  - History of Spark and Spark Versions/Releases
  - Spark Architecture
  - Spark Components
  - Usage and Management of Spark with HDFS
  - Spark Practical
  - Spark Streaming
  - Spark MLlib

❖ Day 15

- **Flume**
  - Introduction to Flume
  - History of Flume
  - Flume Architecture
  - Flume Components
  - Usage and Management of Flume

- Data Fetching from many resources in HDFS using Flume
- Flume Practical and Troubleshooting

❖ Day 16

- **Oozie**
  - Introduction to Oozie
  - History of Oozie
  - Oozie Architecture
  - Oozie Components
  - Oozie Work Flow
  - Scheduling with Oozie
  - Oozie with Hive, HBase, Pig, Sqoop, Flume
  - Oozie Practical and Troubleshooting

❖ Day 17

- **Zookeeper**
  - Introduction to Zookeeper
  - History of Zookeeper
  - Zookeeper components
  - Zookeeper Architecture
  - Usage and Importance Zookeeper with Hadoop
  - Management of Zookeeper
  - Zookeeper Practical and Troubleshooting

❖ Day 18

- **Cloudera**
  - About Cloudera Manager
  - History of Cloudera Manager
  - Usage and Management of Cloudera Manager
  - Usage and Management of each ecosystem tool with Cloudera manager.

❖ Day 19

- **Kafka**
  - Introduction and Configuration
  - Producer API
  - Consumer API
  - Stream API
  - Connector API
  - Topics and Logs
  - Consumers and Producers
  - Kafka as messaging system
  - Kafka as a storage System
  - Kafka for Stream Processing

❖ Day 20

- **Aws Integration**
  - EC2
  - EMR
  - RDS & Redshift
  - Lambda

- S3 storage
- Elastic Search
- Data Bricks (Azure