Big Data and Hadoop

Big data is high volume, high velocity and high variety information assets that demand cost effective, innovative forms of information processing for enhanced insight and decision making. Big data is the process of examining large amount of data of variety of types to uncover hidden patterns, unknown correlations and other useful information. The primary goal of big data analytics is to help organizations make better business decision.

The objective of this workshop is to understand the challenges in architectures to store, and access the Big data, perform analytics on Big data for data intensive applications. This workshop aims at bringing the students, professionals, academics and research scholars together to share their knowledge for achieving focused development and advancement in the field of BIG DATA ANALYSIS.

Topics t o be covered in BIG DATA Workshop

Session 1: BigData

How Big is this Big Data ?

Definition with Real Time Examples

How BigData is generated with Real Time Generation

Use of BigData-How Industry is utilizing BigData

Traditional Data Processing Technologies

Future of BigData!!!

Session 2: Hadoop

Why Hadoop?

What is Hadoop?

Hadoop vs RDBMS, Hadoop vs BigData

Brief history of Hadoop

Apache Hadoop Architecture

Problems with traditional large-scale systems

Requirements for a new approach

Anatomy of a Hadoop cluster

Hadoop Setup and Installation

Session 3: Hadoop Ecosystem

Brief Introduction about Hadoop EcoSystem (MapReduce, HDFS, Hive, PIG, HBase).

Session 4: HDFS

Concepts & Architecture

Data Flow (File Read , File Write)

Fault Tolerance

Shell Commands

Java Base API

Data Flow Archives

Coherency

Data Integrity

Role of Secondary NameNode

HDFS Programming Basics

Session 5: MapReduce

Theory

MapReduce Architecture

Data Flow (Map – Shuffle - Reduce)

MapRed vs MapReduce APIs

MapReduce Programming Basics

Programming [Mapper, Reducer, Combiner, Partitioner]

Session 6: HIVE & PIG

Architecture

Installation

Configuration

Hive vs RDBMS

Tables

DDL & DML

Partitioning & Bucketing

Hive Web Interface

Why Pig

Use case of Pig

Session 7: HBase

RDBMS Vs NoSQL

HBase Introduction

Duration: The duration of this workshop will be two consecutive days, with eight hour session each day in a total of sixteen hours properly divided into theory and hands on sessions.

Fees: Rs. 1200/- per head inclusive of all taxes.