



GUJARAT TECHNOLOGICAL UNIVERSITY

(Established Under Gujarat Act. No.:20 of 2007)

Date: 05-12-2016

CIRCULAR

Interested faculty members and students may register for the following webinar which is going to be held on Wed, Dec 7, 2016 3:30 PM - 4:30 PM IST.

Virtual Academy: HADOOP Ecosystem; T. Chandra Sekhar Reddy

Wed, Dec 7, 2016 3:30 PM - 4:30 PM IST

Registration URL: <https://attendee.gotowebinar.com/register/4299730759070786049>

Description:

Organizations create 2.5 Quintilian bytes of data. So much that 90% of the data in the world today has been set up in the last two years alone. The main data generation factors in the last two years is growth in hardware configurations of the computers/laptops , growth in the internet speed and rapid growth in the technology like smart phones, digital cameras etc.

What is Big Data? Big Data is large volumes of structured, semi-structured and unstructured data. This data is what organizations collect on a daily basis. The amount of data is not the important part, but the information gathered from that data is key. Collecting and analyzing Big Data gives organizations enhanced insight, decision making, and process automation.

Organizations are making critical predictions by sorting through and analyzing Big Data. Formatting unstructured data makes it suitable for data mining and analysis. Hadoop is the core platform for structuring Big Data. It also solves the problem of formatting it for analytic purposes. Hadoop uses a distributed computing architecture consisting of many servers using commodity hardware. This intern makes it inexpensive to scale and support massive data stores.

Hadoop is a software framework for distributed storing and processing of large datasets across large clusters of computers and it is write once and read many times. The core components of Hadoop are HDFS (Hadoop Distributed File System) and Map Reduce. HDFS is for storage purpose which we can able to store variety (i.e., structured, semi structured and unstructured data) and huge volume of data, whereas Map Reduce is for processing this huge volume and variety of data according to the user requirements.

Hadoop has its own family for processing different thing which is tied up in one umbrella called as Hadoop ecosystem; the members of an ecosystem are a PIG, Sqoop, Hive, Hbase, Flume and Oozie.

Presenter:

Dr.Chandrasekhara Reddy Y
Associate Professor, Department of CSE,
MLR Institute of Technology, Dundigal
Hyderabad

Sd/-
Registrar (I/c)