UE18CS322 Lecture Notes : Topic – Introduction to Big Data

Question Bank for Lecture

- T1 – Self assessment linked to LO 1.1 - #1, 4, 5
- What are common pitfalls in analysis of data? Give examples
- How is error estimation of model done in Big Data problems? How is this different from the traditional approach?
- In the Peter Norvig video he defines the following function to compute the best segmentation for a sentence?

UE18CS322 Lecture Notes : Topic – Introduction to Big Data 2

Question Bank for Intro2

- T1: Self assessment for LO1.3 – 1, 2
- T2: Multiple choice questions – 1-7, Short answer : 1-7
- What are the 4Vs characteristics of Big Data?
- What is the difference between volume and velocity of data?
- What differences in data generation have resulted in higher volume?
- Give some example of issues of veracity in Big Data?
- What modules does a typical Big Data architecture have?
- How Big Data and Hadoop are related to each other?
- What are the steps one must adopt to deploy a big data solution?

UE18CS322 Lecture Notes : Topic – Hadoop Distributed File System

Question Bank

- T2 : MCQs 1-10. Short answers 1-10
- What is Hadoop HDFS cluster and in how many different modes can you configure a Hadoop HDFS cluster?
- What is the usage of HDFS for the read operation?
- Described the use of Namenode? Why is it the most important process in HDFS
- The namenode contains the metadata that maps the user file block to the datanodes that contain a copy of the block. It is the most important service as in its absence, the cluster will fail to function. As the namenode is running on the master node, any failure to the master node, renders the entire cluster unusable.
- Can namenode and datanode be on commodity hardware?
- What are the differences between Hadoop and RDBMS?
- Explain the core components of Hadoop.
- What is a block in HDFS and what is its default size in Hadoop 1 and Hadoop 2? Can we change the block size?
- What is Commodity Hardware?
- Why is HDFS only suitable for large data sets and not the correct tool to use for many small files?