K.S. Institute of Technology, Bangalore

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

ASSIGNMENT QUESTIONS

Academic Year	**2020-21**		
Batch	**2018-2022**		
Year/Semester/section	**III/V/A & B**		
Course Code-Title	**18CS72- BIG DATA ANALYTICS**		
Name of the Instructor	**Vaneeta M**	**Dept**	**CSE**

Assignment No: 2 **Total marks:10**

Date of Issue: 17/12/2021 **Date of Submission: 24/12/2021**

Sl. No	Assignment Questions	K Level	CO	Marks
1.	**Identify** the features of HDFS, Explain its components with neat diagram.	K3	**CO2**	1
2.	**Interview** the significance of Apache Pig in Hadoop context. Describe the main components and the working of Apache pig with a simple example.	K3	**CO2**	1
3.	**Construct** the differentiating properties between ACID and BASE.	K3	**CO3**	1
4	**Make use of** suitable example and describe (i) key-value data store, (ii) document data store, (iii) object store.	K3	**CO3**	1
5	**Identify** the features of BigTable, RC, ORC and parquet data store.	K3	**CO3**	1
6	**Utilize** MongoDB Database and list its features and commands for querying the DBs.	K3	**CO3**	1
7	**Identify** four ways for handling Big Data problems.	K3	**CO3**	1
8.	**Model** master-slave and peer-to-peer distribution model with neat diagram.	K3	**CO3**	1
9.	**Construct** MapReduce process diagrammatically to depict a client submitting a job, the workflow of JobTracker and TaskTracker, and TaskTrackers creating the outputs.	K3	**CO4**	1
10.	**Calculate** the number of map tasks required. Assume an input file size of 10 TB and a data block size of 128 MB. Assume that each node does 100 maps. How many nodes are involved in processing? How will you change the number of map tasks per node to 120 using a Java statement?	K3	**CO4**	1

Course in charge **HOD**