

APJ Abdul Kalam Technological University

Ernakulam II Cluster

Third Semester M.Tech Degree Examination December 2017

05CS7041- BIG DATA PROCESSING

Time: 3 hrs.

Max. Marks: 60

I. The foundation of efficient data processing in Hadoop is its data storage model. What are the different options for storing data in Hadoop? Explain the architecture of one of its data storage model. (12 Marks)

II.

a) Write the iterative algorithm for Conjugate Gradient (CG) method to solve the System of Linear equations mentioned as $Ax=b$; Also mention the conditions which A has to be satisfied for the CG to work. (6 Marks)

b) Map-Reduce is a technique used to solve a relatively simple problem in situations where there is a lot of data, and it must be processed in parallel. Mention the conditions that must be true in order for Map-reduce to be applicable. (3 Marks)

c) Mention the main components of Map-Reduce Execution pipeline and explain in short.

(3 Marks)

III.

a) Elaborate the program to calculate the maximum recorded temperature by year for the weather dataset in Pig Latin. (10 Marks)

b) Make a discussion on filter UDF with an example program that removes records with unsatisfactory temperature quality readings. (8 Marks)

OR

IV.

a) What are the differences between external table and Managed table. Write a query for creating external table. (8 Marks)

- b) Discuss the functions of driver in Hive architecture and explain Hive in Hadoop Ecosystem. (10 Marks)

V.

- a) Write a bundle application to monitors and pre-processes raw weather data, which runs every 60 seconds, and to calculate and publishes the weather forecast every 10 minutes. (6 Marks)
- b) How does Spark provide it's speed and avoid Disk I/O while retaining the attractive fault tolerance, locality and scalability properties of MapReduce? (12 Marks)

OR

VI.

- a) Which Coordinator element is used for specifying input and output events for Coordinator applications? What are the different attributes and sub-elements it may contain? Explain. (6 Marks)
- b) How does Oozie describes Workflows? Explain the schema that represents the actions supported by the language, which is a dialect of XML. (12 Marks)