

Total No. of printed pages = 4

CS 1317 E 011

Roll No. of candidate

--	--	--	--	--	--	--	--	--	--

2018

B.Tech. 7th Semester End-Term Examination

BIG DATA ANALYTICS

(Elective I)

Full Marks – 100

Time – Three hours

The figures in the margin indicate full marks
for the questions.

Answer question No. 1 and any *six* from the rest.

1. Say whether the following statements are TRUE or FALSE: (10 × 1 = 10)
 - (i) The fire-and-forget scheme cannot guarantee that all data will be successfully processed.
 - (ii) Speed layer is the only part of Lamda Architecture that needs to be safeguarded from corruption.
 - (iii) Role of Batch Layer of Lamda Architecture affects how data will be written into storage.
 - (iv) One major function of the Speed Layer is to store real-time views of the Database into Batch Layer.
 - (v) Now-a-days data created and used is diverse in nature.

[Turn over

- (vi) Incremental algorithms always perform better than Recomputation algorithms.
 - (vii) In micro-batch stream processing, batches are processed in a strict order.
 - (viii) Batch Layer continuously updates the Serving Layer views.
 - (ix) A Data System computes queries that are functions of all data.
 - (x) A Bloom Filter is a compact data structure that sometimes may return false negatives but no positives.
2. (a) Explain the first principle of Big Data Analytics using examples wherever necessary. (5)
- (b) Explain how the three layers of the Lambda Architecture can serve the requirements of Big Data Analytics. (10)
3. (a) Define the terms Data, Information, Database, Query, and View in connection to Big Data Analytics, giving suitable examples for each. (5)
- (b) Represent the following information with a fact-based model as well as Graph Schema. (5+5= 10)
- “In an organization there are five employees named Mahesh, Suresh, Ruchika, Sachin and Smita. Mahesh, Suresh and Sachin are Male while Ruchika and Smita are Female. Age of Suresh, Sachin and Smita are 34, 32 and 27 respectively. Mahesh Stays at Ganeshguri, Guwahati while Suresh and Smita are from Beltola, Guwahati.”

4. (a) Write a Note on the Low-level nature of Distributed Filesystems that affect storage of the Master Dataset. (5)
- (b) Explain in detail the working of a HDFS, with reference to storing and reading files. (10)
5. (a) Give a comparison between Recomputation versus Incremental algorithms for computing batch views, along with the key trade-offs between these two approaches. (10)
- (b) Justify why the naïve procedure of pre-computing all possible queries in advance do not work in case of batch views. (5)
6. (a) Explain the terms ‘Eventual Accuracy’ and CAP theorem and their significance in Big Data Analytics. (10)
- (b) Compare and Contrast Asynchronous versus Synchronous updates in Speed layer. (5)
7. (a) Explain the Queues and Workers model for stream processing. (8)
- (b) Explain the consequences of System architecture without Queues in the context of Big Data Analytics. (7)
8. (a) Explain the two important issues related to Micro-batch stream processing, viz. Batch-local computation and Stateful computation. (7)
- (b) Explain the significance of achieving exactly-once semantics in micro-batch stream processing. (8)

9. (a) Explain the concepts of Partial Recomputation and its pros and cons in connection to Batch view computation. (7)
- (b) Write a note on Bloom Filter and Bloom Join.(8)