

**VI SEMESTER B.TECH. (COMPUTER SCIENCE & ENGG)**  
**GRADE IMPROVEMENT/MAKE UP EXAMINATIONS, AUG 2021**  
**SUBJECT: BIG DATA INTEGRATION AND PROCESSING [CRA - 4056]**  
**REVISED CREDIT SYSTEM**  
**(12-08-2021)**

Time: 2 Hours

MAX. MARKS: 40

**Instructions to Candidates:**

- ❖ Answer **Any FOUR FULL** questions.
- ❖ Missing data may be suitable assumed.

	Marks
1A. Explain the requirements for Big Data systems.	5
1B. What are the subqueries and aggregate functions? Explain them in detail with the help of an example.	5
2A. With the help of a diagram, explain the data model of Aerospike. Also Explain how the Aerospike Query Language queries fast.	5
2B. List any ten MangoDB operators with its description.	5
3A. What is meant by data exchange problem? Explain with an example.	5
3B. Explain data fusion. How the voting system and copy detection methods are used to resolve conflicts that may occur because of data fusion.	5
4A. What is the purpose of Analytical operations? Give 4 examples of analytical operations that can be performed on big data to provide insight for marketing decision making.	5
4B. What are transformations? Explain cross/Cartesian, make/join and filter operations? What is the difference between cross and match operations?	5
5A. Explain the difference between a narrow transformation and wide transformation. Describe map, flatmap, filter and coalesce as narrow transformations. List two wide transformations.	5
5B. Explain the steps of a SPARK pipeline, ending with a collect action with a diagram.	5
6A. Explain SPARK MLlib and SPARK streaming.	5
6B. Explain how to export data from Mango DB to CSV file.	5