

Roll No

IO-703 (D) (GS)
B.Tech., VII Semester
 Examination, November 2023
Grading System (GS)
Data Engineering
Time : Three Hours

Maximum Marks : 70

Note: i) Attempt any five questions.

किन्हीं पाँच प्रश्नों को हल कीजिए।

ii) All questions carry equal marks.

सभी प्रश्नों के समान अंक हैं।

iii) In case of any doubt or dispute the English version question should be treated as final.

किसी भी प्रकार के संदेह अथवा विवाद की स्थिति में अंग्रेजी भाषा के प्रश्न को अंतिम माना जायेगा।

1. a) Explain the concept of data-driven decisions. Why it is essential in today's business landscape?

डाटा-संचालित निर्णयों की अवधारणा की व्याख्या करें। आज के व्यावसायिक परिदृश्य में यह क्यों आवश्यक है?

b) Compare and contrast batch processing and stream processing in data pipelines. Explain with a suitable example.

डाटा पाइप लाइनों में बैच प्रोसेसिंग और स्ट्रीम प्रोसेसिंग की तुलना उपयुक्त उदाहरण सहित समझाइये।

2. a) Define modern data strategies and explain their significance in today's business environment, along with their advantages.

आधुनिक डाटा रणनीतियों को परिभाषित करें और आज के कारोबारी माहौल में उनके लाभों के साथ-साथ उनके महत्व की व्याख्या करें।

b) Why data pipelines are important in data engineering? Describe three common design patterns used in data pipelines.

डाटा इंजीनियरिंग में डाटा पाइपलाइन क्यों महत्वपूर्ण हैं? डाटा पाइपलाइनों में उपयोग किए जाने वाले तीन सामान्य डिज़ाइन पैटर्न का वर्णन करें।

3. a) Compare and contrast the modern data architecture offerings of two major cloud platforms (e.g., AWS and Azure) along with the strengths and weaknesses of each platform.

प्रत्येक प्लेटफॉर्म की ताकत और कमजोरियों के साथ दो प्रमुख क्लाउड प्लेटफॉर्म (उदाहरण के लिए, AWS और Azure) की आधुनिक डाटा आर्किटेक्चर पेशकशों की तुलना करें और अंतर करें।

b) Define the concept of ML (Machine Learning) security and its significance in data-driven organizations.

ML (मशीन लर्निंग) सुरक्षा की अवधारणा और डाटा-संचालित संगठनों में इसके महत्व को परिभाषित करें।

4. a) Define data enrichment and its role in enhancing the quality and value of data.

डाटा संवर्धन और डाटा की गुणवत्ता और मूल्य बढ़ाने में इसकी भूमिका को परिभाषित करें।

- b) Describe the importance of data validation in the data preparation process.
डाटा तैयार करने की प्रक्रिया में डाटा सत्यापन के महत्व का वर्णन करें।
5. a) Describe the process of batch data ingestion, including its key steps and characteristics.
इसके प्रमुख चरणों और विशेषताओं सहित बैच डाटा अंतर्ग्रहण की प्रक्रिया का वर्णन करें।
- b) Discuss the role and benefits of data lake storage in data management.
डाटा प्रबंधन में डाटा लेक स्टोरेज की भूमिका और लाभों पर चर्चा करें।
6. a) Explain the significance of securing data storage in a data-driven organization.
डाटा-संचालित संगठन में डाटा भंडारण सुरक्षित करने के महत्व को स्पष्ट करें।
- b) Describe the key components and architecture of Apache Hadoop.
Apache Hadoop के प्रमुख घटकों और वास्तुकला का वर्णन करें।
7. a) Describe the stages of the ML lifecycle, from problem formulation to model deployment and explain why each stage is crucial?
समस्या निर्माण से लेकर मॉडल परिनियोजन तक ML जीवनचक्र के चरणों का वर्णन करें और बताइए कि प्रत्येक चरण महत्वपूर्ण क्यों है?

- b) Explain the significance of data pre-processing in machine learning and list common techniques used for data cleaning and transformation.
मशीन लर्निंग में डाटा प्री-प्रोसेसिंग के महत्व को समझाइए और डाटा सफाई और परिवर्तन के लिए उपयोग की जाने वाली सामान्य तकनीकों की सूची बनाइए।

8. Write short notes on any two.
- Deploying a machine learning model and the challenges associated with it.
 - Amazon EMR for big data processing.
 - Processing and consumption phases in a modern data architecture pipeline.
 - Define the five V's of data and their significance.
- किन्हीं दो पर संक्षिप्त टिप्पणियाँ लिखें :
- मशीन लर्निंग मॉडल का परिनियोजन और उस से जुड़ी चुनौतियाँ।
 - बड़े डाटा प्रोसेसिंग के लिए अमेज़ॉन EMR
 - आधुनिक डाटा आर्किटेक्चर पाइप लाइन में प्रसंस्करण और उपभोग चरण।
 - डाटा के पाँच V's और उनके महत्व को परिभाषित करें।
