

Minor Exam
CSL7090 Software and Data Engineering

Department of Computer Science and Engineering
Indian Institute of Technology Jodhpur

Duration: 2 Hour

Date: 21-Sep-2024

Time: 07:00 PM

Marks: 90

Instructions:

1. Any kind of plagiarism may result in zero marks in the Minor component.
2. Write your answers concisely and precisely. Unnecessary content may lead to negative marking. It is recommended to draw diagrams, schematics and explain your solutions in bullet points.
3. It is an open book examination, you can refer to books, the internet, or any other source except discussing with each other and LLM models such as ChatGPT, etc. Any amount of copying from any source, including a book, colleagues' answers, etc., will result in zero marks in the whole paper.

S. No.	Question	Marks
How to answer	In the context of the given scenarios, propose an appropriate solution using the concepts covered during the class while answering the following questions. In each answer, clearly highlight the concepts and software architectural elements used by you in the solution.	
Scenario I	Public Mobile Connection to Employees Without Knowing the Mobile Number	
1	How can an organization ensure privacy for its employees while allowing the public to connect with them using only their names on their personal mobile number without sharing their personal mobile numbers?	10
2	What kind of infrastructure is needed to map a name to a specific employee's mobile phone in real-time, and how can this be implemented seamlessly for both the public and employees?	10
3	How should the organization manage scalability and ensure that multiple employees with the same name can still be uniquely identified by the system?	10

Scenario II	Improving Existing RDBMS	
4	Which features of the existing RDBMS can create performance bottlenecks, and what alternative technologies or architectural changes should be considered to remove these inefficiencies?	10
5	What additional features can be integrated into the RDBMS to support modern use cases like real-time analytics, distributed computing, or integration with machine learning models?	10
6	How can the RDBMS be enhanced to handle unstructured or semi-structured data while still maintaining the integrity and advantages of relational data?	10
Scenario III	IoT-Based Attendance System for 200+ Students	
7	What IoT technologies (e.g., RFID, Bluetooth, facial recognition) can be used to collect attendance from 200+ students within one minute, ensuring minimal disruption and no chaos during the process?	10
8	How would the system handle scenarios where students forget or lose their identification devices (like RFID cards) to ensure accurate attendance records?	10
9	How can the attendance system ensure real-time updates and prevent system failures during class, ensuring that all data is captured within the one-minute window?	10

CSL7090 Software and Data Engineering
Department of Computer Science and Engineering
Indian Institute of Technology Jodhpur

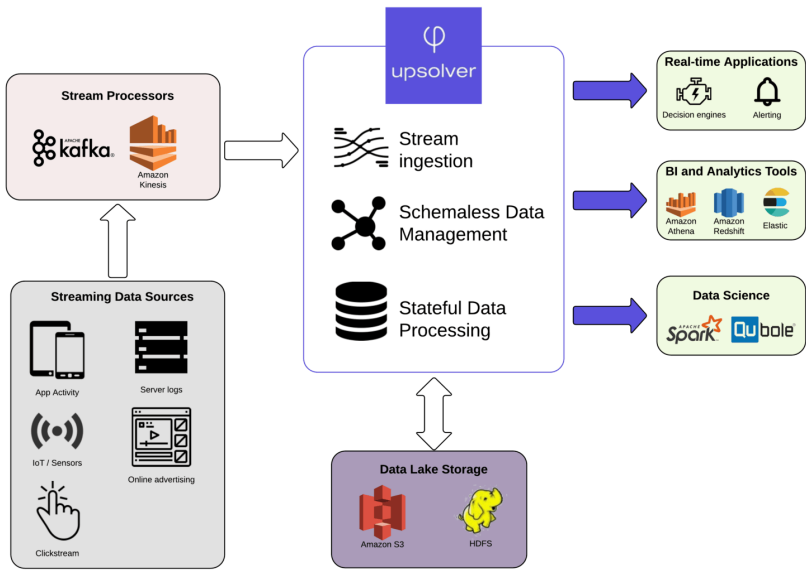
Duration: 3 Hour
Date: 21-Nov-2024

Time: 6:00 PM to 9:00 PM
Marks: 100

Please carefully follow the guidelines below for your examination:

1. **Plagiarism Policy:** Any instance of plagiarism will result in a zero grade for the entire course. Ensure that all your work is original and reflects your understanding of the material.
2. **Answer Submission:** Your answers must be concise and to the point. Submissions must be uploaded as a typed or scanned PDF file on Google Classroom. This is mandatory.
3. **Open-Book Examination Rules:** You may refer to books, the internet, or any other resources, but collaboration or discussion with others, including using tools like ChatGPT or similar AI models, is strictly prohibited. Any evidence of copying from books, peers, or any other sources will lead to a zero grade for the entire paper.
4. **Scope of Answers:** Use only the concepts and topics discussed during the course to answer the questions. Content outside the covered topics as part of the course will not be evaluated.

Your adherence to these instructions ensures fairness and integrity in the examination process.
Good luck!

S. No.	Question	Marks
1	<p>Study the provided data streaming diagram carefully. This diagram is sourced from the following article: https://www.upsolver.com/blog/streaming-data-architecture-key-components</p>  <pre> graph TD subgraph Sources [Streaming Data Sources] direction TB A[App Activity] B[Server logs] C[IoT / Sensors] D[Online advertising] E[Clickstream] end subgraph Processors [Stream Processors] direction TB F[Kafka] G[Amazon Kinesis] end subgraph Central [Upsolver] direction TB H[Stream ingestion] I[Schemaless Data Management] J[Stateful Data Processing] end subgraph Destinations direction TB K[Real-time Applications: Decision engines, Alerting] L[BI and Analytics Tools: Amazon Athena, Amazon Redshift, Elastic] M[Data Science: Spark, Qubole] end subgraph Storage [Data Lake Storage] direction TB N[Amazon S3] O[HDFS] end Sources --> Processors Processors --> Central Central --> Destinations Central <--> Storage </pre>	

	<p>Your task is to design an architectural diagram to implement the system illustrated in the referenced figure, using the concepts discussed during the course. The resulting architecture must support real-time data processing and analytics, as highlighted in the original diagram.</p> <p>Additionally, provide a detailed justification for the inclusion of each architectural element in your design by creating a comparative table that evaluates alternatives. Ensure that all discussions and solution points remain focused on the context of the given diagram.</p> <p>Marks will be awarded based on the following criteria:</p> <ul style="list-style-type: none"> a. Use of standardized notations for the architectural diagram. b. Inclusion of 4+1 architectural views. c. Appropriate selection and application of architectural elements in the diagrams. d. Accuracy and depth of the comparative analysis of alternatives in the table format 	<p>5</p> <p>10</p> <p>10</p> <p>15</p>
2	<p>In the context of Big Data and IoT, revisit the discussion on the key characteristics of Big Data—Volume, Velocity, Veracity, Variety, and Value. For a Big Data system designed for real-time IoT-driven monitoring and predictive analytics, assume a structure consisting of five modules, each containing two components, resulting in ten interacting components in total.</p> <p>Now, consider yourself in the roles of Programmer, Designer, Architect, and Project Manager. For each role, identify and list the specific items or concerns you would focus on during the project review or analysis process. Provide brief descriptions of these concerns for each role.</p> <p>Marks will be allocated based on the following criteria:</p> <ul style="list-style-type: none"> 1. Identification of relevant modules and components, with justification. 2. Comprehensive listing of appropriate concerns with brief descriptions for the Programmer role. 3. Comprehensive listing of appropriate concerns with brief descriptions for the Designer role. 4. Comprehensive listing of appropriate concerns with brief descriptions for the Architect role. 5. Comprehensive listing of appropriate concerns with brief descriptions for the Project Manager role. 	<p>20</p> <p>10</p> <p>10</p> <p>10</p> <p>10</p>