

CSL7090

Software and Data Engineering

FINAL WRITTEN EXAMINATION

Duration: 3 Hours

21-Nov-2025

Maximum Marks: 50

SECTION A – SHORT ANSWER ($7 \times 2 = 14$ marks)

(All questions are NOT compulsory. You can SKIP any TWO. Keep answers concise.)

- ✓ 1. Define *microservices* and list any two advantages of using them in modern software design.
- ✗ 2. State any two differences between Type-1 and Type-2 virtualization.
- ✓ 3. What is the purpose of the 4+1 architectural view model? Mention any two views.
- ✗ 4. Define *NoSQL database*. Give one example and one suitable use-case.
- ✓ 5. What is a column-oriented database and where is it preferred?
- ✗ 6. What are RDDs in Apache Spark? List two properties.
- ✓ 7. Write the significance of a Directed Acyclic Graph (DAG) in data processing.
- ✗ 8. Mention any two differences between JSON and XML.
- ✗ 9. What is the main role of the NameNode in HDFS?

SECTION B – MEDIUM LENGTH ($4 \times 5 = 20$ marks) (Attempt any four questions.)

10. Scenario-based:

You are hired by a company moving from a monolithic architecture to a microservices-based architecture. Explain:

- ✓ a. How modularity improves fault isolation.
- ✗ b. Why API gateways and service discovery become essential.
- ✗ c. Any two challenges the company may face during migration.

11. Draw the 4+1 architectural view model diagram and briefly explain each view using a simple example such as an online examination system or library management system.

12. Compare the following four databases in a tabular form: Relational DB, Document DB (MongoDB), Column DB (Cassandra/HBase), Graph DB (Neo4j)

Use the following comparison parameters: Data model, Schema, Query pattern, Scalability approach

- X 13.** Explain the difference using simple examples between:
- a. Virtual machines vs Containers
 - b. Cluster computing vs Cloud computing
 - c. Shared computation vs Partial computation

- 14.** Write short notes on the following (any two):

- a. Map-Reduce workflow
- b. Dynamic resource allocation in Spark
- c. Columnar storage architecture

SECTION C - LONG ANSWER / CASE STUDY (2 × 8 = 16 marks) (Attempt any two questions.)

15. Case Study: Cloud + Data Architecture Design (8 marks)

A startup wants to build a **scalable food delivery platform** similar to Zomato/Swiggy. They expect: 5 million users, Real-time order tracking, Daily analytics reports, Fault tolerance and quick recovery

Task: Design a complete architecture using the following components:

Microservices, Containerization (Docker/Kubernetes), Cloud data center/cluster, NoSQL + SQL combination, Data lake or HDFS, Streaming engine (Kafka + Spark/Flink/Storm)

Draw a block diagram and justify design choices for: Scalability, Fault tolerance, Real-time processing, Storage strategy

16. Case Study: Large-Scale Data Processing (8 marks)

An electricity board collects hourly consumption readings from **2 million smart meters**. They want to build a pipeline that: Ingests raw meter data, Stores it cost-effectively, Detects anomalies, Produces daily usage summaries

Explain how you will use: HDFS, Spark (RDD or DataFrame model), DAG scheduling, Map-Reduce style batch processing, NoSQL database for fast lookups

Add a small sketch/flow diagram.

X 17. Analytical / Open-Ended (8 marks) Choose any one:

(A) Explain the complete architecture of a **fault-tolerant data center**. Include concepts such as: Redundancy, Load balancing, Failover, Virtualization, Container orchestration, Energy efficiency

OR

(B) Discuss the evolution from Hadoop → Spark → Storm/Flink in data-intensive systems. Explain how each solves the limitations of the previous generation.