



CSL4030: Data Engineering (July-December, 2023)
Minor-2 Answer Outlines
Department of Computer Science and Engineering
Indian Institute of Technology Jodhpur, Rajasthan, India 342030

Duration: 1 hour

Full Marks: 32

Answer 1.

[2+3+1+1+6+2]=15

- i. Consistency and availability.
- ii. According to the CAP theorem, consistency and availability both can be achieved together if there are no network partitions. However, network partitions are inevitable. Hence, it is practically impossible to provide consistency and availability together.
- iii. Exactly once.
- iv. Yes, if we can find another route, then it is a link failure, not a network partition.
- v. The bully algorithm. Its pseudocode is given in the lecture slides.
- vi. The BASE properties.

Answer 2.

[1+3+2]=6

- i. Yes, e.g., we can use AI to generate transcripts of a YouTube video, then extract text keywords from it.
- ii. Property-graph model (Neo4j, Cypher query language or CQL) and triple-store model (RDF4j, SPARQL).
- iii. Pages 76-77, Section 'SSTables and LSM-Tree', Kleppmann.

Answer 3.

[1+1+4+5]=11

- i. A network directory's only function is to store information about its network resources. Hence, it does not need the complete feature set of a distributed database. Moreover, a network directory has a hierarchical tree structure similar to that of a file system directory (folder, subfolder, subsubfolder, etc.). Utilizing the information about the tree structure, we can design simple algorithms for forwarding queries from one node to another node.
- ii. When an entry (i.e. a node) in a server's subtree of a directory information tree (DIT) refers to another server's subtree of the same DIT, it is called a referral.
- iii. For players who are frequent travellers, we need to create replicas and place the replicas strategically to ensure high availability of the player's data at multiple locations. We must take extra measures to ensure consistency between these replicas. For administrators, creating horizontal fragments or shards should suffice. For example, the administrators residing at Jodhpur will almost never make a purchase from outside Jodhpur. Hence, we can create a shard of Jodhpur-based administrators and save the shard only on the Jodhpur server.
- iv. For detecting local deadlocks, we can use the quorum consensus protocol. For detecting global

deadlocks, we can create a special data item called 'ticket'. Each site/server will be assigned a ticket. Any transaction that desires to read or write data of a site, first needs to get a lock on that site's ticket.