

## MODEL QUIZ PAPER FOR EE3801 PART I [T: 40mins

**max;** Use 1 or 2 lines answer to justify your answer, if requested; Each Q ~1.5-2mins; You can use 1 blank white sheet paper to do any calculations and a non-programmable calculator; internet resources and accessing your files on your laptop will not be possible. ]

- (i) In any \_\_\_\_\_, both privacy and security are compromised as outside network traffic may disrupt in an unknown fashion. Choose one the following: (a) Inter-Cluster Communication; (b) Intra-Cluster Communications (c) Both (a) and (b); (d) Depends on the type of Cluster OS used
- (ii) Consider a BDP comprising  $k=12$  stages. A batch of independent tasks  $N = k/2$  tasks need to be processed on this BDP. All tasks take  $T$  units of time and need to be processed at every stage of the BDP. The total completion time for 50% of the tasks is \_\_\_\_\_. Choose one of the following. (a)  $11T$  (b)  $17T$  (c)  $14T$  (d) insufficient information to compute
- (iii) A certain parallel application demands the use of 64K processor cores. A data engineer needs to suggest a Cloud Data Centre (CDC) to be used with  $N$  cabinet nodes. Each cabinet node can accommodate a maximum of 16 node boards. A node board has 8 slots and in each slot a compute card that can hold 4 chips is used. If a chip has 4 processor cores, the number of cabinet nodes that need to be used is \_\_\_\_\_. Assume that all cabinet nodes are to be used fully.
- (iv) A Compact Cluster (CC) is a \_\_\_\_\_. Choose one of the following and state your reason. (a) Tightly-coupled system (b) Subset of a Slack Cluster (SC) system (c) Distributed system using LANs/WANs (d) None of the above
- (v) A \_\_\_\_\_ architecture for a cluster system is suitable when scalability is the key factor. Choose one of the following. (a) Shared memory (b) Shared nothing (c) Shared disk (d) None of the above. State your reason.
- (vi) A chip has 4 processor cores and a compute card in the system comprises 2 such chips. While building a server, a system architect uses a board that can accommodate 32 compute cards. If dedicated local memory needed for computing for each core is 4Mbytes space, determine the total amount of memory that will be used, if the entire server is fully utilized?
- (vii) A distributed computing cluster framework needs to have the following two imperative characteristics - \_\_\_\_\_ & \_\_\_\_\_. Choose the two most relevant from the following.
  - a) Large scale data processing capability
  - b) Low-power computing ability
  - c) Data processing software on all nodes
  - d) Ability to migrate data

- (viii) In general, the type of service that is responsible for facilitating communications and storage between the components of a data pipeline line is generally referred to as \_\_\_\_\_ which is part of the \_\_\_\_\_ engine. Choose from the following.
- Ingestion service, Streaming
  - Data storage, Processing
  - Ingestion Service, Static data processing
  - None of the above pairs are appropriate
- (ix) Consider a **symmetric 4-level binary tree network**. Level 0 has the root node, followed by level-1 with 2 nodes, level-2 with 4 nodes, and level-3 has 8 leaf nodes. Assume all nodes can store data. An optimal chromatic color for this network using T-coloring is 2. **(True/False)? State your reason.**
- (x) For a self-routing network interconnecting 8 processors and 8 memory modules for a small scale cluster discussed in the lecture session, identify all the correct statements. Note that each 2 x 2 switch used can only be in 2 states – parallel connection between the I/O ports or crossed connection between the I/O ports at any point in time.
- A processor has exactly 2 independent paths to reach any memory module
  - A processor has at least 2 independent paths to reach any memory module
  - A processor needs to broadcast twice to send an information to all memory modules.
  - None of the above.
- (xi) Consider a fully-connected topology of a small cluster interconnecting 5 nodes to store data. Then the total number of links in this topology is \_\_\_\_\_ and the minimum number of colors needed to color the nodes using T-coloring is \_\_\_\_\_.
- (xii) Consider a **symmetric 2-dimensional 3 x 3 mesh used as a compute cluster topology**. Assume all 9 nodes can store data. It is desired by the application to fragment three large scale data D1, D2, and D3, from 3 users and store on the nodes to minimize any threat with a criterion that no two adjacent nodes must store the data from the same user. The fragmentation for each data is given as follows. D1= \_\_\_\_\_ chunks, D2 = 2 chunks and D3 = 3 Chunks. It is desired to use one distinct color for each data type. Identify the maximum number of fragments to store D1, if this T-coloring algorithm is to be used.
- (xiii) In a DB present in a DWH, once the changes done by transaction get committed then data should continue to remain in the DB even if the system fails or crashes or any power failure occurs. Which property of ACID insists this behavior? Choose one of the following.
- Isolation
  - Atomicity
  - Durability
  - Consistency

- (xiv) Consider a very large-scale input data from a *DataLake* which demands slow processing. A DE chooses HDFS system for storing data. In this case, which layer in a Lambda architecture is responsible for handling such data? Choose one of the following. State your reason.
- Batch Layer
  - Speed Layer
  - Service Layer
- (xv) For a self-routing network interconnecting 8 processors and 8 memory modules for a small scale cluster discussed in the lecture session, it may be possible that any 2 processors can read simultaneously from any one memory module in this cluster. (True/False)? Justify your answer.
- (xvi) Table below identifies certain DC layer responsibilities. Match all the LHS categories with the RHS categories. Choose the correct mapping.

(a) Access Layer	(1) Provides security services such as firewalls, and intrusion prevention systems
(b) Aggregate Layer	(2) Connects to internet, WAN, platforms.
(c) Core Layer	(3) Facilitates connecting storage devices and compute devices and provides routing functions.
	(4) Provides a platform for high-speed packet switching between multiple aggregation modules

- (b,3), (a,4), (c,1), (d,2)
- (a,3), (b,1), (c,2), (c,4)
- (a,3), (b,2), (c,1), (a,4)
- (a,3), (b,2), (c,1), (c,4)

- (xvii) Match **all** the items on the LHS with **all** the RHS items.

(a) Extract	(1). Validating, Standardizing, and Enriching
(b) Transform	(2). Connecting to the data source; Perform data sampling or profiling to understand the structure and quality of the data; During this process handling the errors and exceptions, if any;
(c) Load	(3). Create or configure the destination database or data warehouse; Define the schema or structure of the target data tables; Ensure data consistency and integrity during this process
	(4). Filtering, Cleansing, De-duplicating, Normalizing, and Aggregating also belongs to this step.