SSN COLLEGE OF ENGINEERING

Department of Computer Science and Engineering

B.E. Computer Science Semester: 7

CS6703 - GRID AND CLOUD COMPUTING

Answer Key

14 AUG 2017, 8 - 9.30 AM **Unit Test 2** **Max Marks: 50**

Staff Dr. K. Vallidevi / Ms. Y. V. Lokeswari

PART A **10 x 2 = 20**

1. List the topologies of Grid. (CO1) (K2)

Intra Grid,
Extra Grid &
Inter Grid

2. Differentiate between Data Grid, Computational Grid and P2P Grid. (CO1) (K2)

Computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities.

Data Grid: Data grid is a grid computing system that deals with data — the controlled sharing and management of large amounts of distributed data.

P2P Grid: This combines p2p nature with Grid infrastructure

3. Specify the motivation of Parallel Programming Paradigm. (CO3) (K2)

To improve productivity for programmers
To decrease programs time to market
To leverage underlying resources more efficiently
To increase system throughput
To support higher levels of abstraction
MapReduce, Hadoop and Dryad are parallel and distributed programming models.

4. What will happen when Name node is failed in HDFS? (CO3) (K2)

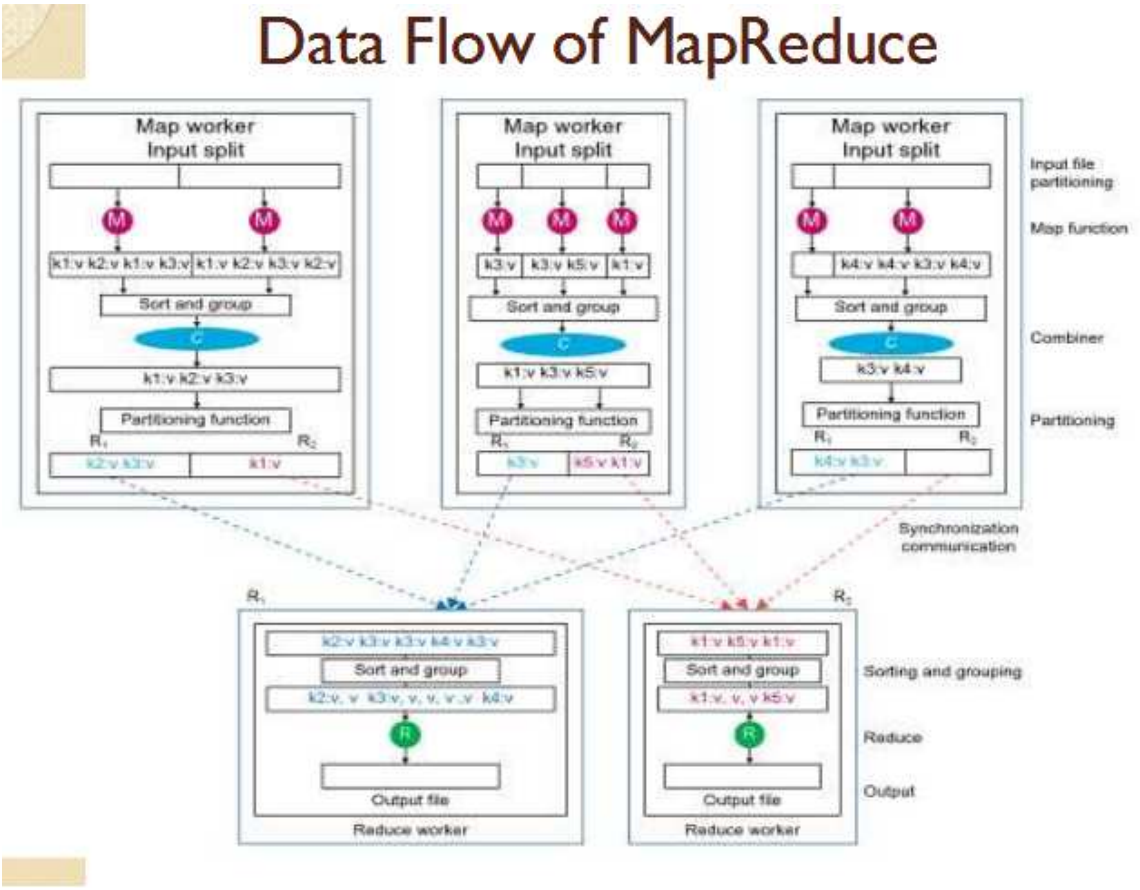
A secondary name node which is one among the data node will take over Name node. Name node takes frequent checkpoints. New secondary name node will resume from recently taken checkpoint.

5. What is the drawback of Hadoop and how to overcome it? (CO3) (K2)

Hadoop is not suitable for iterative processing. Twister overcomes it by using local cache in each worker node.

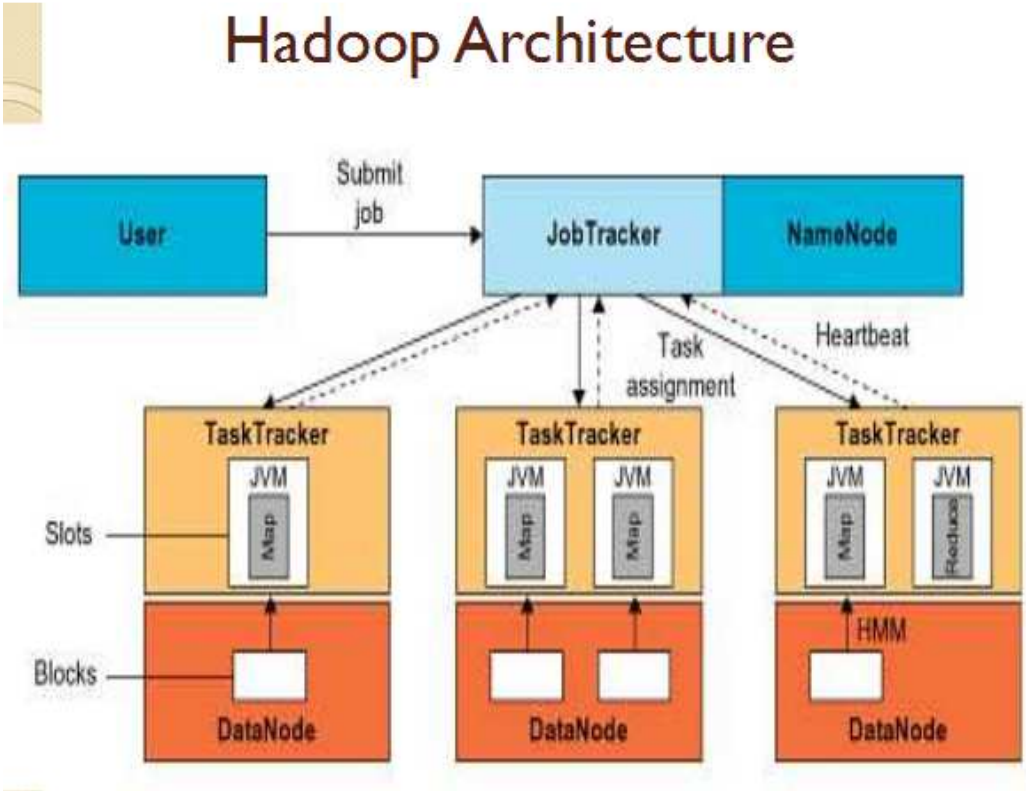
PART B **40 Marks**

6. (a) Sketch and explain the working of Map Reduce data flow and control flow. (CO3) (K2) 16)



OR

(b) Explain the architecture of Hadoop framework with diagram. (CO3) (K2) (16)



7. (a) (i) Briefly explain about how is fault tolerance provided by HDFS. (CO3) (K2) (5)

HDFS makes 3 replicas by default. It places replica and selects replica from the nearest workstation / Data center.

1. Block replication
2. Replica placement
3. Heartbeat and BlockReport messages

(ii) Explain how communication overhead and locality issue is resolved by Map Reduce Library. (CO3) (K2) (5)

Communication overhead : Reduce can start only after all maps finish.

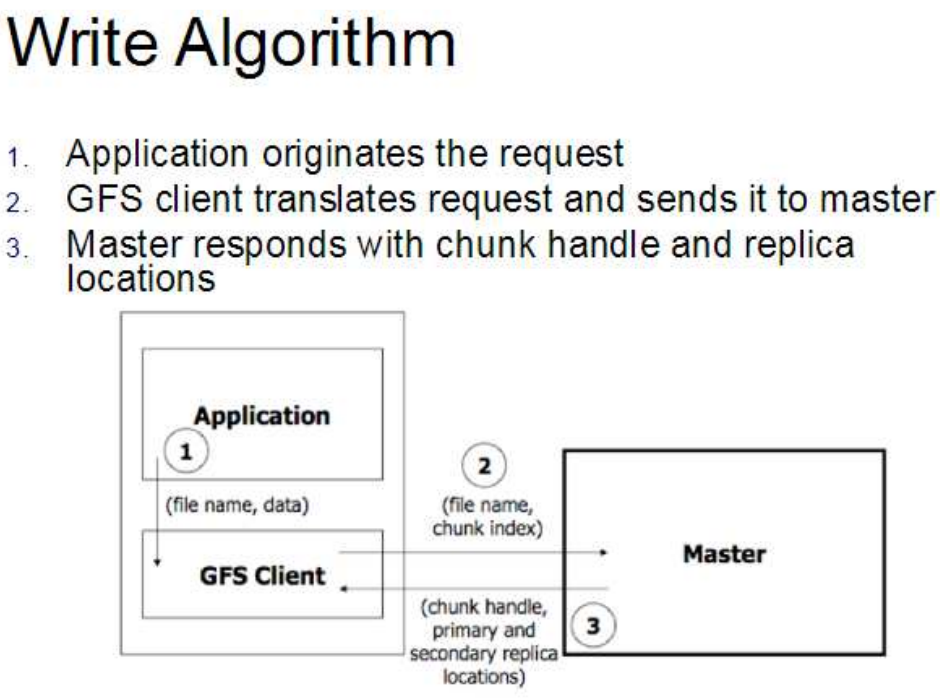
Each Reducer will pick key value pair from the corresponding region of the partitioning function.

Locality Issue: Map has to be assigned to a work station or to a task tracker which could extract data from the same name node in the data center. This could reduce latency in accessing files. Mapper has to communicate the location of key value pair to Reducer.

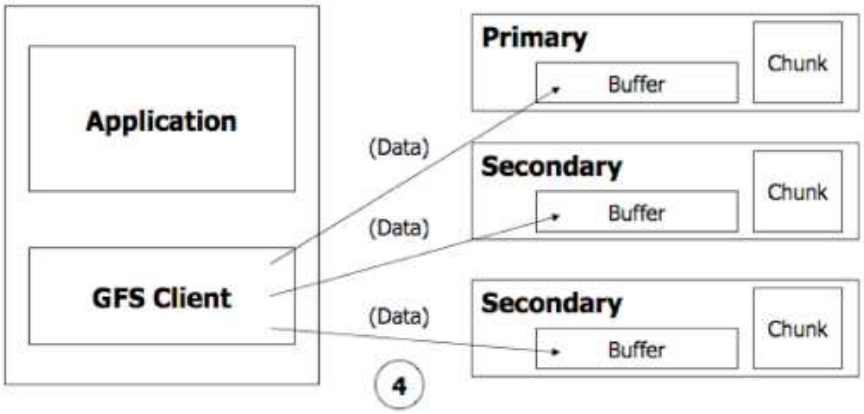
(OR)

(b) Explain about how file mutations are carried out in Google File System (GFS). (CO3) (K2) (10)

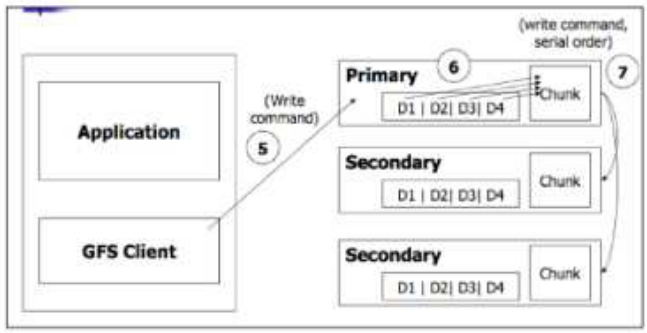
Mutations to a file could be either Record Write or Record append operation.



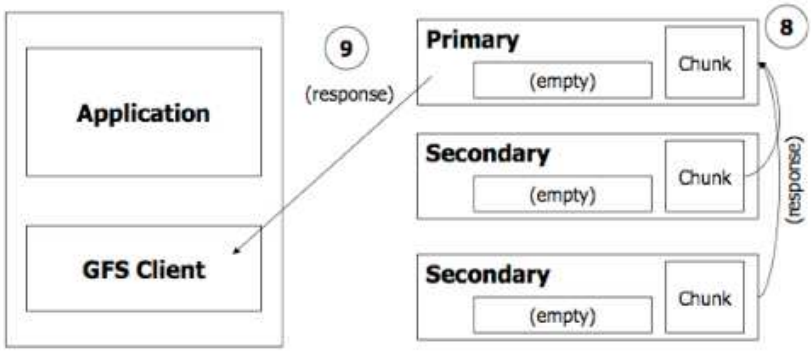
4. Client pushes write data to all locations. Data is stored in chunkserver's internal buffers



5. Client sends write command to primary
6. Primary determines serial order for data instances in its buffer and writes the instances in that order to the chunk
7. Primary sends the serial order to the secondaries and tells them to perform the write



8. Secondaries respond back to primary
9. Primary responds back to the client



Record Append Operation

GFS appends it to the file atomically at least once

GFS picks the offset

Works for concurrent writers

Used heavily by Google apps

e.g., for files that serve as multiple-producer/single-consumer queues

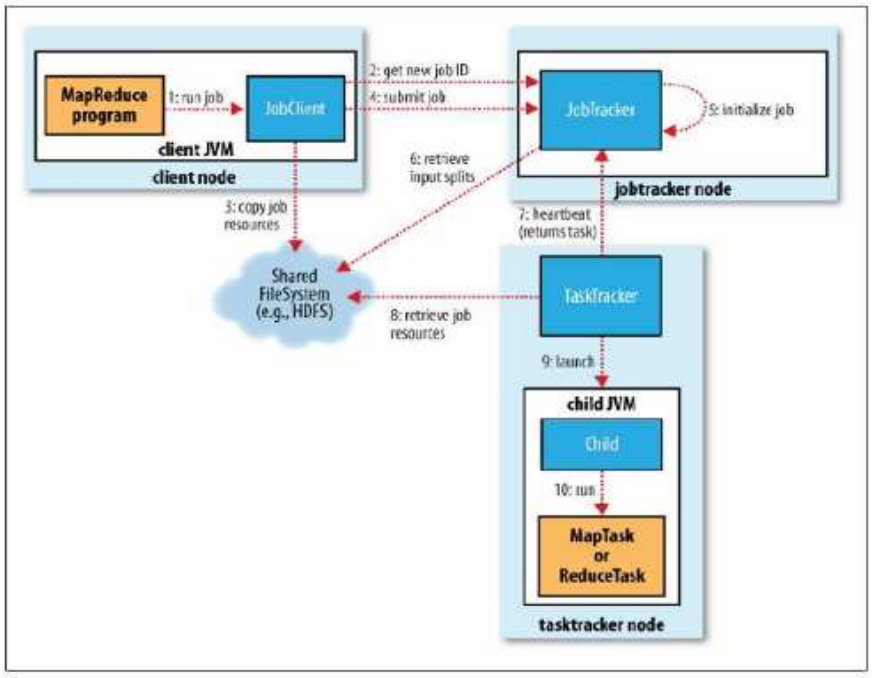
Merge results from multiple machines into one file.

Record Append

Same as write, but no offset and...

1. Client pushes write data to all locations
2. Primary checks if record fits in specified chunk
3. If the record does not fit:
    1. Pads the chunk
    2. Tells secondary to do the same
    3. Informs client and has the client retry
4. If record fits, then the primary:
    1. Appends the record
    2. Tells secondaries to do the same
    3. Receives responses and responds to the client

8. (a) Discuss in detail about how a job is executed in Hadoop Framework. (CO3) (K2) (10)



(OR)

(b) Write about the scope of grid computing in different applications. (CO1) (K3) (10)

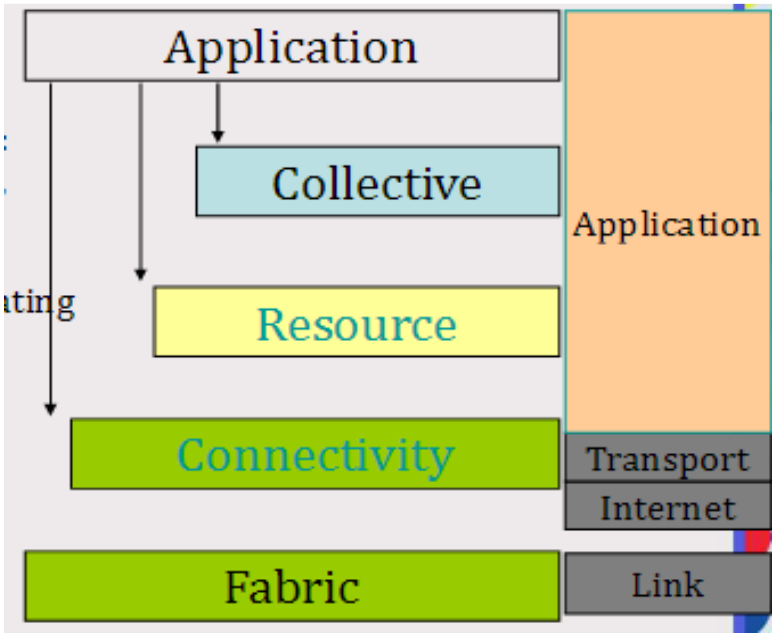
1. Life Sciences
2. Financial Analysis and Services
3. Research Collaboration
4. Engineering and Design
5. Engineering and Design
6. Collaborative Games
7. Government

9. (a) Summarize the technologies available in grid standards. (CO1) (K2) (4)

Web service
OGSA
OGSI
WSRF
OGSA - DAI

(OR)

(b) Illustrate in detail about the various layers in grid architecture. (CO1) (K2) (4)



1. Fabric Layer: Interface to Local Resources
2. Connectivity Layer: Manages Communications
3. Resource Layer: Sharing of a Single Resource
4. The Collective Layer: Coordinating Multiple Resources
5. Application Layer: User-Defined Grid Applications