

# CHAPTER 21



## Parallel and Distributed Storage

### Practice Exercises

- 21.1** In a range selection on a range-partitioned attribute, it is possible that only one disk may need to be accessed. Describe the benefits and drawbacks of this property.
- 21.2** Recall that histograms are used for constructing load-balanced range partitions.
- Suppose you have a histogram where values are between 1 and 100, and are partitioned into 10 ranges, 1 - 10, 11 - 20, ..., 91 - 100, with frequencies 15, 5, 20, 10, 10, 5, 5, 20, 5, and 5, respectively. Give a load-balanced range partitioning function to divide the values into five partitions.
  - Write an algorithm for computing a balanced range partition with  $p$  partitions, given a histogram of frequency distributions containing  $n$  ranges.
- 21.3** Histograms are traditionally constructed on the values of a specific attribute (or set of attributes) of a relation. Such histograms are good for avoiding data distribution skew but are not very useful for avoiding execution skew. Explain why.
- Now suppose you have a workload of queries that perform point lookups. Explain how you can use the queries in the workload to come up with a partitioning scheme that avoids execution skew.
- 21.4** Replication:
- Give two reasons for replicating data across geographically distributed data centers.
  - Centralized databases support replication using log records. How is the replication in centralized databases different from that in parallel/distributed databases?
- 21.5** Parallel indices:

- a. Secondary indices in a centralized database store the record identifier. A global secondary index too could potentially store a partition number holding the record, and a record identifier within the partition. Why would this be a bad idea?
  - b. Global secondary indices are implemented in a way similar to local secondary indices that are used when records are stored in a B<sup>+</sup>-tree file organization. Explain the similarities between the two scenarios that result in a similar implementation of the secondary indices.
- 21.6** Parallel database systems store replicas of each data item (or partition) on more than one node.
- a. Why is it a good idea to distribute the copies of the data items allocated to a node across multiple other nodes, instead of storing all the copies in the same node (or set of nodes)?
  - b. What are the benefits and drawbacks of using RAID storage instead of storing an extra copy of each data item?
- 21.7** Partitioning and replication.
- a. Explain why range-partitioning gives better control on tablet sizes than hash partitioning. List an analogy between this case and the case of B<sup>+</sup>-tree indices versus hash indices.
  - b. Some systems first perform hashing on the key, and then use range partitioning on the hash values. What could be a motivation for this choice, and what are its drawbacks as compared to performing range partitioning on the key?
  - c. It is possible to horizontally partition data, and then perform vertical partitioning locally at each node. It is also possible to do the converse, where vertical partitioning is done first, and then each partition is then horizontally partitioned independently. What are the benefits of the first option over the second one?
- 21.8** In order to send a request to the master replica of a data item, a node must know which replica is the master for that data item.
- a. Suppose that between the time the node identifies which node is the master replica for a data item, and the time the request reaches the identified node, the mastership has changed, and a different node is now the master. How can such a situation be dealt with?
  - b. While the master replica could be chosen on a per-partition basis, some systems support a *per-record master replica*, where the records of a partition (or tablet) are replicated at some set of nodes, but each record's

master replica can be on any of the nodes from within this set of nodes, independent of the master replica of other records. List two benefits of keeping track of master on a per-record basis.

- c. Suggest how to keep track of the master replica for each record, when there are a large number of records.

