1. Which of the following are true for traditional parallel databases

   a. A Database (or a relation) is fragmented and these fragments are distributed between the processing nodes' disk for parallel processing. Thus, each node will process the same SQL-Select query on the fragment stored on it's disk (assume the query is a Select with no join).

   b. Usually small number of processing nodes (not in the thousands) are used for parallel databases.

   c. Homogeneous set of processing nodes (same processor type, disks, etc.)

   e. It is not highly scalable because it is difficult to add another node.

2. Which of the following are true for HadoopDB/Map-Reduce parallel processing architecture

   a. Uses shared nothing architecture of the nodes (i.e., each node has it's own memory and disk and other nodes cannot access another node's memory and disk directly. These nodes can be heterogeneous (different hardware/software) communicating between them through message passing. However, message passing takes more time than communicating through hardwared interconnection connection.

   b. More nodes (i.e., more parallel processing) can be added easily (e.g., with increasing amount of data) because it is a shared nothing architecture.

   c. Nodes are not necessarily homogeneous. Thus, one node can be much slower than another, making the scheduling process dependent on the type of the nodes.

   d. Fault tolerance is important because of thousands of nodes.

   e. HadoopDB is a hybrid system taking advantage of parallel processing and distributed processing of Map-Reduce

3. Indicate which of the following are true.

   a. Analytical queries (e.g., data warehousing) are effectively implemented by HadoopDB because these queries works on very large data sets and the queries are read only. Fault tolerance is easy to handle because a node with a replication can be invoked when a node fails.

      True

   b. Write-queries like SQL INSERTs/DELETEs are not handled by HadoopDB because it is difficult to handle fault tolerance when a node fails.

False

4. In what way are HadoopDB and parallel databases related?

   a. HadoopDB extends an existing parallel database implementation.

   b. HadoopDB builds a parallel database using shared nothing architecture.

5. How is HadoopDB related to MapReduce?

   a. HadoopDB uses Hadoop's open source implementation of MapReduce to build a database.

   b. HadoopDB uses MapReduce to store the database data.