

Quiz Assignment-VII Solutions: Distributed Systems (Week-7)

Q. 1 The number of maps is usually driven by the total size of _____

- A. tasks
- B. inputs
- C. outputs
- D. None of these

Ans: B) inputs

Explanation: Map, written by the user, takes an input pair and produces a set of intermediate key/value pairs.

Q. 2 _____ function is responsible for consolidating the results produced by each of the Map() functions/tasks.

- A. Reducer
- B. Map
- C. Reduce
- D. All of the mentioned

Ans: C) Reduce

Explanation: The MapReduce library groups together all intermediate values associated with the same intermediate key and passes them to the Reduce function.

Q. 3 The Mapreduce framework groups Reducer inputs by key in _____ stage.

- A. sort
- B. shuffle
- C. reduce
- D. None of these

Ans: A) Sort

Explanation: When a reduce worker has read all intermediate data, it sorts it by the intermediate keys so that all occurrences of the same key are grouped together. The sorting is needed because typically many different keys map to the same reduce task. If the amount of intermediate data is too large to fit in memory, an external sort is used.

Q. 4 In the local disk of the HDFS namenode the files which are stored persistently are:

- A. Namespace image and edit log
- B. Block locations and namespace image
- C. Edit log and block locations
- D. Namespace image, edit log and block locations

Ans: A) Namespace image and edit log

Q. 5 In HDFS, application data are stored on other servers called _____. All servers are _____ and communicate with each other using _____ protocols.

- A. NameNodes, fully connected, TCP-based
- B. DataNodes, fully connected, TCP-based
- C. NameNodes, loosely connected, UDP-based
- D. DataNodes, loosely connected, UDP-based

Ans: B) DataNodes, fully connected, TCP-based

Q. 6 In Spark, A _____ is a read-only collection of objects partitioned across a set of machines that can be rebuilt if a partition is lost.

- A. Spark Streaming
- B. FlatMap
- C. Driver
- D. Resilient Distributed Dataset (RDD)

Ans: D) Resilient Distributed Dataset (RDD)

Explanation: Resilient Distributed Dataset (RDD) is a read-only collection of objects partitioned across a set of machines that can be rebuilt if a partition is lost.

Q. 7 In Spark, a dataset with elements of type A can be transformed into a dataset with elements of type B using an operation called _____, which passes each element through a user-provided function of type $A \rightarrow \text{List}[B]$.

- A. Map
- B. Filter
- C. flatMap
- D. Reduce

Ans: C) flatMap

Explanation: A dataset with elements of type A can be transformed into a dataset with elements of type B using an operation called flatMap

Q. 8 Spark also allows programmers to create two restricted types of shared variables to support two simple but common usage patterns such as _____ and _____.

- A. Map, Reduce
- B. Flatmap, Filter
- C. Sort, Shuffle
- D. Broadcast, Accumulators

Ans: D) Broadcast, Accumulators

Explanation: Spark also lets programmers create two restricted types of shared variables to support two simple but common usage patterns such as Broadcast variables and Accumulators.

Spark lets the programmer create a “broadcast variable” object that wraps the value and ensures that it is only copied to each worker once.

Accumulators: These are variables that workers can only “add” to using an associative operation, and that only the driver can read.

Q. 9 Consider the following statements:

Statement-1: Reactive routing protocols ask each host (or many hosts) to maintain global topology information, thus a route can be provided immediately when requested.

Statement-2: Proactive routing protocols have the feature on-demand. Each host computes route for a specific destination only when necessary.

- A. Both statements are false
- B. Both statements are true
- C. Statement 1 is true and statement 2 is false
- D. Statement 1 is false and statement 2 is true

Ans: A) Both statements are false

Explanation: (i) Proactive routing protocols ask each host (or many hosts) to maintain global topology information, thus a route can be provided immediately when requested. But large amount of control messages are required to keep each host updated for the newest topology changes.

(ii) Reactive routing protocols have the feature on-demand. Each host computes route for a specific destination only when necessary. Topology changes which do not influence active routes do not trigger any route maintenance function, thus communication overhead is lower compared to proactive routing protocol.

Q. 10 Most of the on-demand routing protocols use _____ for route discovery. Flooding suffers from _____ problem.

- A. Broadcast, Flooding
- B. Shortest-path, Convergecast
- C. Flooding, Broadcast storm
- D. Broadcast, Convergecast

Ans: C) Flooding, Broadcast storm

Explanation: On-demand routing protocols attract much attention due to their better scalability and lower protocol overhead. But most of them use flooding for route discovery. Flooding suffers from broadcast storm problem.

Broadcast storm problem refers to the fact that flooding may result in excessive redundancy, contention, and collision. This causes high protocol overhead and interference to other ongoing communication sessions.
