2023-CSE5BDC-T5-W - Big Data Management on the Cloud

Big Data Management on the Cloud Week 3. Quizzes or Notes

**Quiz: Check your understanding 1**

**YARN runs a daemon (NodeManager) on each slave node, and a master process (ResourceManager) on the master node. Each NodeManager reports to the ResourceManager how much compute resource and memory are available on each slave node.**

True

**Architecturally, where is the YARN layer in Hadoop?**

Between MapReduce and HDFS

**Which of the following is correct?**

ResourceManager accepts job request (code) from users, initiates application masters and allocates resources

## Quiz: Check your understanding 2

## Which problems of batch processing does Storm solve? (Choose two of the following.)

Batch processing system like MapReduce need to load the data from disk (HDFS) and store the results back on HDFS. This results in a lot of slow Disk IO which is too slow to handle data streams.

Batch processing has a high turnaround time, because it is designed to process large batches of data at once. Thus, requiring the user to wait a long time before it is all process. Hence it is not suitable for processing data streams.

A **sprout** can be imagined as a data adapter that converts the data source into a stream of **tuples**. A **bolt** consumes multiple streams from sprouts and/or other bolts, and then performs intermediate **transformation**, such as filtering,

**aggregation**, read and write, and joins, to emit new streams of data. A

**topology** depicts a network of sprouts and bolts, where a bolt can subscribe to a set of streams from sprouts or each other, and the last layer of bolts are called sinks.

## Quiz: Check your understanding 3

Which of the following is a problem of MapReduce PageRank that Pregel solves? (Choose two answers.)

- The entire state of the graph is shuffled across the network on every iteration instead of shuffling only the new rank contributions.

- The entire graph structure needs to be loaded and stored to disk during each processing iteration.

When implementing page rank using Pregel’s think like a vertex approach, which of the following operations is not correct? (Choose two answers.)

* The program for the current vertex sends its page rank contribution to all nodes in the graph in one step.
* The program for the current vertex updates its page rank with the maximum page rank contributions from its neighbours.