Big Data And Hadoop

Assignment 12.1

Problem Statement: In this assignment you need to select the correct answer for the given questions.

1. What is in-memory processing in Spark?

> a) Processing data in each node
>
> b) Storing data in RDBMS during processing
>
> c) Maximum effective utilization of RAM during the processing
>
> d) Using more no of CPU threads

Answer: **Option c** – Maximum effective utilization of RAM during the processing. Instead of dumping intermediate outputs on disk, Apache Spark caches it in memory, hence, results in optimization of performance.

----

2. What are the features of Apache Spark?

> a) In-memory processing
>
> b) Ease of use APIs
>
> c) Unified high level tools
>
> d) Runs Everywhere (Hadoop, Mesos, standalone, or in the cloud. It can access diverse data sources including HDFS, Cassandra, HBase, S3.)
>
> e) All the above
>
> f) None of the above

Answer: **Option e** – All the above.

Explanation: Each of the options (a) to (d) hold true for Spark. It uses in-memory caching and has a rich set of APIs in Java, Python, etc. Also it supports a variety of cluster managers like Mesos, Hadoop Yarn, etc

----

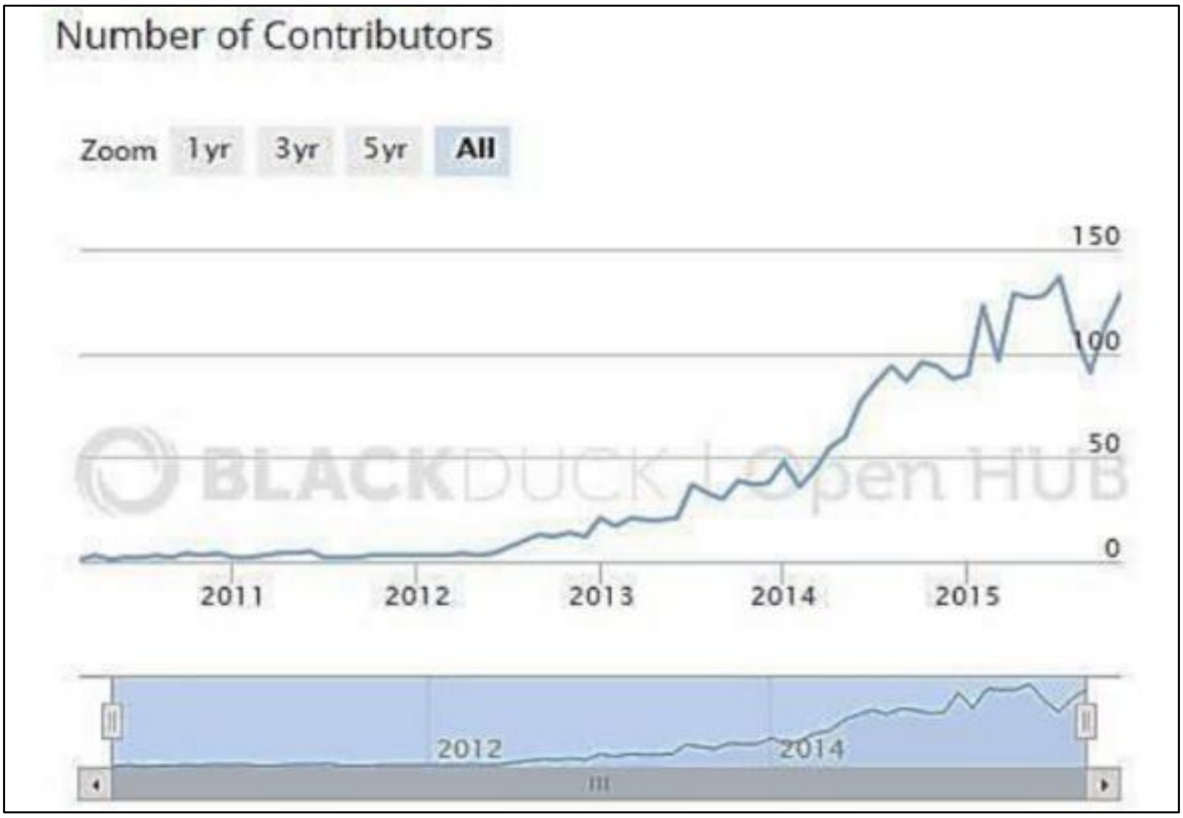
3. Apache Hadoop is more active project than Apache Spark in open source community in the last year.

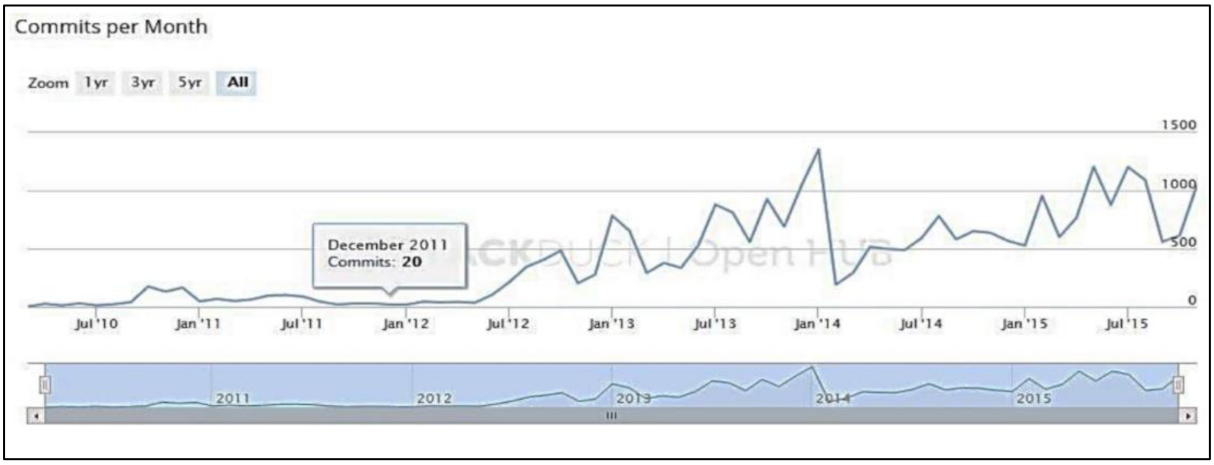
> a) true

b) false

Answer: **Option b** – false.

Explanation: Apache Spark is the most active project.





----

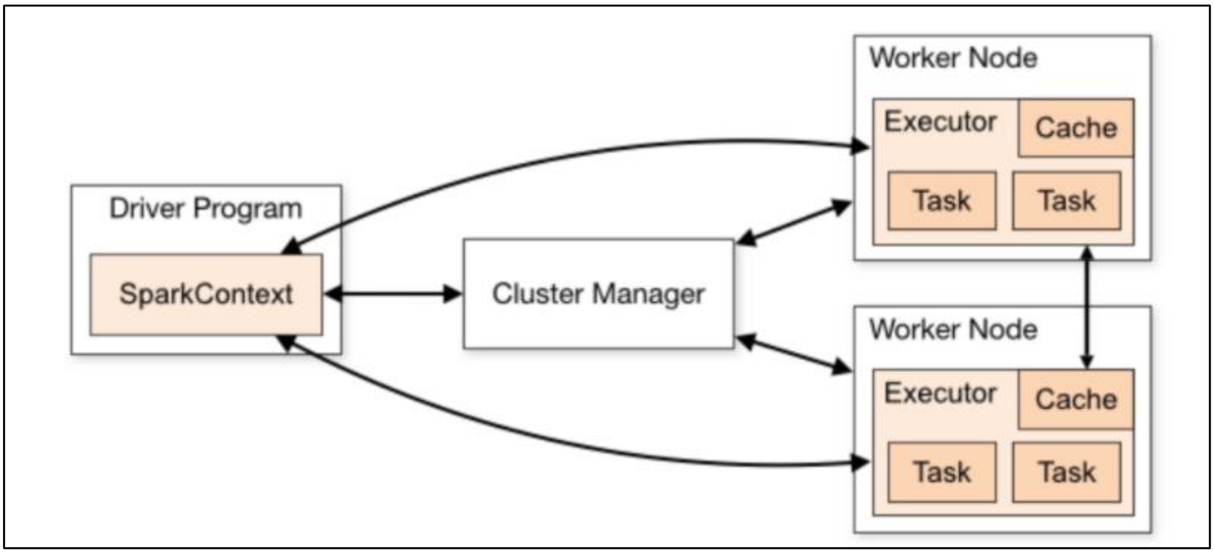
4. Driver program will be launched in every node of the worker.

a) true

b) false

Answer: **Option b** – false.

Explanation: Driver program is not launched on worker node.



----

5. Spark only supports Stream processing.

a. Yes

b. No

Answer: **Option b** – No.

Explanation: Spark supports SQL, R, Machine Learning and Graph computation along with Stream processing.