

**1. The general processing of MapReduce includes the steps...**

- A Attaching chunks from a DFS to a map task which turns each chunk into a key-value pair sequence.
- B Continuous synchronisation of data nodes during the map step.
- C Collecting key-value pairs by a master node which are sorted by key and dividing the keys among all reducer tasks.
- D On each key the reduce tasks work separately and combine all the values associated with a specific key.

**2. The characteristics of data storage operations on HDFS are...**

- A Write Once, Read Often Model
- B Individual files content cannot be modified but new data can be appended to the end of a file
- C Write Often, Read Often Model
- D Modifications of individual files content are possible if the files are locked for the period of modification

**3. What are some important differences between Hadoop and Spark?**

- A Spark writes data between operations on RAM rather than HDD, which makes it orders of magnitude faster than Hadoop.
- B Spark offers high-level transformations (e.g. filter, join) which makes it much easier to use than Hadoop.
- C In contrast to HDFS files, RDDs in Spark cannot be replicated across multiple nodes.
- D Spark can cache results in memory, which is not possible in Hadoop.

**4. What are possible operations on RDDs in Spark?**

- A Lazily evaluated Transformations (e.g. map, reduce, join)
- B Actions (e.g. collect, print) which trigger Transformations
- C Lazily evaluated Actions (e.g. collect, print)
- D Transformations (e.g. map, reduce, join) which trigger Actions (e.g. collect, print)