

## Quiz: 03 Spark Intro

員編 : \_\_\_\_\_

1. Which Transformation is involved in data shuffling between partitions?  
(A) map() (B) flatMap() (C) reduceByKey()
2. Which one of the following is not transformation function?  
(A) join() (B) sortBy() (C) top()
3. Jobs are divided into "stages" based on the \_\_\_\_\_ ?  
(A) shuffle boundary (B) narrow boundary (C) pipelineable boundary
4. In Hadoop ecosystem, which one is most likely to be the cluster manager of spark application?  
(A) Yarn (B) Kubernetes (C) Mesos
5. Which snippet of code is in meaning to save dataframe as a Hive Table?  
(A) `df.write.format("csv").save("demo")`  
(B) `df.write.format("parquet").saveAsTable("demo")`  
(C) None of above



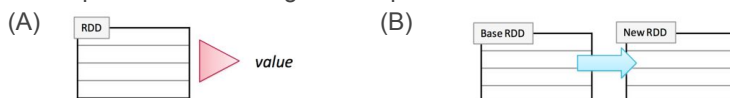
© B0A7-B0BB Athemaster Co. All Rights Reserved

2

6. With `df.na.drop(thresh=C)`, which row would be drop?  
(A) Aewergrqa Mary CA null null  
(B) Cfarffafq4B Joe null null null  
(C) All of above



7. Which picture is illustrating a concept of Action.



8. Which one is not the supported languages by Apache Spark.  
(A) Java (B) Python (C) Go
9. Spark applications run as independent sets of processes on a cluster, coordinated by the SparkContext object in main program. which one of the following is the SparkContext?  
(A) executor (B) container (C) driver
10. \_\_\_\_, the smallest units of work for Spark, in a job can be executed in parallel.  
(A) jobs (B) tasks (C) stages

© B0A7-B0BB Athemaster Co. All Rights Reserved

3