## Apache Spark and parallel data processing

Pregunta 1:

1. What is the key concept of Apache Spark

- (●) Writing code that can be executed in parallel on multiple cluster machines
- ( ) Writing code that runs very fast on a single CPU because Apache Spark optimizes it's assembly language

Pregunta 2:

2. How does Apache Spark achieve parallelization?

- (●) By executing functions in parallel on RDDs
- ( ) By executing queries in parallel on multiple relational databases

Pregunta 3:

3. How does an Apache Spark program look like?

- ( ) It contains code that controls distribution of workload to different cluster members
- (●) Code uses either the RDD or DataFrame API or SQL. All parallelization aspects are hidden from the user and being taken care of by Apache Spark

Pregunta 4:

4. What does BigData mean?

- ( ) Any dataset over 1 TB
- (●) Any dataset which exeeds processing capabilities of a single compute node