

University of California, Santa Cruz - Silicon Valley Extension
Class: DBDA.X407 Hadoop: Distributed Processing of Big Data

Full Name: _____

===== Midterm =====

8 Questions - 40 points

a1. Select the one that **IS NOT** one the 5 V's of Big Data? (5 points)

- a. Velocity.
- b. Variety.
- c. Volume
- d. Veracity.
- e. Validity.

a2. How the Spark context is used in a Spark program? (5 points)

- a. It allows storing cached variables.
- b. Is the handle to connect to a Spark cluster.
- c. It's used as a transformation.
- d. It's used as a type of RDD.
- e. None of the above.

a3. What does the command **distcp** do? (5 points)

- a. Allows to run distributed jobs in the cluster.
- b. Sets the limits on storage in HDFS.
- c. It's a utility program used to perform parallel copy between clusters.
- d. This program is deprecated and it's not used Hadoop anymore.
- e. None of the above.

a4. What's the benefit of Hadoop Job Queues? (5 points)

- a. They make jobs more secure.
- b. They define resource allocation rules and allow several groups to concurrently use the cluster.
- c. By queuing the jobs, Hadoop can make sure only one job will be running at any point in time.
- d. Job Queues are used to store logs about the Hadoop jobs.
- e. None of the above.

a5. How the Attempts, Jobs and Tasks are related? (5 points)

- a. A Job contain an attempt and each attempt is made of tasks.
- b. A job is made of tasks and each task can contain one or more attempts.
- c. An attempt represents a full try of Job execution including its tasks.
- d. Jobs contains tasks and attempts are not related to either jobs or tasks.
- e. None of the above.

a6. a Spark function named 'my_operation' receives an RDD as input and returns an RDD.
What can you say about this function? (5 points)

- a. This function is an action.
- b. This function is a transformation.
- c. It can be an action or transformation.
- d. It's not an action nor a transformation just a function.
- e. None of the above.

a7. Explain in your own words what is a Spark RDD. (1-2 paragraphs). (10 points)
