



! Try again once you are ready
TO PASS 80% or higher

Try again

GRADE
61.90%

Apache Spark and parallel data processing

LATEST SUBMISSION GRADE

61.9%

1.

```
1 rdd = sc.parallelize(range(100))
2 rdd2 = range(100)
3
```

0 / 1 point

Please consider the following code.

Where is the execution of API calls on "rdd" taking place?

- ☐ In the ApacheSpark worker nodes
- ☒ On the local Driver machine

! Incorrect

Incorrect. Please revisit Video "Parallel data processing strategies of ApacheSpark"

2.

```
1 rdd = sc.parallelize(range(100))
2 rdd2 = range(100)
3
```

0 / 1 point

Please consider the following code.

Where is data in "rdd2" stored physically?

- ☒ In main-memory of ApacheSpark worker nodes
- ☐ On the local Driver machine

! Incorrect

Incorrect. Please revisit Video "Parallel data processing strategies of ApacheSpark"

3. What is the parallel version of the following code?

1 / 1 point

```
1 len(range(9999999999))
2
```

- ☒ sc.parallelize(range(9999999999)).count()
- ☐ parallelize(range(9999999999)).count()
- ☐ len(sc.parallelize(range(9999999999)))
- ☐ size(sc.parallelize(range(9999999999)))
- ☐ count(sc.parallelize(range(9999999999)))

✓ Correct
Correct

4. Which storage solutions support seamless modification of schemas? (Select all that apply)

0.667 / 1 point

- ☐ ObjectStorage
- ☒ NoSQL

✓ Correct
Correct

☐ SQL/Relational Databases

You didn't select all the correct answers

5. Which storage solutions support dynamic scaling on storage? (Select all that apply)

0.667 / 1 point

- ☒ ObjectStorage

✓ Correct
Correct

- ☐ NoSQL
- ☐ SQL/Relational Databases

You didn't select all the correct answers

6. Which storage solutions support normalization and integrity checks on data out of the box? (Select all that apply)

1 / 1 point

- ☐ ObjectStorage
- ☐ NoSQL
- ☒ SQL/Relational Databases

✓ Correct
Correct

7. What is the advantage of using ApacheSparkSQL over RDDs? (select all that apply)

1 / 1 point

- ☐ ApacheSparkSQL bypasses the RDD interface which has been proven to be very complicated
- ☐ SQL is simpler than RDD but has some performance drawbacks
- ☒ Catalyst and Tungsten are able to optimise the execution, so are more likely to execute more quickly than if you would had implemented something equivalent using the RDD API.

✓ Correct
Correct

- ☒ The API is simpler and doesn't require specific functional programming skills

✓ Correct
Correct