Congratulations! You passed!
TO PASS 80% or higher

Keep Learning

GRADE
90.47%

Apache Spark and parallel data processing

LATEST SUBMISSION GRADE
90.47%

1. 1 / 1 point

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

Please consider the following code.

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

- ○ On the local Driver machine
- ◉ In the ApacheSpark worker nodes

Correct
Correct

2. 1 / 1 point

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

Please consider the following code.

Where is data in " **rdd2** " stored physically?

- ○ In main-memory of ApacheSpark worker nodes
- ◉ On the local Driver machine

Correct
Correct

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
- [x] 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
- [x] 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

- [x] The API is simpler and doesn't require specific functional programming skills

✓ **Correct**
Correct