

Welcome to exercise one of “Apache Spark for Scalable Machine Learning on BigData”. In this exercise you’ll apply the basics of functional and parallel programming.

Let’s start with a simple example. Let’s consider you have a list of integers.

Let’s find out what the size of this list is.

Note that we already provide an RDD object, so please have a look at the RDD API in order to find out what function to use: <https://spark.apache.org/docs/latest/api/python/pyspark.html#pyspark.RDD>

The following link contains additional documentation: <https://spark.apache.org/docs/latest/rdd-programming-guide.html>

```
In [ ]: rdd = sc.parallelize(range(100))
```

```
In [ ]: # please replace $$ with the correct characters
rdd.c$$$t()
```

You should see "100" as answer. Now we want to know the sum of all elements. Please again, have a look at the API documentation and complete the code below in order to get the sum.

```
In [ ]: rdd.s
```

You should get "4950" as answer.