DataFrames

Objective

Understand DataFrames, Rows and schemas

Read DataFrames from files

Create DataFrames manually

Recap

Creating a SparkSession

```scala
val spark = SparkSession.builder()
  .appName("My Application")
  .config("spark.master", "local")
  .getOrCreate()
```

allows creating, reading and writing DataFrames

Reading a DataFrame from a file

```scala
val firstDF = spark.read()
  .format("json")
  .option("inferSchema", "true")
  .load("path/to/my/file.json")
```

```scala
firstDF.show()
firstDF.printSchema()
firstDF.take(5)
```

Rows = unstructured data

Schema = description of fields aka <u>columns</u> and their type

Spark types

Spark rocks