RDDs are a low-level abstraction of the data. In the first version of Spark, you worked directly with RDDs. You can think of RDDs as long lists distributed across various machines. You can still use RDDs as part of your Spark code although data frames and SQL are easier. This course won't go into the details of RDD syntax, but you can find some further explanation of the difference between RDDs and DataFrames in Databricks' **A Tale of Three Apache Spark APIs: RDDs, DataFrames, and Datasets** blog post.

Here is a link to the Spark documentation's **RDD programming guide**.