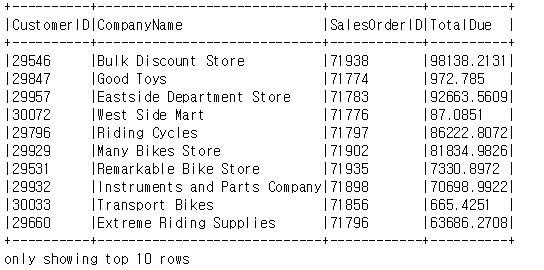
Group5-Github Relation

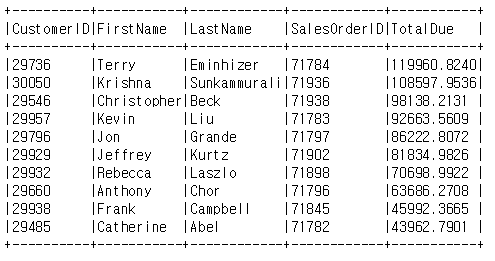
RDD vs Dataframe vs SQL

강준후, 손덕인, 조주영

1. Data:
   1. SalesLTProduct.txt: 15 features, 295 rows
   2. SalesLTSalesOrderHeaders.txt: 22 features, 33rows
   3. SalesLTCustomer.txt: 15 features, 848rows
2. Environment:
   1. Spark 2.3
   2. Hadoop 2.7.5
3. Process:
   1. Select some features that we will use and map(make) lists from several files.



* 1. Do innerjoin to put data together.



* 1. Check that execution time by using time.time()



1. Result:

* RDD is fastest among them. It is 3 times faster than second faster one.
* But Dataframe and SQL code are easier to understand than RDD.