

In [0]:



```
student_records_df.show(truncate=False)
```

```
+-----+-----+-----+-----+-----+-----+
|roll_no|name      |passed|marks                                     |s
ports      |DoB              |
+-----+-----+-----+-----+-----+-----+
|1       |John Doe  |true  |{Chemistry -> 81, Math -> 89, Physics -> 87}|
[chess, football]      |2012-05-01 13:01:05|
|2       |John Smith|false |{Chemistry -> 36, Math -> 29, Physics -> 31}|
[volleyball, tabletennis]|2012-05-12 14:02:05|
+-----+-----+-----+-----+-----+-----+
|-----+-----+-----+-----+-----+-----+
|
```

In [0]:



```
student_records_df.createOrReplaceTempView('records')
```

In [0]:



```
spark.sql("SELECT * FROM records").show()
```

```
+-----+-----+-----+-----+-----+-----+
|roll_no|name|passed|marks                                     |sports|
DoB|
+-----+-----+-----+-----+-----+-----+
|1       |John Doe| true|{Chemistry -> 81,...| [chess, football]|2012
-05-01 13:01:05|
|2       |John Smith| false|{Chemistry -> 36,...|[volleyball, tabl...|2012
-05-12 14:02:05|
+-----+-----+-----+-----+-----+-----+
|-----+-----+-----+-----+-----+-----+
|
```

In [0]:



```
spark.sql('SELECT roll_no, marks["Physics"], sports[1] FROM records').show()
```

```
+-----+-----+-----+
|roll_no|marks[Physics]| sports[1]|
+-----+-----+-----+
|1       |87| football|
|2       |31| tabletennis|
+-----+-----+-----+
```

In [0]:



```
spark.sql("SELECT * FROM records where passed = True").show()
```

```
+-----+-----+-----+-----+-----+-----+
-----+
|roll_no|    name|passed|          marks|          sports|
DoB|
+-----+-----+-----+-----+-----+-----+
-----+
|      1|John Doe|  true|{Chemistry -> 81,...|[chess, football]|2012-05-0
1 13:01:05|
+-----+-----+-----+-----+-----+-----+
-----+
```

In [0]:



```
spark.sql('SELECT * FROM records WHERE
          marks["Chemistry"] < 40').show()
```

```
+-----+-----+-----+-----+-----+-----+
-----+
|roll_no|    name|passed|          marks|          sports|
DoB|
+-----+-----+-----+-----+-----+-----+
-----+
|      2|John Smith| false|{Chemistry -> 36,...|[volleyball, tabl...|2012
-05-12 14:02:05|
+-----+-----+-----+-----+-----+-----+
-----+
```

In [0]:



```
# Creating Global View
```

In [0]:



```
student_records_df.createGlobalTempView('global_record')
```

In [0]:



```
spark.sql("SELECT * FROM global_temp.global_records").show()
```

In [0]:



```
# Dropping Columns from DataFrame
```

In [0]:

▶▶

```
student_records_df.columns
```

```
Out[19]: ['roll_no', 'name', 'passed', 'marks', 'sports', 'DoB']
```

In [0]:

▶▶

```
student_records_df = student_records_df.drop('passed')
```

In [0]:

▶▶

```
# Few More Queries
```

In [0]:

▶▶

```
spark.sql("SELECT round( (marks.Physics+marks.Chemistry+marks.Math)/3) avg_marks FROM re
```

```
+-----+
|avg_marks|
+-----+
|      86.0|
|      32.0|
+-----+
```

In [0]:

▶▶

```
student_records_df=spark.sql("SELECT *, round( (marks.Physics+marks.Chemistry+marks.Math
student_records_df.show()
```

```
+-----+-----+-----+-----+-----+-----+-----+
-----+-----+
|roll_no|    name|passed|          marks|          sports|
DoB|avg_marks|
+-----+-----+-----+-----+-----+-----+-----+
-----+-----+
|      1| John Doe|  true|{Chemistry -> 81,...| [chess, football]|2012
-05-01 13:01:05|      86.0|
|      2|John Smith| false|{Chemistry -> 36,...|[volleyball, tabl...|2012
-05-12 14:02:05|      32.0|
+-----+-----+-----+-----+-----+-----+-----+
-----+-----+
```

In [0]:

▶▶

```
student_records_df.createOrReplaceTempView('records')
```

In [0]:



```
student_records_df = student_records_df.withColumn('status', (when(col('avg_marks') >= 40,
student_records_df.show()
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|roll_no|      name|passed|      marks|      sports|
DoB|avg_marks|status|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|      1| John Doe|  true|{Chemistry -> 81,...| [chess, football]|2012
-05-01 13:01:05|      86.0|passed|
|      2|John Smith| false|{Chemistry -> 36,...|[volleyball, tabl...|2012
-05-12 14:02:05|      32.0|failed|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
```

In [0]:



```
# another table
```

In [0]:



```
employeeData = (('John', 'HR', 'NY', 90000, 34, 10000),
('Neha', 'HR', 'NY', 86000, 28, 20000),
('Robert', 'Sales', 'CA', 81000, 56, 22000),
('Maria', 'Sales', 'CA', 99000, 45, 15000),
('Paul', 'IT', 'NY', 98000, 38, 14000),
('Jen', 'IT', 'CA', 90000, 34, 20000),
('Raj', 'IT', 'CA', 93000, 28, 28000),
('Pooja', 'IT', 'CA', 95000, 31, 19000))
columns = ('employee_name', 'department', 'state', 'salary', 'age', 'bonus')
```

In [0]:



```
employeeDf = spark.createDataFrame(employeeData, columns)
```

In [0]:



```
employeeDf.groupby(col('department')).agg(sum(col('salary'))).show()
```

```
+-----+-----+
|department|sum(salary)|
+-----+-----+
|      HR|      176000|
|    Sales|      180000|
|      IT|      376000|
+-----+-----+
```

In [0]:



```
employeeDf.groupby(col('department')).agg(sum(col('salary')).alias('total_sal')).orderBy
```

```
+-----+-----+
|department|total_sal|
+-----+-----+
|      HR|    176000|
|    Sales|    180000|
|      IT|    376000|
+-----+-----+
```

In [0]:



```
employeeDf.groupby(col('department')).agg(sum(col('salary')).alias('total_sal')).orderBy
```

```
+-----+-----+
|department|total_sal|
+-----+-----+
|      IT|    376000|
|    Sales|    180000|
|      HR|    176000|
+-----+-----+
```

In [0]:



```
employeeDf.groupby(col('department'),col('state')).agg(sum(col('bonus'))).show()
```

```
+-----+-----+-----+
|department|state|sum(bonus)|
+-----+-----+-----+
|      HR|   NY|     30000|
|    Sales|   CA|     37000|
|      IT|   NY|     14000|
|      IT|   CA|     67000|
+-----+-----+-----+
```

In [0]:



```
employeeDf.groupby(col('department')).agg(avg(col('salary')).alias('avarage_salary'),max
```

```
+-----+-----+-----+
|department|avarage_salary|maximum_bonus|
+-----+-----+-----+
|      HR|      88000.0|         20000|
|    Sales|     90000.0|         22000|
|      IT|     94000.0|         28000|
+-----+-----+-----+
```

In [0]:

In [0]:

In [0]: