

Lab BDA2 – Spark SQL – Exercises

Task 1

minimum

year	station	value
1990	147270	-35.0
1990	166870	-35.0
1952	192830	-35.5
1974	166870	-35.6
1974	179950	-35.6
1954	113410	-36.0
1992	179960	-36.1
1975	157860	-37.0
1972	167860	-37.5
2000	169860	-37.6
1995	182910	-37.6
1957	159970	-37.8
1989	166870	-38.2
1983	191900	-38.2
1953	183760	-38.4
2009	179960	-38.5
1993	191900	-39.0
1984	191900	-39.2
1984	123480	-39.2
2008	179960	-39.3

only showing top 20 rows

maximum

year	station	value
1975	86200	36.1
1992	63600	35.4
1994	117160	34.7
2014	96560	34.4
2010	75250	34.4
1989	63050	33.9
1982	94050	33.8
1968	137100	33.7
1966	151640	33.5

2002	78290	33.3
2002	78290	33.3
1983	98210	33.3
1986	76470	33.2
1970	103080	33.2
1956	145340	33.0
2000	62400	33.0
1959	65160	32.8
2006	75240	32.7
1991	137040	32.7
1988	102540	32.6

+-----+-----+-----+

only showing top 20 rows

Task 2

SQL

year	month	value
2014	7	147681
2011	7	146656
2010	7	143419
2012	7	137477
2013	7	133657
2009	7	133008
2011	8	132734
2009	8	128349
2013	8	128235
2003	7	128133
2002	7	127956
2006	8	127622
2008	7	126973
2002	8	126073
2005	7	125294
2011	6	125193
2012	8	125037
2006	7	124794
2010	8	124417
2014	8	124045

+-----+-----+-----+

only showing top 20 rows

with only one count per station:

+-----+-----+-----+

year	month	value
+-----+	+-----+	+-----+
1972	10	378
1973	6	377
1973	5	377
1972	8	376
1973	9	376
1972	5	375
1972	6	375
1972	9	375
1971	8	375
1971	9	374
1972	7	374
1971	6	374
1971	5	373
1973	8	373
1974	8	372
1974	6	372
1970	8	370
1974	9	370
1973	7	370
1971	7	370

+-----+
only showing top 20 rows

built-in API

year	month	count
+-----+	+-----+	+-----+
2014	7	147681
2011	7	146656
2010	7	143419
2012	7	137477
2013	7	133657
2009	7	133008
2011	8	132734
2009	8	128349
2013	8	128235
2003	7	128133
2002	7	127956
2006	8	127622
2008	7	126973
2002	8	126073
2005	7	125294
2011	6	125193
2012	8	125037

```
|2006|      7|124794|
|2010|      8|124417|
|2014|      8|124045|
+----+-----+-----+
only showing top 20 rows
```

with only one count per station:

```
+----+-----+-----+
|year|month|count|
+----+-----+-----+
|1972|  10|  378|
|1973|   6|  377|
|1973|   5|  377|
|1973|   9|  376|
|1972|   8|  376|
|1972|   6|  375|
|1971|   8|  375|
|1972|   9|  375|
|1972|   5|  375|
|1972|   7|  374|
|1971|   9|  374|
|1971|   6|  374|
|1973|   8|  373|
|1971|   5|  373|
|1974|   8|  372|
|1974|   6|  372|
|1970|   8|  370|
|1973|   7|  370|
|1974|   5|  370|
|1974|   9|  370|
+----+-----+-----+
only showing top 20 rows
```

Task 3

```
+----+-----+-----+-----+
|year|month|station|avgMonthlyTemperature|
+----+-----+-----+-----+
|2014|   7|  96000|                26.3|
|1994|   7|  96550|  23.071052631578947|
|1983|   8|  54550|                23.0|
|1994|   7|  78140|  22.970967741935485|
|1994|   7|  85280|  22.87258064516129|
|1994|   7|  75120|  22.858064516129037|
|1994|   7|  65450|  22.856451612903232|
```

1994	7	96000	22.808064516129033
1994	7	95160	22.764516129032252
1994	7	86200	22.711290322580645
2002	8	78140	22.7
1994	7	76000	22.6983870967742
1997	8	78140	22.66612903225807
1994	7	105260	22.659677419354843
1975	8	54550	22.642857142857142
2006	7	76530	22.5983870967742
1994	7	86330	22.548387096774192
2006	7	75120	22.527419354838706
1994	7	54300	22.469354838709677
2006	7	78140	22.45806451612904

+-----+-----+-----+-----+-----+-----+

only showing top 20 rows

Task 4

station	maxTemp	maxDailyPrecipitation
-----	-----	-----
-----	-----	-----
-----	-----	-----

Task 5

year	month	avgMonthlyPrecipitation
2016	7	0.0
2016	6	47.6625
2016	5	29.250000000000007
2016	4	26.900000000000006
2016	3	19.962500000000006
2016	2	21.5625
2016	1	22.325000000000003
2015	12	28.925
2015	11	63.88750000000002
2015	10	2.2625
2015	9	101.3
2015	8	26.987499999999997
2015	7	119.09999999999994
2015	6	78.66250000000002
2015	5	93.225
2015	4	15.337499999999999
2015	3	42.61250000000001

```
|2015|      2|      24.825|
|2015|      1|    59.11250000000003|
|2014|     12|    35.46250000000001|
+-----+
only showing top 20 rows
```

Task 6

```
+-----+
|year|month|      difference|
+-----+
|2014|    12|  0.8238893537957012|
|2014|    11|  2.0635396726928983|
|2014|    10|  1.5225549840378179|
|2014|     9| 0.061058186437215056|
|2014|     8| -0.6426470719706945|
|2014|     7|  2.093910775893086|
|2014|     6| -1.807368619731525|
|2014|     5|  0.2671906501407033|
|2014|     4|  2.066193158991547|
|2014|     3|  4.4867483435745665|
|2014|     2|  5.420311314566042|
|2014|     1|  0.9325880207201753|
|2013|    12|  3.8796729966761174|
|2013|    11|  0.9342517939050201|
|2013|    10|  0.752906890196174|
|2013|     9| -1.0050575160421236|
|2013|     8| -0.3146412068680533|
|2013|     7| 0.008280277359361321|
|2013|     6| -0.5441868015497064|
|2013|     5|  1.5739208554192938|
+-----+
only showing top 20 rows
```