

Window functions

In [2]:

```
from pyspark.sql import SparkSession

spark = SparkSession \
    .builder \
    .appName("Window functions") \
    .getOrCreate()
```

In [3]:

```
products = spark.read\
    .format("csv")\
    .option("header", "true")\
    .load('../datasets/products.csv')
```

In [4]:

```
products.show()
```

```
+-----+-----+-----+
|  product|category|price|
+-----+-----+-----+
|Samsung TX|  Tablet|  999|
|Samsung JX|  Mobile|  799|
|Redmi Note|  Mobile|  399|
|      Mi|  Mobile|  299|
|      iPad| Tablet|  789|
|      iPhone| Mobile|  999|
|  Micromax|  Mobile|  249|
|   Lenovo| Tablet|  499|
|  OnePlus|  Mobile|  356|
|      Xu|  Tablet|  267|
+-----+-----+-----+
```

Window rank function

In [5]:

```
import sys
from pyspark.sql.window import Window
import pyspark.sql.functions as func
```

In [6]:

```
windowSpec1 = Window \
    .partitionBy(products['category']) \
    .orderBy(products['price'].desc())
```

In [7]:

```
price_rank = (func.rank().over(windowSpec1))
```

In [8]:

```
product_rank = products.select(
    products['product'],
    products['category'],
    products['price']
).withColumn('rank', func.rank().over(windowSpec1))

product_rank.show()
```

product	category	price	rank
iPhone	Mobile	999	1
Samsung JX	Mobile	799	2
Redmi Note	Mobile	399	3
OnePlus	Mobile	356	4
Mi	Mobile	299	5
Micromax	Mobile	249	6
Samsung TX	Tablet	999	1
iPad	Tablet	789	2
Lenovo	Tablet	499	3
Xu	Tablet	267	4

Window max function between rows

In [9]:

```
windowSpec2 = Window \
    .partitionBy(products['category']) \
    .orderBy(products['price'].desc()) \
    .rowsBetween(-1, 0)
```

In [10]:

```
price_max = (func.max(products['price']).over(windowSpec2))
```

In [11]:

```
products.select(
    products['product'],
    products['category'],
    products['price'],
    price_max.alias("price_max")).show()
```

product	category	price	price_max
iPhone	Mobile	999	999
Samsung JX	Mobile	799	999
Redmi Note	Mobile	399	799
OnePlus	Mobile	356	399
Mi	Mobile	299	356
Micromax	Mobile	249	299
Samsung TX	Tablet	999	999
iPad	Tablet	789	999
Lenovo	Tablet	499	789
Xu	Tablet	267	499

Window price difference function between ranges

In [12]:

```
windowSpec3 = Window \
    .partitionBy(products['category']) \
    .orderBy(products['price'].desc()) \
    .rangeBetween(-sys.maxsize, sys.maxsize)
```

In [13]:

```
price_difference = \
    (func.max(products['price']).over(windowSpec3) - products['price'])
```

In [14]:

```
products.select(
  products['product'],
  products['category'],
  products['price'],
  price_difference.alias("price_difference")).show()
```

```
+-----+-----+-----+-----+
| product|category|price|price_difference|
+-----+-----+-----+-----+
|   iPhone|  Mobile|  999|           0.0|
|Samsung JX|  Mobile|  799|          200.0|
|Redmi Note|  Mobile|  399|          600.0|
|   OnePlus|  Mobile|  356|          643.0|
|         Mi|  Mobile|  299|          700.0|
|  Micromax|  Mobile|  249|          750.0|
|Samsung TX| Tablet|  999|           0.0|
|       iPad| Tablet|  789|          210.0|
|   Lenovo| Tablet|  499|          500.0|
|        Xu| Tablet|  267|          732.0|
+-----+-----+-----+-----+
```

In [104]:

```
windowSpec4 = Window \
  .partitionBy(products['category']) \
  .orderBy(products['price'].asc()) \
  .rangeBetween(0, sys.maxsize)
```

In [105]:

```
sys.maxsize
```

Out[105]:

```
9223372036854775807
```

In [106]:

```
price_max = (func.max(products['price']).over(windowSpec4))
```

In [107]:

```
products.select(
  products['product'],
  products['category'],
  products['price'],
  price_max.alias("price_max")).show()
```

```
+-----+-----+-----+-----+
| product|category|price|price_max|
+-----+-----+-----+-----+
|  Micromax|  Mobile|  249|       999|
|         Mi|  Mobile|  299|       999|
|   OnePlus|  Mobile|  356|       999|
|Redmi Note|  Mobile|  399|       999|
|Samsung JX|  Mobile|  799|       999|
|   iPhone|  Mobile|  999|       999|
|        Xu| Tablet|  267|       999|
|   Lenovo| Tablet|  499|       999|
|       iPad| Tablet|  789|       999|
|Samsung TX| Tablet|  999|       999|
+-----+-----+-----+-----+
```

In []:

In []:

