

Pandas 데이터 프레임에서 Spark 데이터 프레임으로의 업 컨버전 속도 향상  
 Spark 데이터 프레임에서 Pandas 데이터 프레임으로의 업 컨버전 속도 향상  
 Pandas UDF (일명 벡터화 된 UDF)와 함께 사용  
 Apache Spark로 R 최적화

In [1]:

```
sdf = spark.read.option("inferSchema", "true").option("header", "true")W
.csv("usedcars.csv").show(5)
```

Price	Age	KM	FuelType	HP	MetColor	Automatic	CC	Doors	Weight
7450.0	65.0	82000.0	Petrol	86	1	0	1300	3	1015
7250.0	74.0	130025.0	Petrol	110	1	0	1600	3	1050
8950.0	80.0	64000.0	Petrol	110	0	0	1600	3	1055
11450.0	54.0	62987.0	Petrol	110	0	0	1600	5	1080
null	42.0	38932.0	Petrol	110	1	0	1600	3	1040

only showing top 5 rows

In [5]:

```
sdd_p= spark.read.format("parquet")W
.load("usedcars.parquet").show(5)
```

Price	Age	KM	FuelType	HP	MetColor	Automatic	CC	Doors	Weight
7450.0	65.0	82000.0	Petrol	86	1	0	1300	3	1015
7250.0	74.0	130025.0	Petrol	110	1	0	1600	3	1050
8950.0	80.0	64000.0	Petrol	110	0	0	1600	3	1055
11450.0	54.0	62987.0	Petrol	110	0	0	1600	5	1080
null	42.0	38932.0	Petrol	110	1	0	1600	3	1040

only showing top 5 rows

In [8]:

```
### 속도는 비슷

# 1. pandas 함수
import pandas as pd
%time df = pd.read_parquet('usedcars.parquet', engine='pyarrow')
```

CPU times: user 6.08 ms, sys: 523 µs, total: 6.6 ms  
 Wall time: 4.67 ms

In [12]:

```
# 2.  
import pyarrow.parquet as pq  
%time df = pq.read_pandas('usedcars.parquet').to_pandas()
```

CPU times: user 5.8 ms, sys: 0 ns, total: 5.8 ms

Wall time: 4.63 ms