

Chessa Pandu Aditirta
155610037
STMIK AKAKOM YOGYAKARTA
2018

```

examples/src/main/resources
Welcome to

      /_/_/_/_/_/_/_/_/_/_/_/_/_/_/_/_\
     /  /  /  /  /  /  /  /  /  /  /
    /__/_/_/_/_/_/_/_/_/_/_/_/_/_/_/_\
   /  /  /  /  /  /  /  /  /  /  /
  /__/_/_/_/_/_/_/_/_/_/_/_/_/_/_/_\

version 2.2.0

Using Python version 2.7.13 (default, Jan 19 2017 14:48:08)
SparkSession available as 'spark'.
>>> from pyspark.sql import SparkSession
>>> spark = SparkSession \
...   .builder \
...   .appName("Belajar Spark SQL dengan python") \
...   .config("spark.some.config.option", "some-value") \
...   .getOrCreate()
>>> df = spark.read.json("examples/src/main/resources/driftMonsters.json")
>>> df.show()

```

| KodeMobil | NUM | driftrating | merk | name | weight |
|-----------|-----|-------------|------------|------------------|--------|
| AE86 | 1 | S | Toyota | Sprinter Trueno | 1100 |
| S13 | 2 | B | Nissan | 240sx | 1200 |
| FC3S | 3 | A | Mazda | Savanna RX-7 | 1400 |
| E3 | 4 | C | Mitsubishi | Evolution III | 1500 |
| R32 | 5 | C | Nissan | Skyline BNR32 | 1200 |
| S15 | 6 | S | Nissan | Silvia Monalisa | 1200 |
| JZA80 | 7 | B | Toyota | Celica Supra | 1410 |
| UKMK7 | 8 | A | ford | Fiesta sixth gen | 1000 |
| FD3S | 9 | A | Mazda | Efini RX-7 | 1300 |
| GT86 | 10 | B | Subaru | Toyota 86 | 1200 |
| RX9 | 11 | C | Mazda | Mazda RX-8 | 1200 |

```
>>> df.createOrReplaceTempView("driftMonsters")
```

Hal yang menarik dalam spark sql adalah perintah sql dapat disimpan ke dalam sebuah variabel dan cukup mengetikkan variabel tersebut dan ditambahkan .show()

```
>>> sqlDFAll = spark.sql("SELECT * FROM driftMonsters")
>>> sqlDFAll.show()
```

| KodeMobil | NUM | driftrating | merek | name | weight |
|-----------|-----|-------------|------------|------------------|--------|
| AE86 | 1 | S | Toyota | Sprinter Trueno | 1100 |
| S13 | 2 | B | Nissan | 240sx | 1200 |
| FC3S | 3 | A | Mazda | Savanna RX-7 | 1400 |
| E3 | 4 | C | Mitsubishi | Evolution III | 1500 |
| R32 | 5 | C | Nissan | Skyline BNR32 | 1200 |
| S15 | 6 | S | Nissan | Silvia Monalisa | 1200 |
| JZA80 | 7 | B | Toyota | Celica Supra | 1410 |
| UKMK7 | 8 | A | ford | Fiesta sixth gen | 1000 |
| FD3S | 9 | A | Mazda | Efini RX-7 | 1300 |
| GT86 | 10 | B | Subaru | Toyota 86 | 1200 |
| RX9 | 11 | C | Mazda | Mazda RX-8 | 1200 |

Percobaan

Bayangkan anda memiliki banyak perintah sql yang harus anda kerjakan bagaimana mengatasinya? Spark SQL dibuat untuk itu dengan menyimpannya dalam sebuah variabel. Berikut 5 contoh sql-nya

```
>>> sqlDFDriftS = spark.sql("SELECT NUM, KodeMobil, name, merek FROM driftMonsters WHERE driftrating='S'")
```

```
>>> sqlDFDriftC = spark.sql("SELECT NUM, KodeMobil, name, merek, weight FROM driftMonsters WHERE driftrating='C'")
>>> sqlDFLWeight = spark.sql("SELECT NUM, KodeMobil, name, merek, driftrating FROM driftMonsters WHERE weight<1200")
>>> sqlDFToyota = spark.sql("SELECT NUM, KodeMobil, name, driftrating, weight FROM driftMonsters WHERE merek='Toyota'")
>>> sqlDFMazdaAND = spark.sql("SELECT NUM, KodeMobil, name, driftrating, weight FROM driftMonsters WHERE merek='Mazda' AND weight>1300")
```

maka kelimanya dapat dipanggil dengan perintah berikut

```
>>> sqlDFDriftS.show()
```

| NUM | KodeMobil | name | merek |
|-----|-----------|-----------------|--------|
| 1 | AE86 | Sprinter Trueno | Toyota |
| 6 | S15 | Silvia Monalisa | Nissan |

```
>>> sqlDFDriftC.show()
```

| NUM | KodeMobil | name | merek | weight |
|-----|-----------|---------------|------------|--------|
| 4 | E3 | Evolution III | Mitsubishi | 1500 |
| 5 | R32 | Skyline BNR32 | Nissan | 1200 |
| 11 | RX9 | Mazda RX-8 | Mazda | 1200 |

```

>>> sqlDFLWeight.show()
+---+-----+-----+-----+-----+
|NUM|KodeMobil|          name| merek|driftrating|
+---+-----+-----+-----+-----+
|  1|    AE86|Sprinter Trueno|Toyota|          S|
|  8|   UKMK7|Fiesta sixth gen|  ford|          A|
+---+-----+-----+-----+-----+

>>> sqlDFToyoda.show()
+---+-----+-----+-----+-----+
|NUM|KodeMobil|          name|driftrating|weight|
+---+-----+-----+-----+-----+
|  1|    AE86|Sprinter Trueno|          S|  1100|
|  7|   JZA80|  Celica Supra|          B|  1410|
+---+-----+-----+-----+-----+

>>> sqlDFMazdaAND.show()
+---+-----+-----+-----+-----+
|NUM|KodeMobil|          name|driftrating|weight|
+---+-----+-----+-----+-----+
|  3|    FC3S|Savanna RX-7|          A|  1400|
+---+-----+-----+-----+-----+

```

spark sql memungkinkan mengerjakan sql dengan lebih cepat