

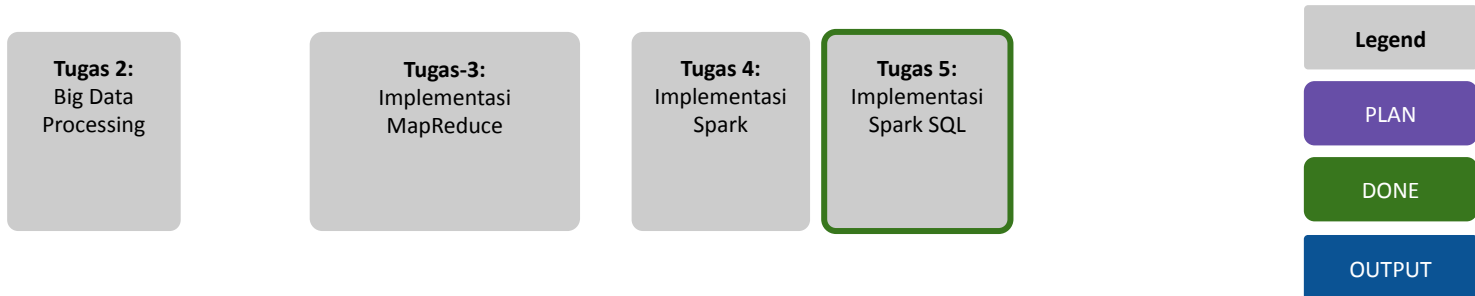
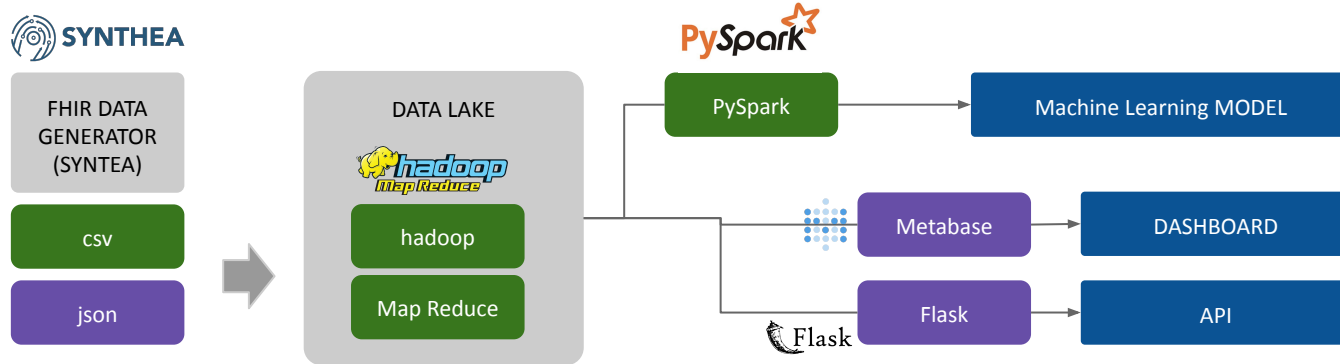
# LAPORAN



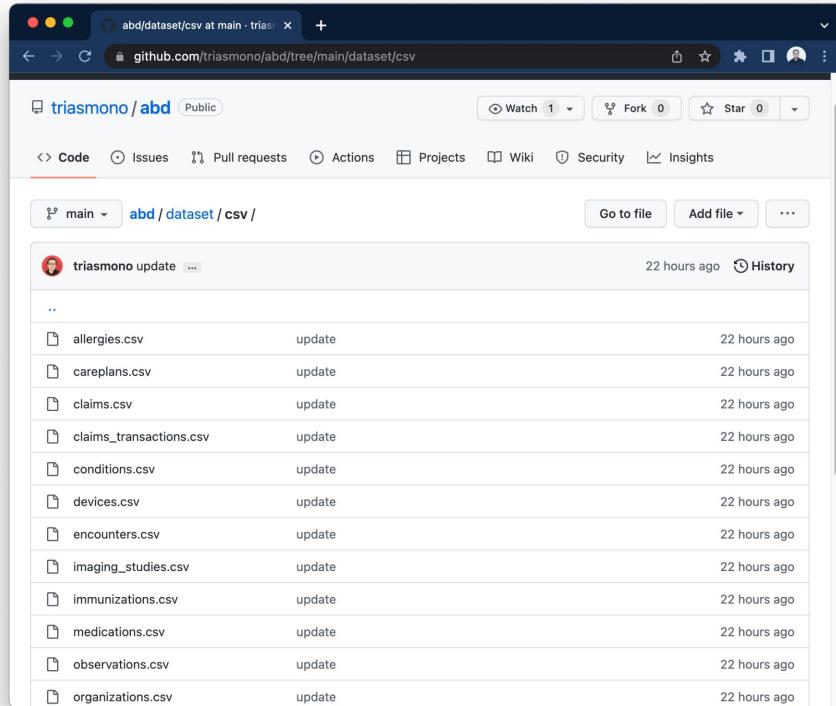
## TUGAS ANALISIS BIG DATA

TRIASMONO (21/475796/PPA/06138)

# Pipeline Data Processing



# Tugas-2



Link :

<https://github.com/triasmono/abd/tree/main/dataset/csv>

- Data di generate dari Synthea

# Tugas-3

```
(venv) triasmono@triasmacbook abd % python mapReduce.py dataset/HepatitisCdata.csv
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /var/folders/5g/d1xb8j6d6c9g4_snjzsjsnm6c0000gn/T/mapReduce.triasmono.20220610.060005.462654
Running step 1 of 1...
job output is in /var/folders/5g/d1xb8j6d6c9g4_snjzsjsnm6c0000gn/T/mapReduce.triasmono.20220610.060005.462654/output
Streaming final output from /var/folders/5g/d1xb8j6d6c9g4_snjzsjsnm6c0000gn/T/mapReduce.triasmono.20220610.060005.462654/output...
"0s=suspect Blood Donor"          7
"1=Hepatitis"                    24
"2=Fibrosis"                     21
"3=Cirrhosis"                    30
"Category"                       1
"0=Blood Donor"                  533
Removing temp directory /var/folders/5g/d1xb8j6d6c9g4_snjzsjsnm6c0000gn/T/mapReduce.triasmono.20220610.060005.462654...
(venv) triasmono@triasmacbook abd % █
```

Output saat  
dijalankan di  
localhost

Link pada repository github : <https://github.com/triasmono/abd/mapReduce.py>

- Dataset trial sementara menggunakan HepatitisCdata.csv

# Tugas-4

```
22/06/10 13:11:05 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeSystemBLAS
22/06/10 13:11:05 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeRefBLAS
+-----+-----+-----+-----+-----+
|      features|   Category|   rawPrediction|   probability|prediction|
+-----+-----+-----+-----+-----+
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[16.8659999122222...|[0.99999999713294...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[16.6972748438798...|[0.99999999637431...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[16.9749012032268...|[0.99999999744246...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[16.9063606025137...|[0.99999999667374...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[16.5157851429802...|[0.99999999566777...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[16.5895969347311...|[0.99999999586296...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[16.7244889135414...|[0.99999999620717...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[16.4388356918784...|[0.99999999520268...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[16.1590785471911...|[0.99999999304878...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[15.8170447401296...|[0.999999998951661...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[16.6806147119589...|[0.99999999644018...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[15.8477759590231...|[0.999999998967104...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[16.6008486946049...|[0.99999999596483...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[16.0583569358174...|[0.99999999223021...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[16.5373985507710...|[0.99999999568303...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[17.4481221311405...|[0.99999999850756...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[16.5500199027248...|[0.99999999524594...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[16.4596877596279...|[0.99999999525499...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[15.9649865493678...|[0.999999998951217...|    0.0|
|(1047,[0,1,2,3,4,...]|0=Blood Donor|[15.6290177850385...|[0.99999998572352...|    0.0|
+-----+-----+-----+-----+-----+
only showing top 20 rows
```

Link :

<https://github.com/triasmono/abd/spark.py>

- Dataset trial sementara menggunakan HepatitisCdata.csv

Output saat dijalankan di localhost

# Tugas-5: Spark SQL

```
Terminal: Local x Local (2) x +
STATISTIK DATA PASIEN
Berdasarkan Etnik
+-----+-----+
| ETHNICITY|count|
+-----+-----+
|nonhispanic| 109|
| hispanic| 9|
+-----+-----+

Berdasarkan Gender
+-----+-----+
|GENDER|count|
+-----+-----+
| F| 58|
| M| 60|
+-----+-----+

Berdasarkan Status Pernikahan
+-----+-----+
|MARITAL|count|
+-----+-----+
| null| 41|
| M| 58|
| S| 19|
+-----+-----+
```

```
Terminal: Local x Local (2) x +
Berdasarkan Kota
+-----+-----+
| COUNTY|count|
+-----+-----+
| Norfolk County| 14|
| Worcester County| 17|
| Bristol County| 8|
| Franklin County| 1|
| Hampshire County| 1|
| Plymouth County| 6|
| Suffolk County| 10|
| Essex County| 15|
| Hampden County| 13|
| Middlesex County| 28|
| Berkshire County| 5|
+-----+-----+

Berdasarkan Negara
+-----+-----+
| STATE|count|
+-----+-----+
| Massachusetts| 118|
+-----+-----+
```

```
Terminal: Local x Local (2) x +
Daftar pasien dengan fasilitas kesehatan
+-----+-----+-----+-----+
| FIRST| LAST|HEALTHCARE_EXPENSES|HEALTHCARE_COVERAGE|
+-----+-----+-----+-----+
| Lanny564| BartoLetti150| 604126.44| 4697.97|
| Rubin812| Gulgowski1816| 54666.47| 33950.5|
| Gaston250| Moen819| 94069.78| 800.7|
| Karma832| Spencer878| 229386.87| 989380.51|
| Alejandra902| Adorno791| 2846160.05| 23723.55|
| Freeman822| Lindgren255| 51656.85| 74.95|
| Bruno518| Fritsch593| 123246.63| 96.06|
| Deena887| Crenin516| 669038.6| 57.86|
| Emmanuel930| Prohaska837| 1602695.76| 3259.21|
| Arnoldo445| Dietrich576| 1077165.32| 118460.51|
| Malorie969| Lesch175| 394204.08| 305293.7|
| Louise871| Kautzer186| 78057.2| 106.01|
| Romona686| Treute1973| 2110503.88| 310716.25|
| Mose244| Heidenreich818| 612804.64| 3855.85|
| Clarence5| Bashirian201| 1644695.93| 20671.58|
| Catalina187| Villaseñor332| 2048801.49| 460488.11|
| Lavon272| Boehm581| 1710580.93| 13226.56|
| Senaida415| Jacobson885| 1310753.4| 375396.93|
| Harris789| Daniel959| 1688778.67| 53556.7|
| Marlen929| Kuhlman484| 52278.6| 0.0|
+-----+-----+-----+-----+
only showing top 20 rows
```

```
Daftar rencana perawatan pasien
+-----+-----+-----+-----+
| PATIENT| CODE| DESCRIPTION|
+-----+-----+-----+-----+
|4dbddb44-6dab-6a6...|225358003| Wound care|
|4dbddb44-6dab-6a6...|718361005|Weight management...|
|2a1214a2-7488-3b1...| 53950000| Respiratory therapy|
|2a1214a2-7488-3b1...|734163000|Care plan (record...|
|b191cd76-2dbf-5af...|276239002|Therapy (regime/t...|
|b191cd76-2dbf-5af...|711282006| Skin condition care|
|ceaaa849-8b9a-48e...|134435003|Routine antenatal...|
|ceaaa849-8b9a-48e...|134435003|Routine antenatal...|
|ceaaa849-8b9a-48e...|698360004|Diabetes self man...|
|46df91f7-d789-de1...|443402002|Lifestyle educati...|
|46df91f7-d789-de1...|734163000|Care plan (record...|
|46df91f7-d789-de1...|773513001|Physiotherapy car...|
|46df91f7-d789-de1...|736376001|Infectious diseas...|
|46df91f7-d789-de1...|736376001|Infectious diseas...|
|0716a41b-b232-dd6...|384758001|Self-care interve...|
|0716a41b-b232-dd6...|225358003| Wound care|
|bb990949-195f-0fe...|134435003|Routine antenatal...|
|bb990949-195f-0fe...|276239002|Therapy (regime/t...|
|bb990949-195f-0fe...|385691007| Fracture care|
|bb990949-195f-0fe...|134435003|Routine antenatal...|
+-----+-----+-----+-----+
only showing top 20 rows
```

Output saat dijalankan di localhost

- Code pada repository github : <https://github.com/triasmono/abd/sparksql.py>
- Dataset trial sementara menggunakan patients.csv dan careplans.csv
- Simulasi menampilkan statistik data patients dan rencana perawatan