

CARA MENJALANKAN PROGRAM
ARSITEKTUR DATA PIPELINE UNTUK SISTEM INFORMASI
AKADEMIK DI AKADEMI TEKNIK BIAK

1. Menjalankan Apache Spark

```
(base) C:\Users\heman>pyspark
```

Gambar 5.3 Menjalankan *apache spark*

```
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to

  ____  __
 / ___/ /  \
 \___ \|  /
  ___/ | /
     \|_/

version 3.3.2

Using Python version 3.9.12 (main, Apr  4 2022 09:22:27)
Spark context Web UI available at http://DESKTOP-UUJ960M:4040
Spark context available as 'sc' (master = local[*], app id = local-1689140806203).
SparkSession available as 'spark'.
>>> 23/07/12 14:33:37 WARN ProofsMetricsGetter: Exception when trying to compute pagesize, as a result reporting of ProofsTree metrics is stopped
23/07/12 14:33:37 WARN GarbageCollectionMetrics: To enable non-built-in garbage collector(s) List(G1 Concurrent GC), users should configure it(them) to spark.eventlog.gcMetrics.youngGenerationGarbageCollectors or spark.eventlog.gcMetrics.oldGenerationGarbageCollectors
```

Gambar 5.4 Tampilan *apache spark*

2. Proses ETL Data Dosen

Setelah apache spark berhasil dijalankan, untuk proses ETL data dosen dimulai dengan menjalankan perintah sebagai berikut:

- sc = spark.sparkContext
- path="csv/dt_dosen.csv"
- df = spark.read.csv(path)
- df = spark.read.options(delimiter=";", header=True).csv(path)
- df.show()

```
Anaconda Prompt (miniconda3) - pyspark
>>> path="csv/dt_dosen.csv"
>>> df = spark.read.options(delimiter=";", header=True).csv(path)
>>> df.show()
+----+-----+-----+-----+
| NO|      NAMA|      NIDN|      NIDK|
+----+-----+-----+-----+
| 1| Rahman Daud Tuasa...|197401022005011000|      null|
| 2| Ir. Zacharias Yaw...|      null|8834470018|
| 3| Felisitas Kanan,...|      null|8842470018|
| 4| Bruri Marwano Wut...|1227087401|      null|
| 5| Irma Dwi Kusuma,...|1411089003|      null|
| 6| Benedictus Sutaso...|1218026501|      null|
| 7| Wonsiri Heppy Nag...|1230087001|      null|
| 8| Yuliana Helena Ma...|0024058308|      null|
| 9| Eprosina Jarmida ...|1418068301|      null|
|10| Muhammad Nurwahyu...|1428058401|      null|
|11| Riyanti Ohorella...|1420018602|      null|
|12| Utami Aryanti., S...|1402048801|      null|
|13| Daniel Wambrauw., ST|1208116402|      null|
|14| MAXI GUNTUR LALUJ...|      null|9912001199|
|15| DOMINGGUS P LEKAT...|      null|9912000137|
|16| DARMA SETIYAWAN., ST|      null|9912000528|
|17| ERIC HUTAJULU., S.Pd|      null|9914011945|
|18| Mario Agapito Ari...|0026058902|      null|
|19| Taufiq Akbar., S...|1429089003|      null|
|20| Frinda Wahyu Nurh...|1401078501|      null|
+----+-----+-----+-----+
only showing top 20 rows
>>>
```

Gambar 5.5 Dataframe Dosen

3. Proses ETL Data Mahasiswa

Untuk proses ETL data mahasiswa dimulai dengan menjalankan perintah sebagai berikut :

- sc = spark.sparkContext
- path="csv/mhs.csv"
- df = spark.read.csv(path)
- df = spark.read.options(delimiter=";", header=True).csv(path)
- df.show()

```
Anaconda Prompt (miniconda3) - pyspark
rs should configure it(them) to spark.eventLog.gcMetrics.youngGeneration
dGenerationGarbageCollectors
sc= spark.sparkContext
>>> path="csv/mhs.csv"
>>> df = spark.read.options(delimiter=";", header=True).csv(path)
>>> df.show()
```

No	Nama	Status	TTL
1	Daniel Ringgo Koibur	Cuti	Biak, 11 Desember...
2	Natanael Korwa	Cuti	Opiaref, 23 Agust...
3	Ekawanti F. Radam...	Aktif	Layani , 21, Mare...
4	Lemleri leunufna	Aktif	Layani ,15,Ferbua...
5	Marcelino,A,Lendr...	Aktif	Layani ,15 ,maret...
6	Titus Rumarar	Aktif	
7	Ferdinand Randi D...	Aktif	Ambon, 2 Juli 1998
8	Demianus Mnuwon	Aktif	null
9	Hans Tonci Rumabur	Aktif	null
10	Silas Kiam	Aktif	null
11	Oskar A. Maninwarba	Aktif	Mos, 13 April 1996
12	Musa Saflembolo	Aktif	Biak, 17 Juni 2001
13	Julan Jhosua Kbarek	Aktif	Biak, Juni 2002
14	Alberth Karim Karma	Aktif	null
15	Yany Marnes Sahun...	Aktif	null
16	Demianus Kiambo	Aktif	Urmboridori, 25 J...
17	Timus Wospakrek	Aktif	Swaipak, 24 Desem...
18	Sofyan Korwa, S.AN	Cuti	Ujung Pandang. 6 ...
19	Arman S. Boserem	Aktif	Maryendi, 27 Sept...
20	Yulianus Eneas Ra...	Aktif	Biak, 27 Oktober ...

Gambar 5.6 Dataframe Mahasiswa

4. Proses ETL Data Mata Kuliah

Untuk proses ETL data mata kuliah dimulai dengan menjalankan perintah sebagai berikut :

- sc = spark.sparkContext
- path="csv/dt_matkul.csv"
- df = spark.read.csv(path)
- df = spark.read.options(delimiter=";", header=True).csv(path)
- df.show()

```
>>> path="csv/dt_matkul.csv"
>>> df = spark.read.options(delimiter=";", header=True).csv(path)
>>> df.show()
```

NO	KDMK	NAMA MATAKULIAH	SKS
1	MKK-225705	Akuntansi Dasar	2
2	MKK-225706	Statistik	2
3	MKK-225707	Aljabar Linier	3
4	MKK-225708	Algoritma Pemrogr...	2
5	MKK-225709	Prak. Algoritma P...	2
6	MKK-225710	Pengenalan Linux	3
7	MKB-325703	Paket Program Nia...	2
8	MKB-325704	Praktek Paket Pro...	2
9	MKB-325705	Basis Data I	2
10	MKB-325706	Prak. Basis Data I	2
null	null	Jumlah SKS	22

Gambar 5.7 Dataframe Mata Kuliah

5. Proses ETL Data KRS

Untuk proses ETL data KRS dimulai dengan menjalankan perintah sebagai berikut:

- `sc = spark.sparkContext`
- `path="csv/dt_krs.csv"`
- `df = spark.read.csv(path)`
- `df = spark.read.options(delimiter=";", header=True).csv(path)`
- `df.show()`

```
>>> sc = spark.sparkContext
>>> path="csv/dt_krs.csv"
>>> df = spark.read.options(delimiter=";", header=True).csv(path)
>>> df.show()
```

NO	KODE MK	MK	SKS	B/U	HARI	JAM	DOSEN
1	MKK-225710	PENGENALAN LINUX	3	B	SENIN	14.00-16.00	HABEL KOIBUR., S.Kom
2	MKK-225705	AKUNTANSI DASAR	2	B	SELASA	14.00-16.00	SITI CAHYANINGSIH...
3	MKB-325703	PAKET PROGRAM NIA...	2	B	RABU	14.00-18.00	HABEL KOIBUR., S.Kom
4	MKB-325704	PRAKTEK PAKET PRO...	2	B	RABU	14.00-18.00	HABEL KOIBUR., S.Kom
5	MKK-225707	ALJABAR LINEAR	2	B	KAMIS	14.00-16.00	Ir. ZAKARIAS YAWA...
6	MKK-225706	STATISTIK	2	B	KAMIS	16.00-18.00	RAHMAN D.T., ST., MT
7	MKB-325707	BASIS DATA I	1	B	JUMAT	14.00-16.00	MARIO A.A. GOBEL...
8	MKB-325706	PRAKTEK BASIS DATA I	2	B	JUMAT	16.00-18.00	MARIO A.A. GOBEL...
9	MKK-225708	ALGORITMA PEMROGR...	1	B	SABTU	14.00-16.00	HABEL KOIBUR., S.Kom
10	MKK-225709	PRAKTEK ALGORITMA...	2	B	SABTU	16.00-18.00	HABEL KOIBUR., S.Kom

Gambar 5.8 Dataframe KRS

6. Proses ETL Data Beasiswa

Untuk proses ETL data beasiswa dimulai dengan menjalankan perintah sebagai berikut:

- `sc = spark.sparkContext`
- `path="csv/dt_penerima_beasiswa.csv"`
- `df = spark.read.csv(path)`
- `df = spark.read.options(delimiter=";", header=True).csv(path)`
- `df.show()`

```
>>> path="csv/dt_penerima_beasiswa.csv"
>>> df = spark.read.options(delimiter=";", header=True).csv(path)
>>> df.show()
+----+-----+-----+
| No | Nama | NIM |
+----+-----+-----+
| 1 | Karel Istia | 144006214012015 |
| 2 | Robertho Hommy | 144006214012016 |
| 3 | Ekawanti Radamusa | 144006224012001 |
| 4 | Lemlemri Ari Leun... | 144006224012002 |
| 5 | Marcelino A. Lend... | 144006224012003 |
| 6 | Musa Saflombolo | 144006224012011 |
| 7 | Silas Kiam | 144006224012008 |
| 8 | Julian Josua Kbarek | 144006224012013 |
| 9 | Hans Tonci Rumabur | 144006224012007 |
| 10 | Ferdinan R. D. La... | 144006224012005 |
| 11 | Fredinand Cevin Sada | 144006574012002 |
| 12 | Fransiska Kolbur | 144006574012006 |
| 13 | Elias O. Seyauw R... | 144006574012011 |
| 14 | Maria Desi Kafiar | 144006574012008 |
| 15 | Nigel Kbarek | 144006574012001 |
| 16 | Yosias Teurupun | 144006574012018 |
| 17 | Desy Natalia Mnum... | 144006574012020 |
| 18 | Levinus Theo Rump... | 144006214012101 |
| null | Edwin E. W. Rando... | 144006214012105 |
| 3 | Herman Keys Mansawan | 144006214012108 |
+----+-----+-----+
only showing top 20 rows
```

Gambar 5.9 Dataframe Penerima Beasiswa