

BAB V

HASIL DAN PEMBAHASAN

5.1 Data

Pada proses pembuatan arsitektur data pipeline ini menggunakan dua data penting yaitu data internal kampus Akademi Teknik Biak dan data Lembaga Layanan Pendidikan Tinggi Wilayah XIV.

5.1.1 Data Kampus Akademi Teknik Biak

Data yang digunakan pada penelitian ini adalah dataset APTOS 2019 Blindnes Detectioin 5.590 gambar dari Retinopathy Diabetic.

1. Data Dosen
2. Data Mahasiswa
3. Data Mata Kuliah
4. Data KRS
5. Data Beasiswa

5.1.2 Data Lembaga Layanan Pendidikan Tinggi Wilayah XIV

1. PPDIKTI
2. SISTER
3. SINTA
4. SIMCITABMAS
5. ARJUNA
6. REG SDM

5.2 Instalasi dan Konfigurasi *Apache Spark*

Adapun beberapa tahapan dalam persiapan proses ETL menggunakan *apache spark* yaitu :

1. Penginstalan *Apache Spark*

Apache Spark yang digunakan pada penelitian ini versi 3.3.2 dimana penulis langsung mengambil dari alamat resminya <https://spark.apache.org/>



Gambar 5.1 Alamat *website apache spark*

Untuk menjalankan apache spark diwajibkan untuk memasang JDK terlebih dahulu. JDK yang dipakai adalah java versi 20.0.1 2023-04-18

2. *Anaconda spark*

Untuk mengelola dan mendistribusi paket dalam pemrograman python, anaconda yang digunakan pada penelitian ini versi 4.14.0

3. *Scala*

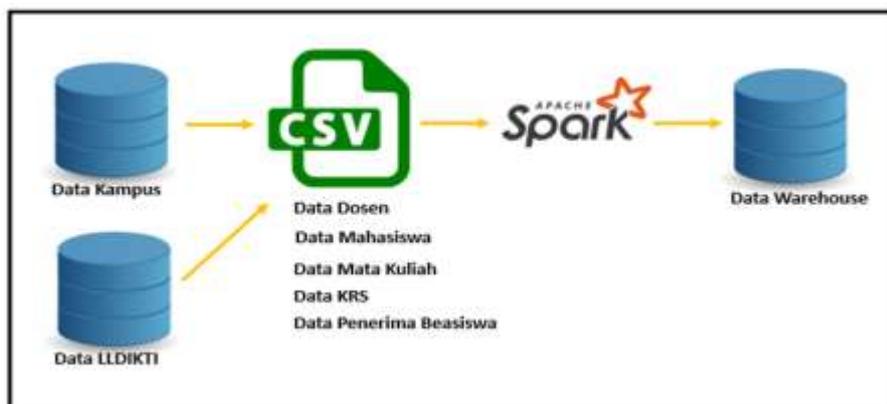
Sedangkan untuk menggabungkan pemrograman berorientasi objek dan fungsional dalam satu bahasa tingkat tinggi yang ringkas penulis menggunakan scala versi 2.13.6. Setelah mendapatkan gambaran umum dari *software* yang digunakan pada system, maka dilakukan analisis untuk memenuhi kebutuhan sistem yang akan digunakan mulai dari spesifikasi *hardware* yang akan digunakan serta perangkat lunaknya, daftar dari spesifikasi sistem diantaranya adalah sebagai berikut:

Tabel 5.1 Spesifikasi sistem

| Spesifikasi Sistem | | |
|--------------------|----------------|--|
| 1 | Laptop | 1 Buah |
| 2 | Processor | Intel(R) Core(TM) i3-3217U CPU @ 1.80GHz 1.80 GHz |
| 3 | RAM | 4GB |
| 4 | OS | Windows 10 64 Bit |
| 5 | Apache spark | Versi 3.3.2 |
| 6 | JDK | Versi 20.0.1 2023-04-18 |
| 7 | Anaconda spark | Versi 4.14.0 |
| 8 | Scala | Versi 2.13.6 |

5.3 Extract-Transform-Load (ETL)

Dalam penelitian ini data source yang diambil sebagai sampel pada proses ETL yaitu data kampus Akademi Teknik Biak. File dari data kampus Akademi Teknik Biak ini penulis mengubah dengan ekstensi csv sebelum proses extract pada apache spark. Untuk proses extract data sendiri dibutuhkan library pendukung pada apache spark dengan menggunakan library sparkContext. SparkContext ini dilakukan agar pemanggilan data mahasiswa dengan ekstensi csv dapat dijalankan pada apache spark. Setelah proses extract dilakukan, apache spark akan membaca data mahasiswa.csv dan melakukan proses transform data. Dari proses transform data mahasiswa.csv ini akan di load ke dalam sebuah Data warehouse sebagai output dari proses ETL. Untuk proses ETL yang dilakukan dapat dilihat pada gambar 5.2



Gambar 5.2 Arsitektur Data Pipeline Akademi Teknik Biak

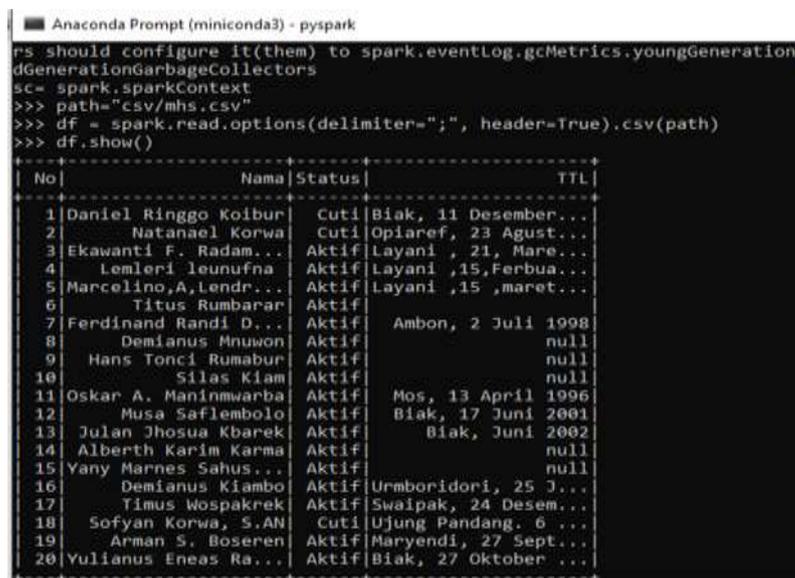
5.4 Proses Extract-Transform-Load (ETL) Pada Apache Spark

Untuk proses ETL pada penelitian ini, penulis menggunakan data pada kampus Akademi Teknik Biak yang terdiri dari data dosen, data mahasiswa, data mata kuliah, data KRS, dan data beasiswa. Data ini merupakan data mentah dalam bentuk file CSV, yang disimpan pada sebuah direktori dan kemudian akan dilakukan proses ETL pada *Apache Spark*. Dari penjelasan pada gambar 5.3 berikut tahapan dan proses ETL pada apache spark Berikut Tahapan proses ETL pada *Apache Spark*:

5.4.3 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 Rumarbar | Aktif | |
| 7 | Ferdinand Randi D... | Aktif | Ambon, 2 Juli 1998 |
| 8 | Demianus Mnuwon | Aktif | null |
| 9 | Hans Tonci Rumabur | Aktif | null |
| 10 | Silas Klam | 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 | Urmboridoni, 25 J... |
| 17 | Timus Wospakrek | Aktif | Swaipak, 24 Desem... |
| 18 | Sofyan Korwa, S.AN | Cuti | Ujung Pandang. 6 ... |
| 19 | Arman S. Bosenen | Aktif | Maryendi, 27 Sept... |
| 20 | Yulianus Eneas Ra... | Aktif | Biak, 27 Oktober ... |

Gambar 5.6 Dataframe Mahasiswa

5.4.4 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

5.4.5 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 YAMA... |
| 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

5.4.6 Proses ETL Data Beasiswa

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

- sc = spark.sparkContext
- path="csv/dt_penerima_basiswa.csv"
- df = spark.read.csv(path)

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

```
>>> path="csv/dt_penerima_basiswa.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 Koibur | 144006574012006 |
| 13 | Elias D. 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