

TESIS

ARSITEKTUR DATA PIPELINE UNTUK SISTEM

INFORMASI AKADEMIK DI AKADEMI TEKNIK BIAK



HEMAN KORERI ISRAEL MNSN

21/2006/0056/TSD/13

PROGRAM STUDI TEKNOLOGI INFORMASI

PROGRAM MAGISTER

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA

YOGYAKARTA

2024

TESIS
ARSITEKTUR DATA PIPELINE UNTUK SISTEM
INFORMASI AKADEMIK DI AKADEMI TEKNIK BIAK

Diajukan sebagai salah satu syarat untuk menyelesaikan studi

Program Magister
Program Studi Teknologi Informasi
Fakultas Teknologi
Universitas Teknologi Digital Indonesia
Yogyakarta

Disusun Oleh

HEMAN KORERI ISRAEL MNSEN

21/2006/0056/TSD/13

PROGRAM STUDI TEKNOLOGI INFORMASI
PROGRAM MAGISTER
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA

2024

HALAMAN PERSETUJUAN
UJIAN TESIS

Judul : **Arsitektur Data Pipeline Untuk Sistem Informasi Akademik Di Akademi Teknik Biak**
Nama : **Heman Koreri Israel Mnsen**
NIM : **21/2006/0056/TSD/13**
Program Studi : **Teknologi Informasi**
Program : **Magister**
Semester : **Gasal**
Tahun Akademik : **2023/2024**

Telah diperiksa dan disetujui untuk diujikan di hadapan Dewan Penguji Tesis

Yogyakarta, 16 Febuari 2024

Dosen Pembimbing,



Dr. Bambang P.D.P, S.E., Ak., S.Kom., MMSI

0525087201

HALAMAN PENGESAHAN


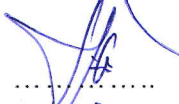
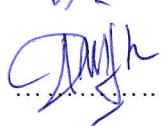
TESIS

ARSITEKTUR DATA PIPELINE UNTUK SISTEM INFORMASI AKADEMIK DI AKADEMI TEKNIK BIAK

Telah dipertahankan di depan Dewan Penguji Tesis dan dinyatakan diterima untuk memenuhi sebagian persyaratan guna memperoleh Gelar

Magister Komputer
Program Studi Teknologi Informasi
Fakultas Teknologi Informasi
Universitas Teknologi Digital Indonesia
Yogyakarta

Yogyakarta, 31 Januari 2024

Dewan Penguji	NIDN	Tandatangan
1. Dr. Bambang P.D.P, S.E., Ak., S.Kom., MMSI (Ketua)	0525087201	
2. Dr. Widyastuti Andriyani, S.Kom., M.Kom (Penguji 1)	0217038201	
3. Dr. Dommy Kristomo, S.T., M.Eng. (Penguji 2)	0530078302	

Mengetahui

Ketua Program Studi Teknologi Informasi

08 MAR 2024

Dr. Widyastuti Andriyani, S.Kom., M.Kom

NPP 121176

PERNYATAAN KEASLIAN TESIS

Dengan ini saya menyatakan bahwa naskah tesis ini belum pernah diajukan untuk memperoleh gelar Magister Komputer di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara sah diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 31 Januari 2024



Heman Koreri Israel Mnsen

21/2006/0056/TSD/13

INTISARI

Dalam membangun sebuah sistem informasi yang terintegrasi, perencanaan arsitektur adalah langkah awal yang harus dilakukan. Perencanaan pengembangan sebuah sistem informasi sangat diperlukan agar sistem yang dikembangkan dapat berjalan sesuai dengan kebutuhan. Data yang digunakan pada penelitian ini yaitu data internal kampus Akademi Teknik Biak dan data eksternal dari Lembaga Layanan Pendidikan Tinggi Wilayah XIV.

Adapun tujuan dari penelitian ini adalah merancang arsitektur data *pipeline* sistem informasi Akademi Teknik Biak. Arsitektur *pipeline* digunakan untuk membawa sumber daya yang besar dari satu lokasi ke lokasi lain yang jaraknya berjauhan dengan lebih efisien. Metode yang digunakan pada penelitian ini adalah proses *extract-transform-load* (ETL).

Untuk proses *extract* data sendiri dibutuhkan *library* pendukung pada apache spark dengan menggunakan *library SparkSession*. *SparkSession* ini dilakukan agar pemanggilan data Akademi Teknik Biak dengan ekstensi csv dapat dijalankan pada *apache spark*. Setelah proses *extract* dilakukan, *apache spark* akan membaca data dengan ekstensi.csv dan melakukan proses *transform* data. Dari proses *transform* data ekstensi.csv ini akan di load ke dalam sebuah *Dataframe* sebagai output dari proses ETL. Hasil penelitian yang diperoleh adalah teknologi *apache spark* dapat mempermudah penulis dalam proses perancangan sistem informasi Akademi Teknik Biak dan menjadi salah satu solusi yang tepat dalam proses *Extract Load Transform (ETL)* data dengan skala yang besar dan *real-time*.

Kata Kunci: Apache Spark, Arsitektur Pipeline, ETL, Sistem Informasi, DataFrame.

ABSTRACT

In building an integrated information system, architectural planning is the first step that must be taken. Planning the development of an information system is needed so that the system developed can run as needed. The data used in this study are internal data from the Biak Engineering Academy campus and external data from the Higher Education Service Institution Region XIV.

The purpose of this research is to design the data pipeline architecture of the Biak Technical Academy information system. The pipeline architecture is used to bring large resources from one location to another location that is far apart more efficiently. The method used in this research is the extract-transform-load (ETL) process.

For the data extract process itself, a supporting library is needed in apache spark using the SparkSession library. This SparkSession is done so that the Biak Technical Academy data call with the csv extension can be run on apache spark. After the extract process is done, apache spark will read the data with the extension.csv and perform the data transform process. From the process of transforming the extension.csv data, it will be loaded into a Dataframe as the output of the ETL process. The results of the research obtained are apache spark technology can facilitate the author in the process of designing the Biak Technical Academy information system and become one of the right solutions in the Extract Load Transform (ETL) process with large-scale and real-time data.

Keywords: Apache Spark, Pipeline Architecture, ETL, Information Systems, DataFrame.

DAFTAR ISI

TESIS	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TESIS	v
INTISARI.....	vi
ABSTRACT.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Tujuan.....	3
1.5 Manfaat.....	3
BAB II TINJAUAN PUSTAKA.....	4
BAB III LANDASAN TEORI.....	20
3.1 Arsitektur Pipeline.....	20
3.2 <i>Apache Spark</i>	21
3.3 Resilient Distributed Dataset (RDD).....	22
BAB IV METODOLOGI PENELITIAN	25
4.1 Studi Literatur	25
4.2 Jenis Penelitian dan Tempat Penelitian	26
4.3 Alat dan Teknologi Yang Digunakan Pada Penelitian	26
4.4 Perancangan Arsitektur Data Pipeline Akademi Teknik Biak.....	26
4.4.1 Data Kampus Akademi Teknik Biak	27
BAB V HASIL DAN PEMBAHASAN.....	29

5.1	Data	29
5.1.1	Data Kampus Akademi Teknik Biak.....	29
5.1.2	Data Lembaga Layanan Pendidikan Tinggi Wilayah XIV.....	29
5.2	Instalasi dan Konfigurasi <i>Apache Spark</i>	29
5.3	<i>Extract-Transform-Load</i> (ETL)	31
5.4	Proses <i>Extract-Transform-Load</i> (ETL) Pada <i>Apache Spark</i>	31
5.4.1	Menjalankan <i>Apache Spark</i>	32
5.4.2	Proses ETL Data Dosen	32
5.4.3	Proses ETL Data Mahasiswa.....	33
5.4.4	Proses ETL Data Mata Kuliah.....	33
5.4.5	Proses ETL Data KRS	34
5.4.6	Proses ETL Data Beasiswa.....	34
BAB VI KESIMPULAN DAN SARAN		36
6.1	Kesimpulan.....	36
6.2	Saran	36

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR GAMBAR

Gambar 3.1 <i>Arsitektur Pipeline</i>	21
Gambar 3.2 <i>Arsitektur Apache Spark</i>	22
Gambar 3.3 <i>Parallel Operationn Across Data In Nodes In Cluster</i>	23
Gambar 3.4 <i>Dataframe</i>	24
Gambar 4.1 <i>Arsitektur data pipeline akadmi teknik biak</i>	27
Gambar 5.1 <i>Alamat website apache spark</i>	30
Gambar 5.2 <i>Arsitektur Data Pipeline Akademi Teknik Biak</i>	31
Gambar 5.3 <i>Menjalankan apache spark</i>	32
Gambar 5.4 <i>Tampilan apache spark</i>	32
Gambar 5.5 <i>Dataframe Dosen</i>	32
Gambar 5.6 <i>Dataframe Mahasiswa</i>	33
Gambar 5.7 <i>Dataframe Mata Kuliah</i>	34
Gambar 5.8 <i>Dataframe KRS</i>	34
Gambar 5.9 <i>Dataframe Penerima Beasiswa</i>	35

DAFTAR TABEL

Tabel 2.1 Tinjauan Pustaka	11
Tabel 5.1 Spesifikasi Sistem	30