

**IMPLEMENTASI *EXTRACT TRANSFORM LOAD* PADA
SENTRALISASI DATA AKADEMIK *BOOTCAMP* GUNA
MENINGKATKAN EFISIENSI PROSES PEMBELAJARAN DI
HACKTIV8**

**Oleh
NATHANIA SANTA NIGEL SIMBOLON**

Skripsi

**Sebagai Salah Satu Syarat Untuk Mencapai Gelar
SARJANA ILMU KOMPUTER**

Pada

**Program Studi Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG**

2025

ABSTRAK

IMPLEMENTASI EXTRACT TRANSFORM LOAD PADA SENTRALISASI DATA AKADEMIK BOOTCAMP GUNA MENINGKATKAN EFISIENSI PROSES PEMBELAJARAN DI HACKTIV8

Oleh

NATHANIA SANTA NIGEL SIMBOLON

Penelitian ini membahas implementasi Extract, Transform, Load (ETL) pada sentralisasi data akademik di PT. Hacktivate Teknologi Indonesia (Hacktiv8) untuk meningkatkan efisiensi proses pembelajaran. Pertambahan data yang pesat dalam data akademik yang tidak terorganisir, diperlukan sistem otomasi yang terintegrasi untuk memonitor dan mengevaluasi perkembangan peserta bootcamp. Proses ETL yang diterapkan mengumpulkan data dari berbagai Google Spreadsheet, mentransformasikannya menjadi bentuk normal, dan menyimpannya dalam data warehouse yang terpusat menggunakan Google Cloud Platform. Penelitian ini menggunakan metodologi 4-steps-Kimball untuk mendesain skema *data warehouse* dan melakukan pengujian kualitas data berdasarkan standar ISO/IEC 25012. Hasil evaluasi menunjukkan bahwa sistem ETL yang diimplementasikan sangat efisien, dengan penilaian positif dari pengguna, meskipun terdapat beberapa saran untuk peningkatan, terutama dalam hal dokumentasi dan adaptabilitas sistem. Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam proses pengambilan keputusan berbasis data di *environment* pendidikan non-formal.

Kata Kunci: ETL, *data warehouse*, Google Cloud Platform, ISO/IEC 25012

ABSTRACT

IMPLEMENTATION OF EXTRACT TRANSFORM LOAD ON BOOTCAMP ACADEMIC DATA CENTRALIZATION TO IMPROVE LEARNING PROCESS EFFICIENCY AT HACKTIV8

By

NATHANIA SANTA NIGEL SIMBOLON

This study explores the implementation of the Extract, Transform, Load (ETL) process for academic data centralization at PT. Hacktivate Teknologi Indonesia (Hacktiv8) with the aim of enhancing learning process efficiency. Given the rapid growth of disorganized academic data, an integrated automation system is required to monitor and evaluate bootcamp participants' progress. The implemented ETL process consolidates data from various Google Spreadsheets, transforms it into a normalized format, and stores it in a centralized data warehouse using Google Cloud Platform. This research employs the 4-Steps-Kimball methodology to design the data warehouse schema and evaluates data quality based on the ISO/IEC 25012 standard. The evaluation results indicate that the implemented ETL system is highly efficient, receiving positive user feedback despite suggestions for improvement, particularly regarding system documentation and adaptability. This study is expected to make a significant contribution to data-driven decision-making processes within non-formal educational environments.

Keyword: ETL, data warehouse, Google Cloud Platform, ISO/IEC 25012

**IMPLEMENTASI *EXTRACT TRANSFORM LOAD* PADA
SENTRALISASI DATA AKADEMIK *BOOTCAMP* GUNA
MENINGKATKAN EFISIENSI PROSES PEMBELAJARAN DI
HACKTIV8**

Oleh:

NATHANIA SANTA NIGEL SIMBOLON

SKRIPSI

**Sebagai Salah Satu Syarat Untuk Mencapai Gelar
SARJANA ILMU KOMPUTER**

Pada

**JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG**

2025

Judul Skripsi

: **IMPLEMENTASI EXTRACT
TRANSFORM LOAD PADA
SENTRALISASI DATA AKADEMIK
BOOTCAMP GUNA
MENINGKATKAN EFISIENSI
PROSES PEMBELAJARAN DI
HACKTIV8**

Nama Mahasiswa

: **Nathania Santa Nigel Simbolon**

Nomor Pokok Mahasiswa

: 2117051009

Program Studi

: **S1 Ilmu Komputer**

Fakultas

: **Matematika dan Ilmu Pengetahuan Alam**



MENYETUJUI

1. Komisi Pembimbing

Didik Kurniawan, S.Si., M.T.
NIP 1981021 5200812 2 001

Rahman Taufik, S.Pd, M. Kom.
NIP 19600421 198603 2 001

2. Ketua Jurusan Ilmu Komputer

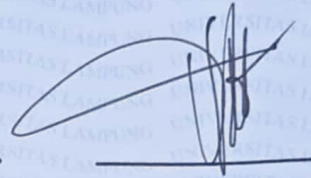
Dwi Sakethi, S. Si., M.Kom.
NIP. 19680611 199802 1 001

MENGESAHKAN

1. Tim Penguji

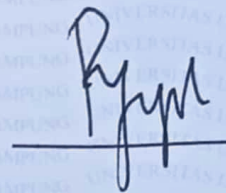
Ketua

: Didik Kurniawan, S.Si., M.T.



Sekretaris/Anggota

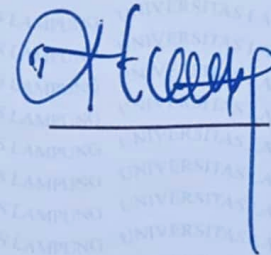
: Rahman Taufik, S.Pd, M. Kom.



Penguji

Bukan Pembimbing

: Tristiyanto, S.Kom., M.I.S., Ph.D.



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Heri Satria, S.Si., M.Si.

NIP 1971100120050110002

Tanggal Lulus Ujian Skripsi: 25 April 2025

PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Nathania Santa Nigel Simbolon

NPM : 2117051009

Program Studi : S1 Ilmu Komputer

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa skripsi saya yang berjudul **“Implementasi Extract Transform Load Pada Sentralisasi Data Akademik Bootcamp Guna Meningkatkan Efisiensi Proses Pembelajaran Di Hacktiv8”** adalah benar-benar hasil karya sendiri dan bukan hasil plagiat sebagaimana telah diatur dalam Pasal 27 Peraturan Akademik Universitas Lampung dengan Keputusan Rektor Nomor 3187/H26/DT/2010. Semua hasil tulisan yang tertuang dalam Skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung.

Bandar Lampung, 25 April 2025

Penulis



Nathania Santa Nigel Simbolon

NPM 2117051009

RIWAYAT HIDUP



Penulis bernama lengkap Nathania Santa Nigel Simbolon, lahir di Teluk Betung pada 25 Februari 2003. Penulis merupakan anak pertama dari tiga bersaudara dari pasangan Bapak Bilma Simbolon dan Ibu Rosni Maryana Malau. Penulis menyelesaikan pendidikan di Taman Kanak-Kanak Bodhisattva pada tahun 2009, SDN 5 Talang pada tahun 2015, SMPN 1 Bandar Lampung pada tahun 2018 dan SMAN 2 Bandar Lampung pada tahun 2021.

Penulis kemudian diterima sebagai mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung pada tahun 2021, melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN). Penulis telah mengikuti kegiatan Kuliah Kerja Nyata (KKN) Periode 2 Tahun 2024 selama 40 hari di Desa Bandar Negeri, Kecamatan Labuhan Maringgai, Kabupaten Lampung Timur, Provinsi Lampung.

Selama menjadi mahasiswa, penulis aktif mengikuti beberapa kegiatan antara lain:

1. Anggota bidang Keilmuan Himakom Universitas Lampung.
2. Asisten Dosen mata kuliah Matematika dan Pemrograman Terstruktur.
3. Pada tahun 2023 mendapatkan kesempatan untuk melakukan kegiatan kerja praktik di SMAS YP Unila sebagai *Full Stack Web Developer*
3. Pada tahun 2024 mendapatkan kesempatan untuk mengikuti kegiatan magang MSIB 6 di PT. Hacktivate Teknologi Indonesia dan mengikuti kegiatan Merdeka Belajar Kampus Merdeka (MBKM) riset penelitian Deep Learning: Face Recognition untuk Sistem Kehadiran Mahasiswa di Jurusan Ilmu Komputer Universitas Lampung di tahun 2023.
4. Menjadi Finalis untuk kompetisi Sains Data Nasional 2024 dan menjadi prosposal kedua terbaik.

Selain aktif berorganisasi, penulis juga aktif mengikuti lomba akademik, seperti penulisan esai, dan lomba di bidang sains data. Penulis berhasil menorehkan beberapa prestasi, diantaranya mendapatkan Juara II Esai Nasional DPM Unila, Finalis Turnamen Sains Data Nasional 2024 yang diselenggarakan Lembaga Sertifikasi Profesi (LSP).

MOTO

“Mengasihi sesamamu dengan seluruh kemampuanmu dan lakukan yang terbaik”

(Penulis)

“Segala perkara dapat kutanggung di dalam Dia yang memberi kekuatan kepadaku.”

(Filipi 4:13)

“Trust in the LORD with all thine heart; and lean not unto thine own understanding.

In all thy ways acknowledge him, and he shall direct thy paths.”

(Proverbs 3:5-6)

PERSEMBAHAN

Puji syukur kepada Allah Bapa Maha Kasih yang telah memberikan kekuatan dan hikmat kebijaksanaan bagiku untuk menyelesaikan karya ini.

Dengan segala kerendahan hati, Kupersembahkan skripsi ini kepada:

Kedua Orangtuaku Terkasih,

Bapak **Bilma Simbolon** dan Ibu **Rosni Maryana Malau**

Terima kasih atas kasih sayang yang selalu kalian berikan, namaku yang selalu disebutkan dalam doa, serta segala dukungan yang tak henti-hentinya kalian berikan dalam perjalanan dan perjuanganku selama melangkah menuju keberhasilan.

Semoga Allah Bapa di Surga senantiasa memberkati setiap langkah hidup kalian.

Dan terakhir khususnya bagi,

Almamater tercinta Universitas Lampung

Serta semua orang yang mendukungku selama ini.

SANWACANA

Puji syukur kepada Allah Bapa di Surga, karena atas rahmat dan kasih karunia penyertaan-Nya, Penulis dapat menyelesaikan skripsi yang berjudul **“Implementasi Extract Transform Load Pada Sentralisasi Data Akademik Bootcamp Guna Meningkatkan Efisiensi Proses Pembelajaran Di Hacktiv8”**. Skripsi ini merupakan salah satu syarat untuk memperoleh gelar Sarjana Ilmu Komputer di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

Penyelesaian skripsi ini tidak terlepas dari bantuan, bimbingan, *feedback*, dan dukungan dari berbagai pihak, baik secara langsung maupun tidak langsung. Maka, pada kesempatan ini, penulis mengucapkan terima kasih kepada semua pihak yang telah membantu dalam penyusunan skripsi ini sehingga dapat diselesaikan dengan baik kepada:

1. Bapak Bilma Simbolon dan Ibu Rosni selaku kedua orang tua yang selalu memberikan dukungan, kasih sayang, dan doa serta membantu dalam segala hal yang tidak terhitung nilainya.
2. Bapak Didik Kurniawan, S.Si., M.T selaku Dosen Pemimbing I dan Bapak Rahman Taufik, S.Pd., M. Kom selaku Dosen Pemimbing II yang telah meluangkan waktunya untuk membantu penulis dengan sabar dalam membimbing, memberikan saran, masukan, motivasi dan pengarahan serta pengertiannya selama proses pengerjaan skripsi sehingga penulis dapat menyelesaikan skripsi ini dengan baik;
3. Bapak Tristiyanto, S.Kom., M.I.S., Ph.D selaku dosen pembahas serta ketua Prodi S1 Ilmu Komputer Universitas Lampung yang telah memberikan kritik, saran serta masukan kepada penulis;
4. Dr. Aristoteles, S.Si., M.Si. selaku Dosen Pembimbing Akademik dan Wakil Dekan Bidang Umum dan Keuangan FMIPA Universitas Lampung yang selalu membantu memberikan arahan kepada penulis selama menempuh pendidikan di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung;

5. Bapak Dwi Sakethi, S.Si., M. Kom selaku Ketua Jurusan Ilmu Komputer Universitas Lampung.
6. Para dosen, Ibu Ade Nora Maela, dan karyawan yang bertugas di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung, yang selama ini telah memberikan ilmu dan membantu proses administrasi penulis dalam studi.
7. Mas Fahmi dan Mas Rachman selaku mentor di Hacktiv8, terima kasih sudah selalu mempermudah segala kebutuhan penulis dalam penyusunan karya ini;
8. Tante Jordan yang selalu memberikan semangat dan dukungan
9. Kak Ester sebagai kakak rohani dan teman-teman Perkantas Lampung yang selalu memberikan semangat dan doa dalam setiap pergumulan penulis;
10. Keluarga KKN Bandar Negeri 2024 Mas Eka, Mas Nani, Mas Eenind, Mas Devi, Mas Nopel, dan Mas Hafidz.
11. Sahabat KMK UNILA, Gebi, Tata, Zita, Avila, Mba Tere, Febe, Tepang, Moms Emil, Cinta, yang senantiasa memberikan semangat kepada penulis sejak menjadi mahasiswa baru hingga sekarang;
12. Keluarga menuju S. Kom Annisa, Azzahra, dan Kurnia yang memberi motivasi untuk tekun;
13. Teman magang MSIB 6 Mohammad Aziz Riza, Delfrin Adiyatma Situmeang, dan Desvin Sitohang yang memberikan saran dalam penelitian ini;
14. Kepada dr. Cindy Riannie yang mendukung penulis dalam proses dukungan psikologi penulis;
15. Teman-teman “Asikkknya” sedari awal masuk FMIPA Unila, menemani penulis selama masa perkuliahan ini sedari awal;
16. Kepada seluruh pihak yang mendukung jalannya perkuliahan dan proses penulisan skripsi ini, mohon maaf karena keterbatasan penulis yang tidak dapat menyebutkan satu persatu, terima kasih.

Semoga Allah Bapa selalu senantiasa memberikan berkat-Nya dan rahmat yang berlimpah sebagai balasan atas kebaikan-kebaikan yang diberikan kepada penulis selama penyusunan skripsi ini. Akhir kata, penulis menyadari masih terdapat kekurangan dalam penulisan skripsi ini dan masih jauh dari

kesempurnaan, akan tetapi penulis berharap semoga skripsi ini dapat berguna dan bermanfaat bagi semua pihak khususnya bagi pengembangan ilmu hukum pada umumnya.

Bandar Lampung, 25 April 2025

Nathania Santa Nigel Simbolon

DAFTAR ISI

Halaman

DAFTAR TABEL.....	vii
DAFTAR GAMBAR.....	ix
I. PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian.....	3
1.4. Batasan Masalah.....	4
1.5. Manfaat Penelitian.....	4
II. TINJAUAN PUSTAKA.....	5
2.1. <i>Data engineering</i>	5
2.2. <i>Data mart</i>	5
2.3. <i>Data warehouse</i>	6
2.4. Python.....	7
2.5. Sentralisasi Data	8
2.6. <i>Extract Transform Load</i>	8
2.7. Google Cloud Platform.....	9
2.8. <i>4-Steps-Kimball-Methodology</i>	10
2.9. ISO/IEC 25012	10
2.10. Skala Likert	11
2.11. Penelitian yang Relevan	12
III. METODE PENELITIAN.....	14
3.1. Waktu dan Tempat Penelitian	14
3.1.1. Tempat Penelitian.....	14
3.1.2. Waktu Penelitian	14
3.2. Perangkat Penelitian	15
3.2.1. Perangkat Keras	15

3.2.2.	Perangkat Lunak.....	15
3.3.	Objek Penelitian	16
3.4.	Tahapan Penelitian	17
3.4.1.	Analisis Kebutuhan	17
3.4.2.	Sumber Data.....	18
3.4.3.	Implementasi ETL.....	18
3.4.4.	Evaluasi	38
IV.	HASIL DAN PEMBAHASAN.....	46
4.1.	Analisis Kebutuhan	46
4.2.	Sumber Data	47
4.3.	Implementasi ETL	50
4.3.1.	Skema <i>Data warehouse</i>	50
4.3.2.	Skema <i>Extract Transform Load</i>	66
4.3.3.	Penjadwalan	75
4.4.	Evaluasi	77
V.	SIMPULAN DAN SARAN.....	88
5.1.	Simpulan.....	88
5.2.	Saran	89
	DAFTAR PUSTAKA.....	91

DAFTAR TABEL

Tabel	Halaman
1. Skala Likert	12
2. Penelitian Terdahulu	12
3. Waktu Penelitian	14
4. Proses Pemodelan <i>Data warehouse</i> Data Akademik FTDS Hacktiv8.....	21
5. <i>Attributes</i> Data Peserta <i>Bootcamp Past batch</i>	28
6. <i>Attributes</i> Data Peserta <i>Bootcamp Current batch</i>	30
7. <i>Attributes</i> Data Aktivitas Mengajar Instruktur.....	32
8. <i>Attributes</i> Data Aktivitas Kehadiran Peserta <i>Current batch</i>	34
9. <i>Attributes</i> Data Peserta <i>Bootcamp All batch</i>	35
10. Tahap <i>Transformasi</i> Data	36
11. Beberapa Metrik <i>Data Quality Dimensions ISO/IEC 25012 Standard</i>	39
12. Daftar Kuesioner Evaluasi Kualitas Data	40
13. Contoh Pengisian Kuesioner Variabel Kualitas Data	42
14. Distribusi Kelas	45
15. <i>Bus Dimension Conformance Matrix</i>	52
16. Skema Dim_student	52
17. Skema Dim_instructor	53
18. Skema Dim_batch.....	54
19. Skema Tabel Dim_ct.....	55
20. Skema Tabel Dim_week	56
21. Skema Tabel Dim_Date	57
22. Skema Tabel Dim_Kelulusan	58
23. Skema fact_currentBatchPerformance_table	60
24. Skema fact_absensiCurrentBatch_table.....	60
25. Skema fact_pastBatchPerformance_table.....	61
26. Skema fact_instructorActivity_table.....	61

27. Skema fact_allBatchPerformance_table	61
28. Hasil Evaluasi <i>Standard Data Quality Dimensions</i> ISO/IEC 25012.	84
29. Distribusi Penilaian Keseluruhan dengan Skala Likert.....	85
30. Distribusi Kelas Setiap Indikator Evaluasi	86

DAFTAR GAMBAR

Gambar	Halaman
1. <i>Data warehouse</i>	7
2. Struktur Folder Data Akademik Hacktiv8	16
3. Tahapan Penelitian	17
4. <i>Four Step Kimball Methodology</i>	20
5. Proses ETL	23
6. <i>Worksheet Penilaian</i>	24
7. <i>Worksheet Peserta Past Batch</i>	25
8. <i>Extract Data Past Batch</i>	26
9. <i>Data Worksheet Past Batch</i>	27
10. Kolom Status dari <i>Data Source</i>	27
11. <i>Extract Data Current Batch</i>	29
12. <i>Data Worksheet Current Batch</i>	30
13. <i>Worksheet Teachers Simulations</i>	31
14. <i>Extract Data Aktivitas Instruktur</i>	32
15. Contoh <i>Worksheet Absensi Peserta Current batch</i>	33
16. <i>Worksheet Mentoring Log</i>	34
17. Proses <i>Transform</i>	36
18. <i>Spreadsheet Teachers Simulation</i>	48
19. <i>Spreadsheet Hasil Penilaian Peserta Bootcamp Current Batch</i>	49
20. <i>Spreadsheet Hasil Penilaian Peserta Bootcamp Past Batch</i>	50
21. Dimensi Student	53
22. Dimensi Instructor	54
23. Dimensi Batch	55
24. Dimensi Challenge Type	56
25. Dimensi Week	57
26. Dimensi Date	58

27. Dimensi Kelulusan	59
28. Skema Star Tabel fact_currentBatchPerformance_table	62
29. Skema Star Tabel fact_absensiCurrentBatch_table	62
30. Skema Star Tabel fact_pastBatchPerformance	63
31. Skema Star Tabel fact_allBatchPerformance.....	63
32. Skema Star Tabel fact_instructor_activity	64
33. <i>Dashboard</i> ETL pada Google Cloud Function	67
34. Hasil <i>Extract</i> Tabel New_Batch.....	68
35. Hasil <i>Extract</i> Tabel Past_Batch	69
36. Hasil <i>Extract</i> Tabel New_Batch_Absensi.....	69
37. Hasil <i>Extract</i> Tabel Sentralisasi_Aktivitas_Instruktur.....	69
38. Hasil <i>Extract</i> Tabel All_Batch	70
39. Tahapan <i>Cleaning Transformation</i>	71
40. Tahapan <i>Standardization Transformation</i>	72
41. Hasil <i>transformasi</i> standarisasi	73
42. Tahapan <i>Joining Transformation</i>	73
43. Hasil <i>Load</i> pada <i>Data warehouse</i>	74
44. Implementasi Cloud Scheduler	75
45. Implementasi Workflows	76
46. Hasil Metrik Akurasi <i>Rules</i> ‘a’	77
47. Hasil Metrik Akurasi <i>Rules</i> ‘a’	78
48. Hasil Metrik Akurasi <i>Rules</i> ‘b’	78
49. Hasil Metrik Akurasi <i>Rules</i> ‘C’.....	79
50. Hasil Metrik Akurasi <i>Rules</i> ‘d’	79
51. Hasil Metrik Konsistensi <i>Rules</i> ‘e’	80
52. Hasil Metrik Konsistensi <i>Rules</i> ‘f’	80
53. Hasil Metrik Kelengkapan <i>Rules</i> ‘g’	82
54. Hasil Metrik Keterkinian <i>Rules</i> ‘h’	83
55. Hasil Skor Kuesioner Metode Skala Likert	85

I. PENDAHULUAN

1.1. Latar Belakang

Teknologi informasi berperan penting dalam meningkatkan efisiensi dan efektivitas birokrasi dan dunia usaha (Mirza dkk., 2023). Lebih lanjut, pengembangan birokrasi digital juga merupakan upaya untuk meningkatkan efisiensi dan efektivitas birokrasi. Pembangunan birokrasi dapat dilakukan melalui pemanfaatan teknologi informasi dan pengolahan (Lilik Wienarni, 2019). Data merupakan bagian dari teknologi informasi yang menjadi elemen kunci dalam proses pengambilan keputusan dan pengembangan strategi bisnis. Data historis yang terkumpul dalam jumlah besar membuka kesempatan untuk melakukan analisis dari berbagai dimensi, yang dikenal sebagai analisis multidimensi (Agapito dkk., 2020). Selain itu, data bisa menjadi aspek penggerak dalam pengambilan keputusan bisnis dan memahami kondisi saat ini dari organisasi dengan mengekstrak nilainya melalui analisis data (Dollah & Aris, 2018). Aktivitas pengumpulan dan analisis data tidak terbatas pada organisasi yang berorientasi pada keuntungan, tetapi juga di bidang pendidikan (Yulianto, 2019).

PT. Hacktivate Teknologi Indonesia (Hacktiv8) yang merupakan salah satu perusahaan di bidang pendidikan non formal. Pada saat ini, dominasi teknologi informasi serta komputerisasi dan otomasi pekerjaan terlihat di hampir semua proses yang dilakukan di semua institusi. Fenomena ini menyebabkan peningkatan signifikan dalam jumlah data yang dihasilkan dan diproses, akan tetapi belum dimanfaatkan dengan baik, terutama untuk mengevaluasi pelaksanaan proses pembelajaran pada perusahaan *bootcamp* ini.

Saat ini data siswa dan aktivitas akademik pada program Full-Time Data Science (FTDS) tidak tersentralisasi dan masih menggunakan metode manual *data processing* atau spreadsheet sederhana. Meskipun lebih murah, metode ini sebenarnya tidak efektif untuk melakukan analisis data yang lebih mendalam dan

pengambilan keputusan yang berbasis data (Alsibhawi dkk., 2023). Hal ini menyebabkan kesulitan dalam *tracking* performa siswa dan keberjalanan *bootcamp* dengan baik. Data-data ini tersebar di berbagai *Google Spreadsheet* yang terpisah berdasarkan angkatan *bootcamp* dan lokasi cabang *bootcamp* yang diikuti oleh peserta. Setiap *sheet* berisi informasi yang spesifik untuk data siswa satu angkatan dan satu lokasi, yang menyebabkan aktivitas *report*, *monitoring*, dan *evaluation* menjadi tidak efisien. Tim akademik perusahaan harus memeriksa setiap *sheet* secara manual untuk mendapatkan insight dari data, yang memakan waktu dan rentan terhadap human *error*. Oleh karena itu, perlu adanya proses ETL untuk migrasi data menjadi *data warehouse*. Data yang diproses melalui proses ETL (*Extract Transform Load*) disimpan dalam *data warehouse* secara sistematis dan terorganisir sehingga mudah dipahami (Fernandez dkk., 2017).

Perusahaan mengusulkan ETL dengan penyediaan informasi untuk memudahkan pemahaman data yang telah dikumpulkan menggunakan *business intelligence tools* yaitu platform Google Cloud. Kebutuhan informasi mengenai aktivitas akademik dan progres belajar siswa melatarbelakangi perlunya ETL yang berfungsi untuk mengintegrasikan berbagai sumber data. Sumber data dianalisis sesuai kebutuhan dan tujuan yang ingin dicapai. Dalam artikel jurnal yang ditulis oleh Revathy Sreekumar tentang penelitiannya, "ETL Scheduling in Real-Time Data Warehousing", penulis membahas arsitektur *data warehouse*, biaya, dan pentingnya *scheduling*. Untuk menyelesaikan masalah tersebut, proses ETL yang efektif harus menggunakan *scheduler*. Karena penelitian ini berfokus pada institusi pendidikan dan membutuhkan penyajian data yang cepat, penghemat waktu sangat penting (Berliantara dkk., 2017).

Data yang dihasilkan akan disimpan menjadi *data warehouse* dan digunakan untuk memvisualisasikan setiap performa siswa dan keberjalanan *bootcamp* dengan baik. Seperti yang dilakukan oleh Dorina Kabakchieva menggunakan *Business Intelligence* untuk menyajikan informasi sebagai indikator statistik dan peramalan yang dapat memecahkan banyak masalah akademik dan mendukung proses pengambilan keputusan yang berkaitan dengan strategi untuk meningkatkan kualitas pembelajaran (Kabakchieva, 2015). Selanjutnya, penyimpanan data juga harus diperhatikan mengurangi kesalahan dan memudahkan analisis terutama

dalam penggunaan *insight* (Broman & Woo, 2018).

Lavalle melakukan sebuah studi yang melibatkan 3.000 eksekutif dan manajer bisnis. Survei ini menunjukkan bahwa 50% dari responden menganggap peningkatan manajemen data dan BI&A (*Business Intelligence and Analytics*) sebagai prioritas mereka. Survei ini juga mengungkapkan bahwa 20% dari mereka menyebutkan kekhawatiran terhadap kualitas data sebagai hambatan utama dalam sistem BI&A. Analisis data telah digunakan selama bertahun-tahun untuk mendukung pengambilan keputusan bisnis (Lavalle dkk., 2011). Saat ini, permasalahan yang dimiliki perusahaan yaitu belum memiliki data yang tersentralisasi untuk menunjang *monitoring*, antisipasi, dan pengambilan keputusan yang lebih cepat dan tepat dalam evaluasi. Penggabungan semua data siswa dan aktivitas akademik dari setiap *sheet* dari data *source* yang merupakan *raw data*. Sentralisasi data dengan implementasi ETL juga memungkinkan otomatisasi pembaharuan data untuk mendapatkan informasi kegiatan akademik *bootcamp* dan analisis kemudian digunakan untuk pengambilan *insight* serta keputusan pengembangan kurikulum.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka permasalahan yang dapat dirumuskan pada penelitian ini yaitu sebagai berikut: Bagaimana mengimplementasikan *extract, transform, load* pada sentralisasi data akademik dapat meningkatkan efisiensi *monitoring*, pelaporan, dan pengambilan keputusan?

1.3. Tujuan Penelitian

Tujuan dari penelitian ini sebagai berikut:

1. Mengimplementasikan *Extract, Transform, Load* pada sentralisasi data untuk menyatukan informasi siswa dan aktivitas akademik.
2. Mengefisiensikan proses penyediaan data yang akurat dan real time
3. Mempermudah pemantauan performa siswa *bootcamp* melalui data yang tersentralisasi

1.4. Batasan Masalah

Implementasi sentralisasi memindahkan data dari sumber terpisah ke *data warehouse* terpusat akan menggunakan teknologi Google Cloud Platform, Python, dan *Spreadsheet*. Penelitian yang dilakukan terbatas data transaksi akademik yang terdiri dari:

1. Data siswa dari *batch* awal (*Past batches*) 2021-2023
2. Data siswa yang enroll di tahun 2024 (per Januari 2024) hingga penelitian dimulai.
3. Data aktivitas Instruktur mentoring dan mengajar
4. Pembuatan *pipeline ETL* pada struktur *data warehouse* menggunakan skema *star* dengan metodologi *Kimball 4 Steps*.

1.5. Manfaat Penelitian

Beberapa manfaat yang diharapkan dengan adanya penelitian ini adalah sebagai berikut:

1. Data akademik yang tersentralisasi meningkatkan efisiensi operasional, proses *monitoring*, pelaporan, dan pengambilan keputusan menjadi lebih cepat dan efisien, mengurangi waktu dan usaha yang dibutuhkan untuk mengelola informasi.
2. Peningkatan kualitas laporan untuk meningkatkan program *bootcamp* dan mencapai tujuan strategis organisasi.

II. TINJAUAN PUSTAKA

2.1. *Data engineering*

Data engineering adalah proses pengembangan, implementasi, dan pemeliharaan sistem dan proses yang mengambil *raw data* dan menghasilkan informasi berkualitas tinggi yang konsisten, mendukung berbagai kasus penggunaan di hilir, seperti analisis dan *machine learning*. *Data engineering* berfokus di antara keamanan data, manajemen data, *dataOps*, *data architecture*, dan *software engineering*. Seorang *data engineer* mengelola siklus hidup *data engineering*, dimulai dari mendapatkan data dari *resources* hingga menyajikan data untuk digunakan dalam kasus seperti analisis atau *machine learning* (Reis & Housley, 2022).

Tahapan dalam siklus *data engineering* adalah sebagai berikut (Kretz, 2019):

1. *Data generation*: proses mengumpulkan data dari berbagai sistem sumber.
2. *Data storage*: menyimpan data dengan aman untuk pemrosesan dan analisis di masa mendatang.
3. *Data ingestion*: proses memasukkan data ke dalam sistem yang tersentralisasi.
4. *Data transformation*: pengubahan data ke dalam format yang berguna untuk analisis.
5. *Data serving*: penyajian data kepada *end user* untuk pengambilan keputusan dan keperluan operasional.

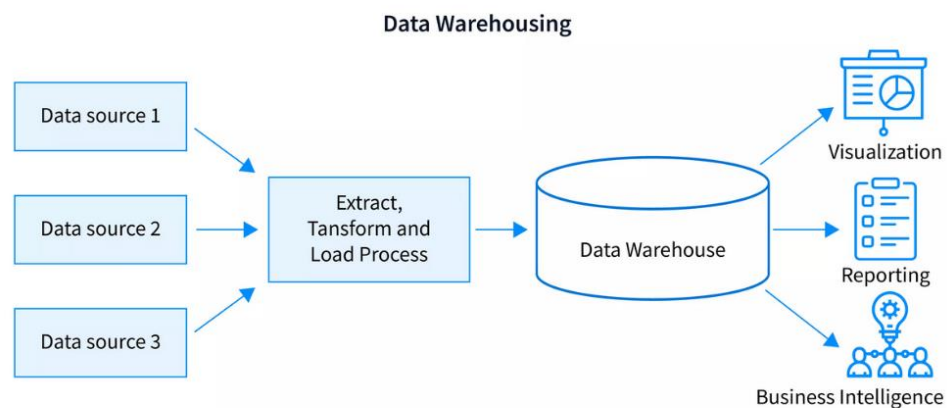
2.2. *Data mart*

Data mart adalah subset yang lebih terperinci dari *data warehouse* yang dirancang untuk melayani kebutuhan analitik dan pelaporan. *Data mart* berfokus pada satu sub organisasi, departemen, atau lini bisnis tertentu; setiap departemen

memiliki *data mart*-nya sendiri yang disesuaikan dengan kebutuhannya. Ini berbeda dengan *data warehouse* penuh yang melayani organisasi atau bisnis secara lebih luas. *Data mart* ada karena dua alasan. Pertama, *data mart* membuat data lebih mudah diakses oleh analis dan pengembang laporan. Kedua, *data mart* menyediakan tahap *transformasi* tambahan di luar yang disediakan oleh *pipeline* ETL atau ELT awal. Ini bisa meningkatkan kinerja secara signifikan jika laporan atau kueri analitik membutuhkan gabungan dan agregasi data yang kompleks, terutama ketika data mentah berukuran besar. Proses *transformasi* dapat mengisi *data mart* dengan data yang sudah digabungkan (Reis & Housley, 2022).

2.3. *Data warehouse*

Repositori *big data* ada dalam banyak bentuk, sesuai dengan kebutuhan perusahaan (Nambiar & Mundra, 2022). Repositori data yang efektif perlu menyatukan, mengatur, mengevaluasi, dan menyebarkan sejumlah besar sumber daya data untuk meningkatkan analitik dan kinerja *query* (Khine & Wang, 2018). Berdasarkan sifat dan skenario pengaplikasiannya, ada banyak jenis repositori data yang berbeda selain database relasional tradisional. *Data warehouse* digunakan untuk memberikan informasi tentang kinerja organisasi secara keseluruhan dan membantu proses pengambilan keputusan lebih efektif dan tepat (Moscoso-Zea dkk., 2018). Dua repositori data yang populer di antaranya adalah *data warehouse* (*enterprise data warehouse*) dan *data lake* (El Aissi dkk., 2022). *Data warehouse* (DW) adalah repositori data yang menyimpan data terstruktur, terfilter, dan diproses yang telah diolah untuk tujuan tertentu, sedangkan *data lake* (DL) adalah kumpulan data yang luas yang tujuannya tidak ditentukan (Nambiar & Mundra, 2022). Hal ini berarti, *data warehouse* menyimpan sejumlah besar data yang dikumpulkan oleh berbagai sumber, biasanya menggunakan skema yang telah ditentukan sebelumnya. Biasanya, *data warehouse* adalah *database* relasional yang dibangun khusus dan berjalan pada perangkat keras khusus baik di tempat atau di cloud (Rehman dkk., 2018).



Gambar 1. *Data warehouse*

Data warehouse telah digunakan secara luas untuk menyimpan data perusahaan dan mendukung aplikasi *business intelligence* dan analitik (Garani dkk., 2019). Dengan demikian, tujuan dari *data warehouse* adalah membantu pengguna data membuat analisis dan keputusan menjadi lebih cepat dan lebih baik (Homayouni H., 2018).

2.4. Python

Python adalah bahasa pemrograman skrip yang sangat mudah dipelajari dan sangat mudah dibaca, karena sintaksis kodenya sangat mirip dengan bahasa Inggris. Python diciptakan pada tahun 1989 oleh Guido van Rossum, Python sebenarnya sangat sederhana dalam sifatnya, tetapi juga dianggap oleh industri sebagai sangat kuat, cepat, dan dapat dijalankan di hampir semua lingkungan. Meskipun Python adalah bahasa pemrograman serbaguna yang dapat dijalankan di hampir semua tempat, ia cukup fleksibel dan gesit untuk memberikan kemampuan dalam menciptakan solusi yang sangat khusus untuk mengatasi masalah unik yang terkait dengan *big data* (Miller, 2017). Selain itu, Python dapat digunakan pada hampir semua *platform* sistem operasi dan mendukung paradigma pemrograman prosedural, fungsional, dan *object oriented programming* (OOP). Python adalah pilihan yang sangat baik untuk manipulasi *big data* seperti dalam hal menangani *outlier* dalam *big data* dalam memproses data besar (x, 2017).

2.5. Sentralisasi Data

Sentralisasi data adalah pendekatan repositori data secara terpusat yang memungkinkan *user* untuk mengakses data secara bersama-sama untuk berbagai file dan suatu instansi (Filiana dkk., 2020). Sesuai dengan kegunaannya cloud storage mampu memberikan beberapa kemudahan dalam mengakses data kapan saja dan dimana saja (Kholil & Mu'min, 2018). *User* dapat dengan mudah mengakses dan mengelola data dari satu tempat, yang meningkatkan efisiensi dan keamanan informasi bisnis

Keuntungan sentralisasi data adalah sebagai berikut (Jamaludin dkk., 2022):

1. Integritas data: data yang tersimpan di satu lokasi lebih mudah dikelola dan dijaga konsistensinya.
2. Keamanan tinggi: protokol akses dapat diterapkan untuk mengontrol siapa yang dapat mengakses data, mengurangi risiko kebocoran informasi.
3. Minim redundansi: dengan menyimpan data di satu tempat, kemungkinan terjadinya data yang berlebihan dapat diminimalkan.
4. Efisiensi biaya: pengelolaan dan pemeliharaan yang lebih sederhana dapat mengurangi biaya operasional.

Data yang tersentralisasi sangat cocok untuk organisasi yang memerlukan pengelolaan data yang terpusat dan terstruktur, tetapi juga memerlukan perhatian terhadap potensi risiko yang terkait dengan akses dan pemeliharaan sistem.

2.6. *Extract Transform Load*

ETL berfungsi sebagai metode dasar untuk sentralisasi data dari berbagai sumber, meningkatkan aksesibilitas dan kualitas data bagi organisasi. Evolusi terus-menerus dari teknologi dan alat ETL memainkan peran penting dalam mengatasi tantangan yang terkait dengan integrasi data, sehingga mendukung inisiatif *business intelligence* yang efektif (Yulianto & Ardhian, 2019).

ETL pada dasarnya diusulkan sebagai metode integrasi dan pemuatan data untuk analisis dan perhitungan. Ketika basis data semakin meluas pada tahun 1970-

an, akhirnya hal ini menjadi mekanisme utama untuk pemrosesan data dalam operasi *data warehouse*. ETL dapat mengorganisir data menggunakan seperangkat *business rules* untuk memenuhi kebutuhan *business intelligence* tertentu, seperti *monthly report*, tetapi juga dapat menangani analisis yang lebih rumit untuk meningkatkan operasi belakang atau pengalaman *end user*. Sebuah organisasi akan sering menggunakan ETL untuk mengekstrak data dari sumber sistem, membersihkan data untuk meningkatkan kualitas data dan menetapkan konsistensi, serta memuat data ke dalam basis data target (Khan dkk., 2024).

2.7. Google Cloud Platform

Google Cloud Platform merupakan berbagai layanan Google yang saling terintegrasi untuk mendukung kebutuhan pengembangan aplikasi dan analisis data. Penggunaanya pada penelitian ini meliputi fitur-fitur Cloud Scheduler, Cloud Functions, BigQuery, dan Workflows, GCP menawarkan solusi fleksibel dan efisien bagi bisnis dalam mengelola operasional mereka di era digital saat ini. Cloud Scheduler adalah layanan yang memungkinkan pengguna untuk menjadwalkan tugas secara otomatis. Layanan ini dapat digunakan untuk memicu fungsi atau alur kerja pada waktu tertentu atau secara berkala.

Google Cloud Functions adalah layanan *serverless* yang memungkinkan pengembang untuk menjalankan kode sebagai respons terhadap berbagai peristiwa atau pemicu tertentu tanpa perlu mengelola infrastruktur *server*. Cloud Functions secara dinamis mengalokasikan sumber daya sesuai dengan jumlah pemanggilan fungsi yang diterima, sehingga ideal untuk beban kerja yang bersifat berbasis peristiwa seperti *Internet of Things* (IoT), ETL (*Extract, Transform, Load*), *webhook*, integrasi dengan *Apache Kafka*, dan analitik data. Kemudian, BigQuery adalah platform Google yang dioptimalkan untuk kueri analitis, menampilkan data dalam format tabel, baris, dan kolom dan mendukung semantik transaksi *database*. Terakhir adalah Google Workflow adalah *tools* yang dapat menggabungkan layanan API Google Cloud yang memungkinkan integrasi aplikasi, otomatisasi proses, *pipeline* data, dan *machine learning*. Workflows di Cloud Run Functions menangani pemformatan permintaan, *retrial*, dan *cron jobs* untuk menyelesaikan

operasi yang berjalan lama. Penggunaannya adalah untuk mengotomatiskan alur kerja proses ETL (Google, 2024).

2.8. 4-Steps-Kimball-Methodology

Metode Kimball adalah pendekatan untuk membangun *data warehouse* yang diusulkan oleh Ralph Kimball pada tahun 1980-an. Metode ini juga dikenal sebagai *dimensional modelling*. Selain metode inmon, metode ini sudah menjadi standar pengembangan *data warehouse*. Metode ini menggunakan proses pengembangan *data warehouse* dimulai dari *bottom to top*, yaitu dengan membuat *data mart* di divisi yang sudah memiliki kebutuhan. Setelah itu, proses integrasi *data mart* dilakukan untuk membuat data yang tersimpan di *data warehouse* berbentuk multidimensional dan telah dilakukan denormalisasi.

Terdapat empat Langkah. Pertama, Kimball berpendapat bahwa pertama kali pengguna harus memilih salah satu proses bisnis yang akan digunakan sebagai obyek dari *data warehouse*. Proses bisnis ini harus menjadi elemen yang mendukung aspek analisis dan pengambilan keputusan organisasi yang khusus. Pendeklarasian level detail atau *grain* data adalah langkah kedua yang harus diambil Pengguna dalam menentukan data yang akan dimasukkan sebagai data fakta. Ini dapat berupa level harian, bulanan, atau transaksi per kata. Langkah ketiga adalah identifikasi dari dimensi-dimensi yang akan menempatkan fakta-fakta ini dalam konteks. Dimensi ini biasanya adalah atribut-atribut deskriptif, seperti waktu, lokasi, atau demografi siswa, dan lain-lain. Langkah keempat diambil pengguna adalah dimensi data-data yang akan diukur atau direkam dalam tabel fakta. Pada penelitian biasanya berupa nilai numerik yang dapat diolah (Moscoso-Zea dkk., 2018)

2.9. ISO/IEC 25012

ISO/IEC 25012 merupakan standar internasional menetapkan standar untuk kualitas produk data dengan Peraturan Perlindungan Data Umum (GDPR) (Vesjolijs, 2024). Tujuan dari penelitian ini adalah untuk mengevaluasi kualitas

data pada *data marts* akademik Hacktiv8 dengan menggunakan standar ISO/IEC 25012:2008 sebagai acuan penilaian, dengan fokus pada identifikasi kekuatan dan kelemahan data dalam memenuhi karakteristik kualitas yang ditetapkan oleh standar tersebut. Standar ini digunakan mengevaluasi kualitas data inheren ETL. Kualitas data inheren mengacu pada sejauh mana karakteristik kualitas data memiliki potensi intrinsik untuk memenuhi kebutuhan yang dinyatakan dan tersirat ketika data digunakan dalam kebutuhan yang ditentukan. Dari sudut pandang yang melekat, kualitas data merujuk pada data itu sendiri. Beberapa pendekatan yang menjadi metrik evaluasi penelitian ini adalah akurasi, kelengkapan, konsistensi, dan terkini. Hasil evaluasi ini diharapkan dapat memberikan rekomendasi konkret untuk meningkatkan kualitas data, sehingga memungkinkan pengembang kurikulum lebih efisien menganalisis kegiatan akademik *bootcamp*.

2.10. Skala Likert

Skala likert adalah metode yang mengonversi nilai kualitatif menjadi nilai kuantitatif untuk memungkinkan analisis statistik yang optimal. Skala ini umum digunakan untuk mengukur sikap atau perilaku responden dengan menanyakan sejauh mana mereka setuju atau tidak setuju dengan suatu pernyataan tertentu. Namun, dalam penelitian-penelitian yang disebutkan pada penelitian ini, tidak ada metode atau pendekatan khusus yang diusulkan untuk menganalisis persyaratan kualitas produk dengan standar ISO/IEC 25012 untuk evaluasi berdasarkan skala Likert (Iqbal & Babar, 2016). Melalui pendekatan ini, kita dapat mengklasifikasikan berbagai evaluasi kualitas data dan efisiensi proses dari berbagai sudut pandang *stakeholders* dan *end user*, untuk menentukan seberapa efisien persyaratan kualitas tersebut apakah sangat sudah efisien atau belum. Tabel berikut merupakan contoh representasi skala likert 5 poin.

Tabel 1. Skala Likert

Skor	Deskripsi
5	Sangat Setuju
4	Setuju
3	Cukup Setuju
2	Kurang Setuju
1	Tidak Setuju

2.11. Penelitian yang Relevan

Beberapa penelitian yang telah dilakukan dan berkaitan dengan implementasi ETL pada penyimpanan data terpusat dapat dilihat pada tabel 2 berikut.

Tabel 2. Tabel Penelitian Terdahulu

No	Penulis	Judul	Metode	Hasil
1.	Muryjas dkk., 2021	<i>Managing the Process of Evaluation of the Academic Teachers with the Use of Data mart and Business Intelligence</i>	ETL Research dan interview	Penelitian ini menghasilkan model Konseptual <i>Data mart</i> asli untuk mengevaluasi pekerjaan jarak jauh dosen yang mengimplementasikan pendekatan analitis dengan <i>Dashboard</i> yang menyajikan nilai-nilai KPI untuk evaluasi.
2.	Yulianto, 2019	<i>Extract Transform Load (ETL) process in distributed database academic data warehouse</i>	Kimball's four-step dimensional design process	Perancangan ETL yang mengikuti alur ETL: Ekstraksi, Pembersihan, Konformasi, dan Pemuatan. Proses ETL yang dirancang dengan alat seperti Pentaho <i>Data Integration</i> , dapat memastikan integrasi data yang efektif dan meningkatkan kinerja <i>Data warehouse</i> .

No	Penulis	Judul	Metode	Hasil
3.	Soriano & Dela Cruz (2020)	<i>Implementing an ETL-driven Data Integration Framework for State Universities and Colleges</i>	ETL-driven Data Integration	<i>Framework data integration</i> efektif untuk universitas di Filipina yang memiliki <i>data source</i> yang beragam, serta hasil implementasi pentingnya teknologi ETL dalam proses tersebut.
4.	(Wijaya dkk., 2017)	Perancangan <i>Data warehouse</i> Nilai Mahasiswa Dengan <i>Kimball Nine-Step Methodology</i>	Kimball nine-step methodology	Rancangan <i>Data warehouse</i> nilai mahasiswa merangkum data operasional kedalam bentuk yang lebih ringkas dan bermanfaat.
5.	(Vesjolijs, 2024)	<i>The E(G)TL Model: A Novel Approach for Efficient Data Handling and Extraction in Multivariate Systems</i>	Memodifikasi proses ETL yang ada menggunakan GenAI dengan evaluasi ISO/IEC 25012	Model EGTL (ekstrak, hasilkan, transfer, muat), <i>framework</i> teoritis yang dirancang untuk meningkatkan proses ETL tradisional dengan mengintegrasikan langkah ‘generate’ yang memanfaatkan kecerdasan buatan generatif (GenAI).

III. METODE PENELITIAN

3.1. Waktu dan Tempat Penelitian

3.1.1. Tempat Penelitian.

Penelitian ini dilakukan di PT Hacktivate Teknologi Indonesia yang beralamat di Jl. Sultan Iskandar Muda Nomor 7, RT 5/RW 9, Kebayoran Lama Selatan, Kebayoran Lama, Jakarta Selatan, DKI Jakarta, serta di Jurusan Ilmu Komputer, Universitas Lampung, yang beralamat di Jl. Prof. Dr. Ir. Sumantri Brojonegoro Nomor 1, Gedong Meneng, Kecamatan Rajabasa, Kota Bandar Lampung, Provinsi Lampung.

3.1.2. Waktu Penelitian

Penelitian ini dilakukan pada semester genap Tahun Ajaran 2023/2024. Waktu penelitian ditampilkan dalam tabel berikut ini.

Tabel 3. Waktu Penelitian

Tahapan Sentralisasi	September	Oktober	November	Desember
Analisis Sumber Data				
Analisis Kebutuhan				
Implementasi ETL				
Evaluasi				
Pembuatan Laporan				

Tabel 3 menunjukkan bahwa waktu yang dihabiskan untuk penelitian dibagi menjadi tujuh tahapan selama bulan September hingga Desember. Analisis sumber data dilakukan selama dua minggu pertama di bulan September. Selanjutnya, analisis kebutuhan dimulai dari minggu kedua hingga minggu ketiga bulan September. Tujuannya adalah untuk menganalisis kebutuhan proses dan teknis

terkait proses sentralisasi data. Implementasi ETL, berlangsung dari minggu terakhir September hingga minggu pertama November. Implementasi berlangsung selama enam minggu. Evaluasi berlangsung di minggu ketiga dan keempat di bulan November. Tahap ini mencakup peninjauan dan pengukuran efektivitas implementasi ETL. Lalu tahapan terakhir pembuatan laporan hasil penelitian berlangsung selama di minggu keempat November hingga minggu kedua Desember.

3.2. Perangkat Penelitian

3.2.1. Perangkat Keras

Adapun perangkat keras yang digunakan dalam penelitian ini adalah sebagai berikut:

- a. System Manufacturer: ASUS
- b. System Model: ASUS Vivobook 14 K413E
- c. Processor: Intel Core I5 Gen 11
- d. GPU: Intel Iris XE MAX
- e. RAM: 8.00 GB
- f. System Type: 64 bit
- g. Storage: SSD 512 GB

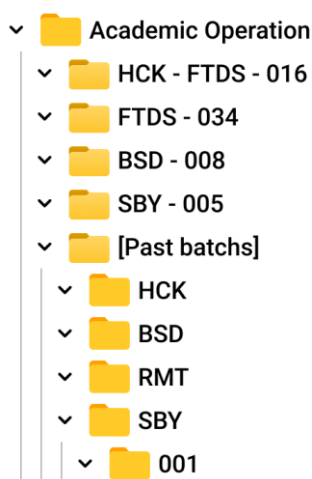
3.2.2. Perangkat Lunak

Beberapa perangkat lunak atau alat pada pengembangan sentralisasi ini, antara lain:

- a. Python 3.9.12 sebagai bahasa pemrograman untuk melakukan *scraping* data.
- b. Database PostgreSQL 13.
- c. *Google Spreadsheet* sebagai *data source*
- d. Visual Studio Code sebagai Code Editor.
- e. Github sebagai Version Control Tools.

3.3. Objek Penelitian

Fokus penelitian ini adalah data akademik dan aktivitas instruktur dari Perusahaan *Bootcamp* Data Science Hacktiv8 Indonesia yang disimpan dalam bentuk *Google Spreadsheet* terpisah setiap file dan folder sesuai angkatan peserta. Data ini juga disimpan dalam *google drive* pada folder *Academic Operations* yang terdiri 2 jenis data peserta *bootcamp* yaitu *Past batch* dan *Current batch*. Berikut adalah gambar dari struktur *database* akademik perusahaan *bootcamp* yang disimpan dalam *platform google drive*:

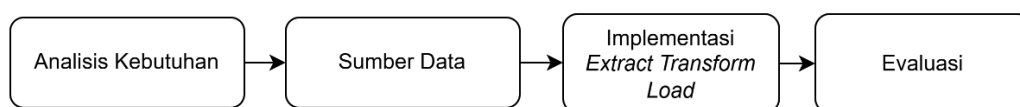


Gambar 2. Struktur Folder Data Akademik Hacktiv8

Gambar 2 yang terlampir merupakan struktur folder data *bootcamp data science* di *Hacktiv8* terorganisasi dengan baik dalam struktur folder *Google Drive*. Setiap folder mewakili angkatan atau *batch* tertentu dengan kode (misalnya HCK, BSD, SBY), dan data ini dibagi antara *batch* yang sudah selesai ("*Past batchs*") dan *batch* yang sedang berlangsung. Pada penelitian folder yang digunakan hanya folder yang memiliki *sheet* penilaian saja terbatas hanya untuk HCK -FTDS - 016, FTDS - 034, BDS - 008, SBY – 005, [*Past batchs*] termasuk folder lain di dalamnya. Folder dengan nama HCK -FTDS - 016, FTDS - 034, BDS - 008, SBY – 005 merupakan kategori *current batch* tersimpan pada *root* folder.

3.4. Tahapan Penelitian

Penelitian ini mengumpulkan data dengan dua cara: observasi dan wawancara sebagai data sekunder. Data primer berasal dari data akademik operasional Hacktiv8. Data Primer dikumpulkan dengan cara membuat *program scraping value* dari *spreadsheet* yang ada pada semua folder ‘Academic Operation’. Moscoso-Zea, Paredes-Gualtor, dan LujánMora dalam penelitiannya menyarankan empat tahapan untuk mengembangkan sentralisasi data dengan *data warehouse* untuk penelitian sejenis ini (Moscoso-Zea dkk., 2018). Berikut gambaran tahapan penelitian dalam bentuk diagram.



Gambar 3. Tahapan Penelitian

Penelitian dimulai dengan analisis sumber data dengan mengidentifikasi sumber data yang akan disimpan terpusat, termasuk memahami format, struktur. Setelah itu, dilanjutkan dengan analisis kebutuhan yang bertujuan untuk mengumpulkan dan memahami kebutuhan data yang diperlukan untuk *reporting* dan *monitoring*. Tahap selanjutnya adalah Implementasi *Extract Transform Load* (ETL), di mana dilakukan proses dilakukan *scraping* dan mengekstraksi data dari sumber-sumber yang telah diidentifikasi, mentransformasi data tersebut menjadi format yang lebih terstruktur dan dinormalisasi, kemudian dimuat ke dalam sistem penyimpanan data yang baru. Terakhir, tahap evaluasi merupakan proses yang telah dilakukan ditinjau untuk menilai efektivitasnya serta memastikan bahwa sentralisasi data memenuhi kebutuhan yang telah ditetapkan, dan benar bahwa dapat memudahkan *monitoring*.

3.4.1. Analisis Kebutuhan

Tahapan analisis kebutuhan dalam penelitian ini dilakukan menggunakan observasi. *Stakeholders* membutuhkan informasi dalam mendukung pengelolaan

data akademik *bootcamp*. Metode ini diterapkan untuk memahami secara lebih jelas proses pengelolaan data akademik serta kebutuhan dalam meningkatkan efisiensi operasional. Proses ini mencakup pemilihan, pengecekan, dan identifikasi jenis dan struktur data yang diperlukan untuk mengetahui jenis dan jenis analisis yang diperlukan. Observasi dilakukan untuk mengidentifikasi detail kebutuhan data untuk kebutuhan *monitoring* laporan analisis proses akademik di *bootcamp*, meliputi pengelolaan *sheet* penilaian, pembagian *buddy* untuk setiap murid, *log mentoring*, jadwal *bootcamp*, dan jadwal terkait dengan aktivitas peserta maupun instruktur. Tahapan ini juga untuk mengetahui komparasi kualitas kurikulum yang ada.

3.4.2. Sumber Data

Sumber data diperoleh dari operasional akademik *Full Time Data Science* Hacktiv8. Data tersebut merupakan data aktivitas peserta *bootcamp* dari Tahun 2020-2024. Data meliputi penilaian pembelajaran peserta *bootcamp* yang terdapat pada *spreadsheet* dari salah satu *sheet* bernama '*cummulative*'.

3.4.3. Implementasi ETL

Proses ETL (*Extract Transform Load*) merupakan bagian penting proses integrasi *data warehouse* agar menjadi format yang lebih terstruktur dan terpusat (Yulianto, 2019). Implementasi ETL bertujuan mengekstrak data dari berbagai data akademik Hacktiv8 yang memiliki format *spreadsheet*, mengubahnya menjadi format yang sesuai dengan aturan bisnis, dan kemudian memuatnya ke dalam struktur *data warehouse* yang sudah dirancang. Proses ETL sangat penting untuk integrasi *data warehouse* karena memungkinkan data dari operasional dimasukkan ke dalam *data warehouse*. Strategi sentralisasi data seperti ini dapat diakses dengan *cost* lebih murah untuk mempermudah akses ke data instansi. Penelitian ini bertujuan untuk mengembangkan teknologi penyimpanan data secara terpusat yang memungkinkan pengguna mengakses data secara bersama untuk berbagai file dan untuk sentralisasi data instansi. Sesuai dengan kegunaannya, penyimpanan *cloud* mampu memungkinkan akses data kapan saja dan dimana saja (Kholil & Mu'min, 2018).

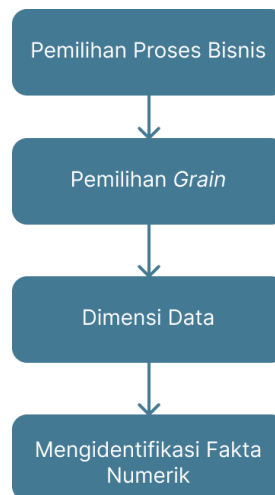
ETL yang dibangun oleh bahasa Python dapat bekerja secara otomatis menggunakan *scheduler* di dalamnya. Metode ini dapat menjalankan proses *extract*, *transform*, dan *load* dari basis data berupa *flat file (spreadsheet)* ke *data warehouse* dengan waktu proses yang cepat dan terpusat.

Sentralisasi data juga membantu meningkatkan kinerja dan mempermudah navigasi data. Salah satu tujuan utama dari *data warehouse* adalah menyediakan data agregat dan deret waktu yang siap digunakan, sehingga mengurangi beban perhitungan pada saat eksekusi *query*. Dengan demikian, *data warehouse* dapat memberikan efisiensi yang lebih besar dalam mendukung analisis dan pengambilan keputusan.

Sentralisasi data dengan implementasi ETL ini secara otomatis akan terjadwal melakukan proses ETL yang akan dijalankan secara berkala. Dalam konteks ETL (*Extract, Transform, Load*), penggunaan *scheduler* berperan sangat penting dalam mengotomasikan proses ETL. Hal ini dilakukan setiap minggu sekali, dengan pertimbangan akan selalu ada angkatan peserta *bootcamp* yang baru dalam rentang waktu tersebut dan *reporting* tim operasional. Sentralisasi data sering dilakukan untuk mengurangi jumlah join yang diperlukan dalam pengambilan data pada *database* relasional, dengan menambahkan data yang seharusnya ditempatkan di tabel terpisah. Ini bisa memudahkan pengambilan data tetapi mengorbankan efisiensi penyimpanan dan potensi inkonsistensi data. Hasil sentralisasi dapat dimodelkan kembali untuk proses penyimpanan data yang terstruktur.

3.4.3.1. Skema *Data warehouse*

Penelitian ini pada prosesnya menggunakan pendekatan bottom-up dari Kimball dalam mengembangkan model multidimensi (Yulianto, 2019). Berdasarkan proses desain dimensional empat langkah dari Kimball, aktivitas desain model dibangun melalui langkah-langkah berikut:



Gambar 4. *Four Step Kimball Methodology*

Berdasarkan gambar tersebut menunjukkan alur proses yang terdiri dari empat tahap dalam perancangan *data warehouse* , yaitu:

- a. Pemilihan proses bisnis: yang akan dimodelkan, dengan mempertimbangkan tujuan analisis bisnis dan sumber data yang tersedia.
- b. Menentukan *grain*: mendefinisikan detail representasi tabel fakta individu yang berkaitan dengan pengukuran proses bisnis yang diputuskan.
- c. Memilih dimensi: menentukan atribut yang mendeskripsikan pengukuran tabel fakta. Dimensi yang umum adalah waktu dan tanggal
- d. Mengidentifikasi fakta numerik: mendefinisikan ukuran yang akan dimasukkan dalam tabel fakta. Fakta harus konsisten dengan granularitas yang telah ditetapkan.

Keuntungan penting dari pendekatan ini adalah terdiri dari *data mart* sebagai representasi unit bisnis dalam hal ini memberikan *insight* terhadap kualitas kurikulum, yang mudah diintegrasikan antar *data mart*. Untuk pertimbangan pada skala *data warehouse* perusahaan, metode ini juga mudah diperluas dari *data warehouse* yang sudah ada serta menyediakan kemampuan pelaporan.

Metode *Kimball 4 steps methodology* untuk membangun *data warehouse* dimulai dengan membuat tabel fakta dan dimensi. Tabel fakta mengandung metrik yang biasanya berasal dari tabel transaksi di *relational database*, dan tabel dimensi

mengandung atribut yang biasanya berasal dari tabel master di relational database. Metode ini mengabaikan standar normalisasi (Yulianto, 2019).

Tabel 4. Proses Pemodelan *Data warehouse* Data Akademik FTDS Hacktiv8

No	Tahap	Metode
1.	Pemilihan proses bisnis	• Analisis progres belajar peserta <i>bootcamp current_batch</i> • Analisis progres belajar peserta <i>bootcamp past_batch</i> • Analisis aktivitas mengajar instruktur (<i>mentoring, technical class</i>) • Analisis <i>bootcamp</i> informasi kehadiran tiap peserta <i>current batch</i> • Analisis peserta aktif dan alumni (<i>all_batch</i>)
2.	Pemilihan <i>grain</i>	• Informasi progres belajar dari penilaian setiap peserta <i>bootcamp past batch</i> setiap minggu, <i>challenge type</i> dan <i>batch</i>. • Informasi progres belajar dari penilaian setiap peserta <i>bootcamp current batch</i> untuk setiap siswa di setiap minggu, <i>challenge type</i> dan <i>batch</i> nya. • Informasi kehadiran setiap peserta <i>current batch</i> dalam seminggu • Informasi aktivitas jadwal mengajar instruktur setiap hari • Informasi peserta setiap aktif dan alumni keseluruhan <i>batch</i> termasuk tanggal mulai pembelajaran setiap <i>batch</i>
3.	Dimensi	Dimensi data terkait: • Dimensi Peserta: berisi informasi terkait nama peserta <i>bootcamp</i>. • Dimensi Instruktur: berisi informasi terkait id dan nama instruktur. • Dimensi <i>Batch</i>: berisi informasi terkait <i>batch, phase</i>, dan <i>location</i>, date of enrolment • Dimensi Week: berisi Week. • Dimensi Status kelulusan: berisi <i>Weekly Status</i> dan Status. • Dimensi <i>Challenge_type</i>: berisi tipe ujian. • Dimensi date: berisi informasi per tanggal tertentu yang menjelaskan kegiatan pada saat itu.

No	Tahap	Metode
4.	Fakta Numerikal	• Tabel fakta progres belajar peserta <i>bootcamp current batch</i> • Tabel fakta progres belajar peserta <i>bootcamp past_batch</i> • Tabel fakta kehadiran tiap <i>peserta current batch bootcamp</i> • Tabel fakta aktivitas mengajar instruktur • Tabel fakta peserta aktif dan alumni (<i>all_batch</i>)

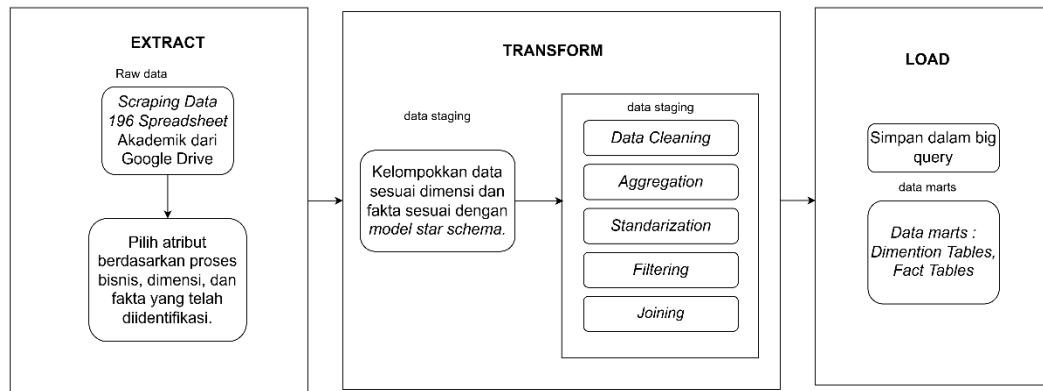
Membuat *output data warehouse* harus dapat dipahami oleh pengguna non-teknis yang sebagian besar terdiri dari *lead operational* program dan rekan-rekannya adalah salah satu tujuan utama dari pekerjaan ini. *Output data warehouse* harus merender data kompleks dari banyak *sheet* terpisah yang tidak terstruktur ke dalam format yang sesuai untuk tabel, di mana setiap baris menggabungkan skor, dan fitur lainnya dari data aktivitas *bootcamp*. Karena itu, peneliti memilih format *spreadsheet* untuk menyusun data secara terstruktur atau berbentuk tabel.

Pada tahapan desain tabel fakta numerikal sentralisasi digunakan untuk data. Dalam *Kimball 4-step design*, tabel fakta dan tabel dimensi didesain menggunakan skema *star* yang berkategori. Tabel fakta dalam skema ini menghubungkan dimensi-dimensi secara langsung, sehingga memungkinkan akses cepat tanpa perlu melakukan banyak “*join*” seperti dalam model ternormalisasi. Pada perancangan *data warehouse* kali ini, skema *star* dipilih karena penelitian ini memerlukan lebih dari 2 tabel fakta yang menggunakan tabel dimensi yang sama yang mempertimbangkan proses bisnis yang cukup banyak dan terorganisir yang telah dinormalisasikan sehingga data tidak mengalami redundansi dan *query*-nya lebih sederhana (Dharayani dkk., 2015). Masing-masing skema mewakili satu data mart yang akan mewakili satu tujuan bisnis. Data mart dalam hal ini satu skema *star* akan digabungkan menjadi skema *data warehouse* yang utuh.

3.4.3.2. Skema *Extract Transform Load*

Proses secara berkala diekstrak dan diintegrasikan ke dalam *data warehouse* disebut proses ETL (Andriansyah, 2022). Berikut merupakan penjelasan mengenai

proses implementasi ETL (*Extract Transform Load*) yang akan digunakan oleh peneliti dalam rancang bangun *data warehouse* yang berasal dari data *staging* sentralisasi data akademik *bootcamp*. Berikut gambaran umum proses ETL yang dilakukan pada penelitian ini.



Gambar 5. Proses ETL

A) *Extract*

Tahapan *extract* dalam penelitian ini dilakukan untuk mengumpulkan data yang diperlukan agar dapat diproses ke tahap *transformation*. Alat yang digunakan adalah program yang dibuat dengan bahasa pemrograman Python. Otomasi ekstraksi ini dilakukan untuk menangani *manual processing* dengan data yang tidak terstruktur dan terpusat, sehingga tidak perlu menyalin setiap atribut secara manual dari spreadsheet yang dipisah per *batch* untuk digunakan dalam proses ETL ini. Ekstraksi data yang diperlukan juga harus berdasarkan analisis kebutuhan yang telah dilakukan sebelumnya. Data proses ekstraksi hanya mencakup data yang dibutuhkan saja untuk implementasi ETL ini.

			Total Score			Min Score			Total Score			Min Score			Total Score		
			31			21.7			42			29.4			25		
No	Name	Buddy	LC 1			LC 2			LC 3			LC 4			LC 5		
			Score	Score (%)	Status	Score	Score (%)	Status	Score	Score (%)	Status	Score	Score (%)	Status	Score	Score (%)	Status
1	Ahmad Dani Rifai	Yuda	15.5	50.00%	Danger	31	73.81%	Safe	21	73.81%	Safe	21	73.81%	Safe	21	73.81%	Safe
2	Gracia Valerine	Yuda	22	70.97%	Safe	31	73.81%	Safe	21	73.81%	Safe	21	73.81%	Safe	21	73.81%	Safe
3	Muhammad Fiqih Al-ayubi	Yuda	29	93.55%	Safe	39	92.86%	Safe	25	92.86%	Safe	25	92.86%	Safe	25	92.86%	Safe
4	Ryan Trisnadi	Yuda	18.5	59.68%	Danger	32	76.19%	Safe	23	76.19%	Safe	23	76.19%	Safe	23	76.19%	Safe
5	Syihabuddin Ahmad Satisma	Lis	28.25	91.13%	Safe	33.6	80.00%	Safe	24.98	80.00%	Safe	24.98	80.00%	Safe	24.98	80.00%	Safe
6	Vicky Belario	Lis	14.6	47.10%	Danger	26.8	63.81%	Danger	21.94	63.81%	Danger	21.94	63.81%	Danger	21.94	63.81%	Danger
7	Romano Oka Cleodora	Lis	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger
8	Vincent Kaunang	Lis	10.9	35.16%	Danger	28.617	68.14%	Danger	23	68.14%	Danger	23	68.14%	Danger	23	68.14%	Danger
9			0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger
10			0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger
11			0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger
12			0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger
13			0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger
14			0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger
15			0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger	0	0.00%	Danger

Gambar 6. *Worksheet* Penilaian

Proses *extract* yang dilakukan oleh peneliti dimulai dengan mengiterasi folder dan file dalam Google Drive “Academic Operations” dan memilih folder serta *spreadsheet* yang akan diintegrasikan. Setiap *spreadsheet* yang berisi penilaian harian dan mentoring akan di-*scrape* menggunakan bahasa pemrograman Python dengan lingkungan Google Cloud Storage yang dapat menghubungkan berbagai *tools* Google. Ekstraksi data dilakukan dengan cara *scraping* sumber data berupa file .xlsx yang disimpan di Google Drive dan masih memiliki format seperti gambar 6. Data dipilih berdasarkan kebutuhan yang telah diidentifikasi sebelumnya dari *worksheet* yang sesuai. Terdapat tiga *worksheet* yang diperlukan dalam proses ETL, yaitu “Cumulative,” “Mentoring Log,” dan “Teachers Simulation.” Sumber data diambil menggunakan *library* yang tersedia pada Python, seperti Pandas dan Scikit-Learn. Selain itu, pustaka lain yang digunakan mencakup Google Spreadsheet untuk mengakses Google Sheets, Google Workspace API, dan *google.auth.default* untuk menangani kredensial dan autentikasi pada Google Drive. Data ini hanya diambil dari setiap nilai *cell*-nya dengan membuat metode *scraping* sesuai kebutuhan tabel untuk setiap jenis data yang akan diambil dari masing-masing *spreadsheet*. Berikut adalah tahapan *extract* untuk setiap data yang diintegrasikan.

1) Data *Past batch*

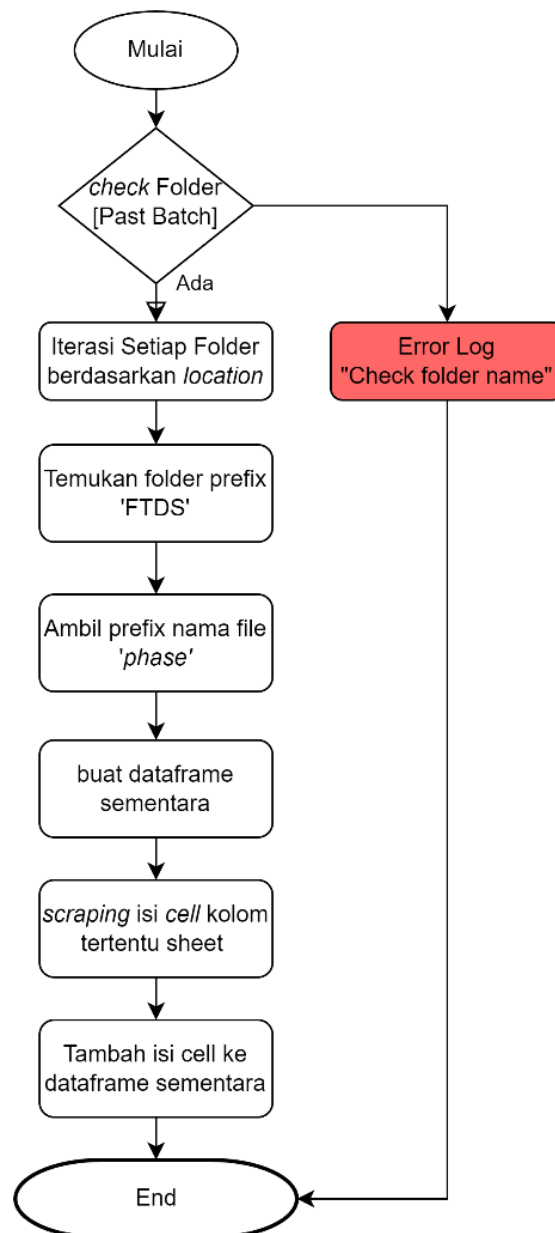
Tahapan *extract* pada data *past batch* dilakukan dengan mengiterasi seluruh

folder yang pada folder google drive yang bernama ‘[*PAST BATCH*]’ pada *directory* ‘Academic Operations’. Data yang diperlukan untuk proses ekstraksi dan sentralisasi data ini memerlukan *worksheet* penilaian siswa *bootcamp* yang bernama ‘*cummulative*’.

Data Science Fulltime Program Phase 1 Cumulative Score								
			320	224	290			
No	Name	Buddy	Milestones 1			Milestones 2		
			Score	Score (%)	Status	Score	Score (%)	Status
1	Dandi Alfikri	Raka	350	109,38%	Safe	293	101,03%	Safe
2	Arya Bandoro	Raka	273	85,31%	Safe	225	77,59%	Safe
3	Aldo Dwicahyo	Afif	300	93,75%	Safe	243	94,14%	Safe
4	Nisa Aprilia M	Afif	300	93,75%	Safe	254	97,93%	Safe
5	Brian Budiarto	Danu	301	101,88%	Safe	274	103,10%	Safe
6	Anggito Zainul	Fahmi	290	90,63%	Safe	283	107,24%	Safe
7	Michael Andrian	Fahmi	313	107,19%	Safe	273	104,48%	Safe
8	Dennis Herdiawan	Irfan	286	96,88%	Safe	276	95,17%	Safe
9	Dinda Tirta Rahayu	Irfan	273	92,81%	Safe	278	95,86%	Safe
10	Irpan Habibi Sunarya Putra	Danu	305	95,31%	Safe	275	94,83%	Safe
11	Muhammad Rizki Herfian	Afif	300	93,75%	Safe	278	106,21%	Safe
			n	n	n			

Gambar 7. *Worksheet Peserta Past Batch*

Proses *extract* dari data *source* berformat seperti gambar 7 di atas memerlukan kolom MileStone 1, Milestone 2 ; akan menjadi sebuah kategori untuk atribut *Challenge_type*, *Name*, *Buddy*, *Date of Enrollment*, *Week*, *Weekly Status*, *Score*, *Status*. Kemudian *Phase*, *Batch*, *Location*; diambil dari *prefix* nama di tiap spreadsheet akademik. Sehingga akan dimanipulasi dalam format baru yang lebih normal. Untuk lebih lanjut proses *extract* ditampilkan diagram alur program pada gambar berikut.



Gambar 8. *Extract Data Past Batch*

Program akan mengecek folder *Past batch*. Ini adalah kumpulan direktori atau tempat penyimpanan data *batch* lama yang akan diekstraksi. Setiap folder google drive di dalam *Past batch* akan diiterasi sehingga sistem mencari folder yang memiliki *prefix* tertentu dalam hal ini 'FTDS' memanggil fungsi `get_folder_path` dan `get_indeks`. Langkah ini memastikan bahwa setiap lokasi direktori diperiksa untuk mendapatkan informasi yang diperlukan dengan memanggil fungsi `get_batch_loc_phase`.

			610		70,00%				
No	Name	Buddy	Total		Status	Week 1		Week 2	
			Score	Score (%)		Score	Status	Score	Status
1	Dandi Alfikri	Raka	643	105,20%	Pass	109,38%	Saved	105,20%	Saved
2	Arya Bandoro	Raka	498	81,45%	Pass	85,31%	Saved	81,45%	Saved
3	Aldo Dwicahyo	Afif	543	93,94%	Pass	93,75%	Saved	93,94%	Saved
4	Nisa Aprilia M	Afif	554	95,84%	Pass	93,75%	Saved	95,84%	Saved
5	Brian Budiarto	Danu	575	102,49%	Pass	101,88%	Saved	102,49%	Saved
6	Anggito Zainul	Fahmi	573	98,93%	Pass	90,63%	Saved	98,93%	Saved
7	Michael Andrian	Fahmi	586	105,84%	Pass	107,19%	Saved	105,84%	Saved
8	Dennis Herdiawan	Irfan	562	96,02%	Pass	96,88%	Saved	96,02%	Saved
9	Dinda Tirta Rahayu	Irfan	551	94,34%	Pass	92,81%	Saved	94,34%	Saved
10	Irpan Habibi Sunarya Putra	Danu	580	95,07%	Pass	95,31%	Saved	95,07%	Saved
11	Muhammad Rizki Herfian	Afif	578	99,98%	Pass	93,75%	Saved	99,98%	Saved

Gambar 9. Data *Worksheet Past Batch*

Fungsi `get_batch_loc_phase` dijalankan untuk mendapatkan *attributes* kode *batch*, *location*, dan *phase*. Kemudian fungsi `get_week_status_score_buddy` akan *scraping* isi *cell* baris keempat yang akan menjadi *attributes week* yaitu kolom Week 1 dan Week 2, *weekly_status* pada kolom Status di bawah baris Week, *status* di bawah baris Total seperti yang tertampil pada gambar 9 untuk mendapatkan *value cell* tersebut.

Buddy	Status
Danu	Graduate
Danu	Graduate
Danu	Graduate
Danu	Honors
Danu	Graduate
Danu	Graduate
Danu	Graduate
Irfan	Participant
Irfan	Graduate
Irfan	Teachers Award
Irfan	Participant
Irfan	Fail

Gambar 10. Kolom Status dari *Data Source*

Kolom pada *data source* spreadsheet *akademik* untuk *phase 2* kolom status adalah status kelulusan peserta *bootcamp* dalam satu periode belajar. Hal ini akan berbeda dengan kategori status untuk *phase 1* dan *phase 0* yang memiliki 3 jenis kategori sedangkan *phase 2* memiliki 5 kategori peringkat kelulusan.

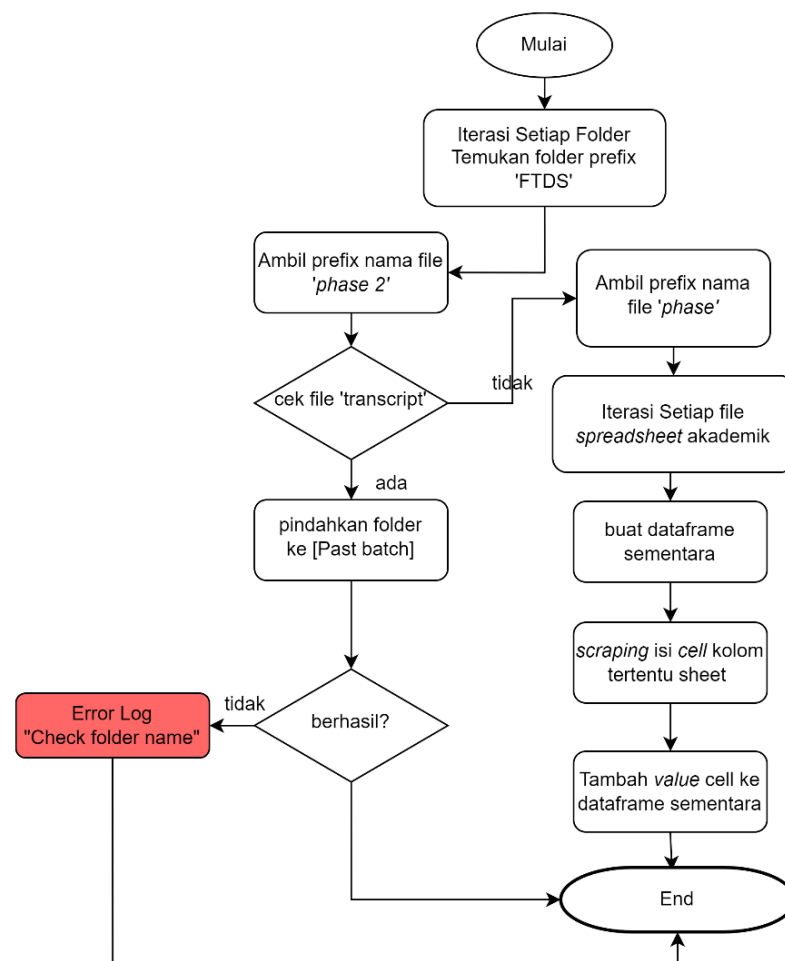
Tabel berikut adalah *attributes* yang akan dihasilkan oleh sentralisasi dan proses ETL pada data *past batch*. Terdapat 11 *attributes* pada tahapan *extract* data *past batch*.

Tabel 5. *Attributes* Data Peserta *Bootcamp Past batch*

No	<i>Attributes</i>	Keterangan
1	Name	Berisikan data nama peserta <i>bootcamp</i>
2	Buddy	Berisikan data nama instruktur <i>bootcamp</i>
3	<i>Phase</i>	Berisikan kategori <i>phase</i> dengan
4	<i>Batch</i>	Berisikan kategori angkatan peserta <i>bootcamp</i>
5	<i>Location</i>	Berisikan data lokasi <i>bootcamp</i>
6	Date of Enrollment	Berisikan tanggal mulai melaksanakan <i>bootcamp</i> pada <i>phase</i> tertentu
7	Week	Berisikan kategori <i>week</i> tiap <i>phase</i>
8	Weekly Status	Berisikan kategori status penilaian <i>challenge</i> peserta <i>bootcamp</i>
9	Challenge Type	Berisikan kategori tugas dan <i>challenge</i> penilaian
10	Score	Berisikan nilai peserta <i>bootcamp</i>
11	Status	Berisikan kategori status peserta <i>bootcamp</i> (berdasarkan akumulasi penilaian tiap <i>phase</i>)

2) Data *Current batch*

Tahap *extract* pada data *current batch* dilakukan dengan mengiterasi seluruh folder yang pada folder google drive pada direktori '*Academic Operations*'. Data yang diperlukan untuk proses ekstraksi dan sentralisasi data ini memerlukan *worksheet* penilaian siswa *bootcamp* yang bernama '*cummulative*' sama seperti data *past batch*.



Gambar 11. *Extract Data Current Batch*

Program yang ada pada *script* python bernama *new_batch.py* untuk *extract* data peserta *current batch*. Pertama, program mengecek folder *root* yaitu 'Academic Operation'. Setiap di dalamnya akan diiterasi sehingga sistem mencari folder yang memiliki *prefix* tertentu dalam hal ini 'FTDS' memanggil fungsi *get_folder_path* dan *get_indeks*. Langkah ini memastikan bahwa setiap lokasi direktori diperiksa untuk mendapatkan informasi yang diperlukan dengan memanggil fungsi *get_batch_loc_phase*. Kemudian pada program juga akan mengecek keterbaruan data jika ada file data terbaru dengan mengatur pengambilan data setiap 1 minggu. Berikut contoh *worksheet* untuk peserta *current batch*.

			Total Score			Min Score		2		
			36			25,2		2		
								1		
										Week 1
No	Name	Buddy	Presentasi P2M3							
			Score	Score (%)	Status	Score (%)	Status	Weekly Ca		
1	Ahmad Dani Rifai	Irfan	28	77,78%	Safe	62,64%	Danger	Slow		
2	Gracia Valerine	Irfan	29	80,56%	Safe	78,74%	Saved	Safe		
3	Muhammad Fiqih Al-ayubi	Irfan	31	86,11%	Safe	87,60%	Saved	Lazy		
4	Ryan Trisnadi	Irfan	33	91,67%	Safe	68,20%	Danger	Slow		
5	Syihabuddin Ahmad Satisma	Irfan	30	83,33%	Safe	83,85%	Saved	Slow		
6	Vicky Belario	Irfan	28	77,78%	Safe	66,40%	Danger	Safe		
7	Vincent Kaunang	Irfan	29	80,56%	Safe	63,61%	Danger	Safe		
8	Michael Parsaoran	Irfan	29	80,56%	Safe	58,78%	Danger	Safe		

Gambar 12. Data *Worksheet Current Batch*

Gambar di atas merupakan salah satu contoh *worksheet* untuk data *current batch*. Tahapan *extract* sama seperti data *past batch* yaitu dengan *scraping* tiap cell yang diperlukan untuk menjadi *attributes* data *current batch*. *Spreadsheet* penilaian untuk *phase 2* pada data *current batch* memiliki kategori yang sama untuk *attribute* status. Tabel berikut adalah *attributes* yang akan dihasilkan oleh sentralisasi dan proses *extract* pada data *current batch*. Terdapat 15 *attributes* pada tahapan *extract* data *current batch*.

Tabel 6. *Attributes* Data Peserta *Bootcamp Current batch*.

No	Attributes	Keterangan
1	Name	Berisikan data nama peserta <i>bootcamp</i>
2	Buddy	Berisikan data nama instruktur <i>bootcamp</i>
3	Phase	Berisikan kategori <i>phase</i>
4	Batch	Berisikan kategori angkatan peserta <i>bootcamp</i>
5	Location	Berisikan data lokasi <i>bootcamp</i>
6	Date of Enrollment	Berisikan tanggal mulai melaksanakan <i>bootcamp</i> pada <i>phase</i> tertentu
7	Week	Berisikan kategori <i>week</i> tiap <i>phase</i>
8	Weekly Status	Berisikan kategori status penilaian <i>challenge</i> peserta <i>bootcamp</i>
9	Weekly Cat	Berisikan kategori performa belajar <i>challenge</i> per <i>week</i> peserta <i>bootcamp</i>
10	Challenge Type	Berisikan kategori tugas dan <i>challenge</i> penilaian
11	Total score	Nilai skor maksimum untuk <i>tugas/challenge</i> tertentu bertipe integer

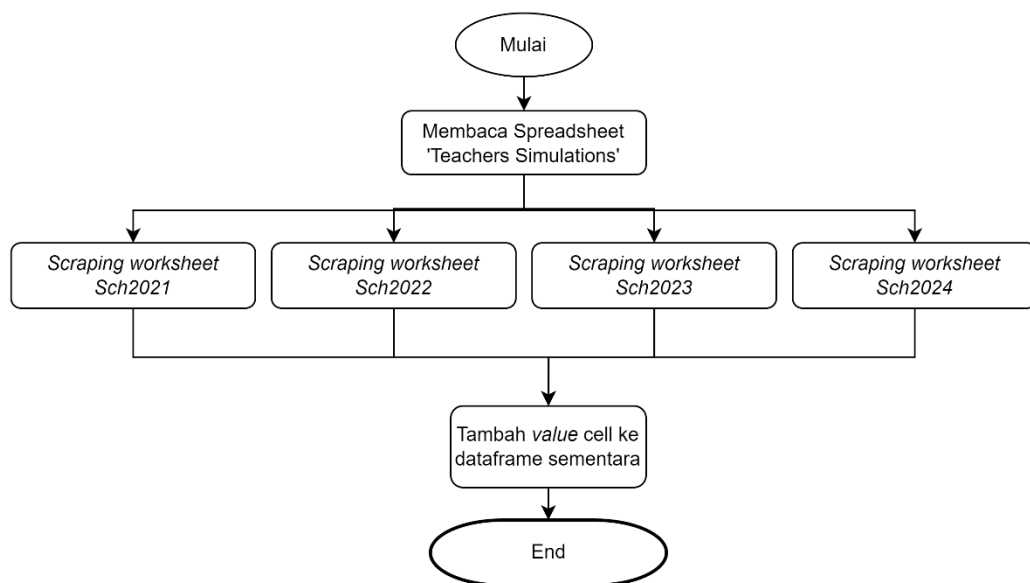
3) Data Instruktur

Proses ekstraksi data aktivitas mengajar instruktur dari spreadsheet "*Teachers Simulations*". Pada tahap ini, sistem membaca spreadsheet berjudul "*Teachers Simulations*" yang berisi data terkait aktivitas mengajar instruktur. Gambar berikut ini adalah contoh *worksheet* yang memuat data jadwal instruktur mengajar.

	A	B	C	D	E	F	G	H	I	J	K
300											
301											
302		Week 1	Monday, May 13, 2024		Tuesday, May 14, 2024		Wednesday, May 15, 2024		Thursday, May 16, 2024		Friday, May 17, 2024
303			Morning	Afternoon	Morning	Afternoon	Morning	Afternoon	Morning	Afternoon	Morning
304	Batch 17 HCK	Phase 0	Hana	Hana	Riki	Riki	Yuda	Yuda	Lis	Lis	Lis
305	Batch 16 HCK	Phase 1	Lis	Lis	Elita	Elita	Riki	Riki	Hana	Hana	Elita
306	Batch 15 HCK	Phase 2	Irfan	Irfan	Irfan	Irfan	Elita	Elita	Rachman	Rachman	Irfan
307	Batch 32 RMT	Phase 0	Yuda	Yuda	Lis	Lis	Hana	Hana	Yuda	Yuda	Rachman
308	Batch 31 RMT	Phase 1	Elita	Elita	Hana	Hana	Rachman	Rachman	Elita	Elita	Hana
309	Batch 30 RMT	Phase 2	Riki	Riki	Danu	Danu	Danu	Danu	Irfan	Irfan	Danu
310	Batch 6 BSD	Phase 0	Vincent	Vincent	Vincent	Vincent	Vincent	Vincent	Ali	Ali	Vincent
311		Phase 1									
312	Batch 5 BSD	Phase 2	Ali	Ali	Ali	Ali	Ali	Ali	Vincent	Vincent	Ali
313	Batch 5 SBY	Phase 0	Dyco	Dyco	Yosef	Yosef	Yosef	Yosef	Yosef	Yosef	Dyco
314		Phase 1									
315	Batch 4 SBY	Phase 2	Yosef	Yosef	Dyco	Dyco	Dyco	Dyco	Dyco	Dyco	Yosef
316		Standby Jira		Riki		Elita		Fahmi		Hana	
317		SJB Offline	HCK	Lis				SJB Online			
318			SBY	Yosef							
319			BSD	Ali							
320											
321		Week 2	Monday, May 20, 2024		Tuesday, May 21, 2024		Wednesday, May 22, 2024		Thursday, May 23, 2024		Friday, May 24, 2024
322			Morning	Afternoon	Morning	Afternoon	Morning	Afternoon	Morning	Afternoon	Morning
323	Batch 17 HCK	Phase 0	Riki	Riki	Yuda	Yuda	Riki	Riki	Libur	Libur	Irfan
324	Batch 16 HCK	Phase 1	Hana	Hana	Riki	Riki	Elita	Elita	Libur	Libur	Elita
325	Batch 15 HCK	Phase 2	Yuda	Yuda	Irfan	Irfan	Yuda	Yuda	Libur	Libur	Rachman

Gambar 13. *Worksheet Teachers Simulations*

Proses *extract* dilakukan dengan cara *scraping* tiap cell pada *worksheet* dengan mengambil cell yang berisi nama instruktur, week, *batch*, *location*, yang berada pada kolom A-B. Program dibuat untuk dapat mengakses nilai di setiap baris *worksheet*, memecah string untuk mendapatkan informasi tambahan, serta mengakses data *batch* dan *phase* yang di-*scraping* secara vertikal. Saat mendapatkan nilai-nilai dari baris yang berbeda. Hal ini dilakukan guna mengotomasi ekstraksi data dibandingkan harus dengan cara manual. Berikut diagram alur *extract* data aktivitas instruktur.



Gambar 14. *Extract* Data Aktivitas Instruktur

Terdapat 4 fungsi untuk melakukan *scraping* yaitu fungsi `sch_2021` untuk *worksheet* jadwal tahun 2021, `sch_2022` untuk *worksheet* jadwal tahun 2022, `sch_2023` untuk *worksheet* jadwal tahun 2023, dan `sch_2024` untuk *worksheet* jadwal tahun 2024. Setelah itu, sistem menjalankan fungsi `'isi_tanggal'` untuk mengisi tanggal sesuai rentang kolom yang diekstrak. Tahap ini kemudian dilanjutkan dengan pengecekan ketersediaan data terbaru menggunakan fungsi `'cek_keterbaruan'`, yang bertujuan untuk memastikan bahwa data apa yang ada. Tabel berikut ini adalah *attribute* yang akan dihasilkan *extract* data aktivitas instruktur mengajar.

Tabel 7. *Attributes* Data Aktivitas Mengajar Instruktur

No	<i>Attributes</i>	Keterangan
1	<i>Instructor Name</i>	Berisikan data nama instruktur <i>bootcamp</i>
2	<i>Location</i>	Berisikan data lokasi mengajar <i>bootcamp</i>
3	<i>Batch</i>	Berisikan kategori angkatan peserta <i>bootcamp</i>
4	<i>Phase</i>	Berisikan kategori <i>phase</i>
5	<i>Date</i>	Berisikan tanggal mulai melaksanakan <i>bootcamp</i> pada <i>phase</i> tertentu
6	<i>Week</i>	Berisikan kategori <i>week</i> tiap <i>phase</i>
7	<i>Day</i>	Berisikan kategori hari mengajar (ex. day1, day2)
8	<i>Section</i>	Berisikan kategori sesi mengajar (AM atau PM)

4) Data Kehadiran *Current batch*

Tahapan *extract* data absensi *current batch* ini dikumpulkan dari berbagai file *spreadsheet* yang disimpan dalam google drive yang sama seperti data penilaian *current batch* pada poin 2 sub bab ini. Folder yang menjadi sumber data adalah “3. Academic Operations” dan di dalamnya terdapat folder yang mengandung kata kunci "FTDS", yang mewakili *batch* tertentu dari program *bootcamp*. *Worksheet* yang diakses ini adalah *worksheet* yang bernama ‘absensi’. Berikut contoh gambar dari *worksheet* ‘absensi’ untuk mendapatkan data kehadiran peserta *bootcamp current batch* pada *sheet* di *data source* akademik.

Data Science Fulltime Program Phase 1 Attendance													
Nama	Buddy	Week 1						Week 2					
		Day 3		Day 4		Day 5		Day 1		Day 2		Day 3	
		AM	PM	AM	PM	AM	PM	AM	PM	AM	PM	AM	PM
Ahmad Dani Rifai	Elita	H	H	H	H	H			H	H	H		
Gracia Valerine	Elita	H	H	H	H	H			H	H	H		
Muhammad Fiqih Al-ayubi	Elita	H	H	A	H	H			H	H	H		
Ryan Trisnadi	Elita			S	S	H			H	H	H		
Syihabuddin Ahmad Satisma	Rachman	H	H	H	H	H			H	H	H		
Vicky Belario	Rachman	H	H	H	H	H			H	H	H		
Vincent Kaunang	Rachman	H	H	H	H	I			H	H	H		
	O	O											

Gambar 15. Contoh *Worksheet* Absensi Peserta *Current batch*

Pada *worksheet* ‘absensi’ akan dilakukan *scraping* menggunakan *script* python `new_absensi.py` untuk dijalankan *extract* absensi peserta *current batch* untuk jenis kehadiran kelas biasa (kelas teknikal). Selanjut akan dilakukan *scraping* untuk jenis kehadiran mentoring. Data kehadiran yang berjenis ini didapatkan dari *worksheet* ‘Mentoring Log’. Gambar berikut ini adalah contoh *worksheet* ‘Mentoring Log’.

Data Science Fulltime Program Phase 1 Mentoring Log						
No	Nama	Buddy	Week 1		Week 2	
			Mentoring 1	Mentoring 2	Mentoring 1	Mentoring 2
1	Ahmad Dani Rifai	Elita				
2	Gracia Valerine	Elita				
3	Muhammad Fiqih Al-ayubi	Elita				
4	Ryan Trisnadi	Elita				
5	Syihabuddin Ahmad Satisma	Rachman				
6	Vicky Belario	Rachman				
7	Vincent Kaunang	Rachman				
8		0	0			
9		0	0			
10		0	0			
11		0	0			

Gambar 16. *Worksheet Mentoring Log*

Fungsi mencari indeks *cell* yang mengandung string 'Week'. Setelah menemukan indeks tersebut, kode mengakses nilai di setiap *row* yang relevan, memecah string untuk mendapatkan informasi tambahan, serta mengakses data *batch* dan *phase* yang diambil secara vertikal. Saat mendapatkan nilai-nilai dari baris yang berbeda, kemudian mengambil informasi dari kolom-kolom yang diperlukan. Setelah mengumpulkan data yang diperlukan, program dapat menggabungkan hasil ekstraksi data. Kolom yang digunakan untuk *extract* data kehadiran *mentoring* adalah *Nama*, *Buddy*, *Week 1-4*, *Mentoring 1-2* setiap minggunya. Data ini akan dibentuk menjadi format *dataframe* seperti yang tersedia pada *library* di Python itu sendiri. Dengan cara ini, proses *scraping* memastikan bahwa data yang diambil terorganisir dengan baik, memudahkan analisis dan penyimpanan dalam format yang diinginkan untuk dilanjutkan ke tahap *transformation*. Tabel berikut ini adalah *attributes* yang akan dihasilkan *extract* data aktivitas instruktur mengajar.

Tabel 8. *Attributes* Data Aktivitas Kehadiran Peserta *Current batch*

No	Attributes	Keterangan
1	<i>Name</i>	Berisikan data nama peserta <i>bootcamp</i>
2	<i>Buddy</i>	Berisikan data nama instruktur <i>bootcamp</i>
3	<i>Phase</i>	Berisikan kategori <i>phase</i>
4	<i>Batch</i>	Berisikan kategori angkatan peserta <i>bootcamp</i>
5	<i>Location</i>	Berisikan data lokasi <i>bootcamp</i>
6	<i>Week</i>	Berisikan kategori <i>week</i> tiap <i>phase</i>

No	Attributes	Keterangan
7	<i>Day</i>	Berisikan kategori <i>week</i> tiap <i>phase</i>
8	<i>sectionI</i>	Berisikan kategori kehadiran <i>section 1</i> (H/S/I/A)

5) Data All Batch

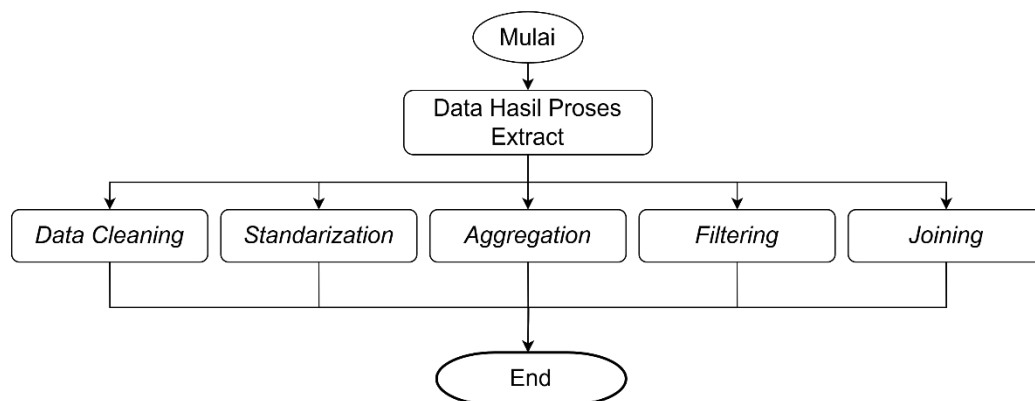
Pada tahapan *extract* data *All Batch*, semua data berasal dari data *past batch* dan *current batch*. Proses ini dilakukan dengan melakukan *merge* pada atribut-atribut yang sama dari kedua jenis data tersebut, sehingga hanya data dengan fitur yang sama pada kedua *batch* yang digabungkan ke dalam *dataframe* sementara. Proses penggabungan ini memastikan konsistensi dan kesesuaian antar *batch* dalam data yang akan digunakan untuk tahap selanjutnya, yaitu *transformasi* dan pemuatan ke dalam sistem terpusat.

Tabel 9. *Attributes* Data Peserta *Bootcamp All batch*

No	Attributes	Keterangan
1	<i>Name</i>	Berisikan data nama peserta <i>bootcamp</i>
2	<i>Buddy</i>	Berisikan data nama instruktur <i>bootcamp</i>
3	<i>Phase</i>	Berisikan kategori <i>phase</i> dengan
4	<i>Batch</i>	Berisikan kategori angkatan peserta <i>bootcamp</i>
5	<i>Location</i>	Berisikan data lokasi <i>bootcamp</i>
6	<i>Date of Enrollment</i>	Berisikan tanggal mulai melaksanakan <i>bootcamp</i> pada <i>phase</i> tertentu
7	<i>Week</i>	Berisikan kategori <i>week</i> tiap <i>phase</i>
8	<i>Weekly Status</i>	Berisikan kategori status penilaian <i>challenge</i> peserta <i>bootcamp</i>

B) Transform

Setelah data *extraction* dan disimpan, langkah berikutnya adalah melakukan sesuatu dengan data tersebut. Tahapan berikutnya dalam *lifecycle* rekayasa data adalah *transformasi*, yang berarti data perlu diubah dari bentuk aslinya menjadi sesuatu yang berguna untuk keperluan di tahap berikutnya. Tanpa *transformasi* yang tepat, data akan tetap pasif dan tidak berguna untuk *reporting*, analisis, atau *machine learning* (ML).

Gambar 17. Proses *Transform*

Transformasi yang dilakukan pada penelitian ini, membersihkan dari nilai *null* (*data cleaning*), dengan mengubah data ke tipe yang benar (misalnya, mengubah data *string* yang diambil menjadi tipe numerik dan tanggal). *Attributes* tanggal akan dikonversi menjadi format yyyy/mm/dd ini merupakan bagian tahapan *standardization*, serta menempatkan catatan ke dalam format standar, dan mengurutkan data kehadiran pada tanggal tertentu (*filtering*). Kemudian dilakukan penggabungan beberapa kolom seperti di tabel *current batch* untuk menghitung total persen dari skor penilaian. Terutama dalam kasus atribut '*name*' harus yang berformat huruf pertama dari setiap kata menjadi huruf kapital. Pada tahap *transformasi* berikutnya, skema data dapat diubah dan normalisasi diterapkan.

Tabel 10. Tahap *Transformasi* Data

No	Tahap	Keterangan
1	<i>Data cleaning</i>	Menghilangkan nilai <i>null</i> , duplikasi
2	<i>Standardization</i>	Menyesuaikan tipe data masing masing atribut Standarisasi atribut tanggal dalam format YYYY/MM/DD
3	<i>Joining</i>	Menggabungkan data dari berbagai sumber yang kohesif berdasarkan atribut yang sama.
4	<i>Aggregation</i>	Membandingkan performa peserta di berbagai lokasi. Tingkat kelulusan atau keberhasilan peserta

No	Tahap	Keterangan
		berdasarkan mentor yang ditugaskan dan perhitungan persentase.
4	<i>Filtering</i>	Memfilter data siswa yang berkategori 'safe' dan 'danger'

C) Load

Data yang telah diekstraksi dan ditransformasi kemudian dimuat ke tujuan akhir (*user interface*) untuk pembuatan grafik yang lebih mudah dipahami. Pada tahapan *load* ini, semua data yang akan diproses untuk kebutuhan yang lebih jauh seperti *data mart*, penyimpanan data operasional, *data warehouse* perusahaan akan disimpan ke menggunakan layanan *cloud* yaitu *bigquery*. Penelitian ini menggunakan solusi *Cloud Business Intelligence Google*. Solusi ini menggunakan produk-produk yang tersedia di *Google Cloud Platform* (GCP) untuk menawarkan arsitektur. Produk utama dalam arsitektur ini adalah *Google BigQuery*, yang berfungsi sebagai pengelola *Data warehouse* dan *Big data*. *Applications front-end* untuk *business intelligence* dan *decision support* memungkinkan *knowledge workers* melakukan analisis dan memvisualisasikan hasilnya.

Berdasarkan tahapan dan deskripsi di atas, proses mengekstraksi, mengubah, dan memuat data ke sumber *host* baru, seperti *data warehouse*, disebut ETL. Ini adalah prosedur yang diperlukan untuk mengoptimalkan implementasi data analitik.

3.4.3.3. Penjadwalan

Penjadwalan atau *scheduling* dilakukan untuk mengotomatisasi alur kerja ETL (*Extract, Transform, Load*) secara periodik, sehingga data dapat dikelola secara efisien dan konsisten. *BigQuery* dipilih sebagai *data warehouse* untuk menyimpan dan mengelola data yang telah melalui proses ETL (*Extract, Transform, Load*) karena kemampuannya dalam menangani volume data yang besar dan mendukung query SQL dengan performa tinggi. Proses ETL secara

otomatis sesuai jadwal yang ditentukan menggunakan Google Cloud Scheduler. Tools ini memungkinkan penjadwalan tugas (*jobs*) secara teratur, seperti menjalankan script Python untuk mengekstraksi, mentransformasi, dan memuat data ke BigQuery.

Selain itu, Google Workflows digunakan untuk merancang alur proses yang terintegrasi dengan baik. Workflows mengelola langkah-langkah otomatisasi, mulai dari autentikasi API, pengolahan data, hingga penyimpanan data ke dalam tabel di BigQuery. Kombinasi antara Google Cloud Scheduler dan Workflows memastikan proses berjalan sesuai jadwal dengan alur yang jelas, terstruktur, dan dapat diandalkan. Automasi ini mengurangi kebutuhan intervensi manual sekaligus memastikan konsistensi dan akurasi dalam pembaruan data.

3.4.4. Evaluasi

Evaluasi merupakan tahapan penting untuk memastikan kualitas dan efektivitas data dalam proses ETL yang telah dilakukan. Tujuan utama dari pengujian ETL adalah untuk memverifikasi dan menghilangkan cacat data serta kesalahan umum yang terjadi sebelumnya sebelum diproses untuk analisis dan pelaporan (Yulianto, 2019). Beberapa pertimbangan dilakukan evaluasi data hasil ETL sebagai berikut (Abd Al-Rahman dkk., 2023).

Evaluasi untuk data yang dihasilkan dari implementasi ETL dilakukan dengan validasi menggunakan *tools data quality inherent* yaitu *Great Expectation* untuk menilai kualitas dari data itu sendiri. *Great Expectation* dalam penelitian ini digunakan membantu mengevaluasi kualitas data dari proses ETL serta memvalidasi format *attribute* tanggal, memeriksa tipe data, memastikan konsistensi data, misalnya di bagian *attribute* nama tidak terdapat karakter unik selain huruf (Mandrzzato, 2022). Kemudian evaluasi kedua berupa kuesioner yang disebarakan kepada para *stakeholder* yang terlibat langsung dalam proses *monitoring* dan *reporting*, seperti tim akademik, instruktur dengan mengadopsi standar ISO/IEC 25012:2008 untuk mengukur efisiensi proses (Syafrianto, 2024). Pengukuran ini menggunakan skala likert. Berikut metrik evaluasi dengan *tools* untuk menilai *data quality inherent* dengan ISO/IEC 25012.

Tabel 11. Beberapa Metrik *Data Quality Dimensions ISO/IEC 25012 Standard*

Metrik	Definisi	Rules	Threshold
Akurasi	Tingkat di mana data memiliki atribut yang benar-benar dengan persentase data tanpa kesalahan	a) Atribut (student_name) pada tabel dimensi dim_student dan (instructor_name) pada tabel harus bebas dari tanda baca.	97%
		b) Atribut Weekly_Status pada tabel dimensi dim_kelulusan hanya berisi status kelulusan.	97%
		c) • Atribut location hanya berisi SBY, HCK, RMT, BSD.	100%
		d) Atribut batch tabel dimensi dim_Batch memiliki 3 digit angka	97%
Konsistensi	Tingkat konsistensi yang dimiliki atribut dengan format yang disepakati	e) Atribut tanggal pada tabel dim_Date memiliki format YYYY/MM/DD	100%
		f) Jumlah row sama dengan referensi tabel sesudah dan sebelum <i>transformation</i>	97%
Kelengkapan	Tingkat nilai <i>null</i> atau kosong	g) Atribut bebas dari nilai null	98%
Terkini	Jumlah data	h) Atribut tanggal date min 28/06/2021 dan max (tanggal terkini)	TRUE

Tujuannya evaluasi kedua adalah mendapatkan *feedback* tentang pengalaman dan tantangan end user dalam menggunakan data dari *data marts* dari implementasi ETL. Hasil dari kuesioner ini akan dianalisis untuk mengidentifikasi masalah kualitas data dan efisiensi proses yang paling signifikan dan untuk rekomendasi *development* yang relevan dalam implementasi ETL. Pertanyaan penelitian yang dirumuskan adalah bagaimana perspektif pengguna data berkorelasi dengan standar kualitas data berdasarkan ISO/IEC 25012:2008 (Syafrianto, 2024).

Metode wawancara untuk menilai kualitas data Instrumen wawancara diklasifikasikan berdasarkan standar ISO/IEC 25012:2008. Berikut adalah isi dari

kuesioner yang dirancang penulis untuk mendapatkan evaluasi dan validitas dari *stakeholder* dan instruktur sebagai *end user* hasil akhir.

Tabel 12. Daftar Kuesioner Evaluasi Kualitas Data

No	Kategori	Pertanyaan	Pilihan Jawaban
1	Kualitas Data	Apakah data yang dihasilkan sudah lengkap dan tidak ada data yang hilang? (P1)	☐ Sangat Setuju ☐ Setuju ☐ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju
		Apakah data yang dihasilkan oleh proses ETL sudah akurat dan bebas dari kesalahan? (P2)	☐ Sangat Setuju ☐ Setuju ☐ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju
		Apakah data yang dihasilkan sudah konsisten dengan sumber data aslinya? (P3)	☐ Sangat Setuju ☐ Setuju ☐ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju
		Apakah data yang dihasilkan sudah terstruktur dengan baik dan mudah dipahami? (P4)	☐ Sangat Setuju ☐ Setuju ☐ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju
		Apakah yang dihasilkan sudah sesuai dengan kebutuhan <i>reporting</i> , <i>monitoring</i> dan analisis <i>progres</i> belajar peserta <i>bootcamp</i> ? (P5)	☐ Sangat Setuju ☐ Setuju ☐ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju

No	Kategori	Pertanyaan	Pilihan Jawaban
2	Efisiensi Proses	Apakah proses <i>monitoring</i> data menjadi lebih mudah dan efisien? (P6)	☐ Sangat Setuju ☐ Setuju ☐ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju
		Apakah proses ETL dapat dengan mudah diadaptasi untuk mengakomodasi perubahan data? (P7)	☐ Sangat Setuju ☐ Setuju ☐ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju
		Apakah waktu yang dibutuhkan untuk menghasilkan laporan menjadi lebih singkat? (P8)	☐ Sangat Setuju ☐ Setuju ☐ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju
		Apakah informasi yang diperoleh dari data menjadi lebih relevan untuk mendukung pengambilan keputusan? (P9)	☐ Sangat Setuju ☐ Setuju ☐ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju
		Apakah dokumentasi proses ETL sudah lengkap dan mudah dipahami? (P10)	☐ Sangat Setuju ☐ Setuju ☐ Cukup Setuju ☐ Kurang Setuju ☐ Sangat Setuju ☐ Tidak Setuju

Dari desain angket pertanyaan di atas, pengolahan data nilai skor jawaban responden dalam hal ini *end user* dari *pipeline* ETL dibedakan menjadi tiga kategori yaitu setuju, netral, tidak setuju. Setuju berarti responden benar mengaku untuk

pertanyaan yang diberikan. Netral berarti responden tidak merasa ada perbedaan. Kemudian tidak setuju berarti responden tidak setuju atas pertanyaan yang diajukan. Perhitungan hasil kuesioner yang telah diberikan kepada responden, data dianalisis menggunakan model skala Likert. Lihat tabel 13 berikut ini Click or tap here to enter text..

Tabel 13. Contoh Pengisian Kuesioner Variabel Kualitas Data

No	Kategori	Pertanyaan	Pilihan Jawaban
1	Kualitas Data	Apakah data yang dihasilkan sudah lengkap dan tidak ada data yang hilang? (P1)	☐ Sangat Setuju ☐ Setuju ☒ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju
		Apakah data yang dihasilkan oleh proses ETL sudah akurat dan bebas dari kesalahan? (P2)	☐ Sangat Setuju ☐ Setuju ☐ Cukup Setuju ☒ Kurang Setuju ☐ Tidak Setuju
		Apakah data yang dihasilkan sudah konsisten dengan sumber data aslinya? (P3)	☐ Sangat Setuju ☒ Setuju ☐ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju
		Apakah data yang dihasilkan sudah terstruktur dengan baik dan mudah dipahami? (P4)	☐ Sangat Setuju ☐ Setuju ☒ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju
		Apakah yang dihasilkan sudah sesuai dengan kebutuhan <i>reporting</i> , <i>monitoring</i> dan analisis	☐ Sangat Setuju ☐ Setuju ☐ Cukup Setuju

		<i>progres belajar peserta bootcamp?</i> (P5)	☒ Kurang Setuju ☐ Tidak Setuju
--	--	---	--

No	Kategori	Pertanyaan	Pilihan Jawaban
2	Kualitas Data	Apakah proses <i>monitoring</i> data menjadi lebih mudah dan efisien? (P6)	☐ Sangat Setuju ☒ Setuju ☐ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju
		Apakah proses ETL dapat dengan mudah diadaptasi untuk mengakomodasi perubahan data? (P7)	☐ Sangat Setuju ☒ Setuju ☐ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju
		Apakah waktu yang dibutuhkan untuk menghasilkan laporan menjadi lebih singkat? (P8)	☐ Sangat Setuju ☒ Setuju ☐ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju
		Apakah informasi yang diperoleh dari data menjadi lebih relevan untuk mendukung pengambilan keputusan? (P9)	☐ Sangat Setuju ☐ Setuju ☒ Cukup Setuju ☐ Kurang Setuju ☐ Tidak Setuju
		Apakah dokumentasi proses ETL sudah lengkap dan mudah dipahami? (P10)	☐ Sangat Setuju ☐ Setuju ☐ Cukup Setuju ☒ Kurang Setuju ☐ Tidak Setuju

Misalkan seorang responden mengisi kuesioner seperti pada Tabel 13.

Maka, perlu dicari terlebih dahulu nilai rata-ratanya karena pada variabel penilaian pengukuran masing-masing terdapat lima pertanyaan yaitu, kualitas data (P1-P5) dan efisiensi (P6-P10). Cara untuk mencari nilai rata-rata pada penelitian ini adalah sebagai berikut (Aditya Santika dkk., 2023):

Kualitas data = $(3 + 2 + 4 + 3 + 2) / 5 = 2,80$. Nilai tersebut diperoleh berdasarkan jawaban responden pada kuesioner. Maka dari itu responden tersebut memperoleh nilai 2,80 dan nilai tersebut nantinya perlu dibulatkan menjadi 3, begitupun pada variabel metrik lain. Berdasarkan perhitungan skala likert dilakukan dengan langkah-langkah sebagai berikut:

1. Menghitung total jawaban kuesioner
2. Menghitung total minimal

Pada penelitian ini skor minimal yaitu 1 dan total pertanyaan pada kuesioner sebanyak 10. Perhitungan skor maksimal dilakukan dengan cara:

$$Total \text{ min} = skor \text{ minimal skala likert} * total \text{ pertanyaan}$$

$$Total \text{ min} = 1 * 10 = 10$$

3. Menghitung total maksimal

Pada penelitian ini skor minimal yaitu 5. Perhitungan skor maksimal dilakukan dengan cara:

$$Total \text{ maks} = skor \text{ maksimal skala likert} * total \text{ pertanyaan}$$

$$Total \text{ maks} = 5 * 10 = 50$$

4. Menentukan rentang (range)

Untuk menentukan rentang dilakukan perhitungan:

$$Range = Total \text{ maks} - Total \text{ min}$$

$$Range = 50 - 10 = 40$$

5. Menghitung nilai panjang kelas interval (p)

Untuk menghitung nilai interval dilakukan perhitungan:

$$p = \frac{40}{5}$$

$$p = 8$$

6. Menentukan kategori berdasarkan jawaban

Pada penelitian ini terdapat 5 kelas skala dengan kategori sangat setuju, setuju, cukup setuju, kurang setuju, dan tidak setuju. Maka dari itu dengan panjang kelas interval (p) = 8 dan dengan dimulai batas bawah 10. Proses hitung rentang p

dikurangi dengan 1. Hal ini memastikan bahwa ada "batas atas" dari suatu kategori tidak tumpang tindih dengan batas bawah kategori berikutnya, maka distribusi kelas dapat dilihat pada Tabel 12.

Tabel 14. Distribusi Kelas

Perhitungan Batas Atas	p	Kategori
$10 + 8 - 1 = 17$	$10 - 17$	Sangat Tidak Setuju
$18 + 8 - 1 = 25$	$18 - 25$	Kurang Setuju
$26 + 8 - 1 = 33$	$26 - 33$	Cukup Setuju
$34 + 8 - 1 = 41$	$34 - 41$	Setuju
$42 + 8 - 1 = 50$	$42 - 50$	Sangat Setuju

Hasil kategori berdasarkan contoh pengisian kuesioner variabel kualitas data, total skor penilaian responden x jika dijumlahkan menghasilkan 31, yang muncul dari skor kumulatif pada kategori kualitas data (14) dan efisiensi proses (17). Total skor ini berada pada interval (p) kategori 'cukup setuju'. Hal ini menunjukkan bagaimana responden menilai keakuratan dan efisiensi sistem yang digunakan keandalan datanya dan seberapa baik kerjanya. Selanjutnya, setiap kategori yang diperoleh setiap responden akan dilihat dari persentase tiap kategori untuk semua responden. Evaluasi selanjutnya diperlukan untuk menilai setiap indikator pengukuran dengan menghitung rata-ratanya sehingga akan diperoleh kesimpulan kualitas data dan efisiensi dengan adanya implementasi ETL pada sentralisasi data dengan skema *data warehouse* yang dirancang (Ghita & Trisminingsih, 2021).

Metodologi ini sangat penting untuk menggambarkan pemahaman yang lebih luas tentang kualitas informasi dan produktivitas *workflow* setelah implementasi *Extract, Transform, Load* (ETL) untuk penggabungan data melalui *blueprint data warehouse* yang sudah ada. Dengan demikian, penilaian hasil akan membantu dalam menentukan keberhasilan sistem dalam memenuhi kebutuhan pengguna dan menentukan bagian-bagian yang perlu ditingkatkan.

V. SIMPULAN DAN SARAN

5.1. Simpulan

Penelitian ini menghasilkan sistem *pipeline* ETL pada sentralisasi data dengan skema *data warehouse*. Skema ini menghasilkan *data warehouse* yang bersumber dari database akademik *bootcamp* dengan tools bantuan untuk implementasi ETL yaitu Google Cloud Platform. Penggunaan *scheduler* membantu melakukan *Extract, Transform, Load* dalam melakukan otomatisasi. Data yang dihasilkan dapat digunakan untuk kebutuhan efisiensi dalam *monitoring*, pelaporan, dan pengambilan keputusan akibat data akademik yang terfragmentasi di berbagai sumber.

Hasil evaluasi data proses ETL menunjukkan bahwa validasi kualitas data sesuai dengan standar ISO/IEC 25012 memenuhi ambang batas. Pada metrik akurasi, atribut seperti "student_name", "weekly_status", "lokasi", dan "batch" memenuhi aturan yang ditentukan dengan akurasi rata-rata di atas 98%; atribut tanggal memenuhi format yang diharapkan dengan akurasi rata-rata di atas 100%. Menurut evaluasi kelengkapan dan konsistensi data, hampir semua tabel memiliki tingkat kelengkapan mendekati atau sama dengan 100%, kecuali tabel aktivitas. Secara keseluruhan, kriteria yang ditetapkan menunjukkan bahwa data tersebut valid, konsisten, lengkap, dan terkini.

Selain itu, evaluasi yang dilakukan melalui kuesioner kepada pengguna mayoritas pengguna memberikan skor 4 (setuju) berdasarkan hasil rata-rata skala Likert, yang menunjukkan bahwa data yang dihasilkan dari proses ETL dianggap memiliki kualitas yang baik, dengan aspek konsistensi dan struktur mendapatkan nilai tertinggi. Namun, terdapat beberapa catatan perbaikan pada kesesuaian data untuk dokumentasi proses ETL yang dinilai hanya sekadar cukup oleh beberapa pengguna. Sebaliknya, proses ETL secara umum dianggap efisien. Hasil evaluasi ini dapat digunakan sebagai dasar untuk upaya lebih lanjut untuk memperbaiki

instruksi dan *pipeline* ETL.

5.2. Saran

Hal yang perlu diperhatikan adalah pentingnya bagi peneliti untuk meningkatkan pemahaman *end user* terhadap sistem yang dihasilkan, terutama melalui pelatihan bagi pengguna terkait penggunaan sistem serta peningkatan kualitas dokumentasi proses ETL serta memastikan mekanisme kesalahan skenario ETL diatasi dengan baik seperti pesan kesalahan, dan lain-lain. Langkah ini bertujuan untuk memastikan bahwa semua *end user* dapat memanfaatkan sistem secara efektif dan efisien. Selanjutnya, pada penelitian selanjutnya dapat mengembangkan sistem visualisasi interaktif untuk mendukung proses dalam lingkup analitik yang *scalable*.

DAFTAR PUSTAKA

DAFTAR PUSTAKA

- Abd Al-Rahman, S. Q., Hasan, E. H., & Sagheer, A. M. (2023). Design and implementation of the web (extract, transform, load) process in data warehouse application. *IAES International Journal of Artificial Intelligence*, 12(2), 765–775. <https://doi.org/10.11591/ijai.v12.i2.pp765-775>
- Aditya Santika, A., Hamonangan Saragih, T., Kartini, D., & Ramadhani, R. (2023). Penerapan Skala Likert Pada Klasifikasi Tingkat Kepuasan Pelanggan Agen BRILink Menggunakan Random Forest. *Jurnal Sistem dan Teknologi Informasi*, 11(3). <https://doi.org/10.26418/justin.v11i3>
- Alsibhawi, I. A. A., Yahaya, J. B., & Mohamed, H. B. (2023). Business Intelligence Adoption for Small and Medium Enterprises: Conceptual Framework. Dalam *Applied Sciences (Switzerland)* (Vol. 13, Nomor 7). MDPI. <https://doi.org/10.3390/app13074121>
- Andriansyah, D. (2022). Implementasi Extract-Transform-Load (ETL) Data Warehouse Laporan Harian Pool. *Jurnal Teknik Informatika*, 8(2), 45–49. <https://doi.org/10.51998/jti.v8i2.486>
- Berliantara, A. Y., Wicaksono, S. A., & Pinandito, A. (2017). *Optimasi Scheduling untuk Proses Extract, Transform, Load (ETL) pada Data Warehouse Menggunakan Metode Round Robin Data Partitioning (Studi Kasus: Universitas XYZ)* (Vol. 1, Nomor 11). <http://j-ptiik.ub.ac.id>
- Broman, K. W., & Woo, K. H. (2018). Data Organization in Spreadsheets. *The American Statistician*, 72(1), 2–10. <https://doi.org/10.1080/00031305.2017.1375989>
- Dharayani, R., Ayu Laksitowening, K., & Putri Yanuarfiani, A. (2015). Implementasi ETL (Extract, Transform, Load) Pangkalan Data Perguruan Tinggi dengan Menggunakan State-Space Problem. 2(Vol. 2 No. 1 (2015): April, 2015), 1159.

- El Aissi, M. E. M., Benjelloun, S., Loukili, Y., Lakhrissi, Y., Boushaki, A. El, Chougrad, H., & Elhaj Ben Ali, S. (2022). *Data Lake Versus Data Warehouse Architecture: A Comparative Study* (hlm. 201–210). https://doi.org/10.1007/978-981-33-6893-4_19
- Filiana, A., Prabawati, A. G., Rini, M. N. A., Virginia, G., & Susanto, B. (2020). Perancangan Data Warehouse Perguruan Tinggi untuk Kinerja Penelitian dan Pengabdian kepada Masyarakat. *Jurnal Teknik Informatika dan Sistem Informasi*, 6(2). <https://doi.org/10.28932/jutisi.v6i2.2557>
- Garani, G., Chernov, A., Savvas, I., & Butakova, M. (2019). A Data Warehouse Approach for Business Intelligence. *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, 70–75. <https://doi.org/10.1109/WETICE.2019.00022>
- Ghita, F., & Trisminingsih, R. (2021). Pengujian Data Warehouse SOLAP untuk Komoditas Pertanian Indonesia. *Jurnal Ilmu Komputer dan Agri-Informatika*, 8(1), 42–56. <https://doi.org/10.29244/jika.8.1.42-56>
- Google Teams. (2024, November 12). *Cloud Run functions overview*. <https://cloud.google.com/functions/docs/concepts/overview>
- Gualo, F., Moisés, R., Javier, V., Ismael, C., & Mario, P. (2021). *Data Quality Certification using ISO/IEC 25012: Industrial Experiences*. <https://www.aqclab.es>
- Homayouni H. (2018). *An approach for testing the Extract-Transform-Load process in data warehouse systems*. Colorado State University.
- Iqbal, H., & Babar, M. (2016). An Approach for Analyzing ISO / IEC 25010 Product Quality Requirements based on Fuzzy Logic and Likert Scale for Decision Support Systems. Dalam *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 7, Nomor 12). www.ijacsa.thesai.org
- Jamaludin, Khairunnisa Samosir, Wahyuddin S, Elmi Devia, Leo Willyanto Santoso, Yuniansyah, Junaidi, Sri Rezeki Candra Nursari, Noor Azizah, & Muhamad Hadi Saputra. (2022). *Sistem Basis Data* (M. S. Mila Sari, Ed.; 1 ed., Vol. 2). PT Global Eksekutif Teknologi.

- Kabakchieva, D. (2015). Business Intelligence Systems for Analyzing University Students Data. *Cybernetics and Information Technologies*, 15(1), 104–115. <https://doi.org/10.1515/cait-2015-0009>
- Khan, B., Jan, S., Khan, W., & Chughtai, M. I. (2024). An Overview of ETL Techniques, Tools, Processes and Evaluations in Data Warehousing. *Journal on Big Data*, 6(1), 1–20. <https://doi.org/10.32604/jbd.2023.046223>
- Khine, P. P., & Wang, Z. S. (2018). Data lake: a new ideology in big data era. *ITM Web of Conferences*, 17, 03025. <https://doi.org/10.1051/itmconf/20181703025>
- Kholil, M., & Mu'min, S. (2018). Pengembangan Private Cloud Storage sebagai Sentralisasi Data Universitas Nahdlatul Ulama Sidoarjo Berbasis Open Source Owncloud. *Jurnal Ilmu Komputer dan Desain Komunikasi Visual*, 3(Vol. 3 No. 1 (2018): Jurnal Ilmu Komputer dan Desain Komunikasi Visual (JIKDISKOMVIS)), 34–42. <https://journal.unusida.ac.id/index.php/jik/article/view/42/31>
- Kretz, A. (2019). The Data Engineering Cookbook Mastering The Plumbing Of Data Science. Dalam Andreas Kretz (Ed.), *Mastering The Plumbing Of Data Science* (1 ed., Vol. 1, Nomor 1). CookBook.
- Lavalle, S., Lesser, E., Shockley, R., Hopkins, M., & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review*, 52, 21–32.
- Mandruzzato, L. (2022). *Ensuring High Data Quality Standards: A Framework for Single and Cross-Enterprise Platforms* [Politecnico di Milano]. <https://hdl.handle.net/10589/210452>
- Miller, J. D. . (2017). *Big data visualization : learn effective tools and techniques to separate big data into manageable and logical components for efficient data visualization* (Laxmi Subramanian & Tushar Gupta, Ed.; 1 ed., Vol. 1). Packt Publishing Ltd. <https://unidel.edu.ng/focelibrary/books/big-data-visualization.9781785281945.78390.pdf>
- Moscoso-Zea, O., Paredes-Gualtor, J., & Luján-Mora, S. (2018). A Holistic

- View of Data Warehousing in Education. *IEEE Access*, 6, 64659–64673.
<https://doi.org/10.1109/ACCESS.2018.2876753>
- Muryjas, P., Wawer, M., & Rzemieniak, M. (2021). Managing the Process of Evaluation of the Academic Teachers with the Use of Data Mart and Business Intelligence. *European Research Studies Journal*, XXIV(Special Issue 2), 127–140. <https://doi.org/10.35808/ersj/2196>
- Nambiar, A., & Mundra, D. (2022). An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management. *Big Data and Cognitive Computing*, 6(4), 132. <https://doi.org/10.3390/bdcc6040132>
- Rehman, K. U., Ahmad, U., & Mahmood, S. (2018). A Comparative Analysis of Traditional and Cloud Data Warehouse. *VAWKUM Transactions on Computer Sciences*, 15, 34. <https://doi.org/10.21015/vtcs.v15i1.487>
- Reis, J., & Housley, M. (2022). *Fundamentals of Data Engineering Plan and Build Robust Data Systems* (Jessica Haberman & Michele Cronin, Ed.; 1 ed., Vol. 1). O'Reilly Media, Inc. <http://oreilly.com>
- Soriano, L. T., & Dela Cruz, J. S. (2020). Implementing an ETL-driven Data Integration Framework for State Universities and Colleges. *IOP Conference Series: Materials Science and Engineering*, 803(1). <https://doi.org/10.1088/1757-899X/803/1/012041>
- Syafrianto, A. (2024). Data Quality Analysis on Open Government Data Portals: A Qualitative Study Using ISO/IEC 25012:2008 Standards. *Journal of Information Systems and Informatics*, 6(2). <https://doi.org/10.51519/journalisi.v6i2.862>
- Vesjolijs, A. (2024). The E(G)TL Model: A Novel Approach for Efficient Data Handling and Extraction in Multivariate Systems. *Applied System Innovation*, 7(5). <https://doi.org/10.3390/asi7050092>
- Wijaya, G., Nusa, S., & Jakarta, M. (2017). Perancangan Data Warehouse Nilai Mahasiswa Dengan Kimball Nine-Step Methodology. *Jurnal Informatika*, 4(1).
- Yulianto, A. A. (2019). Extract Transform Load (ETL) Process in Distributed Database Academic Data Warehouse. *APTIKOM Journal on Computer Science and Information Technologies*, 4(2), 61–68.

<https://doi.org/10.11591/APTIKOM.J.CSIT.36>