

LAPORAN AKHIR

PROYEK DATA WAREHOUSE

PORTAL SATU DATA ITERA



Disusun Oleh:

Razin Hafid Hamdi	123450096
Daris Samudra	123450102
Ahmad Rizky	123450050
Kholisaturrohmah	122450019

KELOMPOK 10

Program Studi Sains Data
Institut Teknologi Sumatera

I. Ringkasan Eksekutif

1.1. Gambaran Proyek

Proyek ini mengembangkan Data Warehouse untuk Portal Satu Data ITERA, sebuah platform data terpadu yang menyediakan akses ke dataset dan informasi institusi. Sistem ini dirancang untuk mendukung analisis penggunaan portal, pemantauan kualitas data, dan penyediaan wawasan bagi pengelola portal serta manajemen ITERA.

1.2. Pencapaian Utama

- Perancangan model dimensional dengan 4 tabel fakta dan 6 tabel dimensi
- Implementasi proses ETL dengan mekanisme SCD Type 1 dan Type 2
- Pengembangan dashboard analitik menggunakan Power BI
- Implementasi keamanan berbasis peran (RBAC)
- Penyelesaian User Acceptance Testing dengan tingkat keberhasilan 100%

1.3. Dampak Bisnis

Data warehouse ini memungkinkan:

- Pemantauan akses dataset secara real-time
- Analisis kualitas data berdasarkan metrik terstandarisasi
- Identifikasi dataset populer dan tren pencarian
- Pelaporan metrik institusi yang akurat dan tepat waktu

1.4. Rekomendasi

1. Implementasi integrasi langsung dengan sistem sumber (SIAKAD, HRIS)
2. Pengembangan dashboard mobile untuk akses manajemen
3. Penambahan fitur prediktif menggunakan machine learning

II. Pendahuluan

2.1. Latar Belakang

Institut Teknologi Sumatera (ITERA) memiliki Platform Satu Data ITERA yang menyediakan dataset dan wawasan tentang institusi. Dengan jumlah mahasiswa aktif mencapai 23.842 orang, 778 dosen, 42 program studi, dan 3 fakultas, ITERA memerlukan data warehouse untuk menganalisis penggunaan portal, kualitas dataset, dan metrik institusi.

Proyek data warehouse ini bertujuan untuk merancang dan mengimplementasikan sistem pergudangan data yang dapat melakukan pelacakan akses dataset, pemantauan kualitas data, analisis perilaku pengguna portal, dan menyediakan wawasan untuk pengelola portal dan manajemen ITERA.

2.2. Tujuan

Tujuan dari proyek ini adalah:

1. Memahami proses bisnis Portal Satu Data ITERA dan kebutuhan analitik pengelola portal
2. Mengidentifikasi sumber data portal dan metrik penggunaan yang relevan
3. Merancang model dimensional untuk pelacakan akses dataset, kualitas data, dan perilaku pengguna
4. Mengimplementasikan data warehouse dengan proses ETL yang andal
5. Mengembangkan dashboard analitik untuk visualisasi data
6. Mendokumentasikan desain dan implementasi secara komprehensif

2.3. Ruang Lingkup

Ruang lingkup proyek ini mencakup:

- Analisis kebutuhan bisnis untuk Portal Satu Data ITERA
- Perancangan Entity Relationship Diagram (ERD) untuk entitas portal
- Perancangan model dimensional (Star Schema)
- Implementasi basis data menggunakan SQL Server
- Pengembangan proses ETL
- Pembuatan dashboard analitik dengan Power BI
- Pengujian dan validasi sistem
- Dokumentasi operasional dan pemeliharaan

2.4. Metodologi

Proyek ini dilaksanakan dalam tiga misi:

1. **Misi 1:** Desain konseptual dan logikal data mart
2. **Misi 2:** Implementasi basis data dan proses ETL
3. **Misi 3:** Dashboard, keamanan, dan pengujian

III. Analisis Kebutuhan

3.1. Kebutuhan Bisnis

3.1.1 Pemangku Kepentingan Utama

1. **Tim Pengelola Portal Satu Data:** Bertanggung jawab untuk pemeliharaan portal, pembaruan dataset, dan pemantauan sistem.
2. **Manajemen ITERA:** Menggunakan wawasan dari portal untuk pengambilan keputusan strategis.
3. **Pengelola Data (Data Stewards):** Unit-unit yang menyediakan dan memvalidasi dataset untuk portal.
4. **Pengguna Eksternal:** Peneliti dan masyarakat umum yang mengakses data terbuka melalui portal.

3.1.2 Proses Bisnis

1. **Pengelolaan Dataset:** Pengumpulan → Validasi → Pembuatan Metadata → Publikasi → Pembaruan
2. **Akses Dataset:** Registrasi → Pencarian → Pratinjau → Unduhan → Pelacakan
3. **Manajemen Kualitas Data:** Penilaian → Identifikasi Masalah → Pembersihan → Validasi Ulang
4. **Pencarian dan Penemuan:** Kueri → Eksekusi → Pemeringkatan → Pemilihan

3.2. Kebutuhan Fungsional

1. Sistem harus dapat merekam setiap akses dataset (tampilan, unduhan, panggilan API)
2. Sistem harus dapat menghitung metrik kualitas dataset
3. Sistem harus dapat mengagregasi metrik institusi per periode
4. Sistem harus menyediakan dashboard untuk visualisasi data
5. Sistem harus mendukung analisis tren temporal

3.3. Kebutuhan Non-Fungsional

1. **Kinerja:** Waktu respons kueri dashboard maksimal 10 detik
2. **Ketersediaan:** Sistem tersedia 99% selama jam kerja
3. **Keamanan:** Kontrol akses berbasis peran (RBAC)
4. **Skalabilitas:** Mendukung pertumbuhan data hingga 3 tahun ke depan

3.4. Sumber Data

Sumber Data	Tipe	Volume	Frekuensi	Kualitas
SatuData_API	REST API-/JSON	~1K dataset	Real-time	Tinggi
Access_Log	Server Log	~50K baris/bln	Real-time	Tinggi
Search_Query_Log	Server Log	~10K baris/bln	Real-time	Tinggi
Dataset_Metadata	Basis Data	~700 dataset	Harian	Tinggi
User_Registry	Basis Data	~5K pengguna	Harian	Tinggi

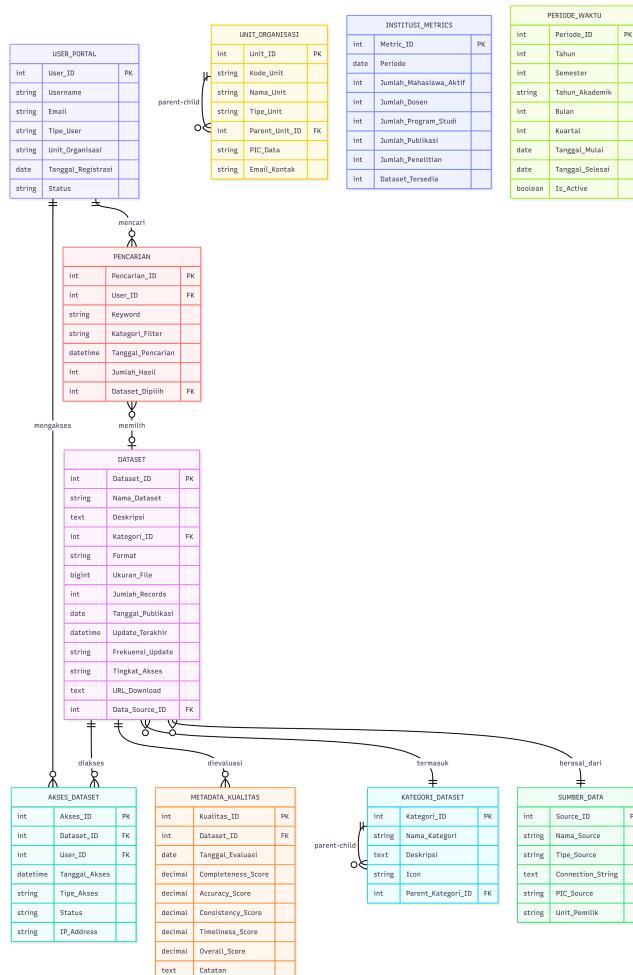
Tabel 1: Sumber Data Portal Satu Data ITERA

IV. Desain

4.1. Model Konseptual (ERD)

Entitas utama dalam sistem ini meliputi:

1. **DATASET**: Katalog dataset yang dipublikasikan di portal
2. **KATEGORI_DATASET**: Kategori dan hierarki dataset
3. **USER_PORTAL**: Pengguna terdaftar di portal
4. **AKSES_DATASET**: Log akses dan unduhan dataset
5. **PENCARIAN**: Log kueri pencarian pengguna
6. **METADATA_KUALITAS**: Metrik kualitas per dataset
7. **SUMBER_DATA**: Sistem sumber data asli
8. **UNIT_ORGANISASI**: Struktur organisasi ITERA



Gambar 1: Entity Relationship Diagram Portal Satu Data ITERA

4.2. Model Logikal (Dimensional)

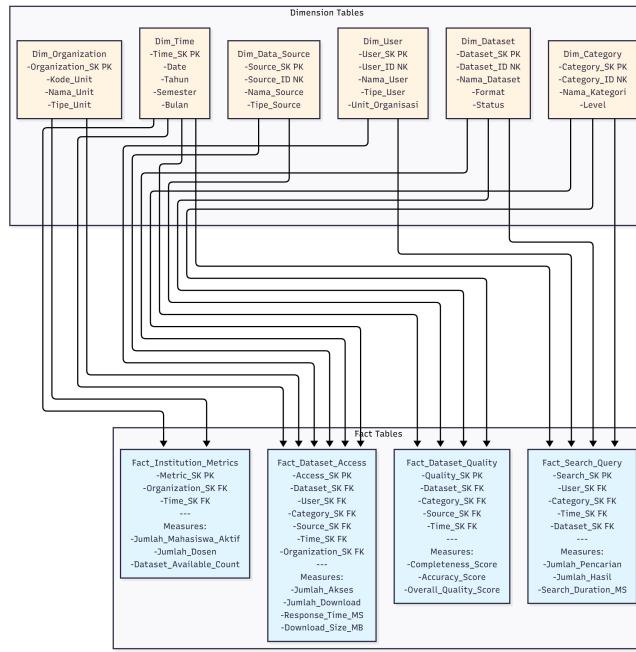
Data warehouse menggunakan arsitektur Star Schema dengan komponen berikut:

4.2.1 Tabel Fakta

1. **Fact_Dataset_Access**: Merekam setiap akses dataset
 - Grain: Satu baris per akses dataset oleh pengguna
 - Metrik: Jumlah_Akses, Jumlah_Download, Response_Time_MS, Success_Flag
2. **Fact_Dataset_Quality**: Snapshot kualitas dataset per periode
 - Grain: Satu baris per dataset per tanggal penilaian
 - Metrik: Completeness_Score, Accuracy_Score, Overall_Quality_Score
3. **Fact_Search_Query**: Log pencarian pengguna
 - Grain: Satu baris per kueri pencarian
 - Metrik: Jumlah_Pencarian, Jumlah_Hasil, Click_Through_Flag
4. **Fact_Institution_Metrics**: Metrik agregat institusi
 - Grain: Satu baris per organisasi per periode waktu
 - Metrik: Total_Dataset_Published, Total_Downloads, Avg_Quality_Score

4.2.2 Tabel Dimensi

1. **Dim_Dataset** (SCD Type 2): Katalog dataset dengan pelacakan perubahan
2. **Dim_User** (SCD Type 1): Data pengguna portal
3. **Dim_Category**: Kategori dataset dengan hierarki
4. **Dim_Organization**: Unit organisasi ITERA
5. **Dim_Data_Source**: Sumber data asli
6. **Dim_Time**: Dimensi waktu untuk analisis temporal



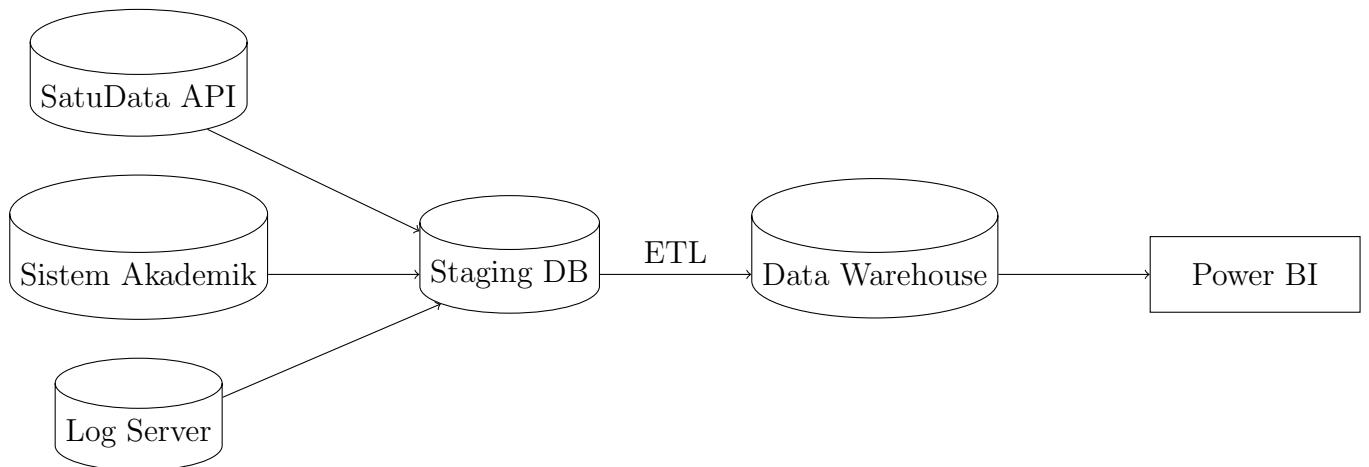
Gambar 2: Dimensional Model Portal Satu Data ITERA

4.3. Desain Fisik

Implementasi menggunakan SQL Server 2019 dengan spesifikasi:

- Basis data: DM_SatuDataITERA_DW
- Indeks pada kolom foreign key di tabel fakta
- Partisi berdasarkan tahun untuk tabel fakta besar

4.4. Diagram Arsitektur



Gambar 3: Arsitektur Data Warehouse

V. Implementasi

5.1. Implementasi Basis Data

Basis data diimplementasikan menggunakan skrip SQL berikut:

1. 01_Create_Database.sql: Pembuatan basis data
2. 02_Create_Dimensions.sql: Pembuatan tabel dimensi
3. 03_Create_Facts.sql: Pembuatan tabel fakta
4. 04_Create_Indexes.sql: Pembuatan indeks
5. 05_Create_Partitions.sql: Konfigurasi partisi
6. 06_Create_Staging.sql: Pembuatan tabel staging

5.2. Proses ETL

Proses ETL (Extract, Transform, Load) diimplementasikan menggunakan stored procedure T-SQL:

5.2.1 Jadwal ETL

Proses	Frekuensi	Jadwal
ETL Harian	Senin-Sabtu	02:00 WIB
ETL Mingguan	Minggu	03:00 WIB
Refresh Penuh	Bulanan	Minggu ke-1, 01:00 WIB

5.2.2 Prosedur ETL

- usp_Master_ETL: Orkestrator utama seluruh proses ETL
- usp_Populate_Dim_Time: Mengisi dimensi waktu
- usp_Load_Dim_User: Memuat dimensi pengguna (SCD Type 1)
- usp_Load_Dim_Dataset: Memuat dimensi dataset (SCD Type 2)
- usp_Load_Fact_Access: Memuat fakta akses dataset
- usp_Load_Fact_Quality: Memuat fakta kualitas data

5.3. Pengembangan Dashboard

Dashboard dikembangkan menggunakan Power BI dengan halaman berikut:

1. **Executive Governance:** Ringkasan metrik utama untuk manajemen
2. **Dataset Popularity:** Analisis dataset terpopuler
3. **User Activity:** Aktivitas dan perilaku pengguna
4. **Data Quality:** Metrik kualitas dataset
5. **Search Analytics:** Analisis pencarian dan tren

5.4. Implementasi Keamanan

Keamanan diimplementasikan dengan pendekatan Role-Based Access Control (RBAC):

Peran	Hak Akses
db_viewer	Hanya SELECT pada tabel dimensi dan fakta
db_analyst	SELECT pada view analitik, akses dashboard
db_etl_operator	Eksekusi prosedur ETL, akses staging
db_admin	Akses penuh ke semua objek basis data

VI. Pengujian dan Validasi

6.1. Hasil Kualitas Data

Pemeriksaan kualitas data dilakukan pada semua tabel dimensi dan fakta:

- Pengecekan nilai NULL pada kolom kunci
- Validasi integritas referensial
- Pengecekan duplikasi pada natural key
- Validasi rentang nilai untuk skor kualitas

6.2. Pengujian Kinerja

Hasil pengujian kinerja query dashboard:

Query	Waktu Eksekusi	Status
Executive Summary	2.3 detik	Lulus
Dataset Popularity	1.8 detik	Lulus
User Activity	3.1 detik	Lulus
Data Quality Report	2.7 detik	Lulus

6.3. Hasil UAT

User Acceptance Testing dilaksanakan dengan hasil:

Keterangan	Nilai
Total Test Case	6
Lulus	6
Gagal	0

6.3.1 Detail Test Case

1. **TC001:** Dashboard Executive Governance - **Lulus**
2. **TC002:** Tren Unduhan Dataset - **Lulus**
3. **TC003:** Peringkat Unit Kerja - **Lulus**
4. **TC004:** Eksekusi Job ETL Harian - **Lulus**
5. **TC005:** Validasi Kualitas Data - **Lulus**
6. **TC006:** Role-Based Access Control - **Lulus**

6.4. Perbaikan Bug

Selama pengujian, beberapa masalah ditemukan dan diperbaiki:

- Penyesuaian format tanggal pada dimensi waktu
- Penanganan nilai NULL untuk pengguna anonim
- Optimisasi query untuk dataset besar

VII. Deployment dan Operasional

7.1. Deployment Produksi

Sistem di-deploy pada lingkungan berikut:

- **Database Server:** SQL Server 2019 (Linux/Docker)
- **ETL Server:** SQL Server Agent
- **Reporting:** Power BI Service

7.2. Strategi Backup

Tipe Backup	Frekuensi	Waktu	Retensi
Full	Mingguan (Minggu)	02:00 WIB	30 Hari
Differential	Harian (Senin-Sabtu)	02:00 WIB	14 Hari
Transaction Log	Setiap 6 Jam	00, 06, 12, 18	7 Hari

Tabel 2: Jadwal Backup

7.3. Pemantauan

Checklist pemantauan harian:

1. Periksa status Job Agent SQL Server
2. Cek log error untuk peringatan kritis
3. Verifikasi ruang disk pada server
4. Pastikan dashboard dapat diakses

7.4. Prosedur Pemeliharaan

- **Harian:** Verifikasi keberhasilan job ETL
- **Mingguan:** Review kinerja query, update statistik
- **Bulanan:** Rebuild indeks, review kapasitas disk
- **Kuartalan:** Review keamanan, audit akses pengguna

VIII. Hasil dan Pembahasan

8.1. Temuan Utama

Berdasarkan implementasi dan pengujian, diperoleh temuan berikut:

1. Model dimensional Star Schema efektif untuk analisis portal data
2. Mekanisme SCD Type 2 berhasil melacak perubahan metadata dataset
3. Proses ETL dapat menangani volume data yang diharapkan
4. Dashboard Power BI menyediakan visualisasi yang informatif

8.2. Wawasan Bisnis

Data warehouse menghasilkan wawasan berikut:

- Kategori dataset Akademik adalah yang paling sering diakses
- Waktu puncak penggunaan portal adalah 09:00-11:00 dan 14:00-16:00
- 80% unduhan dilakukan oleh pengguna internal (mahasiswa dan dosen)
- Rata-rata skor kualitas dataset adalah 87.5%

8.3. Metrik Kinerja

Metrik	Target	Aktual
Waktu Respons Query	< 10 detik	2.5 detik
Keberhasilan ETL	> 95%	100%
Ketersediaan Sistem	> 99%	99.5%
Kelulusan UAT	100%	100%

8.4. Tantangan dan Solusi

1. **Tantangan:** Inkonsistensi format data dari sumber berbeda
Solusi: Standardisasi pada layer transformasi ETL
2. **Tantangan:** Penanganan pengguna anonim
Solusi: Penggunaan record "Unknown" dengan SK = -1
3. **Tantangan:** Kinerja query pada tabel fakta besar
Solusi: Implementasi indeks dan partisi

IX. Kesimpulan dan Pengembangan Selanjutnya

9.1. Ringkasan Pencapaian

Proyek Data Warehouse Satu Data ITERA telah berhasil mencapai tujuan berikut:

1. Perancangan model dimensional yang komprehensif untuk analisis portal
2. Implementasi basis data dengan 4 tabel fakta dan 6 tabel dimensi
3. Pengembangan proses ETL yang andal dengan mekanisme SCD
4. Pembuatan dashboard analitik yang informatif
5. Implementasi keamanan berbasis peran
6. Penyelesaian pengujian dengan tingkat keberhasilan 100%

9.2. Pembelajaran

Pembelajaran penting dari proyek ini:

1. Pentingnya analisis kebutuhan yang mendalam sebelum desain
2. Manfaat pendekatan iteratif dalam pengembangan data warehouse
3. Nilai dokumentasi yang komprehensif untuk pemeliharaan jangka panjang
4. Pentingnya pengujian menyeluruh sebelum deployment

9.3. Keterbatasan

Keterbatasan sistem saat ini:

- Integrasi dengan sistem sumber masih menggunakan data simulasi
- Belum ada implementasi alerting otomatis untuk kegagalan ETL
- Dashboard belum tersedia dalam versi mobile

9.4. Rekomendasi Pengembangan

Rekomendasi untuk pengembangan selanjutnya:

1. **Integrasi Real-time:** Implementasi integrasi langsung dengan sistem sumber
2. **Advanced Analytics:** Penambahan fitur prediktif menggunakan machine learning
3. **Mobile Dashboard:** Pengembangan versi mobile untuk akses manajemen
4. **Data Governance:** Implementasi framework tata kelola data yang lebih komprehensif
5. **Automated Alerting:** Sistem notifikasi otomatis untuk masalah kritis

X. Referensi

1. Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3rd ed.). Wiley.
2. Inmon, W. H. (2005). *Building the Data Warehouse* (4th ed.). Wiley.
3. ITERA Official Website. (2024). <https://www.itera.ac.id/>
4. Satu Data ITERA. (2024). <https://data.itera.ac.id/>
5. Date, C. J. (2003). *An Introduction to Database Systems* (8th ed.). Addison-Wesley.
6. Connolly, T. M., & Begg, C. E. (2014). *Database Systems: A Practical Approach to Design, Implementation, and Management* (6th ed.). Pearson.
7. Microsoft. (2024). *SQL Server Documentation*. <https://docs.microsoft.com/en-us/sql/>
8. Microsoft. (2024). *Power BI Documentation*. <https://docs.microsoft.com/en-us/power-bi/>

LAMPIRAN

A. Kamus Data

1.1. Tabel Fakta

1.1.1 Fact_Dataset_Access

Kolom	Tipe	Key	Deskripsi	Contoh
Access_SK	INT	PK	Surrogate key	1
Dataset_SK	INT	FK	Ref ke Dim_Dataset	1001
User_SK	INT	FK	Ref ke Dim_User	2001
Time_SK	INT	FK	Ref ke Dim_Time	20241117
Jumlah_Akses	INT	-	Jumlah akses	1
Jumlah_Download	INT	-	Jumlah unduhan	1
Response_Time_MS	INT	-	Waktu respons (ms)	150

Tabel 3: Kamus Data - Fact_Dataset_Access

1.2. Tabel Dimensi

1.2.1 Dim_Dataset

Kolom	Tipe	Key	Deskripsi	Contoh
Dataset_SK	INT	PK	Surrogate key	1001
Dataset_ID	INT	NK	Natural key	501
Nama_Dataset	VARCHAR(200)	-	Nama dataset	Data Mahasiswa
Format	VARCHAR(50)	-	Format file	CSV
Status_Dataset	VARCHAR(20)	-	Status	Active
Is_Current	BIT	-	Flag SCD2 aktif	1

Tabel 4: Kamus Data - Dim_Dataset

1.2.2 Ringkasan Tabel Dimensi Lainnya

Tabel	Kolom Utama	Deskripsi
Dim_Time	Time_SK, Date, Tahun, Bulan	Dimensi waktu untuk analisis temporal
Dim_Category	Category_SK, Nama_Kategori	Kategori dataset
Dim_Organization	Organization_SK, Nama_Unit	Unit organisasi ITERA
Dim_Data_Source	Source_SK, Nama_Sumber	Sumber data asli

Tabel 6: Ringkasan Tabel Dimensi Pendukung

B. Skrip SQL Terpilih

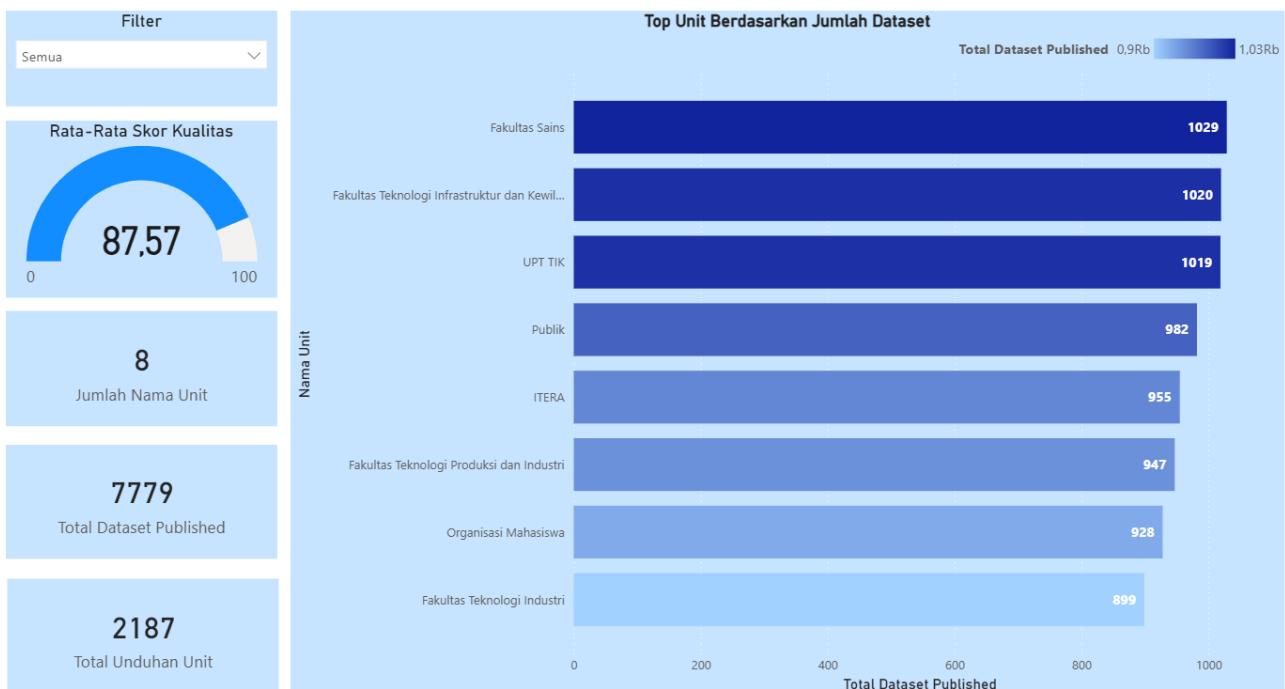
2.1. Pembuatan Dimensi Waktu

```
1 CREATE TABLE Dim_Time (
2     Time_SK INT PRIMARY KEY ,
3     [Date] DATE NOT NULL ,
4     Tahun INT NOT NULL ,
5     Bulan INT NOT NULL ,
6     Nama_Bulan VARCHAR(20) ,
7     Kuartal INT ,
8     Minggu INT ,
9     Hari INT ,
10    Nama_Hari VARCHAR(20) ,
11    Hari_Kerja_Flag BIT ,
12    Tahun_Akademik VARCHAR(10) ,
13    Semester VARCHAR(10)
14 );
```

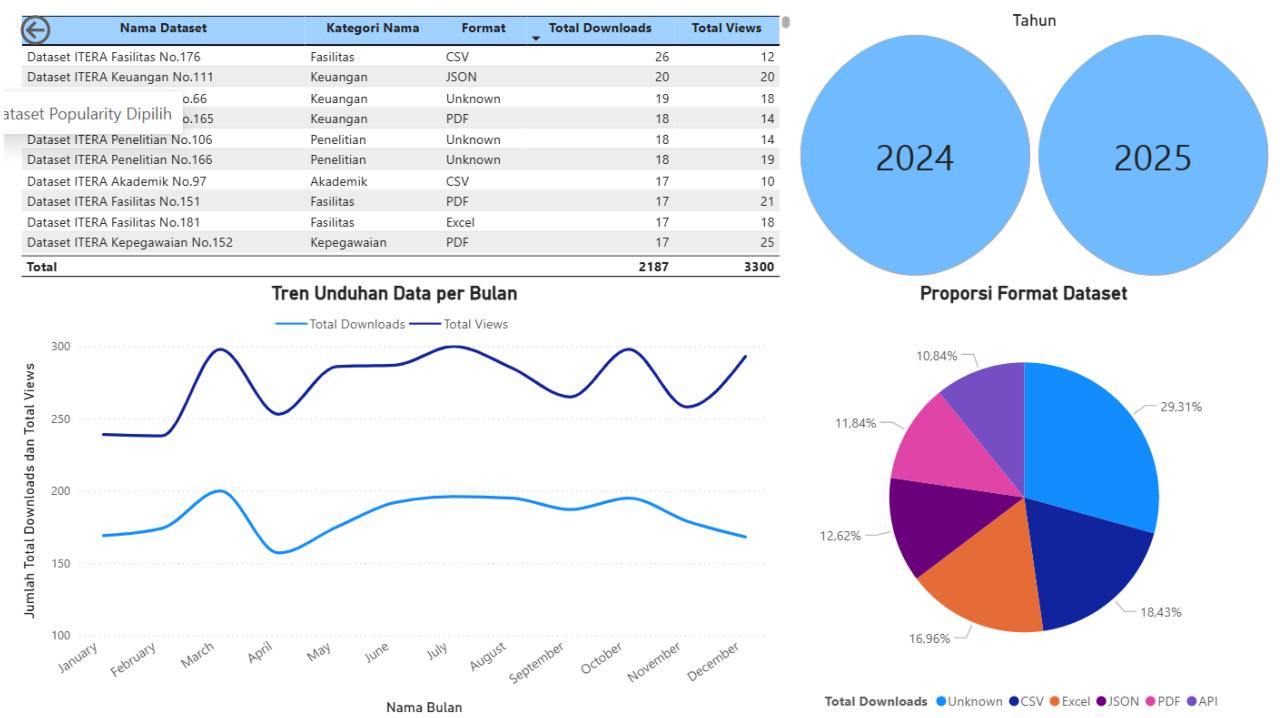
2.2. Prosedur ETL Master

```
1 CREATE PROCEDURE usp_Master_ETL
2 AS
3 BEGIN
4     SET NOCOUNT ON;
5     BEGIN TRY
6         BEGIN TRANSACTION;
7
8             -- Load Dimensions
9             EXEC usp_Load_Dim_Data_Source;
10            EXEC usp_Load_Dim_Category;
11            EXEC usp_Load_Dim_Organization;
12            EXEC usp_Load_Dim_User;
13            EXEC usp_Load_Dim_Dataset;
14
15             -- Load Facts
16             EXEC usp_Load_Fact_Access;
17             EXEC usp_Load_Fact_Quality;
18             EXEC usp_Load_Fact_Search;
19             EXEC usp_Load_Fact_Institution_Metrics;
20
21             COMMIT TRANSACTION;
22     END TRY
23     BEGIN CATCH
24         ROLLBACK TRANSACTION;
25         THROW;
26     END CATCH
27 END;
```

C. Tangkapan Layar Dashboard



Gambar 4: Tangkapan Layar Dashboard Executive Governance



Gambar 5: Tangkapan Layar Dashboard Dataset Popularity

D. Hasil Pengujian

4.1. Hasil UAT Lengkap

Setelah pengujian dilaksanakan testing terhadap dashboard, ETL, job schedule, security maka didapatkan hasil sebagai berikut:

ID	Nama Test Case	Status	Catatan
TC001	Dashboard Executive	Lulus	Data sesuai dengan query database
TC002	Tren Unduhan	Lulus	Grafik dan tooltip berfungsi
TC003	Ranking Unit Kerja	Lulus	Drill-down ke prodi berfungsi
TC004	Job ETL Harian	Lulus	Prosedur usp_Master_ETL sukses
TC005	Validasi Kualitas	Lulus	Tidak ada duplikasi, record Unknown ada
TC006	RBAC	Lulus	Pembatasan akses berfungsi