

LAPORAN TUGAS PERGUDANGAN DATA

Analisis Kebutuhan Data Warehouse untuk Sekolah Dasar Bombardilo Crocodilo



Disusun Oleh Kelompok 9 - RA

Dhea Amelia Putri	:122450004
Marleta Cornelia Leander	:122450092
Berliyana Kesuma Hati	:121450086
Najla Juwairia	:122450037
Nurul Alfajar Gumel	:122450127
Tarisah	:121450141

**PROGRAM STUDI SAINS DATA
FAKULTAS SAINS
INSTITUT TEKNOLOGI SUMATERA
LAMPUNG SELATAN
2025**

1. Pendahuluan

Laporan ini menyusun analisis kebutuhan dan desain konseptual sistem Data Warehouse untuk Sekolah Dasar Bombardilo Crocodilo. Pada misi ini, fokus diarahkan pada perancangan dan implementasi lebih lanjut, mencakup desain logikal dan fisik dari skema multidimensi yang telah dirancang. Termasuk di dalamnya adalah desain tabel dimensi dan fakta secara rinci, strategi indexing, struktur penyimpanan data, serta partisi dan views untuk optimasi kinerja sistem pergudangan data.

2. Desain Konseptual

2.1 Deskripsi Data

Dataset yang digunakan pada penelitian ini yaitu data dummy yang terdiri dari 1000 baris di setiap tabelnya. Data ini menyajikan gambaran menyeluruh tentang sistem penilaian siswa di Sekolah Dasar Bombardilo Crocodilo dimana setiap siswa memiliki informasi lengkap seperti nama, jenis kelamin, kelas, tahun masuk, dan nama orang tua. Data siswa ini terhubung dengan data mata pelajaran yang mencakup nama pelajaran, kategori, tingkat kelas, serta nilai ambang kelulusan (KKM). Selain itu, terdapat data guru yang mencatat siapa pengampu mata pelajaran tersebut, lengkap dengan bidang ajar, status kepegawaian, dan latar belakang pendidikan mereka. Seluruh informasi ini kemudian dikaitkan dalam data nilai, yang mencatat hasil ujian dan tugas siswa untuk setiap mata pelajaran yang diajarkan oleh guru tertentu. Nilai-nilai ini dihitung rata-ratanya untuk menentukan status kelulusan siswa dalam mata pelajaran tersebut. Dengan keterhubungan antar data ini, sistem dapat memberikan gambaran menyeluruh mengenai performa siswa, kompetensi guru, serta efektivitas pengajaran berdasarkan hasil belajar yang tercatat.

2.2 Tabel Dimensi

Struktur data pada penelitian ini terdiri dari tiga tabel dimensi yaitu Dimensi Siswa, Dimensi Mata Pelajaran, dan Dimensi Guru. Dimensi Siswa mencakup informasi identitas siswa seperti nama, NISN, jenis kelamin, kelas, dan nama orang tua. Dimensi Mata Pelajaran mencatat detail pelajaran seperti nama, kategori, tingkat, KKM, dan tahun mulai diajarkan. Sementara itu, Dimensi Guru berisi data pengajar seperti nama, NIP, bidang ajar, status, tahun mengajar, dan pendidikan terakhir. Ketiga dimensi ini saling terhubung dan mendukung analisis data nilai siswa.

Tabel 1. Struktur Tabel Dimensi

Dimensi	Atribut	Tipe Data
Siswa	ID_Siswa	VARCHAR
	Nama_Siswa	VARCHAR
	NISN	INT

Mata Pelajaran	Jenis_Kelamin	VARCHAR
	Kelas	VARCHAR
	Tahun_Masuk	INT
	Nama_Orangtua	VARCHAR
	ID_MatPel	VARCHAR
	Nama_Matpel	VARCHAR
	Kategori	VARCHAR
	Tingkat	INT
Guru	KKM	INT
	ID_Guru	VARCHAR
	Nama_Guru	VARCHAR
	NIP	INT
	Bidang_Ajar	VARCHAR
	Status	VARCHAR
	Tahun_Mengajar	INT
	Pendidikan	VARCHAR

2.3 Tabel Fakta

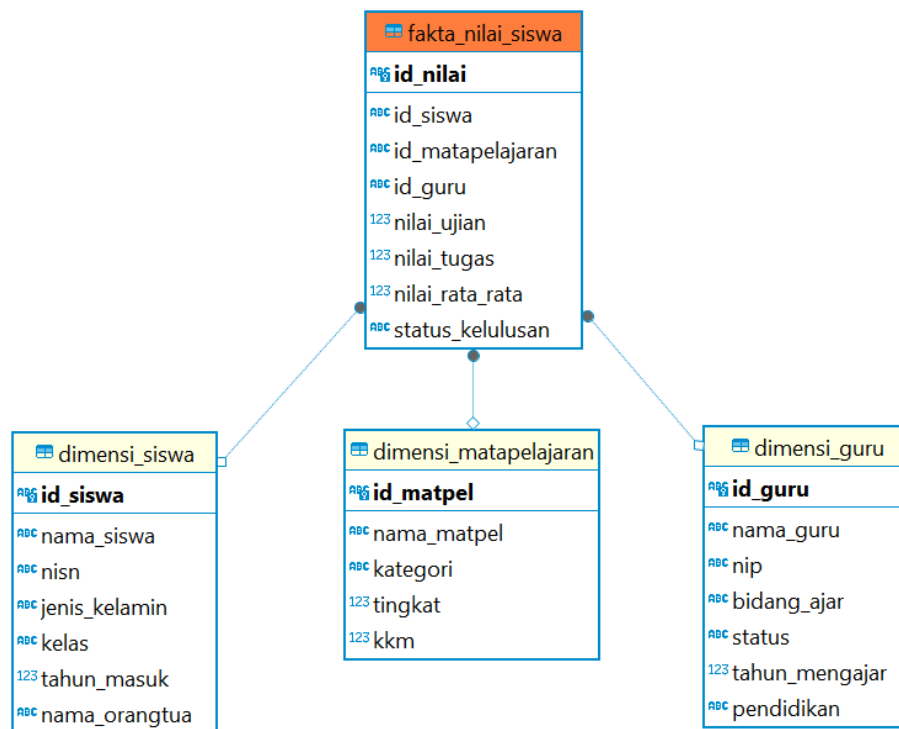
Tabel fakta pada penelitian ini yaitu Fakta_Nilai_Siswa yang menyimpan data hasil belajar siswa dengan menghubungkan tiga dimensi utama yaitu siswa, guru, dan mata pelajaran. Tabel ini mencatat nilai ujian, nilai tugas, nilai rata-rata, serta status kelulusan siswa untuk setiap mata pelajaran yang diajarkan oleh guru tertentu.

Tabel 2. Struktur Tabel Fakta

Atribut	Tipe Data	Keterangan
ID_Nilai	VARCHAR	Kode unik untuk tiap entri nilai
ID_Siswa	VARCHAR	Kode yang menghubungkan ke identitas siswa
ID_MataPelajaran	VARCHAR	Kode yang menghubungkan ke data mata pelajaran

ID_Guru	VARCHAR	Kode yang menghubungkan ke data guru
Nilai_Ujian	INT	Nilai ujian yang diperoleh siswa
Nilai_Tugas	INT	Nilai tugas yang diperoleh siswa
Nilai_Rata-Rata	FLOAT	Rata-rata dari nilai ujian dan tugas
Status_Kelulusan	VARCHAR	Status apakah siswa lulus atau tidak

2.4 Skema Data



Gambar 1. Star Schema

Star schema di atas digunakan untuk menganalisis data nilai siswa, dengan tabel fakta di pusat yang mencatat nilai ujian, nilai tugas, rata-rata, dan status kelulusan. Tabel fakta ini terhubung ke tiga tabel dimensi, yaitu Dimensi Siswa yang memuat data identitas siswa, Dimensi Mata Pelajaran yang berisi informasi pelajaran dan KKM, serta Dimensi Guru yang mencakup data guru seperti nama, NIP, bidang ajar, status, tahun mengajar, dan pendidikan terakhir. Skema ini digunakan karena sederhana, mudah dipahami, dan efisien untuk melakukan analisis berbasis kuantitas seperti perbandingan nilai antar siswa, pelajaran, atau guru. Star schema juga dapat mempermudah dalam pelaporan dan pengambilan keputusan secara cepat di lingkungan pendidikan.

3. Desain Logikal dan Fisik

3.1 Translasi Desain Konseptual ke Model Relasional

Translasi desain konseptual ke model relasional adalah proses mengubah rancangan awal basis data (biasanya berupa diagram seperti ERD) menjadi bentuk tabel-tabel yang bisa digunakan langsung di sistem manajemen basis data (DBMS). Tujuannya adalah agar struktur data yang telah dirancang secara logis dapat diimplementasikan secara nyata dalam bentuk tabel relasional yang saling terhubung, sehingga bisa menyimpan, mengelola, dan mengakses data dengan efisien sesuai kebutuhan aplikasi atau organisasi. Query yang digunakan yaitu sebagai berikut.

```
-- Tabel DIMENSI SISWA
CREATE TABLE Dimensi_Siswa (
  ID_Siswa VARCHAR(10) PRIMARY KEY,
  Nama_Siswa VARCHAR(100),
  NISN VARCHAR(20),
  Jenis_Kelamin CHAR(1),
  Kelas VARCHAR(10),
  Tahun_Masuk INT,
  Nama_Orangtua VARCHAR(100)
);

-- Tabel DIMENSI MATA PELAJARAN
CREATE TABLE Dimensi_MataPelajaran (
  ID_MatPel VARCHAR(10) PRIMARY KEY,
  Nama_MatPel VARCHAR(100),
  Kategori VARCHAR(50),
  Tingkat INT,
  KKM INT
);

-- Tabel DIMENSI GURU
CREATE TABLE Dimensi_Guru (
  ID_Guru VARCHAR(10) PRIMARY KEY,
  Nama_Guru VARCHAR(100),
  NIP VARCHAR(20),
  Bidang_Ajar VARCHAR(100),
  Status VARCHAR(50),
  Tahun_Mengajar INT,
  Pendidikan VARCHAR(100)
);

-- Tabel FAKTA NILAI SISWA
CREATE TABLE Fakta_Nilai_Siswa (
  ID_Nilai VARCHAR(10) PRIMARY KEY,
```

```

ID_Siswa VARCHAR(10),
ID_MataPelajaran VARCHAR(10),
ID_Guru VARCHAR(10),
Nilai_Ujian INT,
Nilai_Tugas INT,
Nilai_Rata_Rata DECIMAL(5,2),
Status_Kelulusan VARCHAR(20),

-- Foreign Keys
FOREIGN KEY (ID_Siswa) REFERENCES Dimensi_Siswa(ID_Siswa),
FOREIGN KEY (ID_MataPelajaran) REFERENCES
Dimensi_MataPelajaran(ID_MatPel),
FOREIGN KEY (ID_Guru) REFERENCES Dimensi_Guru(ID_Guru)
);

```

3.2 Strategi Perancangan dan Implementasi Index

Dalam perancangan basis data menggunakan model star schema, salah satu strategi optimasi yang diterapkan adalah pembuatan indeks pada atribut foreign key di tabel fakta. Tabel fakta berfungsi sebagai pusat integrasi data yang berelasi langsung dengan beberapa tabel dimensi melalui atribut kunci asing (foreign key), seperti ID_Siswa, ID_MataPelajaran, dan ID_Guru. Untuk meningkatkan efisiensi saat melakukan proses join antara tabel fakta dan tabel dimensi, maka diterapkan indeks pada ketiga atribut tersebut. Penerapan indeks ini bertujuan untuk mempercepat eksekusi query yang sering melakukan penggabungan data (join) antara tabel fakta dan dimensi, serta penyaringan data berdasarkan entitas siswa, mata pelajaran, atau guru. Dengan adanya indeks, sistem basis data dapat menelusuri baris-baris yang relevan secara lebih efisien tanpa perlu melakukan pencarian secara penuh (full table scan). Berikut merupakan kode SQL yang digunakan untuk membuat indeks pada atribut foreign key di tabel Fakta_Nilai_Siswa.

```

CREATE INDEX idx_fk_id_siswa ON Fakta_Nilai_Siswa(ID_Siswa);
CREATE INDEX idx_fk_id_matpel ON Fakta_Nilai_Siswa(ID_MataPelajaran);
CREATE INDEX idx_fk_id_guru ON Fakta_Nilai_Siswa(ID_Guru);

```

3.3 Desain Partisi Tabel dan View untuk Akses Efisien

Desain partisi tabel dan penggunaan view merupakan strategi penting dalam mengoptimalkan akses dan pengelolaan data pada sistem basis data skala besar, seperti data warehouse berbasis star schema. Partisi tabel merupakan teknik untuk memecah satu tabel besar menjadi beberapa bagian lebih kecil berdasarkan kriteria tertentu, sehingga sistem dapat membaca hanya bagian data yang relevan tanpa harus memproses seluruh tabel. Hal ini secara langsung meningkatkan kecepatan query dan efisiensi sumber daya. Pada konsep kali ini, tabel Fakta_Nilai_Siswa yang menyimpan ribuan data nilai siswa dipartisi berdasarkan status kelulusan. Sementara itu, view adalah representasi virtual dari

satu atau lebih tabel yang digunakan untuk menyederhanakan query yang kompleks dan menyajikan data secara tematik atau terfokus. Penelitian ini mengkombinasikan view dengan partisi untuk memberikan akses lebih cepat ke subset data tertentu, seperti view yang hanya menampilkan siswa yang lulus atau view khusus untuk data per guru atau mata pelajaran. Query SQL yang digunakan sebagai berikut.

```
-- PARTISI
PARTITION BY LIST (Status_Kelulusan) (
  PARTITION p_lulus VALUES IN ('Lulus'),
  PARTITION p_tidak_lulus VALUES IN ('Tidak Lulus')
);

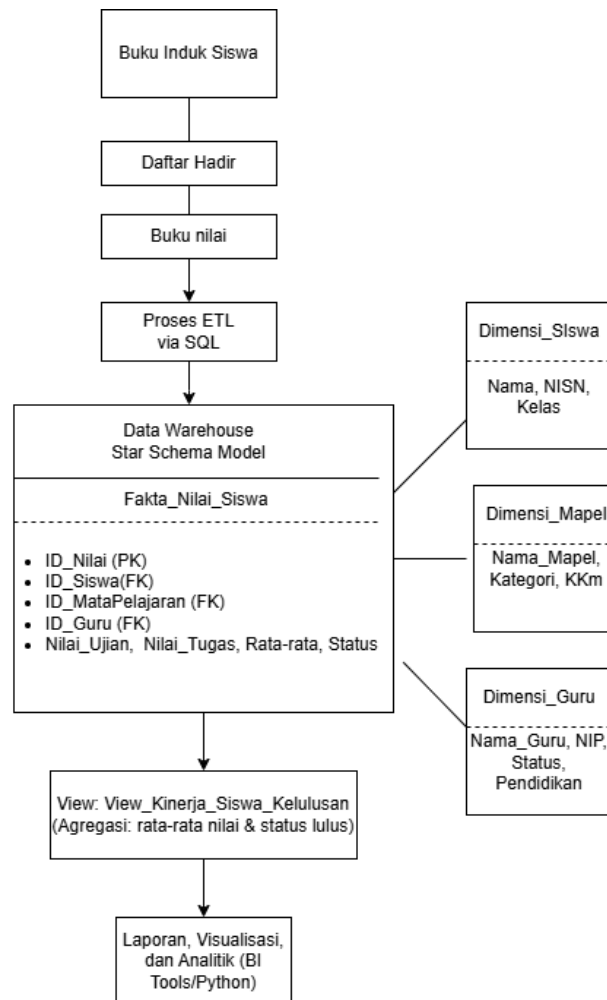
-- VIEW
CREATE VIEW View_Kinerja_Siswa_Kelulusan AS
SELECT
  Status_Kelulusan,
  AVG(Nilai_Ujian) AS RataRata_Ujian,
  AVG(Nilai_Tugas) AS RataRata_Tugas,
  AVG(Nilai_Rata_Rata) AS RataRata_Akhir,
  COUNT(*) AS Jumlah_Nilai
FROM Fakta_Nilai_Siswa
GROUP BY Status_Kelulusan
ORDER BY Status_Kelulusan;
```

3.4 Optimalisasi Penyimpanan dan Organisasi Data

Optimalisasi penyimpanan dan organisasi data merupakan aspek penting dalam pembangunan sistem data warehouse yang efisien. Dalam penelitian ini, data disusun menggunakan pendekatan star schema, di mana tabel fakta menjadi pusat penyimpanan data transaksional dan tabel dimensi berfungsi untuk menyimpan atribut deskriptif seperti informasi siswa, mata pelajaran, dan guru. Untuk meningkatkan efisiensi penyimpanan dan kinerja query, tabel fakta diimplementasikan dengan teknik partisi berdasarkan atribut status kelulusan. Partisi ini memungkinkan sistem basis data membatasi proses pembacaan data hanya pada segmen yang relevan, sehingga mempercepat eksekusi query dan mengurangi beban sistem. Selain itu, penggunaan indeks pada atribut foreign key yang menghubungkan tabel fakta dan dimensi sangat membantu dalam mempercepat proses join, terutama ketika data berukuran besar. Di sisi lain, penerapan view dimaksudkan untuk menyederhanakan proses pengambilan data oleh pengguna akhir, dengan menyediakan antarmuka virtual yang merepresentasikan query kompleks dalam bentuk yang lebih mudah diakses. Keseluruhan strategi ini tidak hanya meningkatkan performa sistem, tetapi juga mendukung kebutuhan analisis data secara cepat, konsisten, dan terstruktur, yang menjadi tujuan utama dari implementasi data warehouse.

4. Arsitektur Sistem dan ETL

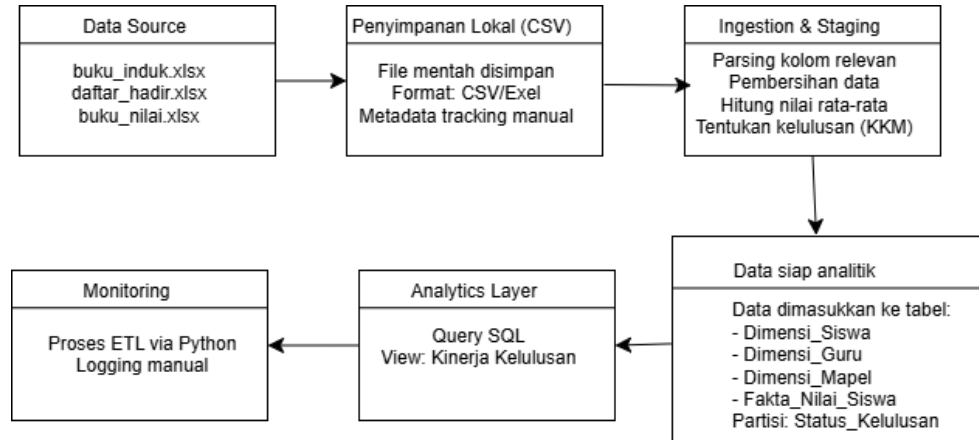
4.1 Arsitektur Data Warehouse



Gambar 2. Arsitektur Data

Diagram diatas menggambarkan arsitektur *logic* Data Warehouse yang menggunakan star schema, di mana data dari Buku Induk Siswa, Daftar Hadir, dan Buku Nilai diekstrak melalui proses ETL berbasis SQL dan dimuat ke dalam tabel fakta utama Fakta_Nilai_Siswa. Tabel ini menghubungkan tiga dimensi yaitu Dimensi_Siswa, Dimensi_Mapel, dan Dimensi_Guru, yang masing-masing menyimpan atribut deskriptif untuk mendukung analisis performa akademik. Fakta yang disimpan mencakup nilai ujian, nilai tugas, rata-rata, dan status kelulusan. Data tersebut kemudian diringkaskan dalam view View_Kinerja_Siswa_Kelulusan yang menyajikan agregasi nilai dan status lulus, dan akhirnya digunakan dalam proses analitik serta pelaporan menggunakan SQL, Python, atau BI tools.

4.2 Pipeline

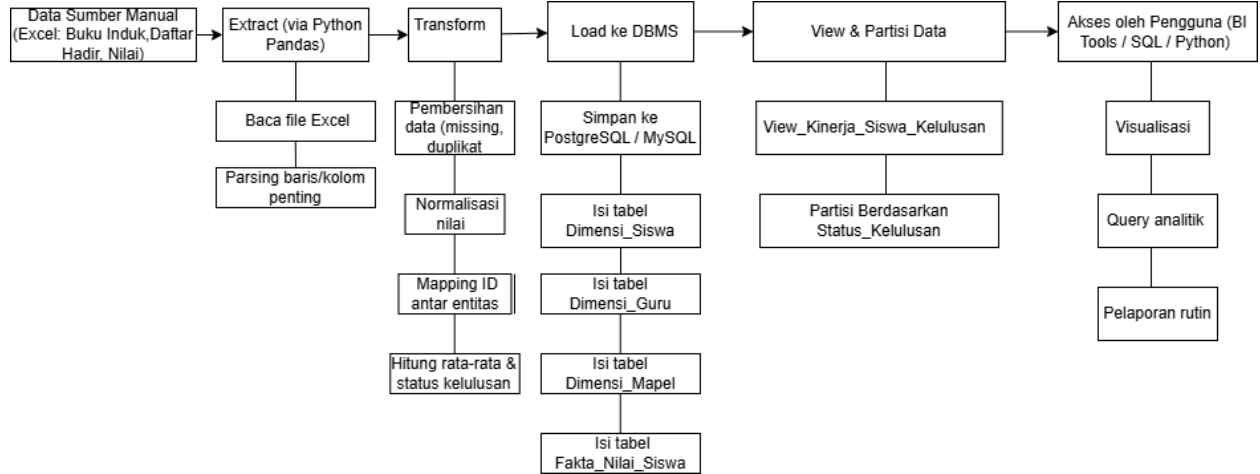


Gambar 3. Pipeline Data

Pipeline merupakan rangkaian proses otomatis yang digunakan untuk memindahkan data dari sistem sumber ke dalam data warehouse sekolah dasar bombardilo crocodilo secara terstruktur. Pipeline ini dirancang untuk mendukung proses ETL (Extract, Transform, Load) agar data dari siswa, guru dan mata pelajaran dapat dimuat ke dalam skema star schema yang telah dirancang sebelumnya. Pipeline ini menjadi bagian penting dari arsitektur data warehouse karena memastikan bahwa data yang dianalisis di sekolah dasar bombardilo crocodilo selalu akurat, lengkap, dan dapat digunakan untuk menghasilkan laporan kinerja siswa, efektivitas guru, serta kualitas mata pelajaran berdasarkan data nilai yang telah terintegrasi.

Pipeline ETL pada sistem Data Warehouse Sekolah Dasar Bombardilo Crocodilo terdiri dari tahapan ekstraksi data dari berbagai sumber (buku induk, daftar hadir, dan buku nilai), penyimpanan awal dalam format CSV/Excel, hingga proses staging yang melibatkan pembersihan data, parsing kolom relevan, perhitungan rata-rata, dan penentuan status kelulusan berdasarkan KKM. Data yang telah bersih dimuat ke dalam tabel dimensi dan fakta menggunakan skema star, dengan partisi berdasarkan status kelulusan untuk efisiensi akses. Selanjutnya, lapisan analitik memungkinkan query SQL dan visualisasi performa siswa, sementara proses ETL dijalankan via Python dengan pencatatan manual sebagai bagian dari monitoring sistem.

4.3 Proses ETL



Gambar 4. Proses ETL

ETL (Extract, Transform, Load) merupakan proses inti dalam pembangunan data warehouse yang bertujuan untuk mengintegrasikan data dari berbagai sumber ke dalam satu sistem penyimpanan terpusat. Dalam konteks penelitian ini, data warehouse dirancang untuk mengelola dan menganalisis informasi nilai siswa secara terstruktur menggunakan model star schema.

4.3.1 Extract

Proses extract merupakan tahap awal dalam ETL yang bertujuan untuk mengambil data mentah dari berbagai sumber operasional, seperti file Excel, database sekolah, atau sistem manajemen akademik lainnya. Pada penelitian ini, data yang diekstrak mencakup informasi siswa, guru, mata pelajaran, dan nilai siswa. Setiap data tersebut kemudian dikumpulkan dan disimpan dalam bentuk sementara untuk diproses lebih lanjut pada tahap transformasi. Proses ini penting untuk memastikan bahwa seluruh data yang dibutuhkan tersedia secara lengkap sebelum masuk ke tahap selanjutnya.

4.3.2 Transform

Transformasi merupakan tahap krusial di mana data mentah yang telah diekstrak selanjutnya dibersihkan, dikonversi, dan disesuaikan dengan struktur yang telah ditentukan dalam star schema. Dalam penelitian ini, proses transformasi mencakup standarisasi format data (penulisan nama, jenis kelamin, dll), penghapusan data duplikat, kalkulasi nilai rata-rata dari nilai ujian dan tugas, serta penentuan status kelulusan berdasarkan nilai KKM tiap mata pelajaran. Transformasi juga melibatkan penggabungan data antar entitas seperti siswa, guru, dan mata pelajaran agar saling terhubung secara logis.

4.3.3 Load

Setelah data berhasil ditransformasikan, tahap terakhir yaitu load dilakukan dengan cara memasukkan data ke dalam sistem data warehouse. Data dimensi seperti siswa, guru, dan mata pelajaran dimuat terlebih dahulu untuk memastikan integritas referensial. Kemudian, data nilai siswa dimuat ke dalam tabel fakta yang menghubungkan ketiga dimensi tersebut. Pada tahap ini juga diterapkan partisi tabel berdasarkan status kelulusan untuk optimasi penyimpanan, serta indeks pada atribut foreign key untuk mempercepat proses query. Hasil akhirnya adalah data yang tersimpan secara terstruktur dan siap digunakan untuk analisis dan pelaporan.

5. Penutup

Dengan menyelesaikan desain konseptual, logikal, dan fisik untuk Data Warehouse Sekolah Dasar Bombardilo Crocodilo, sistem ini diharapkan mampu mengintegrasikan data akademik dan administrasi secara efisien. Implementasi skema star dengan optimalisasi index, storage, dan partisi memungkinkan analisis data yang lebih cepat, akurat, dan relevan bagi para stakeholder seperti guru, kepala sekolah, hingga Dinas Pendidikan.

LAMPIRAN

https://bit.ly/DatasetSQL_DW_Kel9_RA