

Laporan Data Quality

Misi 2 Step 6: ETL Testing & Validation

Kelompok 10

1 Kesimpulan Utama

Berdasarkan pengujian kualitas data (*Data Quality Checks*) yang dilakukan terhadap tabel *Staging* dan *Data Warehouse*, kami melaporkan hasil sebagai berikut:

- **Status ETL:** Sukses. Proses *Extract, Transform, Load* (ETL) berhasil memindahkan 100% data transaksi ke dalam Data Warehouse.
- **Volume Data:** Total 18.324 baris data akses berhasil dimuat.
- **Temuan Utama:** Ditemukan beberapa anomali data yang bersumber dari sistem hulu (*upstream*), meliputi ketidaklengkapan atribut wajib (*NULL values*) dan ketidakakuratan rentang nilai (*Invalid Range*). Isu integritas referensial (*Orphan Records*) berhasil ditangani secara otomatis oleh sistem ETL.

2 Quality Metrics Summary

Tabel berikut merangkum hasil pengecekan kualitas data berdasarkan dimensi kualitas: *Completeness*, *Accuracy*, *Consistency*, dan *Uniqueness*.

Tabel 1: Ringkasan Metrik Kualitas Data

ID	Dimensi	Deskripsi Pengecekan	Error	Status
DQ-01	Completeness	Cek Username kosong pada tabel Dim_User	1	FAIL
DQ-02	Completeness	Cek Nama_Dataset kosong pada tabel Dim_Dataset	1	FAIL
DQ-03	Consistency	Cek <i>Orphan Records</i> (Anak tanpa Induk) di Fact_Access	0	PASS
DQ-04	Accuracy	Cek Skor Kualitas di luar rentang (0-100)	1	FAIL
DQ-05	Accuracy	Cek File Size bernilai negatif	0	PASS
DQ-06	Uniqueness	Cek Duplikasi pada User_ID	0	PASS

ID	Dimensi	Deskripsi Pengecekan	Error	Status
DQ-07	Uniqueness	Cek Duplikasi Record Aktif (SCD Type 2)	0	PASS

3 Detail Temuan dan Analisis

1	Dim_User	505	Username is NULL
1	Dim_Dataset	106	Nama_Dataset is NULL
	Access_SK	Dataset_SK	Issue
	Quality_SK	Dataset_SK	Issue
1	3	150.00	Score Out of Range
	Access_SK	File_Size_Downloaded	Issue
	User_SK	Tanggal_Registrasi	Issue
	User_ID	Duplicate_Count	
	Dataset_ID	Active_Count	

Gambar 1: Screenshot hasil data quality

3.1 Kelengkapan Data (Completeness)

Ditemukan adanya data master yang tidak memiliki atribut wajib.

- **Temuan:** Terdapat 1 Dataset dengan Dataset_ID = 106 yang memiliki Nama_Dataset bernilai NULL.
- **Dampak:** Dataset ini akan muncul sebagai ”(Unknown)” atau kosong pada dashboard laporan, menyulitkan identifikasi.

3.2 Akurasi Data (Accuracy)

Ditemukan data numerik yang tidak masuk akal secara logika bisnis.

- **Temuan:** Terdapat 1 record pada Fact_Dataset_Quality yang memiliki Overall_Quality_Score sebesar **150.00**.
- **Analisis:** Batas maksimum skor seharusnya adalah 100.00. Hal ini mengindikasikan kesalahan input atau kalkulasi pada sistem sumber.

3.3 Konsistensi (Consistency)

Pengujian integritas referensial antara tabel Fakta dan Dimensi.

- **Hasil:** 0 Error.
- **Analisis:** Meskipun terdapat data transaksi log yang merujuk ke Dataset ID 999 (tidak ada di master), prosedur ETL `usp_Load_Fact_Access` telah berhasil memfilter data tersebut menggunakan klausa `WHERE IS NOT NULL`. Data kotor berhasil dicegat dan tidak masuk ke Data Warehouse.