

Dokumentasi Data Quality DAG

1. Tujuan DAG

DAG kualitas data hanya digunakan untuk validasi dan mengecek kualitas data harian dari file CSV yang berisi data pasar saham Brasil. DAG ini akan mendeteksi masalah seperti data hilang, nilai tidak valid, duplikat, dan inkonsistensi antar tabel.

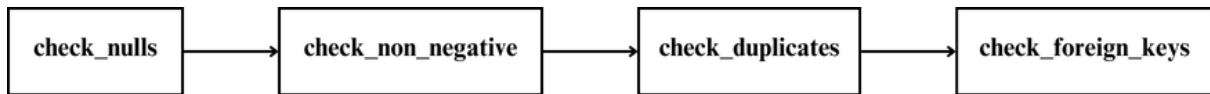
2. Penjadwalan DAG

Dataset: Brazil Stock Market

Nama DAG: data_quality_brazil_stock

Jadwal: Harian (@daily)

3. Struktur Task dan Dependensi



4. Task Detail

a. check_nulls

Deskripsi: Memastikan kolom penting tidak berisi NULL.

Kolom dicek: keyTime, keyCompany, openValueStock, closeValueStock, highValueStock, lowValueStock, quantityStock.

Ambang batas: Tidak boleh ada NULL.

b. check_non_negative

Deskripsi: Memastikan semua nilai numerik bernilai positif/0.

Kolom dicek: openValueStock, closeValueStock, highValueStock, lowValueStock, quantityStock.

Ambang batas: Tidak boleh ada nilai < 0 .

c. check_duplicates

Deskripsi: Mengecek tidak ada data duplikat per saham per waktu.

Kombinasi dicek: keyTime + keyCompany.

Ambang batas: Kombinasi harus unik.

d. check_foreign_keys

Deskripsi: Memastikan keyCompany di factStocks.csv ada di dimCompany.csv

Ambang batas: Tidak boleh ada keyCompany tidak dikenal.