

Dokumen Strategi Penjadwalan DAG

Deskripsi DAG

DAG ini bertanggung jawab untuk menjalankan pipeline ETL yang mengolah dan memuat data dari pasar saham Brazil ke dalam data warehouse berbasis DuckDB. Pipeline terdiri dari tiga task utama:

1. **run_full_pipeline**: menjalankan pipeline ETL penuh atau inisialisasi penuh (full load), meliputi ekstraksi data mentah, transformasi, dan load awal ke data warehouse.
2. **update_coin_data**: Memperbarui tabel dimensi untuk data koin dengan data baru (kolom `keyCoin` pada tabel `dimCoin`),
3. **update_stock_data**: Memperbarui tabel dimensi untuk data perusahaan saham dengan data baru (kolom `keyCompany` pada tabel `dimCompany`).

Strategi Penjadwalan dan Dependensi

Dataset berisi data historis saham yang tersedia dalam file CSV. Meskipun bersifat statis (tidak diupdate *real-time*), skenario yang dibangun mensimulasikan kondisi operasional nyata di mana:

- Data saham umumnya dirilis setelah jam pasar tutup
- Update harian dilakukan dari sumber eksternal
- Data historis dapat diperbarui jika ditemukan koreksi.

Sehingga strategi penjadwalan sebagai berikut:

- Initial Load (run_full_pipeline)
 - Frekuensi: Sekali saja saat awal deployment
 - Penjadwalan: Tidak perlu dijadwalkan otomatis, cukup dijalankan manual. Initial Load dipisahkan dari update harian menghindari duplikasi atau reprocessing yang tidak perlu.
- Daily Update Tasks (update_coin_data & update_stock_data)
 - Frekuensi: Sekali sehari (simulasi update real-time).
 - Waktu ideal: Tengah malam (00:00) waktu server.
 - Penjadwalan (schedule_interval): DAG dijadwalkan harian pada pukul 00:00 memungkinkan proses berjalan saat server dalam beban rendah, dan data untuk hari sebelumnya telah tersedia sepenuhnya.

Daily update task baru dapat dijalankan setelah run_full_pipeline berhasil terproses. Task update_coin_data dan update_stock_data dapat dijalankan secara parallel setelah run_full_pipeline berhasil, namun untuk menghindari proses write database DuckDB secara bersamaan (race condition), daily update task dibuat serial dengan update_coin_data yang dijalankan terlebih dahulu.