

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Sistem pemantauan harga pasar dan kebutuhan pokok merupakan salah satu sistem penting yang digunakan oleh Kementerian Perdagangan untuk mengawasi dan menganalisis perkembangan harga komoditas secara nasional. Sistem ini menghimpun data harga harian berbagai komoditas penting dari seluruh wilayah Indonesia. Data tersebut selanjutnya diolah guna mendukung analisis dan pengambilan keputusan oleh pemangku kebijakan. Namun, sistem ini menghasilkan volume data yang sangat besar dan kompleks, sehingga memerlukan solusi penyimpanan dan pengelolaan yang efisien.

Data berukuran besar yang tersebar secara terfragmentasi menyulitkan proses pengolahan dan analisis data secara efektif. Untuk mengatasi hal ini, diperlukan solusi penyimpanan terpusat berupa data *warehouse* yang berfungsi sebagai repositori terintegrasi [1]. Data *warehouse* secara khusus dirancang untuk mengonsolidasikan dan menyusun data transaksi dari berbagai sumber sistem operasional ke dalam struktur dimensional yang terstandarisasi [2].

Kimball mengatakan bahwa proses ETL (*Extract, Transform, Load*) merupakan komponen vital dalam membangun sistem data *warehouse* yang terintegrasi. Tahapan ini meliputi pengambilan data dari berbagai sumber yang beragam, termasuk sistem basis data seperti *SQL Server* dan *MySQL*, pembersihan data dari anomali dan ketidaksesuaian, serta konversi data ke dalam format yang seragam sebelum akhirnya disimpan dalam struktur dimensional yang terstandarisasi [3].

Dalam implementasinya, Kementerian Perdagangan menghadapi tantangan signifikan dalam sistem *business intelligence* mereka. Saat ini, instansi tersebut mengandalkan *Tableau* sebagai platform *dashboard* pemantauan harga, namun biaya lisensinya yang tinggi menjadi kendala utama. Untuk mengatasi hal ini, Kementerian berencana beralih ke *Apache Superset* sebagai solusi alternatif. Sebuah platform visualisasi data *open source* yang tidak hanya menyediakan tampilan data interaktif, tetapi juga menawarkan fleksibilitas tinggi dalam pengelolaan dan pengembangan *dashboard*.

Namun, dalam penggunaannya, *Apache Superset* melakukan *query* langsung ke *database*. Jika dihubungkan ke *database production*, hal ini tentu akan sangat membebani sistem yang sedang berjalan, terutama saat proses input data harga yang membutuhkan performa *real-time*. Oleh karena itu, diperlukan adanya *database warehouse* yang terpisah dari *database production* sebagai sumber data utama untuk *Superset*. Agar data *warehouse* tetap terbarui secara rutin, diperlukan *workflow ETL* yang terstruktur dan otomatis.

Penelitian ini akan mengimplementasikan *workflow ETL* menggunakan *Apache Airflow* dan *Python*. Solusi otomatisasi dengan *Apache Airflow* telah berhasil diimplementasikan oleh Wahyudi et al. untuk akuisisi data prediksi curah hujan BMKG [4]. *Airflow* memungkinkan penjadwalan ETL 2x sehari dengan *success rate* 98%, serta monitoring *real-time* melalui antarmuka DAG (*Directed Acyclic Graph*). *Python* dipilih karena merupakan bahasa pemrograman yang kaya akan pustaka untuk pengolahan data. Kemampuan *Python* dalam integrasi dengan berbagai *tools* dan *platform* (seperti *Apache Airflow*) menjadikannya pilihan utama untuk implementasi *workflow* data *mining* yang terstruktur [5]. Kombinasi keduanya memberikan solusi yang efisien, terstruktur, dan dapat diskalakan sesuai kebutuhan.

Penelitian ini diharapkan dapat membantu Kementerian Perdagangan dalam proses migrasi sistem *dashboard* ke platform yang lebih efisien dan fleksibel, serta memastikan bahwa kinerja sistem utama tidak terganggu. Selain itu, solusi ini juga memberikan kontribusi terhadap efisiensi biaya dan peningkatan ketersediaan informasi bagi seluruh pemangku kepentingan di berbagai wilayah Indonesia.

1.2. Identifikasi Masalah

Berdasarkan latar belakang di atas, maka dapat diidentifikasi beberapa permasalahan sebagai berikut:

1. Volume data harga kebutuhan pokok yang besar dan terfragmentasi menyulitkan proses pengolahan dan analisis data secara efektif.
2. Sistem saat ini masih bergantung pada *database production* untuk visualisasi data di *Apache Superset*, yang menyebabkan beban sistem meningkat dan mengganggu kinerja operasional.
3. Belum tersedia *workflow ETL* yang terstruktur dan otomatis untuk mendukung pembaruan data *warehouse* secara berkala dan terintegrasi.
4. Kebutuhan untuk beralih dari platform *dashboard* berbayar (*Tableau*) ke platform *open source* (*Apache Superset*) tanpa mengorbankan performa dan integritas data.

1.3. Rumusan Masalah

Berdasarkan identifikasi masalah, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana merancang dan mengimplementasikan *workflow ETL* yang terstruktur dan otomatis untuk mendukung sistem data *warehouse*?

2. Bagaimana *Apache Airflow* dan *Python* dapat digunakan untuk mengelola dan memantau proses ETL secara efisien?
3. Bagaimana integrasi data *warehouse* dengan *Apache Superset* dapat mengurangi beban *database production* tanpa mengganggu operasional input data?

1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Mengetahui tingkat kepuasan atau ketidakpuasan terhadap *dashboard Superset* yang menggunakan *database production (live)*.
2. Merancang dan mengimplementasikan *workflow ETL* yang terstruktur dan otomatis menggunakan *Apache Airflow* dan *Python*.
3. Membangun sistem data *warehouse* yang terpisah dari *database production* untuk mendukung kebutuhan analisis dan visualisasi data.
4. Mengintegrasikan data *warehouse* dengan *Apache Superset* untuk menyediakan *dashboard* interaktif tanpa membebani sistem produksi.

1.5. Ruang Lingkup

Penelitian ini memiliki ruang lingkup sebagai berikut:

1. Data yang digunakan terbatas pada data harga kebutuhan pokok dari *website Sistem Pemantauan Pasar dan Kebutuhan Pokok (SP2KP)*.
2. Implementasi *workflow ETL* dibatasi pada proses ekstraksi, transformasi dasar, dan pemuatan data ke dalam data *warehouse* berbasis *PostgreSQL*.
3. *Apache Airflow* digunakan sebagai alat utama untuk penjadwalan, pengelolaan, dan monitoring *workflow ETL*.

4. Integrasi dilakukan dengan *Apache Superset* sebagai platform visualisasi data.
5. Penelitian tidak membahas visualisasi data atau analisis data secara mendalam, melainkan fokus pada pembangunan *workflow ETL* dan integrasi data *warehouse*.

