

Nama : Kharisma Novi Chandramukti

Kelas/Batch : Data Engineering 4

Task 1

1. Jelaskan apa itu DBT !

JAWAB :

DBT (Data Build Tool) adalah sebuah alat open-source yang digunakan untuk transformasi data dalam proses analisis data (data warehouse) menggunakan bahasa SQL. DBT menggunakan proses ETL (Extract, Transform, Load) menjadi lebih terstruktur dan mudah dikelola, dan memungkinkan kolaborasi yang lebih baik antara tim analisis data dan tim engineering. Fungsi utama DBT adalah :

- a. Mengubah data mentah menjadi format yang lebih berguna untuk analisis
- b. Mengelola hubungan antar tabel dan transformasi
- c. Menyediakan dokumentasi yang jelas untuk model data
- d. Menguji kualitas data untuk memastikan konsistensi dan keakuratan

2. Apa keuntungan menggunakan DBT ?

JAWAB :

Keuntungan menggunakan DBT adalah :

a. Keterbacaan dan Pemeliharaan Kode

DBT menggunakan SQL yang memungkinkan pengguna untuk menulis transformasi data dengan cara yang mudah dibaca dan dipahami

b. Version Control

DBT memungkinkan integrasi dengan system version control seperti git, sehingga perubahan pada model data dapat dilacak dan dikelola dengan lebih baik

c. Modularitas dan Reusibilitas

DBT dalam pembuatan model data yang modular yang dapat digunakan kembali, sehingga mengurangi duplikasi dan meningkatkan efisiensi pengembangan

d. Dokumentasi Otomatis

DBT dapat menghasilkan dokumentasi otomatis dari model data, yang membantu tim memahami struktur dan hubungan data dengan lebih baik

e. Pengujian Kualitas Data

DBT menyediakan kemampuan untuk menulis dan menjalankan tes untuk memverifikasi kualitas data, mengurangi resiko kesalahan dalam data analisis

f. Integrasi dengan Platform Data Modern

DBT terintegrasi dengan berbagai platform data seperti Snowflake, BigQuery, dan Redshift yang memungkinkan pengguna untuk bekerja dalam ekosistem yang sudah ada

3. Jelaskan dependency tree dan versioning pada DBT !

JAWAB :

- a. Dependency Tree adalah struktur yang menunjukkan bagaimana model – model data saling berhubungan dan tergantung satu sama lain. Dalam DBT merujuk pada cara transformasi data diatur berdasarkan urutan eksekusi dan ketergantungan antar model
- b. Versioning merujuk pada pengelolaan versi dari model, skrip, dan konfigurasi DBT. Versioning digunakan untuk melacak perubahan yang dibuat seiring waktu dan memastikan bahwa tim dapat mengelola berbagai versi dari model yang sama dengan mudah