

CAPSTONE PROJECT MODULE 2

GENERAL INFORMATION

Pada Modul 2 siswa telah mempelajari Database dan SQL, beberapa materi yang telah dipelajari di antaranya adalah :

- PostgreSQL dan datamart concept.
- Insert data into postgresql.
- SQL Basic
- Data Manipulation
- Data Filtering and Aggregating
- SQL Joins dan Subqueri
- SQL Views and CTE's
- SQL Window Functions
- SQL Ranking and Filtering

Pengerjaan project ini akan melatih siswa untuk lebih siap dan lebih memahami bahasa pemrograman Python dan SQL function ketika diaplikasikan ke dalam bidang pekerjaan Data Engineering.

Setiap siswa akan mengerjakan semua case study, dengan tahapan pengerjaan sebagai berikut :

1. Download Guideline Module 2 guna mengetahui apa yang harus dikerjakan oleh siswa.
2. Download folder **master_capstone2** dan **capstone_postgress** dari Gdrive folder berikut :

<https://drive.google.com/drive/folders/1DMZ1R76DJOeXoHrHb5F7P3awEY1sA3tM>

Setelah download folders, lakukan beberapa hal berikut :

- a. Open aplikasi docker dari local computer Anda, jalankan docker-compose.yaml yang tersedia di folder capstone_postgress.
 - b. Open aplikasi VSCode, buat koneksi baru menggunakan postgres, sesuaikan nama user, database, dan password yang ada pada docker-compose.yaml pada koneksi database di DBeaver
3. Setelah database berhasil terkoneksi di DBeaver, Gunakan sumber data dari GDrive Soal no 2 dan kerjakan soal-soal berikut menggunakan bahasa SQL :
 - a. Hitunglah ada berapa banyak transaksi di table Order?
Aggregate COUNT pada kolom Id.
 - b. Hitunglah berapa jumlah Order dari seluruh transaksi di Table Order?
Aggragate SUM pada kolom Total.
 - c. Hitung 10 product yang sering memberikan discount di tabel Order berdasarkan produk Title dari table Products?

Join table Order dengan table produk untuk mendapatkan produk Title, aggregate Count untuk mendapatkan total order dan grouping berdasarkan kolom Title dari Table Products. Urutkan dari total transaksi tertinggi - terendah. Tampilkan 10 product teratas.

- d. Hitung berapa jumlah Order dari transaksi di Table Order, berdasarkan kolom category dari tabel produk?

Gunakan Konsep CTE, Join table Order dengan table Product, Hitung total menggunakan aggregate SUM, dan grouping berdasarkan kolom Category dari table Product. Urutkan dari total transaksi tertinggi - terendah.

- e. Hitung berapa jumlah total Order di table Order, dari setiap title di table Product yang memiliki rating ≥ 4 ?

Gunakan Konsep CTE, Join table Order dengan table Product, Hitung total menggunakan aggregate SUM, dan grouping berdasarkan kolom Title dari table Product. Urutkan dari total transaksi tertinggi - terendah.

- f. Dapatkan list reviews berdasarkan kategori produk = 'Doohickey', dimana rating dari table reviews ≤ 3 ? Dan urutkan berdasarkan table created_at di table reviews.

Gunakan Konsep CTE, Join table Reviews dengan table Product, select kolom created_at, body dan rating dari table reviews, dan filter product category dan rating reviews ≤ 3 . Urutkan dari kolom created terbaru - terlama.

- g. Ada berapa source di table Users?

Ambil unik data di kolom source dari table Users.

- h. Hitung total user di table Users yang memiliki email dari gmail.com.

Rename kolom alias total_user_gmail.

- i. Dapatkan list id, title, price, and created at dari table Products, dengan kriteria price antara 30 - 50, dan urutkan dari transaksi terbaru - terlama?

- j. Dapatkan list Name, Email, Address, Birtdate dari table Users yang lahir diatas tahun 1997? Buatlah dalam format database views.

- k. Dapatkan list id, created_at, title, category, dan vendor dari table products, yang memiliki title yang sama / title muncul dengan value yang sama sebanyak lebih dari 1 kali.

Gunakan format cte dan row_number dalam mengerjakan soal berikut.

4. Gunakan data source [NY Taxi Trip](#) yang ada di Capstone 1. Kemudian kerjakan proyek data berikut :

- Pilihlah salah satu data source, seperti : Yellow Taxi Trip Records, Green Taxi Trip Records, dan lainnya.
- Buatlah skrip Python untuk ekstrak data, kemudian insert data ke dalam PostgreSQL.
- Untuk proses ekstraksi data, setiap 1 kali program running, gunakan proses filter data per hari / per minggu / perbulan.

- Kemudian data hasil ekstraksi, masukkan / insert data ke dalam postgresql. Gunakan skema insert data / incremental proses. Data baru selalu di append / ditambahkan ke dalam tabel. Pastikan tidak ada hasil data yang duplicate.
 - Proses filter data berdasarkan rantang waktu bisa di set manual / otomatis.
 - Buatlah index kolom, dan tentukan kolom mana saja yang ingin di indexing. Buat dengan fungsi python.
 - Buatlah partition tabel jika data tabel memiliki kolom data dengan format date, datetime, dan timestamp.
 - Sediakan error handling dan validasi data dari setiap step / blok kode.
 - Buatlah 3 query SQL untuk agregasi data dengan konsep window function, CTE, dan aggregate. Jelaskan temuan anda.
5. Untuk soal pertama, berikan catatan di setiap blok kode query jika diperlukan. Dan juga pada setiap fungsi sql diusahakan semua query fungsi menggunakan huruf Capital / CapsLock.
 6. Untuk soal kedua, buatlah fungsi / catetan fungsi / catatan blok kode yang di perlukan dengan format file python (.py) / Jupyter Notebook (.ipynb).
 7. Buatlah dokumentasi hasil dari project yang anda kerjakan dengan format Google Document / lainnya.
 8. Buatlah video dokumentasi / penjelasan hasil capstone project, dengan maksimal durasi selama 30 menit. Manfaatkan file dokumentasi google docs / di sesuaikan dengan cara and menjelaskan project.

Setelah mengikut tahapan pengerjaan, setiap tahapan akan memiliki bobot penilaian, diantaranya sebagai berikut :

Poin Penilaian

- Video Penjelasan : Maksimal 20 Point
 - Kelengkapan project.
 - Google Document
 - Penjelasan skrip SQL dan python, serta hasil pengerjaan project.
- Hasil Proyek Studi Kasus no 3 / proyek pertama: Maksimal 60 Point
 - Kerjakan setiap soal query SQL
- Hasil Proyek Studi Kasus no4 / proyek kedua : Maksimal 20 Point

Case Study

- Data Master Capstone 2
- NY Taxi Trip

Waktu Pengerjaan

Lama waktu pengerjaan Capstone Project Module 2 adalah **17 hari kerja**. Pengerjaan akan dihitung sejak H+1 setelah pengumuman Guidline Project.

Contoh :

- Pengumuman capstone tanggal 30 Oktober 2025, Hari Kamis.
- Pengerjaan mulai dari 31 Oktober 2025, Hari Jumat.
- Pengumpulan maksimal 24 November 2025 23:59:59, Hari Senin.

Metode Pengumpulan

Pengumpulan dilakukan dengan cara:

- Unggah video penjelasan project ke dalam cloud storage (Youtube, Google Drive, Dropbox) masing-masing siswa. Buka hak akses untuk publik. Kirim ke email Mentor : Samsudiney@gmail.com dan melalui google form yang akan di sediakan oleh team purwadika.

- Video penjelasan maksimal berdurasi 30 menit dan wajib mengaktifkan kamera depan atau webcam, sehingga wajah siswa ada dalam rekaman video.

- Mengisi Google Forms yang telah disediakan oleh operasional untuk mencantumkan link video, link google document / cloud storage / link github.

- Pastikan siswa menerima email konfirmasi bahwa siswa telah sukses melakukan pengisian dan pengumpulan Google Forms Capstone Project Module 2 yang dikirim secara otomatis oleh sistem. Cek folder spam apabila e-mail tidak ada di folder inbox.

Catatan

- Jika siswa mengumpulkan Capstone Project Module 2 melewati tenggat waktu yang sudah ditentukan, maka akan ada pengurangan poin untuk nilai akhir sebagai berikut:

- Telat 1 detik sampai 24 jam: nilai akhir dikurangi 10 poin
- Telat 24 jam sampai 72 jam: nilai akhir dikurangi 20 poin
- Telat lebih dari 72 jam: nilai akhir menjadi 0

- Segala bentuk plagiarisme tidak akan ditoleransi dan mutlak diberikan nilai 0.

- Di larang plagiat dan copy paste secara saklek dari AI. artinya pahami blok kode nya, dan jelaskan sesuai pemahaman anda.