Nama : Syarifa Ulya Fidyana

Kelas : H

NIM : 40011423650293

**PERTEMUAN 4 SISTEM INFORMASI AKUNTANSI**

**TUGAS AKHIR SIA**

**PETUNJUK**

Anda adalah model AI yang mengkhususkan diri dalam menghasilkan dan mengoptimalkan kode SQL untuk tugas analitik data di Google BigQuery. Peran Anda adalah membantu pengguna dalam membuat kueri SQL yang efisien, dapat diskalakan, dan dapat dipelihara yang disesuaikan dengan platform BigQuery.

Ikuti petunjuk terperinci ini untuk memenuhi persyaratan pengguna:

1. Pahami kebutuhan pengguna:
   * Mengklarifikasi tujuan analitik spesifik (misalnya, meringkas data, analisis tren, analitik prediktif).
   * Tentukan set data yang dikueri (misalnya, nama tabel, skema, lokasi data di BigQuery).
   * Identifikasi filter, agregasi, gabungan, atau fungsi jendela tertentu yang diperlukan.
   * Periksa preferensi seperti pengoptimalan biaya, penyetelan kinerja, atau fitur SQL tertentu.
2. Hasilkan Kode SQL:
   * Tulis kueri SQL yang bersih dan terstruktur dengan baik yang dioptimalkan untuk arsitektur BigQuery.
   * Gabungkan praktik terbaik seperti:
   * Operasi JOIN yang efisien: Gunakan jenis dan kunci JOIN yang sesuai.
   * Partisi dan Pengklusteran: Manfaatkan partisi dan pengklusteran BigQuery untuk kueri yang lebih cepat.
   * Klausa WITH: Untuk keterbacaan dan modularitas, gunakan Ekspresi Tabel Umum (CTE) untuk subkueri.
   * Gunakan alias yang jelas dan konsisten untuk tabel dan bidang.
   * Menangani nilai NULL secara efektif dalam kueri.
3. Optimalkan untuk BigQuery:
   * Minimalkan biaya kueri dengan:
   * Memilih hanya kolom yang diperlukan (SELECT bidang tertentu, bukan SELECT \*).
   * Memfilter data di awal kueri untuk mengurangi byte yang dipindai (klausa WHERE dan LIMIT).
   * Memanfaatkan ARRAY, STRUCT, atau bidang berlapis untuk data hierarkis.
   * Terapkan strategi caching jika berlaku (misalnya, menggunakan tampilan materialisasi atau tabel sementara).
   * Berikan petunjuk atau penjelasan eksekusi untuk kueri yang kompleks.
4. Gabungkan Fitur Analitik:
   * Gunakan fungsi dan alat SQL lanjutan BigQuery, seperti:
   * Fungsi Analitik: Untuk peringkat, jumlah bergulir, rata-rata, dll. (misalnya, ROW\_NUMBER(), OVER(), LAG()).
   * Analisis Geospasial: Manfaatkan fungsi geografi BigQuery jika data spasial terlibat.
   * Machine Learning: Integrasikan dengan BigQuery ML untuk analitik prediktif.
   * Fungsi JSON: Ekstrak dan analisis data JSON jika ada.
5. Validasi dan Pengujian:
   * Berikan contoh input dan output untuk verifikasi.
   * Sertakan kueri untuk memvalidasi asumsi (misalnya, memeriksa distribusi data atau outlier).
   * Sarankan penggunaan EXPLAIN atau QUERY PLAN BigQuery untuk menilai performa.
6. Dukungan dan Dokumentasi:
   * Sertakan komentar sebaris yang menjelaskan setiap komponen kueri.
   * Tawarkan saran tentang mengelola kumpulan data (misalnya, kedaluwarsa tabel, format penyimpanan).
   * Tautkan ke dokumentasi BigQuery yang relevan untuk fungsi atau fitur lanjutan yang digunakan dalam kueri.

Berdasarkan petunjuk dokumen diatas, dan dokumen tabel yang tersedia:

Tolong bantu saya dalam menggunakan BigQuery untuk menganalisis data penjualan dan produksi produk dari dua tabel yang telah saya unggah, yaitu tabel\_produksi dan tabel\_penjualan, yang ada dalam dataset perusahaan di project saya bernama ulya-project. Saya ingin mencari id\_produk dengan jumlah penjualan tertinggi selama bulan Januari tahun 2024. Kriteria lainnya adalah id\_produksi nya bernomor ganjil.

Catatan untuk membantu anda mengerjakan:

• Kolom id\_produk, id\_produksi, dan tanggal produksi ada di tabel\_produksi,

• Kolom jumlah\_terjual ada di tabel\_penjualan

• Untuk cari penjualan terbesar: jumlah\_terjual nya bisa diurutkan saja dari yang terbesar ke terkecil

• Kolom yang dipakai hanya id\_produk, id\_produksi, tanggal\_produksi, dan jumlah terjual saja

• Untuk menghubungkan 2 tabel tersebut menggunakan foreign key, foreign key nya yaitu kolom id\_produk

Dapatkah Anda memberikan query SQL khusus BigQuery untuk menjalankan analisis ini?

**Narasi Penjelasan Hasil Kueri SQL**

Kueri SQL yang telah disusun bertujuan untuk menganalisis data penjualan dan produksi produk dari dua tabel yang terdapat dalam dataset perusahaan di proyek **ulya-project**. Fokus analisis ini adalah untuk mengidentifikasi produk dengan jumlah penjualan tertinggi selama bulan Januari tahun 2024, dengan kriteria tambahan bahwa produk tersebut harus memiliki **id\_produksi** yang bernomor ganjil.

**Proses Analisis**

1. **Pengumpulan Data Penjualan**:
   * Kueri pertama, yang didefinisikan dalam Common Table Expression (CTE) bernama **penjualan\_terfilter**, memfilter data dari tabel **tabel\_penjualan**. Dalam langkah ini, kami mengambil semua penjualan yang terjadi antara tanggal 1 Januari 2024 hingga 31 Januari 2024.
   * Untuk setiap **id\_produk**, kami menghitung total jumlah terjual (**jumlah\_terjual**) dengan menggunakan fungsi agregasi **SUM()**. Hasilnya adalah tabel yang menunjukkan total penjualan untuk setiap produk selama bulan tersebut.
2. **Pengumpulan Data Produksi**:
   * Kueri kedua, yang didefinisikan dalam CTE bernama **produksi\_terfilter**, memfilter data dari tabel **tabel\_produksi**. Di sini, kami hanya memilih produk yang memiliki **id\_produksi** bernomor ganjil, yang dilakukan dengan menggunakan fungsi **MOD()** untuk memeriksa sisa bagi.
   * Hasil dari CTE ini adalah tabel yang berisi **id\_produk**, **id\_produksi**, dan **tanggal\_produksi** untuk produk yang memenuhi kriteria.
3. **Penggabungan Data**:
   * Setelah kedua CTE disiapkan, kami melakukan penggabungan (join) antara **produksi\_terfilter** dan **penjualan\_terfilter** berdasarkan kolom **id\_produk**. Kami menggunakan **LEFT JOIN** untuk memastikan bahwa semua produk dari **produksi\_terfilter** ditampilkan, bahkan jika tidak ada penjualan yang tercatat.
   * Dalam hasil akhir, kami juga menggunakan fungsi **COALESCE()** untuk menangani kemungkinan nilai NULL pada total penjualan, sehingga jika tidak ada penjualan untuk produk tertentu, nilai total terjual akan ditampilkan sebagai 0.
4. **Pengurutan Hasil**:
   * Hasil akhir dari kueri diurutkan berdasarkan total penjualan (**total\_terjual**) dari yang tertinggi ke terendah. Ini memungkinkan kita untuk dengan mudah mengidentifikasi produk mana yang memiliki penjualan tertinggi selama periode yang ditentukan.

**Hasil yang Diharapkan**

Hasil dari kueri ini akan memberikan daftar produk yang memenuhi kriteria yang telah ditetapkan, lengkap dengan informasi mengenai **id\_produk**, **id\_produksi**, **tanggal\_produksi**, dan total penjualan. Dengan informasi ini, manajemen dapat mengambil keputusan yang lebih baik terkait strategi penjualan dan produksi, serta mengidentifikasi produk yang paling laris di pasar.

WITH penjualan\_terfilter AS (

    -- Memfilter tabel\_penjualan untuk mendapatkan data penjualan selama bulan Januari 2024

    SELECT

        produk\_nama,

        SUM(jumlah\_terjual) AS total\_penjualan

    FROM

        `ulya-project.COMPANY\_ulya.tabel penjualan`

    WHERE

        tanggal\_penjualan BETWEEN '2024-01-01' AND '2024-01-31'  -- Filter tanggal

    GROUP BY

        produk\_nama

),

produksi\_terfilter AS (

    -- Memfilter tabel produksi untuk mendapatkan produksi\_id yang bernomor ganjil

    SELECT

        produk\_nama,

        produksi\_id,

        tanggal\_produksi

    FROM

        `ulya-project.COMPANY\_ulya.tabel produksi`

    WHERE

        MOD(produksi\_id, 2) = 1  -- Memastikan produksi\_id adalah ganjil

)

-- Menggabungkan hasil dari kedua tabel berdasarkan produk\_nama

SELECT

    p.produk\_nama,

    p.produksi\_id,

    p.tanggal\_produksi,

    COALESCE(j.total\_penjualan, 0) AS total\_penjualan  -- Menggunakan COALESCE untuk menangani nilai NULL

FROM

    produksi\_terfilter p

LEFT JOIN

    penjualan\_terfilter j ON p.produk\_nama = j.produk\_nama  -- Menghubungkan tabel berdasarkan produk\_nama

ORDER BY

    j.total\_penjualan DESC  -- Mengurutkan berdasarkan total penjualan tertinggi

