**Latihan Data Warehouse**

**Nama : Mario Orlando**

**NIM : 2602078291**

1. Berdasarkan dengan tabel yang sudah dibuat pada Latihan SSIS, Jelaskan mengenai dengan poin berikut:
   1. Nama tabel Fact yang sudah dibuat di Data Warehouse

Jawab:

**factNetflix** 🡪 Agar konsisten, maka penamaan tabel menggunakan penamaan **camelCase**, yang **mengkapitalisasi huruf pertama setiap kata tanpa spasi/tanda baca**. Mengapa? Karena **meningkatkan tingkat keterbacaan** sehingga memudahkan pengembang untuk memahami skema yang ada. Penamaan camelCase juga umumnya didasarkan pada **standar sebuah perusahaan/organisasi**.

* 1. List kolom apa saja yang terdapat pada tabel fact yang sudah dibuat

Jawab:

**factId, showId, totalMovies, totalTvShow, averageDurationMinutes, averageDurationSeasons, averageReleaseYear, averageDateAdded, averageCountry, averageGenre** 🡪 Berisikan **data yang ingin diperiksa** dan yang memiliki ***value* numerikal**. Salah satu karakteristik *fact table* adalah **memiliki banyak *foreign key*** untuk menghubungkan ke *dimension table*.

* 1. Tipe data yang terdapat pada setiap kolom

Jawab:

**factId INT PRIMARY KEY 🡪** *Unique key* dalam *table* factNetflix

**showId INT FOREIGN KEY 🡪** Menunjuk ke *dimension table* lainnya

**totalMovies INT 🡪** Menghitung jumlah *movies* yang ada

**totalTvShow INT 🡪** Menghitung jumlah TV *show* yang ada

**averageDurationMinutes INT 🡪** Menghitung rata-rata durasi *movies*

**averageDurationSeasons INT 🡪** Menghitung rata-rata durasi TV *show* setiap *season*-nya

**averageReleaseYear INT 🡪** Menghitung rata-rata tahun perilisan

**averageDateAdded INT 🡪** Menghitung rata-rata tanggal perilisan

**averageCountry INT 🡪** Menghitung rata-rata negara terbanyak merilis *movies*/TV *show*

**averageGenre INT 🡪** Menghitung rata-rata *genre* yang ada

* 1. Gambar Table Fact yang sudah dibuat

Jawab:

**A screenshot of a computer

Description automatically generated**

1. Kalian merupakan seorang data Engineer yang sedang bekerja di Perusahaan Netflix, dan tabel yang sudah kalian buat akan dipakai oleh dashboard yang digunakan oleh Board of Director Perusahaan. Berikut adalah kondisi mengenai penggunaan dashboard:
   1. Dashboard akan dipakai untuk keperluan meeting monthly. Namun anggota Board of Director kerap melihat dashboard setiap harinya untuk melihat update dari list film yang ditayangkan. Berdasarkan kondisi tersebut, jelaskan frekuensi refresh data yang akan digunakan pada tabel fact yang sudah dibuat beserta dengan alasannya.

Jawab:

Menurut saya, mengingat anggota *Board of Director* sering memeriksa *dashboard* setiap hari, maka lebih baik *refresh data* dilakukan per-hari karena dalam Netflix, konten dapat berubah setiap hari agar dapat menyajikan informasi yang paling *up to date*. Tetapi, apabila data dalam *table* memiliki data yang kompleks atau membutuhkan waktu pemrosesan yang lama, maka sebaiknya frekuensi *refresh data* per-bulan, karena kalau dilakukan *refresh data* per-hari, hal tersebut dapat membebani sumber daya sistem dan menyebabkan penurunan kinerja. Sebenarnya, kita harus **mempertimbangkan dalam menentukan frekuensi *refresh data* dari segi kompleksitas data, performa, dan manfaat yang akan didapatkan**. Tidak menutup kemungkinan juga kalau *refresh data* dilakukan per-minggu sehingga data lebih *up to date*, dan performa akan jauh lebih baik daripada per-hari.

* 1. Setelah perundingan dengan tim aplikasi, diketahui bahwa sumber data akan ditambahkan status audit untuk menandai bahwa pada row data tersebut terjadi perubahan. Berdasarkan dengan informasi tersebut. Tentukan apakah metode yang dipakai pada saat transfer data menggunakan Incremental Load atau Full Refresh beserta dengan alasannya.

Jawab:

Menurut saya, ***incremental load* lebih cocok** daripada *full refresh*. Mengapa? Metode *incremental load* cocok digunakan ketika kita hanya perlu memperbarui **data sebagian** (dari soal hanya mau memperbarui ***row***, yang berarti tidak mempengaruhi data yang lain) karena lebih efisien (mengurangi beban pada sistem) dan pembaruan yang dilakukan lebih cepat. Metode *full refresh* sebaiknya digunakan apabila data yang mau diubah sangat kompleks atau data mengalami perubahan yang signifikan. Namun, kita memang harus **mempertimbangkan lagi berdasarkan kompleksitas data, waktu yang dibutuhkan, dan sumber daya yang tersedia**. Jika, ketiga hal tersebut terpenuhi, *full refresh* dapat dipertimbangkan lagi.

1. Perusahaan Netflix mempunyai 3 Business Unit yang berbeda, Divisi Operasional yang berhubungan dengan penyiaran film, Divisi Sales yang berhubungan dengan penjualan subscription, dan Divisi Legal yang mengurusi lisensi dari film yang beredar. Setiap divisi mempunyai data mart yang berisi tabel untuk menunjang kebutuhan tiap divisi. Dari kondisi tersebut, tentukan:
   1. Pada Data Mart mana tabel yang kalian buat berada, beserta dengan alasannya

Jawab:

Pada *data mart* divisi operasional karena pada divisi operasional **berhubungan dengan penyiaran film**, sama seperti data SSIS yang sudah dibuat, yaitu tentang penayangan Netflix. Isinya terdapat banyak tabel, seperti netflix\_titles, netflix\_titles\_directors, netflix\_titles\_countries, netflix\_titles\_cast, dan netflix\_titles\_category.

* 1. Selain tabel yang sudah kalian buat, terdapat dua tabel lain Bernama FactSubscriptionSales dan FactMovieLicense. Tentukan pada data mart mana kedua tabel tersebut berada

Jawab:

factSubscriptionSales 🡪 *data mart* divisi *sales* karena **berhubungan dengan penjualan subscription**

factMovieLicense 🡪 *data mart* divisi *legal* karena **berhubungan dengan lisensi dari film yang beredar**

* 1. Buatlah Flow data dari Aplikasi sampai ke Business Unit.

Jawab:

A diagram of data storage

Description automatically generated