Latihan Data Warehouse

Notes & Pre requisite:

1. Berdasarkan dengan Penamaan nama tabel di Binus, Setiap tabel yang dibuat harus menggunakan **Camel Case**
2. Untuk pembuatan diagram tabel, bisa menggunakan Draw.IO dan Microsoft Visio.
3. Berdasarkan dengan tabel yang sudah dibuat pada Latihan SSIS, Jelaskan mengenai dengan poin berikut:
   1. **Nama** tabel Fact yang sudah dibuat di Data Warehouse

Netflix Chronicles: The 2015-2019 Surge in Film & TV Show Productions

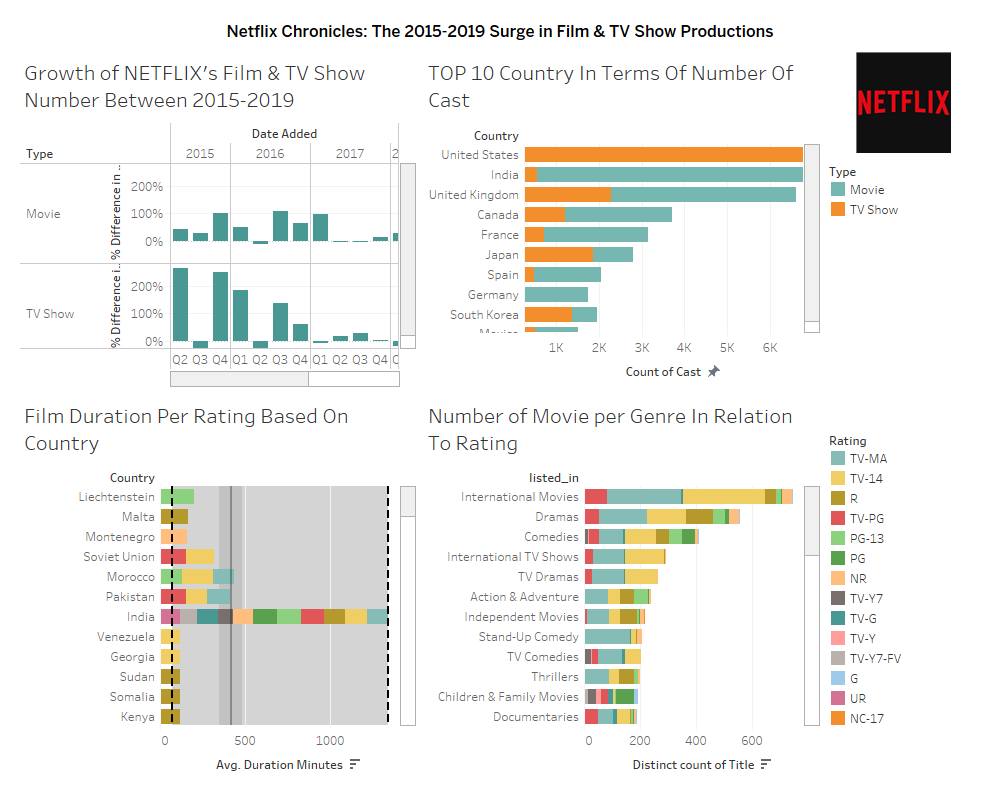
* 1. List **kolom** apa saja yang terdapat pada tabel fact yang sudah dibuat

Pada tabel fact yang sudah dibuat menggunakan SSIS, teradapat 13 kolom dengan atribut masing-masing. Kolom-kolom tersebut berupa:

* duration\_seasons
* type
* title
* date\_added
* release\_year
* rating
* description
* show\_id
* cast
* listed\_in
* country
* director
  1. **Tipe data** yang terdapat pada setiap kolom

Tipe data yang digunakan pada tiap kolom adalah sebagai berikut:

* duration\_seasons: float
* type: varchar
* title: varchar
* date\_added: datetime
* release\_year: float
* rating: varchar
* description: varchar
* show\_id: float
* cast: varchar
* listed\_in: varchar
* country: varchar
* director: varchar
  1. **Gambar** Table Fact yang sudah dibuat

****

Soal Case:

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

Refresh data secara *daily* dapat dilakukan untuk memastikan bahwa data list film yang ditayangkan selalu terbaru. Alternatifnya, kita bisa melakukan suatu kondisi untuk mengecek apabila hari tersebut tidak ada pergerakan dalam list film (baik berkurang, bertambah, atau update) maka tampilkan di dashboard bahwa tidak ada pergerakan film pada hari itu. Daily refresh dapat dilakukan secara otomatis dengan memanfaatkan tools seperti cron job atau dashboard scheduler lainnya.

* 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

Metode yang cocok untuk digunakan pada saat transfer data dalam keperluan menambahkan status audit adalah Incremental Load. Hal tersebut disebabkan karena Incremental Load lebih efisien dan tidak mengganggu data yang lain sebab tidak memerlukan pembaruan data secara keseluruhan (Full Refresh). Penambahan status audit memungkinkan proses transfer data dengan hanya memuat data yang berubah saja, sehingga menjadi lebih efisien, cepat, dan tidak membutuhkan *resource* berlebih.

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

Tabel yang saya buat dapat dikategorikan menjadi data mart Divisi Sales. Hal tersebut disebabkan karena data yang saya tampilkan di dashboard memberikan insight seputar

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

FactSubscriptionSales dapat diklasifikasikan pada data mart bagian Divisi Sales karena berasal dari namanya, tabel tersebut kemungkinan membahas data seputar subscription Netflix. FactMovieLicense bisa diklasifikasikan pada bagian Divisi Legal, sebab kemungkinan membahas seputar lisensi setiap film yang ada di Netflix.

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

A black screen with white text

Description automatically generated