IMDB DATA WAREHOUSE

Design and Implementation of a Data Warehouse for Online Movie Database

Abstract

A concise data warehouse solution to fulfil the business intelligence needs of IMDB

Group 4

ISYS 637-603

Contents

1. Executive Summary

The entertainment industry is part of the technological advancements, which has transformed the way business is being conducted. Data driven decisions give great insights into the business viability. Implementation of Data Warehouse can help businesses utilize data to make decisions about future movie productions.

Our Data Warehouse leverages the data residing in the IMDB databases and other transactional data to find out the movie performance. This data can be used for Business Intelligence requirements, for e.g. to find which is the most profitable genre, which Actor- Director pair works best for various genres etc. A data warehouse gets data from various sources and once the collation and analyses of such data has been performed, it can be used by production houses.

This report outlines the process of creating a data warehouse from scratch using SSIS, SSAS and SSRS. Sample reports that satisfy Business Intelligence questions are shown as well as the method for creating customizable reports.

The Business Intelligence reports obtained from implementing this Data Warehouse would ultimately increase revenue by leveraging previous movie data. Using a consolidated Data Warehouse instead of separately housed data sources greatly improve the efficiency of creating Business Intelligence reports and aid the business in delivering top performance.

2. About IMDB

The Internet Movie Database (abbreviated IMDb) is an online database of information related to films, television programs and video games, including cast, production crew, fictional characters, biographies, plot summaries, trivia and reviews, operated by IMDb.com, Inc., a subsidiary of Amazon.

IMDb originated with a Usenet posting by British film fan and computer programmer Col Needham entitled "Those Eyes", about actors with beautiful eyes. Others with similar interests soon responded with additions or different lists of their own. Needham subsequently started an "Actors List", while Dave Knight began a "Directors List", and Andy Krieg took over "THE LIST" from Hank Driskill, which would later be renamed the "Actress List". Both lists had been restricted to people who were alive and working, but soon retired people were added, so Needham started what was then (but did not remain) a separate "Dead Actors/Actresses List". The goal of the participants now was to make the lists as inclusive as possible.

By late 1990, the lists included almost 10,000 movies and television series correlated with actors and actresses appearing therein. On October 17, 1990, Needham developed and posted a collection of Unix shell scripts which could be used to search the four lists, and thus the database that would become the IMDb was born. At the time, it was known as the "rec.arts.movies movie database".

3. Why is Data Warehouse Needed?

A goal of every business is to make better business decisions than their competitors. That is where business intelligence (BI) comes in. BI turns the massive amount of data from operational systems into a format that is easy to understand, current, and correct so decisions can be made on the data.

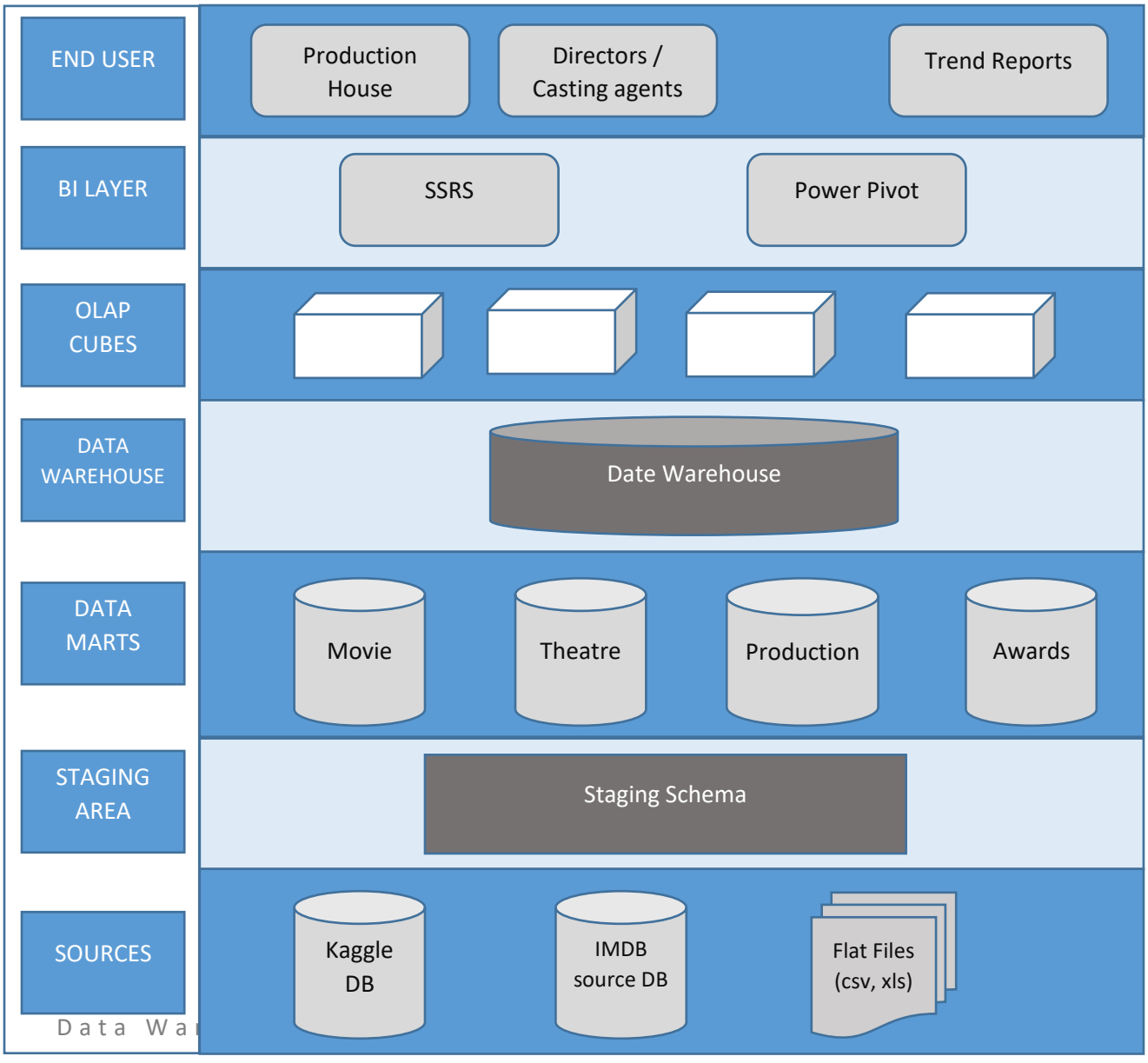
The idea is to create a permanent storage space for the data needed to support reporting, analysis, and other BI functions. While it may seem wasteful to store data in multiple places (source systems and the data warehouse), the many advantages of doing that more than justify the effort and expense.

Data warehouses reside on servers dedicated to this function running a database management system (DBMS) such as SQL Server and using Extract, Transform, and Load (ETL) software

such as SQL Server Integration Services (SSIS) to pull data from the source systems and into the data warehouse.

In respect to IMDB, it is usually considered that there is no universal way to claim the goodness of movies. Many people rely on critics to gauge the quality of a film, while others use their instincts. However, it takes the time to obtain a reasonable amount of critics review after a movie is released. Moreover, human instinct sometimes is unreliable. Thus, a data warehouse solution can be used to derive facts over word of mouth. The data collection reflects lists of movies and associated information. The amount of data stored in the database can be used to answer multiple business intelligence question such as performance of a movie, theatres and the production houses. The data can also be used as a basis to decide on a new project.

4. Data Warehouse Architecture



4.1 Star Schema

Multiple data marts are created to answer the different business questions. This section gives an overview of the data-marts created and the fact, dimension tables used to create them along with their relationship. All the data marts created follow start schema.

Also, note: All attributes of the dimensions are of SCD type 1 (contents are overwritten on change) except for the ones denoted as SCD 2. These attributes need the data to be preserved even after they change. This is handled by use of start_date and end_date attributes in the appropriate dimensions. If end_date is null then no change has happened whereas, if end_date is populated then it denotes that specific entry as old (not current) and end date also represents the effective date of the change just the way start_date denotes the day when entry happened.

- Awards

  The data mart for analyzing the winners of different awards is shown in the below figure. It contains:

  - Awards Fact table
  - Actor, Movie and Director Dimension tables

  The grain of the fact table is an Award. A transactional entry happens every time an award is announced / awarded.

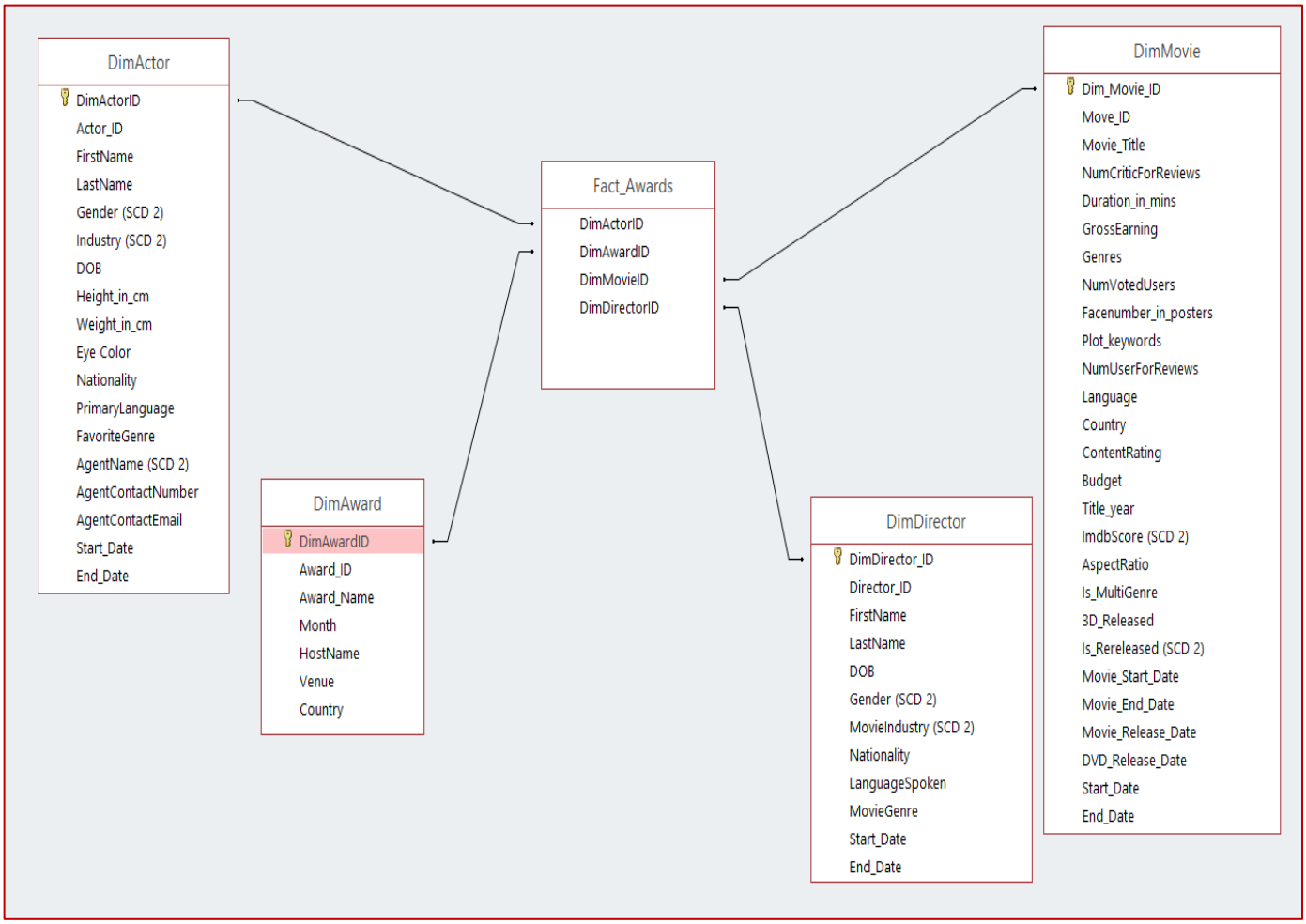


Figure: Awards Data mart

- Movie Performance

  The data mart for analyzing a movie's performance is shown in the below figure. It contains:

  - Movie Performance Fact table
  - Date, Movie, Actor and Director Dimension tables

The grain of the fact table is a Movie. A transactional entry happens every time a movie is release and at the end of each week.

Here, Role Playing is used to represent actor in different views. Actor dimension plays the roles of LeadActor and SupportingActor in the fact table.

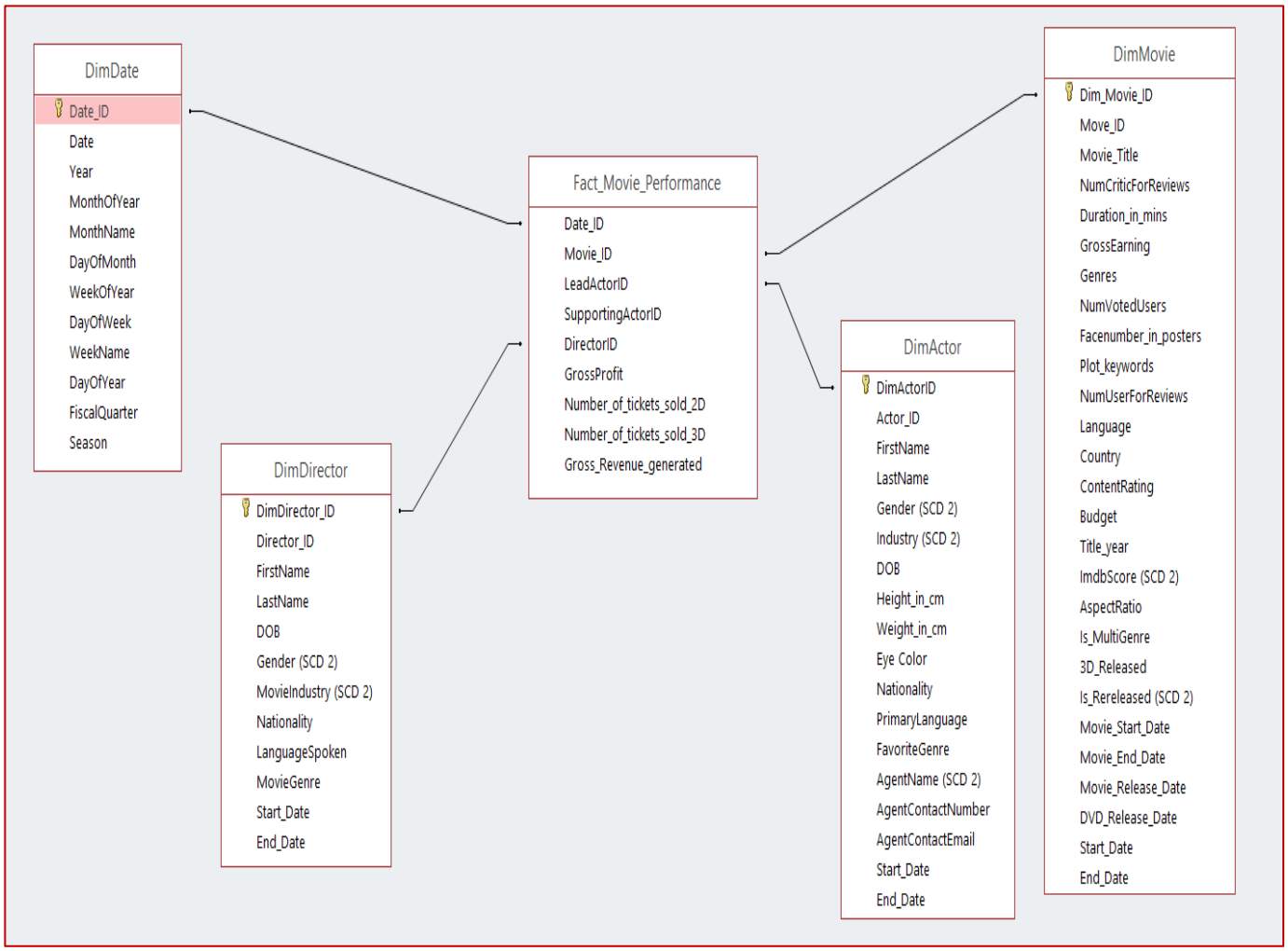


Figure: Movie Performance data mart

- Movie Production

  The data mart for analyzing the performance of a movie production-house is shown in the below figure. It contains:

- Movie Production Fact table
- Date, Movie, Actor and Production House Dimension tables

The grain of the fact table is a Movie. A transactional entry happens every time a movie is release and weekly from that point on.

Here, Role Playing is used to represent date and actor in different views. Date dimension (DimDate) plays the roles of Movie start date, end date, release date and DVD release date in the fact table. And Actor dimension plays the roles of LeadActor and SupportingActor in the fact table.

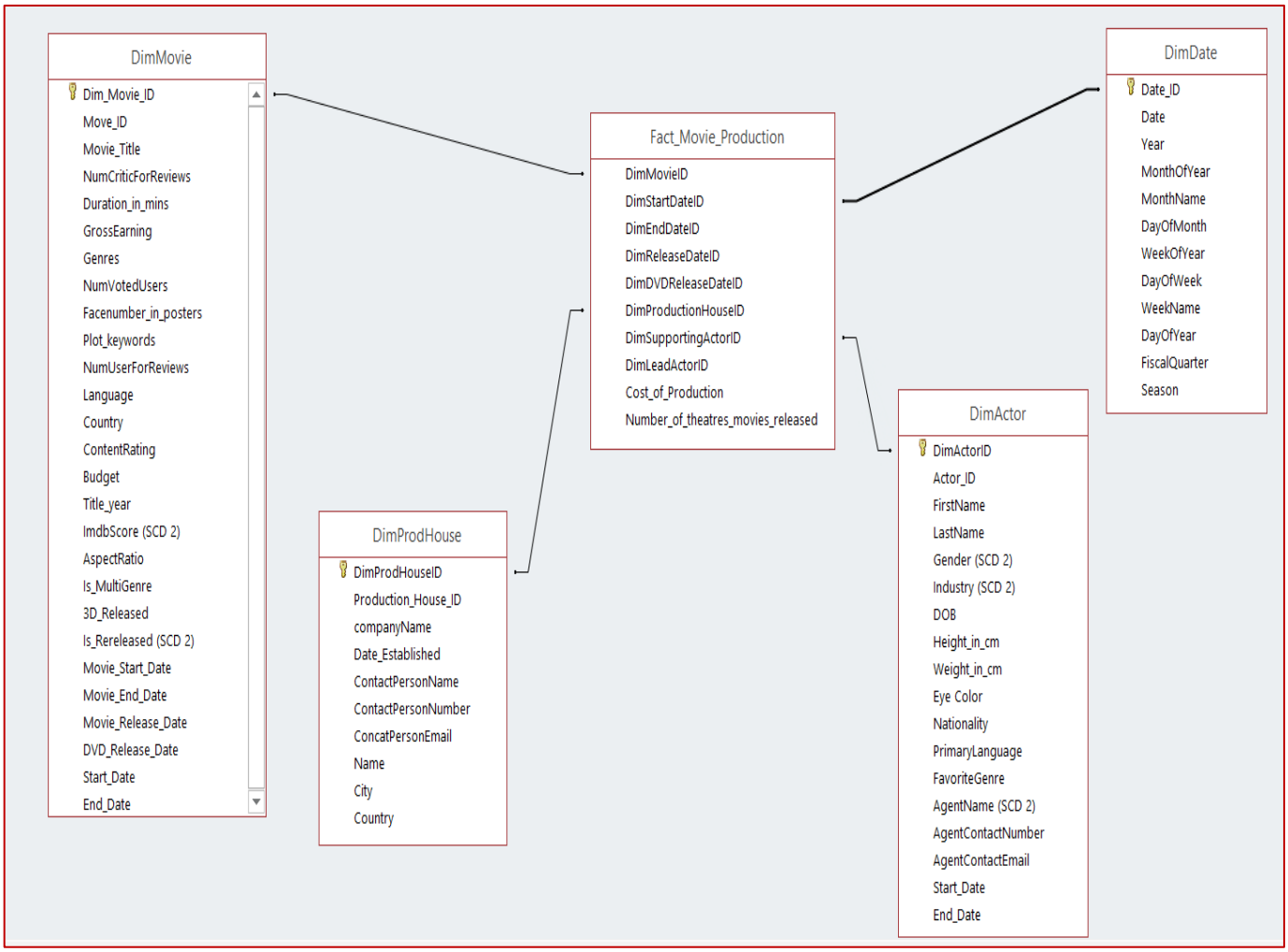


Figure: Movie Production data mart

- Theatre Performance

The data mart for analyzing the performance of a movie theatre is shown in the below figure. It contains:

- Movie Theatre Fact table
- Date, Movie and Theatre Dimension tables

The grain of the fact table is a Movie release in a theatre. A transactional entry happens every time a movie is release and weekly from that point on.

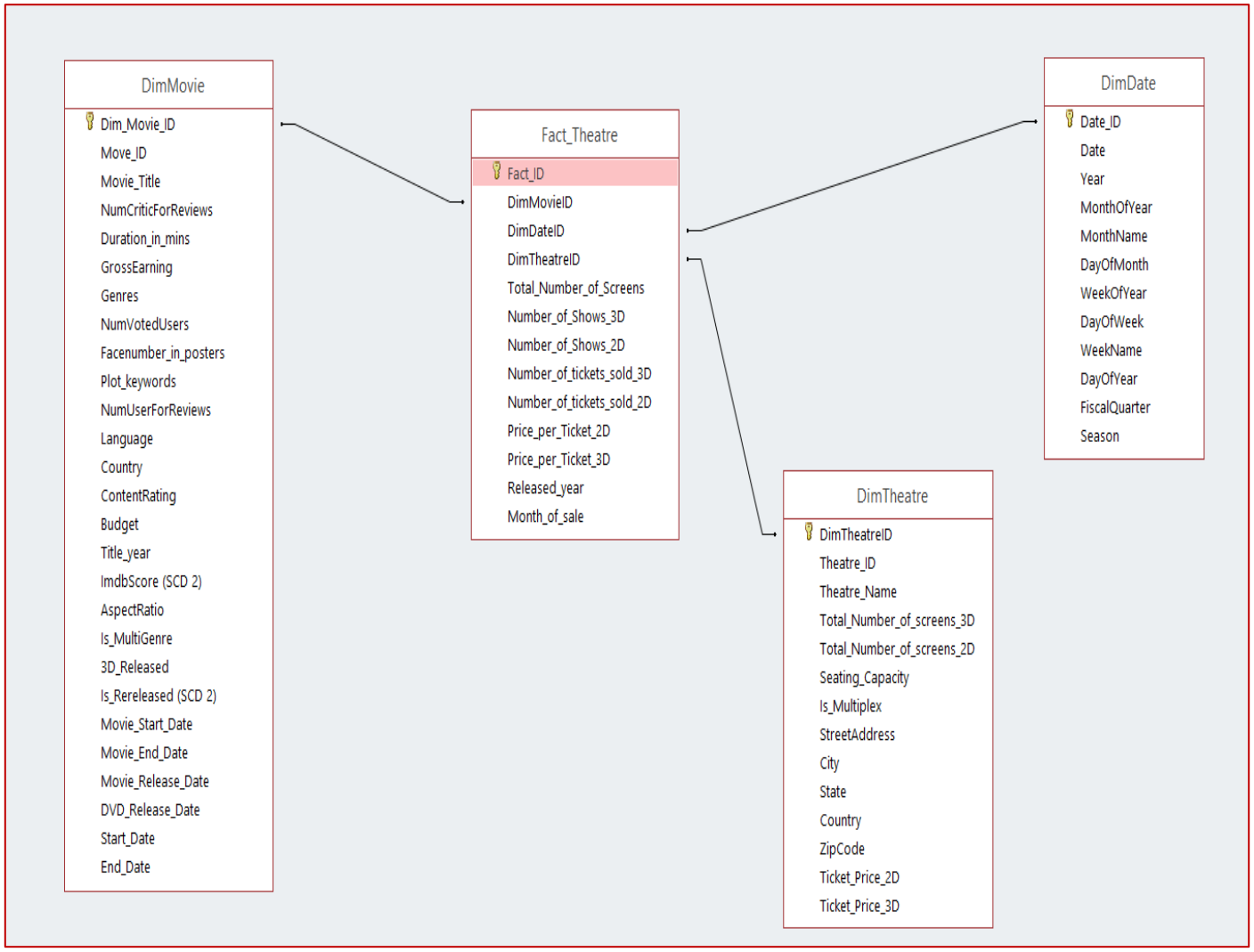


Figure: Theatre performance data mart

5. Data Warehouse Matrix

Business Process	*Date*	*Actor*	*Theatre*	*Movie*	*Production House*	*Director*	*Awards*
Production House Performance	**X**	**X**		**X**	**X**		
Movie Performance	**X**	**X**		**X**		**X**	
Theatre Performance	**X**		**X**	**X**			
Awards Distribution		**X**		**X**		**X**	**X**

6. Meta Data

6.1 Dimensions

Dimension Meta Data	Description
Name of Dimension	DimAward
Business Definition	This dimension holds the data that contains the various attributes for the awards that could be won by any movie, director, actor etc.

Attributes	<table><tr><th>Attributes</th><th>Format</th></tr><tr><td>DimAwardId</td><td>AutoNumber</td></tr><tr><td>Award_Name</td><td>ShortText</td></tr><tr><td>Country</td><td>ShortText</td></tr><tr><td>Venue</td><td>ShortText</td></tr><tr><td>Host_name</td><td>ShortText</td></tr><tr><td>Award_id</td><td>Number</td></tr><tr><td>Month</td><td>ShortText</td></tr></table>
SCD	All dimensions are of SCD type 1.
Hierarchy	No Hierarchy
Load Frequency	Weekly
Source	CSV
Conformed	Yes. The event attributes remains same across all fact tables
Role Playing	No Roles

Dimension Meta Data	**Description**
Name of Dimension	DimProductionHouse
Business Definition	This dimension holds the data that contains the various attributes for the Production house which produces any movie.
Attributes	<table><tr><th>Attributes</th><th>Format</th></tr><tr><td>DimProdHouseId</td><td>AutoNumber</td></tr><tr><td>CompanyName</td><td>ShortText</td></tr><tr><td>City</td><td>ShortText</td></tr><tr><td>Country</td><td>ShortText</td></tr><tr><td>Name</td><td>ShortText</td></tr><tr><td>ProdHouse_id</td><td>Number</td></tr><tr><td>Date_Established</td><td>Date/Time</td></tr><tr><td>ContactPersonEmail</td><td>ShortText</td></tr><tr><td>ContactPersonNumber</td><td>ShortText</td></tr><tr><td>ProdHouse_id</td><td>ShortText</td></tr><tr><td>Date_Established</td><td>ShortText</td></tr></table>
SCD	All dimensions are of SCD type 1.
Hierarchy	No Hierarchy
Load Frequency	Weekly

Source	Relational Database
Conformed	Yes. The event attributes remains same across all fact tables
Role Playing	No Roles

Dimension Meta data	**Description**
Name of Dimension	Dim_Actor
Business Definition	This dimension holds the data that contains the various attributes for the actor (Lead actor and Supporting actors).
Attributes Format	<table><tr><th>Attributes</th><th>Format</th></tr><tr><td>DimActorID</td><td>AutoNumber</td></tr><tr><td>Actor_ID</td><td>ShortText</td></tr><tr><td>FirstName</td><td>ShortText</td></tr><tr><td>LastName</td><td>ShortText</td></tr><tr><td>Gender (SCD 2)</td><td>ShortText</td></tr><tr><td>Industry (SCD 2)</td><td>ShortText</td></tr><tr><td>DOB</td><td>Date/Time</td></tr><tr><td>Height_in_cm</td><td>Number</td></tr><tr><td>Weight_in_cm</td><td>Number</td></tr><tr><td>Eye Color</td><td>ShortText</td></tr><tr><td>Nationality</td><td>ShortText</td></tr><tr><td>PrimaryLanguage</td><td>ShortText</td></tr><tr><td>FavoriteGenre</td><td>ShortText</td></tr><tr><td>AgentName (SCD 2)</td><td>ShortText</td></tr><tr><td>AgentContactNumber</td><td>ShortText</td></tr><tr><td>AgentContactEmail</td><td>ShortText</td></tr><tr><td>Start_Date</td><td>Date/Time</td></tr><tr><td>End_Date</td><td>Date/Time</td></tr></table>
SCD	All dimensions are of SCD type 1 except the below: Gender, Industry and AgentName are SCD type2. These are handled by use of start_date and end_date attributes.
Hierarchy	No Hierarchy
Load Frequency	Weekly
Source	Relational Database and CSV
Conformed	Yes. The event attributes remains same across all fact tables
Role Playing	DimActor plays the role of lead actor and supporting actor as a part of Fact_Movie_Performance and Fact_Movie_Production

<table>
<tr><th>Dimension meta data</th><th>Description</th></tr>
<tr><td>Name of Dimension</td><td>DimDirector</td></tr>
<tr><td>Business Definition</td><td>This dimension holds the data that contains the various attributes for the movie director.</td></tr>
<tr><td>Attributes Format</td><td>
<table>
<tr><th>Attributes</th><th>Format</th></tr>
<tr><td>DimDirectorID</td><td>int</td></tr>
<tr><td>Director_ID</td><td>float</td></tr>
<tr><td>FirstName</td><td>varchar</td></tr>
<tr><td>LastName</td><td>varchar</td></tr>
<tr><td>DOB</td><td>datetime</td></tr>
<tr><td>Gender</td><td>varchar</td></tr>
<tr><td>MovieIndustry</td><td>float</td></tr>
<tr><td>Nationality</td><td>varchar</td></tr>
<tr><td>Language Spoken</td><td>varchar</td></tr>
<tr><td>MovieGenre</td><td>varchar</td></tr>
</table>
</td></tr>
<tr><td>SCD</td><td>All dimensions are of SCD type 1.</td></tr>
<tr><td>Hierarchy</td><td>No Hierarchy</td></tr>
<tr><td>Load Frequency</td><td>Weekly</td></tr>
<tr><td>Source</td><td>Relational Database and CSV</td></tr>
<tr><td>Conformed</td><td>Yes. The event attributes remains same across all fact tables</td></tr>
<tr><td>Role Playing</td><td>No Roles</td></tr>
</table>

<table>
<tr><th>Dimension Meta data</th><th>Description</th></tr>
<tr><td>Name of Dimension</td><td>Dim_Theatre</td></tr>
<tr><td>Business Definition</td><td>This dimension holds the data that contains the various attributes about a theatre.</td></tr>
<tr><td>Attributes Format</td><td>
<table>
<tr><th>Attributes</th><th>Format</th></tr>
<tr><td>DimTheatreID</td><td>AutoNumber</td></tr>
<tr><td>Theatre_ID</td><td>ShortText</td></tr>
<tr><td>Theatre_Name</td><td>ShortText</td></tr>
<tr><td>Total_Number_of_screens_3D</td><td>Number</td></tr>
<tr><td>Total_Number_of_screens_2D</td><td>Number</td></tr>
<tr><td>Seating_Capacity</td><td>Number</td></tr>
<tr><td>Is_Multiplex</td><td>Yes/No</td></tr>
<tr><td>StreetAddress</td><td>ShortText</td></tr>
</table>
</td></tr>
</table>

	City: ShortText State: ShortText Country: ShortText ZipCode: Number Ticket_Price_2D: Number Ticket_Price_3D: Number
SCD	All dimensions are of SCD type 1.
Hierarchy	StreetAddress < City < State < Country
Load Frequency	Weekly
Source	Relational Database
Conformed	Yes. The event attributes remains same across all fact tables
Role Playing	No Roles

Dimension Meta data	**Description**
Name of Dimension	Dim_Movie
Business Definition	This dimension holds the data that contains the various attributes about a movie.
Attributes Format	Target table Attributes: Data Types DimMovieID: int Movie_ID: int Num_critic_for_reviews: float sDuration_in_minutes: float GrossEarning: float Genres: varchar Movie_title: varchar Num_voted_users: float Facenumber_in_poster: float Plot_keywords: varchar Num_user_for_reviews: float Language: varchar Country: varchar

	Content_rating: varchar Budget: float Title_year: float Imdb_score: float Aspect_ratio: float Is_Multigenre: varchar 3D_Released: varchar Is_Rereleased: varchar Movie_Start_Date: datetime Movie_End_Date: datetime Movie_Release_Date: datetime DVD_Release_Date: datetime
SCD	All dimensions are of SCD type 1.
Hierarchy	No Hierarchy
Load Frequency	Weekly
Source	Relational Database
Conformed	Yes. The event attributes remains same across all fact tables
Role Playing	No Roles

Dimension Meta data	**Description**
Name of Dimension	DimDate
Business Definition	This dimension holds the data about Date and time.
Attributes Format	Attributes: Format DimdateID: int Date: datetime Year: int MonthofYear: int MonthName: varchar DayofMonth: float WeekofYear: float DayofWeek: float

WeekName	varchar
DayofYear	float
Fiscal Quarter	varchar
Season	varchar
DimdateID	int
Date	datetime

SCD	All dimensions are of SCD type 1.
Hierarchy	Date < Week < Month < Year
Load Frequency	Weekly
Source	Relational Database
Conformed	Yes. The event attributes remains same across all fact tables
Role Playing	Yes. There are role playing dimensions of Movie_start_date, Movie_end_date, Movie_release_date, DVD_release_date

6.2 Facts

Fact Table Meta Data	**Description**
Name of Fact table	Fact_Movie_Production
Business definition	This transactional fact table is created to capture the performance of the production houses.
Dimensions	DimDate, DimMovie, DimProdHouse, DimActor
Grain	Record is created for each movie capturing the production house Of the movie, cost of production and the number of thetares movie was released in.
Load Frequency	Weekly
Source	Transactional Database
Measures / Facts	Cost Of Production Number_of_theatres_movies_released

Fact Table Meta Data	**Description**
Name of Fact table	Fact_Theatre
Business definition	This transactional fact table is created to capture the performance of the theaters worldwide.
Dimensions	Dim_Date, DimMovie, DimTheatre
Grain	Record is created for each movie released in each theatre capturing the all theatre attributes for each movie such as ticket price, no of shows etc.
Load Frequency	Weekly
Source	Transactional Database
Measures / Facts	Total_Number_of_Screens Number_of_Shows_3D Number_of_Shows_2D

	Number_of_tickets_sold_3D Number_of_tickets_sold_2D Price_per_Ticket_2D Price_per_Ticket_3D

Fact Table Meta Data	**Description**
Name of Fact table	Fact_Award
Business definition	This transactional fact table is created to capture the winners of the various awards.
Dimensions	DimAward, DimMovie, DimDirector, DimActor
Grain	Record is created for each award capturing the winner of that year/month's winner.
Load Frequency	Weekly
Source	Transactional Database
Measures / Facts	Factless Fact table

Fact Table Meta Data	**Description**
Name of Fact table	Fact_Movie_Performance
Business definition	This transactional fact table is created to capture the performance of the movies worldwide.
Dimensions	Dim_Date, DimDirector, DimActor, DimMovie
Grain	Record is created for each movie released capturing the movie's performance such as number of tickets sold, profit price, etc.
Load Frequency	Weekly
Source	Transactional Database
Measures / Facts	Gross Profit, Number_of_tickets_sold_2D, Number_of_tickets_sold_3D, Gross_Revenue_Generated

7. ETL Plan

ETL refers to 'Extract, Transform and load'. The various steps in the ETL process are as follows:

- Extracting data from databases: Data presented in the .csv files is extracted and loaded into the staging area by using Data flow tasks.
- Transforming the extracted data: For storing the data in relevant formats to enable query execution Data present in the staging area is cleaned and transformed to cater to the Business Intelligence questions.
- Loading the data into a final target database: The final tables that have been created are loaded into the Facts and Dimension tables.

Following is the layout of the **ETL plan** for Data Warehouse implementation:

- Preparation of Data mappings of the Data from sources in Excel to staging area and from the staging area to the data warehouse.
- Determine the Data extraction rules
- Determine the Data transformation and cleansing rules
- Implementation plan : Plan and execute procedures for extraction and loading

7.1 Data Mappings for Data Warehouse (including sources, staging and target details and transformations)

Target table	Target table Attributes	Data Types	Source File	Transformation Rule
DimActor	DimActorID	int	Actor_table.xlsx	Surrogate key of the dimension. Inserted as incremental key while loading data.
DimActor	Actor_Id	float	Actor_table.xlsx	Primary Key
DimActor	First_name	varchar	Actor_table.xlsx	
DimActor	Last_name	varchar	Actor_table.xlsx	
DimActor	Gender	varchar	Actor_table.xlsx	
DimActor	Industry	varchar	Actor_table.xlsx	
DimActor	DOB	datetime	Actor_table.xlsx	
DimActor	Height_in_cm	float	Actor_table.xlsx	
DimActor	Weight_in_lb	float	Actor_table.xlsx	
DimActor	EyeColor	varchar	Actor_table.xlsx	
DimActor	Nationality	varchar	Actor_table.xlsx	
DimActor	PrimaryLanguage	varchar	Actor_table.xlsx	
DimActor	AgentName	varchar	Actor_table.xlsx	
DimActor	AgentContactNumber	varchar	Actor_table.xlsx	
DimActor	AgentContactEmail	varchar	Actor_table.xlsx	
DimActor	FavoriteGenre	varchar	Actor_table.xlsx	

Target table	Target table Attributes	Data Types	Source File	Transformation Rule
DimAward	DimAwardID	int	award.csv	Surrogate key of the dimension. Inserted as incremental key while loading data.
DimAward	Award_Id	float	award.csv	Primary key
DimAward	Award_name	varchar	award.csv	
DimAward	Country	varchar	award.csv	
DimAward	Venue	varchar	award.csv	
DimAward	Host_name	varchar	award.csv	
DimAward	Month	varchar	award.csv	

Target table	Target table Attributes	Data Types	Source File	Transformation Rule
DimDate	DimdateID	int	Date.xlsx	Primary key
DimDate	Date	datetime	Date.xlsx	
DimDate	Year	int	Date.xlsx	
DimDate	MonthofYear	int	Date.xlsx	
DimDate	MonthName	varchar	Date.xlsx	
DimDate	DayofMonth	float	Date.xlsx	
DimDate	WeekofYear	float	Date.xlsx	
DimDate	DayofWeek	float	Date.xlsx	
DimDate	WeekName	varchar	Date.xlsx	
DimDate	DayofYear	float	Date.xlsx	
DimDate	Fiscal Quarter	varchar	Date.xlsx	
DimDate	Season	varchar	Date.xlsx	

Target table	Target table Attributes	Data Types	Source File	Transformation Rule
DimDirector	DimDirectorID	int	Director.xlsx	Surrogate key of the dimension. Inserted as incremental key while loading data.
DimDirector	Director_ID	float	Director.xlsx	Primary key
DimDirector	FirstName	varchar	Director.xlsx	
DimDirector	LastName	varchar	Director.xlsx	
DimDirector	DOB	datetime	Director.xlsx	
DimDirector	Gender	varchar	Director.xlsx	
DimDirector	MovieIndustry	float	Director.xlsx	
DimDirector	Nationality	varchar	Director.xlsx	
DimDirector	Language Spoken	varchar	Director.xlsx	
DimDirector	MovieGenre	varchar	Director.xlsx	

Target table	Target table Attributes	Data Types	Source File	Transformation Rule
DimMovie	DimMovieID	int	Movie.xlsx	Surrogate key of the dimension. Inserted as incremental key while loading data.
DimMovie	Movie_ID	int	Movie.xlsx	Primary key
DimMovie	Num_critic_for_reviews	float	Movie.xlsx	
DimMovie	Duration_in_minutes	float	Movie.xlsx	
DimMovie	GrossEarning	float	Movie.xlsx	
DimMovie	Genres	varchar	Movie.xlsx	
DimMovie	Movie_title	varchar	Movie.xlsx	
DimMovie	Num_voted_users	float	Movie.xlsx	

DimMovie	Facenumber_in_poster	float	Movie.xlsx	
DimMovie	Plot_keywords	varchar	Movie.xlsx	
DimMovie	Num_user_for_reviews	float	Movie.xlsx	
DimMovie	Language	varchar	Movie.xlsx	
DimMovie	Country	varchar	Movie.xlsx	
DimMovie	Content_rating	varchar	Movie.xlsx	
DimMovie	Budget	float	Movie.xlsx	
DimMovie	Title_year	float	Movie.xlsx	
DimMovie	Imdb_score	float	Movie.xlsx	
DimMovie	Aspect_ratio	float	Movie.xlsx	
DimMovie	Is_Multigenre	varchar	Movie.xlsx	
DimMovie	3D_Released	varchar	Movie.xlsx	
DimMovie	Is_Rereleased	varchar	Movie.xlsx	
DimMovie	Movie_Start_Date	datetime	Movie.xlsx	
DimMovie	Movie_End_Date	datetime	Movie.xlsx	
DimMovie	Movie_Release_Date	datetime	Movie.xlsx	
DimMovie	DVD_Release_Date	datetime	Movie.xlsx	

Target table	Target table Attributes	Data Types	Source File	Transformation Rule
DimProdHouse	DimProdHouseID	int	production.xlsx	Surrogate key of the dimension. Inserted as incremental key while loading data.
DimProdHouse	ProdHouse_ID	float	production.xlsx	Primary key
DimProdHouse	CompanyName	varchar	production.xlsx	
DimProdHouse	City	varchar	production.xlsx	
DimProdHouse	Country	varchar	production.xlsx	
DimProdHouse	Name	varchar	production.xlsx	

DimProdHouse	ContactPersonEmail	varchar	production.xlsx	
DimProdHouse	DateEstablished	datetime	production.xlsx	
DimProdHouse	ContactPersonNumber	varchar	production.xlsx	

Target table	Target table Attributes	Data Types	Source File	Transformation Rule
DimTheatre	DimTheatreID	int	Theatre.xlsx	Surrogate key of the dimension. Inserted as incremental key while loading data.
DimTheatre	theatre_ID	Int	Theatre.xlsx	Primary key
DimTheatre	Theatre_Name	varchar	Theatre.xlsx	
DimTheatre	Total_num_of_Screens_2D	float	Theatre.xlsx	
DimTheatre	Total_num_of_Screens_3D	float	Theatre.xlsx	
DimTheatre	Seating_Capacity	float	Theatre.xlsx	
DimTheatre	IsMultiplex	bit	Theatre.xlsx	
DimTheatre	Street_Address	varchar	Theatre.xlsx	
DimTheatre	City	varchar	Theatre.xlsx	
DimTheatre	State	varchar	Theatre.xlsx	
DimTheatre	Country	varchar	Theatre.xlsx	
DimTheatre	ZipCode	float	Theatre.xlsx	
DimTheatre	Ticket_price_2D	float	Theatre.xlsx	
DimTheatre	Ticket_price_3D	float	Theatre.xlsx	

Fact tables

Target table	Target table Attributes	Data Types	Staging table attributes	Transformation Rule
fact_Awards	DimActorID	int	Actor_Id	Foreign key of dimension table corresponding to Actor ID
fact_Awards	DimDirectorID	Int	Director_ID	Foreign key of dimension table corresponding to Director ID
fact_Awards	DimMovieId	int	Movie_Id	Foreign key of dimension table corresponding to movie ID
fact_Awards	DimAwardID	int	Award_Id	

Target table	Target table Attributes	Data Types	Staging table attributes	Transformation Rule
fact_Movie_Performance	DimMovieId	int	Movie_Id	Foreign key of dimension table corresponding to movie ID
fact_Movie_Performance	DimActorID	int	Actor_ID	Foreign key of dimension table corresponding to actor ID
fact_Movie_Performance	DimActorID	int	Actor_ID	Foreign key of dimension table corresponding to actor ID
fact_Movie_Performance	DimDirectorID	int	Director_Id	Foreign key of dimension table corresponding to director ID
fact_Movie_Performance	DimDateID	int	Date_Id	Foreign key of dimension table corresponding to date ID

fact_Movie_Performance	Number_of_tickets_sold_2D	float		
fact_Movie_Performance	Number_of_tickets_sold_3D	float		
fact_Movie_Performance	Gross_Revenue_Generated	float		
fact_Movie_Performance	GrossProfit	float		

Target table	Target table Attributes	Data Types	Staging table attributes	Transformation Rule
fact_movie_production	DimMovieID	int	Movie_Id	Foreign key of dimension table corresponding to Movie ID
fact_movie_production	DimStartDateID	Int	Date_ID	Foreign key of dimension table corresponding to Date ID
fact_movie_production	DimEndDateId	int	Date_ID	Foreign key of dimension table corresponding to Date ID
fact_movie_production	DimReleaseDateId	int	Date_ID	Foreign key of dimension table corresponding to Date ID
fact_movie_production	DimDVDReleaseDateId	int	Date_ID	Foreign key of dimension table corresponding to Date ID
fact_movie_production	DimProdHouseID	int	ProdHouse_ID	Foreign key of dimension table corresponding to ProdHouse_ID
fact_movie_production	DimLeadActorID	int	Actor_Id	Foreign key of dimension table corresponding to Actor_ID
fact_movie_production	DimSupportingActorID	int	Actor_id	Foreign key of dimension table corresponding to Actor_ID
fact_movie_production	Cost_of_production	float		

fact_movie_production	Number_of_Theatres_Movie_Released_In	int		

Target table	Target table Attributes	Data Types	Staging table attributes	Transformation Rule
fact_Theatre	DimMovieID	int	Movie_Id	Foreign key of dimension table corresponding to Movie ID
fact_Theatre	DimTheatreID	Int	Date_ID	Foreign key of dimension table corresponding to Theatre ID
fact_Theatre	DimDateId	int	Date_ID	Foreign key of dimension table corresponding to Date ID
fact_Theatre	Number_of_tickets_sold_2D	int		
fact_Theatre	Number_of_tickets_sold_3D	int		
fact_Theatre	Number_of_shows_2D	int		
fact_Theatre	Number_of_shows_3D	int		
fact_Theatre	Total_number_of_Screens	int		
fact_Theatre	Released_year	float		
fact_Theatre	Month_of_sale	int		
fact_Theatre	Price_per_ticket_2D			
fact_Theatre	Price_per_ticket_3D			

7.2 Data Extraction Rules

The process of retrieving data out from data sources for processing or storage is known as Data Extraction. Data extraction is the initial step of data transforming, loading and then designing the data warehouse. Data present in data files is often poorly structured. The import to the staging system of such data is usually followed by data transformation before moving ahead.

To achieve data extraction, we employed the following steps:

- Source data that is present in the Comma Separated Value (.csv) and other formats is extracted and imported into Microsoft SQL server as tables.
- This data is used for Data transformation and further loading. Once data has been put extracted into the staging area, it is cleaned and transformed to create Dimension and fact tables. Once these Dimension and fact tables have been verified with respect to the Business Intelligence needs, it is loaded into the Data Warehouse area.

7.3 Data transformation and cleaning rules

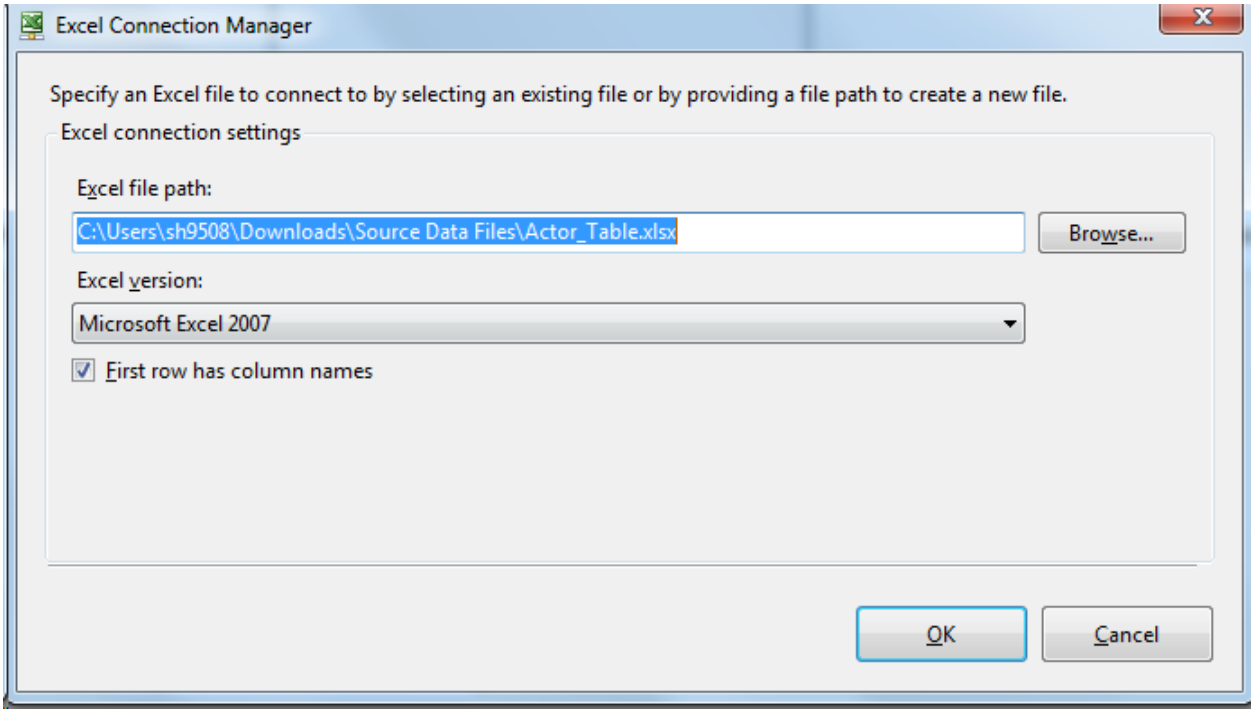
The next step is to clean the data that has been extracted from the source files. In order to maintain the consistency of the data throughout, it is essential to clean and transform the data. It ensures that all dirty data is removed and does not lead to any false results. The clean data is then loaded into individual data marts for further analysis. The following actions were performed to clean the data:

- **Removal of Dirty Data:** Attributes in data sources which were irrelevant to the business questions asked, were ignored while extracting data. Records having strange values were deleted. For example, records having gibberish special character values "%%^$#%" were deleted.
- **Removal of Null Values**: All the null values present in various tables are deleted.
- **Surrogate Key Creation**: All dimension and fact tables have surrogate keys created before the data is loaded in the data warehouse.
- **Derived Attributes:** The derived attributes in the Dimensional Table and Fact Tables are as follows:
  - In the Movie_Performance fact table derived column is added with name Gross Profit.

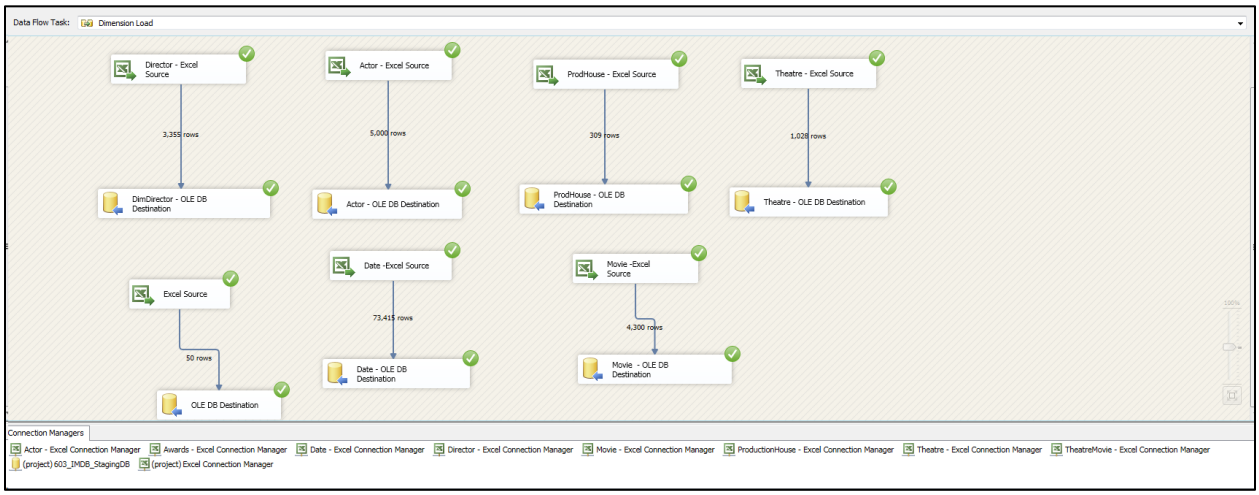
7.4 Implementation plan

Dimension Creation:

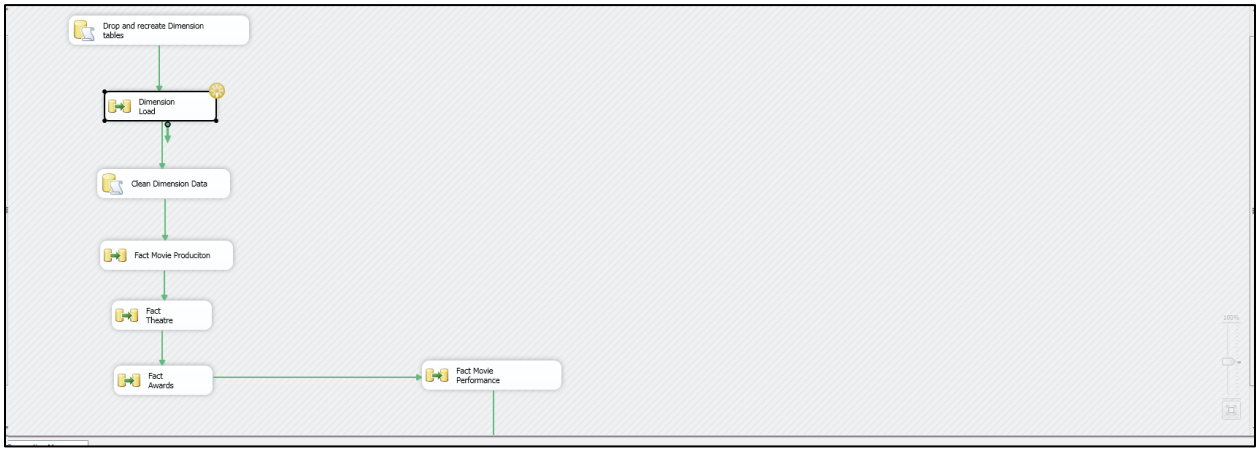
Step1: Establishing the Connection Manager



Step 2: Establish Connection Manager for all Dimensions:



Step 3: Executing Dimension Load Control Flow:

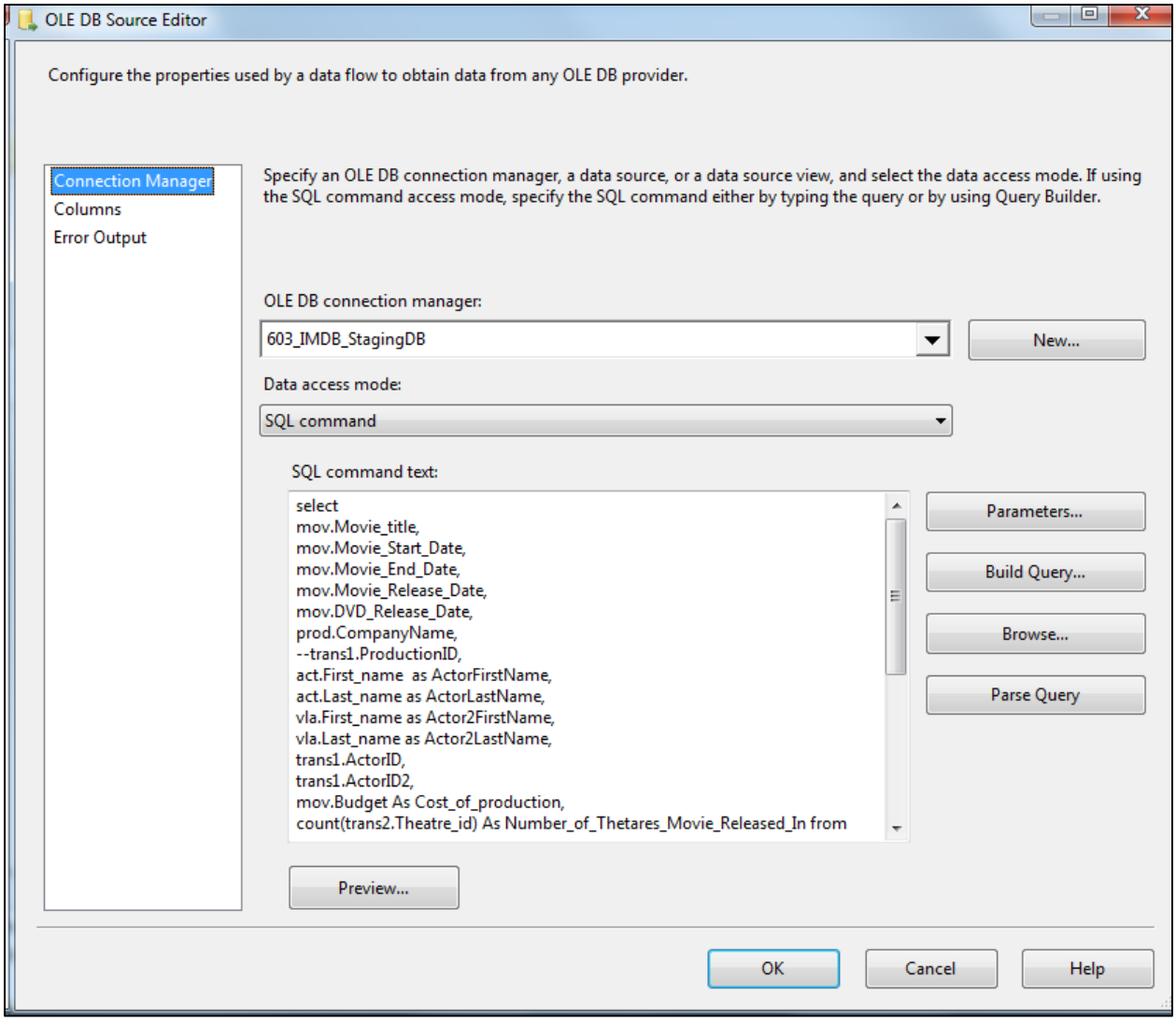


Facts Creation:

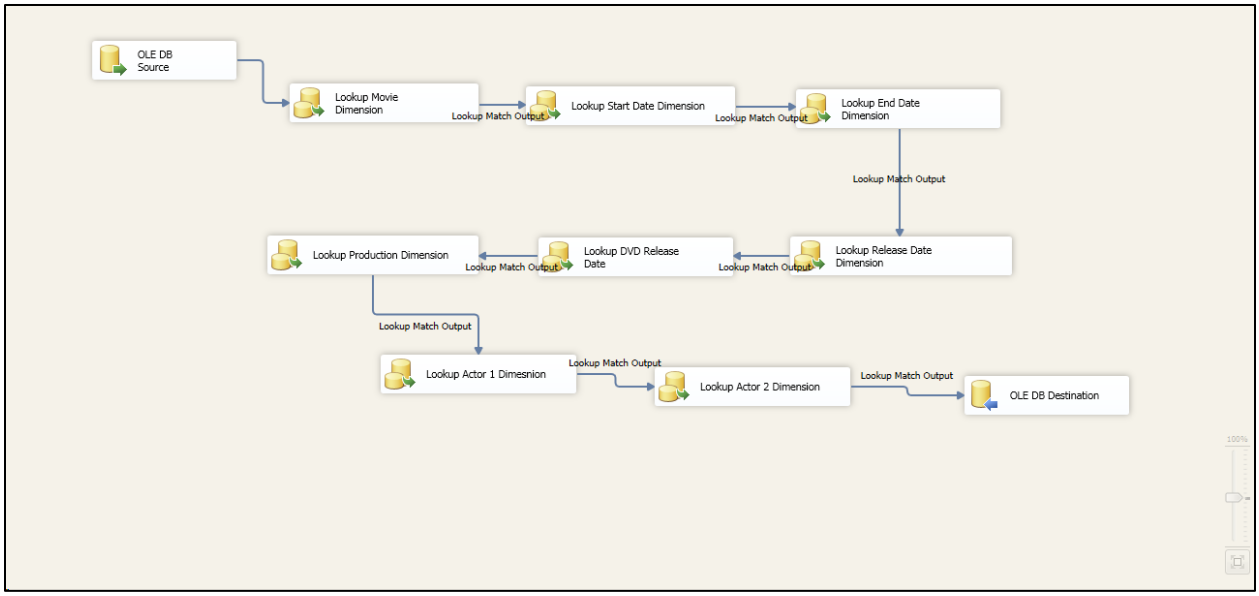
1. Fact Movie Production:

Step 1: Establish Fact Movie Production Control Flow

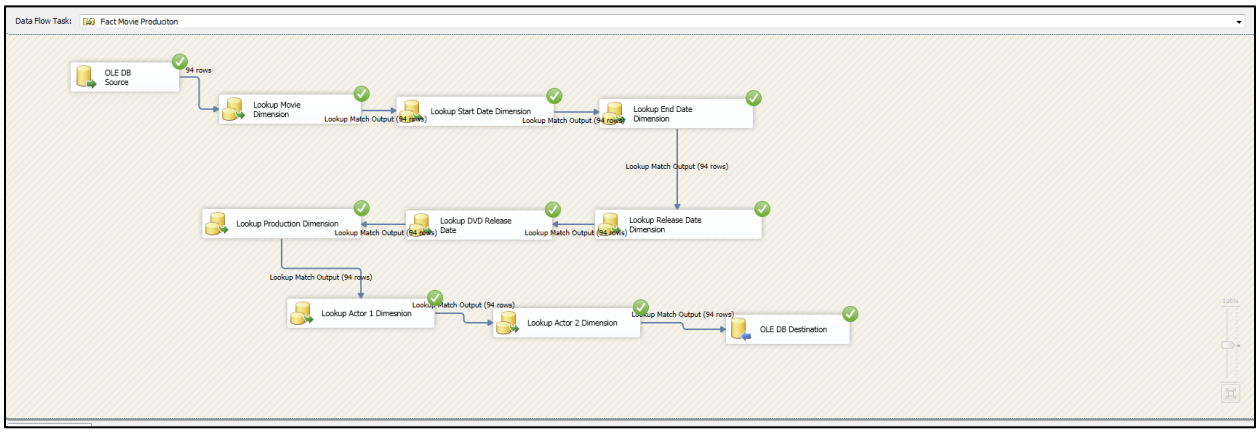
Step 2: Establishing Pre fact query and providing it as OLE DB data source:



Step 3: Making Lookups on all the Dimension Tables which are required:



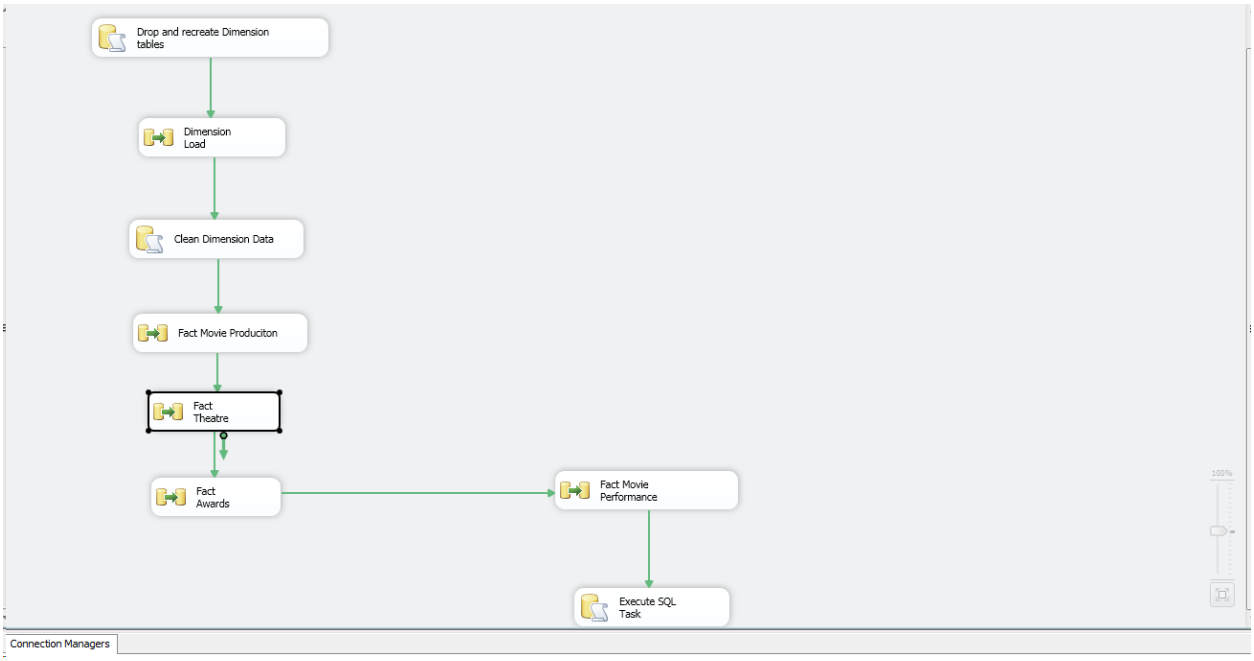
Step 4: Successful execution of all Lookups:



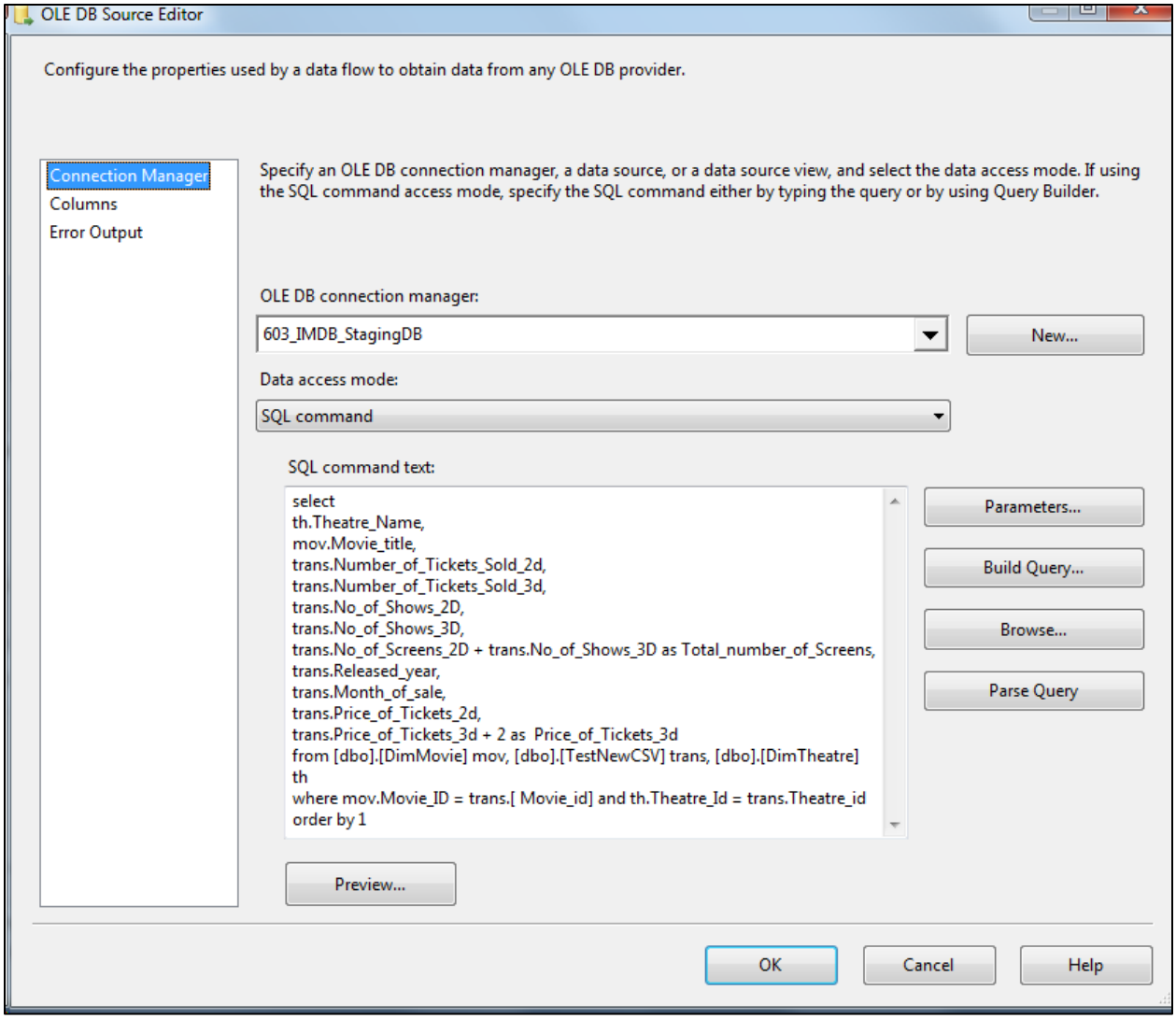
Step 5: Executing fact- Movie Production Control Flow:

2. Fact Theatre

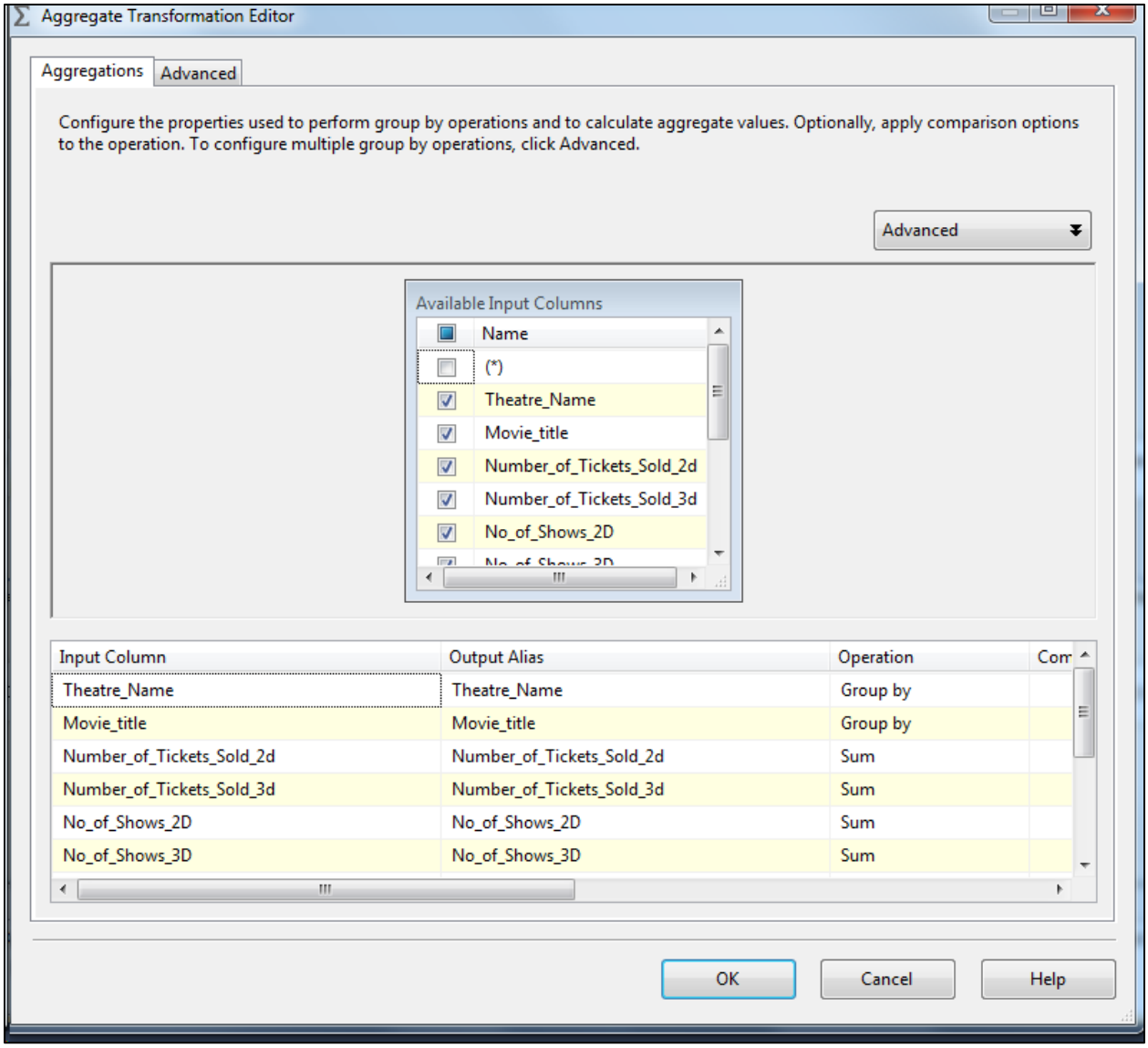
Step 1: Establish Fact Theatre Control Flow



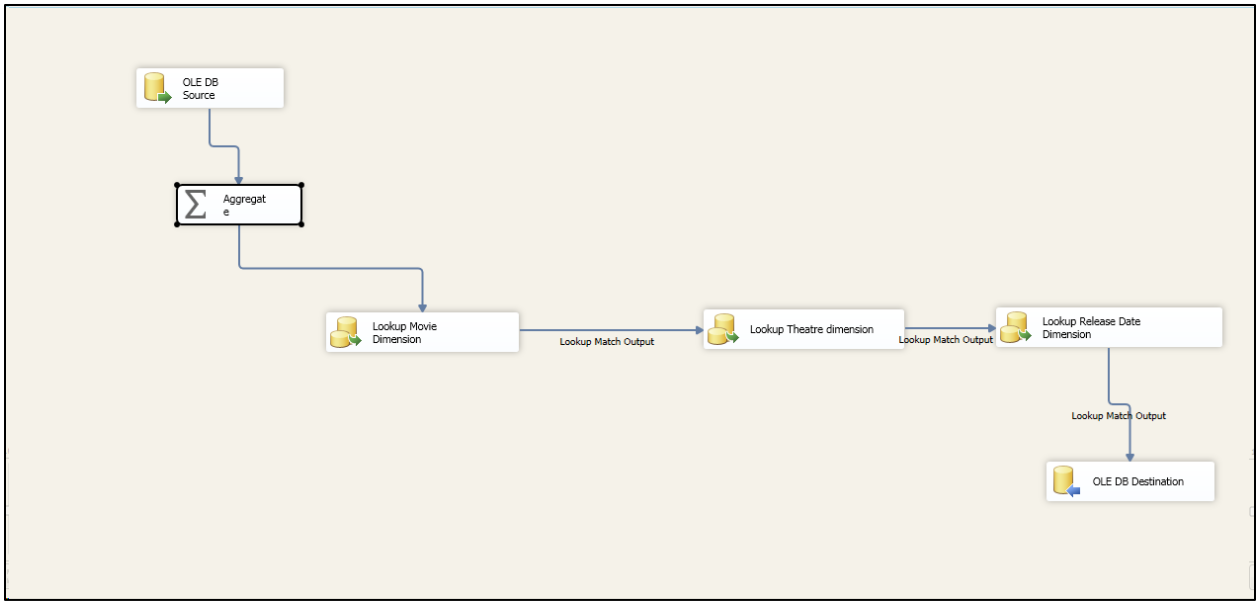
Step 2: Establishing Pre fact query and providing it as OLE DB data source:



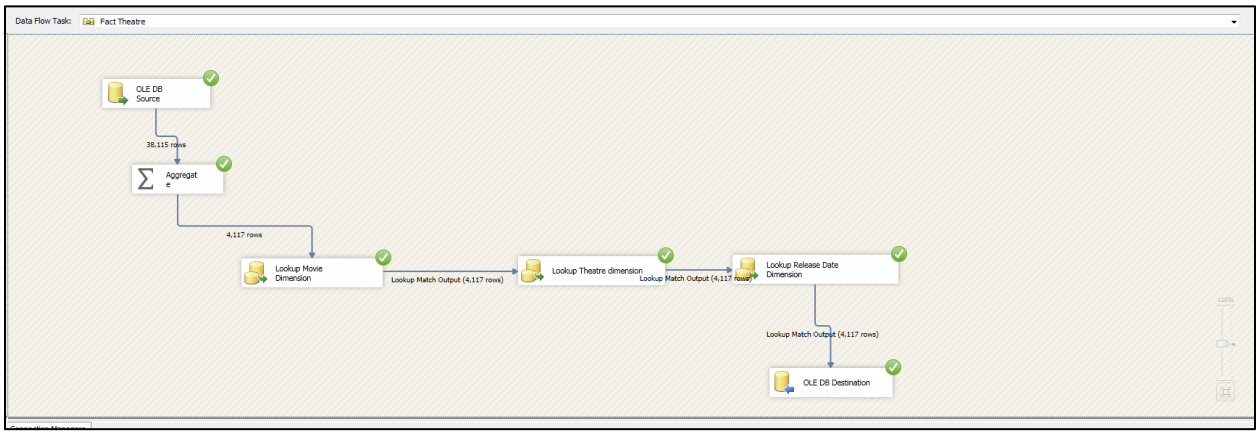
Step 3: Performing aggregation on the pre Fact query:



Step 4: Making Lookups on all the Dimension Tables which are required:



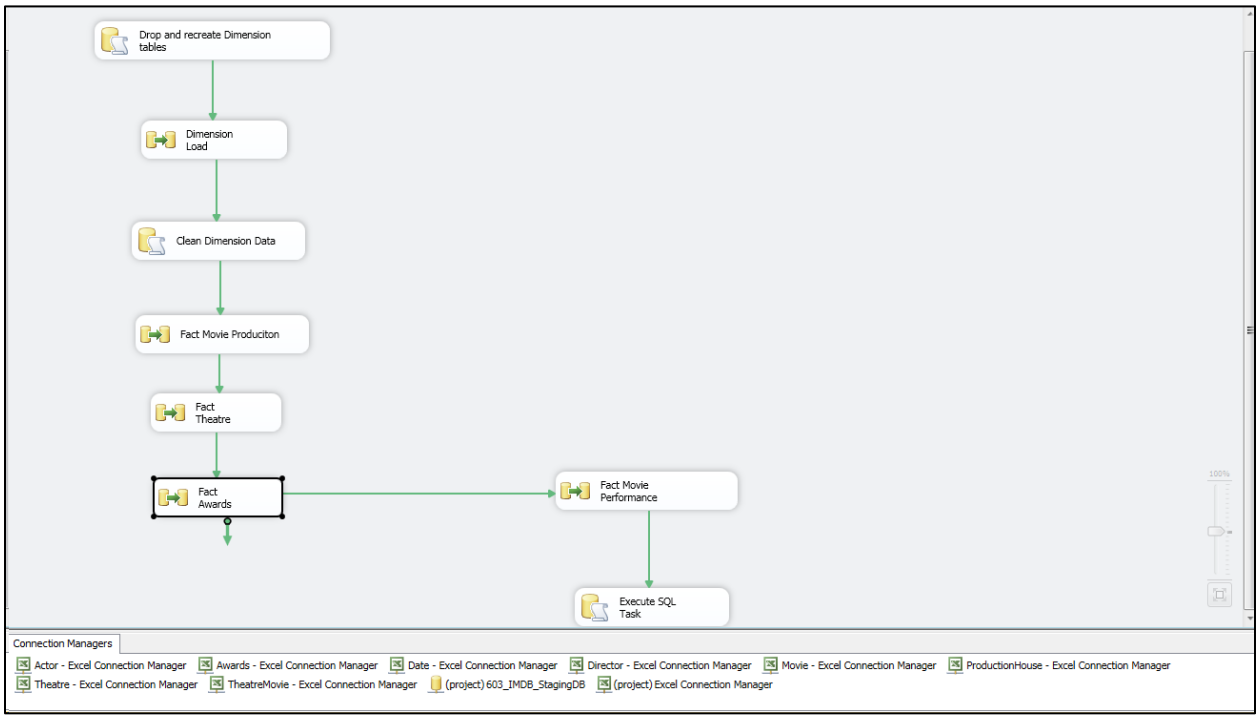
Step 5: Successful execution of all Lookups:



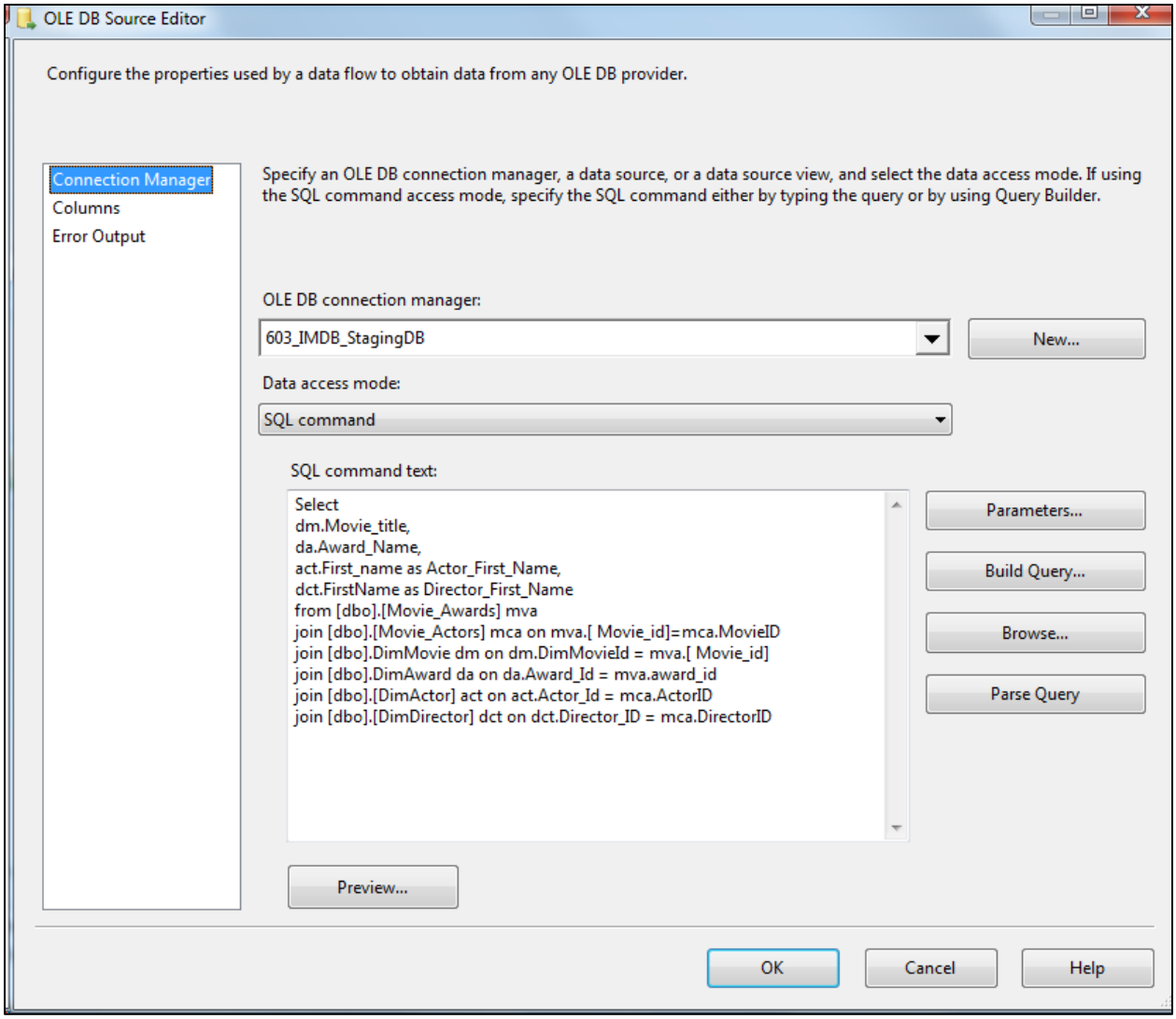
Step 6: Executing fact- theatre Control Flow:

3. Fact Awards

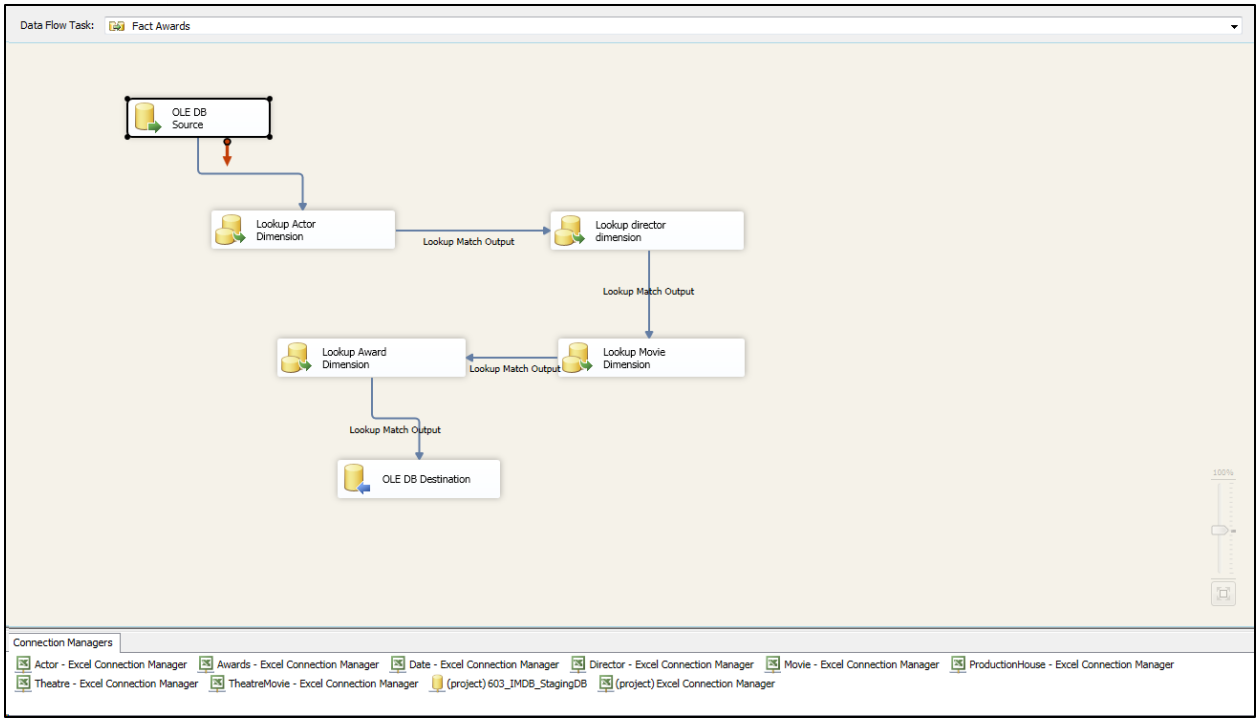
Step 1: Establish Fact Awards Control Flow



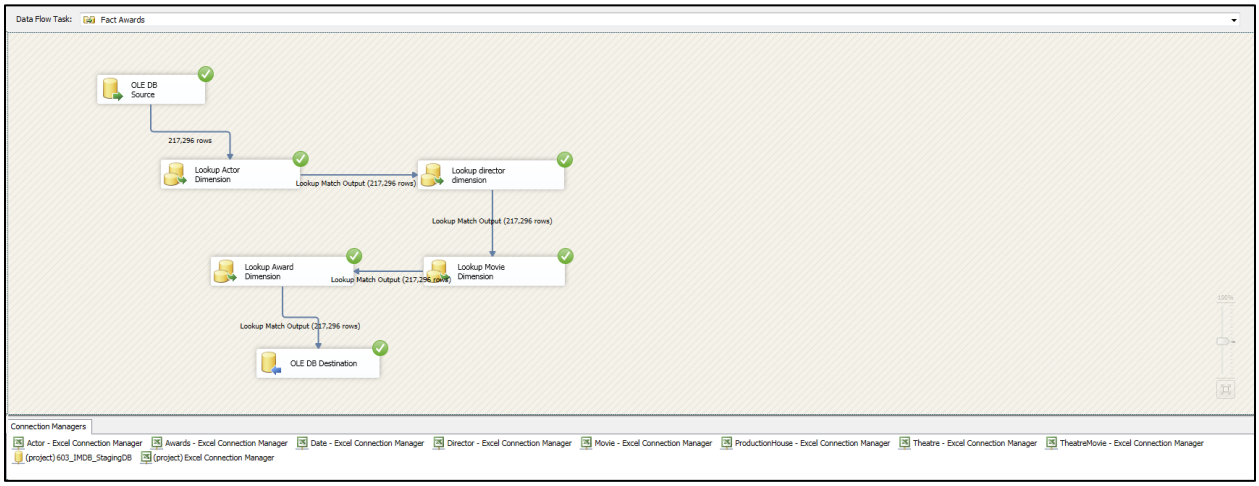
Step 2: Establishing Pre fact query and providing it as OLE DB data source:



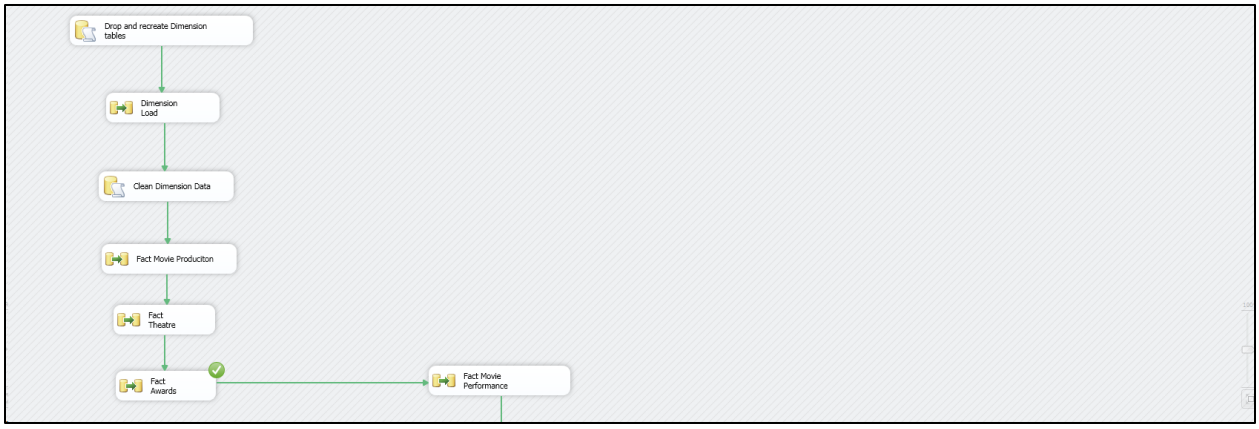
Step 3: Making Lookups on all the Dimension Tables which are required:



Step 4: Successful execution of all Lookups:

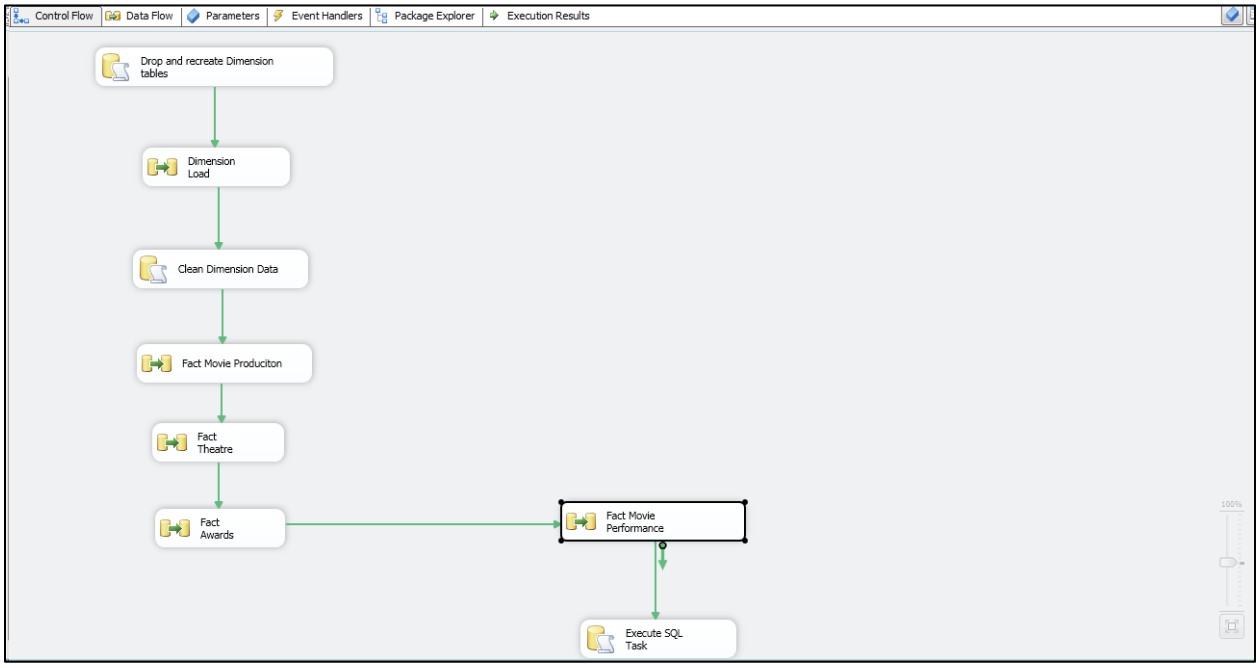


Step 5: Executing fact- Awards Control Flow:

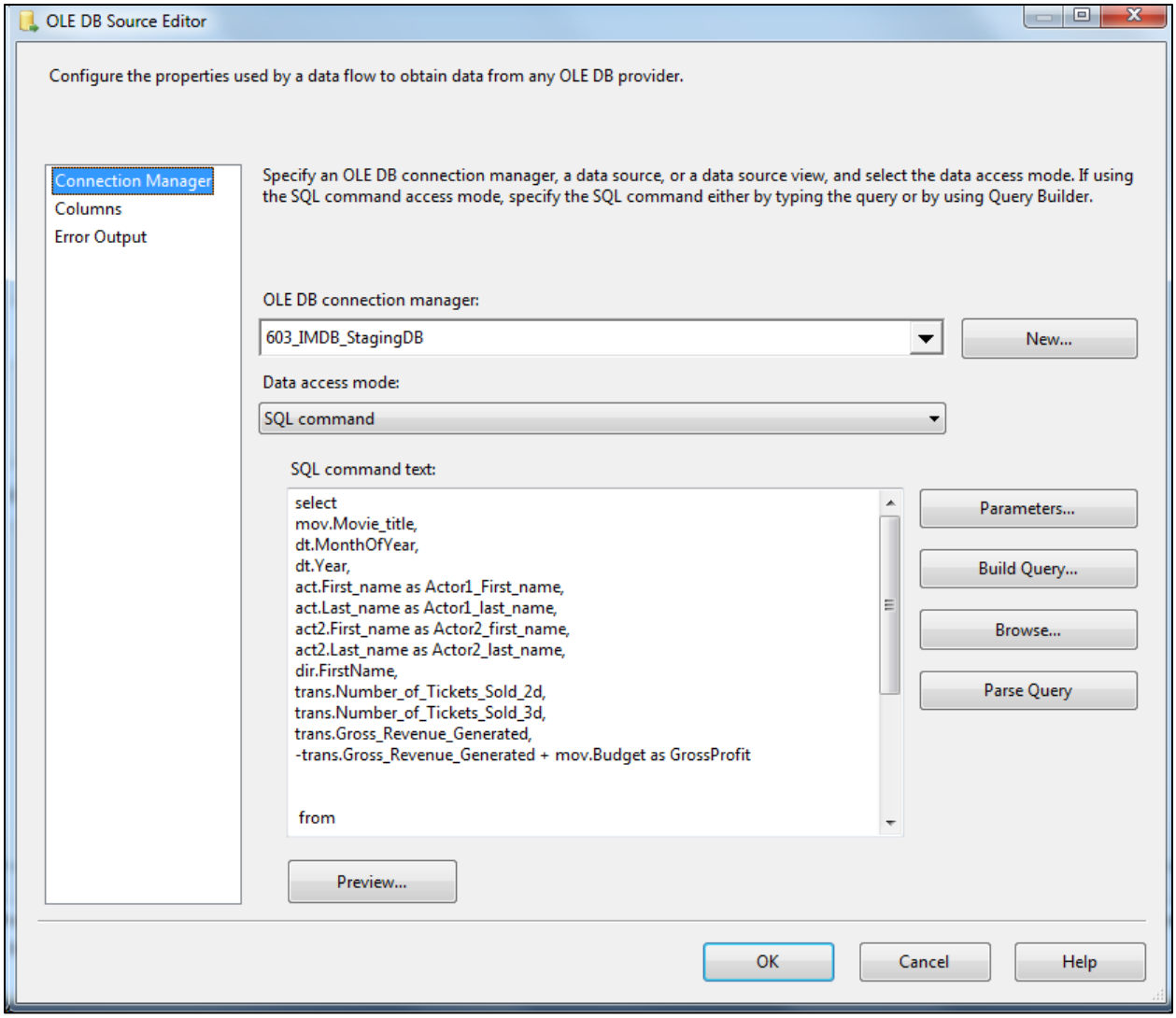


4. Fact Movie Performance

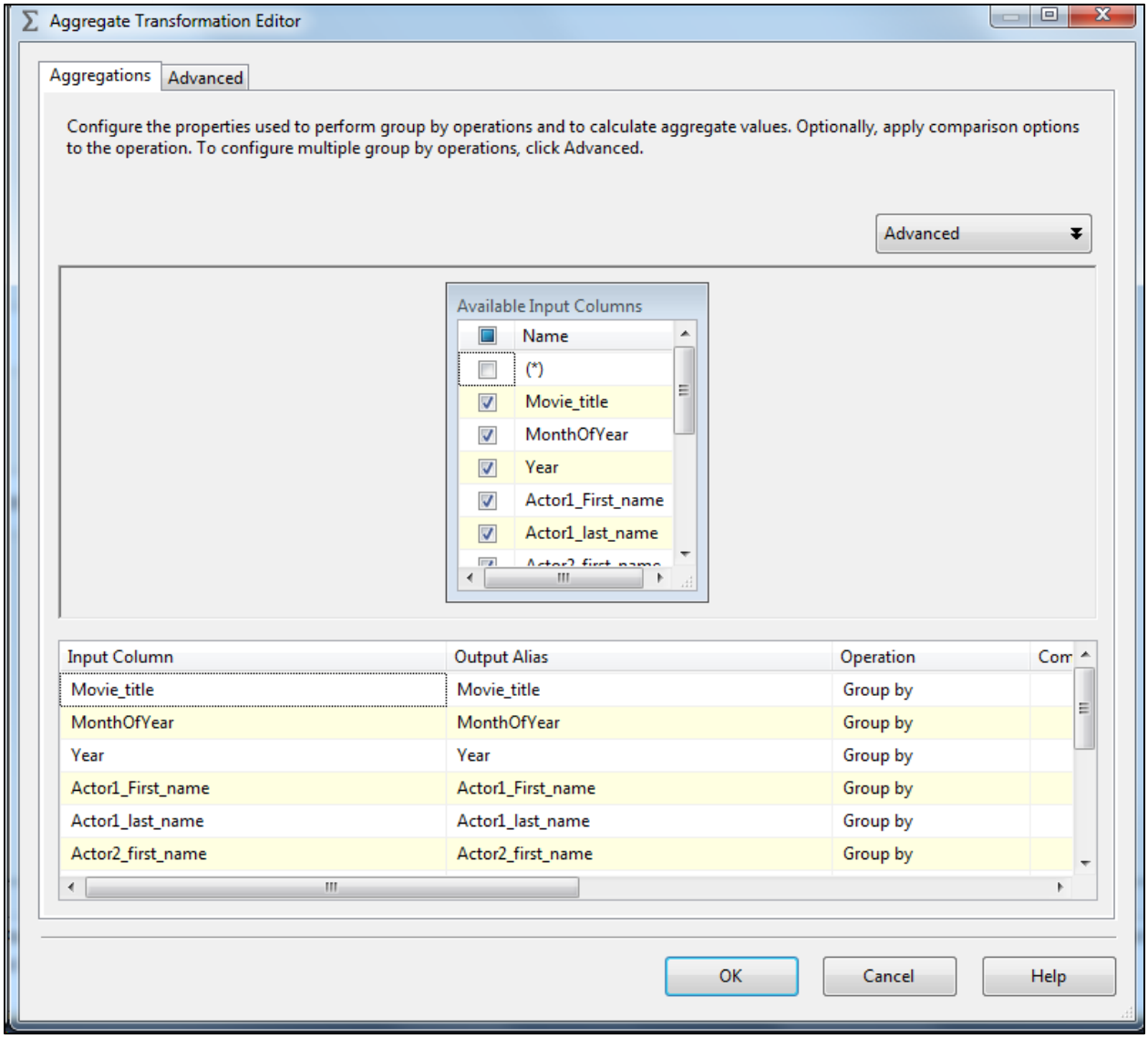
Step 1: Establish Fact Awards Control Flow



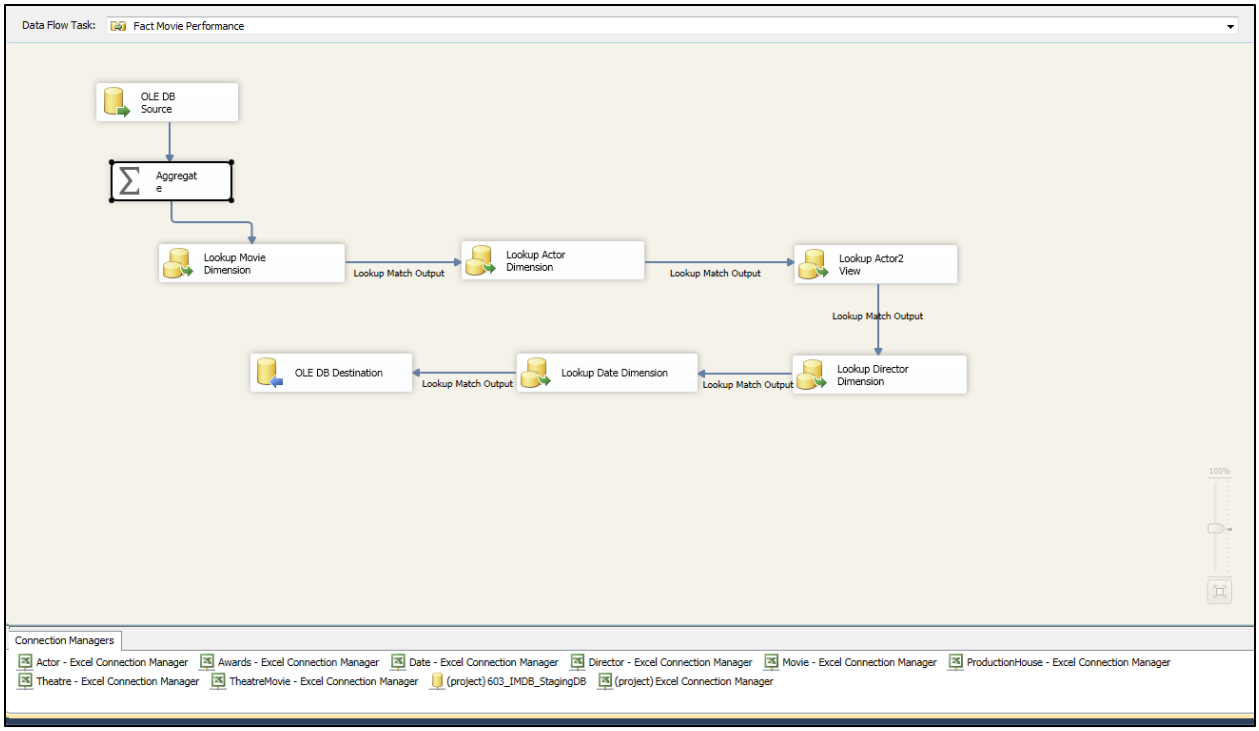
Step 2: Establishing Pre fact query and providing it as OLE DB data source:



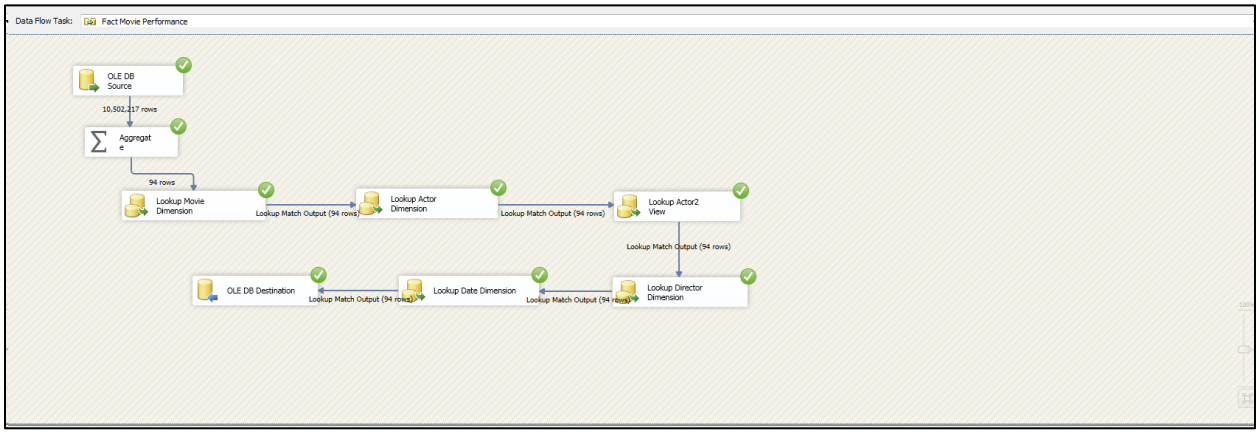
Step 3: Performing aggregation on the pre Fact query:



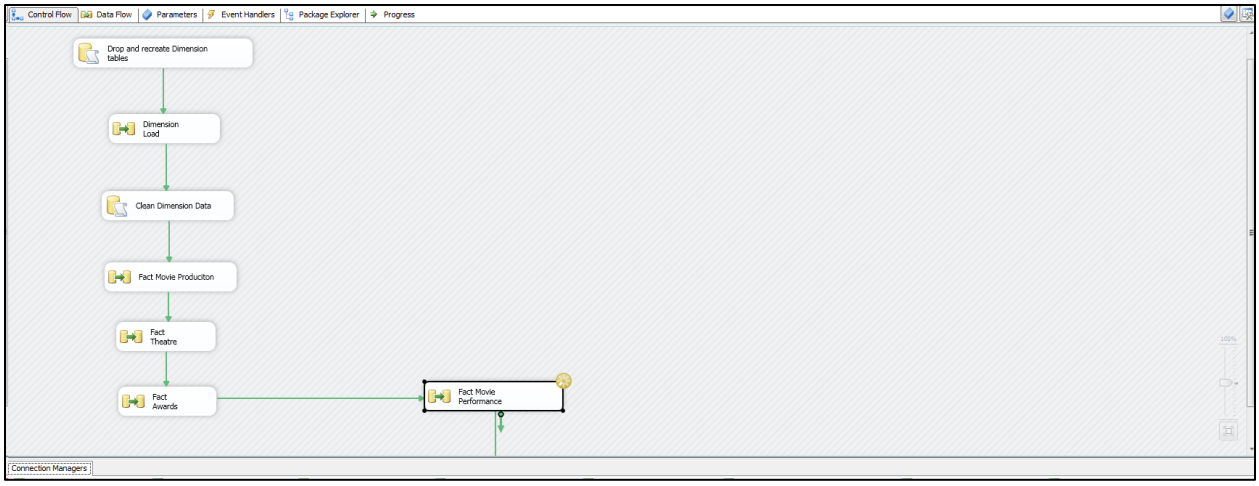
Step 4: Making Lookups on all the Dimension Tables which are required:

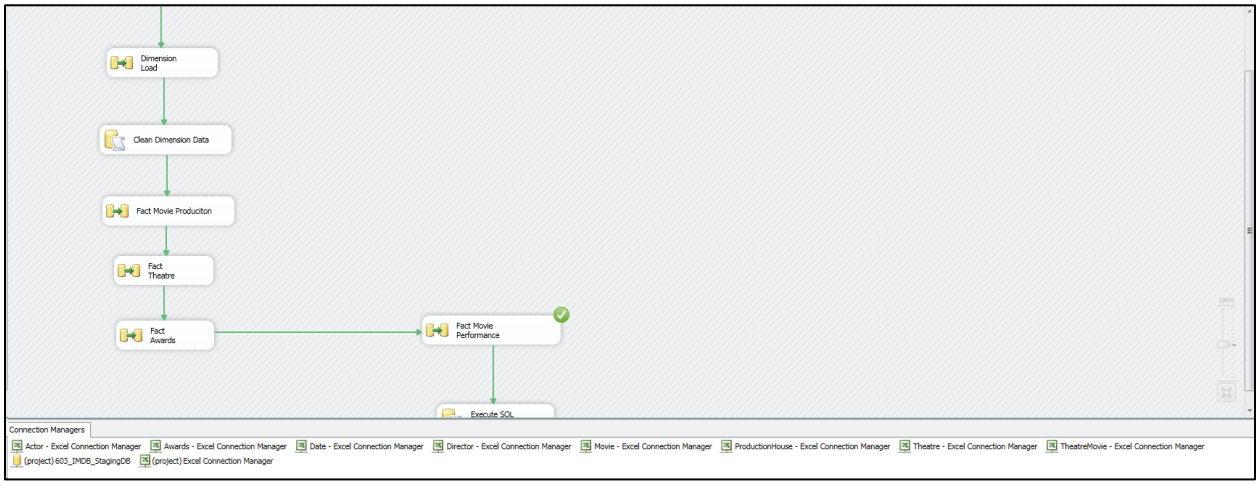


Step 5: Successful execution of all Lookups:



Step 6: Executing fact- Awards Control Flow:





8. Business Intelligence Reporting

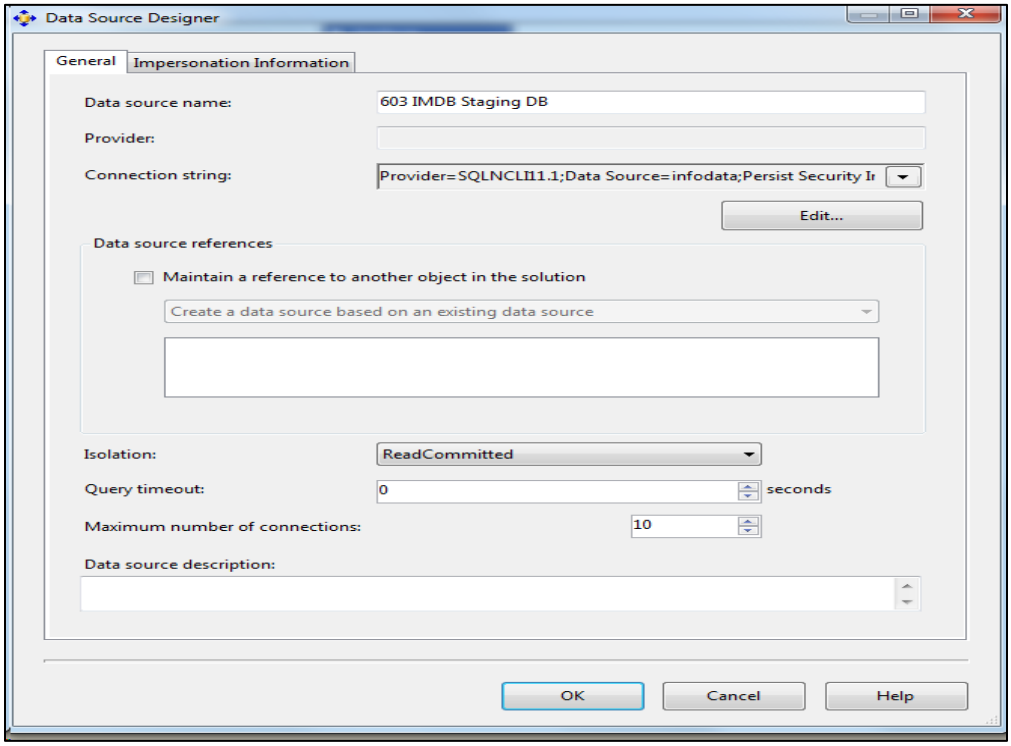
“Business intelligence (BI) is a technology-driven process for analyzing data and presenting actionable information to help corporate executives, business managers and other end users make more informed business decisions. BI encompasses a variety of tools, applications and methodologies that enable organizations to collect data from internal systems and external sources, prepare it for analysis, develop and run queries against the data, and create reports, dashboards and data visualizations to make the analytical results available to corporate decision makers as well as operational workers.”[4]

Reporting Tool	Data Mart	Questions Answered
SSRS over SSAS	Movie Performance	4
SSRS over SSAS	Production Performance	1
SSRS over SSAS	Awards Distributions	1
SSRS over SSAS	Theatre Performance	2

8.1 Data Mart creation using SSAS

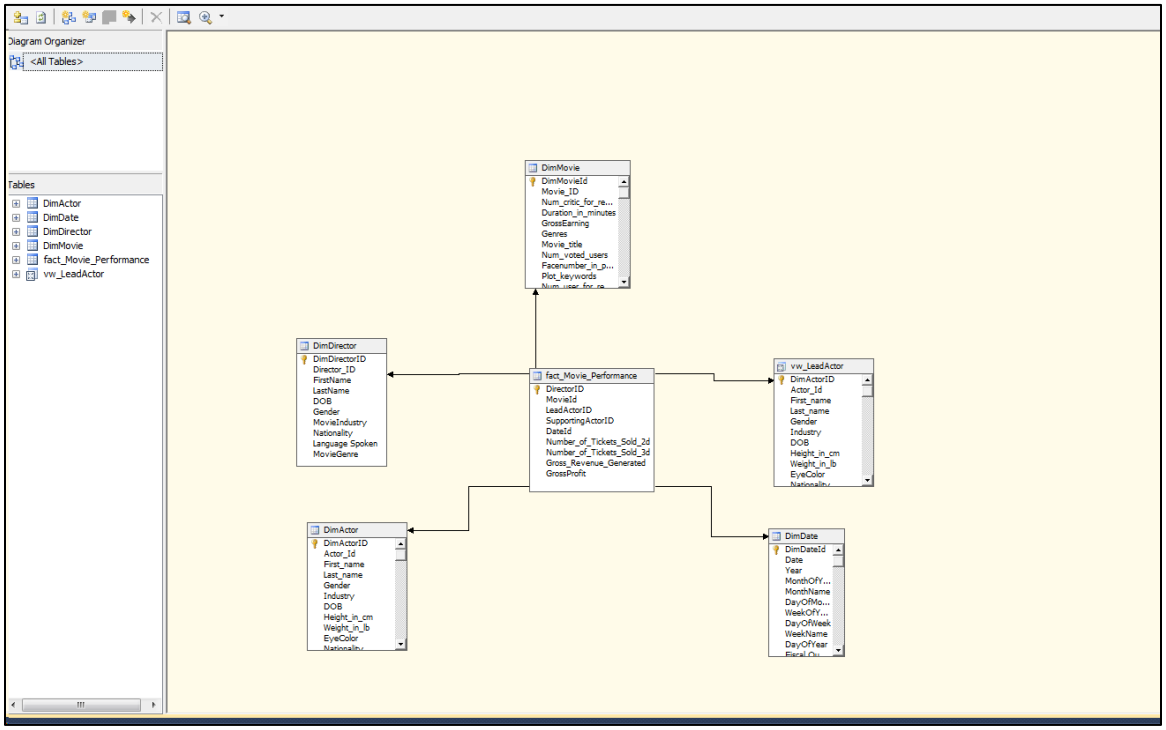
Following are the screenshots for systematic creation of a data mart.

Establishing Data Source for making cubes

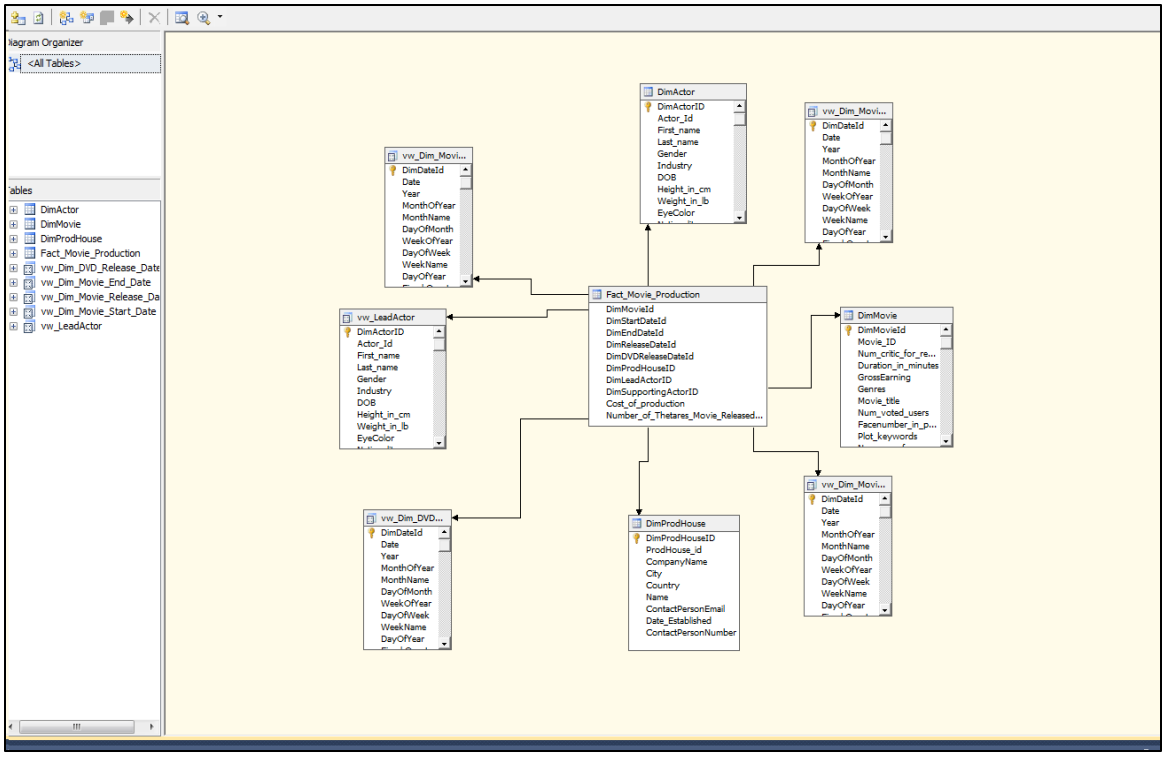


Creating Data Source views for 4 Data Marts

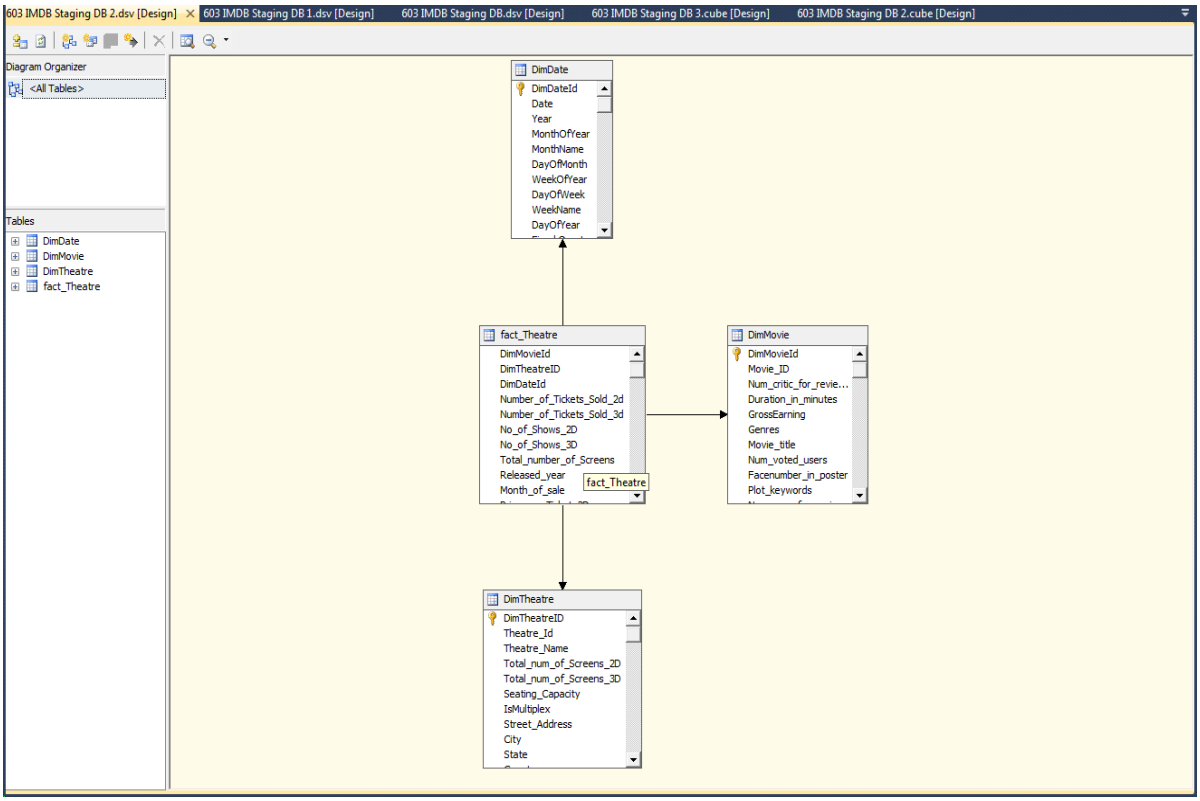
Movie Performance



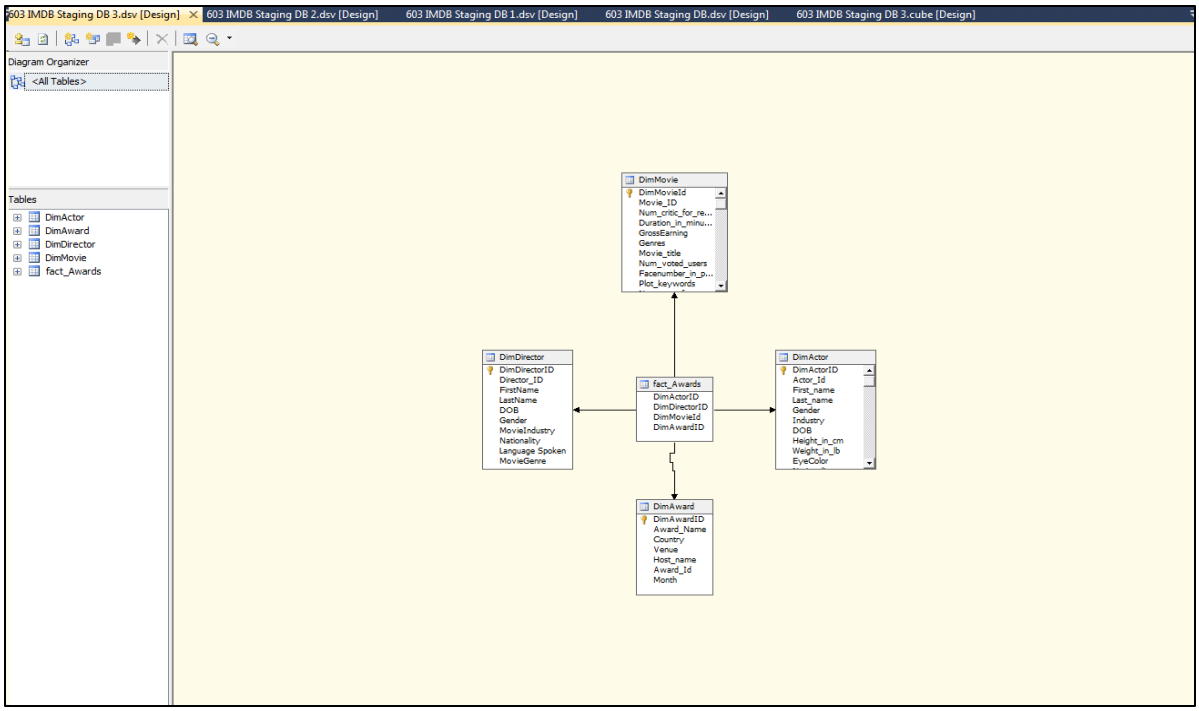
Production House Performance



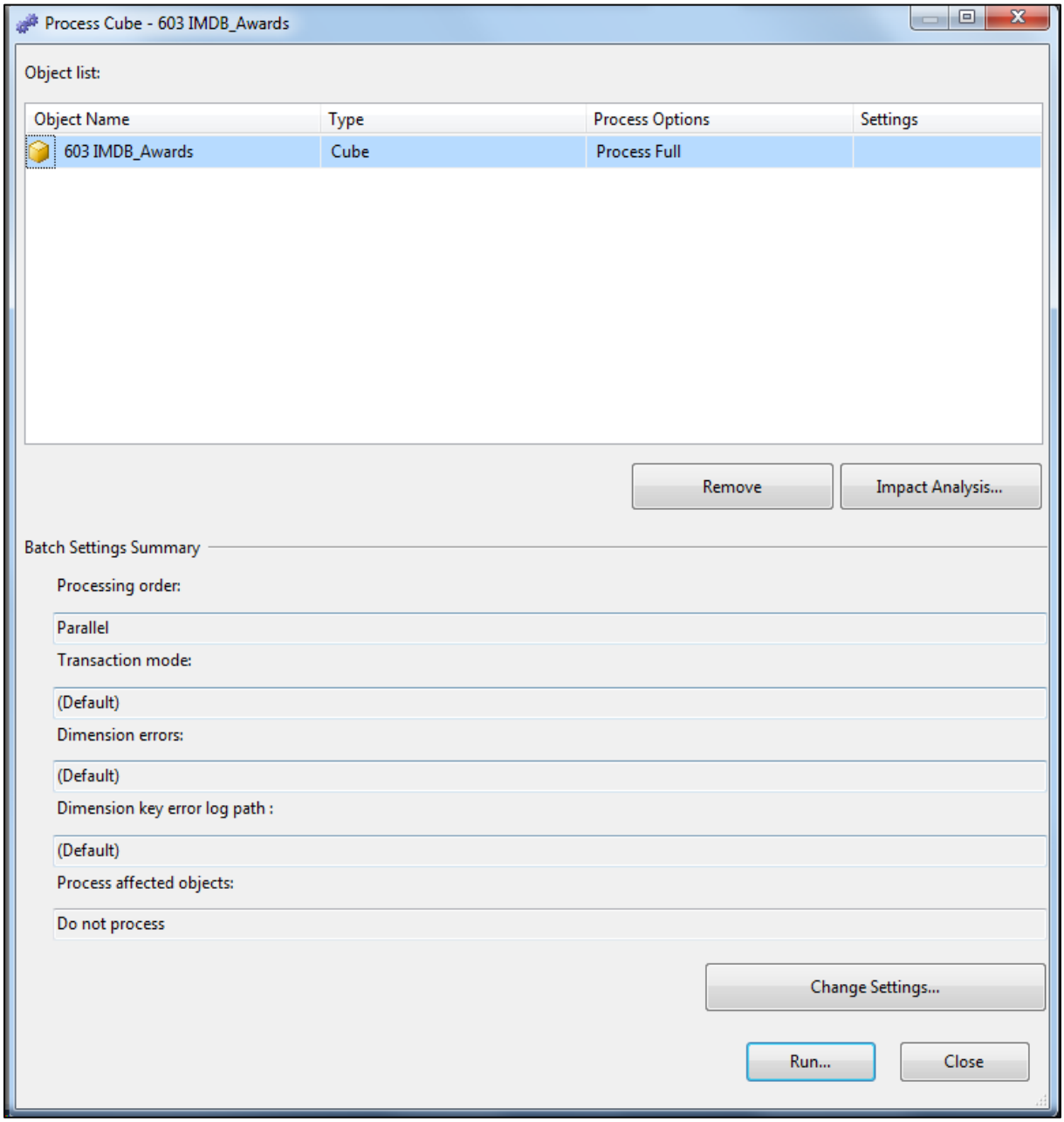
Theatre Performance

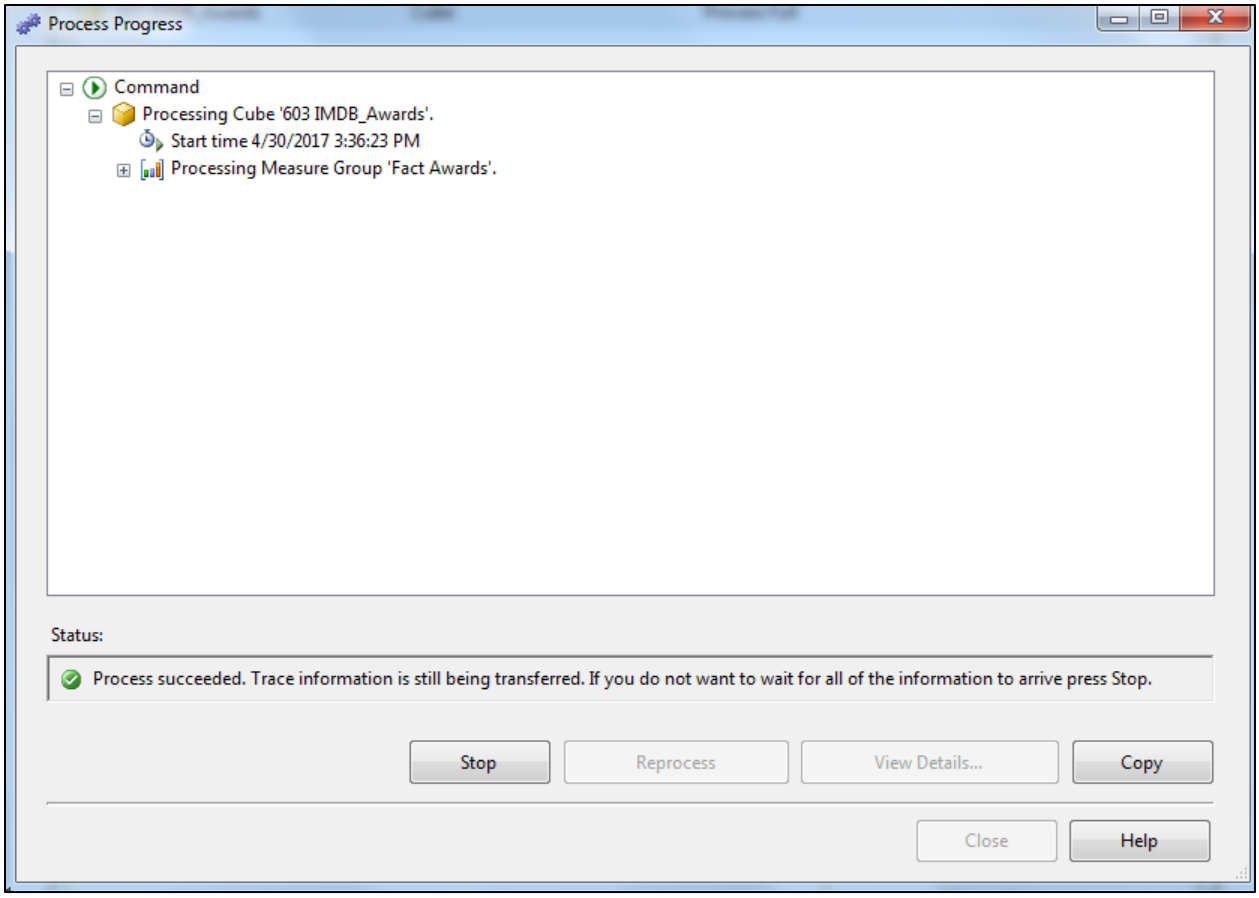


Awards Distribution



Following steps are followed for the processing of the cubes:

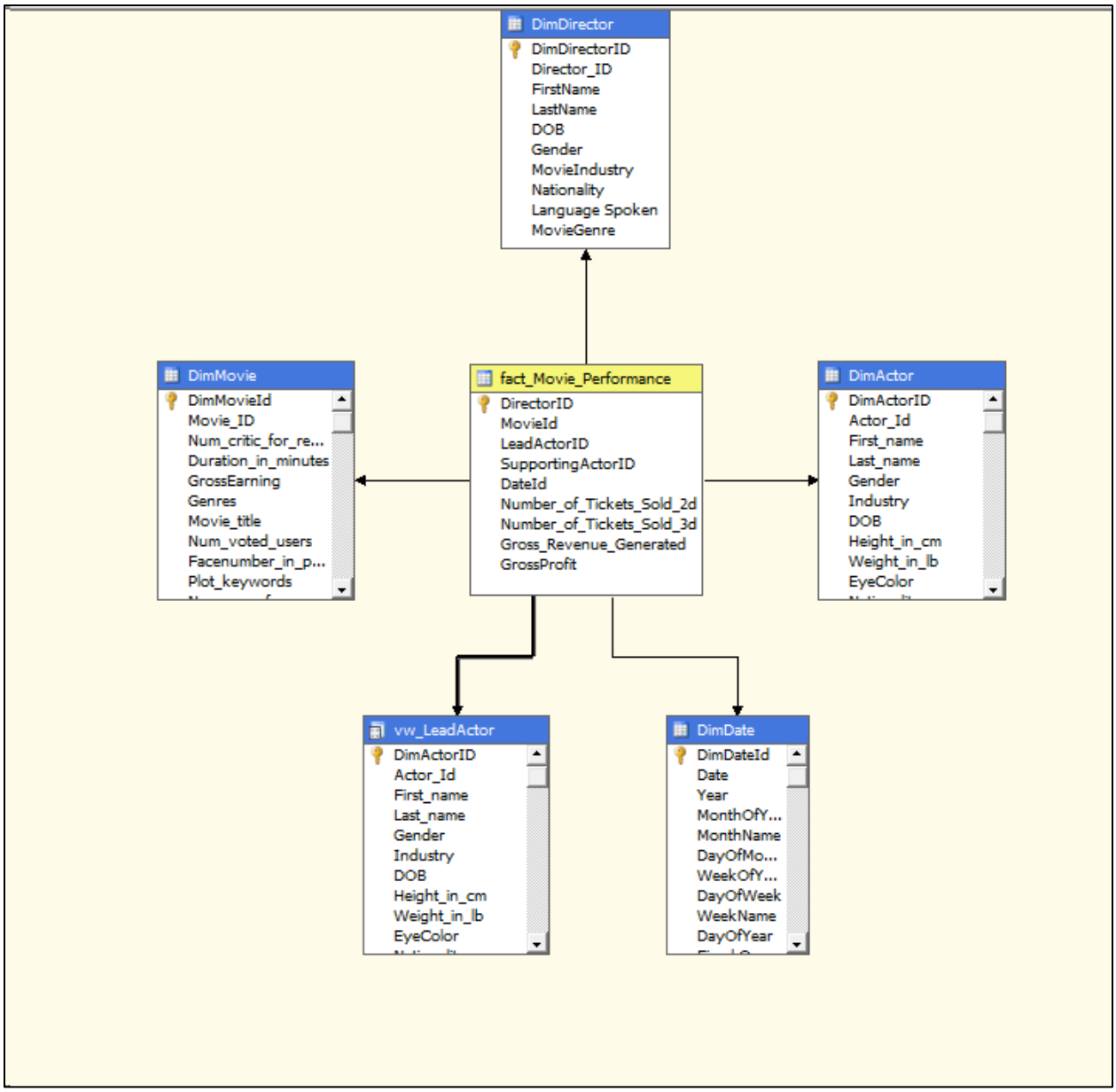



Process Progress
Command
Processing Cube '603 IMDB_Awards'.
Start time 4/30/2017 3:36:23 PM
Processing Measure Group 'Fact Awards'.
Status:
Process succeeded. Trace information is still being transferred. If you do not want to wait for all of the information to arrive press Stop.
Stop
Reprocess
View Details...
Copy
Close
Help

8.2 Report building from individual Data Mart is SSRS

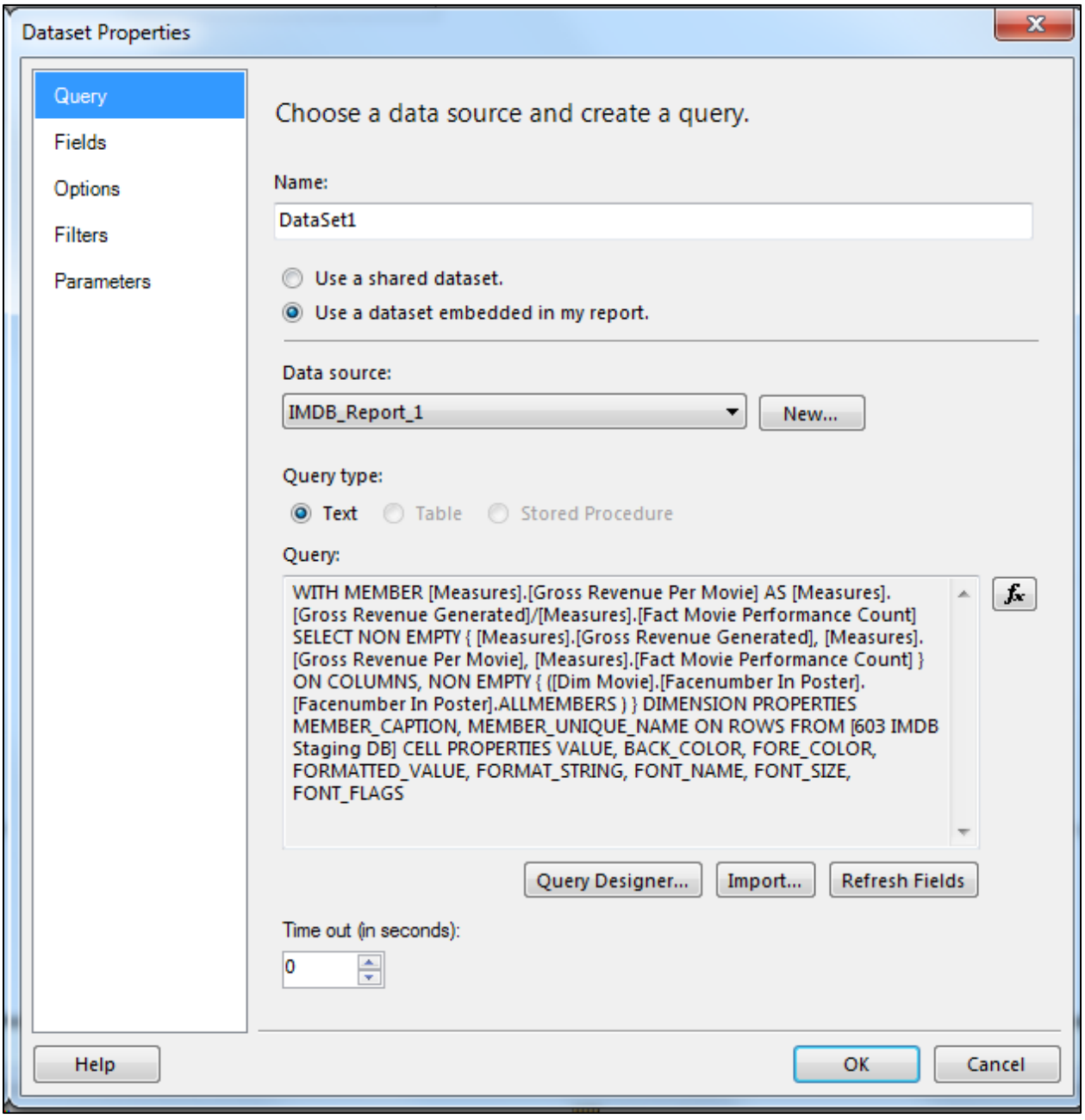
8.2.1 Movie Performance

Cube Name – IMDB_Movie_Performance.cube



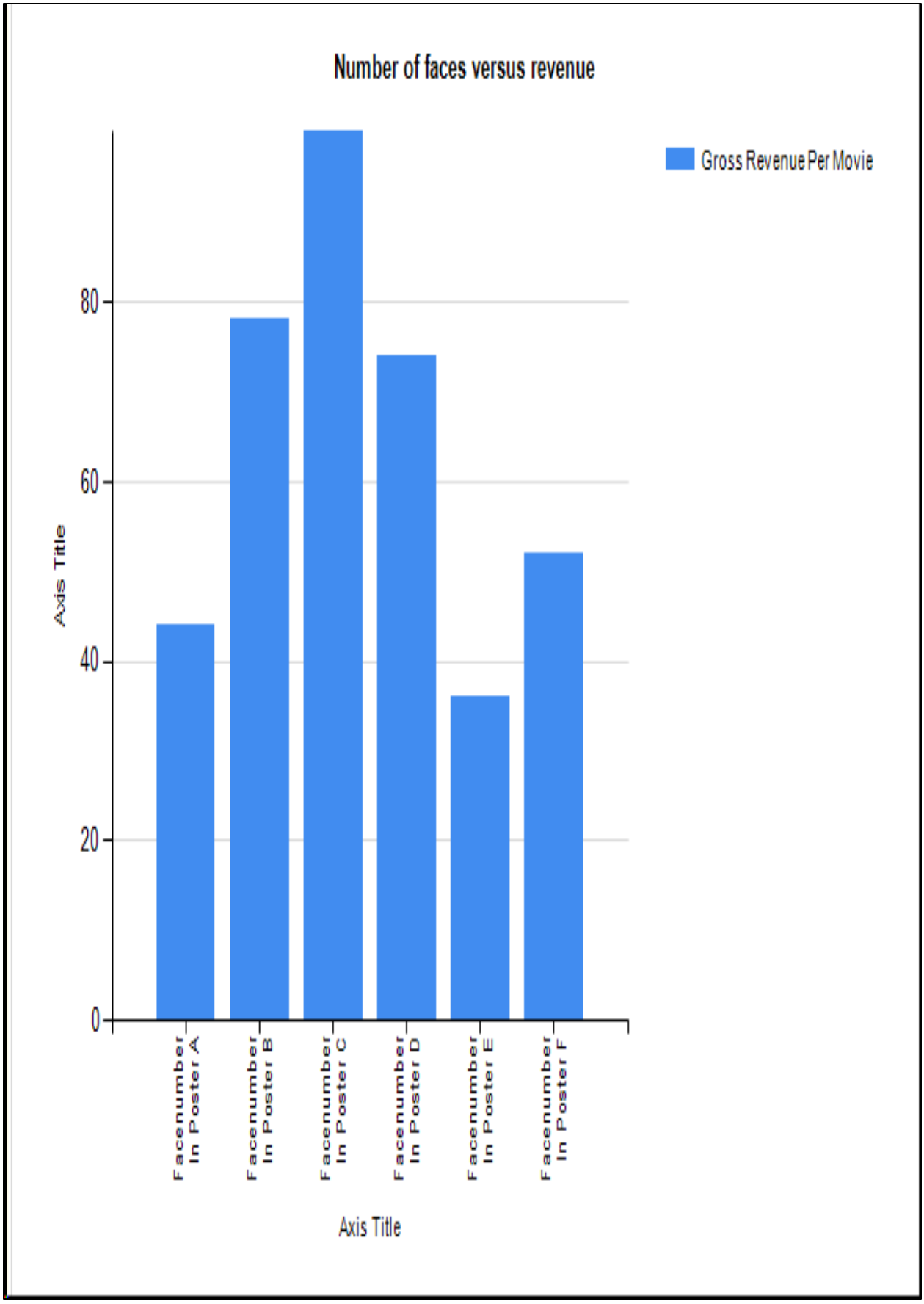
Business Questions Catered

1. **What is the impact of number of faces in a poster on generated revenue generated per movie?**

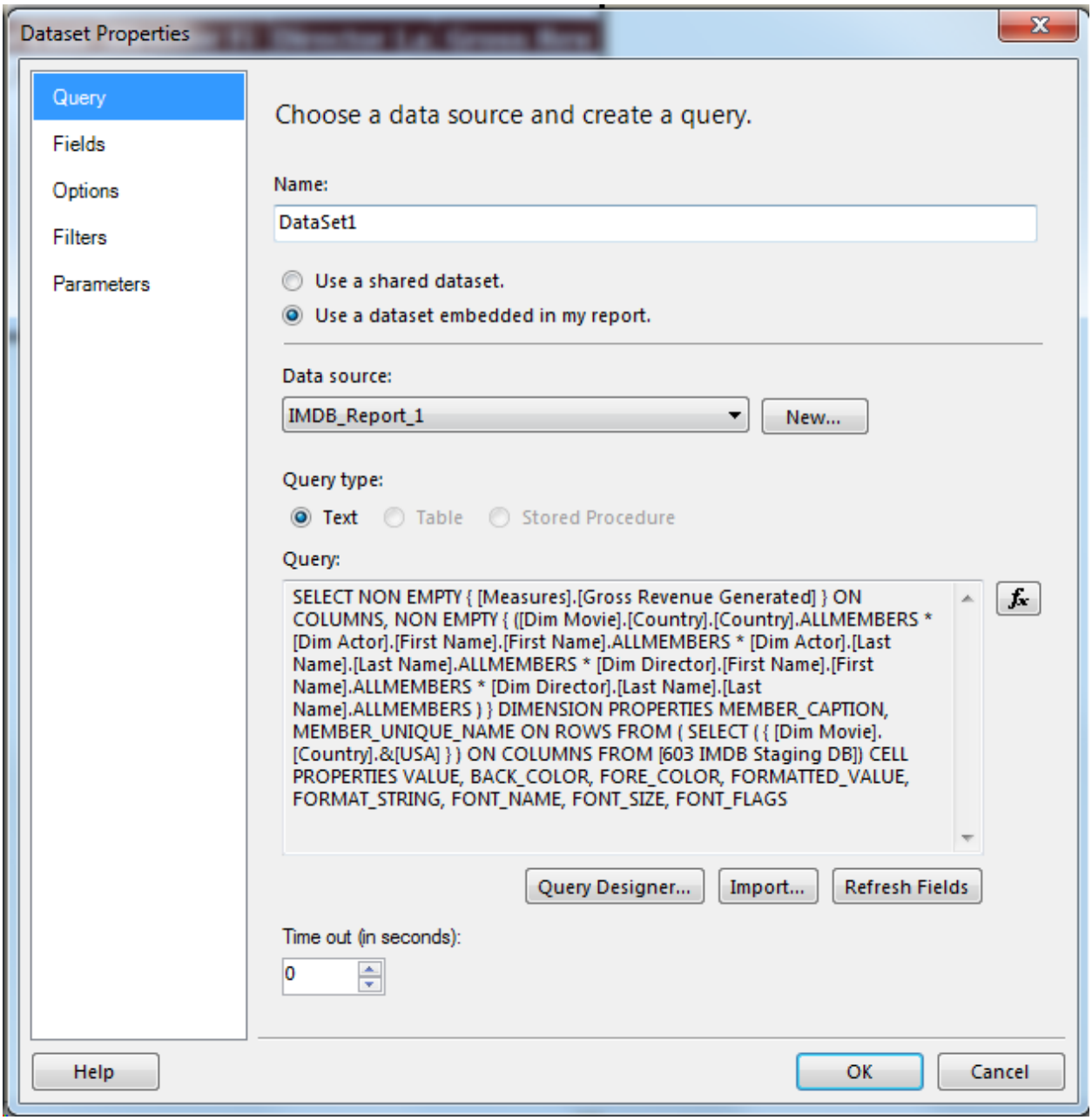


Impact of Number of faces on posters

Facenumber In Poster	Gross Revenue Per Movie
0	
	1085057389.43182
1	
	1141149531.53333
2	
	1098048246.625
3	
	954025864
4	
	1026812223.25
5	
	1000615800
6	
	1319895106.5
7	
	1271518971
8	
	1402777497


Number of faces versus revenue
Gross Revenue Per Movie
Axis Title
80
60
40
20
0
Facenumber In Poster A
Facenumber In Poster B
Facenumber In Poster C
Facenumber In Poster D
Facenumber In Poster E
Facenumber In Poster F
Axis Title

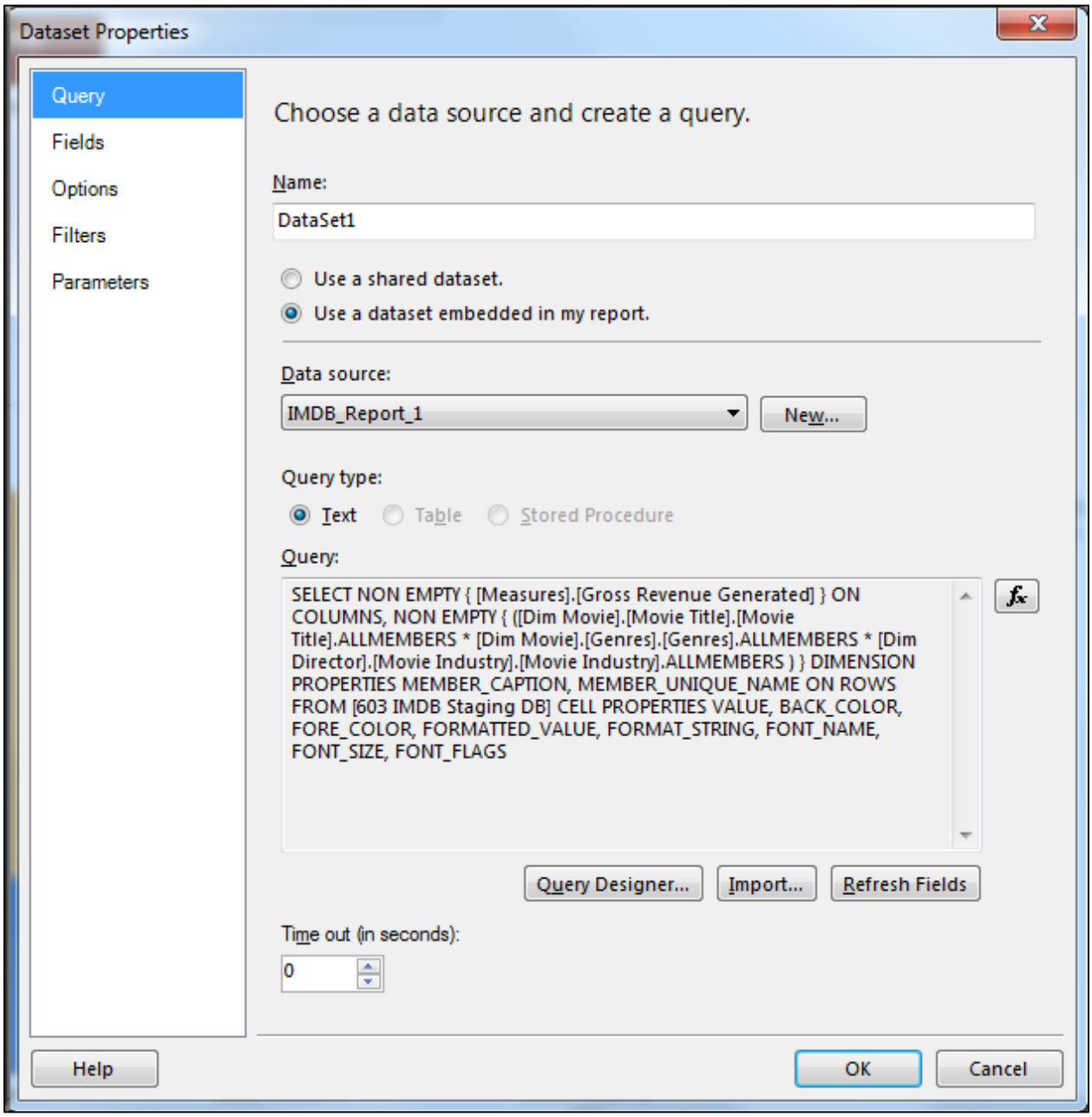
2. **What are the best actor director combination for a movie generating highest revenue based on each country?**



Best Actor-Director Pair

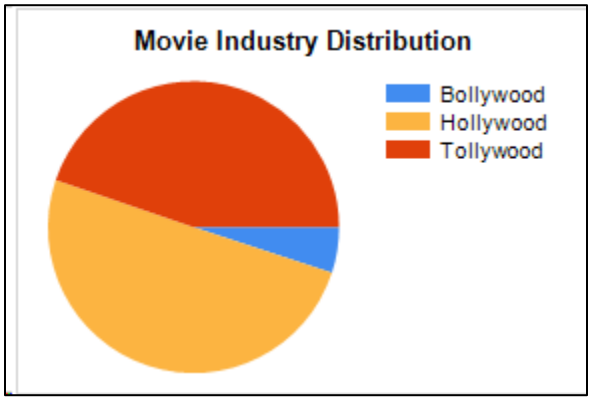
Country	Actor First Name	Actor Last Name	Director First Name	Director Last Name	Gross Revenue Generated
USA	Aaron	Arnold	Joe	Camp	810874750
USA	Alice	Green	Emile	Ardolino	1066016104
USA	Amanda	Campbell	Ildik\^o	Enyedi	262938280
USA	Andrea	Alvarez	Joel	Coen	286631580
USA	Anne	Elliott	Claire	Denis	1330752190
USA	Annie	Gibson	\Un	Axelman	1353773968
USA	Brandon	Robinson	Jaques W.	Benoit	1465013996
USA	Brian	Wilson	Ethan	Coen	548069430
USA	Carl	Jenkins	Nora	Ephron	1355309646
USA	Carol	Warren	Ren\'e jr.	Cardona	914269918
USA	Catherine	Riley	Andy	Cadiff	256007641
USA	Christina	Andrews	Colin	Bucksey	1267882206
USA	Christopher	Miller	Stanley	Donen	1403880043
USA	Cynthia	Cruz	Albert R.	Broccoli	1256749889
USA	Cynthia	Morris	Samuel	Bischoff	1210286588
USA	Daniel	James	Jack	Arnold	1331985835
USA	Deborah	James	Ray	Austin	1300498380
USA	Deborah	Perry	John G.	Adolfi	1402777497
USA	Dennis	Banks	Roland	Emmerich	1346841097
USA	Donna	Greene	Assi	Dayan	688996855
USA	Dorothy	Parker	Rachel	Liebling	747167084
USA	Douglas	Stone	Cecil B.	DeMille	286224798

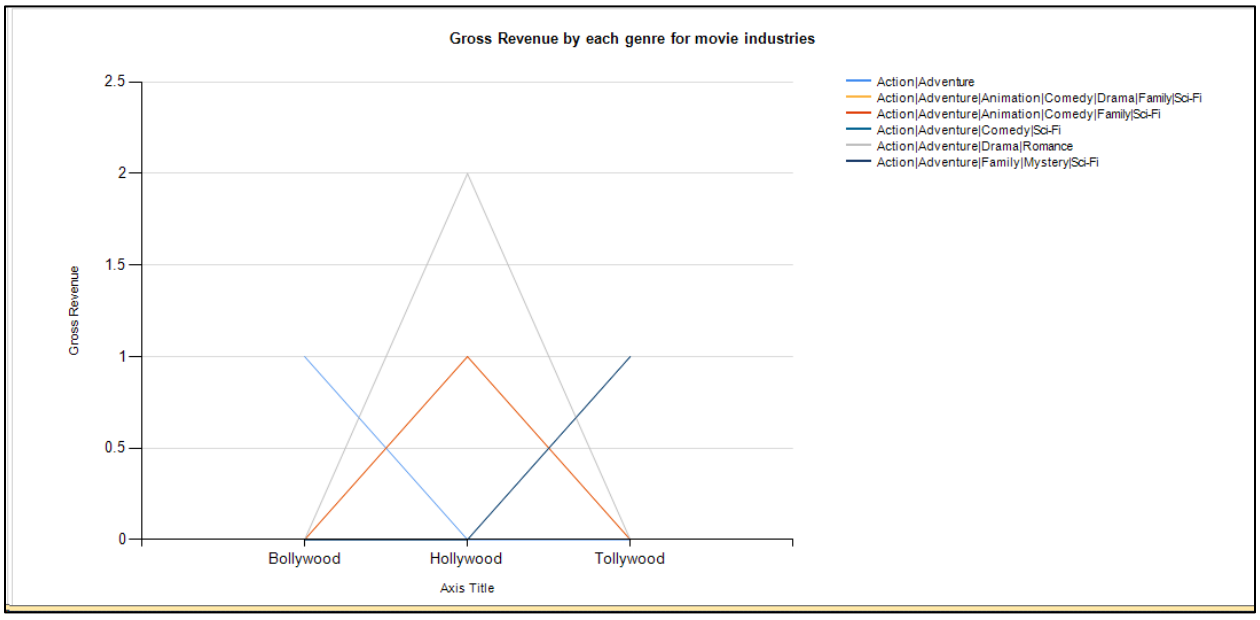
3. **What are the highest grossing movie industries and what genres are the highest grossers in each industry?**



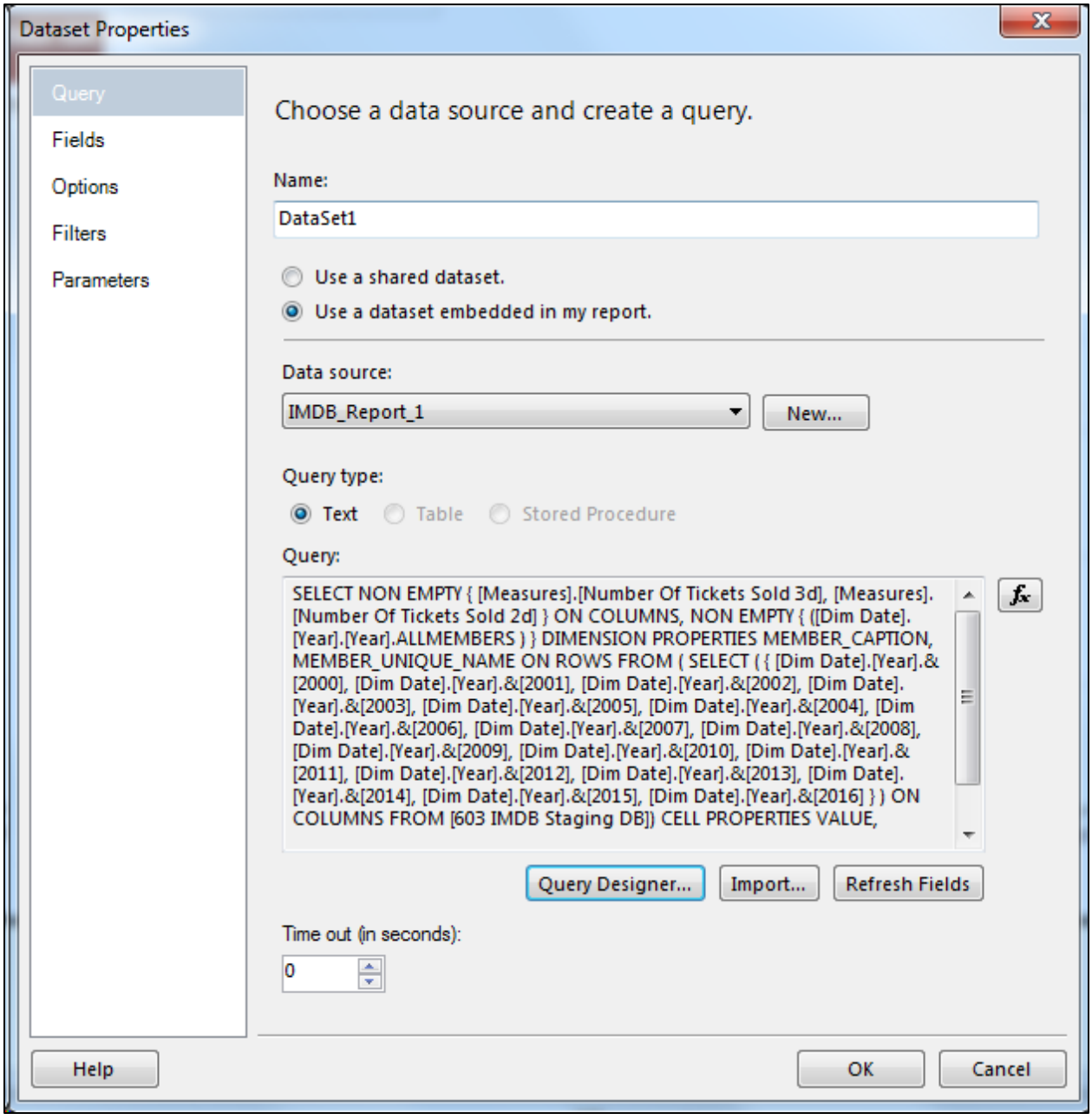
Highest grossing movie industry

Movie Industry	Gross Revenue Generated	Genres
⊞ Bollywood	31468526391	
⊞ Hollywood	41655447371	
⊞ Tollywood	29557921743	



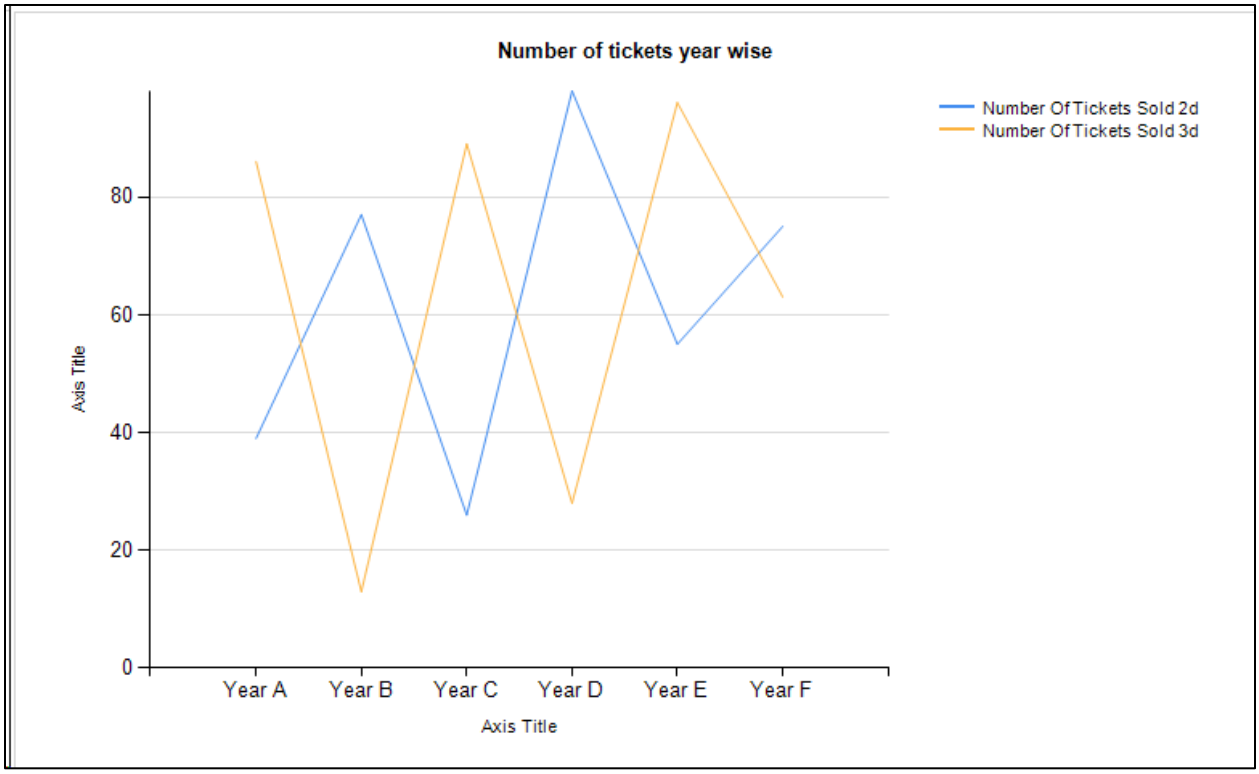


4. **What is the popularity trend for 2D movies vs 3D movies in last one decade?**



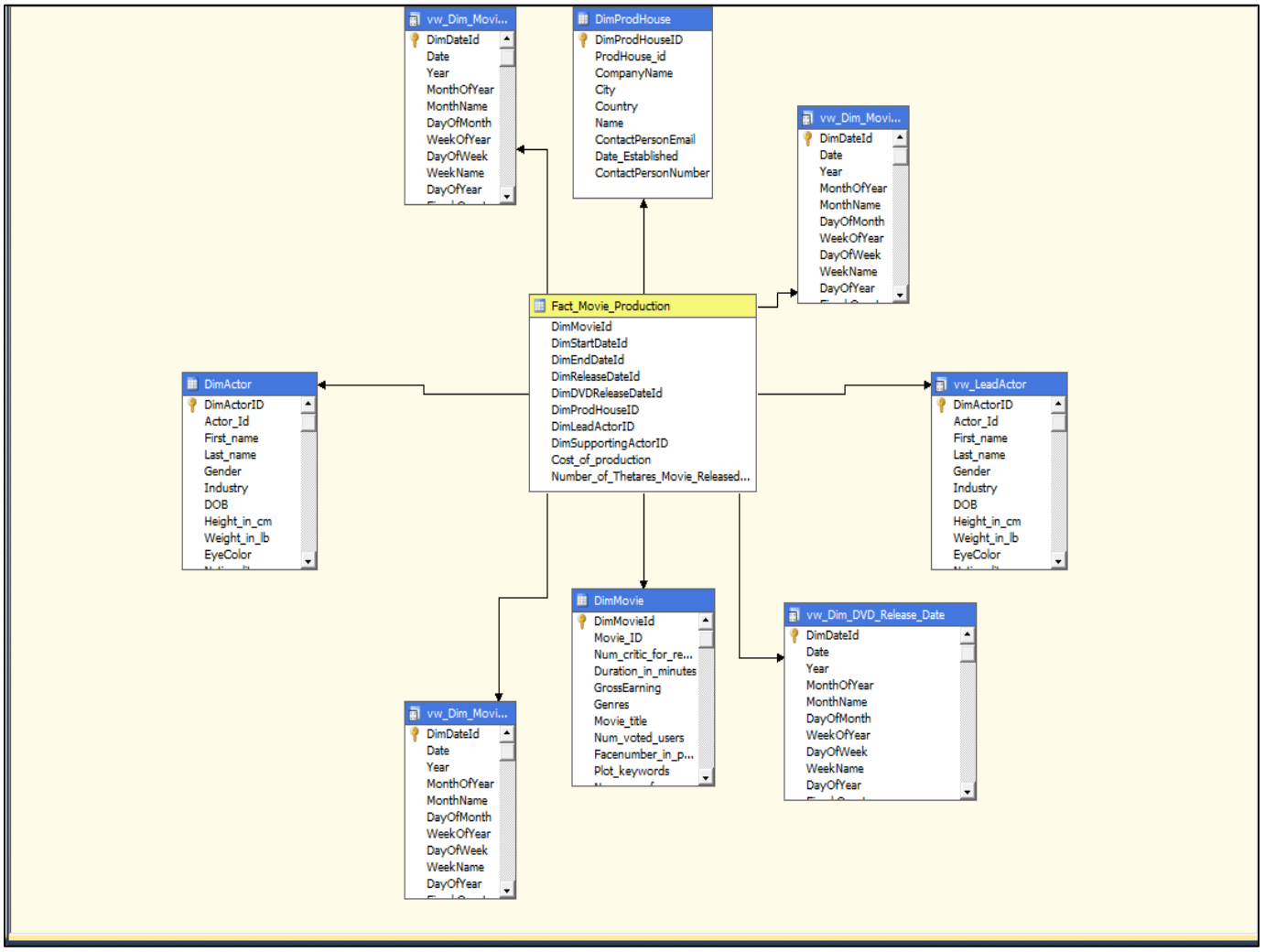
Yearly sale of 2D versus 3D tickets

Year	Number Of Tickets Sold 2d	Number Of Tickets Sold 3d
2000		
	101833574	76381489
2001		
	102270488	75348011
2002		
	199285292	142502961
2003		
	20141196	16031309
2005		
	40335030	28910700
2007		
	102291940	75014327
2008		
	120602490	88965504
2009		
	142977417	107376531
2010		
	203700876	156387870
2011		
	97423740	76451790
2015		
	101208955	76126948
2016		
	101191626	77199176


Number of tickets year wise
Number Of Tickets Sold 2d
Number Of Tickets Sold 3d
Axis Title
80
60
40
20
0
Year A
Year B
Year C
Year D
Year E
Year F
Axis Title

i. Production House Performance

Cube Name – IMDB_Production_House_Performance.cube



Business Questions Catered

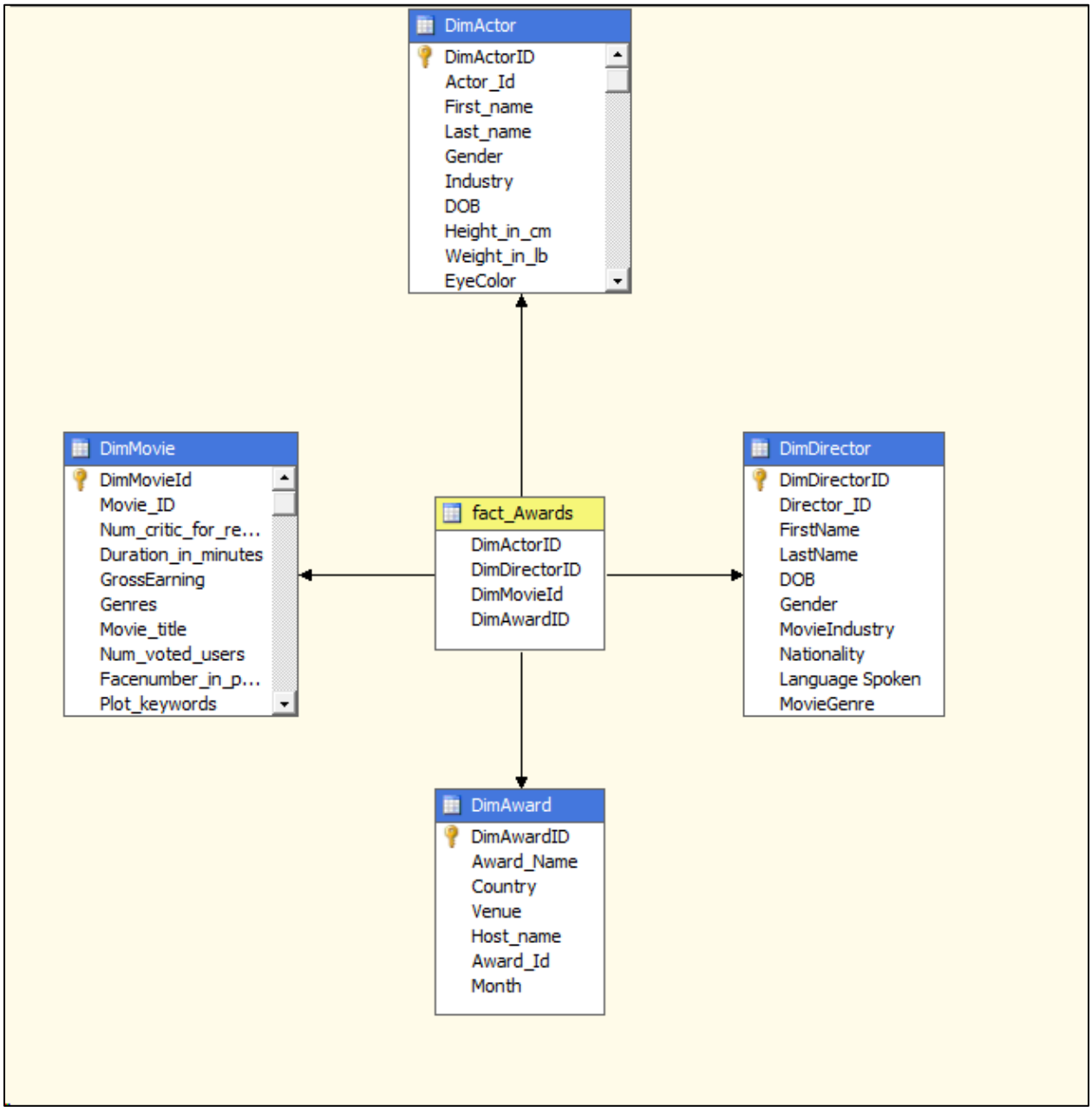
1. **What pair of actors casted in a movie result in higher cost of production?**

Cost of Production for Actors Combination

Lead Actor First Name	Lead Actor Last Name	Supporting Actor First Name	Supporting Actor Last Name	Cost Of Production
Aaron	Arnold	Emily	Wells	225000000
Alan	Campbell	Lori	Medina	250000000
Alice	Green	Eric	Carroll	200000000
Amanda	Campbell	Gloria	Burton	175000000
Andrea	Alvarez	David	Perez	175000000
Anna	Black	Maria	Bryant	200000000
Anne	Elliott	Beverly	Jackson	170000000
Annie	Gibson	Kevin	Ross	250000000
Billy	Wagner	Chris	Nichols	180000000
Bobby	Allen	Fred	Mccoy	200000000
Brian	Wilson	Sandra	Medina	170000000
Carl	Jenkins	Gerald	Green	175000000
Carol	Warren	Stephen	Franklin	225000000
Catherine	Riley	Douglas	Stone	200000000
Christina	Andrews	Rebecca	Morgan	180000000
Christopher	Miller	Kathleen	Simmons	185000000
Cynthia	Cruz	Harold	Elliott	258000000
Cynthia	Morris	Jean	Little	190000000
Deborah	James	Mildred	Hughes	140000000
Deborah	Perry	Joseph	Sullivan	170000000

ii. Awards Distribution

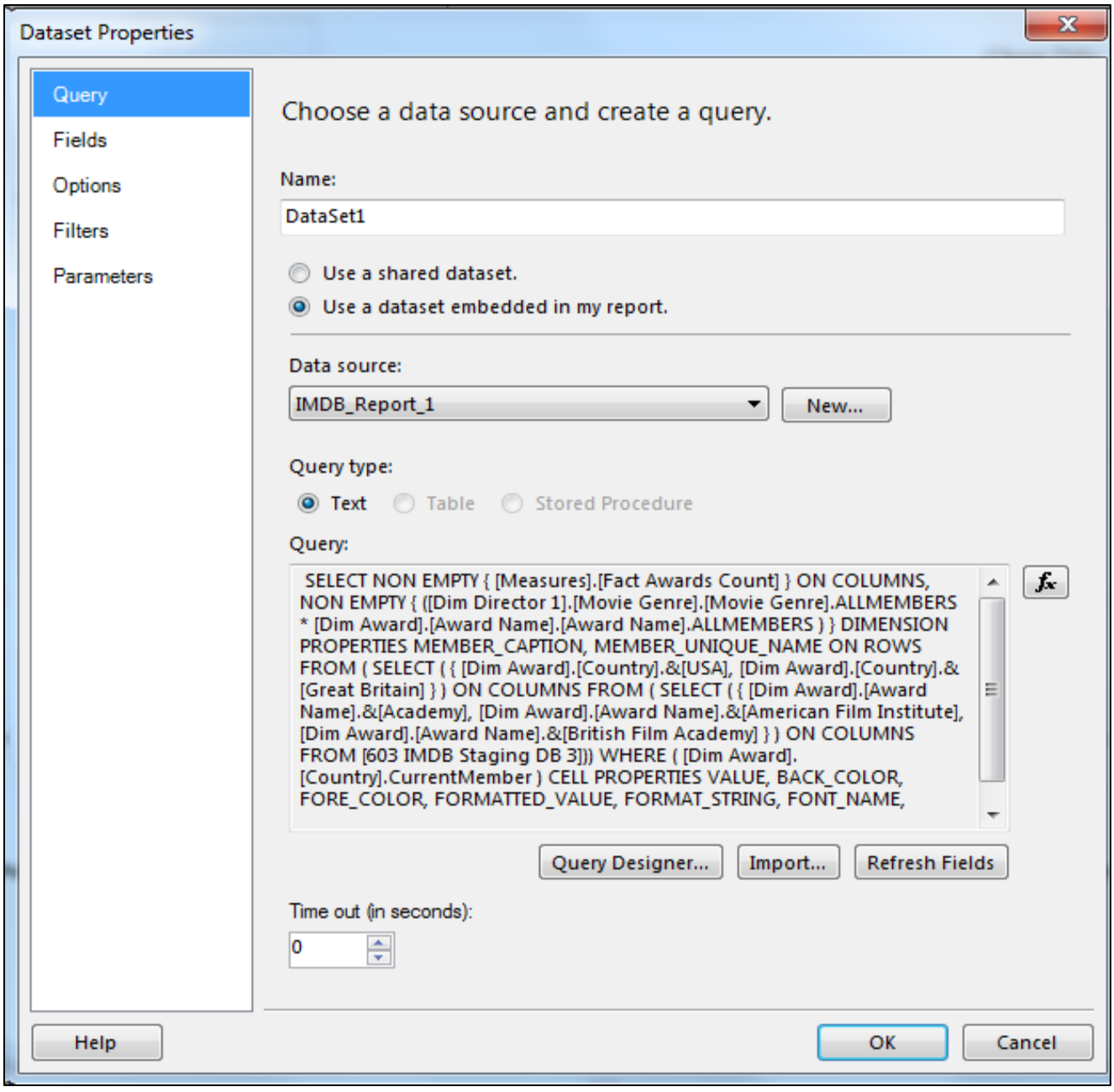
Cube Name – IMDB_Awards_distribution.cube



Business Questions Catered

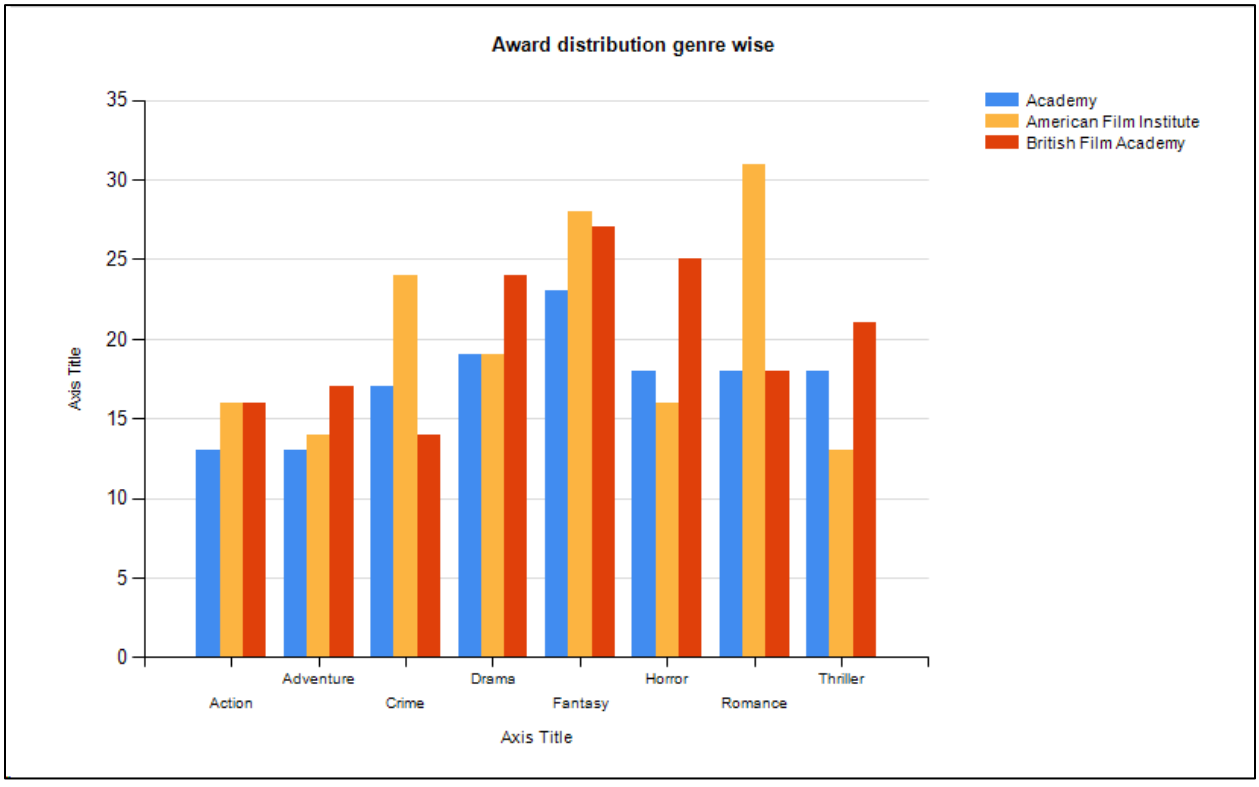
1. **What are the most popular genre, which won for major awards in a particular country?**

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Award	Award Name	Equal	(Academy, American Film Institute, British Film Academy)	
Dim Award	Country	Equal	(USA, Great Britain)	
<Select dimension>				



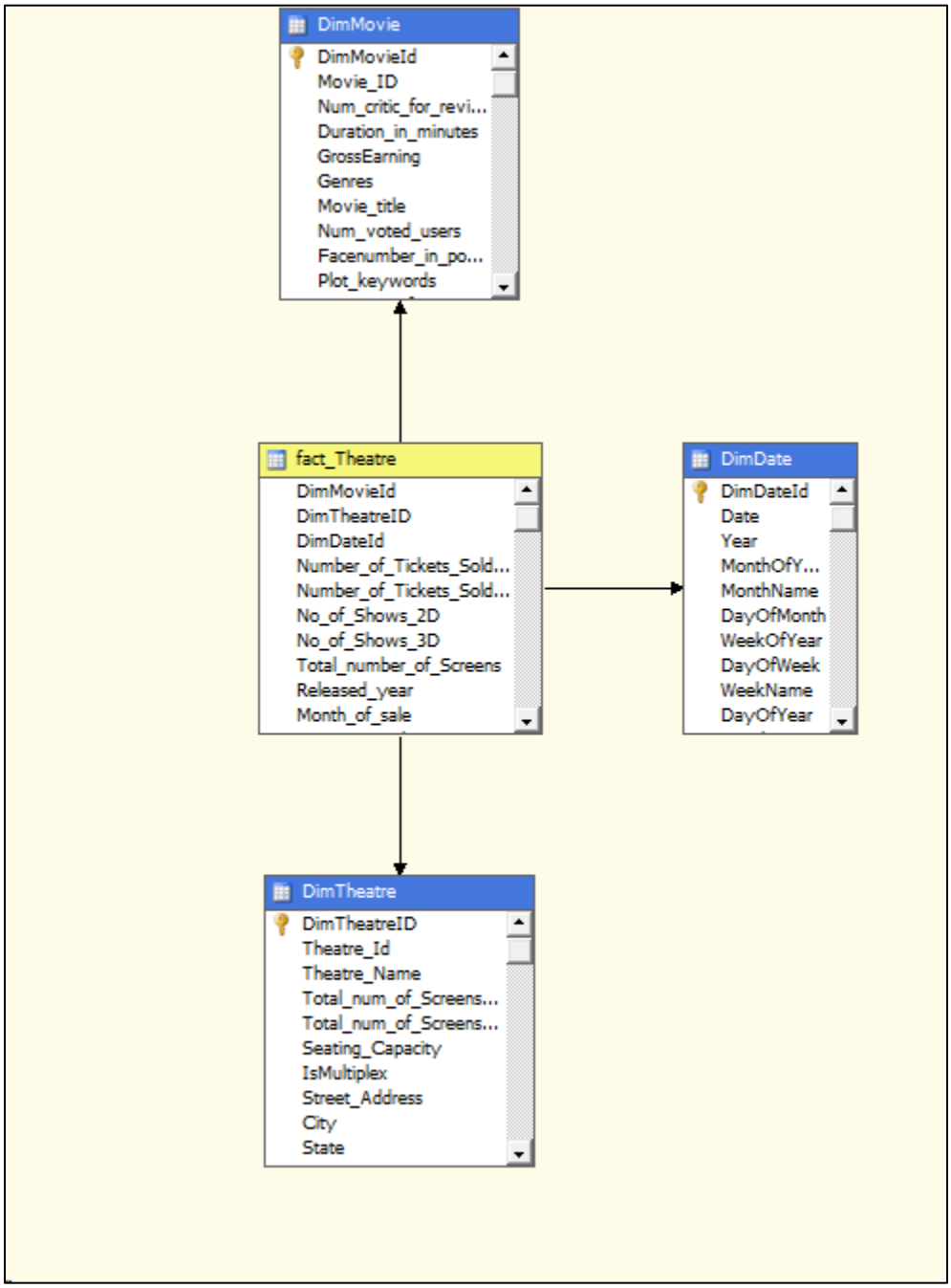
Award Distribution

Award Name	Awards Count
⊞ Academy	13
⊞ American Film Institute	16
⊞ British Film Academy	16



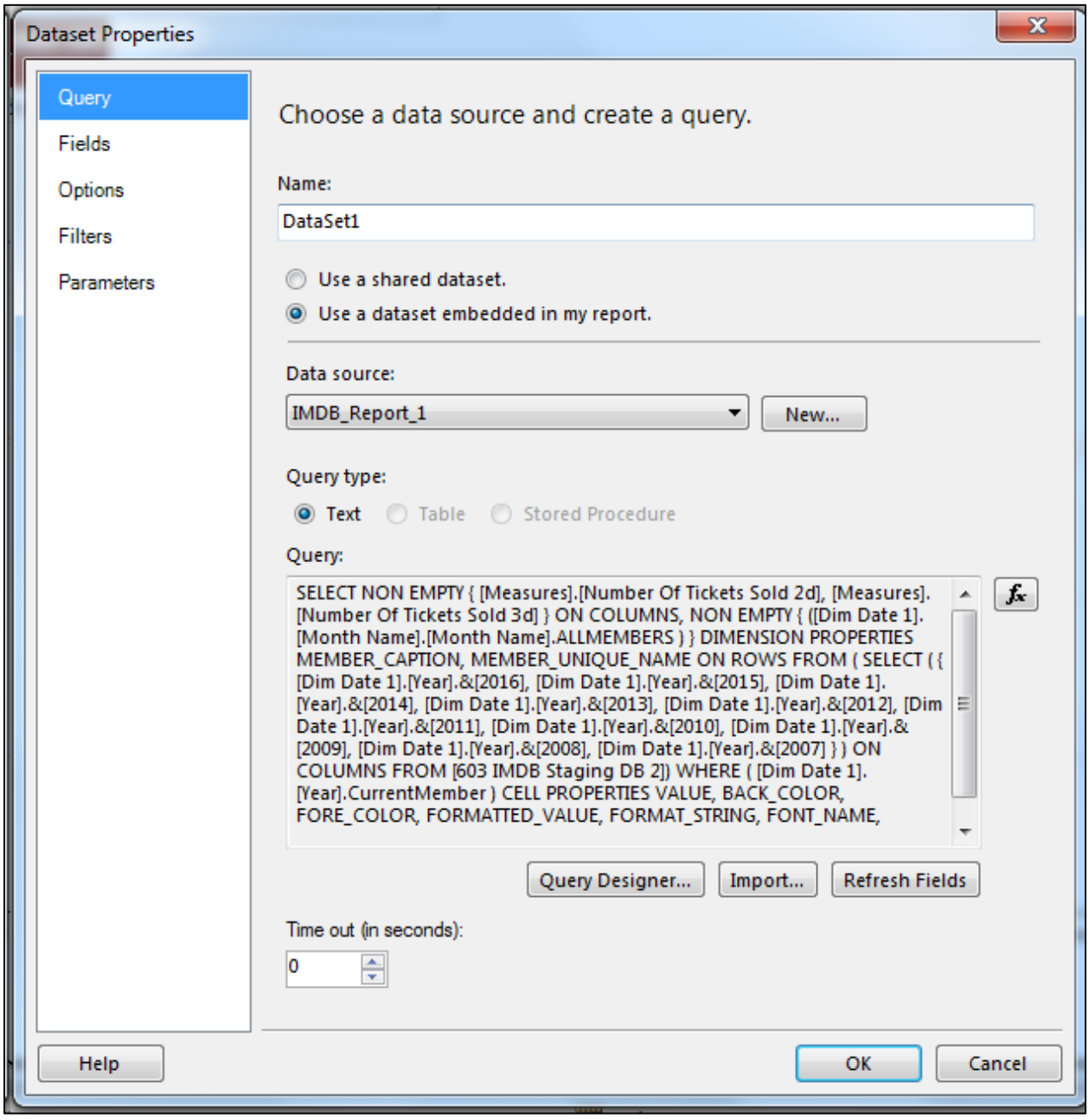
iii. Theatre Performance

Cube Name – IMDB_Theatre_Performance.cube



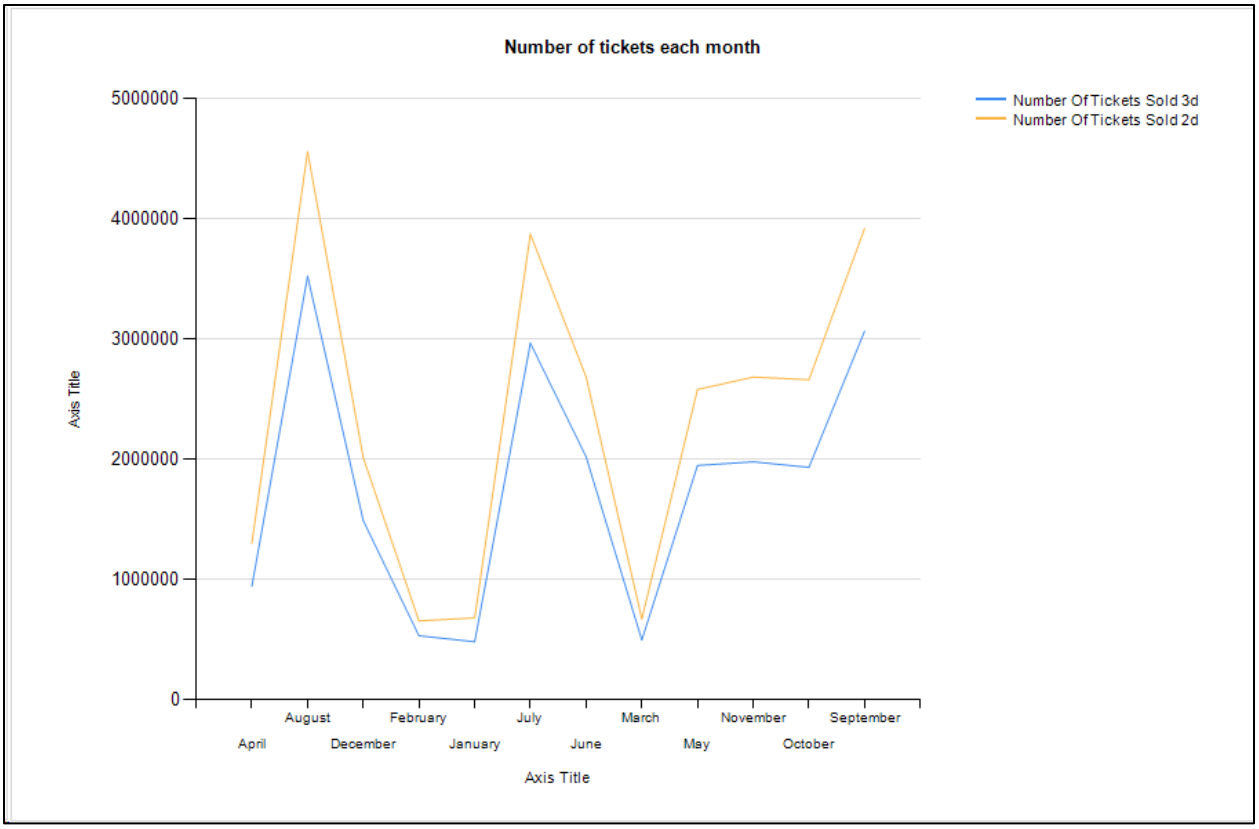
Business Questions Catered

1. **What is the month wise trend of tickets sold across theatres? Which month is the most popular in terms of 2D vs 3D tickets?**

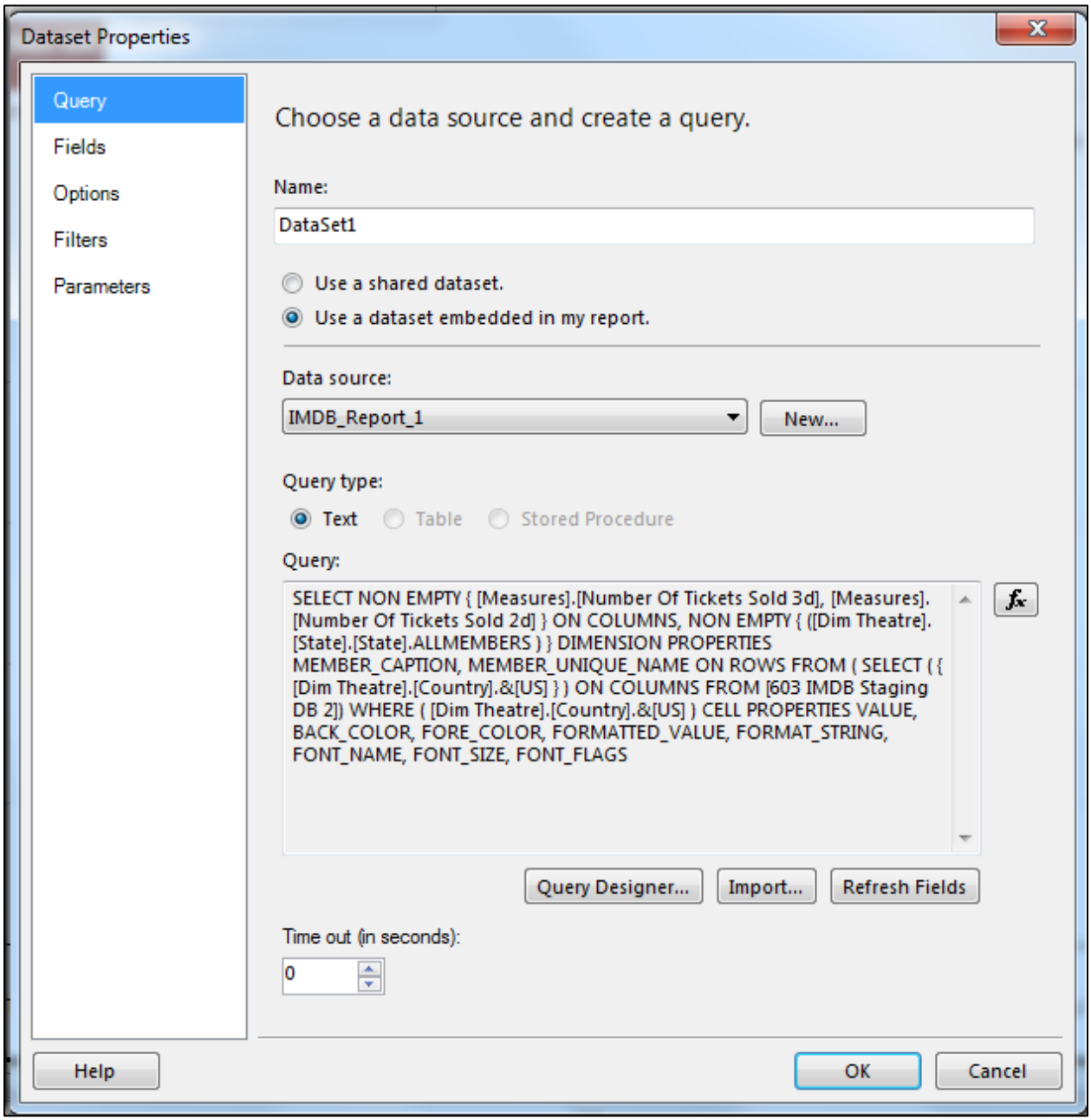


2D versus 3D sale month wise

Month Name	Number Of Tickets Sold 3d	Number Of Tickets Sold 2d
April		
	944611	1301058
August		
	3524055	4558851
December		
	1492056	2014745
February		
	535836	658954
January		
	485215	683766
July		
	2965909	3872678
June		
	2022003	2685414
March		
	500224	675130
May		
	1950705	2582446
November		
	1980565	2684712
October		
	1934984	2663063
September		
	3067484	3920101

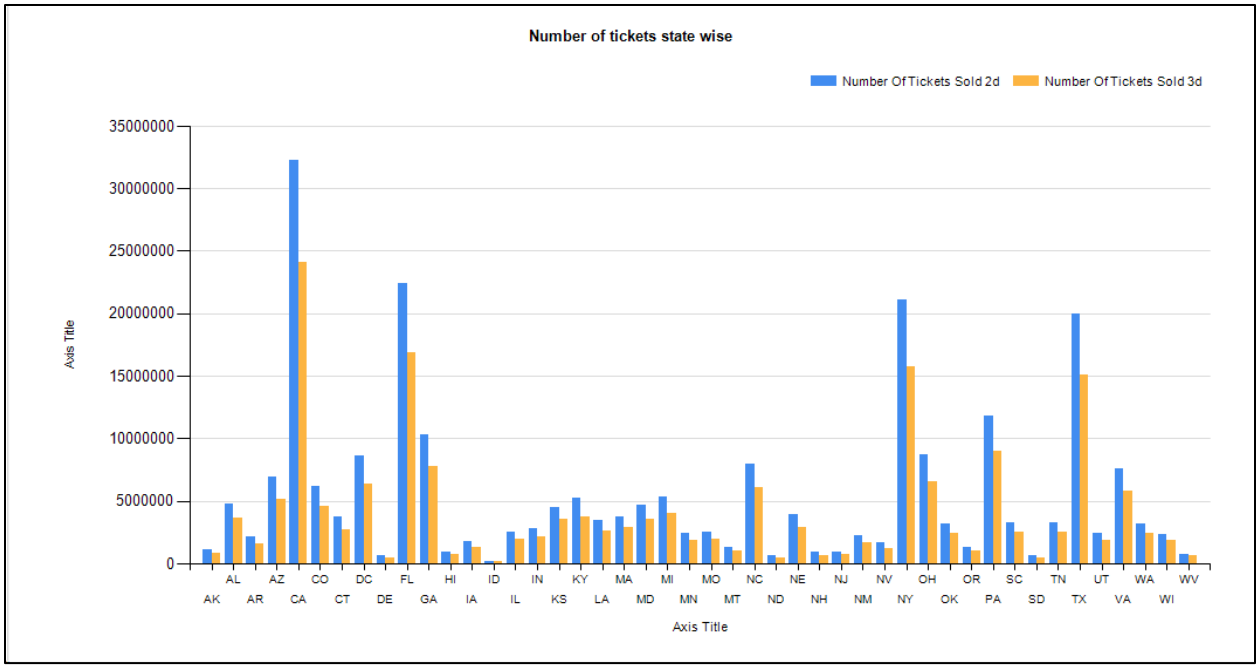

Number of tickets each month
Number Of Tickets Sold 3d
Number Of Tickets Sold 2d
5000000
4000000
3000000
2000000
1000000
0
Axis Title
April
August
December
February
January
July
June
March
May
November
October
September
Axis Title

2. What are the popularity trends of 2D tickets sold versus 3D tickets sold across different states of a country?



2D versus 3D tickets state wise

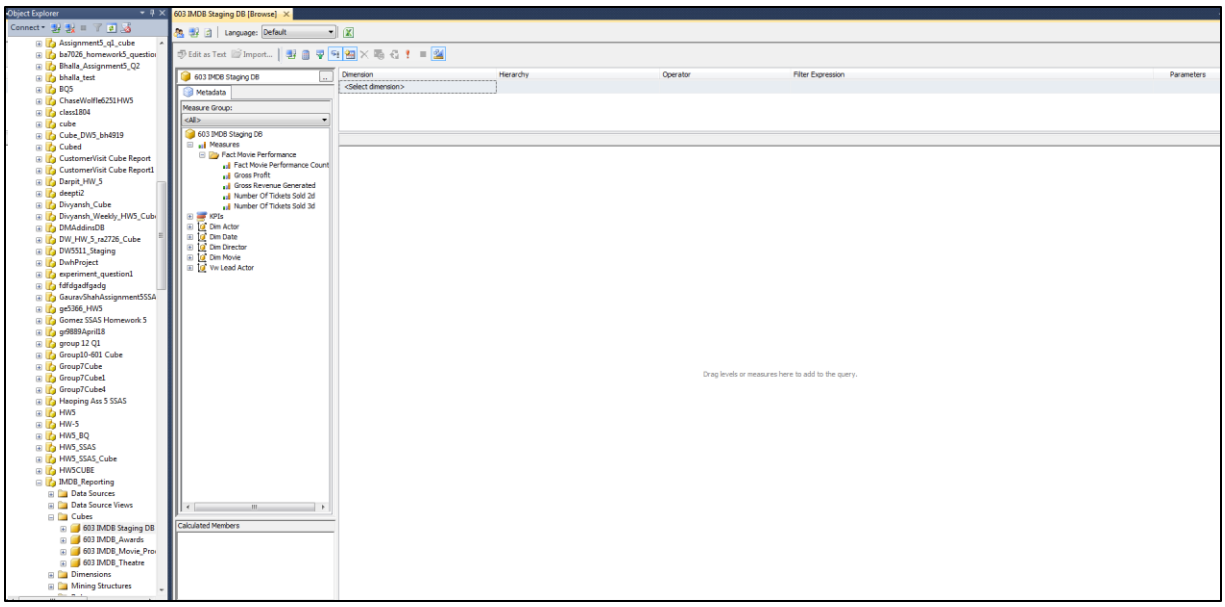
State	Number Of Tickets Sold 2d	Number Of Tickets Sold 3d
AK		
	1124197	876172
AL		
	4751417	3670068
AR		
	2108099	1597006
AZ		
	6893483	5133416
CA		
	32274401	24105487
CO		
	6140241	4571541
CT		
	3715547	2689123
DC		
	8592977	6345123
DE		
	600073	440366
FL		
	22387272	16883630
GA		
	10310191	7766111
HI		


Number of tickets state wise
Number Of Tickets Sold 2d
Number Of Tickets Sold 3d
35000000
30000000
25000000
20000000
15000000
10000000
5000000
0
Axis Title
AL AZ CO DC FL HI ID IN KY MA MI MO NC NE NJ NV OH OR SC TN UT WA WV
AK AR CA CT DE GA IA IL KS LA MD MN MT ND NH NM NY OK PA SD TX VA WI
Axis Title

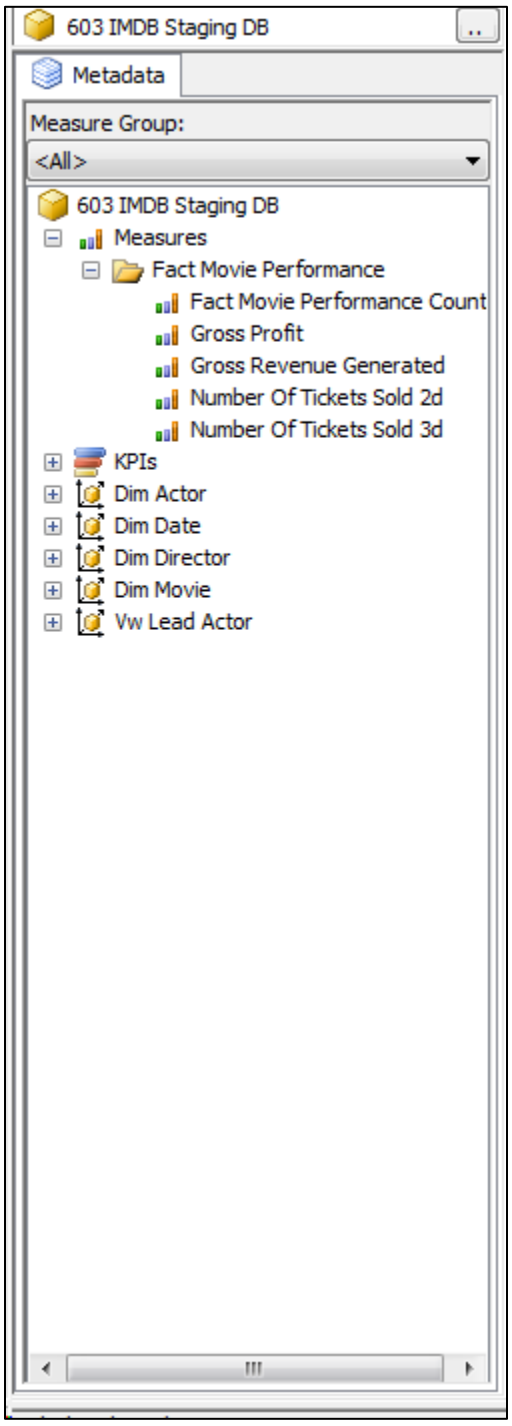
8.3 User Manual for creating Dynamic Reports

8.3.1 Browsing cubes via SSRS

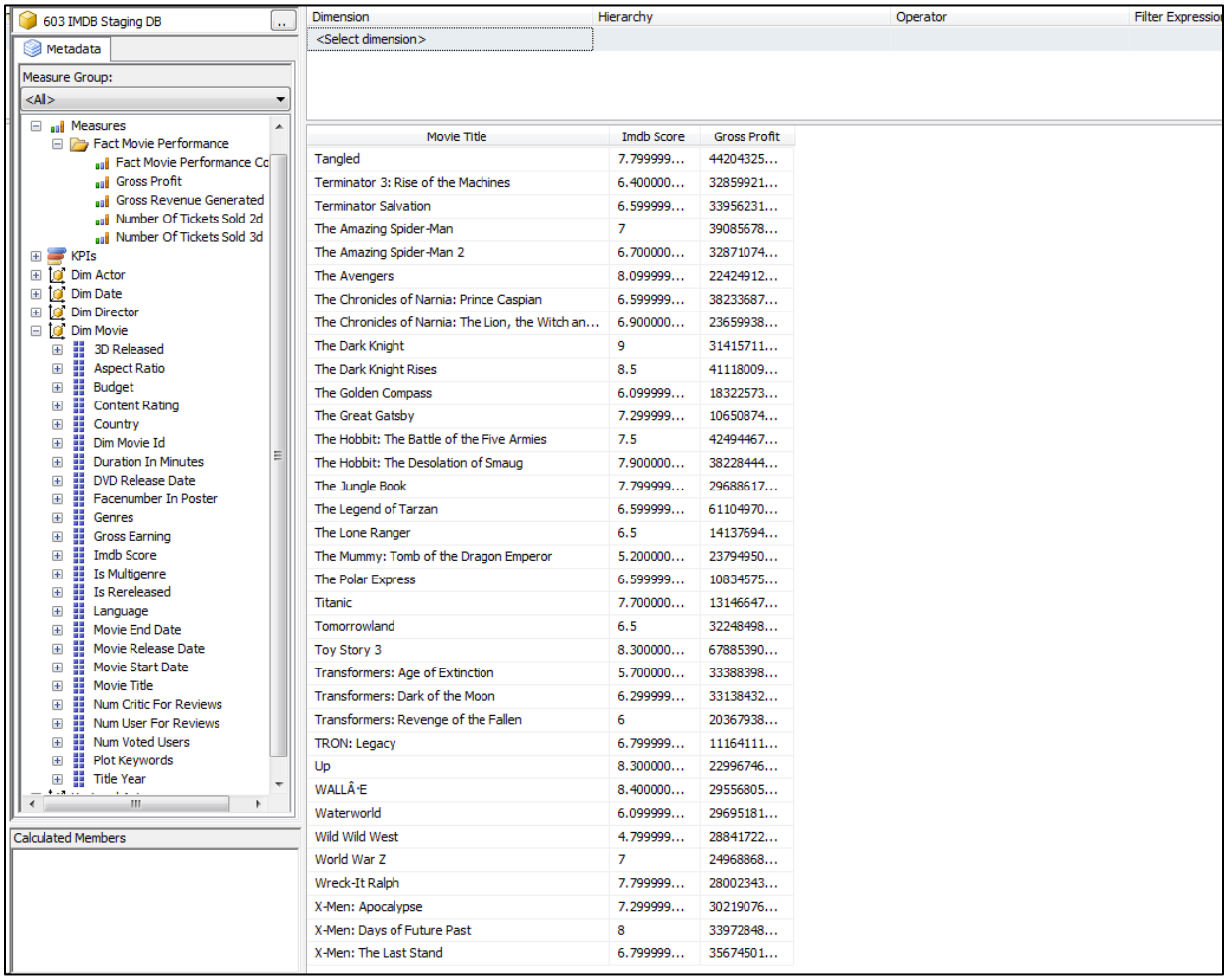
Browse the desired cube. The cubes have been deployed on server and can be browsed easily.



On the left pane, user will see all the measures and dimensions for a particular data mart.



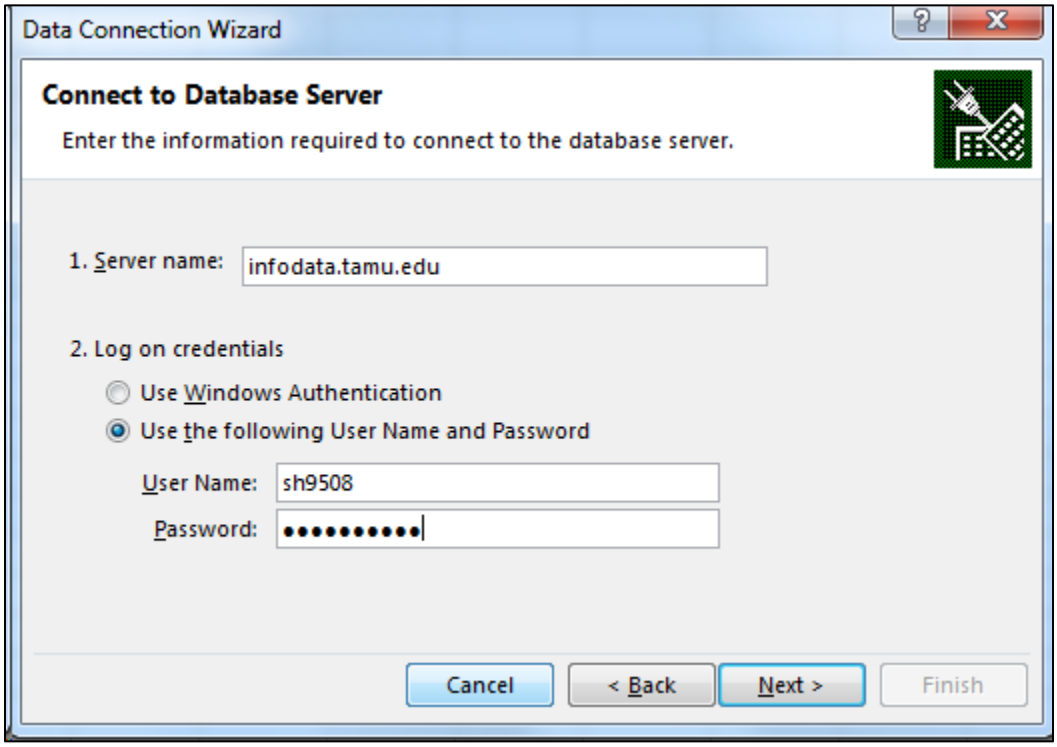
User can drag and drop the measures and corresponding fields from the dimensions which the user requires.

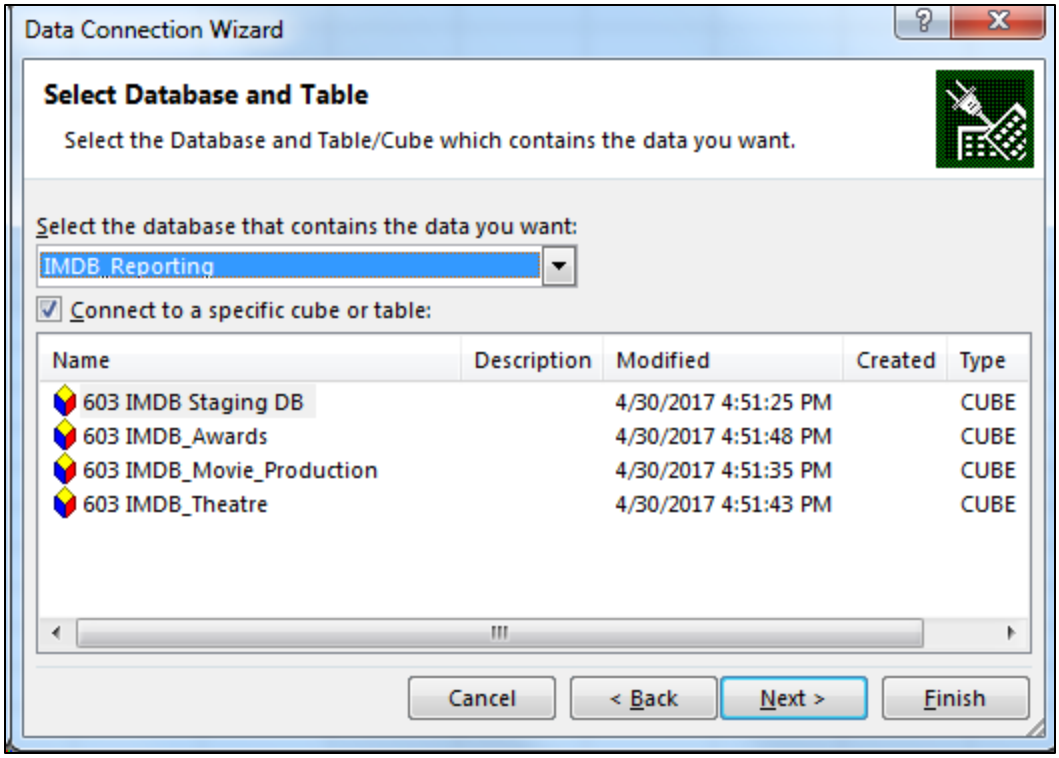


Similarly, user can browse any of the 4 cubes to determine trends and generate reports he/she needs.

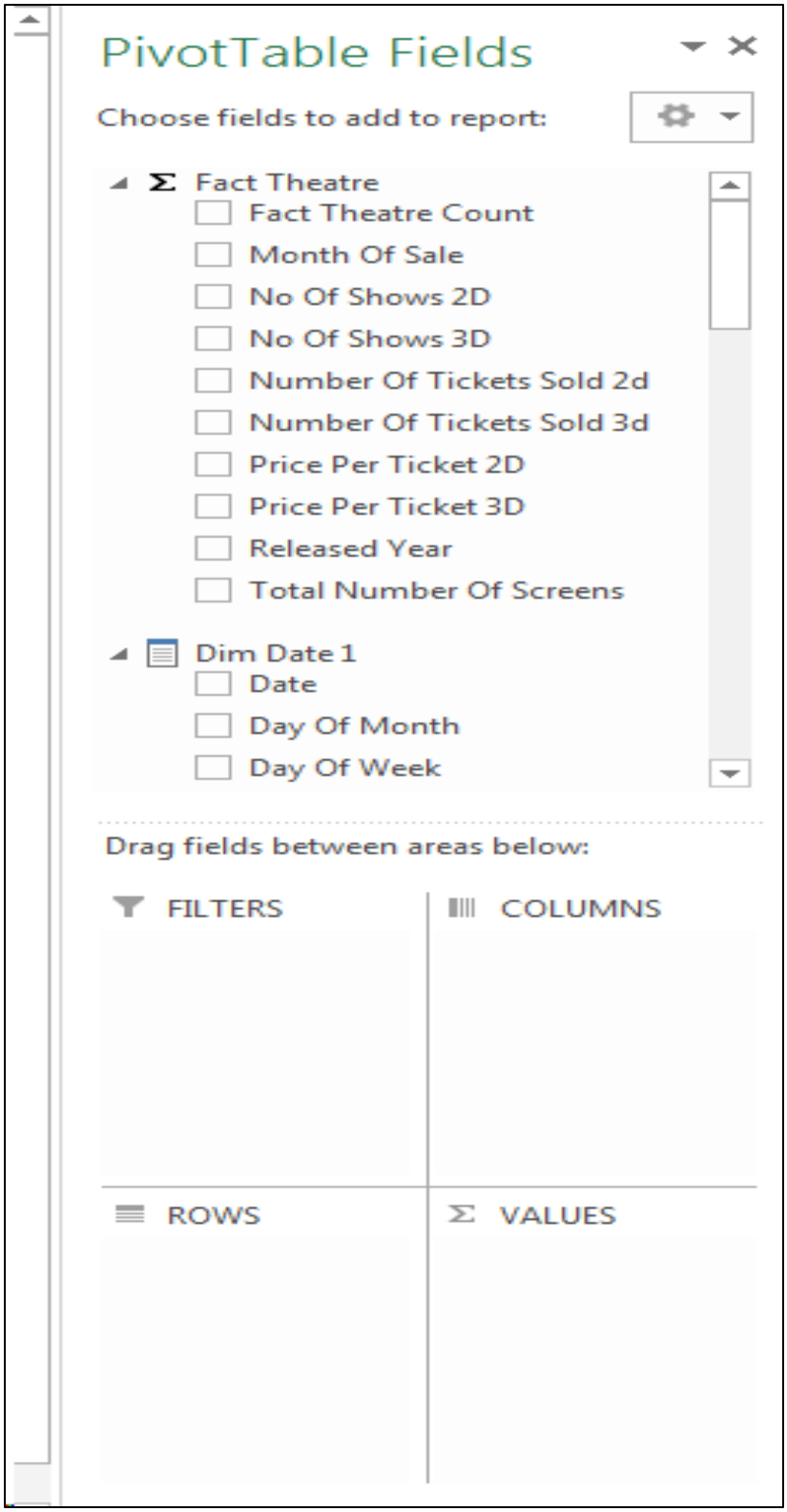
8.3.2 Browsing cubes using pivot

Connecting to the cubes

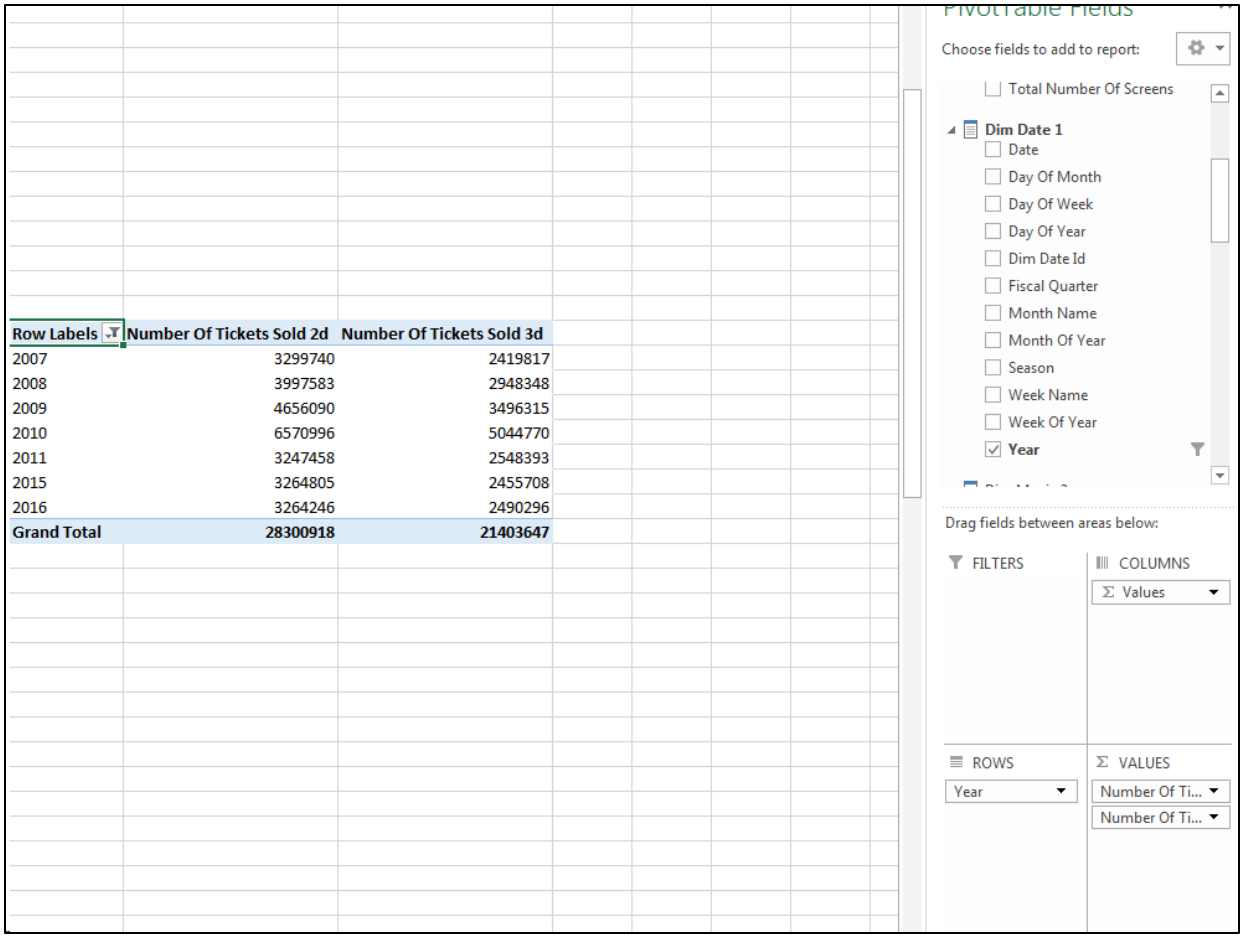




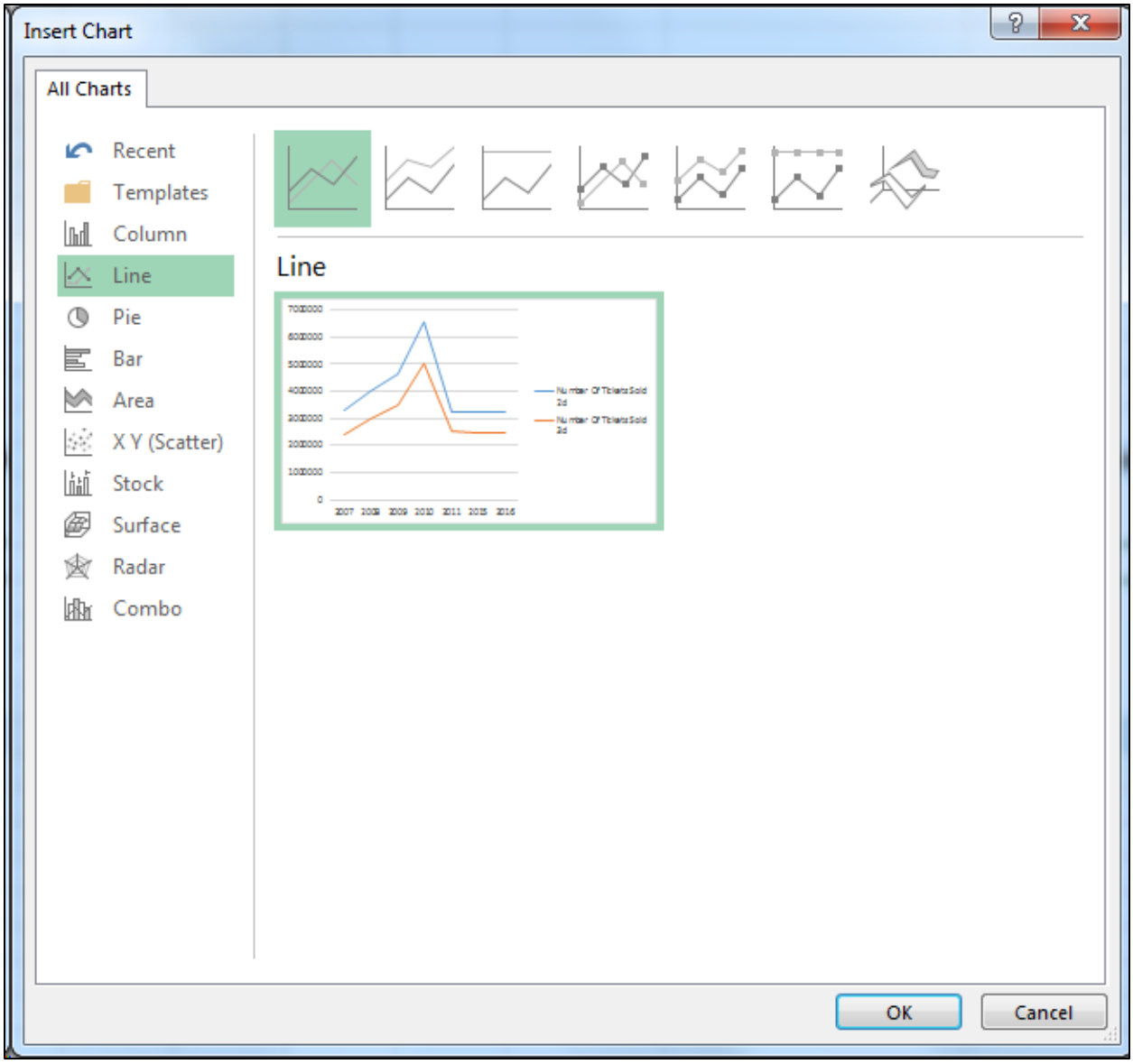
User can drag and drop the measures and corresponding fields from the dimensions which the user requires.



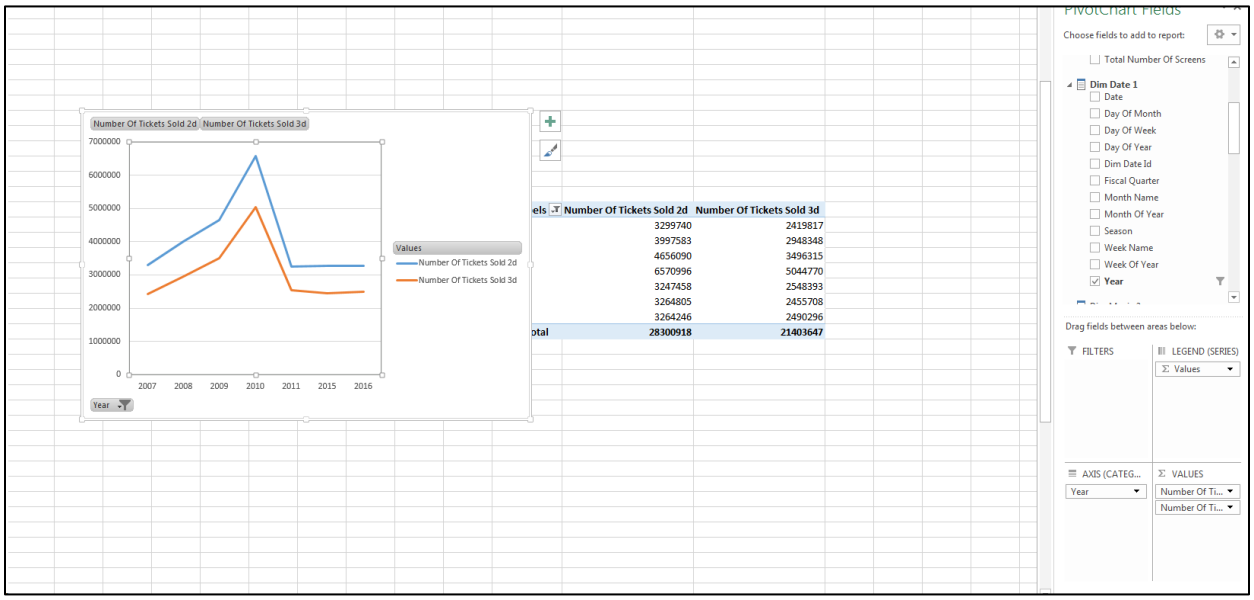
Sample report generated after user selection:



Row Labels	Number Of Tickets Sold 2d	Number Of Tickets Sold 3d
2007	3299740	2419817
2008	3997583	2948348
2009	4656090	3496315
2010	6570996	5044770
2011	3247458	2548393
2015	3264805	2455708
2016	3264246	2490296
Grand Total	**28300918**	**21403647**

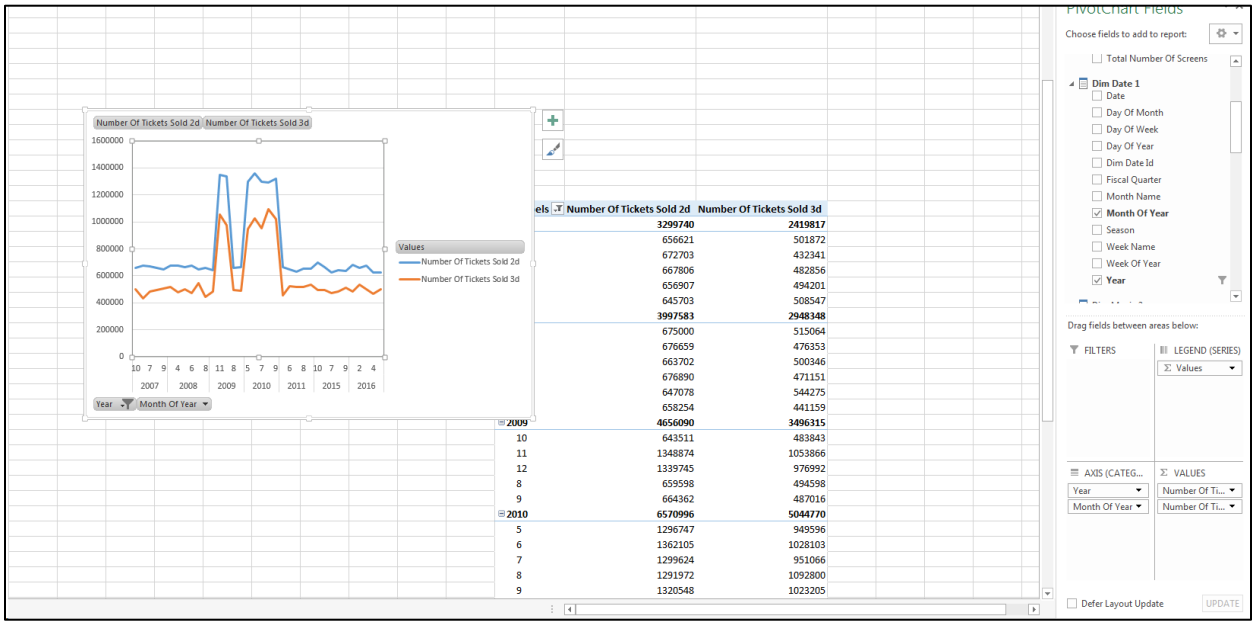

Insert Chart
All Charts
Recent
Templates
Column
Line
Pie
Bar
Area
X Y (Scatter)
Stock
Surface
Radar
Combo
Line
OK
Cancel

Trends are generated as per the figure below after user selection



	Number Of Tickets Sold 2d	Number Of Tickets Sold 3d
	3299740	2419817
	3997583	2948348
	4656090	3496315
	6570996	5044770
	3247458	2548393
	3264805	2455708
	3264246	2490296
	28300918	21403647

User can now change the dimensions or measures and corresponding changes will be reflected in the graph.



	Number Of Tickets Sold 2d	Number Of Tickets Sold 3d
	3299740	2419817
	656621	501872
	672703	432341
	667806	482856
	656907	494201
	645703	508547
	3997583	2948348
	675000	515064
	676659	476353
	663702	500346
	676890	471151
	647078	544275
	658254	441159
2009	4656090	3496315
10	643511	483843
11	1348874	1053866
12	1339745	976992
8	659598	494598
9	664362	487016
2010	6570996	5044770
5	1296747	949596
6	1362105	1028103
7	1299624	951066
8	1291972	1092800
9	1320548	1023205

This way either by deploying cubes directly to the server or implementing the cubes using pivots would enable the user to generate reports and trends from a particular data mart. This will help users who do not have prior knowledge of SQL servers to go ahead and generate the reports they desire.

9. Glossary of terms

A

Aggregation: One way of speeding up query performance. Facts are summed up for selected dimensions from the original fact table. The resulting aggregate table will have fewer rows, thus making queries that can use them go faster.

Attribute: Attributes represent a single type of information in a dimension. For example, year is an attribute in the Time dimension.

C

Conformed Dimension: A dimension that has exactly the same meaning and content when being referred to from different fact tables.

D

Data Cleansing: The transformation of data in its current state to a pre-defined, standardized format using packaged software or program modules.

Data Extraction: The process of pulling data from operational and external data sources in order to prepare the source data for the data warehouse environment.

Data Integration: The movement of data between two co-existing systems. The Interfacing of this data may occur once every hour, once a day, etc.

Data Integrity: The quality of the data residing in the database objects. The measurement which users consider when analyzing the value and reliability of the data.

Data Mart: A data warehouse data class organized for a business functional area or department. The database contains data summarized at multiple levels of granularity and may be designed using relational or multidimensional database structures.

Data Mart Data Model: The logical representation of the specific information requirements organized around a department of functional area.

Data Migration: The movement of data from one database to another database -- but not necessarily to a working application or subsystem tables.

Data Model: A representation of the specific information requirements of a business area.

Derived Attribute: A value that is derived by some algorithm from the values of other attributes; for example, profit, which is the difference between revenue and expense.

Data Mart: Data marts have the same definition as the data warehouse (see below), but data marts have a more limited audience and/or data content.

Data Warehouse: A warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process (as defined by Bill Inmon).

Data Warehousing: The process of designing, building, and maintaining a data warehouse system.

Dimension: A multidimensional structure, which represents a side of a multidimensional cube. Each dimension represents a different category, such as region, time, product type etc.

Dimensional Model: A type of data modeling suited for data warehousing. In a dimensional model, there are two types of tables: dimensional tables and fact tables. Dimensional table records information on each dimension, and fact table records all the "fact", or measures.

Dimension Table: A table that contains discrete values (usually a countable text field like school or degree). Also, see fact table. Imagine viewing a spreadsheet. The row and column names would be the dimensions and the numeric data within would be the facts.

Drill Across: Data analysis across dimensions.

Drill Down: Data analysis to a child attribute.

Drill Through: Data analysis that goes from an OLAP cube into the relational database.

Drill Up: Data analysis to a parent attribute.

E

ETL: Stands for Extraction, Transformation, and Loading. The movement of data from one area to another.

Extraction, Transformation and Loading (ETL) Tool: Software that is used to extract data from a data source like an operational system or data warehouse, modify the data and then load it into a data mart, data warehouse or multi-dimensional data cube.

F

Fact Table: A type of table in the dimensional model. A fact table typically includes two types of columns: fact columns and foreign keys to the dimensions.

G

Grain: A term used to describe how finally broken down a fact is in a table. For example, we might have wages individually recorded per employee in one table but we might have another table with wages aggregated by department.

H

Hierarchy: A hierarchy defines the navigating path for drilling up and drilling down. All attributes in a hierarchy belong to the same dimension.

M

Metadata: Data about data. For example, the number of tables in the database is a type of metadata.

Measure: A quantifiable variable or value stored in a multi-dimensional OLAP cube. It is a value in the cell at the intersection of two or more dimensions.

Metric: A measured value. For example, "Total movie ticket sales" is a metric.

MOLAP: Multidimensional OLAP. MOLAP systems store data in the multidimensional cubes.

O

OLAP: On-Line Analytical Processing. OLAP should be designed to provide end users a quick way of slicing and dicing the data.

S

SSIS: SQL Server Integration Services is a platform for building enterprise-level data integration and data transformations solutions. You use Integration Services to solve complex business problems by copying or downloading files, sending e-mail messages in response to events, updating data warehouses, cleaning and mining data, and managing SQL Server objects and data

SSAS: SQL Server Analysis Services is an analytical data engine used in decision support and business analytics, providing the analytical data for business reports and client applications such as Power BI, Excel, Reporting Services reports, and other data visualization tools.

SSRS: SQL Server Reporting Services is a solution that customers deploy on their own premises for creating, publishing, and managing reports, then delivering them to the right users in different ways, whether that is viewing them in web browser, on their mobile device, or as an email in their in-box.

Star Schema: A common form of dimensional model. In a star schema, a single dimension table represents each dimension.

9. Bibliography

[1] Business analysis framework for the data warehouse design and architecture of a data warehouse: http://www.tutorialspoint.com/dwh/dwh_architecture.htm

[2] Kimball Technical DW/BI System Architecture: http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/technical-dw-bi-system-architecture/

[3] Data Warehousing and Knowledge Discovery: 12th International Conference, DaWaK 2010, Bilbao, Spain, August 30 - September 2, 2010, Proceedings:https://books.google.com/books?id=IBZtCQAAQBAJ&printsec=frontcover#v=onepage&q&f=false

[4] Data Warehouse Architecture: http://www.1keydata.com/datawarehousing/data-warehouse-architecture.html

[5] Concepts of dimensional data modeling: https://www.ibm.com/support/knowledgecenter/en/SSGU8G_12.1.0/com.ibm.whse.doc/ids_ddi_350.htm

[6] The Data Warehouse Toolkit, Third Edition: The Definitive Guide to Dimensional Modeling by Ralph Kimball, Margy Ross. John Wiley & Sons, Jul 1, 2013

[7] The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data by Ralph Kimball, Joe Caserta. John Wiley & Sons, Apr 27, 2011

[8] Business Intelligence: Practices, Technologies, and Management by Rajiv Sabherwal, Irma Becerra-Fernandez. John Wiley & Sons, 2011

[9] SQL Server Technical Documentation -SQL Server Integration Services: https://docs.microsoft.com/en-us/sql/sql-server/sql-server-technical-documentation

[10] SQL Server Analysis Services: https://docs.microsoft.com/en-us/sql/analysis-services/analysis-services

10. Author of the Report

Name	UIN
Aditya Dakur Rudraiah	126003208
Ajay Thomas	825008104
Apurva Shrivastava	925009508
Isha Arora	825008103
Poonam Tare	825008406

11. Date Report Created

3rd May, 2017

12. Contribution of each group member to the project

Collective work as a group:

- Collecting the Movie data and preparing transactional data.
- Finalizing BI questions that the data warehouse will address.
- Drafting and finalizing the dimensional model for the data set.
- Documentation for the respective tasks performed.

Individual Contributions:

Student	Staging	Data warehouse	Reporting
Aditya Dakur	Creating the data sets and loading the data from source CSV tables to staging	• ETLTransfomations and loading the dimensions from Staging to data warehouse. • Loading the Facts from staging to data warehouse.	Generated Movie performance reports using SSRS and SSAS
Ajay Thomas	Creating the data sets and loading the data from source CSV tables to staging	• ETL Transformations and loading the dimensions from Staging to data warehouse. • Loading the Facts from staging to data warehouse.	Generated Theatre performance reports using SSRS and SSAS.
Apurva Shrivastava	Deformalizing the data provided and creating source tables and staging databases	• Loading the dimensions from Staging to data warehouse. • Loading the Facts from staging to data warehouse.	Implementation of SSAS and SSRS reporting. Also created power pivots.
Isha Arora	Creating the data sets and loading the data from source CSV tables to staging	• Loading the dimensions from Staging to data warehouse. • Loading the Facts from staging to data warehouse.	Generated Awards performance reports using SSRS and SSAS
Poonam Tare	Creating the data sets and loading the data from source CSV tables to staging	• Loading the dimensions from Staging to data warehouse. • Loading the Facts from staging to data warehouse.	Generated Production house performance reports using SSRS and SSAS