

Today in Film

Intro:

I want to make an application that uses the IMDB dataset to display all the films and shows that were released on today's date in history. I will use Microsoft Excel to build the application. My motivation behind this is that I graduated undergrad from the College of Business, and I'm really interested in becoming more familiar with database integration with Excel.

Database Overview:

Using MySQL, I will create a database schema only with relations essential to the application and then populate them with the LIST files provided by the IMDB website. I don't have the disk space or memory necessary to manage a full local IMDB instance on my personal computer. Each relation will have all tuples consistent with the IMDB source files; however, This database will not have every single relation that IMDB has. This database will contain data relating to:

- Movies
 - Genres
 - Ratings
 - Release Dates
 - Run-Times
- Actors
- Directors

Due to the format of the IMDB source files (.list), I expect customization of relations to be limited. Editing the data or formatting in those files before importing isn't practical due to their massive scale. For example, the 'Actors' file contains 3,981,379 tuples.

Because of the size, I will prioritize speed over normalization.

Application Overview:

Platform: MS Excel / VBA 64bit

Connector: ODBC Data Sources 64bit

The application will take the current date as an excel formula, pass it into a query to find a list of films that were released on today's date at anytime in the past. Today's date will automatically refresh. ODBC will bridge the connection from my local MySQL database to MS Excel.

Eg. If today's date = 4/11, then it would return any films that were released on April, 11th.

Input: There is no user input. The data will refresh and the query will run upon opening the excel file.

Output: Attributes: Title, Year, imdbdate

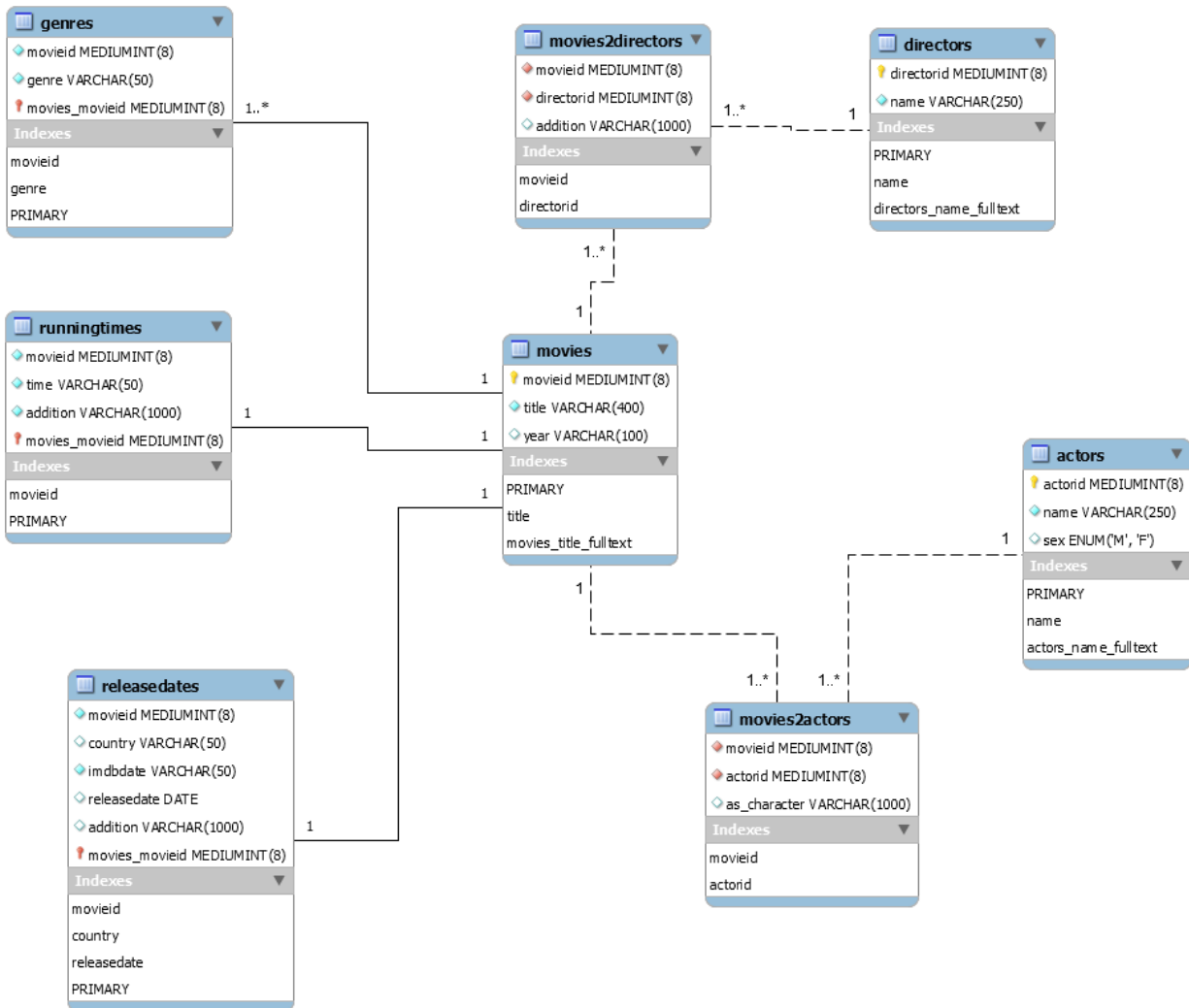
I'm not going to worry about SQL injection or any other security issues, as this will all only work locally. My personal computer will act as both client side and server side for the purpose of testing.

Environment:

- Windows 10
- MySQL 5.7
- MySQL Workbench 6.3 CE
- ODBC Data Sources (64bit) - The 64 bit version will work with MySQL
- Microsoft Office Excel 2016 (64bit) - Must use the 64 bit version to connect with ODBC

Conceptual and Logical Database Design:

E-R Diagram:



E-R Diagram Assumptions:

- A film does not have multiple release dates
- An actor does not play multiple characters in the same film/show

E-R Description:

Entities:

- movies: movie titles and the year they were released.
- directors: names of movie directors.
- movies2directors: what directors directed what movie.
- actors: names of male and female actors.
- movies2actors: what actors are in what movies.
- releasedates: when a movie was released
- runningtimes: how long a movie is.
- genres: genres of movies.

Relationships:

- movies & releasedates
 - 1-to-1
- movies & runningtimes
 - 1-to-1
- movies & genres
 - 1-to-1
- movies & movies2actors
 - 1-to-many
- movies & movies2directors
 - 1-to-many
- actors & movies2actors
 - 1-to-many
- directors & movies2directors
 - 1-to-many

Attributes:

- movieid
 - Belongs to: genres, runningtimes, releasedates, movies, movies2directors, movies2actors
 - Unique identifier for a single movie
 - Integer
 - PK/FK
 - Cannot be NULL
 - Not Derived
- Genre
 - Belongs to: genres

- Genre category that a movie belongs to
 - String
 - Not NULL
 - Not Derived
- Time
 - Belongs to: runningtimes
 - How long a movie is (time)
 - String
 - Not NULL
 - Not Derived
- Addition
 - Belongs to: runningtimes
 - Additional notes pertaining to run-time
 - String
 - Not NULL
 - Not Derived
- Country
 - Belongs to: releasedates
 - Country that a film was released in, pertaining to its release date
 - String
 - NULL
 - Not Derived
- Imdbdate
 - Belongs to: releasedates
 - Release date formatted differently for query purposes (1 January 1999)
 - String
 - Not NULL
 - Not Derived
- Releasedate
 - Belongs to: releasedates
 - Release date of a film
 - Date
 - NULL
 - Not Derived
- Addition
 - Belongs to: releasedates
 - Additional noted related to release date
 - String
 - NULL
 - Not Derived
- Directorid
 - Belongs to: directors, movies2directors
 - Unique identifier for a movie director

- Integer
 - PK/FK
 - Not NULL
 - Not Derived
- Name
 - Belongs to: directors
 - The name of a movie director
 - String
 - Not NULL
 - Not Derived
- title
 - Belongs to: movies
 - The title of a film
 - String
 - Not NULL
 - Not Derived
- Year
 - Belongs to: movies
 - The year that a film was released
 - String
 - NULL
 - Not Derived
- Actorid
 - Belongs to: actors, movies2actors
 - Unique identifier for an actor
 - Integer
 - PK/FK
 - Not NULL
 - Not Derived
- As_character
 - Belongs to: movies2actors
 - The character that an actor played in a specific movie
 - String
 - NULL
 - Not Derived
- Sex
 - Belongs to: actors
 - Sex of an actor (M, F)
 - Enum
 - NULL
 - Not Derived
- Name
 - Belongs to: actors

- The name of an actor
- String
- Not NULL
- Not Derived

Relational Schema:

```

movies(
    movieid: int, FK,
    title: string,
    year: string
)
directors(
    directorid: int, FK,
    name: string
)
movies2directors(
    movieid: int,
    directorid: int
    addition: string
)
actors(
    actorid: int, FK,
    name: string,
    sex: enum
)
movies2actors(
    movieid: int,
    actorid: int
    As_character: string
)
releasedates(
    movieid: int,
    country: string,
    imdbdate: string,
    releasedate: date,
    addition: string
)
runningtimes(
    movieid: int,
    time: string,

```

```

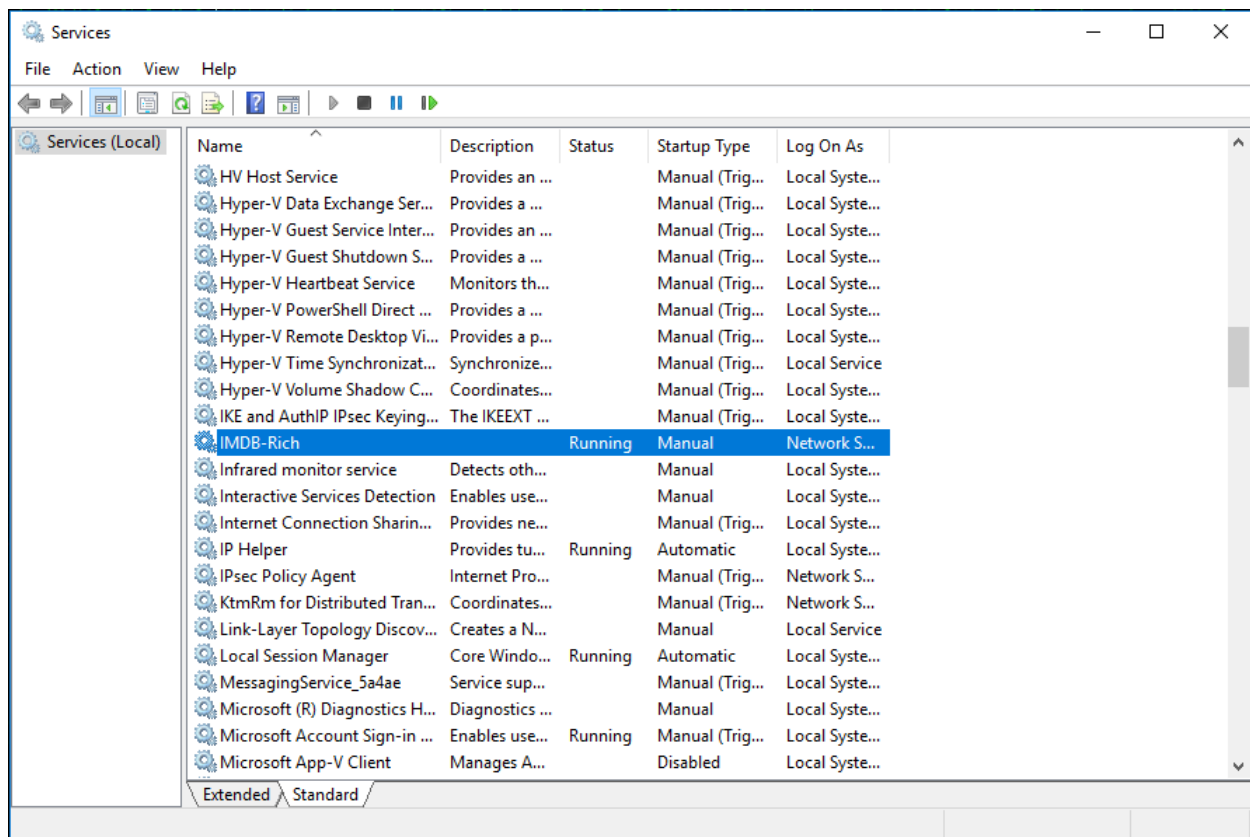
        addition: string
    )
    genres(
        movieid: int
        genre: string
    )

```

Database and Database Application Implementation and Demonstration:

I intend this to be in a narrative format, so I will start with the service.

Starting this service will start my instance of my MySQL database.



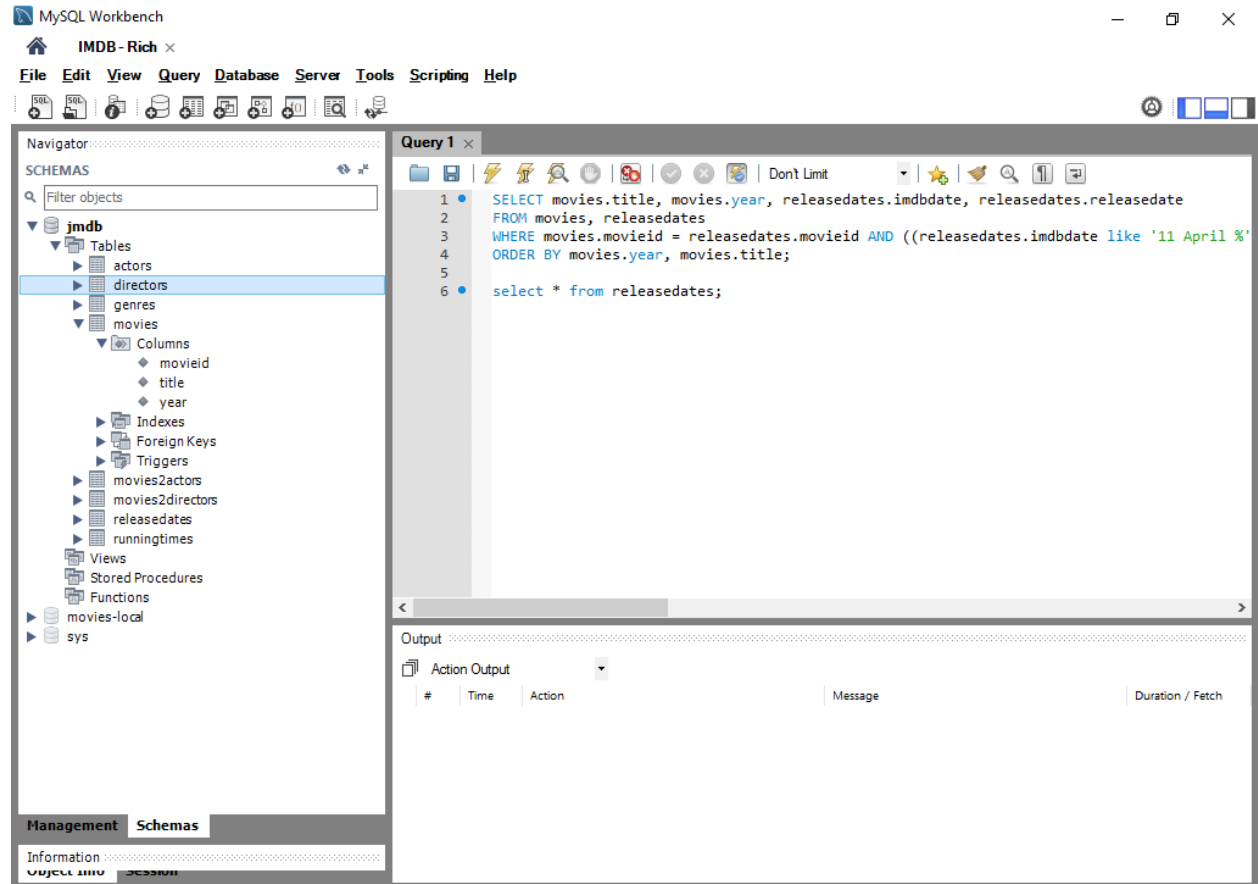
I will now take you through MySQL.

My connection is local.

Database: jmdb

Here is my schema:

You can see each attribute better in my Diagram above.



Here are some queries:

All batman movies and shows, their year, and their genres.

The screenshot shows the MySQL Workbench interface. The left sidebar displays the 'Navigator' pane with a tree view of the 'jmdb' database schema, including tables like 'actors', 'directors', 'genres', 'movies', 'movies2actors', 'movies2directors', 'releasedates', 'runningtimes', 'Views', 'Stored Procedures', 'Functions', 'movies-local', and 'sys'. The main window shows 'Query 1' with the following SQL code:

```
1 SELECT movies.title, movies.year, genres.genre
2 FROM movies, genres
3 WHERE movies.movieid = genres.movieid AND (movies.title like '%Batman%');
```

The 'Result Grid' displays the query results in a table with columns 'title', 'year', and 'genre'. The results list various Batman movies and their associated genres.

title	year	genre
"Batman: The Animated Series" (1992)	1992-1995	Action
"Batman: The Animated Series" (1992)	1992-1995	Adventure
"Batman: The Animated Series" (1992)	1992-1995	Animation
"Batman: The Animated Series" (1992)	1992-1995	Familv
"Batman: The Animated Series" (1992)	1992-1995	Sci-Fi
"Batman: The Brave and the Bold" (2008)	2008-2011	Action
"Batman: The Brave and the Bold" (2008)	2008-2011	Adventure
"Batman: The Brave and the Bold" (2008)	2008-2011	Animation
"Batman: The Brave and the Bold" (2008)	2008-2011	Crime
"Batman: The Brave and the Bold" (2008)	2008-2011	Familv
"Batman: The Brave and the Bold" (2008)	2008-2011	Fantasv
"Batman: The Brave and the Bold" (2008)	2008-2011	Sci-Fi
"Batman: The Darkest Night" (2017)	2017-????	Action
"Batman: The Darkest Night" (2017)	2017-????	Drama
"Batman: The Darkest Night" (2017)	2017-????	Mysterv
"Batman: The Hunt for Red Hood" (2016)	2016-????	Action
"Batman: The Long Halloween" (2015)	2015-????	Adventure

The 'Output' pane at the bottom shows the execution details: 'Action Output' is selected, and the message indicates '92 row(s) returned' with a duration of '1.500 sec / 0.000 sec'.

Due to the size of the database, complicated queries result in connection timeouts or if I increase the timeout threshold, then it crashes. For instance when I try to find a list of horror films released in the month of October, then I get this.

23:05:16 SELECT movies.title, movies.year, releasedates.imdbdate, genres.genre FROM movies, releasedates, genres WHERE movies.movieid = releasedates.movieid = genres.movieid AND (genres.genre = 'Horror') ORDER BY movies.year, movies.title Error Code: 2013. Lost connection to MySQL server during query 30.000 sec

However when I take it down to a list of Horror films, I get:

MySQL Workbench

IMDB - Rich x

File Edit View Query Database Server Tools Scripting Help

Navigator

SCHEMAS

Filter objects

jmdb

- Tables
 - actors
 - directors
 - genres
 - Columns
 - movieid
 - genre
 - Indexes
 - Foreign Keys
 - Triggers
- movies
 - Columns
 - movieid
 - title
 - year
 - Indexes
 - Foreign Keys
 - Triggers
- movies2actors
- movies2directors
- releasedates
- runningtimes

- Views
- Stored Procedures
- Functions
- movies-local
- sys

Query 1 x SQL File 3*

```
1 SELECT movies.title, movies.year, genres.genre
2 FROM movies, genres
3 WHERE movies.movieid = genres.movieid
4 AND (genres.genre = 'Horror')
5 ORDER BY movies.year, movies.title;
```

Result Grid

title	year	genre
"13th Hour" (1958)	1958-1959	Horror
"Double Shock" (1958)	1958-1959	Horror
"Shock Theater" (1958)	1958-1959	Horror
"Shock" (1958/I)	1958-1959	Horror
"Shock" (1958/II)	1958-1959	Horror
"Shock!" (1958)	1958-1960	Horror
"Creature Feature" (1958)	1958-1961	Horror
"The Witching Hour" (1958)	1958-1962	Horror
"Haunted" (1958)	1958-????	Horror
"Quatermass and the Pit" (1958)	1958-????	Horror
"The Unforeseen" (1958)	1958-????	Horror
"The Vell" (1958)	1958-????	Horror
"Starline" (1959/I) (The Turn of the Screw (#1...	1959	Horror
"13 Demon Street" (1959)	1959-1960	Horror
"The Advisor's Mysterv Theater" (1959)	1959-1960	Horror

Result 1 x Read Only

Output

Action Output

#	Time	Action	Message	Duration / Fetch
5	23:03:33	SELECT movies title, movies year, releasedates imdb...	Error Code: 2013. Lost connection to MySQL server du...	30.000 sec
6	23:04:22	SELECT movies title, movies year, releasedates imdb...	Error Code: 2013. Lost connection to MySQL server du...	30.016 sec
7	23:05:16	SELECT movies title, movies year, releasedates imdb...	Error Code: 2013. Lost connection to MySQL server du...	30.000 sec
8	23:05:57	SELECT movies title, movies year, genres.genre FROM...	2324 row(s) returned	0.078 sec / 0.000 sec

This is the query I will be using for my application (except for the exact date):

The screenshot shows the MySQL Workbench interface. The top menu bar includes File, Edit, View, Query, Database, Server, Tools, Scripting, and Help. The left sidebar shows the 'Navigator' pane with a tree view of the 'jmdb' database schema, including tables like 'actors', 'directors', 'genres', 'movies', 'movies2actors', 'movies2directors', 'releasedates', and 'runningtimes'. The main window displays 'Query 1' with the following SQL code:

```
SELECT movies.title, movies.year, releasedates.imdbdate
FROM jmdb.movies movies, jmdb.releasedates releasedates
WHERE movies.movieid = releasedates.movieid AND ((releasedates.imdbdate like '5 April %'))
ORDER BY movies.year, movies.title;
```

The 'Result Grid' shows the following data:

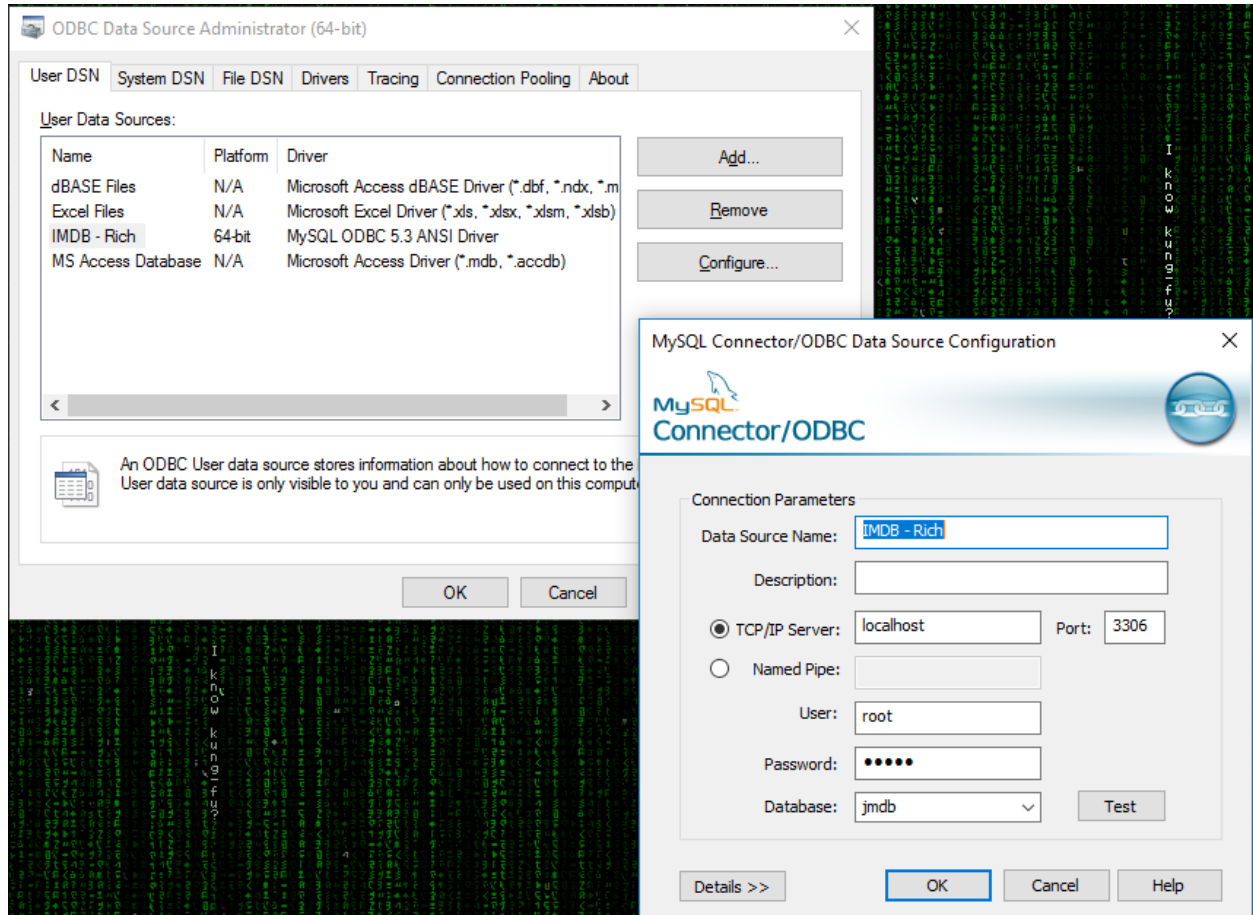
title	year	imdbdate
"Cosmopolitan Cafe" (1937) {{1937-04-05}}	1937	5 April 1937
"Sea Stories" (1936) {{#1.4}}	1937	5 April 1937
"Memorv Lane" (1947) {{1948-04-05}}	1948	5 April 1948
"The Teen-Age Show" (1946) {{#1.4}}	1948	5 April 1948
"Fireside Theatre" (1949) (Friend of the Familv ...	1949	5 April 1949
"Memorv Lane" (1947) {{1949-04-05}}	1949	5 April 1949
"Suspense" (1949) (The Man Upstairs (#1.4))	1949	5 April 1949
"Fireside Theatre" (1949)	1949-1955	5 April 1949
"How Do You View?" (1949) {{#2.1}}	1950	5 April 1950
"Kraft Television Theatre" (1947) (A Doll's Hous...	1950	5 April 1950
"The Plainclothesman" (1949) (Alibi (#1.26))	1950	5 April 1950
"Treasury Men in Action" (1950) {{#1.13}}	1950	5 April 1950
"Art Ford on Broadwav" (1950)	1950-????	5 April 1950
"Jim and Judy in Teleland" (1950)	1950-????	5 April 1956
"Joev Fave's Frolics" (1950)	1950-????	5 April 1950

The 'Output' pane at the bottom shows the execution log:

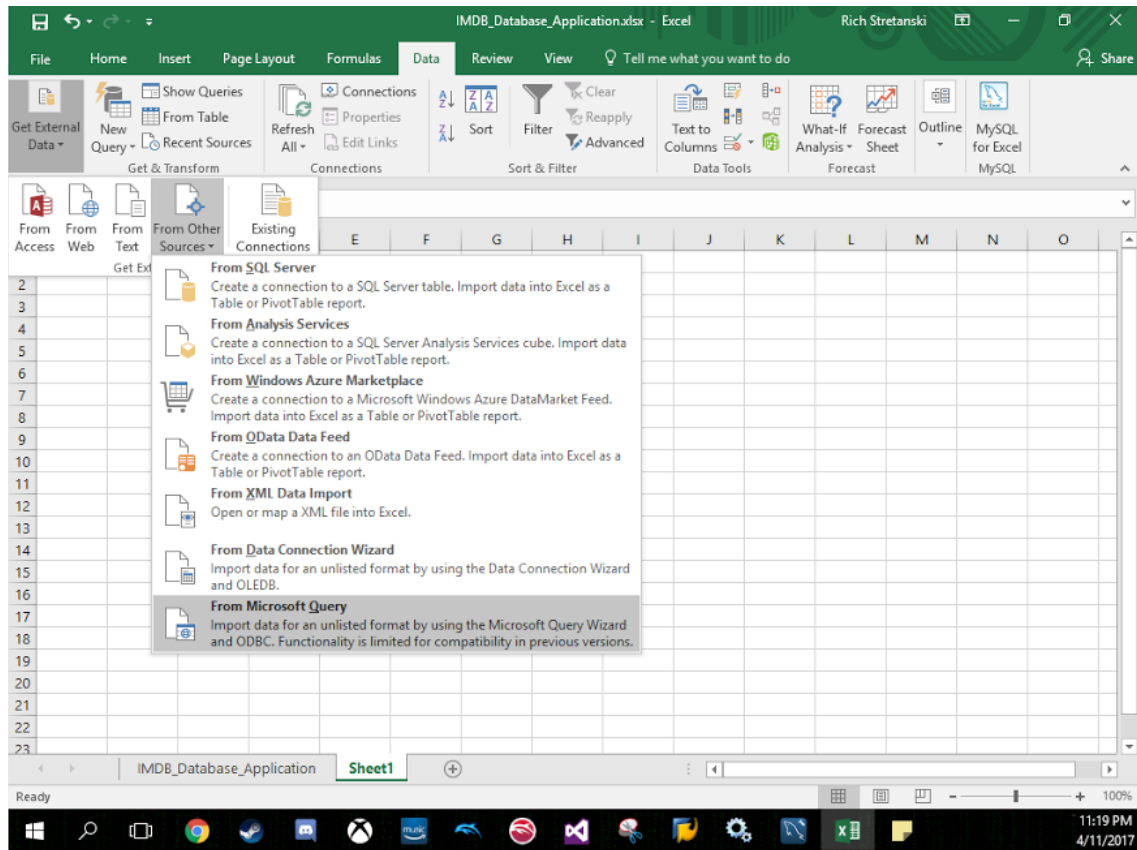
#	Time	Action	Message	Duration / Fetch
6	23:04:22	SELECT movies.title, movies.year, releasedates.imdbd...	Error Code: 2013. Lost connection to MySQL server du...	30.016 sec
7	23:05:16	SELECT movies.title, movies.year, releasedates.imdbd...	Error Code: 2013. Lost connection to MySQL server du...	30.000 sec
8	23:05:57	SELECT movies.title, movies.year, genres.genre FROM...	2324 row(s) returned	0.078 sec / 0.000 sec
9	23:13:05	SELECT movies.title, movies.year, releasedates.imdbd...	7580 row(s) returned	1.922 sec / 0.000 sec

Moving on to the database application:

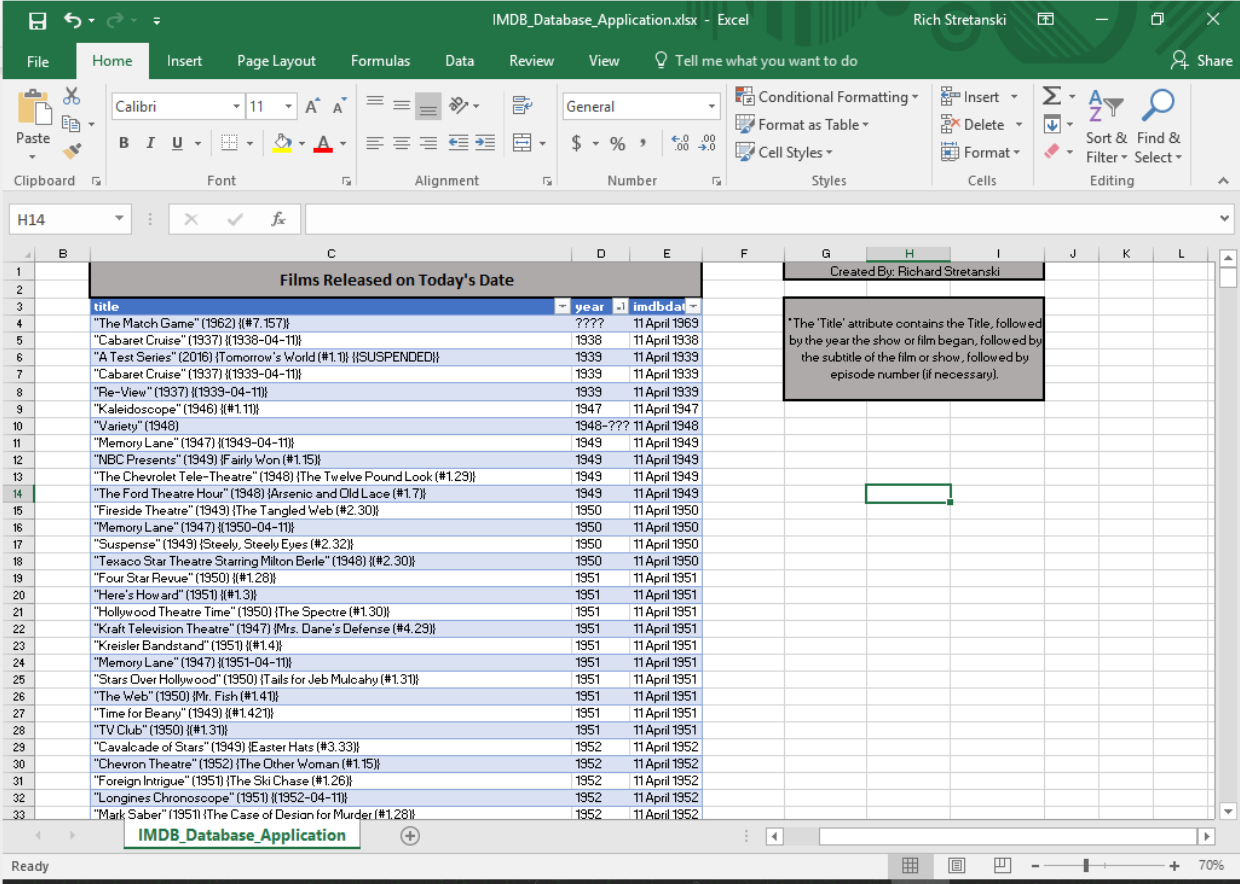
I used the ODBC data source connector to bridge MySQL and Excel.



I established the connection through ODBC in Excel.



Here is my finished application:
It runs the query and displays the results upon opening.



IMDB_Database_Application.xlsx - Excel

Rich Stretanski

File Home Insert Page Layout Formulas Data Review View Tell me what you want to do

Clipboard Font Alignment Number Styles Cells Editing

H14

Films Released on Today's Date			Created By: Richard Stretanski	
title	year	imdbdate		
"The Match Game" (1962) (#7.157)	1962	11 April 1969		
"Cabaret Cruise" (1937) (#1938-04-11)	1938	11 April 1938		
"A Test Series" (2016) (Tomorrow's World (#1.1)) (SUSPENDED)	1939	11 April 1939		
"Cabaret Cruise" (1937) (#1939-04-11)	1939	11 April 1939		
"Re-View" (1937) (#1939-04-11)	1939	11 April 1939		
"Kaleidoscope" (1946) (#1.11)	1947	11 April 1947		
"Variety" (1948)	1948-???	11 April 1948		
"Memory Lane" (1947) (#1949-04-11)	1949	11 April 1949		
"NBC Presents" (1949) (#Fairly Won (#1.15))	1949	11 April 1949		
"The Chevrolet Tele-Theatre" (1948) (#The Twelve Pound Look (#1.29))	1949	11 April 1949		
"The Ford Theatre Hour" (1948) (#Arsenic and Old Lace (#1.7))	1949	11 April 1949		
"Fireside Theatre" (1949) (#The Tangled Web (#2.30))	1950	11 April 1950		
"Memory Lane" (1947) (#1950-04-11)	1950	11 April 1950		
"Suspense" (1949) (#Steely, Steely Eyes (#2.32))	1950	11 April 1950		
"Texaco Star Theatre Starring Milton Berle" (1948) (#2.30)	1950	11 April 1950		
"Four Star Revue" (1950) (#1.28)	1951	11 April 1951		
"Here's Howard" (1951) (#1.3)	1951	11 April 1951		
"Hollywood Theatre Time" (1950) (#The Spectre (#1.30))	1951	11 April 1951		
"Kraft Television Theatre" (1947) (#Mrs. Dane's Defense (#4.29))	1951	11 April 1951		
"Kreislerr Bandstand" (1951) (#1.4)	1951	11 April 1951		
"Memory Lane" (1947) (#1951-04-11)	1951	11 April 1951		
"Stars Over Hollywood" (1950) (#Tails for Jeb Mulcahy (#1.31))	1951	11 April 1951		
"The Web" (1950) (#Mr. Fish (#1.41))	1951	11 April 1951		
"Time for Beany" (1949) (#1.421)	1951	11 April 1951		
"TV Club" (1950) (#1.31)	1951	11 April 1951		
"Cavalcade of Stars" (1949) (#Easter Hats (#3.33))	1952	11 April 1952		
"Chevron Theatre" (1952) (#The Other Woman (#1.15))	1952	11 April 1952		
"Foreign Intrigue" (1951) (#The Ski Chase (#1.26))	1952	11 April 1952		
"Longines Chronoscope" (1951) (#1952-04-11)	1952	11 April 1952		
"Mark Saber" (1951) (#The Case of Design for Murder (#1.28))	1952	11 April 1952		

IMDB_Database_Application

Ready

70%

"The 'Title' attribute contains the Title, followed by the year the show or film began, followed by the subtitle of the film or show, followed by episode number (if necessary).

Here is where the query runs

The screenshot shows the 'Connection Properties' dialog box in Microsoft Excel. The dialog is titled 'Connection Properties' and has a 'Definition' tab selected. The 'Connection name' is 'Query from IMDB - Rich'. The 'Connection type' is 'Database Query'. The 'Connection string' is 'DSN=IMDB - Rich;'. The 'Command type' is 'SQL'. The 'Command text' is a SQL query:
`SELECT movies_0.title, movies_0.year, releasedates_0.imdbdate
FROM imdb.movies movies_0, imdb.releasedates releasedates_0
WHERE movies_0.movieid = releasedates_0.movieid AND ((releasedates_0.imdbdate like ?))
ORDER BY movies_0.year, movies_0.title`
The background spreadsheet is titled 'IMDB Database Application' and contains a list of movie titles and release dates. The status bar at the bottom shows 'Average: 42836 Count: 22753 Sum: 42836' and a zoom level of 70%.

title	year	releasedates_0.imdbdate
"The Match Game" (1962) (#7.157)	1962	11 April 1952
"Cabaret Cruise" (1937) (#1938-04-11)	1937	11 April 1952
"A Test Series" (2016) (Tomorrow's)	2016	
"Cabaret Cruise" (1937) (1939-04-11)	1937	11 April 1952
"Re-View" (1937) (#1939-04-11)	1937	11 April 1952
"Kaleidoscope" (1946) (#1.11)	1946	
"Variety" (1948)	1948	
"Memory Lane" (1947) (#1949-04-11)	1947	11 April 1952
"NBC Presents" (1949) (#Fairly Won)	1949	
"The Chevrolet Tele-Theatre" (1949)	1949	
"The Ford Theatre Hour" (1948) (#Ari)	1948	
"Fireside Theatre" (1949) (#The Tang)	1949	
"Memory Lane" (1947) (#1950-04-11)	1947	11 April 1952
"Suspense" (1943) (#Steely, Steely E)	1943	
"Texaco Star Theatre Starring Mito		
"Four Star Revue" (1950) (#1.28)	1950	
"Here's Howard" (1951) (#1.3)	1951	
"Hollywood Theatre Time" (1950) (#1)	1950	
"Kraft Television Theatre" (1947) (#M)	1947	
"Kreiser Bandstand" (1951) (#1.4)	1951	
"Memory Lane" (1947) (#1951-04-11)	1947	11 April 1952
"Stars Over Hollywood" (1950) (#Tails for J)	1950	
"The Web" (1950) (#Mr. Fish (#1.41))	1950	
"Time for Beany" (1949) (#1.421)	1949	
"TV Club" (1950) (#1.31)	1950	
"Cavalcade of Stars" (1949) (#Easter Hats)	1949	
"Chevron Theatre" (1952) (#The Other Vol...	1952	
"Foreign Intrigue" (1951) (#The Ski Chase (#1.26))	1951	11 April 1952
"Longines Chronoscope" (1951) (#1952-04-11)	1951	11 April 1952
"Mark Saber" (1951) (#The Case of Design for Murder (#1.28))	1951	11 April 1952

I did a bit of formatting in column A to get the =TODAY() formula to fit into the query.

The screenshot shows an Excel spreadsheet with the following data:

Column	Header	Value
A	11 April	11 April
B	4/11/17	1938
C	title	"The Match Game" (1962) (#7.157)
D	year	1938
E	imdbdate	11 April 1938

The text box explains the 'Title' attribute format: "The 'Title' attribute contains the Title, followed by the year the show or film began, followed by the subtitle of the film or show, followed by episode number (if necessary)."

The Excel application will be included in the upload.