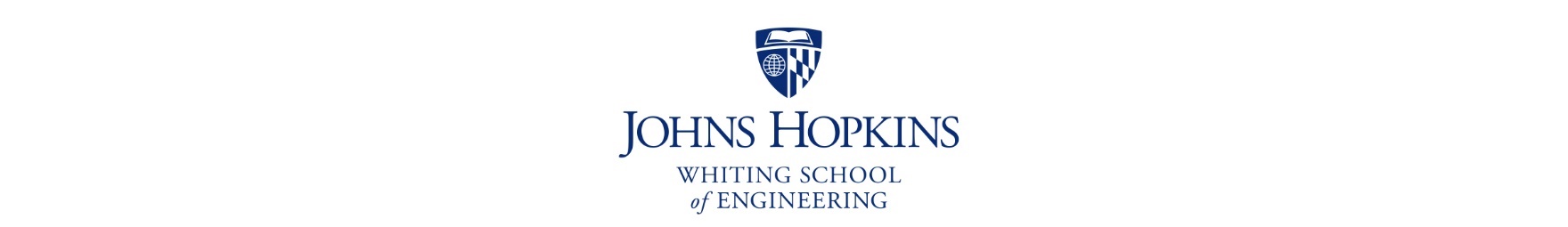
|  |
| --- |
|  |
| MovieDbLite |
| Course Section: CS605.641  Spring, 2020 |
| Prepared by |
| **Anderson, Steven** |
| **04/22/2020** |



|  |
| --- |
| Database Design Project Document |

**Table of Contents**

[1. Introduction 4](#_Toc37004267)

[1.1. Scope and Purpose of Document 4](#_Toc37004268)

[1.2. Project Objective 4](#_Toc37004269)

[2. System Requirements 4](#_Toc37004270)

[2.1 Hardware Requirements 5](#_Toc37004271)

[2.1.1 Database Server Requirements 5](#_Toc37004272)

[2.1.2 Web Application Server Requirements 5](#_Toc37004273)

[2.1.3 Website Front-End Access Requirements 6](#_Toc37004274)

[2.2 Software Requirements 6](#_Toc37004275)

[2.2.1 Database Server Requirements 6](#_Toc37004276)

[2.2.2 Web Application Server Requirements 7](#_Toc37004277)

[2.2.3 Website Front-End Access Requirements 7](#_Toc37004278)

[2.3 Functional Requirements 7](#_Toc37004279)

[2.4 Database Requirements 8](#_Toc37004280)

[2.5 Additional Development Details 9](#_Toc37004281)

[3. Database Design Description 9](#_Toc37004282)

[3.1 Design Rationale 17](#_Toc37004283)

[3.2 E/R Model 18](#_Toc37004284)

[3.4.1 General Movie Information 19](#_Toc37004285)

[3.4.2 User Access and Reviews 20](#_Toc37004286)

[3.4.3 Movie Cast and Crew 21](#_Toc37004287)

[3.4.4 Award Shows and Winners 22](#_Toc37004288)

[3.4.5 Entities 23](#_Toc37004289)

[3.4.6 Relationships 23](#_Toc37004290)

[3.4.7 E/R Diagram 29](#_Toc37004291)

[3.5 Relational Model 29](#_Toc37004292)

[3.5.6 Data Dictionary 29](#_Toc37004293)

[3.5.7 Integrity Rules 33](#_Toc37004294)

[3.5.8 Operational Rules 36](#_Toc37004295)

[3.5.9 Operations 36](#_Toc37004296)

[3.6 Security 36](#_Toc37004297)

[3.4.1. Password Storage 36](#_Toc37004298)

[3.4.2 SQL Injection 37](#_Toc37004299)

[3.4.3 Database Connection Strings 39](#_Toc37004300)

[3.4.4 Network Connection 39](#_Toc37004301)

[3.5 Database Backup and Recovery 39](#_Toc37004302)

[3.5.1 Automated Database Backups 40](#_Toc37004303)

[3.5.2 Full Database Recovery Model 41](#_Toc37004304)

[3.5.3 Source Control of Database and Application 41](#_Toc37004305)

[3.6 Using Database Design or CASE Tool 41](#_Toc37004306)

[3.7 Other Possible E/R Relationships 41](#_Toc37004307)

[4. Implementation Description 42](#_Toc37004308)

[4.1 Data Dictionary 42](#_Toc37004309)

[4.2 Advanced Features 42](#_Toc37004310)

[4.3 Queries 42](#_Toc37004311)

[4.3.1 Get Movie Details Report (Primary Details with Languages/Genres) 42](#_Toc37004312)

[4.3.2 Get All Cast and Crew of a Movie (Table-Valued Function) 43](#_Toc37004313)

[4.3.3 Get All Cast and Crew Member Names and Roles 45](#_Toc37004314)

[4.3.4 Get all Movies that Film Member Participates in (Table-Valued Function) 46](#_Toc37004315)

[4.3.5 Report Movie and Role Information for Film Member 47](#_Toc37004316)

[4.3.6 Report the User Movie Reviews/Ratings 48](#_Toc37004317)

[4.3.7 Report the Helpful Vote Details of a Particular Review 49](#_Toc37004318)

[4.3.8 List defective videos 49](#_Toc37004319)

[4.3.9 List twenty popular videos by category for customers’ recommendations 50](#_Toc37004320)

[5. CRUD Matrix 50](#_Toc37004321)

[5.1 List of Entity Types 50](#_Toc37004322)

[5.2 List of Functions 50](#_Toc37004323)

[6. Concluding Remarks 50](#_Toc37004324)

[Appendices 51](#_Toc37004325)

[References 51](#_Toc37004326)

(Note: Please provide some inputs for the sections in your DB design document template even though some of them may not be suitable for your project. The purpose for all different sections is to ask you to think about different perspectives of a database project. If a section doesn't apply to your project, you should make something up. Please don't skip them.

For instance, Database Backup and Recovery is commonly implemented in a real database project. Even if you don’t implement it, most RDBMSs may have incorporated backup and recovery tools into their interfaces and infrastructure. Give a short description of your RDBMS solution to demonstrate your thought process.)

# Introduction

Movies are a common form of entertainment in the lives of many people. With the evolution of the internet and various streaming services, the access to films are more available than ever. The MovieDbLite project aims to be simple, easy-to-use website for viewing the details about movies and credits, as well as allowing users to rate and review the movies they have seen. In alignment with the website, the database itself is intended to be simple and intuitive for people with database experience.

For myself personally, I am a huge movie buff, and still go to the theaters multiple times a month with my wife to watch new releases. I chose this project because of my interest in movies and the people involved in delivering these films. In addition, the ownership of a movie database allows for me to personalize the reports or features that may not be found in other popular movie database applications. Furthermore, it opens up the possibility to create APIs (Application Programming Interfaces) for consumption of the data by other systems, use cases, and applications (e.g. trivia games).

## Scope and Purpose of Document

The purpose of this document is to describe design, development, requirement, and implementation details of the MovieDbLite database and website application, with the primary focus being on the database that is used to store movie and film industry data. Specifically, when it comes to the design of the database, the focus is on the conceptual and logical design and implementation. An existing, industry-standard Relational Database Management System (RDBMS) environment that handles the physical implementation of the database will be used. Therefore, lower-level details about physical storage and implementation of the database can be found via online resources of the respective RDBMS environment.

Overall, this document is bounded to describing the MovieDbLite database and front-end website for accessing basic details and reports of data within database. Accordingly, information about external applications that may consume the database data (e.g. through an API) is not intended to be covered by this document.

## Project Objective

The objective of this project is to design and implement a simple, yet useful movie database that tracks information related to the film industry. Further, the project intends to provide a front-end website to deliver an easy-to-use user interface for allowing users to view data and rate movies. This project shall be built using modern RDBMS and web-based technologies to captures the data elements and functional requirements that are mentioned in the upcoming sections. Ultimately, this database is intended to serve as a stable and maintainable back-end to enable future opportunities for APIs to be built for external applications to consume.

# System Requirements

The MovieDbLite is composed of three high-level system components that have their distinct set of system requirements:

1. The SQL Server hosted MovieDbLite Database
   1. Hosted on a remote server
2. The MovieDbLite Web Application Environment (Back-End)
   1. Hosted on a remote server
3. The MovieDbLite Website (Front-End)
   1. Accessed by an end-user via Web Browser

This section will cover the System Requirements (software and hardware) for the three major system components of the MovieDbLite system, as well as the functional requirements that make apart building these three components.

Although the Database (system component #1) and Web Application (system component #2) are hosted remotely, this does not necessarily mean they will be hosted on the same remote server/machine. In other words, the database and web application may be hosted on different server environments. Therefore, the following sections may distinguish and discretely separate the requirements for these respective components as appropriate.

## Hardware Requirements

This section describes the hardware requirements for the three components of the MovieDbLite system. In some of the hardware listings, a minimum hardware requirement is listed, as well as a recommended. The minimum hardware requirement is meant to indicate the minimum possible hardware specification to run or host the component. The recommended listing is the suggested minimum to mitigate deterioration of component performance. Overall, the system will support hardware specifications with larger amounts than the minimum/recommended listings, which would subsequently increase or maintain component performance and/or reliability.

### Database Server Requirements

The database server will require hardware able to host SQL Server Standard Edition 2017. The full server hardware requirements can be found in Microsoft’s documentation [here](https://docs.microsoft.com/en-us/sql/sql-server/install/hardware-and-software-requirements-for-installing-sql-server?view=sql-server-ver15#hardware-requirements).

This is the overall breakdown of the hardware requirements for SQL Server Standard Edition:

* **Memory** - Minimum of 1GB, recommended of 4GB
* **Hard Drive**
  + Computer with NTFS or ReFS file format for security reasons.
    - TODO: Provide links for NTFS and ReFS
  + Minimum of 6GB of available hard-disk space
    - Note: [Hard Disk Space Requirements](https://docs.microsoft.com/en-us/sql/sql-server/install/hardware-and-software-requirements-for-installing-sql-server?view=sql-server-ver15#HardDiskSpace) will apply depending on which components are installed
* **Monitor -** Super-VGA (800x600) or higher resolution monitor
* **Internet -** Internet access is required
* **Processor**
  + Speed: Minimum of 1.4GHz, recommended of 2.0+ GHz
  + Type: x64 Processor (e.g. AMD Opteron, AMD Athlon 64, Intel Xeon with Intel EM64T support, Intel Pentium IV with EM64T support)
* **Drive** – A DVD drive is required if installation is from disc

### Web Application Server Requirements

The Web Application environment (back-end to the website) will require an environment that can host an ASP.NET Website targeting .NET Core v3.1. The baseline hardware requirements can be found [here](https://help.syncfusion.com/aspnet-core/installation-and-upgrade/system-requirements#hardware-environment).

* **Memory** - Minimum of 512MB, recommended of 1GB
* **Hard Drive**
  + Up to 2GB of available space may be required
  + 300MB of free space is required in boot drive
* **Internet** – Internet access is required
* **Processor**
  + x86 and x64 processors are supported
  + Although specifications on speed is not provided, modern web servers recommend processor of 2.0+ GHz

### Website Front-End Access Requirements

The end-user will require hardware compatible to run a modern web browser (such as Chrome, FireFox, Edge) in order to access the website.

The following is the minimum hardware requirements for Chrome, which is generally the minimum requirements amongst other browser types:

* **Hard Drive** – Minimum 100MB of free hard drive space is required
* **Memory** – Minimum 128MB of RAM
* **Internet –** Network access is required to connect
* **Processor –** Pentium 4 and above
* **Keyboard (or equivalent) –** Typing will be required to perform searches
* **Mouse (or equivalent) –** Clicking will be required to navigate the web pages

More information can be found at this [link](https://d.docs.live.net/df1cecd6e646b923/Documents/Grad_School/Databases_605_641/Project/•%09https:/smallbusiness.chron.com/google-chrome-software-requirements-48820.html).

TODO: Android/iOS requirements?

TODO: Development environment requirements (Visual Studio and SSMS)

## Software Requirements

This section covers the software required to operate the components of MovieDbLite. In addition to Software application requirements, this section may list the Operating System (OS) requirements – as appropriate.

### Database Server Requirements

The database server will require software installed to host SQL Server Standard Edition 2017. The full server software requirements can be found in Microsoft’s documentation [here](https://docs.microsoft.com/en-us/sql/sql-server/install/hardware-and-software-requirements-for-installing-sql-server?view=sql-server-ver15#hwswr).

In summary, the following two software components are required to be installed on the server hosting the database:

1. .NET Framework (v4.6)
2. Network Software that support the following protocols: Shared Memory, Named Pipes, TCP/IP, and VIA.

The following [link](https://docs.microsoft.com/en-us/sql/sql-server/install/hardware-and-software-requirements-for-installing-sql-server?view=sql-server-ver15#operating-system-support) outlines the supported OS versions to host SQL Server 2017. Overall, the operating system must be a Windows Server 2012+ (2012, 2016, 2019) environment. Although Windows 8 and Windows 10 are listed as well, these are not intended to be used for remote hosting of the database.

### Web Application Server Requirements

The Web Application environment (back-end to the website) will require an Operation System environment that can host an ASP.NET Website targeting .NET Core v3.1. The full supported OS versions can be found [here](https://github.com/dotnet/core/blob/master/release-notes/3.1/3.1-supported-os.md). This section will focus on the remote server environments hosted by Windows, to remain consistent with Windows hosted database environment using SQL Server. In addition, the security model used for authentication to SQL Server will utilize Integrated Security which is Windows specific. More information on the connection string security is mentioned in section 3.4.3 Database Connection Strings

The following are the recommended operating system environments:

* Windows Server 2012 R2+
  + x64 and x86
* Windows Nano Server 1803+
  + x64 and ARM32

### Website Front-End Access Requirements

A modern web browser that supports HTML5 is required for the end-user to access the website.

The following are the preferred browsers to access the website:

* Google Chrome
* Microsoft Edge
* Mozilla Firefox
* Safari

Note: Although some features of the website may work in Internet Explorer, it is recommended to use one of the listed browsers above for full support.

## Functional Requirements

The MovieDbLite website will contain basic functionality to allow users to view details about movies and members of the film industry. Users may also create user accounts and rate/review movies. Administrative access will be available to allow for modification of movie and film industry details to keep the database’s data current.

This section covers the different functional requirements of the MovieDbLite web application in detail below.

* Track movie/film basic information (e.g. title, description, release date, duration, etc.)
  + Note: full data details to be described in Data Dictionary section. TODO: Link to this
* Track the different film members that are involved in the film industry, as well as basic information about them (preferred name, gender, date of birth, etc.)
* Track cast and crew members of a movie by their role (director, producer, actor, costume designer, choreographer, etc.)
* Track the different restriction ratings (e.g. G, PG, PG-13, R, etc.) that a film can be assigned to.
* Track the different international languages (English, Spanish, German, etc.) that a film can be assigned to.
* Track different annual award show events and the particular awards they give out
* Track the film members and movies that win awards at annual award show instances
* Track the different movie genres and allow movies to be classified as multiple of these genres
* Allow user to create a user account login (user name and password)
* Allow user access to modify details of their user account (user name, email address, password)
* Allow user to search for movies by movie title
* Allow user to perform an advanced movie search by other fields (description, release date, average user rating, restriction rating)
* Allow user to search for film members by their name
* Allow user to write a review and assign a user-rating (from 1-10) to a movie
* Allow user to mark a review as helpful or unhelpful
* Allow user to sort reviews by date, helpfulness, or rating
* Reports
  + Report a movie’s basic details (title, restriction rating, language, release date, etc.)
  + Report the film members (actors, producers, crew members, etc.) of a movie
  + Report the movies a film member has been a part of
  + Report the awards given to a certain movie
  + Report the awards given to a certain film member
  + Report the movies with the highest user rating
  + Report the user reviews for a movie
* Administratively allow adding/editing movie information
* Administratively allow adding/editing the film members of a movie
* Administratively allow inputting results from award show instances

## Database Requirements

The MovieDbLite Database is hosted using Microsoft’s SQL Server 2017 (Standard Edition).

The development/management environment used to perform DML and DDL operations against the database is Microsoft SQL Server Management Studio 17 (SSMS). The full details of SSMS are seen below.



## Additional Development Details

The ASP.NET Core Web Application is Developed using the Visual Studio 2019 IDE. The programming languages are C#, JavaScript, HTML, and CSS using the MVC (Model-View-Controller) design pattern.



# Database Design Description

This database

* Track Annual Award Shows
* Track Different Awards
* Track Movies and Film Members that receive awards
* Track Movies
* Track what Language a Movie is in
* Track when a Movie is or will be released
* Track the Restriction Rating on a movie
* Track a Movie’s Genres
* Allow Users to Sign-Up for Site
* Allow Users to Review Movies
* Allow Users to mark other reviews as helpful/unhelpful
* Track Actors in Movies
* Track various Crew Members that work on the set of Movies

Standards:

1. Unique Key Constraints and Index Naming
2. Column Naming
   1. PascalCased
   2. Standard Column Naming
      1. {Blank}Name (Required)
      2. Description (Nullable)
3. Indexing Guidelines
   1. Natural Key on each table
4. Table Naming
   1. PascalCased
5. Identity Keys vs. Non-Identity Keys
6. Foreign Key Naming
7. Primary Key Naming
   1. Clustered
8. View Naming
9. Table Naming
10. Surrogate keys vs. non-surrogate
    1. If a table is being FK referenced, a surrogate key is used
11. Why Film vs Movie in some instances?

Assumptions:

1. Award Shows are annual

## Design Rationale

One of my main objectives with my design was to try to be as consistent and standardized as possible with how entities and their relationships were modeled. These are some general design decisions that were made, and their rationale behind them:

1. Do use a single primary key value if the table is referenced by another entity (i.e. as a foreign key). The reasons for this are as follows:
   1. Simplify the maintenance and foreign key reference (child table only requires a single column as the foreign key to the parent table)
   2. This almost always saved on space as well in the child table (as the composite key would typically take up more space)
   3. MovieCastMember is an example of a table that does *not* have a single primary key value. This is because the table is not referenced by other tables, and it simplifies the maintenance of the table to not do so.
2. In conjunction with #1, make this single primary key **artificial** if at least one of the following scenarios apply:
   1. There is no natural key for the table
      1. E.g. The Movie and FilmMember tables are examples of this.
   2. The natural key for the table is greater than 16 bytes
      1. E.g. The “GenreName” column for the Genre table is a natural key, but is varchar(25) – which takes up more than 16 bytes. An artificial smallint could be used instead to significantly reduce the storage necessary for foreign key references.
   3. The artificial key would take up less space than the natural key
      1. E.g. the “Code” column in RestrictionRating is varchar(10), which is equal to 10 bytes + 2 bytes to hold the length (12 bytes total). However, a smallint (2 bytes) artificial key could be used to save 10 bytes per reference.
   4. The key value has a chance of changing (which would require cascading the update to child tables) and or reordering of the table’s clustered index
      1. Typically, Name/Text based columns *do* have a chance of changing, albeit small chance (e.g. GenreName, UserRoleName, etc.), and thus would require cascading updates to references.
      2. MovieImage is an example of a table that contains a Natural key of {MovieId, ImageName}. However, the ImageName has a fair chance of a changing, which would make for a more expensive write to maintain the table’s clustered index order.
   5. Note: The only table that is referenced by another table (meeting criteria #1) and does not have an artificial primary key is the [Language] table. This is because the LanguageIsoCode (ISO Code 639-1) is well-defined/standardized (unlikely to change), and only requires two bytes of storage (Adhering to conditions 2a, 2b, and 2c listed above).
3. Do use ‘smallint’ data type as a primary key for entities that will not have large amounts of data.
   1. The primary reason for this is that smallint is a 2-byte integer, which can hold a value up to 32,767 (more than enough for several tables in the database, such as AwardShow, FilmRole, Genre, ImageType, RestrictionRating, UserRole). Although storage is cheap, a movie database will contain a significant amount of data. Foreign key references back to tables with a smallint will store half the number of bytes than a normal int.
4. Do use an ‘int’ (max 231 value) data type as a primary key for entities with a medium amount of data (e.g. User)
5. Do use a ‘bigint’ (max 263 value) data type as a primary key for entities with potentially large amounts of data (e.g. Movie, MovieImage, MovieCastMember, etc.)
6. Do use non-identity (in other words, assigned artificial IDs) for certain smallint/static tables that regular users cannot modify (e.g. Genres, RestrictionRatings, MovieImageType, UserRoles, AwardShows, etc).
   1. These tables should be managed by the Development Team members, and they will not be updated often. In addition, they may require code references to these certain ID values (as an enum) in order to conditionally add logic or functionality. An auto-incremented ID has no guarantee that the value will remain the same on all environments, since the value is assigned by the DBMS engine.
   2. For example, see how the UserRole table can match an enum class in code. The enum values can match the Id values in the actual table itself. An assigned artificial ID (non-auto-increment) allows for this consistent referencing:





1. Do use identity/auto-increment for all other artificial keys (on dynamic tables or even for certain relatively static tables that are not necessary to be treated with Enum values).
   1. This allows for less maintenance on the developer, and the DBMS has a reliable way of assigning IDs.
   2. E.g. Movie, FilmMember, AwardShowInstance, etc. contain identity/auto-increment artificial keys.
2. Do add non-clustered unique-index constraints on natural key values
   1. This enforces a constraint on the table, as well as makes lookups for the column faster.
3. Do add a clustered index constraint on the primary key of every table (unless there is a good reason not to)
   1. Every table in the database uses the primary key as the clustered key. For artificial and auto-increment keys, this is very easy on the DBMS engine to maintain the physical order of the records on the disk.

MovieDbLite specific rationale behind relationships and entities

1. A single table is used to store all members of the film industry: the FilmMember table.
   1. Did not split up the FilmMember table into separate, role-based tables (e.g. ActorTable, DirectorTable, etc.) because FilmMembers can (and many times do) participate in several different roles of the film industry. For example, Ben Affleck is an actor, but also a screen writer and producer for certain films.
2. The Movie and FilmMember tables contain a fair amount of NULL columns
   1. The Movie table does contain several NULLable columns, however, this is intentional to allow for announced/planned Movies to be tracked in the system before details are known.
   2. FilmMember contains a number of NULL columns as well, but many of these are optional fields for a FilmMember (such as a Prefix, Suffix, Middle Name, Biography, etc.) that either may not be known (i.e. lesser known figure may not have a Bio) or not relevant (i.e. not everyone has a name Suffix).
3. A required PreferredFullName column is added to the FilmMember table
   1. This allows for the film member to specify the stage name they go by (and want to be reported as in the website), while also entering their real First/Middle/LastNames in the respective fields.
   2. E.g. Ben Affleck’s actual name is Benjamin Geza Affleck-Boldt, but he does not go by that.
4. Movie cast and crew are stored in different places. The movie cast members (i.e. actors) are stored in the MovieCastMember table. The movie director is stored in the Movie table. The movie crew members (i.e. all remaining non-directors, non-actors that participated in the movie) are stored in the MovieCrewMember table. This one I went back and forth with, but ultimately there were a few reasons for this design:
   1. MovieCastMember allows us to store additional movie/actor specific information, such as the CharacterName. This CharacterName would not be relevant for any of the Crew members.
   2. MovieCastMembers will typically be searched, looked up, or reported more often than MovieCrewMembers (because users are typically more interested in the figure heads / actors of the movie). Therefore, splicing the data out into its own table makes for faster lookups – as opposed to it all being crammed into one MovieCastAndCrewMember table.
   3. A crew member in a movie can have multiple roles (e.g. producer and screenwriter), so therefore we specify the RoleId in the table and as part of the Primary Key. Also, there are several Movie roles, so it did not make sense to partition n-role amount of tables out (from a scalability standpoint).
   4. Films almost always have one director, and they are credited in that manner for movies (see [link](https://www.quora.com/Approximately-what-percent-of-films-have-more-than-one-director)). Exceptions are situations like established duos such as the Coen brothers, and this is a future requirement that could be added – which I mention in the 6 Concluding Remarks section. In the meantime, situations like that can involve either specifying the directors in the MovieCrewMember table, or specifying the primary director in Movie.DirectorFilmMemberId, and secondary one in the MovieCrewMember table. I still wanted the Director in the Movie table because this is arguably the most-important role in making the movie (outside of the actors), and will frequently be reported. Putting it directly on the Movie table when required less joins/logic to report, and because it meets 99% of the cases, it made sense.

## E/R Model

This section describes the E/R model for the database. The diagrams displayed were created using Visual Paradigm Community Edition, which is a useful tool for creating several types of diagrams (in addition to ERDs).

The following ERD is a diagram for the entire database and contains a Legend that explains the meaning for the visual icons/indicators. We will focus on four main components/portions of the overall diagram in this section, as these components encompass pieces of the primary functional requirements that can be individually explained.



### General Movie Information

The diagram in this section illustrates how general movie information and details are stored in the database. The parent-most entity that contains Movie information in this relationship is the Movie entity. The Movie entity contains a number of attributes that are “one-to-one” relational with a Movie (e.g. Title, Description, Release Date, etc.) Many of the columns of the Movie entity are nullable because data for the Movie may be entered prior to the movie being released (e.g. in early stages when the Movie is announced). However, it is expected that these attributes are eventually filled in as the details are known.

Child entities are created off of Movie (such as MovieImage, Movie\_Language, and Movie\_Genre) to store additional details related to a Movie. Movie\_Genre and Movie\_Language are many-to-many intersection tables; MovieImage is a one-to-many table (a Movie may have many Images). Language, Genre, and ImageType are other Parent entities used by the respective Movie’s child entity tables.



### User Access and Reviews

The diagram in this section illustrates how Users of MovieDbLite are stored, as well as how the ratings/reviews that user’s provide to Movies are stored. A User account can write a review for a Movie (stored in the MovieUserReview Entity), and other users can mark the Review as Helpful or Unhelpful (via MovieUserReviewHelpful). Each User is tied to a UserRole to indicate whether they are an Admin or regular User, which affects the permissions they have over certain areas of the application (although, both Users and Admins can write reviews for movies).

Note: Attributes from the Movie entity in the diagram below were removed for brevity. The important part was to show the relationship between MovieUserReviews and Movie.



### Movie Cast and Crew

The diagram in this section illustrates how the cast and crew of a movie are stored. The FilmMember parent table stores all the persons/members of the film industry that participate as cast or crew members in a movie. The Movie table itself stores the DirectorFilmMemberId (as there can only be one director per movie). The MovieCastMember is used to store the cast members (i.e. actors) of a movie and associated information about them (e.g. their CharacterName).

The MovieCrewMember stores the crew members of a movie. The parent table FilmRole is used to describe what the crew member’s role is for the given movie (e.g. screenwriter, costumer designer, etc.)



### Award Shows and Winners

The diagram in this section illustrates how the awards that are awarded for movies are stored. There are different types of AwardShows that have different types of Awards (every Award is associated with an AwardShow). The winners of awards are stored in the AwardWinner table. Each AwardWinner is associated with an AwardShowInstance, which is a parent entity that keeps track of annual award show events. An AwardWinner also is associated with a Movie that is associated with the award.

Note: Attributes from the Movie and FilmMember entity in the diagram below were removed for brevity. The important part was to show the relationship between AwardWinners and Movies / FilmMembers.



### Entities

This database consists of a set of static (admin/developer managed) and dynamic tables. Static tables can be defined as tables that are pre-defined and seeded with data upfront, and remain mostly static (meaning that the data in it is not added/modified/deleted often). The primary purpose of these tables is to avoid redundantly storing the names/descriptors for the values within other entities in the database (which may vary in length), where the value can simply be referenced by a static identifier. It also serves as a consistent/standard way of setting and viewing data in the database. For example, if these types did not exist, one may enter “Sci Fi” as a genre for a given movie, whereas another person may spell it out as “Science Fiction”. This would result in two different values being stored that mean the same thing.

The semantics about whether a table is static or dynamic is not too important, however, I chose to organize the following by it.

The static tables in the database, as well as their purpose, are as follows:

1. Award
   1. The purpose of this table is to track the different types of movie/film member awards that can be presented at annual award show events. Each award is tied to an Award Show. The Award contains a required AwardName column and optional Description.
   2. A Unique key/index constraint is added to the table for the combination of {AwardShowId, AwardName}. In other words, the AwardName should be unique for the given AwardShow. Consequently, this does also imply the AwardName by itself is not unique. For example, the Oscars and Critics’ Choice Awards both have an award named “Best Actress” (amongst other awards that also share the same names).
   3. This is considered a static table because (for the most part) the records in this table are not modified often, and have historically been fairly consistent on an annual basis when the award shows are hosted.



1. AwardShow
   1. The purpose of this table is to track the different types of annual movie/film award shows that present awards to movies/film members. The AwardShow contains a required (and unique) ShowName column and optional Description.
   2. This is considered a static table because (for the most part) the records in this table are not modified often, and it is not often the case that Award Shows change. The standard ones have been around for a while (e.g. Oscars, Golden Globes, Critics’ Choice Awards, etc.)



1. FilmRole
   1. The purpose of this table is to track the different role types of the personnel/film members that contribute to making films/movies. A FilmRole contains a required (and unique) RoleName column and optional Description.
   2. This is considered a static table because the records of this table will not be modified often, as the film roles in the industry have been well established over the years (e.g. Director, Writer, Producer, Actor, Costume Designer, etc.)



1. Genre
   1. The purpose of this table is to track the different genre types that movies can be classified by. A Genre contains a required (and unique) GenreName and optional Description.
   2. This is considered a static table because the records of this table will not be modified often, as it is uncommon that new movie genres will be created. The existing movie genres have remained relatively unchanged for a while now (e.g. Action, Comedy, Horror, Romance, Drama, etc.)



1. Image Type
   1. The purpose of this table is to track the different types of images that are supported to be uploaded to the database (e.g. for movie images / thumbnails). This contains required Image Extension (unique key) and Name columns. The ImageExtension is used to help determine the type of image file being uploaded, and if any special logic is necessary for saving/downloading the file’s contents.
   2. This is considered a static table because the records of this table will not be modified often. The different types of images have been well-established and stable for a while now.



1. Language
   1. The purpose of this table is to track the languages that the dialogue of a movie can be classified as. The primary key for the table is the Language Iso Code which is an [ISO Code 639-1](https://en.wikipedia.org/wiki/List_of_ISO_639-1_codes) specification (two letter code for denoting languages). This table also contains a required LanguageName column which is a user-friendly name of the column.
   2. This is considered a static table because the records of this table will not be modified often, as it is uncommon that new languages are created or established to be used in the film industry.



1. RestrictionRating
   1. The purpose of this table is to track the different viewer [restriction ratings](https://en.wikipedia.org/wiki/Motion_Picture_Association_of_America_film_rating_system) that a film is classified as by the Motion Picture Association of America (MPAA). This table contains all required fields (Code (Unique key), ShortDescription, LongDescription, and IsActive). IsActive can be used to indicate the rating is no longer used (e.g. rated “M”).
   2. This is considered a static table because the records of this table do not change often. The current MPAA ratings have not changed since 1996.
   3. You will notice how the columns match up to the properties of the described on the MPAA website [here](https://www.motionpictures.org/film-ratings/).



Figure 1: Green boxed indicates the ShortDescription. Red boxed indicates the Code. Blue boxed indicates part of the LongDescription.



1. UserRole
   1. The purpose of this table is to track the different roles that a user can be assigned to. The user’s role will drive the permissions they have for accessing or modifying data in MovieDbLite. This table contains required (and unique) RoleName and Description columns. The Description is intended to briefly describe the purpose of the role and the access restrictions.
   2. This is considered a static table because the records of this table do not change often (changes would normally involve new functionality or significant changes to areas of the website), and are managed by the development team.



Dynamic

1. AwardShowInstance
   1. The purpose of this table is to keep track of each instance of an AwardShow. In other words, the actual award show event that takes place at a particular year. Each AwardShowInstance contains an Award Show it is in an instance of. It also contains a required Year field (the year the awards are being presented for) and DateHosted field (the Date the event instance took place).
   2. A Unique key/index constraint is added to the table for the combination of {AwardShowId, Year}. In other words, there should only be one instance of an AwardShow for a given Year. It has been the standard for a long time that these Award Shows are done on an annual basis, which is the reasoning behind the constraint.
   3. This table could borderline be considered static; however, multiple entries will be added to this on at least an annual basis, and for that reason, it can be classified as dynamic.
2. AwardWinner
   1. The purpose of this table is to keep track of the winners/recipients of Awards of a particular Award Show Instance. Each AwardWinner record tracks the AwardShowInstance it belongs to, the actual Award that was won, the Film Member it was presented to, and the Movie it was for. All of these columns are required.
   2. The combination of {AwardShowInstance, AwardId, FilmMemberId} make up the PK for the table. Multiple film members may be awarded a particular award (at an award show instance), and so this column must make up the unique primary key. For example, sometimes multiple people are awarded Best Costume Design or Best Cinematography for a particular movie.
   3. Both a FilmMember and a Movie are required to be entered for a record. This is because each Award that is presented is toward a particular Movie (there are currently not awards that exist otherwise, or at least are not tracked by this database). In addition, every Award is accepted/presented to a certain Film Member. This includes “Best Picture” award that may seemingly appear to be a “movie-only” award. However, the Producer(s) accept this award on behalf of the Movie.
3. FilmMember
   1. The purpose of this table is to track all the persons participating in the making of movies/films. This table tracks several fields for each FilmMember, including a Prefix (e.g. Dr., Mr., Mrs., Ms., etc.) (optional), FirstName (required), MiddleName (optional), LastName (required), Suffix (e.g. Jr., Sr., IV, etc.) (optional), PreferredFullName (required), Gender (M or F) (required), DateOfBirth (optional), DateOfDeath (optional), and Biography. The PreferredFullName field is an important field because this is what is intended to be used in most areas of the application to display the name of the FilmMember. This field is synonymous with “StageName”, and may differ from the person’s concatenated prefix/first/middle/last/suffix name. E.g. Katheryn Elizabeth Hudson goes by “Katy Perry”.
   2. An index (non-unique) is added to the PreferredFullName field as it will be common for searches to be performed against this field.
4. Movie
   1. The purpose of this table is to track all the movies/films that are tracked in MovieDbLite. This table consists of several of the one-to-one attributes of a Movie, including a required Title and Description. The remaining fields are optional/nullable because they may not be known when the Movie is first entered into the database (e.g. an announced movie that will be released in the future). However, it is intended that these attributes are eventually populated when the information is known. These optional fields include the ReleaseDate, RestrictionRatingId, DirectorFilmMemberId, and DurationInMinutes. There is also one derived/calculated optional field, which is the AverageUserRating column. This contains an aggregated average based off the MovieUserReview table. More details will be discussed in the database triggers section pertaining to this AverageUserRating field.
   2. An index (non-unique) is added to the Title field as it will be common for searches to be performed against this field.
5. Movie\_Genre
   1. The purpose of this table is to track the genres associated with a movie. This is an intersection table used to define a many-to-many relationship between Movies and Genres. A movie may have multiple genres, and a genre may (and is most likely) associated with multiple movies.
6. Movie\_Language
   1. The purpose of this table is to track the different supported languages associated with a movie. For example, if a movie contains a significant amount of English and Spanish dialogue, then it may be classified as both. Similarly, if the movie is recorded and released in different languages, that relationship will be tracked in this table. This is an intersection table used to define a many-to-many relationship between Movies and Languages. A movie may have multiple languages, and a language may (and is most likely) associated with multiple movies.
7. MovieCastMember
   1. The purpose of this table is to track the different cast members (i.e. actors) associated with a movie. The Primary key consists of the Movie and FilmMember. A Movie Cast Member record also contains a required CharacterName attribute and optional Sequence. The CharacterName stores the name of the cast member in the film. The Sequence is used to order the actors in a listed manner if desired (e.g. by importance, significance of role, order of appearance, etc.) This table is not intended to store film members of the movie that do not act in a scene (e.g. crew members).
8. MovieCrewMember
   1. The purpose of this table is to track the different crew members (e.g. writers, producers, costume designers, etc.) that worked on the film. The Primary key consists of the Movie, FilmMember, and FilmRole that the corresponding film member participated in. It is possible that the same FilmMember shows up in this table more than once for a movie, if they participated in more than one role when making the movie (e.g. if a film member was both a producer and screenwriter). In addition, it is possible that more than one film member of a certain film role worked on the movie (e.g. if multiple producers produced a certain movie). It is also possible that a member of the MovieCastMember table is also a member of the MovieCrewMember table (e.g. if one of the actors is also one of the writers).
   2. This table is intended to be used to store all roles aside from Actors and Directors, since those are stored within MovieCastMember and Movie.DirectorFilmMemberId respectively.
9. MovieImage
   1. The purpose of this table is to track the different uploaded images (posters/thumbnails/action photos/etc.) associated with a movie. This table contains a unique composite key of {MovieId, ImageName} such that each uploaded image for a movie must be provided a unique name to distinguish it from others. In addition, the Image Type (required) (e.g. jpg, png, etc.) is stored with the uploaded movie image, as well as the raw binary File Contents (required) (using varbinary(max) in SQL Server). A DateUploaded (required) is stored as well for when the movie’s image was uploaded/associated. An optional Description is available if additional details are used to describe the image.
10. MovieUserReview
    1. The purpose of this table is to track different reviews/ratings that users give to a certain movie. This table contains a unique composite key of {MovieId, UserId} indicating that a certain user can write only one review for a given movie. The required Rating column is a value 1-10 (1 being the worst, 10 being the best) that the user assigns to the film. The Review column is an optional large text field if the user chooses to describe their review. Lastly, a required DatePosted date is stored with the review, which is useful for sorting circumstances.
    2. A database trigger is added to this table whenever a value changes to calculate and update the Movie.AverageUserRating column. More details will be described in the triggers section of this document.
11. MovieUserReviewHelpful
    1. The purpose of this table is to track the helpfulness of user reviews (based off what other users think are helpful reviews). This can be thought of as an upvote/downvote (thumbs up/thumbs down) type system for reviews. The MovieUserReview and UserId make apart the primary key. This means that a certain user could only give one opinion of whether a certain review is helpful or not. In other words, they cannot upvote or downvote a particular review multiple times (though they can change their upvote to a downvote, and vice-versa, if they wish). The required IsHelpful column is used to store whether the user upvoted or downvoted the review. TODO: Can a user upvote their own review?
12. User
    1. The purpose of this table is to track all of the users in MovieDbLite. Each user creates a required (unique key) UserName to use for logins. A required email is also stored to use for email communication with the user. The user also enters a password upon account creation, and the hashed version of it is stored in the database. More details about that are discussed in the security section of this document. Lastly, each user is assigned to one and only one User Role. This user role dictates their permissions to the application.

### Relationships

This section describes the relationships between entities in MovieDbLite.

#### 3.4.6.1 Award to Award Show

* **Description:** An award is associated with one and only one award show. An award show can have zero to many awards.
* **Layman’s Terms:** This relationship tracks how an award show can award out many different awards.
* **Cardinality:** This is a one-to-many relationship.
* **Parent Entity:** Award Show
* **Child Entity**: Award
* **Participation:** The Award (being the child entity) has **total participation** in the relationship, and the Award Show (being the parent entity) has **partial participation** in the relationship.
* **Identification:** This relationship is a **non-identifying** relationship.

#### AwardShowInstance to AwardShow

* **Description:** An award show instance is associated with one and only one award show. An award show can have zero to many award show instances.
* **Layman’s Terms:** This relationship tracks how award shows can have many different events (instances) over time.
* **Cardinality:** This is a one-to-many relationship.
* **Parent Entity:** Award Show
* **Child Entity**: Award Show Instance
* **Participation:** The Award Show Instance (being the child entity) has **total participation** in the relationship, and the Award Show (being the parent entity) has **partial participation** in the relationship.
* **Identification:** This relationship is a **non-identifying** relationship.

#### AwardShowInstance to Award to FilmMember (AwardWinner)

* **Description:** An award show instance can give out zero to many awards to film members.
* **Layman’s Terms:** An award show instance/event can give out many awards. Each award could be given to one to many film members (i.e. some awards are given to multiple persons on the same film).
* **Cardinality:** This is a unique case of a tertiary (three column) primary key. The AwardShowInstance, Award, and FilmMember respectively have a one-to-many relationship with the AwardWinner table. However, the AwardWinner entity represents a many-to-many-to-many relationship between Awards, AwardShowInstances, and FilmMembers respectively.
* **Parent Entities:** Award Show Instance (Strong), Award (Strong), and Film Member (Strong)
* **Child Entity**: AwardWinner (Weak)
* The AwardWinner is an intersection entity used to store the many-to-many-to-many relationship between the three entities.
* The relationships from AwardShowInstance, Award, and FilmMember to child AwardWinner are **identifying** relationships because the child entity could not exist without the three parent entities.
* The relationships from AwardShowInstance, Award, and FilmMember to child AwardWinner are **partial participation** because not necessarily all AwardShowInstances, Awards, and FilmMembers may be a part of the relationship/intersection entity.
* The relationships from AwardWinner to AwardShowInstance, Award, and FilmMember are **total participation**

#### AwardWinner to Movie

* **Description:** An award winner record is associated with one and only one movie. A movie can have zero or many award winners.
* **Layman’s Terms:** A movie can win multiple awards (but may win none).
* **Cardinality:** This is a one-to-many relationship.
* **Parent Entity:** Movie
* **Child Entity**: Award Winner
* **Participation:** The Award Winner (being the child entity) has **total participation** in the relationship, and the Movie (being the parent entity) has **partial participation** in the relationship.
* **Identification:** This relationship is a **non-identifying** relationship. The MovieId is not a part of the AwardWinner’s Primary Key.

#### Movie to FilmMember (Director)

* **Description:** A movie can have zero or one film member that is the director. A director can direct zero or many movies.
* **Layman’s Terms:** A director can direct zero or many movies. A movie can have at most one director. A movie may have no director before production has started.
* **Cardinality:** This is a one-to-many relationship.
* **Parent Entity:** Film Member
* **Child Entity**: Movie
* **Participation:** The Movie has **partial participation** in the relationship, and the Film Member has **partial participation** in the relationship.
* **Identification:** This relationship is a **non-identifying** relationship.

#### Movie to RestrictionRating

* **Description:** A movie can have zero or MPAA viewer restriction rating. A restriction rating can be associated with zero or many movies.
* **Layman’s Terms:** A movie can be assigned at most one restriction rating (e.g. Titanic is PG-13). It may not be assigned (e.g. if not released to theatres or not yet released). There can be many movies of a certain restriction rating (e.g. there are many rated G movies)
* **Cardinality:** This is a one-to-many relationship.
* **Parent Entity:** Restriction Rating
* **Child Entity**: Movie
* **Participation:** The Movie has **partial participation** in the relationship, and the Restriction Rating has **partial participation** in the relationship.
* **Identification:** This relationship is a **non-identifying** relationship.

#### Movie to Genre (Movie\_Genre Entity)

* **Description:** A movie can be classified as zero or many genres. A genre can be associated with zero or many movies.
* **Layman’s Terms:** This tracks the genres that are associated with movies.
* **Cardinality:** This is a many-to-many relationship between movies and genres
* **Parent Entities:** Movie (Strong) and Genre (Strong)
* **Child Entity**: Movie\_Genre (Weak)
* The Movie\_Genre is an intersection entity used to store the many-to-many relationship between the two entities.
* The relationships from Movie and Genre to child Movie\_Genre are **identifying** relationships because the child entity could not exist without the two parent entities
* The relationships from Movie and Genre to child Movie\_Genre are **partial participation** because not necessarily all Movies or Genres may be a part of the relationship/intersection entity.
* The relationships from Movie\_Genre to Movie and Genre are **total participation** (as normal for the required attribute in the child entity of a relationship)

#### Movie to Language (Movie\_Language Entity)

* **Description:** A movie can have dialogue in zero or many languages. A language can be associated with zero or many movies.
* **Layman’s Terms:** This tracks the dialogue languages that are associated with movies.
* **Cardinality:** This is a many-to-many relationship between movies and languages.
* **Parent Entities:** Movie (Strong) and Language (Strong)
* **Child Entity**: Movie\_Language (Weak)
* The Movie\_Language is an intersection entity used to store the many-to-many relationship between the two entities.
* The relationships from Movie and Language to child Movie\_Language are **identifying** relationships because the child entity could not exist without the two parent entities
* The relationships from Movie and Language to child Movie\_Language are **partial participation** because not necessarily all Movies or Languages may be a part of the relationship/intersection entity.
* The relationships from Movie\_Language to Movie and Language are **total participation** (as normal for the required attribute in the child entity of a relationship)

#### Movie to FilmMember (MovieCastMember)

* **Description:** A movie can have zero to many actors in it. An actor can act/cast in zero to many movies.
* **Layman’s Terms:** This relationship tracks the actors that act in movies.
* **Cardinality:** This is a many-to-many relationship between movies and film members (in the form of actors).
* **Parent Entities:** Movie (Strong) and Film Member (Strong)
* **Child Entity**: MovieCastMember (Weak)
* The MovieCastMember is an intersection entity used to store the many-to-many relationship between the two entities.
* The relationships from Movie and FilmMember to child MovieCastMember are **identifying** relationships because the child entity could not exist without the two parent entities
* The relationships from Movie and FilmMember to child MovieCastMember are **partial participation** because not necessarily all Movies or FilmMembers may be a part of the relationship/intersection entity. E.g. some FilmMembers may not be actors.
* The relationships from MovieCastMember to Movie and FilmMember are **total participation**

#### Movie to FilmMember to FilmRole (MovieCrewMember)

* **Description:** A movie can have multiple film/crew members of different roles.
* **Layman’s Terms:** A movie can have multiple crew members, and the crew member could work on one to many film roles for the movie. (e.g. a producer may also be a screenwriter for the film).
* **Cardinality:** This is a unique case of a tertiary (three column) primary key. The Movie, FilmMember, and FilmRole respectively have a one-to-many relationship with the MovieCrewMember table. However, the MovieCrewMember entity represents a many-to-many-to-many relationship between Movies, FilmMembers, and FilmRoles respectively.
* **Parent Entities:** Movie (Strong), Film Member (Strong), and Film Role (Strong)
* **Child Entity**: Movie Crew Member (Weak)
* The MovieCrewMember is an intersection entity used to store the many-to-many-to-many relationship between the three entities.
* The relationships from Movie, FilmMember, and FilmRole to child MovieCrewMember are **identifying** relationships because the child entity could not exist without the three parent entities.
* The relationships from Movie, FilmMember, and FilmRole to child MovieCrewMember are **partial participation** because not necessarily all Movies, FilmMembers, and FilmRoles may be a part of the relationship/intersection entity.
* The relationships from MovieCrewMember to Movie, FilmMember, and FilmRole are **total participation**

#### MovieImage to ImageType

* **Description:** A movie image is associated with one and only one image type. An image type can be associated with zero or many movie images.
* **Layman’s Terms:** This relationship tracks the image type (e.g. .jpg, .png) associated with a movie image.
* **Cardinality:** This is a one-to-many relationship.
* **Parent Entity:** Image Type
* **Child Entity**: Movie Image
* **Participation:** The Movie Image (being the child entity) has **total participation** in the relationship, and the Image Type (being the parent entity) has **partial participation** in the relationship.
* **Identification:** This relationship is a **non-identifying** relationship.

#### MovieImage to Movie

* **Description:** A movie image is associated with one and only one movie. A movie can have zero or many movie images.
* **Layman’s Terms:** This relationship tracks the image associated with a movie.
* **Cardinality:** This is a one-to-many relationship.
* **Parent Entity:** Movie
* **Child Entity**: Movie Image
* **Participation:** The Movie Image (being the child entity) has **total participation** in the relationship, and the Movie (being the parent entity) has **partial participation** in the relationship.
* **Identification:** This relationship is a **non-identifying** relationship.

#### Movie to User (MovieUserReview)

* **Description:** A user can write zero to many reviews for movies. A movie can reviews written by zero to many users.
* **Layman’s Terms:** This relationship tracks the user reviews written for movies.
* **Cardinality:** This is a many-to-many relationship between movies and users to track user movie reviews.
* **Parent Entities:** Movie (Strong) and User (Strong)
* **Child Entity**: MovieUserReview (Weak)
* The MovieUserReview is an intersection entity used to store the many-to-many relationship between the two entities.
* The relationships from Movie and User to child MovieUserReview are **non-identifying** relationships because the MovieUserReview entity has it’s own artificial key.
* The relationships from Movie and User to child MovieUserReview are **partial participation** because not necessarily all Movies or Users may be a part of the relationship/intersection entity. E.g. Some Users may not write reviews; or some movies may not have reviews written for it.
* The relationships from MovieUserReview to Movie and User are **total participation**

#### MovieUserReview to User (MovieUserReviewHelpful)

* **Description:** A movie user review can be marked helpful by zero to many users. A user can mark zero to many movie user reviews helpful.
* **Layman’s Terms:** This relationship tracks the movie user reviews that are marked helpful by users.
* **Cardinality:** This is a many-to-many relationship between movie user reviews and users to track the reviews that are helpful/unhelpful.
* **Parent Entities:** MovieUserReview (Strong) and User (Strong)
* **Child Entity**: MovieUserReviewHelpful (Weak)
* The MovieUserReviewHelpful is an intersection entity used to store the many-to-many relationship between the two entities.
* The relationships from MovieUserReview and User to child MovieUserReviewHelpful are **identifying** relationships because the child entity could not exist without the two parent entities
* The relationships from MovieUserReview and User to child MovieUserReviewHelpful are **partial participation** because not necessarily all MovieUserReviews or Users may be a part of the relationship/intersection entity. E.g. Some reviews may not be marked helpful; and some users may not mark any reviews as helpful.
* The relationships from MovieUserReviewHelpful to MovieUserReview and User are **total participation**

#### User to UserRole

* **Description:** A user is associated with one and only one user role. A user role can be assigned to zero or many users.
* **Layman’s Terms:** This relationship tracks the role associated with each user.
* **Cardinality:** This is a one-to-many relationship.
* **Parent Entity:** User Role
* **Child Entity**: User
* **Participation:** The User (being the child entity) has **total participation** in the relationship, and the User Role (being the parent entity) has **partial participation** in the relationship.
* **Identification:** This relationship is a **non-identifying** relationship.

### E/R Diagram

E/R diagram

## Relational Model

### Data Dictionary

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Column Name** | **Description** | **Data Type** | **Size** | **Constraint Type** | **Not Null?** | **Valid Values** |
| Id | Award Id - Primary surrogate key | int |  | Primary Key | Y |  |
| AwardShowId | The award show that grants this award | smallint |  | Foreign Key Part of Unique Key | Y |  |
| AwardName | Name of the award (e.g. Best Actor) | varchar | 50 | Part of Unique Key | Y |  |
| Description | Description of award | varchar | 200 |  | N |  |
| Id | Award show Id - Primary surrogate key | smallint |  | Primary Key | Y |  |
| ShowName | Name of award show (e.g. Oscars) | varchar | 50 | Unique Key | Y |  |
| Description | Description of award show and its purpose | varchar | 200 |  | N |  |
| Id | Award show instance Id - Primary surrogate key | int |  | Primary Key | Y |  |
| AwardShowId | The award show that this is an instance of. | smallint |  | Foreign Key Part of Unique Key | Y |  |
| Year | The film year the awards are presented for (this is typically the year prior to the host date year) | smallint |  | Part of Unique Key | Y | 4-digit positive value |
| DateHosted | Date the award show instance takes or took place | date |  | Constraint Type | Y |  |
| AwardShowInstanceId | The award show instance the winner won an award from | int |  | Part of Primary Key Foreign Key | Y |  |
| AwardId | The award the winner won | int |  | Part of Primary Key Foreign Key | Y |  |
| FilmMemberId | The winner's film member Id | bigint |  | Part of Primary Key Foreign Key | Y |  |
| MovieId | The movie the award belongs to | bigint |  | Foreign Key | Y |  |
| Id | Film Member Id - Primary Surrogate Key | bigint |  | Primary Key | Y |  |
| Prefix | Optional prefix for name (e.g. Mr., Mrs., Dr., Prince, etc.) | varchar | 10 |  | N |  |
| FirstName | First name of film member | varchar | 50 |  | Y |  |
| MiddleName | Middle name of film member | varchar | 50 |  | N |  |
| LastName | Last name of film member | varchar | 50 |  | Y |  |
| Suffix | Optional suffix for film member (e.g. Jr, Sr., Jnr., Snr., III, IV). | varchar | 5 |  | N |  |
| PreferredFullName | The required preferred name for the film member, which is how they would like to be reported to others (e.g. Katy Perry instead of Katheryn Hudson) | varchar | 150 | Non-Clustered Index | Y |  |
| Gender | The film member's gender | char | 1 |  | Y | 'M' or 'F' |
| DateOfBirth | The film member's date of birth | date |  |  | N |  |
| DateOfDeath | The film member's date of death | date |  |  | N |  |
| Biography | A biography of variable length for the film member | varchar | max |  | N | No maximum length is specified. |
| Id | Film role id - primary surrogate key | smallint |  | Primary Key | Y |  |
| RoleName | The user-friendly role name for a film role (e.g. Actor, Producer, Costume Designer) | varchar | 50 | Unique Key | Y |  |
| Description | Description of the film role's purpose | varchar | 200 |  | N |  |
| Id | Genre Id - primary surrogate key | smallint |  | Primary Key | Y |  |
| GenreName | User-friendly genre name for a film (e.g. Action, Horror, Drama) | varchar | 25 | Unique Key | Y |  |
| Description | Description of the genre | varchar | 500 |  | N |  |
| Id | Image type id - primary surrogate key | int |  | Primary Key | Y |  |
| ImageExtension | File extension for the image (e.g. .jpg, .png) | varchar | 10 | Unique Key | Y | Extension starts with period |
| Name | Image type name (e.g. JPEG, PNG) | varchar | 25 |  | Y |  |
| LanguageIsoCode | ISO-639-1 code that is a standardized nomenclature used to classify languages (e.g. "en", "fr", "es") | char | 2 | Primary Key | Y | Two-letter code |
| LanguageName | User-friendly name for language (e.g. English, French, Spanish) | varchar | 50 |  | Y |  |
| Id | Movie id - primary surrogate key | bigint |  | Primary Key | Y |  |
| Title | The movie title name (e.g. Pulp Fiction) | varchar | 150 | Non-Clustered Index | Y |  |
| Description | A description of the movie that potentially costs of high-level plot details | varchar | 500 |  | Y |  |
| ReleaseDate | The scheduled or a actual release date for the movie | date |  |  | N |  |
| RestrictionRatingId | The restriction rating of the given movie | smallint |  | Foreign Key | N |  |
| DirectorFilmMemberId | The director for the movie | bigint |  | Foreign Key | N |  |
| DurationInMinutes | The duration of the movie, in minutes (e.g. 154) | int |  |  | N | A positive integer value that describes the movie length in minutes. |
| AverageUserRating | An average of all user ratings for the movie (e.g. 8.9) | decimal |  |  | N | A decimal value between 1.0-10.0 |
| MovieId | The movie id that is associated with the given genre | bigint |  | Part of Primary Key Foreign Key | Y |  |
| GenreId | The genre that is associated with the given movie | smallint |  | Part of Primary Key Foreign Key | Y |  |
| MovieId | The movie id that is associated with the given language. | bigint |  | Part of Primary Key Foreign Key | Y |  |
| LanguageIsoCode | The language id that is associated with the given movie. | char | 2 | Part of Primary Key Foreign Key | Y | Two-letter code |
| MovieId | The movie id that the cast member is a part of | bigint |  | Part of Primary Key Foreign Key | Y |  |
| ActorFilmMemberId | The actor/cast film member id | bigint |  | Part of Primary Key Foreign Key | Y |  |
| CharacterName | The actor's character name in the given movie (e.g. Tony Montana) | varchar | 150 |  | N |  |
| Sequence | A sequence number associated the movie cast member to order/rank the cast members of a given movie | int |  |  | N | A positive integer value. |
| MovieId | The movie id that the crew member is a part id. | bigint |  | Part of Primary Key Foreign Key | Y |  |
| FilmMemberId | The crew member's film member id. | bigint |  | Part of Primary Key Foreign Key | Y |  |
| FilmRoleId | The role that the film member participated in when making the movie. | smallint |  | Part of Primary Key Foreign Key | Y |  |
| Id | Move image id - primary surrogate key | bigint |  | Primary Key | Y |  |
| MovieId | The movie id that the image is associated with | bigint |  | Foreign Key Part of Unique Key | Y |  |
| ImageName | The user-friendly name given to the image | varchar | 100 | Part of Unique Key | Y |  |
| ImageTypeId | The image type (e.g. jpg, png) | int |  | Foreign Key | Y |  |
| Description | Description of the image | varchar | 500 |  | N |  |
| FileContents | The actual raw binary contents of the image | varbinary | max |  | Y | Raw binary contents of the file |
| DateUploaded | The date the image was uploaded | datetime2 |  |  | Y |  |
| Id | Movie user review id - primary surrogate key | bigint |  | Primary Key | Y |  |
| MovieId | The movie the review is associated with | bigint |  | Foreign Key Part of Unique Key | Y |  |
| UserId | The user that wrote the review | int |  | Foreign Key Part of Unique Key | Y |  |
| Rating | The rating the user wrote for the review (e.g. 8) | smallint |  |  | Y | An integer value from 1-10 |
| Review | The optional review text that the user wrote | varchar | 8000 |  | N |  |
| DatePosted | The date the review was posted | datetime2 |  |  | Y |  |
| MovieUserReviewId | The movie user review that was deemed helpful. | bigint |  | Part of Primary Key Foreign Key | Y |  |
| UserId | The user that marked the review helpful | int |  | Part of Primary Key | Y |  |
| IsHelpful | Whether the review was actually helpful or unhelpful, respective of the bit field | bit |  |  | Y |  |
| Id | Restriction rating id - primary surrogate key | smallint |  | Primary Key | Y |  |
| Code | The MPAA restriction rating that can be assigned to a movie (e.g. PG-13) | varchar | 10 | Unique Key | Y | A valid MPAA film rating code (G, PG, PG-13, R, etc.) |
| ShortDescription | A short description of the rating (e.g. "General Audiences") | varchar | 50 |  | Y |  |
| LongDescription | A more detailed description of the rating | varchar | 200 |  | Y |  |
| IsActive | An active flag for whether or not the rating is still used. Certain ratings (e.g. "M") are no longer used. | bit |  |  | Y |  |
| Id | User Id - Primary surrogate key | int |  | Primary Key | Y |  |
| UserRoleId | The user's role for the MovieDbLite application | smallint |  | Foreign Key | Y |  |
| UserName | A username/handle chosen by the user (e.g. stevo9510) | varchar | 25 | Unique Key | Y | Letters and numbers only (no special characters) |
| EmailAddress | The user's email address | varchar | 255 |  | Y | Must be in a valid email format. |
| HashedPassword | The user's password secured with a bcrypt hash | varchar | 60 |  | Y | The password must contain a capital letter, lowercase letter, number, and be 10 or more characters. The password is a bcrypt hashed value (up to 60 chars) |
| Id | User role id - primary surrogate key | smallint |  | Primary Key | Y |  |
| RoleName | The role name for the role (e.g. Admin, User, etc.) | varchar | 25 | Unique Key | Y |  |
| Description | A description of the role and its general permissions | varchar | 500 |  | Y |  |

### Integrity Rules

Data integrity of the database involved utilizing some of the features of the RDBMS environment.

The basic and essential features of data integrity started with specifying each field as either required (via NOT NULL) or optional (NULLable). The mindset used during initial design was to start/default with each field as NOT NULL (required), and then figure out which fields may be optional (or may not have a value immediately known). This led to a database where ~80% of the columns (67/85) are specified as NOT NULL, and only ~20% of the columns are NULLable. Generally, the more enforcement of NOT NULL in the database, the better for data integrity – because it means that all the required/essential data must be entered for the record to commit. The majority of the NULLable columns fall into these tables, where it has minimal impact on the functionality of the system when they are null:

* Movie – A number of these columns are NULL because they may not be known when the movie is first announced. However, it is expected these are eventually filled in once the details are known
* FilmMember – A number of these columns are truly optional (e.g. not everyone has a Prefix or Suffix for their name, nor is the date of birth known for every film member that is credited in movies)

In addition to nullability constraints, basic data type constraints are enforced as well on columns based on their appropriate domain. For example, the FilmMember.DateOfBirth column is of the data type of ‘Date’ to ensure correctly formatted dates are entered into the field. In addition, ‘bit’ fields are used for Boolean (yes or no) values, such as MovieUserReviewHelpful.IsHelpful or RestrictionRating.IsActive. If values are numeric, such as Movie.DurationInMinutes, an approach numeric data type (such as ‘int’ in this scenario) is used. For Movie.AverageUserRating, the decimal(5, 2) type is used, because we only want to go to ‘2’ decimal places for the scale. The precision of ‘5’ is more than enough since our range of ratings is 1-10. The data type of char() is used when the amount of characters for the value is fixed (e.g. language iso code is a char(2)). Other data types (such as varchar and varbinary) are used as appropriate for text and binary related fields.

Referential integrity is established via the use of Foreign Key constraints in the RDBMS engine. All relationships described in the 3.4.6 Relationships section are referentially enforced by foreign key constraints in SQL Server. In addition to referential integrity constraints, the database enforces Unique Key constraints on natural key values for each table. For example, an Award’s unique key is a combination of the {AwardShowId, AwardName}. A MovieUserReview’s unique key is a combination of the {MovieId, UserId} because the same user cannot write more than one review for a given movie. Further, many of the static tables have the corresponding entity’s name column as the unique key. For example, the Genre’s GenreName column is the unique natural key. Two genres with the same name would be indistinguishable if reported to the user.

In addition to some of the basic constraints above, more advanced DBMS constraints were used. These were a few examples of constraints used:

1. The FilmMember Gender field allows for either an ‘M’ for Male or ‘F’ or Female. A special constraint was added to the column to enforce it.

ADD CONSTRAINT [CK\_FilmMember\_Gender] CHECK (([Gender]='M' OR [Gender]='F'))

If you try to perform an insert/update that violates this constraint, the DBMS enforces it. For example:

update FilmMember

set Gender = 'B'

where PreferredFullName = 'Leonardo DiCaprio'

Results in this error message, and does not commit to the database:

The UPDATE statement conflicted with the CHECK constraint "CK\_FilmMember\_Gender".

1. The MovieUserReview Rating column allows for a value between 1 and 10 inclusively. A special constraint was added to the column to enforce it.

ADD CONSTRAINT [CK\_MovieUserReview\_Rating] CHECK (([Rating]>=(1) AND [Rating]<=(10)))

Similarly, an insert/update to this with an invalid value will violate the constraint and not commit the change:

INSERT INTO MovieUserReview

(

MovieId,

UserId,

Rating,

Review,

DatePosted

)

VALUES

(

@MovieId,

@UserId,

12, -- Invalid rating (not between 1 and 10)

'great movie 12 out of 10',

SYSUTCDATETIME()

)

1. The AwardShowInstance [Year] column is enforced as a positive number with a length of four (we can reasonably assume that the database will not be used after the year 9999, and there were no Award show instances prior to the year 1000). This constraint could arguably be stricter (e.g. between 1900 and 9999) but, this is good enough to catch some basic typos or mistakes.

ADD CONSTRAINT [CK\_AwardShowInstance\_Year] CHECK (([Year]>(0) AND len([Year])=(4)))

1. The Language LanguageIsoCode column / primary key was created as a special User-Defined Data Type called ‘iso\_code639\_1’. The Language [ISO Code 639-1](https://en.wikipedia.org/wiki/List_of_ISO_639-1_codes) are strictly defined as two-letter codes to identify different languages.

The base-type is defined as this (a two-letter char instead of varchar since the length must be two).

CREATE TYPE [dbo].[iso\_code639\_1] FROM [char](2) NOT NULL

Then, a custom rule was created and applied to the Data-Type to enforce that the specified values are two-letters (and no numeric digits are used).

CREATE RULE [dbo].[iso\_code\_rule]

AS

@iso\_code639\_1 like '[A-Z][A-Z]'

The rule is bound to the data type using the sp\_bindrule system-level stored procedure:

EXEC sp\_bindrule 'iso\_code\_rule', 'iso\_code639\_1';

The Language.LanguageIsoCode column is specified as the iso\_code639\_1 type, as seen below. Similarly, the Movie\_Language.LanguageIsoCode column on that table is defined as the same data type.

CREATE TABLE [dbo].[Language](

[LanguageIsoCode] [dbo].[iso\_code639\_1] NOT NULL,

[LanguageName] [varchar](50) NOT NULL,

…

Now, any inserts/updates to the LanguageIsoCode column that violates the two-letter rule will be enforced by the DBMS. For example:

INSERT INTO [Language]

(

LanguageIsoCode,

LanguageName

)

VALUES

(

'p5', -- This contains a number, which violates the rule

'Pig Latin'

)

Results in the following error output:

A column insert or update conflicts with a rule imposed by a previous CREATE RULE statement. The statement was terminated.

1. Other constraints in the system are handled either at the stored procedure/trigger level, or within higher-level language application logic depending on complexity. For example, email address format can be tricky to validate within the database. A regex within the application can be used to validate the user’s input before it reaches the database to ensure it is a valid email address.

### Operational Rules

What are the constraints for some operations? For example, will the users be able to delete a patron’s information if he/she has outstanding videos? Can a patron be associated multiple records of checked-out videos? Will the application allow the users to enter a patron if the patron has the same first name and last name as an existing patron in the system?

* UserName constraints

### Operations

Describe what operations are involved for a particular use case. For example, does checking out videos involve insert/delete/update/retrieve?

## Security

This section covers the security provisions taken by MovieDbLite to ensure security of data and user operations within the system.

### 3.4.1. Password Storage

Users have the ability to create user accounts to rate and review movies that they have watched. User accounts require a username, email, and password. The password is transmitted over the network using HTTPS (see upcoming section), and is never stored as a plaintext in the database. Instead, a **hashed** version of the password is stored in the database. Hash functions are useful because they cannot be algorithmically reversed back to plaintext.

When the user attempts to authenticate/login to the system, the typed/entered password is transmitted to the server (via HTTPS) and hashed using the same algorithm that was used to store the password in the database. The hashed entered/typed password is then compared against the hashed value in the database to see if they match, and if so, the user has successfully authenticated.

The hash algorithm used is [BCrypt](https://en.wikipedia.org/wiki/Bcrypt), implemented by the BCrypt.Net-Core package/library available in Microsoft’s NuGet Package Repository. BCrypt is a modern hashing algorithm that uses a salt/work factor to mitigate against [rainbow table](https://en.wikipedia.org/wiki/Rainbow_table) attacks (i.e. precomputed hash tables for reverse engineering hashes). It is a recommend password hashing algorithm by many sources, including this [one](https://auth0.com/blog/hashing-in-action-understanding-bcrypt/) on the Auth0 website.

The following is a snippet of the BCrypt library in use for MovieDbLite. This example shows a method that takes the unhashed (plaintext) password, and then returns the hashed version of it using BCrypt and a work factor/salt of 15.



* Note that we would normally not want to share the work factor/salt or make it known to others (for security reasons), however, for the intentions of this project I wanted to show how it worked with a real value.

### SQL Injection

MovieDbLite prevents SQL Injection attacks through the use of reliable .NET framework libraries and tools, such as ADO.NET and Entity Framework ORM (Object Relational Mapping) functionality.

ADO.NET contains [SqlCommand](https://docs.microsoft.com/en-us/dotnet/api/system.data.sqlclient.sqlcommand?view=netframework-4.8) and [SqlParameter](https://docs.microsoft.com/en-us/dotnet/api/system.data.sqlclient.sqlparameter?view=netframework-4.8) classes that prevent SQL Injection attacks by parameterizing any arguments to pass to a stored procedure. When these classes are used for a SQL Server database connection, they [abstract](https://stackoverflow.com/a/4892205) away functionality to call into the sp\_executesql system-level stored procedure in SQL Server, which isolates the parameters and treats them as data.

Here is an example of usage of these libraries in the MovieDbLite web application to insert an award winner (via the usp\_InsertAwardWinner stored procedure).



Additionally, other operations in MovieDbLite utilize the Entity Framework ORM technology to perform CRUD. Entity framework is built with SQL injection protection by way of using its functions, allowing the developer to not have to put too much effort in mitigating the risk. The example below shows how an AwardShowInstance is added in MovieDbLite utilizing Entity Framework’s DbContext object (see MovieDbLiteContext \_context). The new instance is added to the database by simply adding it to the context’s collection (via Add method), and then calling SaveChangesAsync() against the context. No raw SQL is used in this operation, and consequently there is no SQL injection risk.



### Database Connection Strings

The database connection string for SQL Server utilizes Windows Authentication, also known as Integrated Security (SSPI). This security model is tightly integrated with Windows user and group accounts that are assigned and allowed to access the database.

The primary advantage of this security model is that there is no password management involved, and subsequently, no need to store SQL authentication user and password for the application. Instead, the application authenticates via secure, underlying Windows OS functionality based on the user account that is attempting to access the database. The user account groups that are granted access to the database are managed by administrators and permissions to the database will be limited. Further, groups can be assigned to particular SQL Server [Database-Level Roles](https://docs.microsoft.com/en-us/sql/relational-databases/security/authentication-access/database-level-roles?view=sql-server-ver15) which can limit the access the group member has to database object. For example, a group can be setup such that they have read-only access to the database, whereas another group can be setup to have full read/write access to the database.

For the deployed web application that requires access to the database, the Web Server’s Application Pool User (the proxy user that the web application runs on behalf of) will be granted read/write access to the database tables. Additionally, the App Pool user will be restricted from making any DDL changes (e.g. dropping tables, altering columns, etc.) because all of that should be performed by a DBA with allowed permissions.

Generally, Windows Authentication is a recommended authentication technique for SQL Server if it can be supported. More information about it can be found [here](https://docs.microsoft.com/en-us/dotnet/framework/data/adonet/sql/authentication-in-sql-server).

This is an example of the connection string for a Windows Authentication / SSPI based connection. Notice that there is no user name or password specified, as would be if it used a traditional SQL Authentication connection.



### Network Connection

The Website is hosted using HTTPS secure communication protocol over a computer network. This method of communication encrypts the network data while in-transit between the client and the server, preventing “man-in-the-middle” attacks.

## Database Backup and Recovery

### Automated Database Backups

A backup of the MovieDbLite database will be performed on a nightly basis, and output to the database server and a secondary archive server environment (to achieve some redundancy of the data in the event of a database server failure). These backups will be kept for at least 6 months in the event that recovery of the database is necessary. SQL Server has the ability to schedule automatic backups of the database via the SQL Server Agent Jobs function, as seen [here](https://www.sqlshack.com/multiple-methods-for-scheduling-a-sql-server-backup-automatically/).

Here is an example of a SQL script used to perform database backups for MovieDbLite. It timestamps the backup with the date to ensure no collisions with other backups, as well as identifying when the backup took place.



Output from backup:





### Full Database Recovery Model

The [recovery model](https://docs.microsoft.com/en-us/sql/relational-databases/backup-restore/recovery-models-sql-server?view=sql-server-ver15#RMov) for the database will be set to **Full**. This allows the database to be restored to an arbitrary point of time (e.g. before an error occurred) if that is necessary. The Full recovery model relies on transaction log files of the database to be created and stored. [Backups](https://docs.microsoft.com/en-us/sql/relational-databases/backup-restore/transaction-log-backups-sql-server?view=sql-server-ver15#LogBackupSequence) of the transaction log files will also be required by the DBA, otherwise the log files will grow forever. The backup process for a transaction log file is a similar process to the aforementioned process of backing up a database.

This is an example of how to set the Recovery Model to Full via the Database’s properties using SSMS.



### Source Control of Database and Application

## Using Database Design or CASE Tool

Software engineering tools provide automated or semiautomatic support for software development. A CASE (Computer-Aided Software Engineering) tool sometimes is more important than hardware for achieving good quality and productivity. Describe the tools you used to produce the artifacts for your project.

1. Visual Studio 2019
2. SQL Server Management Studio (SSMS) v17.8.1
3. Visual Paradigm v15.2
4. GitHub for Source Control

## Other Possible E/R Relationships

What were the other alternatives you considered when you designed your database?

# Implementation Description

General implementation requirements

## Data Dictionary

Data dictionary after you have your database set up. This dictionary is less detailed than the one in the previous section. Usually, a “DESCRIBE” operation in your database will provide the information needed for this section. If your RDBMS has a tool to show data dictionary; it is also fine too.

## Advanced Features

Describe any triggers, stored procedures, functions, or others used in the project to implement business rules specified in your database project; (You may include user interfaces with screen shots if you have implemented a database application).

## Queries

General query information (Your specific database queries or reports)

### Get Movie Details Report (Primary Details with Languages/Genres)

This query is used to report the primary details of the all movies in the database. Tables that may have multiple values per movie are comma delimited (e.g. genre and languages). These are pre-delimited ahead of time and inserted into temp tables, primarily for readability of the final report. Alternatively, we could have delimited them within final report itself if we grouped by all columns and used STRING\_AGG. We also do some formatting to convert the DurationInMinutes for the movie into Hours/Minutes.

The data that is returned from this query is user-friendly (no artificial IDs) and could be used for displaying a webpage with the movie’s information on it. This could easily be achieved by filtering the query by a provided MovieId.

Note that you may notice this corresponds with details that are populated by the usp\_InsertMovieDetails stored procedure.

-- Step #1: Create Temp tables to comma delimit multiple result tables

-- such as Languages and Genres

SELECT Movie\_Language.MovieId,

-- STRING\_AGG is a built-in function within SQL Server 2017 to delimit

-- the result based on the grouped column(s)

STRING\_AGG([Language].LanguageName, ', ') as Languages

INTO #MovieLanguages\_Delimited

FROM Movie\_Language

INNER JOIN [Language] ON [Language].LanguageIsoCode = Movie\_Language.LanguageIsoCode

GROUP BY Movie\_Language.MovieId

SELECT Movie\_Genre.MovieId,

STRING\_AGG(Genre.GenreName, ', ') as Genres

INTO #MovieGenres\_Delimited

FROM Movie\_Genre

INNER JOIN Genre ON Genre.Id = Movie\_Genre.GenreId

GROUP BY Movie\_Genre.MovieId

-- Step #2: Do actual report query

SELECT Movie.Title,

Movie.[Description],

Movie.ReleaseDate as 'Release Date',

RestrictionRating.Code as 'Restriction Rating',

Director.PreferredFullName as 'Director',

CASE WHEN Movie.DurationInMinutes IS NOT NULL THEN

-- FORMATMESSAGE works like a PRINTF() statement

FORMATMESSAGE('%dh %dmin', HrDuration.Hrs, MinDuration.Mins)

END as 'Duration',

Movie.AverageUserRating as 'Average User Rating',

#MovieLanguages\_Delimited.Languages,

#MovieGenres\_Delimited.Genres

FROM Movie

LEFT JOIN RestrictionRating ON RestrictionRating.Id = Movie.RestrictionRatingId

LEFT JOIN FilmMember Director ON Director.Id = Movie.DirectorFilmMemberId

-- Join on temp tables

LEFT JOIN #MovieLanguages\_Delimited ON #MovieLanguages\_Delimited.MovieId = Movie.Id

LEFT JOIN #MovieGenres\_Delimited ON #MovieGenres\_Delimited.MovieId = Movie.Id

-- Outer applys used here to calculate hours/mins.

-- This is used to keep the SELECT portion of the query cleaner

OUTER APPLY (SELECT Movie.DurationInMinutes / 60 as Hrs) HrDuration

OUTER APPLY (SELECT Movie.DurationInMinutes % 60 as Mins) MinDuration

-- Cleanup temp tables

DROP TABLE #MovieLanguages\_Delimited

DROP TABLE #MovieGenres\_Delimited

Results:



### Get All Cast and Crew of a Movie (Table-Valued Function)

This shows a Table-Valued Function used to combine all the movie film members (cast, crew, and director) of a particular movie. It takes in a MovieId and unions the film members from the corresponding MovieCastMember, MovieCrewMember, and Movie tables. The FilmRoleId is returned with each FilmMember to identify what role they played in the movie. The Director and Actor are implied based on their corresponding column and table (i.e. the Movie.DirectorFilmMemberId is a director role, and any value in the MovieCastMember table is an actor role).

This function is important to help retrieve the cast and credits for a particular movie. For example, a webpage that shows the movie’s cast/crew would utilize this function.

CREATE FUNCTION [dbo].[tvf\_GetAllMovieFilmMembers]

(

-- Add the parameters for the function here

@MovieId bigint

)

RETURNS TABLE

AS

RETURN

(

SELECT

MovieCastMember.ActorFilmMemberId as FilmMemberId,

cast(2 as int) as FilmRoleId, -- @FilmRoleId\_Actor

MovieCastMember.CharacterName,

MovieCastMember.Sequence

FROM MovieCastMember

WHERE MovieCastMember.MovieId = @MovieId

UNION

SELECT

MovieCrewMember.FilmMemberId,

MovieCrewMember.FilmRoleId,

NULL, -- CharacterName

NULL -- sequence

FROM MovieCrewMember

WHERE MovieCrewMember.MovieId = @MovieId

UNION

SELECT

Movie.DirectorFilmMemberId,

3, -- @FilmRoleId\_Director

NULL, -- CharacterName

NULL -- sequence

FROM Movie

WHERE Movie.Id = @MovieId

)

Here is an example of this Table-Valued Function being used to return details about the Titanic film’s cast and crew.

DECLARE @MovieId\_Titanic int = (SELECT id FROM Movie WHERE Title = 'Titanic' AND YEAR(ReleaseDate) = 1997)

SELECT \*

FROM tvf\_GetAllMovieFilmMembers(@MovieId\_Titanic) mFilmMember

Results:



### Get All Cast and Crew Member Names and Roles

This query returns all the cast and crew member names and role information (including character name for Actors). It utilizes the aforementioned tvf\_GetAllMovieFilmMembers table-valued function to get the foreign key ids that are used to retrieve FilmMember and FilmRole details. The results are sorted in order by Director, Actor, Sequence (for Actors), and then FilmRoleId for the remainder. CASE statement is used to prioritize the Directors and Actors first. ORDER BY operations in SQL Server default to sorting in ascending order (if not specified).

This report is a user-friendly view (e.g. no artificial IDs) of the cast and credits of a particular movie, and can be reported using this data on a webpage.

DECLARE @MovieId\_Titanic int = (SELECT id FROM Movie WHERE Title = 'Titanic' AND YEAR(ReleaseDate) = 1997)

SELECT FilmMember.PreferredFullName,

FilmRole.RoleName,

mFilmMember.CharacterName

FROM tvf\_GetAllMovieFilmMembers(@MovieId\_Titanic) mFilmMember

INNER JOIN FilmMember ON FilmMember.Id = mFilmMember.FilmMemberId

INNER JOIN FilmRole ON FilmRole.Id = mFilmMember.FilmRoleId

ORDER BY

-- List Director First

CASE WHEN mFilmMember.FilmRoleId = 3 -- @FilmRoleId\_Director

THEN 0

ELSE 1

END,

-- Prioritize Actors secondly

CASE WHEN mFilmMember.FilmRoleId = 2 -- @FilmMemberId\_Actor

THEN 0

ELSE 1

END,

-- Order Actors by Sequence

mFilmMember.Sequence,

-- Simply order by Role Id after this

mFilmMember.FilmRoleId

Results:



### Get all Movies that Film Member Participates in (Table-Valued Function)

This shows a Table-Valued Function used to combine all the movies a film member has been a part of, as well as the role that film member played in contributing to the movie. It takes in a FilmId and unions the Movies from the corresponding MovieCastMember, MovieCrewMember, and Movie tables. The FilmRoleId is returned with each Movie to identify what role they played in the movie. The Director and Actor are implied based on their corresponding column and table (i.e. the Movie.DirectorFilmMemberId is a director role, and any value in the MovieCastMember table is an actor role).

This function is important to help retrieve the movies a particular film member has participated in. For example, a webpage that shows the film member’s previous work would utilize this function.

-- =============================================

-- Author: Steven Anderson

-- Create date: 04/05/2020

-- Description: Get all Movies that a Film Member plays a role in (Cast, Crew, Director)

-- =============================================

CREATE FUNCTION [dbo].[tvf\_GetAllFilmMemberMovies]

(

-- Add the parameters for the function here

@FilmMemberId bigint

)

RETURNS TABLE

AS

RETURN

(

SELECT

MovieCastMember.MovieId,

cast(2 as int) as FilmRoleId, -- @FilmRoleId\_Actor

MovieCastMember.CharacterName

FROM MovieCastMember

WHERE MovieCastMember.ActorFilmMemberId = @FilmMemberId

UNION

SELECT

MovieCrewMember.MovieId,

MovieCrewMember.FilmRoleId,

NULL -- CharacterName

FROM MovieCrewMember

WHERE MovieCrewMember.FilmMemberId = @FilmMemberId

UNION

SELECT

Movie.Id,

3, -- @FilmRoleId\_Director

NULL -- CharacterName

FROM Movie

WHERE Movie.DirectorFilmMemberId = @FilmMemberId

)

GO

Usage of this function:

DECLARE @FilmMemberId\_Dicaprio bigint =

(SELECT id FROM FilmMember WHERE PreferredFullName = 'Leonardo DiCaprio')

SELECT \*

FROM dbo.tvf\_GetAllFilmMemberMovies(@FilmMemberId\_Dicaprio) f

Results:



### Report Movie and Role Information for Film Member

This query returns all the movie and role information (including character name for Actors) for a given film member (Leonardo DiCaprio in this case). It utilizes the aforementioned tvf\_GetAllFilmMemberMovies table-valued function to get the foreign key ids that are used to retrieve Movie and FilmRole details.

This report is a user-friendly view (e.g. no artificial IDs) of the movies of a particular film member, and can be reported using this data on a webpage.

SELECT Movie.Title,

FilmRole.RoleName,

FilmMemberMovies.CharacterName

FROM FilmMember

CROSS APPLY dbo.tvf\_GetAllFilmMemberMovies(FilmMember.Id) FilmMemberMovies

INNER JOIN Movie ON Movie.Id = FilmMemberMovies.MovieId

INNER JOIN FilmRole ON FilmRole.Id = FilmMemberMovies.FilmRoleId

WHERE FilmMember.PreferredFullName = 'Leonardo DiCaprio'

Results:



### Report the User Movie Reviews/Ratings

This query returns all the user movie reviews/ratings of a particular movie (The Departed in this case). An aggregation is performed to do a sum of all the upvotes/downvotes by users of each review. Normally, you cannot SUM bits in SQL server (you can only call SUM on a number type, such as an integer). However, we do a little trick here to convert the bit to an integer by multiplying the bit \* 1. Also, we use the bitwise negate operator (tilde ~) to sum the downvotes.

This report is a user-friendly display of all the reviews for a given movie.

DECLARE @MovieTitle varchar(150) = 'The Departed'

SELECT

[User].UserName,

MovieUserReview.DatePosted,

MovieUserReview.Rating,

MovieUserReview.Review,

SUM(MovieUserReviewHelpful.IsHelpful \* 1) as Upvotes,

SUM(~MovieUserReviewHelpful.IsHelpful \* 1) as Downvotes

FROM MovieUserReview

INNER JOIN Movie ON Movie.Id = MovieUserReview.MovieId

INNER JOIN [User] ON [User].Id = MovieUserReview.UserId

LEFT JOIN MovieUserReviewHelpful ON MovieUserReviewHelpful.MovieUserReviewId = MovieUserReview.Id

WHERE Movie.Title = @MovieTitle

GROUP BY

MovieUserReview.MovieId,

[User].UserName,

MovieUserReview.DatePosted,

MovieUserReview.Rating,

MovieUserReview.Review

ORDER BY

MovieUserReview.MovieId,

MovieUserReview.DatePosted

Results:



### Report the Helpful Vote Details of a Particular Review

This report provides more specifics on which users upvoted/downvoted a certain review. Using our previous report example, we examine the users that upvoted/downvoted Superman33’s review of the Departed. We first start by gathering the UserId/MovieId of the particular user review we want to lookup, and then use the ReviewId in the final query to gather the details. We sort by helpful (i.e. upvoted) reviews floating to the top.

This query would be useful to drill into more details about a particular movie review.

DECLARE @UserName varchar(25) = 'SuperMan33'

DECLARE @UserId int = (SELECT id FROM [User] WHERE UserName = @UserName)

DECLARE @MovieId bigint = (SELECT id FROM Movie WHERE Title = 'The Departed')

DECLARE @MovieUserReviewId bigint =

(SELECT id FROM MovieUserReview WHERE MovieId = @MovieId AND UserId = @UserId)

SELECT [User].UserName,

MovieUserReviewHelpful.IsHelpful

FROM MovieUserReviewHelpful

INNER JOIN [User] ON [User].Id = MovieUserReviewHelpful.UserId

WHERE MovieUserReviewHelpful.MovieUserReviewId = @MovieUserReviewId

ORDER BY IsHelpful DESC

Results:



### List defective videos

SELECT

FilmMember.FirstName,

FilmMember.MiddleName,

FilmMember.LastName,

MultiFilmMemberRoles.TotalNumberOfFilmRoles,

FilmMember.PreferredFullName,

FilmMember.Gender,

FilmMember.DateOfBirth,

FilmMember.Biography

FROM FilmMember

INNER JOIN (

SELECT FilmMember.Id as FilmMemberId,

COUNT(\*) as TotalNumberOfFilmRoles

FROM FilmMember

CROSS APPLY dbo.tvf\_GetAllFilmMemberMovies(FilmMember.Id) FilmMemberMovies

GROUP BY FilmMember.Id

HAVING COUNT(\*) > 1

) MultiFilmMemberRoles ON MultiFilmMemberRoles.FilmMemberId = FilmMember.Id

Results:



### List twenty popular videos by category for customers’ recommendations

Same as above

# CRUD Matrix

CRUD matrix as described in one presentation by Dr. Kung

## List of Entity Types

## List of Functions

# Concluding Remarks

Lessons learned and strengths and weaknesses, what you may add to the database project if you have more time.

More data fields.

1. AwardShowInstance could have a location column
2. Track Production companies
3. File Attachments
4. Future - Store Nominations
5. Add remaining award types
6. Add data from other systems
7. Movie Trailers

Appendices

Additional information, such as known defects

Appendix A - DDL, INSERT, SELECT Statements

CREATE statements for creating database objects; INSERT statements to populate test data into the database; SELECT statements to display the test data

Appendix B - Data Dictionary Index

Index to the data dictionary (e.g., column\_name in alphabetical order, table\_name))

References

Reference material