**Nutflux**

**Nicolas Di Costanzo 21209060**

nicolas.dicostanzo@ucdconnect.ie

Use a splash page image here [optional]

Use *LaTeX* if you wish, but use the general spacing and font/style you find here (1.5 spacing, 12 point font for text, etc.).

Be sure to submit a PDF (not a .DOC file) as your report. Overall it should be **30 pages or more**, including diagrams and screenshots. A significant portion of the report should be textual. Do not rely on images to write your report for you.

**Remember**, your project this year concerns a database for a streaming platform.

Your database is intended to support a data-driven, knowledge-based approach to content selection. As such, identify the place of the database in the overall platform, and tell us how you would support its operations at the SQL level.

**What to submit**: This report, as a PDF, *and* the necessary SQL files to allow us to examine your database constructs and test your queries.

Contents

[1. Introduction 1](#_Toc99974249)

[2. Database Plan: A Schematic View 2](#_Toc99974250)

[3. Database Structure: A Normalized View 2](#_Toc99974251)

[1NF 2](#_Toc99974252)

[*“Each column has to store only one value”* rule 2](#_Toc99974253)

[*“All the values in a column have to be of the same type”* rule 3](#_Toc99974254)

[“Each column’s name has to be unique.” Rule 3](#_Toc99974255)

[“Every column has to relate to the key” rule 1](#_Toc99974256)

[2NF 1](#_Toc99974257)

[3NF 2](#_Toc99974258)

[BCNF 2](#_Toc99974259)

[4. Database Views 2](#_Toc99974260)

[5. Procedural Elements 2](#_Toc99974261)

[6. Example Queries: Your Database In Action 2](#_Toc99974262)

[7. Conclusions 2](#_Toc99974263)

[Acknowledgements 2](#_Toc99974264)

[References 2](#_Toc99974265)

[Figure 1: content, works and person table relationship 3](#_Toc99974349)

[Figure 2: content table 1](#_Toc99974350)

[Figure 3: award table 1](#_Toc99974351)

[Figure 4: instance of the "content” table 1](#_Toc99974352)

[Figure 5: relationship between a user and a content 1](#_Toc99974353)

# 1. Introduction

This project will try to describe, at the data level, what a visual content database could be for different kind of target audience. There are two main audiences targeted by this application. “Movie nuts” (pro-users) and casual viewers (standard-users). The base must contain information that are relevant for both. All data relevant to standard-users will be relevant to pro-users but all data relevant to pro-users will not be necessarily relevant to standard-users. However, the database must contain both set of data. The difference will be managed in the way the information appears to users. And this information and the way it is displayed will depend on the status of the user (standard or pro user). The database design does not make a difference between data that is specific to the user, this will be managed on the application itself.

This database will be used for a public application (pro-users are also part of the general public) so it does not need to store very specific information that would only be useful for professionals, such as **??????.**

I would like to propose a database which, in addition to giving the classical and necessary information on a video content (date, name, actors etc...) would give information on the ethics of the people involved in its production. For example, if one of the actors has been accused of sexual or physical assault or embezzlement. How far these accusations have gone: whether there were only rumors, or whether there was a trial or conviction.

More and more people are interested in this aspect of industrial and cultural production. We can see this with the development of organic and fair-trade labels on food products for example. In the world of culture, the numerous controversies that take place each time a work by Roman Polanski is released are good examples of this tendency. The director was convicted in 1977 by the American justice system for the rape of a 13-year-old girl and is still considered a fugitive by the USA. As a result, many people boycott his productions and protest when he receives awards.

More recently, the Weinstein affair has had an international impact. It has brought to light practices of intimidation, abuse of power and sexual assault in the film industry. As a result, some moviegoers have decided to boycott Weinstein productions and even more broadly Hollywood productions, as Hollywood is seen as a central player in the trivialization of these practices.

Of course, ethics encompasses many different parameters such as the inclusion of people issue from minorities, people who are gender fluid etc... But in this work, I will not take these aspects into account, I will focus exclusively on rumors, accusations, and convictions of crimes as it is sufficient to showcase the idea's potential.

# 2. Database Plan: A Schematic View

# 3. Database Structure: A Normalized View

## 1NF

For a database to be in first normal form, it has to follow these four rules:

* Each column has to store only one value.
* All the values in a column have to be of the same type.
* Each column’s name has to be unique.
* Every column has to relate to the key.

Let’s look at our main tables to check how they respect these rules.

### *“Each column has to store only one value”* rule

The database has obviously a “content” table which registers the main information about a content. Its name and its released year for example. Another interesting piece of information about content is the identity and role of the people working on it. However, this is not a relevant information to store in a column. If I had to store this information in a column named “actors” for example, I would have to put every actor’s name in it and separate them with commas. But it would violate one of the 1NF rule since only one value has to be stored in a column. To solve this problem, I have a "*person*" and a "*works*" tables that link the content with every single person that worked on it. Each instance of the *person* table stores information about only one person, and each *works’* table instance stores information about the nature of the functions that the person had on this content.

A screenshot of a video game

Description automatically generated with medium confidence

Figure 1: content, works and person table relationship

### *“All the values in a column have to be of the same type”* rule

Let’s take the content table as an example again. The two most important information about a content is its name and its released year since it is thanks to these two properties that we manage to uniquely identify a content. But these two pieces of information have a different type since one if a number and the other one is one or several words. So, we cannot store them in the same column. That is why there is one column of type *varchar* to store the name and one of the type *integer* to store its released year.

### “Each column’s name has to be unique.” Rule

Some columns have the same meaning within a table. It is the case for the *year* property which appears in several different entities. To differentiate them, I simply add the table column after the property name such as follow:

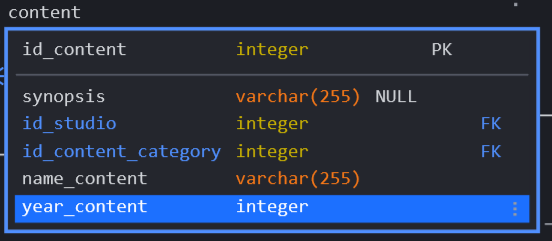


Figure 2: content table

A screenshot of a computer

Description automatically generated with medium confidence

Figure 3: award table

### “Every column has to relate to the key” rule



Figure 4: instance of the "content” table

As we can see in this instance of the *content* table there is a reference to the IDs of the studio and the content’s category. I could have directly stored these two pieces of information as *varchar* but then these columns would not have been connected to a key since several instances of the "content" table can share a same value for these properties. Instances sharing the same value for this property would not have been connected to each other in the database. This is the case if we reference the IDs of the instances of these values.

## 2NF

For a database to be in second normal form, it should not contain any partial dependency. Partial dependency happens when a non-prime attribute only depends on part of a prime key and not on the whole prime key. To remove Partial dependency, we can divide a table, remove the attribute, which is causing partial dependency, and move it to some other table where it fits in well.

A picture containing graphical user interface

Description automatically generated

Figure 5: relationship between a user and a content

The relationship between these tables gives an example on how I manage the second normal form. If I had recorded the note as a property (and not as a foreign key) of the *user* table, it would not have relevant since it is not a sufficient information to know what the rating is. Indeed, a rating is dependent on the user ID AND on the content ID. So, we need to store these two pieces of information in a dedicated table.

## 3NF

For a table to be in third normal form it needs to be in second normal form, and it should not have transitive dependency. Transitive dependency happens when an attribute depends on some non-prime attribute and not on the prime attribute. Let’s take again the example above. When I added the *note* property, I could have added it to the *content* table. But *note* does not dependent on the content ID solely. It also depends on the user ID. If I had done that we would have an instance of each content for each note given. Since *note* property is an attribute in the *rating* table, each content has only one instance in the *content* table and several ratings can be associated to each of these content instances.

## 

## BCNF

For a table to be in Boyce-Codd Normal Form (BCNF) it has to be in the third normal form and for any dependency *A*->*B*, *A* should be a super key. This database is in BCNF since no non-prime attribute

# 4. Database Views

# 5. Procedural Elements

# 6. Example Queries: Your Database In Action

# 7. Conclusions

# Acknowledgements

# References

<https://www.pastemagazine.com/movies/10-mob-movie-actors-with-actual-organized-crime-ti/>

<https://www.rosaleslawfirm.com/blog/2019/08/famous-white-collar-crime-cases-from-gangsters-to-actors-and-television-stars/>

<https://www.independent.co.uk/arts-entertainment/films/news/mark-wahlberg-racist-hate-crimes-wikipedia-history-george-floyd-blm-protests-a9554191.html>

<https://en.wikipedia.org/wiki/Sexual_abuse_in_the_American_film_industry>

<https://www.cbr.com/cannibal-holocaust-director-was-charged-with-murder/>

https://www.glamour.com/gallery/post-weinstein-these-are-the-powerful-men-facing-sexual-harassment-allegations