University of Victoria

Faculty of Engineering

CSC 370
Assignment 2

Submitted by Team: Kaizen Blitz

Ryan Woodward - V00857268
Brendan Ciccone - V00871008
Sarah Sparrow - V00819941

Table of Contents

1. Entity Relationship Diagram

Figure 1 shows the Entity Relationship Diagram for a database for Scenario 1.

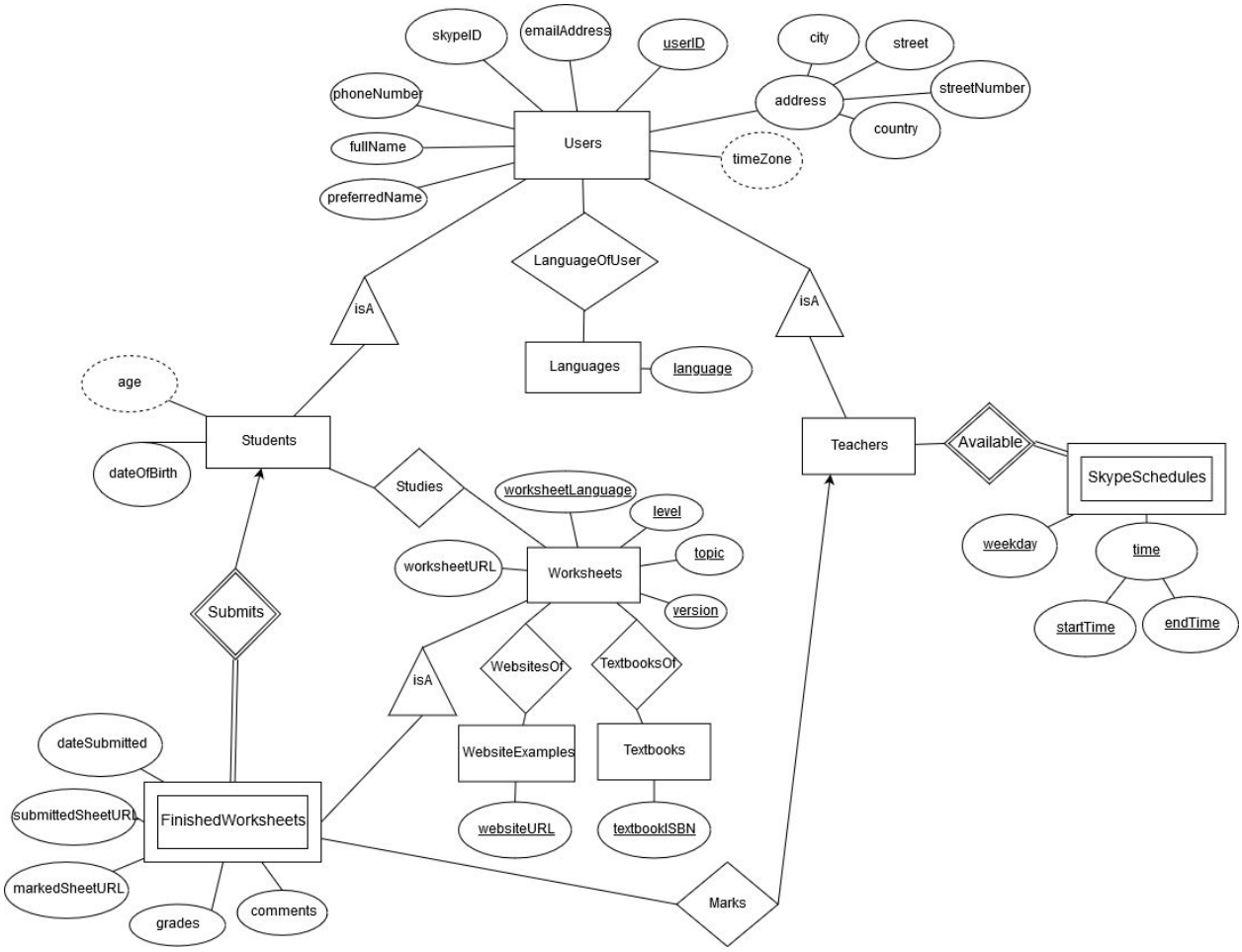

Figure 1: Entity Relationship Diagram for Database

2. Assumptions

In order to design the database, assumptions about the users, teachers, students and worksheets were made.

2.1 User Assumptions

The following assumptions were made about users:

- Users must supply a phone number, email address, and home address when registering for the website.
- Multiple users may live at the same address, share a phone number, and/or share an email address.

2.2 Teacher Assumptions

The following assumptions were made about teachers:

- Teachers are not restricted to what worksheets they can mark.
- Teachers may teach in multiple languages.
- Teachers are not required to have a Skype ID and converse with students over Skype.

2.3 Student Assumptions

The following assumptions were made about students:

- Students must supply their date of birth so that the database may be expanded to disallow users under the legal age where information can be collected on them.
- Students will submit one document/image per worksheet.
- Students may study in multiple languages.
- Students are not required to have a Skype ID.

2.4 Worksheet Assumptions

The following assumptions were made about worksheets:

- There are multiple versions of each worksheet.
- A worksheet's topic, level, version, and language are sufficient to uniquely identify it.
- Worksheets may be viewed by accessing them via their URL
- Teachers may provide feedback by writing on the student's worksheet (which will then be stored in markedSheetURL) and/or by adding text comments.
- Some worksheets require the student and teacher to be able to communicate in a specific language, and some worksheets may be independent of language (for instance, an arithmetic worksheet that uses no words).
- A worksheet is only marked by one Teacher
- A worksheet is only submitted by one Student

3. Relation Schema

The following is the relation schema for the ER diagram with primary keys underlined:

- Users(userID, phoneNumber, fullName, preferredName, skypeID, emailAddress, city, street, streetNumber, country)
- Students(userID, dateOfBirth)
- Teachers(userID)
- SkypeSchedules(userID, weekday, startTime, endTime)
- Marks(teacherUserID, level, topic, version, worksheetLanguage, studentUserID)
- Worksheets(level, topic, version, worksheetLanguage, worksheetURL)
- FinishedWorksheets(userID, level, topic, version, worksheetLanguage, dateSubmitted, submittedSheetURL, markedSheetURL, grades, comments)
- WebsitesOf(websiteURL, level, topic, version, worksheetLanguage)
- TextbooksOf(textbookISBN, level, topic, version, worksheetLanguage)
- WebsiteExamples(websiteURL)
- Textbooks(textbookISBN)
- Languages(language)
- LanguageOfUser(userID, language)

4. Domains and Constraints

The data type and allowable values of each attribute are shown in Table 1.

Table 1: Data type and Allowable Values of Attributes

Attribute	**Representation**	**Allowable-Values**	**NULL Allowed?**
userID, studentUserID,teacherUserID	INT(10)	[0, 1000000000]	No
phoneNumber	INT(15)	[0, 999999999999999]	No
fullName	VAR CHAR(255)	ALL	No
preferredName	VAR CHAR(255)	ALL	Yes
skypeID	VAR CHAR(255)	ALL	Yes
emailAddress	VAR CHAR(255)	ALL	No
city	VAR CHAR(255)	ALL	No
street	VAR CHAR(255)	ALL	No
streetNumber	INT(10)	[0, 1000000000]	No
country	VAR CHAR(255)	ALL	No
dateOfBirth	DATE	ALL	No
level	INT(2)	[0, 99]	No
topic	VAR CHAR(255)	ALL	No
version	INT(4)	[0, 1000]	No
worksheetURL	VAR CHAR(2047)	ALL	No
dateSubmitted	DATE	ALL	No
submittedSheetURL	VAR CHAR(2047)	ALL	No
markedSheetURL	VAR CHAR(2047)	ALL	Yes
grades	FLOAT	[0, 100.0]	Yes
comments	VAR CHAR(2047)	ALL	Yes
language	VAR CHAR(255)	ALL	No
worksheetLanguage	VAR CHAR(255)	ALL	No
websiteURL	VAR CHAR(255)	ALL	No
textbookISBN	INT(13)	[0, 9999999999999]	No
weekday	CHAR(2)	SA,SU,MO,TU,WE,TH,FR	No
startTime	TIME	ALL	No
endTime	TIME	ALL	No

The attribute worksheetLanguage may store the value “All Languages” for worksheets that are language independent. The attribute language may not store the value “All Languages”. Timezone is derived using the user’s address and are stored in UTC. The age attribute is derived from the Student’s dateOfBirth

5. Foreign Keys

The foreign keys used in the database are (userID), (teacherUserID), (studentUserID), (level, topic, version, worksheetLanguage), (websiteURL), (textbookISBN), and (language).

5.1 userID

The userID attribute in each of Students, Teachers, FinishedWorksheets, LanguageOfUser, and SkypeSchedules, and the attributes teacherID and studentUserID in the Marks relation, are foreign keys referencing the primary key (userID) of the Users relation. This can be modeled using relational algebra:

$$\pi_{userID}\ (Students) \subseteq \pi_{userID}\ (Users)$$
$$\pi_{userID}\ (Teachers) \subseteq \pi_{userID}\ (Users)$$
$$\pi_{userID}\ (FinishedWorksheets) \subseteq \pi_{userID}\ (Users)$$
$$\pi_{userID}\ (LanguageOfUser) \subseteq \pi_{userID}\ (Users)$$
$$\pi_{userID}\ (SkypeSchedules) \subseteq \pi_{userID}\ (Users)$$
$$\pi_{teacherUserID}\ (Marks) \subseteq \pi_{userID}\ (Users)$$
$$\pi_{studentUserID}\ (Marks) \subseteq \pi_{userID}\ (Users)$$

The userID attribute will be RESTRICTED in the foreign key entities. If the userID was to be modified it could only be modified in the users entity, at which point it will UPDATE and CASCADE throughout all of the foreign keys.

5.2 Referential Integrity Constraints for level, topic, version, worksheetLanguage

The attributes level, topic, version and worksheetLanguage in relations WebsitesOf, TextbooksOf, FinishedWorksheets, and Marks are a foreign key referencing the primary key of the Worksheets relation. This can be modeled using relational algebra:

$$\pi_{level,topic,version,worksheetLanguage}\ (WebsitesOf) \subseteq \pi_{level,topic,version,worksheetLanguage}\ (Worksheets)$$
$$\pi_{level,topic,version,worksheetLanguage}\ (TextbooksOf) \subseteq \pi_{level,topic,version,worksheetLanguage}\ (Worksheets)$$
$$\pi_{level,topic,version,worksheetLanguage}\ (FinishedWorksheets) \subseteq \pi_{level,topic,version,worksheetLanguage}\ (Worksheets)$$
$$\pi_{level,topic,version,worksheetLanguage}\ (Marks) \subseteq \pi_{level,topic,version,worksheetLanguage}\ (Worksheets)$$

The primary key (level, topic, version, worksheetLanguage) is RESTRICTED from being modified in the Worksheets relation.

5.3 Referential Integrity Constraints for websiteURL

The websiteURL attribute is a foreign key in the WebsitesOf relation that references the primary key of the Websites relation. This can be modeled using relational algebra:

$$\pi_{websiteURL}\ (WebsitesOf) \subseteq \pi_{websiteURL}\ (Websites)$$

If the primary key, websiteURL, is modified in the Websites relation, then those changes will CASCADE to the foreign key, websiteURL, in WebsitesOf.

5.4 Referential Integrity Constraints for textbookISBN

The textbookISBN attribute is a foreign key in the TextbooksOf relation that references the primary key of the Textbook relation. This can be modeled using relational algebra:

$$\pi_{textbookISBN}\ (TextbooksOf) \subseteq \pi_{textbookISBN}\ (Textbooks)$$

If the primary key, textbookISBN, is modified in the Textbook relation, then those changes will CASCADE to the foreign key, textbookISBN, in TextbooksOf.

5.5 Referential Integrity Constraints for language

The language attribute is a foreign key in the LanguageOfUser relation that references the primary key of the Languages relation. This can be modeled using relational algebra:

$$\pi_{language}\ (LanguageOfUser) \subseteq \pi_{language}\ (Languages)$$

If the primary key, language, is modified in the Languages relation, then those changes will CASCADE to the foreign key, language, in the LanguageOfUser relation.

6. Sample Queries

Database tables were created to match the relation schema in order to test some queries. Queries were executed to find the average grade of a worksheet, to determine which teachers are available to Skype at a given time, and to find the most popular languages of all the students.

6.1 Find Average Grade of Worksheet

A database user may wish to see the average grade of a worksheet in order to determine its difficulty.

6.1.2 SQL Query

The following query can be used to determine the average grade for worksheet Algebra 12, version 1, in English:

```
SELECT AVG(grades)
FROM FinishedWorksheets
WHERE level=12 AND topic='Algebra' AND version=1;
```

6.1.3 Sample Database Data

The following sample data was input into the database (see figure 2).

```
kaizen_blitz=> SELECT * FROM FinishedWorksheets;
 userid | level |  topic  | version | worksheetlanguage | datesubmitted | submittedsheeturl |  markedsheeturl   | grades | comments
--------+-------+---------+---------+-------------------+---------------+-------------------+-------------------+--------+----------
 857268 |    12 | Algebra |       1 | English           | 2019-10-25    | www.subUrl.com    | www.markedURL.com |     93 | good job
      1 |    12 | Algebra |       1 | English           | 2019-10-25    | www.subUrl.com    | www.markedURL.com |     43 | bad job
(2 rows)
```

Figure 2: Shows columns and values in FinishedWorksheets table

Executing the sample query from section 6.1.2 will produce the following output (see figure 3).

```
kaizen_blitz=> SELECT AVG(Grades) FROM FinishedWorksheets WHERE level=12 AND topic = 'Algebra' AND version = 1;
 avg
-----
  68
(1 row)
```

Figure 3: Shows the average grade for worksheet Algebra 12 level 1 in English

6.2 Teacher Skype Availability

A database user may need to determine which teachers are available at a specific time for Skype calls.

6.2.1 SQL Query

The following query can be used to list the full name and skype ID of all teachers who are available to Skype at 1PM for at least 30 minutes on a Wednesday:

```
SELECT fullName, skypeID
FROM Users
WHERE userID IN ( SELECT userID
                  FROM SkypeSchedules
                  WHERE weekday='WE'
                        AND startTime <= 1:00PM
                        AND endTime >= 1:30PM )
```

6.2.2 Sample Database Data

The following sample data was input into the Users table in the database (see figure 4).

```
kaizen_blitz=> select * from users
kaizen_blitz-> ;
 userid | phonenumber |    fullname    | preferredname |     skypeid     |     emailaddress      |     city     |    street    | streetnumber |   country
--------+-------------+----------------+---------------+-----------------+-----------------------+--------------+--------------+--------------+-------------
 857268 |  7786788059 | RyanWoodward   | Ryan          | rwoodward84     | rwoodward84@gmail.com | Victoria     | Meares       | 1165         | Canada
      1 |           2 | Fun Guy        | Fun           | FunGuy          | funguy@gmail.com      | Victoria     | Fort         | 256          | Canada
    370 |   298341234 | TeacherMcTeach | Ted           | TedTheTeacher   | callmeted@gmail.com   | Teacherville | 1234 st      | 55           | teacherLand
    477 |       34347 | Fred Friendly  | Fred          | Freddy          | callmefred@gmail.com  | Fredland     | 2345 st      | 66           | Rainville
   5467 |     9581234 | Dr. Phil       | PhilMacGraw   | Phil The Thrill | drphil@gmail.com      | DayTime TV   | Some Channel | 2            | TV
(5 rows)
```

Figure 4: Shows columns and values in Users table

The following data was input into the SkypeSchedules table in the PostgreSQL database (see figure 5).

```
kaizen_blitz=> SELECT * FROM skypeSchedules;
 userid | weekday | starttime | endtime
--------+---------+-----------+----------
    370 | Tu      | 08:00:00  | 12:30:00
   5467 | We      | 10:00:00  | 13:15:00
    477 | We      | 12:00:00  | 14:30:00
(3 rows)
```

Figure 5: Shows columns and values in SkypeSchedules table

Executing the sample query from section 6.2.1 will produce the following output (see figure 6).

```
kaizen_blitz=> SELECT fullName, skypeID FROM Users WHERE UserID IN (SELECT userID FROM SkypeSchedules WHERE weekday = 'We' AND startTime <= '13:00:00' AND endTime >= '13:30:00');
    fullname    | skypeid
----------------+---------
 Fred Friendly | Freddy
(1 row)
```

Figure 6: Shows all teachers available to Skype at 1PM for at least 30 minutes on a Wednesday

6.3 Most Popular Student Languages

A database user may wish to see the most popular languages of students so that they may create more worksheets in those languages.

6.3.3 SQL Query

The following query may be used to list the top 3 most popular languages that students know:

```
SELECT language, COUNT(language) AS occurrences
FROM LanguageOfUser
WHERE userID IN (SELECT userID
                 FROM Students)
GROUP BY language
ORDER BY occurrences DESC, language
LIMIT 3
```

6.3.2 Sample Database Data

The following sample data was input into the Users table in the database (see figure 7).

```
kaizen_blitz=> select * from users
kaizen_blitz-> ;
 userid | phonenumber |    fullname     | preferredname |     skypeid     |     emailaddress      |     city     |    street    | streetnumber |   country
--------+-------------+-----------------+---------------+-----------------+-----------------------+--------------+--------------+--------------+-------------
 857268 |  7786788059 | RyanWoodward    | Ryan          | rwoodward84     | rwoodward84@gmail.com | Victoria     | Meares       | 1165         | Canada
      1 |           2 | Fun Guy         | Fun           | FunGuy          | funguy@gmail.com      | Victoria     | Fort         | 256          | Canada
    370 |   298341234 | TeacherMcTeach  | Ted           | TedTheTeacher   | callmeted@gmail.com   | Teacherville | 1234 st      | 55           | teacherLand
    477 |       34347 | Fred Friendly   | Fred          | Freddy          | callmefred@gmail.com  | Fredland     | 2345 st      | 66           | Rainville
   5467 |     9581234 | Dr. Phil        | PhilMacGraw   | Phil The Thrill | drphil@gmail.com      | DayTime TV   | Some Channel | 2            | TV
 815546 |  1112223333 | Sam Gavrilko    | Sam           | SamGav          | samgav@email.com      | Victoria     | Meares       | 1165         | Canada
     10 |  2223334444 | Sarah Sparrow   | Sarah         | Sarah           | sarspa@emailing.com   | Victoria     | Steetname    | 234          | Canada
     13 |   123456789 | Johnny Gaudreau | Johnny        | Johnny Hockey   | jhock@asdf.com        | Calgary      | Flames       | 1            | Canada
     33 |   234509854 | Justin Bieber   | Big Biebs     | Biebs McGeebs   | jbabs@bigbiebs.com    | LA           | Hollywood    | 2            | USA
     25 |   234509854 | Brendan Ciccone | The Cone      | Brendan         | brenemail@email.com   | Victoria     | 3456         | 4            | Canada
(10 rows)
```

Figure 7: Shows columns and values in Users table

The Language table was filled with the following data (see figure 8).

```
kaizen_blitz=> select * from languages
kaizen_blitz-> ;
 language
----------
 English
 French
 Spanish
 Gibberish
(4 rows)
```

Figure 8: Shows columns and values in Languages table

The LanguageOfUser table was filled with the following data (see figure 9).

```
kaizen_blitz=> select * from languageofUser
kaizen_blitz-> ;
 userid | language
--------+----------
 857268 | English
      1 | French
    370 | Gibberish
    477 | Spanish
   5467 | Gibberish
 815546 | English
     10 | English
     13 | French
     33 | Gibberish
     25 | English
(10 rows)
```

Figure 9: Shows columns and values in LanguagesOfUsers table

The Students table is shown in figure 10.

```
kaizen_blitz=> select * from Students
kaizen_blitz-> ;
 userid | dateofbirth
--------+-------------
 857268 | 1984-09-17
      1 | 1234-12-01
 815546 | 1992-08-06
     10 | 1993-10-23
     13 | 1996-02-14
     33 | 1982-02-28
     25 | 1998-05-15
(7 rows)
```

Figure 10: Shows columns and values in Students table

Executing the sample query from section 6.3.1 will produce the following output (see figure 11).

Figure 11: Shows the top 3 most popular languages that students know

7. Assignment Journal

7.1 ER Diagram Modifications

The following ER diagram (figure 12) is the version used in Assignment 1 for Scenario 1.

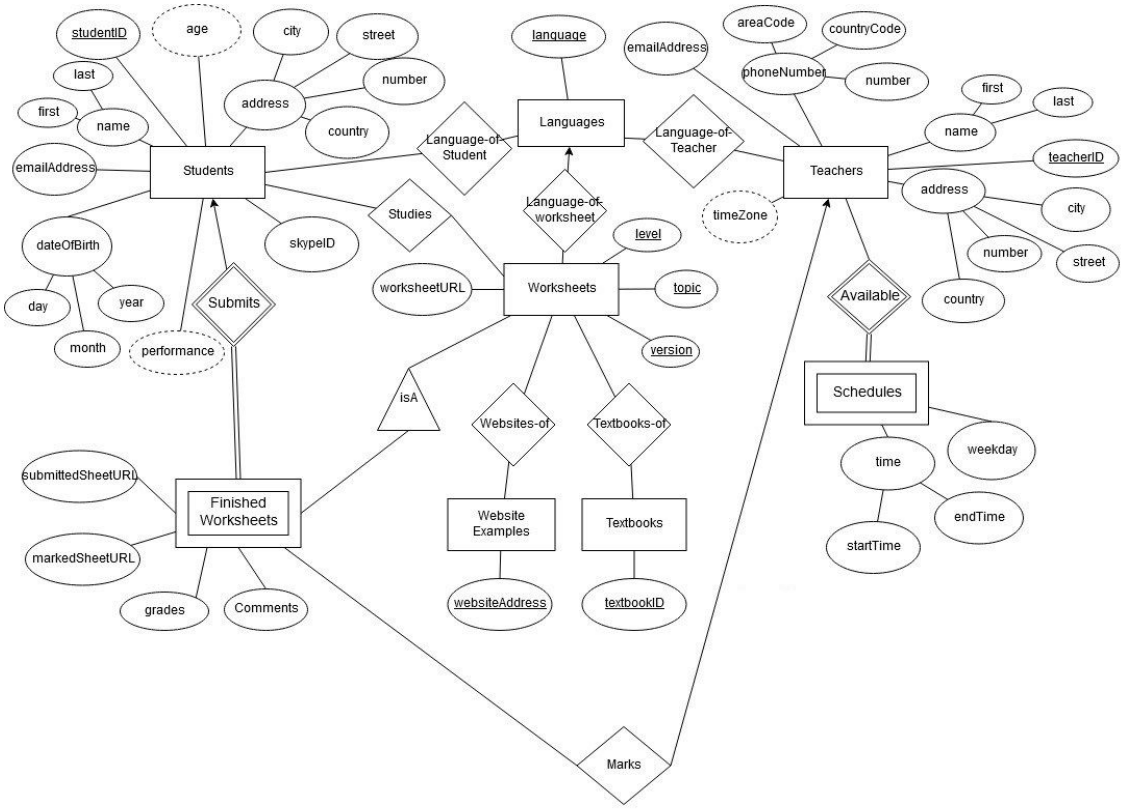

Figure 12: Shows the original entity relationship diagram

The major changes that took place were:

- The User entity was added to reduce the redundancy of keeping attributes, such as name and address, in separate tables for Teachers and Students. The new Entity Relationship Model has the User entity in an isa hierarchy with Students and Teachers.
- The Language-of-worksheet relation was replaced with a worksheetLanguage attribute for each worksheet, because the original relation was not sufficient to store worksheets in multiple languages.
- Full name and phoneNumber were changed from composite attributes to atomic ones, because we realized that setting their type as VAR CHAR would handle the various differences that different countries use.
- The attribute dateOfBirth was changed from a composite attribute to an atomic one, because we discovered that the DATETIME datatype would fit the needs of that attribute.

- The derived attribute, performance, was removed, because we determined that we could query the table to get the same data.

7.2 Group Meetings

7.2.1 Meeting Notes for October 16, 2019

During the interview for assignment 1, Nirmala suggested some changes to the structure of the ER diagram. Brendan and Sarah met after the interview and added the User relation into the ER diagram as an isa hierarchy with Students and Teachers, as well as removed the derived attribute performance.

7.2.2 Meeting Notes for October 24, 2019

After independently learning about the concepts needed for this assignment, we came together on October 24 to complete most of the document.

Ryan's work:
- Created PostgreSQL database
- Added foreign key referential integrity constraints
- Contributed to final relation schema

Brendan's work:
- Created the 'Data type and Allowable Values of Attributes' table
- Determined datatype and constraints for each attribute
- Updated ER diagram to match our changes
- Contributed to final relation schema

Sarah's work:
- Created document that showed the step-by-step process for converting our ER diagram into a relation schema
- Formatted report
- Created assumptions section
- Created three sample queries for the database
- Added foreign key referential integrity constraints
- Contributed to final relation schema

7.2.3 Meeting Notes for October 25, 2019

To finish incomplete report sections, we met up over Slack to complete the screenshots for the sample database queries, the foreign key referential integrity section, the meeting notes section, the assumptions section, and to proofread the document.

Ryan's work:

- Finished creating tables in PostgreSQL database
- Inputted sample data into the PostgreSQL database
- Executed the sample queries on the database and took screenshots of their results
- Discussed and finalized the foreign key referential integrity

Brendan's work:

- Researched issues regarding how we defined the foreign key referential integrity
- Proofread document for errors
- Double checked the relation schemas as compared to the ER diagram
- Discussed and finalized the foreign key referential integrity

Sarah's work:

- Formated the meeting notes
- Fix minor errors in the ER diagram
- Update report formatting
- Discussed and finalized the foreign key referential integrity