Milestone 2:

Revised ER Diagram:

https://drive.google.com/file/d/1trYgUNpfxupiylCzts_EOR909nKYIzDe/view?usp=sharing

Changes made to ER Diagram:

- Changed "is_child_of" relation to be many to one. Any repository object may only have one parent.
- Changed "made_to" relation to be one-to-many instead of many-to-many since each line change may only be made to a single file.
- Repository_object can only be contained by a single repository.
- Added date entity
- added file_name attribute to repository object. This allows for some redundancy.
- Added commit number as attribute of commit since each commit in a repo has an index for the orderings of the commits
- Renamed Issue number to Issue ID so that it is clear this is a database wide ID and doesn't belong to a single repository
- Added Location and FIPPA_Compliance to Organization. Data for companies in Canada must follow FIPPA regulation and be stored in Canada, while companies in other countries need not follow this restriction.

Schema for ER Diagrams:

Here, <u>foregin keys are underlined</u> and **primary keys are in bold.**

Entities:

- *User*(**Username: String**, Password_hash: String)
    - Functional Dependencies:
        - Username → Password_hash
    - Candidate Keys (not primary key): None
    - Constraints:
        - Assertion: Organization NOT NULL
- *Organization*(**Organization_name: String,** Country: String, FIPPA_Compliance: Integer)
    - Functional Dependencies:
        - Organization_name → Country, FIPPA_Compliance
        - Country → FIPPA_Compliance
    - Candidate Keys (not primary key): None
    - Constraints: FIPPA_Compliance is one of 0 or 1.
- *Repository*(**Repository_name: String**, <u>Organization_name: String</u>, <u>Year_created: Integer, Month_created: Integer, Day_created: Integer, Hour_created: Integer, Minute_created: Integer, Second_created: Integer</u>)
    - Functional Dependencies: Repository_name → Organization_name, Year_created, Month_created, Day_created, Hour_created, Minute_created, Second_created
    - Candidate Keys (not primary key): None
    - Constraints:
        - Organization name is NOT NULL
        - Year, Month, Day, Hour, Minute, Second are not null.
- *Commit*(**Commit_SHA: String**, Commit_message: String, Commit_number: Integer, <u>Repository_name: String</u>, <u>File_changed: String</u>, <u>Author: String</u>, <u>Year_created:</u> Integer, <u>Month_created</u>: Integer, <u>Day_created</u>:Integer, <u>Hour_created</u>: Integer, <u>Minute_created</u>: Integer, <u>Second_created</u>: Integer)
    - Functional Dependencies:
        - Commit_SHA → Commit_message, Commit_number, Repository_name, Author, Year_created, Month_created, Day_created, Hour_created, Minute_created, Second_created, File_changed
        - Commit_number and Repository_name → Commit_SHA, Commit_message, Author, Year_created, Month_created, Day_created, Hour_created, Minute_created, Second_created, File_changed
    - Candidate Keys (not primary key):
        - None
    - Constraints:
        - Author, Repository name, File_changed are NOT NULL
        - Year, Month, Day, Hour, Minute, Second are NOT NULL.
        - Commit_number >= 0
- *Line_change*(**<u>Commit_SHA: String</u>, Line_number: Integer**)
    - Functional Dependencies: None

- Candidate Keys (not primary key): None
- Constraints:
  - Line_number: >= 1
- *Issue*(**Issue_ID: Integer**, Issue_message: String, <u>Owning_board: String</u>)
  - Functional Dependencies: Issue_ID → Issue_message, Owning_board
    - Issue_ID -> Issue_message, Owning_board
  - Candidate Keys (not primary key): None
  - Constraints:
    - Owning_board is NOT NULL
- *Repository_Object*(<u>**Repository: String**</u>, **Path: String**, file_extension: String, file_contents: String, Filename: String, Size: Integer, <u>Parent_Directory_Path: String</u>)
  - Functional Dependencies:
    - Path → Filename, Parent_Directory_Path
  - Candidate Keys (not primary key): None
  - Constraints:
    - Default value for parent_directory: "/" (ie root)
    - Size: >= 0
- *Project_board*(**Board_title: String**, <u>Repository_name: String</u>)
  - Functional Dependencies: Board_title → Repository_name
  - Candidate Keys(not primary key): None
  - Constraints:
    - Repository_name is NOT NULL
- *Date*(**Year: Integer, Month: Integer, Day: Integer, Hour: Integer, Minute: Integer, Second: Integer**)
  - Functional Dependencies: None
  - Candidate Keys(not primary key): None
  - Constraints:
    - 31 >= day >= 1, 12 >= month >= 1, year >= 1, 24 >= hour >= 1, 60 >= minute >= 1, 60 >= second >= 1

Relations:
- *Is_Member_of*(<u>**Username: String, Organization_name: String**</u>)
- *Contributor_of*(<u>**Username: String, Repository_name: String**</u>)
- Assigned_to(<u>**Username: String, Issue_ID: Integer**</u>)

Since these relations are all tables of 2 attributes, they are inherently normalized to be in BCNF

Other constraints:
- An assertion that each commit has at least one line change

Normalizations

Repository_Object:

- *Repository_Object*(**Repository: String**, **Path: String**, file_extension: String, file_contents: String, Filename: String, Size: Integer, <u>Parent_Directory_Path: String</u>)
    - Functional Dependencies:
        - Path → Filename
        - Path → Parent_Directory_Path
        - Path → file_extension

Since {Path} is not a key of Repository_Object, this table is not in BCNF and can be normalized

For each FD X → b, new tables are (*Repository_Object - b), (X U b)*
   Repository_Object(**Repository: String**, **Path: String**, file_extension: String, file_contents: String, Filename: String, Size: Integer, <u>Parent_Directory_Path: String)</u>

   Resulting new tables:
   Repository_Object(**Repository: String**, **Path: String**, file_contents: String, Size: Integer)
    - Functional Dependencies: None
    - Candidate Keys(not primary key): None
    - Constraints: None

   RO1(**Path: String**, Filename: String)
    - Functional Dependencies: Path→ Filename
    - Candidate Keys(not primary key): None
    - Constraints: None
   RO2(**Path: String**, Parent_Directory_Path: String)
    - Functional Dependencies: Path→ Parent_Directory_Path
    - Candidate Keys(not primary key): None
    - Constraints: None

   RO3(**Path: String**, file_extension: String)
    - Functional Dependencies: Path→ file_extension
    - Candidate Keys(not primary key): None
    - Constraints: None

Organization

- *Organization*(**Organization_name: String,** Country: String, FIPPA_Compliance: Integer)
    - Functional Dependencies:
        - Organization_name → Country, FIPPA_Compliance
        - Country → FIPPA_Compliance

- Candidate Keys(not primary key): None
- Constraints:
  - FIPPA_Compliance can only take values 0 or 1

Since {Country} is not a key of Repository_Object, this table is not in BCNF and can be normalized

New tables = (Organization - {FIPPA_Compliance}) and
({Country} U {FIPPA_Compliance})

Resulting new tables:
Organization(**Organization_name: String,** <u>Country: String</u>)
- Functional Dependencies:
  - Organization_name → Country
- Candidate Keys(not primary key): None
- Constraints: None

FIPPA_Details(**Country: String**, FIPPA_Compliance:Integer)
- Functional Dependencies:
  - Country → FIPPA_Compliance
- Candidate Keys(not primary key): None
- Constraints:
  - FIPPA_Compliance can only take values 0 or 1

Commit
- *Commit*(**Commit_SHA: String**, Commit_message: String, Commit_number: Integer, <u>Repository_name: String</u>, <u>File_changed: String,</u> <u>Author: Username</u>, <u>Year: Integer, Month: Integer, Day:Integer, Hour: Integer, Minute: Integer, Second: Integer</u>)
  - Functional Dependencies:
    - Commit_SHA → Commit_message, Commit_number, Repository_name, Author, Year, Month, Day, Hour, Minute, Second, File_changed
    - Commit_number, Repository_name → Commit_SHA, Commit_message, Author, Year_created, Month_created, Day_created, Hour_created, Minute_created, Second_created, File_changed

Commit also has more than one FD, however since Commit_SHA and Commit_number, Repository_name both determine all other attributes, and are therefore both superkeys, neither violate 3NF, and therefore no normalization is needed.

SQL DDL Statements

CREATE TABLE User (
        Username CHAR (20),
        Password_hash CHAR (20)
        PRIMARY KEY (Username))

CREATE TABLE Organization (
        Organization_name CHAR (20),
        Country CHAR(60),
        PRIMARY KEY (Organization_name),
        FOREIGN KEY (Country) REFERENCES FIPPA_Details)

CREATE TABLE FIPPA_Details (
        Country CHAR (60),
        FIPPA_Compliance INTEGER,
        PRIMARY KEY (country))

Could not model Constraints: FIPPA_Compliance is one of 0 or 1 (need assertions)

CREATE TABLE Repository (
        Repository_name CHAR (40),
        Organization_name CHAR (20),
        Year_created INTEGER,
        Month_created INTEGER,
        Day_created INTEGER,
        Hour_created INTEGER,
        Minute_created INTEGER,
        Second_created INTEGER
        PRIMARY KEY (Repository_name)
        FOREIGN KEY (Organization_name)
                REFERENCES Organization
        FORGEIGN KEY (Year_created, Month_created, Day_created,Hour_created,
        Minute_created, Second_created)
                REFERENCES Date)
        FOREIGN KEY

CREATE TABLE Commit (
        Commit_SHA CHAR (64),
        Commit_number INTEGER,
        Repository_name CHAR (20) NOT NULL,
        File_changed CHAR (20) NOT NULL,
        Author CHAR (20) NOT NULL,

```
        Year_created INTEGER NOT NULL,
        Month_created INTEGER NOT NULL,
        Day_created INTEGER NOT NULL,
        Hour_created INTEGER NOT NULL,
        Minute_created INTEGER NOT NULL,
        Second_created INTEGER NOT NULL,
        PRIMARY KEY (Commit_SHA),
        FOREIGN KEY (Year_created, Month_created, Day_created, Hour_created,
        Minute_created, Second_created)
                REFERENCES Date
        FOREIGN KEY (Repository_name)
                REFERENCES Repository,
        FOREIGN KEY (File_changed)
                REFERENCES Repository_Object
        FOREIGN KEY (Author)
                REFERENCES User)
```

Could not model constraints: Commit_number >= 0 (need assertions)

```
CREATE TABLE Line_change (
        Commit_SHA CHAR (64),
        Line_number INTEGER
        PRIMARY KEY (Commit_SHA,Line_number))
        FOREIGN KEY (Commit_SHA) REFERENCES Commit)
```

Could not model constraints: Line_number: >= 1

```
CREATE TABLE Issue (
        Issue_ID INTEGER,
        Issue_message CHAR(200),
        Owning_board CHAR(20) NOT NULL,
        PRIMARY KEY (Issue_ID)
        FOREIGN KEY (Owning_board) REFERENCES Project_board)
```


```
CREATE TABLE Repository_Object (
        Repository CHAR(40),
        Path CHAR(40),
        file_extension CHAR(5),
        file_contents CHAR(2000),
        Size INTEGER,
        Parent_Directory_Path CHAR(40) DEFAULT '/',
        PRIMARY KEY (Repository, Path),
        FOREIGN KEY (Repository) REFERENCES Repository,
```

FOREIGN KEY (Parent_Directory_Path) REFERENCES Repository_Object)

Could not model constraint : Size: >= 0 (need assertions)

CREATE TABLE Project_board (
        Board_title CHAR(20),
        Repository_name CHAR(40) NOT NULL,
        PRIMARY KEY (Board_title),
        FOREIGN KEY (Repository_name) REFERENCES Repository)

CREATE TABLE Date (
        Year INTEGER,
        MonthINTEGER,
        Day INTEGER,
        Hour INTEGER,
        Minute INTEGER,
        Second INTEGER
        PRIMARY KEY (Year, Month, Day, Hour, Minute, Second))

Could not model constraints:
        - 31 >= day >= 1, 12 >= month >= 1, year >= 1, 24 >= hour >= 1, 60 >= minute >= 1, 60 >= second >= 1 (assertions needed)

CREATE TABLE Is_Member_of(
        Username CHAR(20),
        Organization_name CHAR(20) NOT NULL,
        PRIMARY KEY (Username, Organization_name)
        FOREIGN KEY (Username) REFERENCES User,
        FOREIGN KEY (Organization_name) REFERENCES Organization)

CREATE TABLE Contributor_of (
        Username CHAR(20),
        Repository_name CHAR(40),
        PRIMARY KEY (Username, Repository_name)
        FOREIGN KEY (Username) REFERENCES User,
        FOREIGN KEY (Repository_name) REFERENCES Repository)

CREATE TABLE Assigned_to (
        Username CHAR(20),
        Issue_ID INTEGER.
        PRIMARY KEY (Username, Issue_ID)
        FOREIGN KEY (Username) REFERENCES User,
        FOREIGN KEY (Issue_ID) REFERENCES Issue)

Examples:

https://docs.google.com/spreadsheets/d/1utVoxrDpu_2lL2zrw7SCRB2HIc_l3cK0GA9wkEAf2yE/edit?usp=sharing