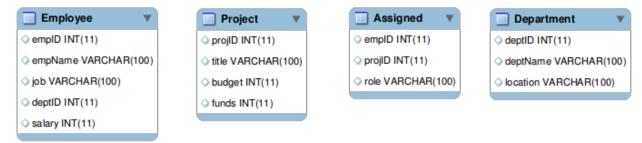# ECE 356 Winter 2020: Assignment 2

Due: 7th February at 5:00 PM

Goals: The goals of this lab are to
(a)   Recognize when a database is not in BCNF
(b)   Decompose a database into BCNF, creating the necessary tables
(c)   Design a BCNF database when just given data
(d)   Analyze the effects of PK/FK/other index on some query for the Baseball DB

Part 1: Consider the following schema:

The constraints on the attributes and functional dependencies between attributes are as follows:
(1) empID is unique per employee
(2) projID is unique per project
(3) The empName is the full name of each employee, and consists of a First and Last Name, and optionally a Middle Name.
(4) The deptID is unique per departments
(5) A department location is a full address, comprising the street number, street name, city name, province, and postal code
(6) An employee may have more than one role on a project
(7) An employee may be assigned to more than one project
(8) A department may have multiple locations
(9) An employee can be in more than one department

SQL that creates a database with these tables and data already exists and will be on Learn.

You are required to take this initial database schema and:
(1) decompose it, per BCNF
(2) determine primary keys
(3) determine foreign keys
(4) write the necessary SQL to create tables for this database, together with the necessary primary and foreign keys
(5) if you decomposition has resulted in the loss of any of the above four tables, write the necessary SQL to create a view that correspond to that table (note: in the case of the empName and department location you should use "concat" to create a single attribute from the atomic components)
(6) write the necessary SQL to migrate the data from these four tables into your schema

Part 1 Submission: For any portions of your solution to this that is SQL, write the SQL in a single file titled `employee.sql` and submit that file to the Assignment 2 Dropbox on `Learn`. Any written component and/or explanation for you schema should be written in a single file titled `employee.pdf` and likewise be submitted to the Assignment 2 Dropbox on `Learn`.

Part 2: The website "kaggle.com" contains roughly 27,000 user-provided datasets, which are an invaluable resource for those wishing to learn about databases with real data sets. Using the "https://www.kaggle.com/datasnaek/chess" dataset, create a BCNF chess database and take the data from the CSV and put it in our BCNF Chess database.

Part 2 Submission: SQL code to create and load the database should be in the file chess.sql, while any explanation you wish to add to justify your design should be in chess.pdf, and both files should be submitted to the DropBox on Learn.

Part 3: In Assignment 1 you had to compute several queries on the Sean Lahman baseball database. There were no explicit indexes on that database, though you should have added primary and foreign keys. Using experiments and/or the **explain** command determine the significance of

       (a) the primary keys on the tables in the query

       (b) the foreign keys on the tables in the query

       (c) whether or not any additional explicit indexes would help in solving those queries.

for the following queries:

(a)
```
select count(playerID) from Master where birthYear is null or birthYear = ""
                                       or birthMonth is null or birthMonth = ""
                                       or birthDay is null or birthDay = "";
```

(b)
```
select playerID,sum(salary) as totalPay from Salaries
    left outer join Appearances using (playerID,yearID,teamID)
    left outer join Managers using (playerID,yearID,teamID)
        where G is null and
              G_all is null
    group by playerID
    order by totalPay desc
    limit 3;
```

Be precise in stating what helps; for example, do not simply state that the PK on Master is helpful; specify what is that PK which helps; if an index would help, specify what the index would be.

Part 3 Submission: You will need to write some amount of SQL to answer this question, as well as writing the necessary explanations. As such, you should submit two files, baseball.sql and baseball.pdf, to the Assignment 2 Dropbox on `Learn`.