**(CPS)**

**Chicago Public Schools**

**2013-2014 Academic Year**

**Team members:**

Ahmed Hassan M. H. Ibrahim    - 160709057

Hasibullah Mahmood                  - 160709073

**Brief description:**

CPS schools for the 2013-2014 academic year database includes various identifiers used to identify school districts, including names; local, state, and federal IDs; and geographic descriptions on the location of each school.

**Changes after phase 1:**

1. The entities Attending Grades and Grade Structure were combined after splitting them.
2. The entities School Type and S type were combined after splitting them.
3. The entities Districts and Location were combined with Address.
4. The entities Community and Geographic area linked to Schools entity through Address entity.
5. The entities Class and Schools were combined.
6. The type of Class attribute inside the Schools entity was changed from tiny integer to Boolean.
7. In charter type entity the entries 0 replaced by null.

**Description about loading the data:**

Dataset were downloaded from Kaggle website. Then, it was cleaned by coding some python scripts. After learning the data, the ER diagram was created. Then, CSV file was created for each entity. Data was inserted by using MySQL workbench (import records from external file) in the inserts choice for each entity and then forward engineering was used to take the script and the empty entries replaced by null and then the script was executed to create the database.

**View:**

There is a view by the name of schools\_per\_geographic\_area which has two attributes 1. Geographic area name 2. Number of schools per geographic area

**Stored Procedure:**

1. The first stored procedure(IN typed) named as SchoolsForGivenCategory accepts one parameter of category name like ‘ES’, ‘MS’, OR ‘HS’ and prints the names of schools in that category.
2. The second stored procedure(IN-OUT typed) named as GovernanceForGivenCPSunit accepts two parameters

1. School cps unit (id of school)

2. A variable to store the returned value

The procedure will return the governance type associated with school.

1. The third stored procedure(OUT typed) named as school\_with\_most\_programs

Accepts one parameter. The procedure will return the name of school which has the most programs among all schools in CPS.

**OUTPUTS:**

OUTPUT OF VIEW:

SELECT \*

FROM schools\_per\_geogrphic\_area;

[View output.csv](View%20output.csv)

Output of first stored procedure:

CALL SchoolsForGivenCategory('MS');

<first_procedure_output.csv>

Output of second stored procedure:

CALL school\_governance(1105, @governanceType);

SELECT @governanceType;

**'Charter'**

Output of third stored procedure:

CALL school\_with\_most\_programs(@schoolName);

SELECT @schoolName;

**'Collins Academy High School'**

**System’s limitations and suggested improvement:**

The CPS dataset entries are not too large so our system works fine. If the data entries were very large then it was need to improve the ram or other components of the system hardware or search for better ways to load and query the data. For example, loading data by using “(import records from external file) in the inserts choice” takes lots of time and workbench may crush but loading by script is much faster.

**load** **data** **local** infile "file location"

**into** **table** tablename fields terminated **by** ',' enclosed **by** '"';

**Full relational table specification (SQL):**

For full relational table specification of our database in DDL please click the following link.

<school_db.sql.txt>

**Codes for acquiring data:**

The python codes which were used for cleaning data, is provided in the bellow link.

<Python_codes.docx>

**SQL\_CODES:**

[**Sql\_codes.docx**](Sql_codes.docx)