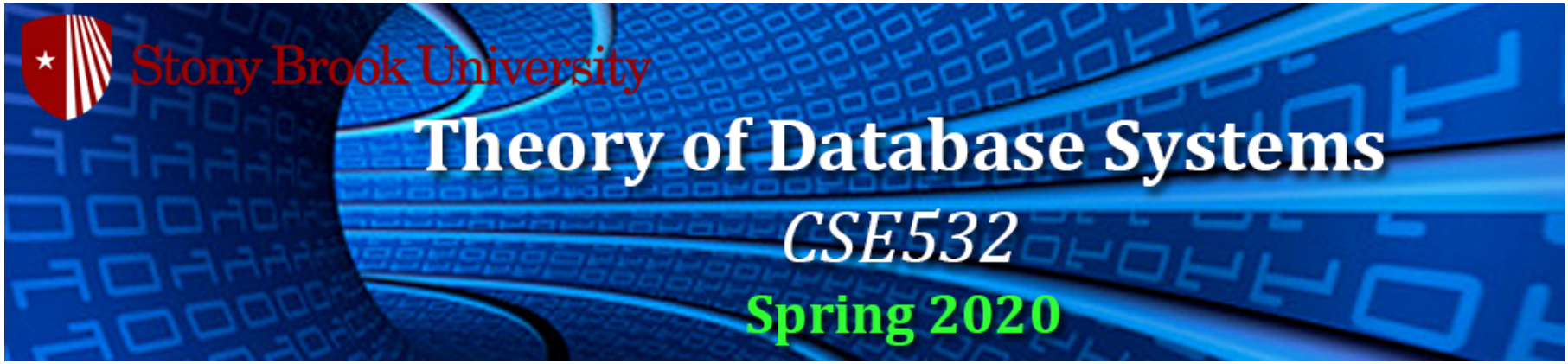

Home

CSE532

Home
Syllabus
Schedule
Homework
Textbooks
Examples
Instructor
Resources

Homework >

Homework3: Spatial Queries

References:

- Please go through DB2 Spatial Tutorial (I, II) before you begin this homework.
- Please also read spatial query examples.
- DB2 Spatial Online Reference (V11.5) PDF Version

Datasets:
There are three datasets to be used in this homework:

1. New York State Health Facility General Information (Download the Health_Facility_General_Information.csv for this homework). This includes the address and geolocation of each healthcare facility:

```
Facility ID, Facility Name, Description, Facility Address 1, Facility Address 2,Facility City, Facility State, Facility Zip Code, Facility County Code, Facility County, Facility Latitude, Facility Longitude
```

2. Health Facility Certification Information (download the Health_Facility_Certification_Information.csv for this homework). This includes information on certifications for services and beds for each facility.

```
Facility ID, Facility Name, Description, Attribute Type, Attribute Value, Measure Value, County
```

3. The US zip code tabulation areas from US Census Bureau, which contains the boundary of each zip code (download tl_2019_us_zcta510.zip). Once you unzip the file, you can run the following command under db2 command line to find metadata for the format. (Documentation of TIGER datasets can be found here.)

```
db2se shape_info -fileName tl_2019_us_zcta510.shp
```

The shapefile has a multipolygon object to represent the boundary of each zip code.

Tasks (10 + 3 extra credit points)

1. (3 points) **Setup the database.**

a. Enable the sample database (or your own database) for spatial support:

```
db2se enable_db sample
```

b. Load the zip code area dataset using the import SQL file:

```
db2 -tf import_zip.sql
```

c. Create two tables for facilities using the createfacilititytable.sql (we create two tables, cse532.facilityoriginal for original data, and cse532.facility with a spatial column).

```
db2 -tf createfacilititytable.sql
```

d. Load Health_Facility_General_Information.csv into cse532.facilityoriginal using script:

```
db2 load from "C:\myfolder\Health_Facility_General_Information.csv" of del MESSAGES load.msg INSERT INTO
cse532.facilityoriginal
```

e. Write a SQL script **facilityinsert.sql** to insert data into cse532.facility by selecting data from cse532.facilityoriginal table and converting (*Latitude, Longitude*) attributes into DB2GSE.ST_POINT type with srs_id 1 for *geolocation* attribute in cse532.facility.

f. Create a SQL script **createfacilititycertificationtable.sql** to create a table:

```
cse532.facilitycertification (FacilityID, FacilityName, Description, AttributeType, AttributeValue, MeasureValue, County)
```

and load the csv file into the table:

```
db2 load from "C:\yourpath\Health_Facility_Certification_Information.csv" of del MESSAGES load.msg INSERT INTO
cse532.facilitycertification
```

g. Update the **createindexes.sql** to add additional indexes besides spatial indexes for the queries below.

```
db2 -tf createindexes.sql
```

2. (3 points) **Write a query nearester.sql to find closest healthcare facility with an ER room (AttributeValue = 'Emergency Department') from "2799 Horseblock Road Medford, NY 11763"(40.824369, -72.993983) (latitude, longitude)**. Please return location and distance in your result. You can use unit 'KILOMETER', 'METER', or 'STATUTE MILE' for distance measurement.

Nearest neighbor search is not directed supported by DB2. You can use ST_BUFFER to create a buffered area (polygon/circle) from a point within a certain distance and search only stores within the buffer. Note that 0.25 degree is roughly 10 miles. For all the datasets, we use spatial reference nad83_srs_1 with srs ID as 1.

You can find information here on functions such as ST_POINT, ST_BUFFER, ST_WITHIN or ST_CONTAINS, and ST_DISTANCE.

3. (3 points) **Write a query noerzips.sql to find zip codes without any "Emergency Department", neither in their neighboring zip codes.**

4. (1 point) **Drop all indexes and perform the two queries again, and compare the query performance in terms of execution time for above two queries.**
Show your time difference with and without indexes in your **README** file.
You can refer to how to estimate query time.

5. (Extra credit: 3 points). **Write SQL queries or stored procedure mergezip.sql to merge zip code areas into large ones with neighboring zip code areas,** so that the new population in each zip code is large than the current average population, using the zip code population table in Homework1. For simplicity, you can remove the duplicates from the population table.

Submission.

Please zip your SQL scripts, results, and a README file. A result file can be a text file or a screenshot.

Please go to blackboard, and submit it under homework 3.

Subpages (1): querytime

Health_Facility_Certification_Information.csv (...	Fusheng Wang, Mar 8, 2020, 5:16 PM	v.1	
Health_Facility_General_Information.csv (659k)	Fusheng Wang, Mar 9, 2020, 5:40 PM	v.3	
createfacilititytable.sql (1k)	Fusheng Wang, Mar 9, 2020, 5:40 PM	v.3	

createindexes.sql (0k)	Fusheng Wang, Mar 9, 2020, 6:10 PM	v.1	
import_zip.sql (0k)	Fusheng Wang, Mar 9, 2020, 6:06 PM	v.2	