**Homework 3 README FILE**

**Tasks**

**1.  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](https://docs.google.com/a/stonybrook.edu/viewer?a=v&pid=sites&srcid=c3Rvbnlicm9vay5lZHV8Y3NlNTMyLXMyMHxneDo1Yjc2ODRkMmI4MDM1Mjg1):

db2 -tf import\_zip.sql

     c. Create two tables for facilities using the [createfacilititytable.sql](https://docs.google.com/a/stonybrook.edu/viewer?a=v&pid=sites&srcid=c3Rvbnlicm9vay5lZHV8Y3NlNTMyLXMyMHxneDo0Yzc1NzRhZWRkMmVjMTA3)(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**](https://docs.google.com/a/stonybrook.edu/viewer?a=v&pid=sites&srcid=c3Rvbnlicm9vay5lZHV8Y3NlNTMyLXMyMHxneDo4MDljYTZhNmJkZmJkYjQ) to add additional indexes besides spatial indexes for the queries below.

db2 -tf createindexes.sql

**2.  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](https://www.google.com/url?q=https%3A%2F%2Fwww.ibm.com%2Fsupport%2Fknowledgecenter%2Fen%2FSS6NHC%2Fcom.ibm.db2.luw.spatial.topics.doc%2Fdoc%2Frsbp4038.html&sa=D&sntz=1&usg=AFQjCNF-9TFBy8zAxnqJJ-9iqsJ0ItFG8A) 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](https://www.google.com/url?q=https%3A%2F%2Fwww.ibm.com%2Fsupport%2Fknowledgecenter%2Fen%2FSSEPGG_11.1.0%2Fcom.ibm.db2.luw.spatial.topics.doc%2Fdoc%2Fdb2sb169.html&sa=D&sntz=1&usg=AFQjCNEbijkfIxoHZwY_3F0pbvE8QGSXWQ) on functions such as ST\_POINT, ST\_BUFFER, ST\_WITHIN or ST\_CONTAINS, and ST\_DISTANCE.

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

**4.  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](https://sites.google.com/a/stonybrook.edu/cse532-s20/homework/homework3-2/querytime).

**Q1d**

load1.sql loads the data from Health\_Facility\_General\_Information.csv

**Q1 f.**

load2.sql loads the data from Health\_Facility\_Certification\_Information.csv

**Q4 .**

We ran query 2 and 3 with and without using indexes.

So the following 2 images will prove that using indexes reduces the time of execution

**Image 1** shows the timestamp of Q1(nearester.sql) and Q2(noerzips.sql) **without using indexes.**

A screenshot of a cell phone

Description automatically generated

So q1 takes 185134 microseconds and q2 takes 3 mins 69 sec 780660 microseconds.

**Image 2** shows the timestamp of Q1(nearester.sql) and Q2(noerzips.sql) **using indexes**.

A screenshot of a cell phone

Description automatically generated

So using indexes q1 takes 68259 microseconds and q2 takes 1 mins 95 sec 766803 microseconds.

So using indexes the execution time of both the queries reduces .