

Teaching Database Concepts using Spatial Data Types

Richard Dargie
UCOL
138 Queen St
Palmerston North, 4442, NZ
+64 06 952 7001
r.dargie@ucol.ac.nz

Aaron Steele
UCOL
138 Queen St
Palmerston North, 4442, NZ
+64 06 952 7001
a.steele@ucol.ac.nz

ABSTRACT

In this paper a process for incorporating core GIS principles into a database development course for level 6 students currently undertaking the Bachelor of Information and Communications Technology degree is described. This paper builds upon previous research and provides a description of how GIS principles were integrated into an existing database development curriculum as well as a set of practical instructions which were developed that utilised GIS and spatial data to enhance the teaching and learning of SQL and database concepts. Initial feedback from students indicates that the immediate visual feedback provided by the spatial data is viewed as beneficial.

Categories and Subject Descriptors

H.2.8 [Database Applications]: Spatial Databases and GIS.

General Terms

Theory, Documentation, Experimentation

Keywords

GIS, database, practical

1. INTRODUCTION

In today's marketplace computing power is no longer restricted to desktop machines. More and more mobile computing devices (laptops, tablets, smart phones, etc.) are emerging with processing and storage capacities comparable to many desktop computing solutions. Most of these devices (whether desktop or mobile) have some form of Geographical Information System (GIS) capability and the uses of this technology are growing. Users can view the location of a particular restaurant, find their way to some destination from their current location, and even capture locations of other online users. The practical application of GIS is growing and the technology is being embraced by everyday consumers without the need for any type of specialised training. It can be argued that this ease of adoption is likely due to the natural form that GIS data takes, i.e. maps, locations, and directions, data which are easily understood by most people. It has also been suggested that utilising the easily understandable data could be beneficial when teaching database concepts [1].

Underlying every GIS is the storage, retrieval and manipulation of spatial data, actions which are now commonly supported by many of the major database management systems. These systems rely on the use of the Structured Query Language (SQL) for the manipulation of both normal relational and spatial data. The teaching and learning of SQL and fundamental database concepts

This quality assured paper appeared at the 4th annual conference of Computing and Information Technology Research and Education New Zealand (CITRENZ2013) incorporating the 26th Annual Conference of the National Advisory Committee on Computing Qualifications, Hamilton, New Zealand, October 6-9, 2013. Mike Lopez and Michael Verhaart, (Eds).

is reasonably well covered within the literature [2, 5, 8]. Although a considerable amount of research exists in the field of GIS [6], it appears that little has been done regarding the utilisation of spatial data for SQL education [1].

Introducing and educating students in the use of GIS and spatial data is dependent on many factors. GIS hardware and software must be available and accessible by the students and educators. The GIS content must be distilled into a form that suits a particular level of student learning, and the GIS content must also fit within the learning outcomes of an existing curriculum.

Accordingly the following sections will describe how GIS and spatial data has been introduced into a level 6 database development course within a Bachelor of Information and Communications Technology degree in order to enhance the teaching and learning of SQL and database concepts.

2. Background

In a previous study, Burt theorised that spatial database software could be used for teaching SQL [1]. Burt argued that the immediate visual feedback afforded by spatial data would be of great benefit to both students and teachers [1].

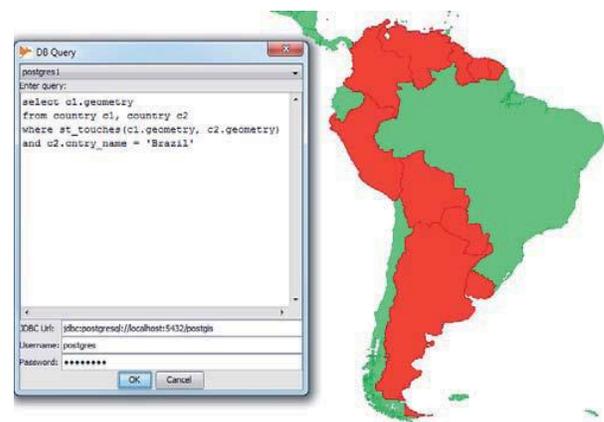


Figure 1. An example of a spatial data query result (reproduced with permission from Burt[1])

Burt provided 'proof of concept' examples that showcased the potential uses of GIS and spatial data within the context of SQL education. Specifically, the select statement, where clause, table joins, aggregate functions, and group by functions were demonstrated [1]. However, his study concluded that formal lesson plans needed to be developed (that expanded on his initial ideas) before the approach could be trialled with real students [1].

Accordingly, this paper builds on the work started by Burt and presents formal lessons that utilise GIS and spatial data for teaching SQL that have been developed for trial with real students

and also presents how they have been integrated within the existing BICT curriculum.

3. METHODOLOGY

In order to introduce GIS and spatial data for the teaching of SQL to students, an analysis of the existing BICT curriculum was undertaken in order to determine the most appropriate level and fit for the content within the programme. The results of this analysis will be presented at the beginning of the following section.

Once a suitable course for the GIS content had been decided upon, a decision needed to be made regarding which GIS software package would be most suitable for the course.

Following from this, the next step was to analyse the existing SQL lesson plans in order to identify which lessons would be most appropriate for enhancement with spatial data and spatial database concepts.

The identified lesson plans were then updated in order to incorporate GIS and spatial data into the learning activities. Subsequently these lessons were trialled with students during 2012. Student feedback was collected via the lecturer acting as a participant observer. An overview of the development of GIS and spatial data SQL lessons and feedback will be presented in the following section.

Based on the successful outcome of the 2012 trial, further SQL lessons were identified and have also been updated in order to utilise GIS and spatial data. Although these new lessons have not yet been trialled with students, an overview will also be presented in the results section.

4. RESULTS AND DISCUSSION

In this section the results of the BICT SQL curriculum analysis are presented. This is followed by the development of initial GIS and spatial data SQL lessons plans trialled during 2012. Finally, additional GIS and spatial data SQL lessons that were developed subsequent to the 2012 trail are presented.

4.1 Course Analysis

Analysis of the BICT curriculum identified three potential courses in which GIS and spatial data could be potentially incorporated into the existing curriculum. A brief description of the three courses will be given below.

D111 Data Fundamentals was a level 5 course focused on providing students with an introduction to databases and database technologies. The main technology used within the course was Microsoft Access, although some SQL content was also included.

D211 Database Development was a level 6 course that focused primarily on SQL and utilised Microsoft SQL Server.

D311 Advanced Database Concepts was a level 7 course that had a primary focus on advanced data warehousing techniques.

Through the analysis of the three potential courses it was decided that the D211 Database Development course would be the most appropriate location for the GIS and spatial data content due to its main focus on generic SQL actions. It was felt the D111 Data Fundamentals paper in its existing form was providing a worthwhile and appropriate introduction to database technologies through the use of Microsoft Access. Although the D111 course did include some SQL content, it was noted that it was only introduced towards the end of the course and that it would not provide enough time to adequately introduce GIS content. Furthermore, the initial exposure to SQL provided by the D111 level 5 was also seen as beneficial to level 6 students beginning

the level 6 D211 course. Conversely, it was felt that the D311 Advanced Database Concepts course would not be suited to the spatial data content due to its more complex focus on specific data warehousing operations and the required prerequisite knowledge of fundamental SQL and database concepts.

Having decided upon the D211 Database Development course as the most suitable place for introducing the GIS and spatial data, the next step was to decide upon a specific software package. In his initial work Burt had suggested that both Microsoft SQL Server and PostGIS were viable candidates for GIS and spatial data education [1]. Using these suggestions as an initial filter, Microsoft SQL Server was selected as the software package to be used due to its existing role with the D211 Database Development course (i.e. it was what the course was already using). This also provided access to the SQL Server Management Studio (SSMS).

4.2 2012 Development and Trial

In 2012 a series of practical instructions were developed to provide students with the spatial data necessary to produce visual results in the form of maps and accompanied data associated with the results.

Initial data for the practical lessons was obtained in the form of shape files from [4]. The shape files were imported into Microsoft SQL Server 2008 using the Shape2SQL tool obtained from [7].

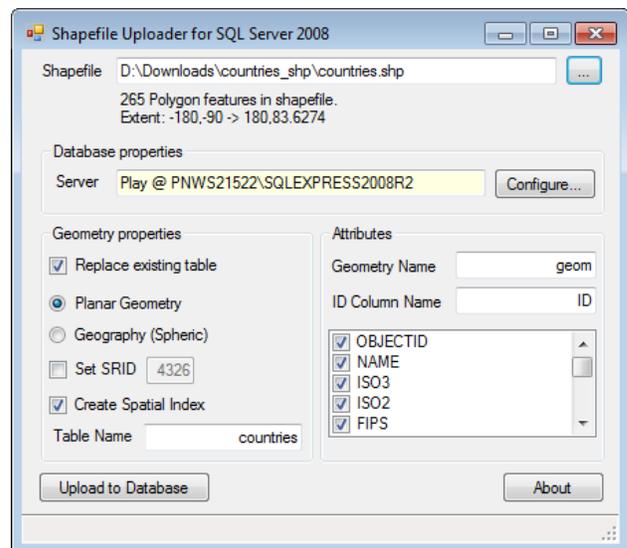


Figure 2. The Shape2SQL tools main interface

A diagram screenshot of the countries table was created for the students. Students would be required to use T-SQL (Transact-SQL, which is Microsoft's specific dialect of SQL) to recreate the countries table based on the screenshot information. This would cause the shape files to be combined into a single table. The students would then be required to apply relational database normalisation. Students were able to create and join multiple tables using T-SQL, additionally they were able to apply *renaming*, *creating*, *dropping* and apply column constraints.

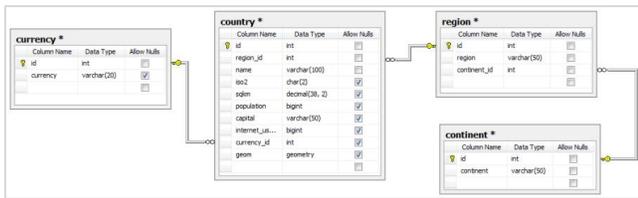


Figure 3. The countries table fully normalised, showing the inclusion of the geometry data type field

A combination of step by step instructions and exercises were developed for the students that required the practice of common database actions such as normalisation, and various other SQL Data Definition Language (DDL) table modifications. A model of the eventual database was provided from which table relationship constraints were established. Students would be required to perform SQL Data Manipulation Language (DML) techniques to populate the normalised tables from existing values. Additional DML practical exercises were provided to teach basic SQL select statements involving JOINS, the use of aggregate functions, the WHERE, GROUP BY, HAVING and ORDER BY clauses. In most cases the inclusion of the GEOMETRY spatial data type was required so a visual result in the form of a map representing countries, continents or regions was outputted.

The developed practical exercises resulted in a series of activities that allowed students to perform common SQL actions on spatial data. Students were able to undertake basic SQL SELECT queries using JOINS, the use of aggregate functions, the WHERE, GROUP BY, HAVING and ORDER BY clauses. Because of the inclusion of the GEOMETRY spatial data type within the countries table, visual results were produced in the form of map boundaries along with associated data to accompany the result. The visual feedback allowed students to quickly determine the accuracy of their SQL queries.

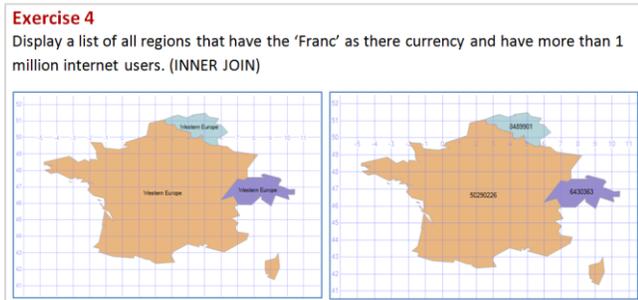


Figure 4. The Spatial results tab of SSMS displaying in visual format; the result of a select query using the GEOMETRY spatial data type

During the 2012 trial students were observed to actively engage with the GIS and spatial data database activities and exercises. Initial feedback from the students was positive and with many noting the immediate visual feedback as a highlight of the exercises. Due to this positive feedback, it was decided that it would be worthwhile to develop additional GIS and spatial data practical lessons for 2013.

4.3 2013 Development

In 2013 it was decided to extend the GIS component further by incorporating it into *stored procedures* and *triggers* (these are advanced database concepts that allow the reuse of collections of database commands). A brief introduction to GIS was added and

included with each set of practical instructions to clarify some of the terminology. This included a brief history of the use of GIS including early manually developed systems used for monitoring the spread of disease.

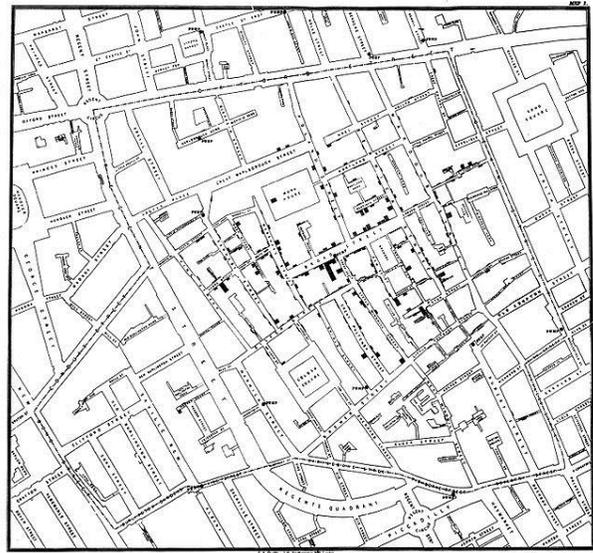


Figure 5. On the Mode of Communication of Cholera (2nd Ed) [9]

For each practical, short descriptions were created describing key GIS terminology that was included in the lesson. This was done to introduce new concepts and to provide valid reasons for including this subject into the lesson.

Point
A *point* can be used to represent single geographical locations, such as addresses, oil wells, or vehicle locations (Geometry, n.d.).



Line
A *line* used to represent linear geographical features such as roads, rivers, power lines, or water pipe (Geometry, n.d.).



Polygon
A *polygon* is used to represent geographical features, such as countries, lakes, or land parcels (Geometry, n.d.).



Figure 6. A brief description of the vector data used within the practical lessons

The database components were included and involved the fundamentals for creating database *stored procedures* and *triggers*. The practical lessons combined the two subjects to produce database objects able to process the GIS spatial data provided.

```

CREATE PROCEDURE udp_points (@point GEOMETRY, @description VARCHAR(20))
AS
    INSERT INTO points VALUES(@point, @description);
GO

CREATE PROCEDURE udp_polygons (@polygon GEOMETRY, @description VARCHAR(20))
AS
    INSERT INTO polygons VALUES(@polygon, @description);
GO

--Add latitude & Longitude values
EXECUTE udp_points 'POINT(175.615936 -40.349102)',
'Aoranqi Building Location'

--Add a polygon
EXECUTE udp_polygons
'POLYGON((175.614899 -40.348236,
175.615921 -40.347786,
175.617004 -40.349194,
175.615982 -40.349644,
175.614899 -40.348236))',
'Aoranqi Hospice Land'

```

Figure 7. A sample of the T-SQL the code for creating stored procedures and passing in vector data as parameters

Lessons were created to demonstrate the three *vector* types POINT, POLYGON and LINESTRING. From these, complex outputs were created by combining them into collections. Initially the vectors were applied to database variables and outputted using SELECT statements on those variables. Students were then shown how to insert the *vectors* into the GEOMETRY data type field. Stored procedures were then introduced and students could then pass *latitude* and *longitude* values into the procedures to demonstrate parameters. Spatial functions were applied to the latitude and longitude values within the stored procedures to demonstrate core GIS principals.

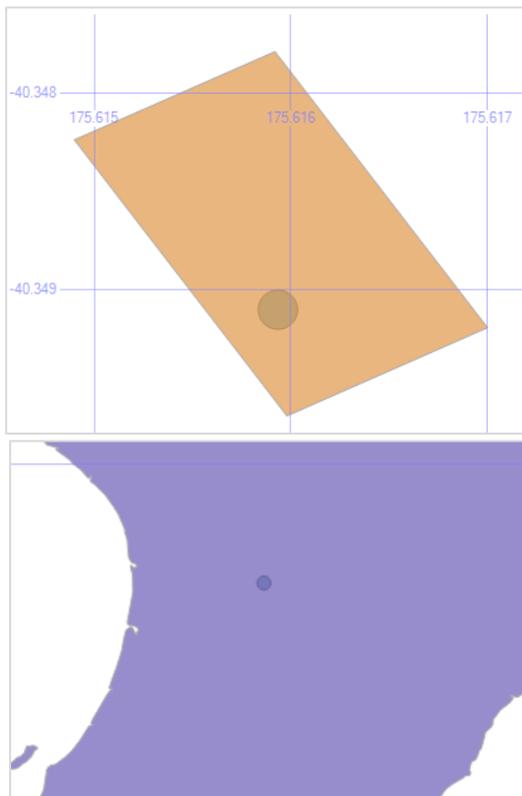


Figure 8. The Spatial results tab of SSMS displaying in visual format; the result of a select query using polygon and point vectors which includes the STBuffer function

The concept of T-SQL DATABASE, AFTER and INSTEAD OF triggers was then introduced. While the triggers did not act directly on the vector data, they enabled the ability to log the insert events and prevent the dropping of tables.

Students apply built in spatial functions to the existing data such as *STBuffer*, *STArea*, *STLength* or *STDistance* to that data. The results can be viewed in tabular or the SSMS visual format.

The coordinate values are stored within a tables GEOMETRY field in hexadecimal format. By making use of the T-SQL *ToString()* function, spatial data was able to be converted from its stored hexadecimal format and viewed in its original form using SSMS.

It was decided to add the spatial data to existing KML (Keyhole Markup Language) [3] files so that this could be imported into Google Maps or Google Earth and viewed as overlays. The KML files were first created within Google Maps and used as a template. This was achieved by creating a new map and adding simple vector shapes to it. The map was exported in the KML format and opened to examine the XML structure. Existing coordinate values were subsequently removed and replaced by values from within the database, this was then resaved. The KML file was then exported back into Google Maps and viewed as an entirely new layer.



Figure 9. A sample of a KML file showing coordinates and the result when viewed within Google Maps

Ultimately this provided a way for students to apply the results of their database activities directly to Google Maps.

5. Conclusions

This paper set out to describe the incorporation of GIS and spatial data for the enhancement of SQL education into an existing database curriculum. In this instance, a level 6 database development course was identified as the most appropriate

location for the incorporation of GIS content. This was due to the courses focus on SQL combined with the fact that enrolled students would also have had some prerequisite SQL knowledge.

This study also presented the development and trial of various GIS and spatial data concepts for SQL education. The development of these lessons built on and enhanced proof on concept examples suggested in a previous study [1]. This study also extended the level of GIS and spatial data application to SQL through the incorporation of stored procedures and triggers. A method for integrating GIS data with Google Maps was also introduced.

Student feedback to the initial trial was positive and provided some confirmation that the immediate visual feedback was seen as beneficial by the students. However, as this result was only captured informally via participant observations on the part of the lecturer it would be premature to generalise this conclusion. The authors note this as a limiting factor within this study. However they are also confident that the developed GIS and spatial data lessons are an asset to student SQL learning.

Nevertheless, it would be worthwhile to conduct further research that could include a trial of the developed lessons with a specific focus on formally collecting student feedback via other data collection methods such as levels of academic achievement, student interviews, and student surveys in order to more accurately capture student perceptions of the lessons and the suggested benefit to student learning.

6. REFERENCES

- 1 Burt, S. 2011. Exploiting Spatial Database Software for use in Teaching. In *Proceedings of the 2nd annual conference of Computing and Information Technology Research and Education New Zealand*. Rotorua, New Zealand: CITRE NZ2011, pp. 35-42.
- 2 Caldeira, C.P., (2008). Teaching SQL: a case study. In *Proceedings of the 13th annual conference on Innovation and technology in computer science education*. Madrid, Spain: ACM, pp. 340-340.
- 3 Google (February, 2012). *Keyhole Markup Language* [Online]. Available <https://developers.google.com/kml/documentation/>
- 4 Hijmans, R. (2013, July 30). *Free Spatial Data* [Online]. Available <http://www.diva-gis.org/Data>
- 5 Kearns, R., Shead, S. & Fekete, A., (1996). A teaching system for SQL. In *Proceedings of the 2nd Australasian conference on Computer science education*. Melbourne, Australia: ACM, pp. 224-231.
- 6 Milson, A., Demirci, A. & Kerski, J (Eds.) (2012). *International Perspectives on Teaching and Learning with GIS in Secondary Schools* Springer.
- 7 Nielsen, M. (2013, July 30). *SQL Server Spatial Tools* [Online]. Available <http://www.sharpgis.net/page/SQL-Server-2008-Spatial-Tools.aspx>
- 8 Raadt, M.D., Dekeyser, S. & Lee, T.Y., (2006). Do students <i>SQLify</i>? improving learning outcomes with peer review and enhanced computer assisted assessment of querying skills. In *Proceedings of the 6th Baltic Sea conference on Computing education research: Koli Calling 2006*. Uppsala, Sweden: ACM, pp. 101-108.
- 9 Vinten-Johansen, P. (2013, July 30). *Online Companion to Cholera, Chloroform, and the Science of Medicine: A Life of John Snow*. [Online] http://johnsnow.matrix.msu.edu/images/online_companion/c_hapter_images/fig12-5.jpg