Gabrielle Sudik
IT1025 CRN 12115 – Spring 2017
Professor Tamerlano
Lab 6
April 17, 2017

Executive Summary

The theme of this week's lab is data and how it can be used by data management tools. It also addresses why data should be standardized, ways that data can be used to improve companies' performances, ways to manipulate data to provide knowledge, and ways that data systems can be exploited.

Data, Information and Knowledge

At the gas station, an example of data is “2.39.” This is a figure, a measurement of something, but without context I don’t know what it measures. “$2.39” is information (ie, adding the dollar sign in front). It takes the data of 2.39 and both converts it to a dollar amount, so I understand what is being counted. “$2.39/gallon” is knowledge. It lets me know how much gas I can buy for that amount (and can therefore be compared to other gas stations that price by the gallon, or compare today’s price with last week, etc.) Finally, I’ll toss in a bit of wisdom (the final step on the data ladder) and note that walking saves money, is good for my health, and is good for the environment.

Data Normalization

The two main purposes of database normalization are to reduce duplication of data between tables, and giving tables as much flexibility as possible. It serves to limit both how many times data must be entered into a database and how many times it will show up when viewing the data.

Using Amazon.com as an example: Once a shopper creates an account, Amazon always knows that person’s name, address, email and (often) credit card information. A shopper can use one log-in to place several orders without having to re-enter all of their information each time. Amazon also allows for the storage of multiple pieces of data in one table (say, multiple addresses or multiple credit cards on file), and can connect all orders in one table. I believe a shopper can even group old orders to see how many times she has ordered a particular item.

If businesses did not normalize their databases, it could drive off a lot of customers, who will have to re-enter information for each order. It could also make tracking orders and returns more difficult, as the company would have to wade through too much information each time they helped a customer. And there would probably be problems tracking inventory, as the large scope of buyers means that inventory is constantly changing.

Important of Data Type Definitions

Correct data type definition is important in the context of large databases and data warehouses because, if the data is labeled incorrectly, the entire database can be rendered useless. For example, consider an online retailer that didn't correctly record zip codes or phone numbers. When customers' information is put into a database, those numbers can get messed up, possible resulting in misdirected packages, and the inability to call the customer to correct it. Another possible way that data type definitions can be messed up is if two different definitions are applied to the same piece of data. For example, a name might be called “customer” when someone places an order but “payee” if they request a refund. In this case, the same person might be in the database twice, which is inefficient and possibly confusing.

SQL

SQL is “structured query language.” It is the primary language users use to manipulate and analyze relational databases. It can do things like identify every individual in a database; list the total distinct items in a database (such as cities), or count the number of items in a group (students in a class, items in an order, etc.) Microsoft Access is an example of an SQL database. But I suspect that Microsoft must use its own interface between the user and the language (sorting, etc.) because I've never had to actually know the language to be able to manipulate data in it.

NoSQL is “not only structured query language.” This is similar to SQL but is designed to cover very large databases that might scan several servers or geographical locations, and might have multiple users analyzing or manipulating data at the same time. Amazon is one example – for instance, dozens of people around the world might simultaneously have the same tea pot in their shopping cart, yet the database will be able to tell new shoppers how many tea pots are left. (In my experience, this is an imperfect system; more than once I’ve been unable to buy an item advertised as “1 item remaining” – the “store” part of Amazon shows the item is available; the “shopping cart” shows it is unavailable.)

So Much Data!

At an airport, there is so much data worth storing. Some of this list might be more accurately called information, since it is data in context: Flight schedules; actual length of time in the air; number of planes leaving per hour or per day; wait times in line for check-in, security, and boarding; time it takes to board each type of plane; number of airport and airline staff needed for various activities; number of bags checked in; the kind or size of checked luggage; number of on time vs late departures and arrivals.

There is also a lot of data related to airport efficiency outside of the flights themselves: Number of parking spaces; seats on shuttles; number of shuttles; time it takes to get from parking to the terminal; number of rental cars; trips made via taxi or public transport; number of restrooms and trash cans; number of security checkpoints; seats in waiting areas; number of shops and restaurants; number of telephones; number of outlets and charging stations.

“Big Data: The Management Revolution”

According to the article, data-driven companies are 5% more productive and 6% more profitable than non-data-driven companies. That fact alone should encourage companies to be data-driven. But of course the underlying reason is that data-driven companies have a better understanding of where their money is being earned and spent, what parts of the company or products are the most valuable, and – especially for very large companies – they have an advantage in learning these things more quickly than competitors.

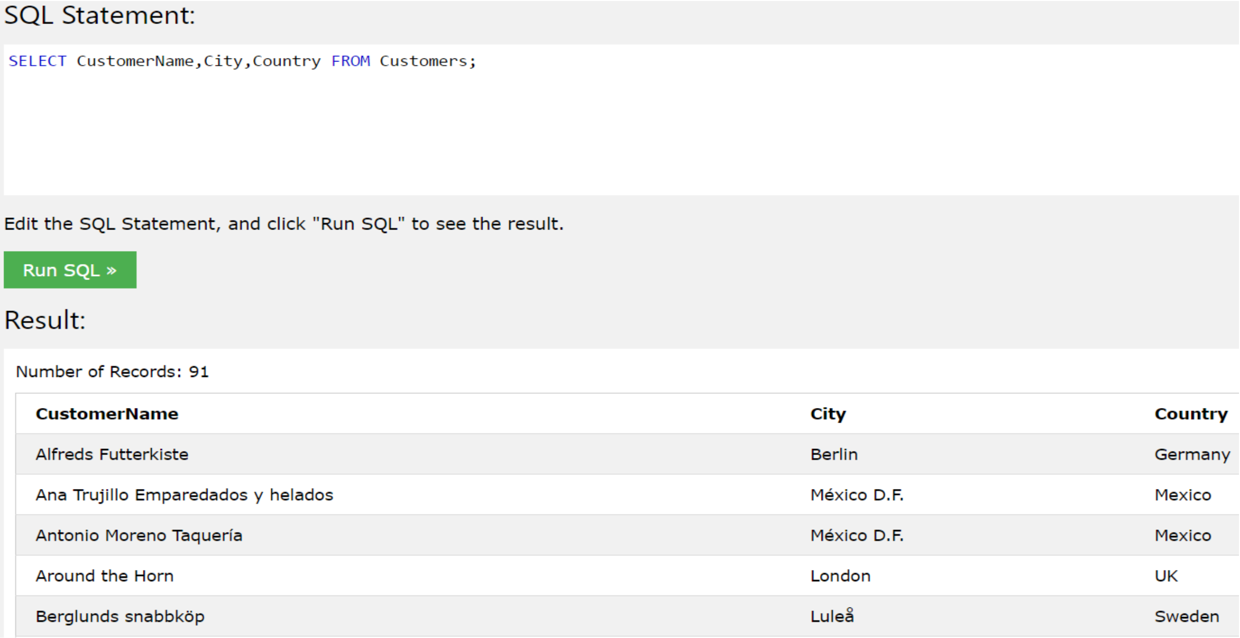
I saw a real-world example several years ago when my part time job in college was at a consignment shop for women’s and men’s clothing. The owner consigned shoes, but she thought they were a pain and (being used shoes) she assumed people didn’t buy them often. So she tended to just stuff the shoes any old place, often letting them get lost under racks of clothes. Then the college’s business school offered several small businesses the opportunity to have their businesses analyzed for free by the MBA students. The students studied all of the shop’s data, ranging from how many items sold and didn’t sell, at what prices, at what times of year, and how much relative profit each group of items made (and also the shop’s expenses). It turned out that shoes were – by far – the biggest money-maker per item sold. And Dena went from just kicking the shoes underneath racks of clothes, to highlighting them on some well-lit shelves, and her shoe sales increased along with the profits.

Using SQL

SQL's main role in web sites is to find, sort and display the data or information that the user wants to see. This can be anything from a Facebook feed that is instructed to display family members' posts, to news aggregators or search engines that will narrow searches (although I suspect there is more to it when it comes to search engines), to looking up books in an online library card catalog.

In the code: “*SELECT * FROM Customers”* the asterisk means “all.” The word customers refers to which table will be searched. So this code instructs the database to show all entries in the customer table.

Here is an example of SQL code to select CustomerName, City and Country from the Customers table of our tutorial:



SQL Statement:

```
SELECT CustomerName,City,Country FROM Customers;
```

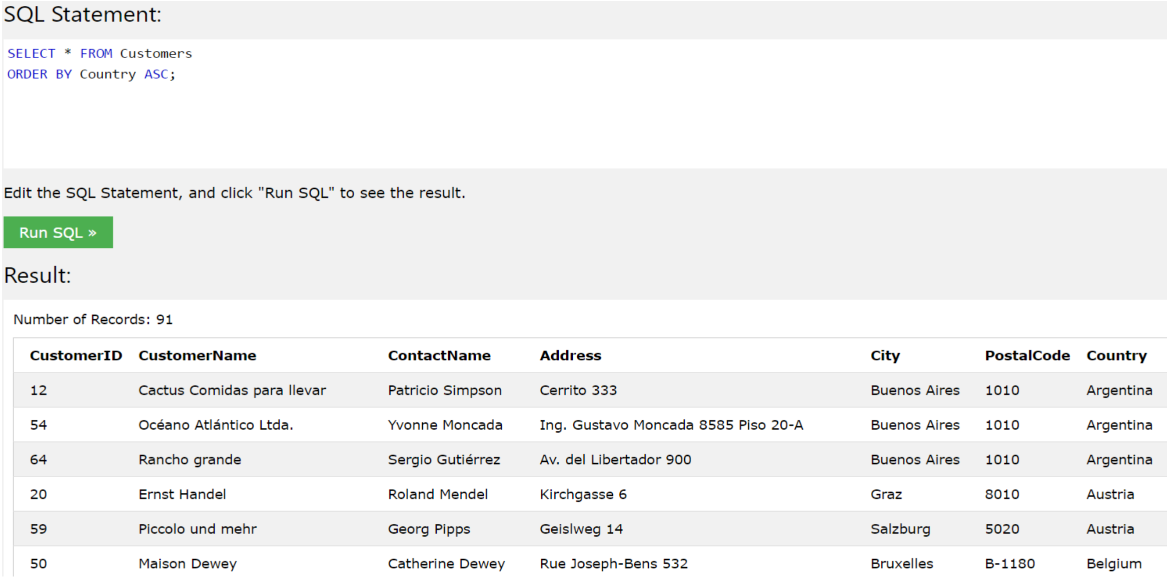
Edit the SQL Statement, and click "Run SQL" to see the result.

Result:

Number of Records: 91

CustomerName	City	Country
Alfreds Futterkiste	Berlin	Germany
Ana Trujillo Emparedados y helados	México D.F.	Mexico
Antonio Moreno Taquería	México D.F.	Mexico
Around the Horn	London	UK
Berglunds snabbköp	Luleå	Sweden

For some reason I am incapable of sorting this list by country, but I can sort all of the fields by country:



SQL Statement:

```
SELECT * FROM Customers
ORDER BY Country ASC;
```

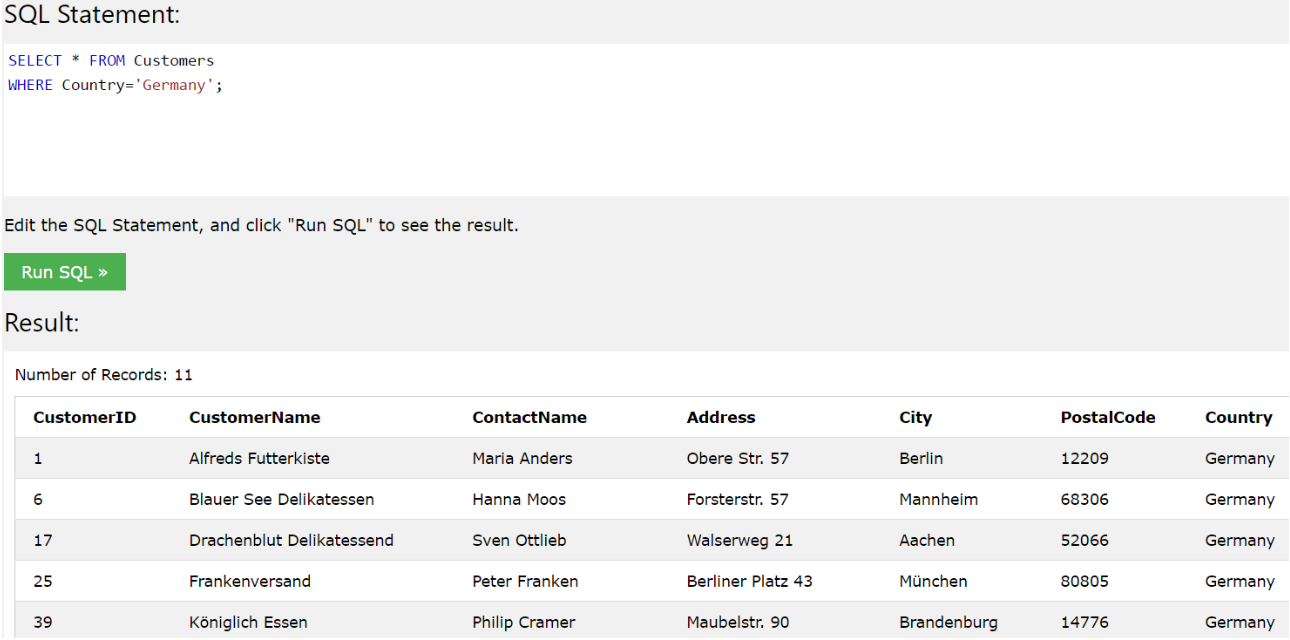
Edit the SQL Statement, and click "Run SQL" to see the result.

Result:

Number of Records: 91

CustomerID	CustomerName	ContactName	Address	City	PostalCode	Country
12	Cactus Comidas para llevar	Patricio Simpson	Cerrito 333	Buenos Aires	1010	Argentina
54	Océano Atlántico Ltda.	Yvonne Moncada	Ing. Gustavo Moncada 8585 Piso 20-A	Buenos Aires	1010	Argentina
64	Rancho grande	Sergio Gutiérrez	Av. del Libertador 900	Buenos Aires	1010	Argentina
20	Ernst Handel	Roland Mendel	Kirchgasse 6	Graz	8010	Austria
59	Piccolo und mehr	Georg Pipps	Geislweg 14	Salzburg	5020	Austria
50	Maison Dewey	Catherine Dewey	Rue Joseph-Bens 532	Bruxelles	B-1180	Belgium

And here is a query selecting only customers from Germany:



SQL Statement:

```
SELECT * FROM Customers
WHERE Country='Germany';
```

Edit the SQL Statement, and click "Run SQL" to see the result.

Result:

Number of Records: 11

CustomerID	CustomerName	ContactName	Address	City	PostalCode	Country
1	Alfreds Futterkiste	Maria Anders	Obere Str. 57	Berlin	12209	Germany
6	Blauer See Delikatessen	Hanna Moos	Forsterstr. 57	Mannheim	68306	Germany
17	Drachenblut Delikatessend	Sven Ottlieb	Walserweg 21	Aachen	52066	Germany
25	Frankenversand	Peter Franken	Berliner Platz 43	München	80805	Germany
39	Königlich Essen	Philip Cramer	Maubelstr. 90	Brandenburg	14776	Germany

And… I attempted a query to select ProductName Unit and Price from the Products table where the price is over 30 dollars sorted by ProductName for a good 15 minutes, and I clearly don't understand using SQL. Bah. Moving on….

The International Organization for Standardization

The purpose of the International Organization for Standardization (ISO) is to promote standardization, raise awareness of the need for standardization, and teach it. The ISO is comprised of 163 countries and all sorts of non-governmental organizations, and works to create standards for materials, products, processes and services.

The organization has created nearly 22,000 standards. They governing everything from transportation, to telecommunication technology, to standardizing ways the handicapped can interact with the world, to safety standards in all industries, to agriculture and the environment. Basically, if things that humans use every day can be made more compatible so they works for more people, there are probably standards governing them.

Standards pertaining to databases include standards on information security, management systems, and codes to standardize common data, such as currency, time and locations.

SQL Injections

SQL injection is the placement of malicious code in SQL statements. It is mostly used in web-based databases, but can be used to attach any SQL database. Basically, to attack the database, instructions are typed into the instruction line that are either designed to avoid security measures, or sometimes are just so sloppy that they produce an unexpected result – one that reveals more than it should. Systems using SQL are vulnerable because people who are able to exploit weaknesses can steal, change or erase data. Some examples include changing one's bank balance, changing a grade in a class, or erasing criminal history. (citations: https://en.wikipedia.org/wiki/SQL_injection, https://www.acunetix.com/websitesecurity/sql-injection/

Conclusion

This lab concerned data, information and knowledge. It also addressed why data should be standardized, ways that data can be used to improve companies' performances, ways to manipulate data to provide knowledge, and ways that data systems can be exploited.