

Andrew Rokoszak  
Professor Alan Labouseur  
Database Systems CMPT 308  
January 26<sup>th</sup> 2016

## Lab 1: PostgreSQL

### **Data vs. Information:**

A good example of a database that takes data and turns it into information is a daily fantasy sports website like Draft Kings or Fan Duel. There are many elements of data stored within the database that must be defined and sorted to provide information to users. The purpose of the site is to create a team of the best players (in the case I will use an NBA team), whom you believe will score the most points for a team total that day. There are various statistics and data values that represent the score each player will get. Some of the statistical categories include points, rebounds, assists, blocks, steals, three-pointers and turnovers. Since all of the “data” is compiled of various numbers it has to be sorted into the different categories seen above, in order to become information for users. Without representing the data in specific categories users may see values of “23.3, 4.8, 5.2, 1.2, and 2.3” which without any definition are irrelevant. When listing those numbers in the same order of the category values (23.3 points, 4.8 rebounds, 5.2 assists, 1.2 blocks, and 2.3 steals, they gain meaning and show what would be a very productive player averaging those figures.

The value of this information is important in determining the pricing of players. Someone who averages figures like the ones listed above would be much more expensive than a player who only averages 12 points, 3 rebounds and 2 assists per game. For a Draft Kings users defining the values of these numbers is critical in determining the price of the players and what players will score the most points for the entire team that night. A player averaging the figures of (23.3 points, 4.8 rebounds, 5.2 assists, 1.2 blocks, and 2.3 steals) would likely have a value or cost of \$9,000. On the other hand, a player who averages the aforementioned 12 points, 3 rebounds and 2 assists per game may only be valued at \$5000. This is another reason why turning the statistical data into information is important for the functionality of the database and the users.

Finally, layers must be associated with certain positions, as the site requires users to fill in positions of point guard, shooting guard, small forward, power forward and center. Without giving each player a role and just having random names in the database, users would not be able to fill in the correct players in the correct positions. It is imperative for the random data to be collected and categorized to create valuable and functional information for users.

**Data Models:**

The original databases were called flat file systems. These “databases” were difficult to maintain and lacked the ability to easily manipulate the data into information and make it accessible for users. This flat file system is defined as a file of records of fields that wasted space recording information more than once. Furthermore there were no explicit relationships between the data inputted. The biggest issue was the fact that many companies used different storage methods making it impossible for one system to transfer data to another. Ultimately this system lacked physical data independence.

Following the flat file system, the IMS method was used, also known as the hierarchical model. This model is organized in a tree-like structure. Each different field entered had some piece of data associated with it. This made it difficult to access the data points further down the tree without going through the earlier branches. Some other issues that exist was adding pieces of data that did not necessarily associate with the top of the hierarchy but still existed as a piece of inventory in the data somewhere. For example, if a user had items A and B and another user had items B and C, there is nowhere to store that an item D still exists. Although this method achieved data independence it allotted for data duplication, which is a waste of space. Similar to the hierarchical model, the network model recued inconsistency and provided users the ability to create a separate space for what would be item D if the users or players did not possess the item. It also enables an item shared between users to be found between them as opposed to in each user’s inventory as a separate item.

The four key components of efficient databases are atomicity, consistency, independence and durability. The relational model eliminates the shortcomings of the previous models. In the relational model, data is managed as tables of rows and columns. This eliminates the need to have the same piece of data written twice, and allows users to manipulate the individual tables to easily change data while not impacting or significantly changing the rest of the database. XML or Extensible Markup Language is a type of file that can be read by both computers and people across all platforms. This is useful for the ever-expanding global market where many people need to be able to use different files across different platforms in a variety of languages.