Roman Bond
Dr. Rivas CMPT 308
1/27/16

Data vs Information

Data and information are two very different things. Data can create information how ever data itself is useless unless you are able to group it and understand it's significance. In our world today we are flooded with data, there are so many people doing so many things. Large companies are able to collect all of this data but the real work is finding a way to take these incomprehensibly large amounts of data and turning them into information. Companies such as Facebook are trying to squeeze out as much data as they possibly can so then they can show their partners and customers that they have so much information on people. Facebook went to the extreme of trying to test a script that would track a user's mouse to see what they hovered over even if they did not click a link. They modified it and now videos will begin playing when your mouse touches it. Facebook separates it's data and information into many different categories such as location, interests and connections. Facebook and other companies use this data to create useful information such as ad targets, friend suggestions, pushing content the user would enjoy and this is all done to grow the network and make more money. Google has made a lot of money because they created one of the biggest data oceans of all time. There are so many searches per second and with every search there is a premium link at the top of the result that companies pay for. An example of useful information that could be useless data is Amazon, all of their barcodes, credit cards, dates and addresses could all be completely useless if amazon didn't have the tools necessary to separate and compile that data.

Data Models

For Big Data Models you essentially have three main categories that a data model breaks down into for data base. The first is the smallest, which is a schema. A schema is comprised of tables, this is where you would group similar things such as addresses, dates, times and other significant numbers. The next layer up is a catalog which is a collection of schemas, this allows sites like amazon to group massive groups of tables into smaller groups so that the user does not need to get all the information, only the information they are looking for at that time. The final layer is a cluster, which is comprised of many catalogs, this allows for even larger separations of data. The issue with a relational model is that they can become way too big and complex to be useful. They begin to have too many references to other databases that the system becomes incredibly slow. XML models are similar to this, they split too much data up into too many subsections that it ends up slowing down processes. Having a large company like Facebook or Google they need instant access to information, they can't have large amounts of directories to sort though that slow down results. No one would use Google if it took 30 seconds to make a search.