Database Versioning

Software projects that require more than one person to contribute to and modify the same database should use some form of database versioning. If they did not, multiple problems would occur: data dependencies would be broken, tables and fields would become unsynchronized, etc. Keeping a single database for the project up to date is difficult. Many people would have to devote time to merge their copies of the database into a single usable one. Therefore, database versioning is becoming more widely used.

Versioning has two main purposes for production: First, versioning allows several parties to manipulate the same material simultaneously. And second, versioning stores previous versions of this material and allows for rolling back to a previous version. [1] The most common use of versioning is to allow multiple people to work on the same software project simultaneously using software versioning. This is accomplished through version control systems such as Git. [2] These software versioning tools allow large teams of programmers to all contribute to the same project safely and effectively. The inherent backup system provided by the tool ensures that the software is kept safe and malicious or unwanted changes can be undone if needed. And, the problem of multiple people accessing the same files is solved by allowing everyone to have their own copy of the shared files while keeping the latest version of the project as a whole in a common, central place - the version control system. Later, individuals’ changes can be merged into the latest version of the project. [2]

Versioning for databases is like software versioning in that files are stored in a version control system to both back them up and to keep the latest updated version of the database in a common, central location. [4] However, unlike software versioning, much more data must be kept describing a database compared to the scripts needed to describe a software project. The files needed to describe a database obviously include schemas, data definition language (DDL) scripts, and data instances. [5] This amount of data is not enough to ensure the dependencies in a database system will still hold, however. This is because databases, unlike source code, consist of dependencies rather than just objects. A change to one field in the database can have lasting effects on the rest of the database. Therefore, database configuration properties, server configuration properties, requirements document, scripts to define users, roles, and permissions, network configuration scripts, etc. are all required. [6] The amount of data that must be stored for each version is one problem with database versioning. Another problem with database versioning is the data dependent nature of databases. And a third problem with database versioning is that culturally many developers see databases as something illogical that cannot be version controlled. To many, because databases are not as simple as source code, version controlling a database seems like a monstrous task not worthy of their time and effort. [7]

Social media companies collect even more information on their consumers than credit card companies. Facebook, for example, collects information on consumers’ personal information, post history, likes and dislikes, comments, friends, and even how much time they spend scrolling through their feed [2]. Like credit card companies, the justification behind maintaining all this data for social media companies is that the data collected regarding users allows for the companies to provide more personalized content. Instagram, for example, has a newsfeed algorithm that gives posts different priorities based on data collected about users’ preferences and engagement history [3]. This personalized content is designed to increase the amount of time spent on a given social media site and hopefully increase the happiness of the user as a result.

Online shopping websites also collect a large amount of user data. Companies like Amazon collect data such as purchasing history, browsing history, method of purchase, spending habits, demographics, etc. This collection of data allows online shopping companies to curate their content towards consumers. In this way, companies can better provide recommendations and individualized coupons based on the browsing and purchasing history of consumers [4]. These recommendations have the potential to influence consumer’s purchasing decisions. This curated view of the online store can also have the effect of seemingly limiting a consumer’s options to a subset based on the consumer’s previous purchases.

There are three types of data that these companies collect when storing data about their consumers: first party data, second party data, and third party data. First party data is the main method of data collection for most companies. It is defined as the information that a company collects regarding its own consumers. This kind of data comes at no cost, is completely legal, and is created and owned by publishers and marketers [5]. An example of using first party data would include Amazon’s recommendation service. The service leverages info based on past purchases and interests to provide the service [6]. The past purchases and interests are first party data in that they are collected by Amazon. Unlike first party data, second party data is not collected by the same company. Instead, a second party collects information that a company uses. This usually occurs when one company makes a deal with a second company to gain access to their First party data. For example, data sold from a social networking site to a credit card company would constitute second party data. The main disadvantage that comes from using second party data is that users must be informed that their data is going to be shared to a different site [7]. Third party data is like second party data in that information is outsourced to a different entity. However, with third party data, information is collected from an organization that does not have any relationship to your company. This usually occurs when companies called data aggregators collect copious amounts of information either by collecting it themselves or buying it from many different publishers. The companies then sell this data to other parties. This kind of data is good for demographic or behavioral targeting. First, second, and third party data can be analyzed to provide information, trends, and statistical probabilities designed to improve services offered by companies.

Two major disadvantages arise from this mass collection of user data by these companies, however. First, opting out of some of this extraneous data collection is usually difficult and sometimes impossible. In this respect, many people agree that Americans have lost control of their personal data [8]. And, this might be costing us money as companies are selling our personal data to other large companies. Second, our personal data might not be stored in a safe way. Cyber-attacks such as the Equifax breach might cause users’ sensitive data such as names, social security numbers, birth dates, etc.to be stolen and used for malicious purposes [9]. In a study done by Pew Research Center, it was found that “a majority of Americans had personally experienced a major data breach” [10]. Privacy and security are the main disadvantages regarding the mass collection of user data.

To protect against these disadvantages, several laws have been enacted. And, while there is no single, comprehensive federal law for data collection in the US, there are several federal and state laws that regulate the collection and use of personal data. In addition, several interest groups and government agencies have developed regulatory tools to keep data collection in check [11]. The federal laws that have been put in place to regulate the collection of personal data include: The Federal Trade Commission Act (FTC Act) which prevents unfair online and offline data security policies and also protects children from online data collection, The Health and Insurance Portability Act (HIPAA) which regulates data about American citizens’ medical information, The Fair Credit Reporting Act which protects American citizens’ information in the files of consumer reporting agencies such as credit card companies, and The Financial Services Modernization Act which regulated data collected about financial information [12]. These laws, while they do provide some basic data protection to American citizens, do not adequately protect all of America’s user data. As such, some state laws have been passed in addition to the federal laws to protect Americans’ data. Although many states have enacted some form of privacy legislation, the best example of states passing this privacy legislation is California [13]. California passed The California Electronic Communications Privacy Act which limits the powers of law enforcement and government officials in the state of California to acquire electronic communication information [14]. In addition to state laws, regulatory organizations such as the Advertising Self-Regulatory Council have been established to create generally accepted procedures to protect and regulate data collection in their respective industries. For example, the ASRC suggests that an icon be placed on any website where tracking data is collected [15]. While many regulatory policies have been enacted both at the federal and state level, more protections need to be established before American citizens’ data is truly private and secure.

Privacy and security of personal data in modern-day America is lacking in that large companies such as Visa, Facebook, and Amazon collect vast quantities of data about citizens and can analyze and sell that data at their discretion. In the future, I hope that more stringent laws will be enacted to limit the powers of these data behemoths. I also plan to limit the amount of personal data that I give websites as I see the detrimental effects it can have towards my data’s privacy and security.

Works Cited

[1] Papandrea, Dawn. “When Your Card's Big Data Collection Goes from Cool to Creepy.”CreditCards.com, Creditcards.com, 15 Aug. 2016, [www.creditcards.com/credit-card-news/big-data-collection-cool-creepy.php](http://www.creditcards.com/credit-card-news/big-data-collection-cool-creepy.php).

[2] Thompson, Cadie. “Companies Aim to Cash in on Your Intimate Social Data.” CNBC, CNBC, 30 Oct. 2013, [www.cnbc.com/2013/10/30/what-companies-are-doing-with-your-intimate-social-data.html](http://www.cnbc.com/2013/10/30/what-companies-are-doing-with-your-intimate-social-data.html).

[3] Hunt, Elle. “New Algorithm-Driven Instagram Feed Rolled out to the Dismay of Users.” The Guardian, Guardian News and Media, 7 June 2016, [www.theguardian.com/technology/2016/jun/07/new-algorithm-driven-instagram-feed-rolled-out-to-the-dismay-of-users](http://www.theguardian.com/technology/2016/jun/07/new-algorithm-driven-instagram-feed-rolled-out-to-the-dismay-of-users).

[4] Carlozo, Lou. “How Online Retailers Collect & Use Your Data.” Dealnews, DealNews, 23 Dec. 2013, [www.dealnews.com/features/How-Online-Retailers-Collect-Use-Consumer-Data/938928.html](http://www.dealnews.com/features/How-Online-Retailers-Collect-Use-Consumer-Data/938928.html).

[5] Lotame. “1st Party Data, 2nd Party Data, and 3rd Party Data | Different Kinds of Data.”LOTAME, 31 Jan. 2018, [www.lotame.com/1st-party-2nd-party-3rd-party-data-what-does-it-all-mean/](http://www.lotame.com/1st-party-2nd-party-3rd-party-data-what-does-it-all-mean/).

[6] Limon, Kathleen. “THE DIFFERENCE BETWEEN FIRST, SECOND, AND THIRD PARTY DATA AND HOW TO USE THEM.” ReTargeter, 6 June 2014, blog.retargeter.com/general/difference-first-second-third-party-data-use.

[7] Dent, Alistair. “An In-Depth Look At Second-Party Data For Digital Marketers.” Marketing Land, 4 June 2015, marketingland.com/second-party-data-digital-marketers-128254.

[8] Madden, Mary. “Public Perceptions of Privacy and Security in the Post-Snowden Era.” Pew Research Center, 12 Nov. 2014, [www.pewinternet.org/2014/11/12/public-privacy-perceptions/](http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/).

[9] O'Brien, Sara Ashley. “Equifax Data Breach: 143 Million People Could Be Affected.” CNNMoney, Cable News Network, 8 Sept. 2017, 9:23 AM ET, money.cnn.com/2017/09/07/technology/business/equifax-data-breach/index.html.

[10] Olmstead, Kenneth, and Aaron Smith. “Americans and Cybersecurity.” Pew Research Center: Internet, Science & Tech, 25 Jan. 2017, [www.pewinternet.org/2017/01/26/americans-and-cybersecurity/](http://www.pewinternet.org/2017/01/26/americans-and-cybersecurity/).

[11] Jolly, Ieuan. “Practical Law.” Practical Law US (New Platform) Signon, 1 July 2017, content.next.westlaw.com/6-502-0467?transitionType=Default&firstPage=true&bhcp=1&contextData=%28sc.Default%29.

[12] Federal Trade Commission. “A Summary of Your Rights Under the Fair Credit Reporting Act.” doi:https://www.consumer.ftc.gov/articles/pdf-0096-fair-credit-reporting-act.pdf.

[13] NCSL. “PRIVACY LEGISLATION RELATED TO INTERNET SERVICE PROVIDERS-2017.”NCSL, 29 Dec. 2017, [www.ncsl.org/research/telecommunications-and-information-technology/privacy-legislation-related-to-internet-service-providers.aspx](http://www.ncsl.org/research/telecommunications-and-information-technology/privacy-legislation-related-to-internet-service-providers.aspx).

[14] Leno, Mark, and Joel Anderson. “California Electronic Communications Privacy Act (CalECPA) - SB 178.” ACLU of Northern CA, 9 May 2017, [www.aclunc.org/our-work/legislation/calecpa](http://www.aclunc.org/our-work/legislation/calecpa).

[15] ASRC. “Advertising Industry Self-Regulation.” Advertising SelfRegulation Council RSS, 2012, [www.asrcreviews.org/](http://www.asrcreviews.org/).