Neal Sakash

Dr. Paul Buhler

DATA 210:02

May 4th, 2015

Question One: Polyglot Persistence

Knowledge of product and user data has become a paramount concern for businesses in recent years and the ability to handle the volume, velocity, and variety of big data has placed a strain on traditional databases. The recent development of diversified NoSQL technologies offer varied solutions to the limits of RDBMS, however they come with their own drawbacks which may prevent companies from investing the time and money to fully utilize their features. One argument is to follow the notion of polyglot persistence and weave NoSQL technologies with an already established RDBMS model to best support different data storage needs. The integration of multiple databases could still create a new set of problems in terms of development and operational costs, therefore a counter argument would be to build multiple models within a single scalable database utilizing ACID properties for synchronicity.

Martin Fowler of Thought Works champions the use of polyglot persistence in database design. He finds fault in the tendency to store various types of data in a single database and performance issues caused by doing so. Using the example of an e-commerce business he breaks down the different data types created by customers and how it is not necessarily in the company’s best interest to store it all in a traditional RDBMS. Customer transactions can be stored neatly in a RDBMS, yet purchase histories, and more importantly the purchase history of related customers, are better kept in a graph store. Not only would that increase the performance of company databases, but valuable business insight could be more easily accessible through easier data analysis. Utilizing NoSQL databases in conjunction with a traditional RDBMS allows a company to choose the right type of storage technology to best serve each aspect of the business. Fowler admits there are drawbacks to implementing such technologies on top of maintaining a company’s existing database, such as deciding exact needs and the costs associated with implementation and employee training, but it seems these problems are minimal when looking at the bigger picture. It would be wise for a company of large-scale capacity to adapt and utilize these new technologies as opposed to losing any form of completive advantage to a business rival.

These problems should not be ignored though. With many database choices comes the resulting confusion for developers. Stephen Pimentel with FoundationDB is critical of Fowler’s praise for polyglot persistence and asserts that building multiple NoSQL databases on top of a traditional RDBMS is a data modeling nightmare. He believes this would restrict the functionality of the additional databases and hinder any grains offered through operational costs. Pimentel’s answer to this problem would be to keep data stored on a single platform but add the ability to create multiple data models for specific business needs. Developers would not have to implement completely new databases to customize this modeling and these models could be synchronized throughout the organization through ACID transactions. The potential problem of users accessing stale data created by eventual consistency in NoSQL systems would be eliminated by ACID support without the need to compromise performance and scalability.

In a perfect world, businesses would be able to execute new efficient technologies with minimal expense and personnel strain. A company considering the implementation of polyglot persistence must weigh its potential benefits with the cost and reorganization that follows. For a startup seeking investment capital they could easily factor this into their budget; however an established company would have to divert more energy and risk to see the project through. Depending on the type of company and products offered, multiple modeling within an establish RDBMS database may be the best course of action to track different data types until NoSQL applications reach a comparable level of maturity.

Question Three: Consumer Trust and Transparency

In the age of big data, the relationship between companies and their consumers has become increasingly intricate. Businesses now have access to unprecedented amounts of customer data giving them a host of advantages in the marketplace. As this immense collection grows larger, the threat of customer backlash increases over privacy concerns and a fear of being exploited. Further, with user connectivity growing from multiple devices, sourcing greater amounts of data to companies, a means of consumer protection is necessary to preserve the integrity of the business-customer relationship. One method would be to place the burden of data use education on businesses themselves, while some go further and argue that there should also be a standard set of rules for each business to follow when collecting user data. Both methods hold it fundamental for customer identity to be preserved and that further integration of user control and activity in the collection process is needed. Businesses would be expected to increase the transparency of what is being collected and used to the consumer, and acknowledge the value of the data to the customer by offering a greater user experience. By not following these recommendations, companies face the threat of consumer distrust and lose the competitive advantage derived from the ability to collect massive amounts of user data to be analyzed for greater customer knowledge and insight.

Former executives of the industrial design firm frog, Theodore Forbath, Timothy Morey, and Allison Schoop believe it is in the best interest and responsibility of companies to clearly state to consumers what user data is being collected and how it will be used during the business relationship. Citing a recent survey study they performed in 2014, the three have ascertained a general idea of what types of data diverse consumers hold to be sacred and what is casually allowed to be parted with. Elements such as financial information and government documentation were more guarded whereas purchase history and demographics were not as likely to be protected by those surveyed. Further, specific companies and organizations were ranked by level of trustworthiness and those with a greater history of user protection scored higher than those with a history of collecting and using user data in secret. Examples would be consumer trust in PayPal as opposed to caution when using Facebook. A pattern observed was that when companies acknowledged the value of the data being sourced and demonstrated how it was being used to offer better products or services to customers, the level of trust in the relationship increased. Consumers were more willing to divest personal information if they knew the benefits of doing so would lead to a better user experience.

To build and maintain this trust Forbath, Morey, and Schoop offer improvements for businesses to implement. For them it begins at the beginning of the relationship. Companies must clearly educate their consumers on what user data is being collected and how it will be used. This proactive approached establishes intent and offers transparency while also minimizing the potential risk of a brand or image being tarnished from undisclosed data collection. Consumers also like to feel in control of the collection process. However tedious, if users are allowed to the view the data being exchanged they will be grateful for the transparency and feel respected as a customer. The final recommendation is for companies to offer a return comparable to the value of the data being collected. Outright payment for user data is counter-productive and may make the consumer feel further exploited. Instead rewards should be intangible, and in the form of improved services. This balances the relationship without worry of the customer feeling patronized.

Some argue that the responsibility of protecting user data should not be left up to companies alone and that consumers must have the added protection of a code of conduct to be followed by all businesses engaged in data collection. Former Information and Privacy Commissioner of Ontario, Canada, Ann Cavoukian lays out the privacy protection features of “The Seven Laws of Identity” in conjunction with an identity metasystem to securely hold private user information. These protections should be allowed to function within business rules as a basis for data collection policy. Many of Cavoukian’s suggestions match up with the recommendations mentioned earlier. Fundamental assurances of identity protection, knowledge of what is being collected and how it will be used, as well as the need for the human component all relate to trust and transparency in a business relationship. The creation of an identity metasystem to serve as a virtual wallet and house protected user data is a lofty goal but may prove cumbersome with implementation if multiple companies offer similar products. Without the backing of a public entity, uniform adoption by users may be hard to come by.

The Obama administration has also weighed in on the need for consumer protections and offered a similar list of seven principles as the basis for data collection. The notions of user control, security, transparency, and accountability should be interpreted as rights disclosed to consumers when entering a business relationship. They should be the backbone of any user agreement that allows access to customer data for peace of mind and greater consumer trust. Enforcing these provisions may prove difficult without regulatory oversight and, as observed by the former members of frog, government confidence with user data lies in the middle of the trust spectrum mentioned earlier. Despite this the seven principles laid out by both parties should offer a basis for companies to follow that facilitate a healthy business-customer relationship, and market factors may prove to be a driving force behind implementation. If a company refuses to offer transparency with the collection and protection of user data, they most certainly will lose consumer trust in the event of an information leak or the revelation of data misuse.

For a healthy business-customer relationship to thrive, companies must be open to transparency and respectful of privacy when collecting user data. These are the demands observed from consumers and policies recommended by those in government. The ability to gather knowledge and insight of consumer habits are the leading competitive advantages companies have available in the age of big data. Businesses that are willing to answer consumer demands and fears are held with greater trust then those engaged in secretive policies of data collection. Businesses should be proactive in the relationship by clearly informing the customer of their data policy and allowing open user interaction during the collection process. Companies should further demonstrate the ability to securely collect user data. By doing so, issues of consumer trust and the need for government oversight would be abated and a greater experience would be had all parties involved.