Thesis Portfolio

Cloud-Based Patient Monitoring System
(Technical Report)

Big Data and Healthcare
(STS Research Thesis)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, VA

In Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Muhammad H. Sareini
Spring, 2019
Department of Computer Science

Table of Contents

Sociotechnical Synthesis

Data is produced today at an alarming rate. Data analysis and machine learning algorithms allow new insights to be pulled from this data, that was not possible until recent years due to the sheer amount being produced. This is allowing data to be able to be leveraged in new, non-traditional markets like as healthcare. Electronic Medical Records (EMRs), Internet of Things devices, and smart-wearables (like the Apple Watch) are generating new sources of continuous data that can be used to provide a more personalized healthcare experience. However, the collection of this personal data leads to the issue of data privacy.

My STS Thesis analyzes the current state of big data healthcare, specifically current uses, future uses, and potential roadblocks. Since this area of research is relatively new, I used information about the adoption of healthcare information technology, such as EMRs, to look at the adoption of this technology though a similar lense to gain insights of what potential roadblocks might be. I use Co-Production of Science and Technology STS framework to show that the increase of big data applications will drive a need for policy reform regarding data storage and privacy. The Social Construction of Technology framework was also used to see the factors of society that is driving the adoption of such technologies.

My technical research focused on creating a prototype of a patient monitoring system for patients who have chronic rhinosinusitis. This goal of this system is to allow patients to enter information about their condition so physicians can perform data analysis. My technical research highlights the design decisions I made regarding the architecture of the application, as well as its current functionalities. Future work on the project is also addressed. Once completed, this application will be a tool physicians can use to leverage data analysis to improve their patient's quality of care and provide a more personalized healthcare experience.

My STS research analyzes how big data can be leveraged in new industries thanks to large volumes of data being produced. Specifically, I focused on the healthcare industry and tried to highlight how this technology will improve upon it. My technical research is one specific example of how big data can, and will, change the healthcare industry for the better. I created a prototype that can be iterated upon to create a final data analysis software that improves the quality of care provided to patients.

Cloud-Based Patient Monitoring System

A Technical Report submitted to the
Department of Computer Science

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, VA

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Muhammad H. Sareini
Spring, 2019

On my honor as a University Student, I have neither given nor received
unauthorized aid on this assignment as defined by the Honor Guidelines
for Thesis-Related Assignments

Signature: ____ Date:_______________

Approved: ____ Date:_______________
Marty Humphrey, Department of Computer Science

Abstract

The future of computing lies in the cloud. The cloud enables companies to quickly and securely stand up virtual machines, data storage, and numerous other managed services. No longer does someone with an idea have to purchase and maintain infrastructure before application development begins, cloud providers like Amazon (Amazon Web Services), Microsoft (Azure), and Google (Google Cloud Platform) allow developers to "rent" virtual machines. Not only has the way we compute changed, but the sheer amount of data we have access to is increasing at enormous rates.

The healthcare industry has been doing well in the recent years by increasing Healthcare Information Technology (HIT) use in forms like Electronic Medical Records. The goal of my technical capstone is to develop a cloud-based patient monitoring System that can be used in hospitals and primary-care offices. This system is a web application that is hosted on AWS so it is available on any device that has access to the internet (computer, mobile, tablet, etc.). To take full advantage of what the cloud has to offer, it was built using a serverless architecture, and developed using the Angular web framework to take advantage of JavaScript's client-side computing. The serverless architecture, combined with the angular framework, enabled me to host the application in an AWS S3 storage bucket for an completely serverless architecture. This system provides an electronic way for doctors and patients to monitor progress, enter new data, and compare against past data and other patients.

Chronic Rhinosinusitis (CRS)

CRS is a chronic disease in which the nasal passages become inflamed and congested for a long period of time. This disease interferes with the natural drainage and causes mucus buildup, and causes pain and pressure (Mayo Clinic, 2018). Currently, if a physician wishes to track the patient’s progress post-treatment, he or she must have the patient fill out a form in office. If any form of data analysis is to be performed on the data (e.g. comparing this patient’s progress to the rest of the patients the physician has treated), the data needs to manually be entered and analyzed.

Inputting the amount of data needed for analysis is not feasible, and, my technical capstone addresses this problem. By having a cloud based system, patients are able to to fill out information on any device that has access to the internet, as well as allow the physician to input the data into the same system during an in-person visit. Once data is inputted into the system, it will all be in one place and ready for data analysis that the physician wishes to perform on a patient's data and could even allow the patient themself to see how their progress compares against other anonymous patients. My project’s goal is to begin on working on a prototype of a web application that allows patients to input their data, and storing the data in a MySQL database all hosted on AWS.

Serverless Architecture (AWS)

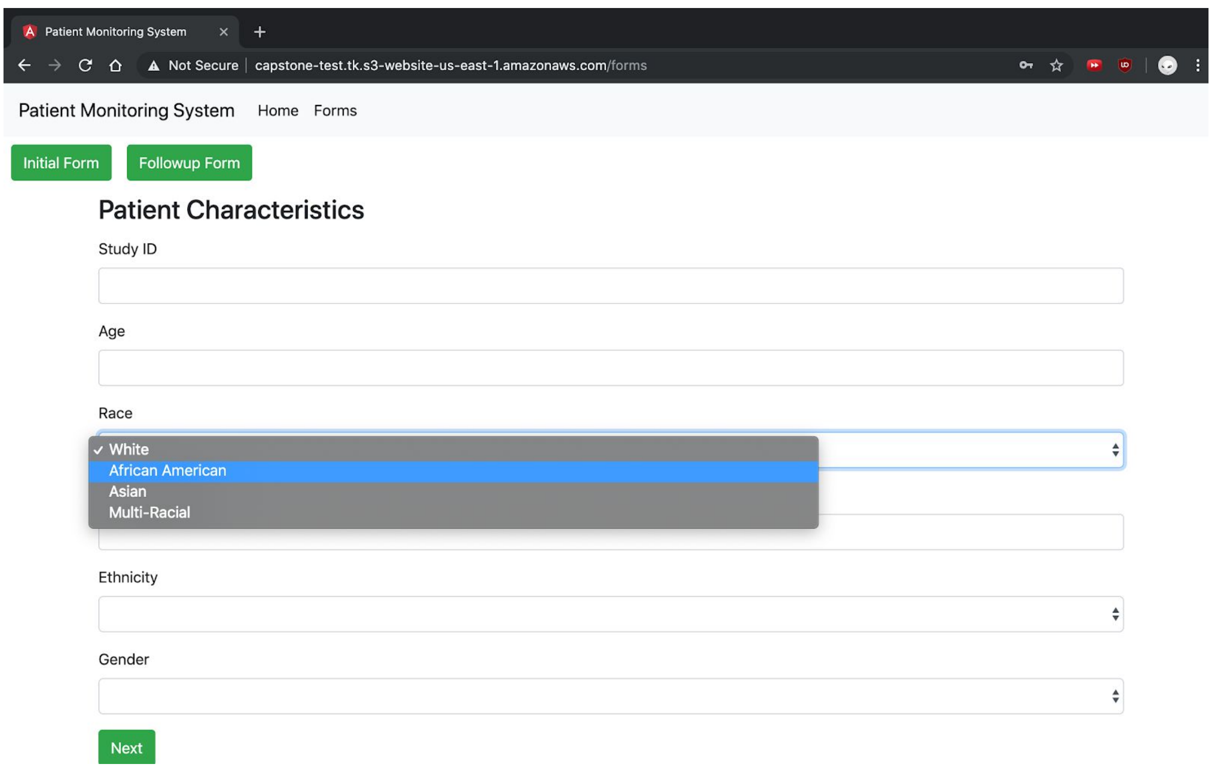
A serverless architecture is one in which an application is not running on the traditional web server, sitting there waiting for web traffic. Instead, application developers make use of managed cloud services (such as AWS lambda functions or Google Cloud Functions) which are stateless, event triggered, and fully managed by the cloud provider that is hosting this service. These serverless options prevent developers from having to worry about scaling and managing costs of infrastructure, just application development (Fowler, M.). With serverless services, developers are not charged for uptime of an application but instead are charged per invocation of a service, allowing developers to pay for only what they use, saving cost at early stages as well as at scale.

My application used web API requests to Amazon's Lambda Functions to store and retrieve data from the database. AWS Lambda Functions are a service that runs code in response to events, and manages the underlying computing resources with regards to scaling, fault tolerance, and many other metrics ("AWS Lambda Features"). For my application, there exists an individual lambda function for each API call interacting with the database. To make these lambda functions accessible from a web application, each was accessed through AWS API Gateway, which allow for a restful web API to have an AWS lambda function as a backend endpoint. Upon invocation of each API, it would run the associated lambda function. The lambda function would then connect to the database and then run queries using the PyMySQL package, returning JSON responses back to the web application.

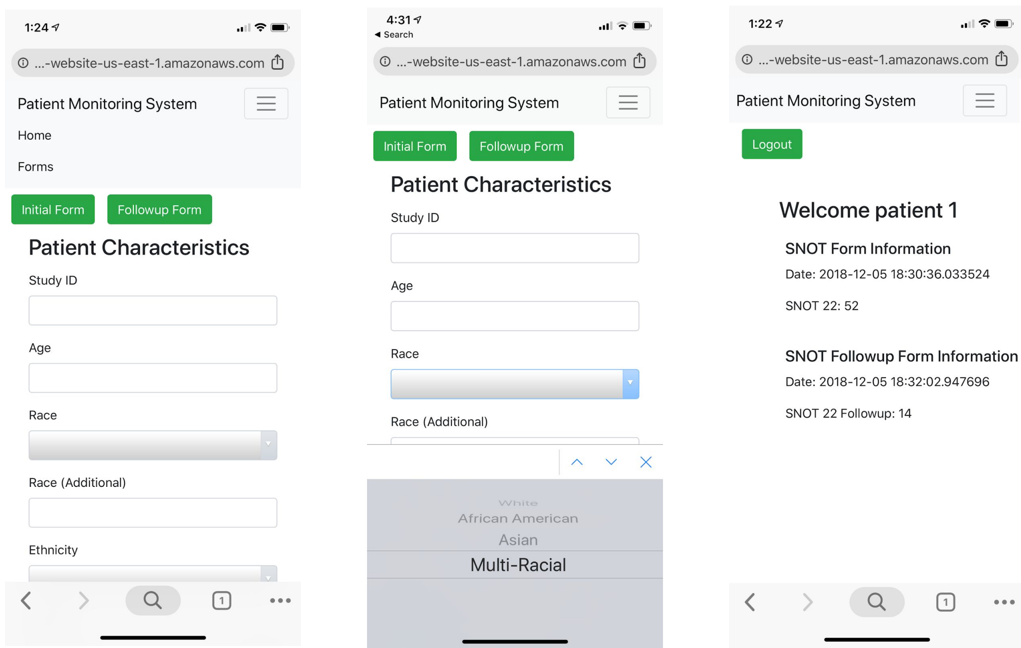
For the database, the application used AWS Relational Database Service (RDS) to store the patient account information and form responses. The database was a db.t2.micro running a MySQL engine. The database has three tables: a table that contains the user's ID and information, a table that contains the user's SNOT form response, and the final table contains the user's SNOT follow up form responses. To associate patients and their SNOT form and follow up form, both of the tables contain the ID key.

Angular Web Framework

Angular is an open-source JavaScript web framework that was originally developed by Google. Angular applications are written in TypeScript, which is a superset of JavaScript, which allows developers to not have to learn a whole new language (Bodrov-Krukowski, 2018). Angular is also a complete web framework, allowing for the reduction of development time by using out-of-the-box functionalities that are baked into the framework. One such functionality is that it is mobile and desktop ready (Bodrov-Krukowski, 2018), which will allow the users of the system to view the application on mobile devices without developers having to write any additional code.



Patient Monitoring System on Google Chrome - MacOS



Patient Monitoring System on Google Chrome - iOS

As mentioned before, TypeScript (and JavaScript) are client-side languages that run in the user's browser, which allow the browser code to trigger the application's serverless functions, without needing a server for the web application itself. Web applications developed with the Angular framework are Single Page Applications (SPA) . SPAs re-render the content of the website without making a request to the server to fetch new HTML (Sherman, 2018). This allows the web application to be hosted statically in an AWS S3 bucket, and all the JavaScript will be pulled into the browser with one single request, allowing no web server to be necessary to host the main web application. So when a patient goes to the website, all of the HTML and CSS is pulled down at one time, and then stored and ran in the browser allowing each patient to only make one request, and then go about using the web application.

Future Work

Since the goal of this project was to make a working prototype, the web application has a single, very small database. In the future, the idea is to have the application use an AWS Serverless Aurora database cluster for high availability and reliability. AWS Serverless Aurora, much like Lambda Functions, are fully managed by Amazon. However, instead of providing serverless computing, this managed service provides serverless database clusters that scale up and down automatically in response to application usage, which means that Amazon allocates more and less resources for the developer automatically based on current traffic. This option allows for developers to pay on a per-second basis for the database capacity in use ("Amazon Aurora Serverless," 2018). Like most other managed serverless resources, when the database is not receiving traffic, no charges are accrued.

Currently, the application is using my own user database table to login and register users. In the future, I would like to integrate AWS Cognito to handle all user login and registering. AWS Cognito is a managed service can scale to millions of users, and allows for managing access in applications as well as sign in with social identity providers like Google and Facebook ("Amazon Cognito"). Having this integrated into the application would allow users to not have to worry about creating new accounts, remembering new passwords, and would save development time by taking advantage of the features the service already provides, like email confirmations for account creation and forgetting passwords.

The current functionality of the application just allows for patients to log in and fill out the forms, and updates their responses in a database. For the full application to be complete, there needs to be a way for physicians to login as well and access information for the patients that are under their care. With this information, there ideally would be a way for the physician to perform some form of statistical analysis if they wish. The finished application would also let the patients see how their responses compare against others.

Conclusion

The patient monitoring system is a serverless web application that is hosted on Amazon Web Services. Currently, it allows for patients to login, register, and complete forms tracking their progress throughout treatment for chronic rhinosinusitis. For the application to be complete, physicians needs to be able to log in and keep track of their patients and patients need to see how they are recovering compared to other patients. By leveraging the cloud in this application, it will provide a highly scalable solution in which all of the infrastructure will be managed by AWS. This allows for physicians to focus on using the application to deliver a better quality of healthcare to their patients, and not have to worry that integrating their application with existing hospital technology or hiring someone to maintain their infrastructure.

References

Amazon Aurora – Auto-Scaling Serverless Database Service - AWS. (2018). Retrieved from https://aws.amazon.com/rds/aurora/serverless/

Amazon Cognito - Simple and Secure User Sign Up & Sign In | Amazon Web Services (AWS). (n.d.). Retrieved from https://aws.amazon.com/cognito/

AWS Lambda – Product Features. (n.d.). Retrieved from https://aws.amazon.com/lambda/features/

Bodrov-Krukowski, I. (2018, April 23). Angular Introduction: What It Is, and Why You Should Use It. Retrieved from https://www.sitepoint.com/angular-introduction/

Chronic sinusitis. (2018, March 03). Retrieved from https://www.mayoclinic.org/diseases-conditions/chronic-sinusitis/symptoms-causes/syc-20351661

Serverless Architectures. (n.d.). Retrieved from https://martinfowler.com/articles/serverless.html

Sherman, P. (2018, April 11). How Single-Page Applications Work – Paul Sherman – Medium. Retrieved from https://medium.com/@pshrmn/demystifying-single-page-applications-3068d0555d46

Big Data and Healthcare

A Research Paper submitted to the
Department of the School of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, VA

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Muhammad H. Sareini
Spring, 2019

On my honor as a University Student, I have neither given nor received
unauthorized aid on this assignment as defined by the Honor Guidelines
for Thesis-Related Assignments

Signature: ____ Date:________________

Approved: ____ Date:________________
S. Travis Elliott, Department of Engineering and Society

Abstract

Everyday, humans are producing more and more data. This allows for data to be used in fields that traditionally are not thought to benefit from large-scale data analysis, like the healthcare. If data is collected on patents using existing Electronic Medical Records (EMR's) and new sources of data, patients can receive healthcare that is personalized to their needs and individual cases. With this new application of patient data, there are ethical concerns to consider as well. I use the Social Construction of Technology framework to analyze some current data privacy laws, and try to figure out where the weak points are that are hiring these types of technologies. I also used the Co-production of Science and Social Order framework to analyze how exactly these data driven healthcare applications can improve the quality of personalized healthcare. At the end of my analysis I concluded that the data regulations are not sufficient in their current state to allow for the rapid growth of these new technologies, and the implications of these technologies, I believe, can greatly improve the quality of healthcare that is currently provided.

Big Data and Healthcare

A Data Driven Society

With the rise of technology and amount of information being produced daily, data is being put to use to better our everyday lives. The data can be analyzed to produce insight that is otherwise invisible to a human by using machine learning and statistical models. The existence of these big data analysis techniques is starting to spread into the healthcare space. The goal of my STS thesis is to analyze the current status of big data in healthcare such as roadblocks, current uses, and future uses. I will use the Co-production of Science and Social Order STS framework to show that big data has been creating a need for policy reform and regulations to be set in place. With the increasing amount of data breaches, data will need regulations that make them feel their data is safe so these healthcare technologies can really take off. I will also use the Social Construction of Technology framework to analyze what factors of modern society drove the adoption, and resistance, of big data techniques. Technological factors like smartphones (as well as other "smart" devices), social media, and Internet of Things (IoT) devices will be analyzed to see if, and how, the existence of this data can be used to increase the quality of healthcare received.

Healthcare Information Technology

Healthcare Information Technology (HIT) refers to the electronic systems that physicians and patients use to store, share, and analyze health-related information. Such systems include

Electronic Health Records (EHRs), Personal Health Records (PHRs), and Electronic Prescribing (E-Prescribing). HIT systems also increase the security and privacy of a patient's data. For example, data can be encrypted so only certain people can view a document, and those who do can be tracked. In contrast, a physical copy which can be accessed by anyone near the copy, and there is no way to tell who read it or when. These HIT solutions allow patients and physicians to focus on health instead of the hassles of managing healthcare information by providing faster and more accurate prescriptions, rapid information sharing, reduced paperwork, and reduction in unnecessary tests ("Health IT: Advancing America's Health Care").

In one study (Chaudhry et al., 2006), the digitalization of records and other patient information allowed for improved clinical care through large-scale screening and the aggregation of data in ways that was not feasible when the information was stored using the traditional paper based system. One study reported that the institutions were using an electronic health record to identify adverse health events, and design plans that would prevent such health events in the future. Identification of such events increased by 2.34%, and the adverse drug events themselves decreased by from 7.6% to 2.2% (Chaudhry et al., 2006). This is just one example of how HIT can increase the quality of care patients receive.

Big Data

Big data consists of large amounts of ever growing-complex data coming from new sources. So much data, in fact, that new ways to analyze this amount of data had to be created. And thanks to this large amount of new data, insight can be drawn from such large amounts of information ("Oracle Big Data"). In the healthcare industry, there is enormous amounts of data that, if analyzed, could lead to increased quality of care for patients and lead to new innovations that detect, prevent, and predict diseases and illness.

Electronic Health Records (EHRs) are one area where big data could really revolutionize the medical field. Mining data from EHRs can be used to increase clinical knowledge and be beneficial in supporting clinical research (Andreu-Perez et al., 2015). Mining data from EHRs, to use for analysis, has proven to be effective on a wide range of challenges like: building models for predicting health risk assessments, enhancing knowledge about survival rates, and pharmacovigilance (the practice of monitoring the effects of medical drugs after they have been licensed for use) (Andreu-Perez et al., 2015).

Another area where big data analysis could improve healthcare is with chronic diseases. For example, some chronic diseases are identifiable by acute events that are very hard to predict solely on data collected during a doctor's visit (Andreu-Perez et al., 2015). There is a rare disease called thoracic aortic dissection which affects between 3-4 per 100,000 people per year. Individuals with connective tissue disorders are more susceptible to aortic aneurysms or tears. By using large-scale population screening for such diseases, at-risk patients could be identified and notified that they could be at-risk for such a rare disease, that otherwise the patient would be unaware of (Andreu-Perez et al., 2015).

Social Health

In 2019, there will be an estimated 2.77 billion social media users around the globe. (“Number of social media users worldwide 2010-2011”). These social media networks, such as Twitter and Facebook, can be a source to obtain data which when analyzed will allow for an improvement in healthcare. For example, around one-fourth of patients with chronic diseases (ex. diabetes, cancer) use social networks to share experiences. One example would be support groups. These posts can be used as one new source of data. Not only the biological information, but the geolocation of applications provide additional ways to understand behaviors, social demographics, and behaviors instead of expensive statistical sampling (Andreu-Perez et al., 2015).

Social networks can also be used to study social factors that influence unhealthy habits, such as smoking. For example, it was discovered that smokers with higher levels of education had a higher influence over their peers compared to less educated smokers (Andreu-Perez et al., 2015). Emotional levels in social media posts on platforms like Twitter can be used to gather information on whole geographical regions. Tweets, for example, that show hostility and stress can serve as predictors of heart disease mortality for a geographical area (Andreu-Perez et al., 2015).

Internet of Things and Mobile Computing

Mobile devices, such as smartphones and smart watches, have increased in popularity in the recent years. Not only has the popularity increased, but the computing power on these devices have as well. Today, a smartphone has more computing power than the entirety of NASA when it first put a man on the moon (Khoso, 2016).

Mobile phones generate data that is highly descriptive and continuous (Andreu-Perez et al., 2015). For example, video cameras can collect data on heart rate and heart rate variability (Andreu-Perez et al., 2015). The new Apple Watch 4 has EKG capability, meaning that it will be able to collect real-time heartbeat data of the wearer, which are stored in the Health Application and can be distributed to healthcare professionals ("New watch can help doctors monitor your heart in real time"). Now, there is even a specific ECG application that has the ability to notify the wearer of an irregular heart rhythm through a watch notification, like any other app notification would work. The watch checks the heart rhythm in the background, and if five abnormal rhythms are detected over a minimum of 65 minutes the ECG app sends out a notification ("ECG app and irregular heart rhythm notification available today on Apple Watch"). This new technology provides a new form of continuous data on heart rhythms that were not possible to collect before. When connected to health providers, a greater level of personalization and responsiveness to healthcare can be applied (Andreu-Perez et al., 2015). This continuous data calls for new forms of analysis and technologies, which will be discussed later in the paper.

Roadblocks to Implementation

Smaller healthcare institutions face roadblocks that are much more difficult for them to overcome when compared to institutions of larger size. For one, these smaller institutions lack funding required for the upfront cost of implementing these systems, and the compounding costs of maintenance. The reality of the fact is that these practices are small, independent businesses that are having a declining income (Poon et al., 2006). One small practice reported that their budget for IT for a whole year was only $1,200 (Poon et al., 2006). Not only is there a lack of funding, but the rate of unsuccessful deployments are much higher than successful implementations. Huge associated risk and lack of funding make it impossible for small institutions to implement HIT systems into their environment, and they are not very lucrative. It really does not make sense to front a large sum of money for something that most likely will fail before it is even up and running.

Age seems to play a factor as well in the acceptance of HIT systems. In the study done by Bhattacherjee & Hikmet (2007), it was found that younger physicians tended to be more open to the acceptance of a Computerized Physician Order Entry (CPOE) system that was being tested at a facility, and the reasoning is believed to be some form of exposure to similar systems during schooling and general technology exposure growing up. On the other side, there was heavy resistance from the older physicians who did not have the same exposure recently, if ever. One younger doctor in particular loved the system, and would regularly log into the system from at home. He claimed that the system would help him optimize his daily rounds while also saving time on processing paperwork, which in turn allowed him to spend more time interacting with his patients (Bhattacherjee & Hikmet, 2007). As much as this one physician loved the system, there were plenty of those who despised it. Instead of learning the system and adapting, they would create intricate ways to avoid the system like smuggling paperwork. Some physicians would go as far as throwing tantrums in protest. In total, only 25% of the physicians used the systems for around 50% of their orders (Bhattacherjee & Hikmet, 2007).

One specific challenge of IoT implementation in healthcare technology is communication. Many of these devices have their own sensors, but many of these devices have their own "languages" for talking to a server (Dimitrov, D. V, 2016). Getting legacy systems to communicate with newly developed big data systems will be a huge issue as well. Health data is largely fragmented across institutions. Being able to communicate between these institutions will be a huge challenge, and would not solve the bigger problem of the data being fragmented across many institutions (Dimitrov, 2016). For data to be useful, it needs to be collected and stored in some uniform structure so that analysis can be properly performed.

These are just some examples of roadblocks that caused the delayed, and painful, acceptance of HIT systems such as EHRs. One could imagine that the acceptance of new forms of HIT, like those that are driven by big data, will face the same types of scrutiny and push back. Not only will these new forms of HIT have to face the previous issues, but new ones as well with an emphasis on issues of privacy and data collection itself. These privacy concerns will be discussed later in the paper.

Science Technology and Society Analysis

Privacy

As mentioned earlier, the use of big data techniques in the healthcare space requires large amounts of data. The data needs to be stored somewhere for useful analysis to take place. Not only does the data need to be stored somewhere, but it cannot be anonymous since links back to the patient must exist so analysis result can be shared with physicians and the patients. Looking at this ethical dilemma using the Co-production of Science and Social Order STS framework, combined with current issues of data privacy, it can be seen that big data and its applications are causing the need for new laws and regulations to be in place to protect people's privacy and hold those who possess data responsible for unforeseen circumstances, such as data breaches.

"Half of all Americans believe their personal information is less secure now than it was five years ago" ("Reforming the U.S. Approach to Data Protection and Privacy"). This makes sense, considering the Equifax breach of 2017 as well as the many other data breaches (Yahoo, Uber, etc.) that have been occurring recently. The United States government has not approached data privacy with a single federal law that regulates the collection and use of personal data. The current approach being taken is a "patchwork" approach by only regulating certain sectors such as health and financial information and creating small fixes without addressing the larger issue at hand ("Reforming the U.S. Approach to Data Protection and Privacy"). At the state level of regulation, the issue is especially prevalent. Currently, all 50 states have enacted laws that require government entities to notify individuals when their personal and or identifiable data has been compromised in a data breach ("SECURITY BREACH NOTIFICATION LAWS"). However, not all state laws are enforced in the same way. It is up to the discretion of each state to decide the time frame that individuals need to be notified by, and even if law enforcement must be notified first ("SECURITY BREACH NOTIFICATION LAWS").

As just described, a privacy issue is at large when people's personal data is in question. The current laws and regulation in place are not robust enough for the public to trust companies with their data, let alone data that is as personal as their health data. The rise of large amounts of data being collected has caused states to implement their own "fixes" that do not always get right to the source of the underlying problem. For big data to be truly accepted in healthcare, these privacy issues need to be addressed so patients feel safe that their data will not be used for any other purposes except to give them a better, more personalized healthcare experience. As these big data techniques are adopted and gain traction, I believe that the forces will cause governing bodies to push for the necessary changes in regulation.

New Technologies

Big data has many uses in healthcare, as was described earlier. For all the data produced to be useful new technologies and forms of analysis are being, and need to be, created as predicted by the Social Construction of Science and Technology (SCOT) STS framework. The creation of big data technologies happened over time, and the need for such technologies have existed even longer.

In 1663, John Graunt recorded and examined information about London's mortality rate to gain an understanding of the Bubonic Plague ("A Short History Of Big Data"). With the rise of machines that can perform computations for people, these tasks became more and more common. In 1937, Franklin D. Roosevelt passed the Social Security Act and the government had to keep around 29 million records, and IBM developed punch card machines for this very task ("A Short History Of Big Data"). Over time, these computing machines developed into the computers we have today. In 2005, Hadoop was developed by Yahoo and is based on Google's MapReduce. The goal of MapReduce was to index the internet, and now Yahoo's Hadoop implementation is still used to this day to process large amounts of data in parallel ("A Short History Of Big Data"). The more data that humans are producing, the more new technologies are developed to use this data.

By 2020 there will be an estimated 6 billion smartphones in the world (Khoso, 2016). In the recent years, there has been research into crowd sourcing computing power from idle smartphones. This is a tremendous amount of idle computing power than can be put to use to analyze big data. Only 1% of data ever produced is analyzed even though 37% would be useful to analyze (Khoso, 2016). Using all the idle smartphones in the world, they can be meshed into distributed computing network to analyze data during times the phone is idle, like when its owner is asleep at night. To match the computing power of one super computer, there needs to be only 1 million smartphones in such a network (Khoso, 2016). In 2014, HTC started the HTC Power to Give Initiative which allows users to download an app that gives access to your phones processor while you sleep to a research grid (Khoso, 2016). There are many other examples of such networks, some others are the BIONIC network at Berkeley and IBM's world community grid. These distributed technologies were created to solve a problem generated by society: how do we analyze all of the new data every day when we have limited supercomputing power? The need of the society drove the development of this distributed network of smart phones to generate supercomputing power.

For big data to really take off in the healthcare field, new forms of technology, or new combinations of existing technologies, are going to be created to securely and effectively handle patient data and allow its analysis. For example, there have been suggestions of using wearable and implantable sensors. Implantable sensors provide a way of capturing critical events while still providing a continuous stream of health information (Andreu-Perez et al., 2015). Having smart devices that provide doctors with continuous health data provide a new set of problems that will need to be solved. Current clinical practices define hypertension with measurements taken sporadically during various hospital visits (Andreu-Perez et al., 2015). Physicians will have more data than they know what to do with, and will need to develop new analysis to be performed to extract the relevant data from a continuous stream to gain insight that will create a more personalized health care.

Conclusion

The healthcare field is the perfect application of big data. Humans are producing more and more data each year, and it can be put to good use. Data coming from sources like social media, your smart phone, and wearables like the Apple Watch collect continuous streams of data that, when collected and analyzed, can provide a more personal healthcare experience than is currently offered. But with the collection of all this data, issues become clear regarding data privacy and the development of new technologies.

With data being as valuable as it has become, it makes sense that companies that store a lot of data on their users are under constant attack by cyber criminals. This health data would be no different, and there needs to be policy in place that ensures data is being collected, stored, and protected the right way. While confidential health data is protected under HIPAA, there is nothing protecting the data that is produced by your smartwatch or smart phone. Ethical issues arise: what will companies do with this data? Sell it? Only pass it on to your healthcare provider?

This era of big data is driving the development of new technologies. Other fields have adapted and created new ways to analyze and store their data, so now it is healthcare's turn. While this is a big task, healthcare that is tailored to the patient will increase the quality of care patients receive, just like the introduction of technology did in the first place.

References

A Short History Of Big Data. (n.d.). Retrieved from https://datafloq.com/read/big-data-history/239

Andreu-Perez, J., C, C., Poon, Y., Merrifield, R. D., Wong, S. T., & Yang, G. (2015). Big Data for Health. *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS,19*(4). Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7154395.

Bhattacherjee, A., & Hikmet, N. (2007). Physicians resistance toward healthcare information technology: A theoretical model and empirical test. *European Journal of Information Systems,16*(6), 725-737. doi:10.1057/palgrave.ejis.3000717

Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., . . . Shekelle, P. G. (2006). Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care. *Annals of Internal Medicine,144*(10), 742. doi:10.7326/0003-4819-144-10-200605160-00125

Dimitrov, D. V. (2016). Medical Internet of Things and Big Data in Healthcare. Healthcare Informatics Research. doi:10.4258/hir.2016.22.3.156

ECG app and irregular heart rhythm notification available today on Apple Watch. (2019, February 20). Retrieved from https://www.apple.com/newsroom/2018/12/ecg-app-and-irregular-heart-rhythm-notification-available-today-on-apple-watch/

Health IT: Advancing America's Health Care(Rep.). (n.d.). The Office of the National Coordinator for Health Information Technology.

Khoso, M. (2016, June 13). Smartphones: A Supercomputer in Your Pocket. Retrieved from https://www.northeastern.edu/levelblog/2016/04/21/smartphones-supercomputer-in-your-pocket/

New watch can help doctors monitor your heart in real time. (n.d.). Retrieved from https://www.heart.org/en/news/2018/09/12/new-watch-can-help-doctors-monitor-your-heart-in-real-time

Number of social media users worldwide 2010-2021. (n.d.). Retrieved from https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

Oracle Big Data. (n.d.). Retrieved from https://www.oracle.com/big-data/guide/what-is-big-data.html

Poon, E. G., Jha, A. K., Christino, M., Honour, M. M., Fernandopulle, R., Middleton, B., . . . Kaushal, R. (2006). Assessing the level of healthcare information technology adoption in the United States: A snapshot. *BMC Medical Informatics and Decision Making,6*(1). doi:10.1186/1472-6947-6-1

Reforming the U.S. Approach to Data Protection and Privacy. (n.d.). Retrieved from https://www.cfr.org/report/reforming-us-approach-data-protection

SECURITY BREACH NOTIFICATION LAWS. (2018, September 29). Retrieved from http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx

Prospectus

Cloud-Based Patient Monitoring System
(Technical Report)

Big Data and Healthcare
(STS Research Thesis)

A Thesis Prospectus Submitted to the
Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, VA

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Muhammad H. Sareini
Fall, 2018
Department of Computer Science

On my honor as a University Student, I have neither given nor received
unauthorized aid on this assignment as defined by the Honor Guidelines
for Thesis-Related Assignments

Signature: ____ Date:________________

Approved: ____ Date:________________
Marty Humphrey, Department of Computer Science

Approved: ____ Date:________________
S. Travis Elliott, Department of Engineering and Society

I. Introduction

The future of computing lies in the cloud. The cloud allows for companies to quickly and securely stand up virtual machines, data storage, and numerous other managed services. No longer does someone with an idea have to purchase and maintain infrastructure before application development beings, cloud providers like Amazon (Amazon Web Services), Microsoft (Azure), and Google (Google Cloud Platforms) allow developers to "rent" virtual machines. Not only has the way we compute changed, but the sheer amount of data we have access to is increasing at enormous rates.

The healthcare industry has been doing well in the recent years by increasing Healthcare Information Technology (HIT) use in forms like Electronic Medical Records. The goal of my technical capstone is to develop a Cloud-Based Patient Monitoring System that can be used in hospitals and primary-care offices. This system will be a web application that is hosted on AWS so it can be accessed on any device that has access to the internet (computer, mobile, tablet, etc.). To take full advantage of what the cloud has to offer, it will be built using a serverless architecture, and developed using the Angular web framework to take advantage of JavaScript's client-side computing. The serverless architecture, combined with the angular framework, will allow me to host the application in an AWS S3 storage bucket for a truly serverless architecture. Once built, this system will provide an electronic way for doctors and patients to monitor progress, enter new data, and compare against past data and other patients.

My socio-technical research project aims to analyze the adoption of Healthcare Information Technology (HIT) as a whole, more specifically the use of big data in healthcare.

Systems, like the one I aim to build for my technical capstone, allow physicians to provide a new level of care that is not possible without such systems in place, since there is currently no way to analyze the data the way it is being stored. I will research and analyze the past trends of healthcare information technology (EMR's etc.) and new ways big data can be used to improve the healthcare industry. I will use STS methodologies to analyze this trend, and gain new insight on the driving factors that led physicians to want to implement big data solutions.

II. Technical Thesis

Chronic Rhinosinusitis (CRS)

CRS is a chronic disease in which the nasal passages become inflamed and congested for a long period of time. This disease interferes with the natural drainage and causes mucus buildup, and causes pain and pressure (Mayo Clinic, 2018). Currently, if a physician wishes to track the patient's progress post-treatment, he or she must have the patient fill out a form in office. If any form of data analysis is to be performed on the data (e.g. comparing this patient's progress to the rest of the patients the physician has treated), the data needs to manually be entered and analyzed.

Inputting the amount of data needed for analysis is not feasible, and, my technical capstone addresses this problem. By having a cloud based system, patients will be able to to fill out data on any device that has access to the internet, as well as allow the physician to input the data into the same system during an in-person visit. Once data is inputted into the system, it will all be in one place and ready for any form of data analysis that the physician wishes to perform on a patient's data, and would even allow the patient themself to see how their progress stacks up against other anonymous patients.

Serverless Architecture (AWS)

A serverless architecture is one in which an application is not running on the traditional web server, sitting there waiting for web traffic. Instead, application developers make use of managed cloud services (such as AWS lambda functions or Google Cloud Functions) which are stateless, event triggered, and fully managed by the cloud provider that is hosting this service. These serverless options prevent developers from having to worry about scaling and managing costs of infrastructure, just application development (Fowler, M.). With serverless services, developers are not charged for uptime of an application but you are charged per invocation of a service, allowing developers to pay for only what they use, saving cost at early stages as well as at scale.

My application will make use of Amazon's managed Lambda Functions and Serverless Aurora Database. AWS Lambda Functions is a service that runs code in response to events, and manages the underlying computing resources with regards to scaling, fault tolerance, and many other metrics ("AWS Lambda Features," 2018). These cloud functions will be used for all computations the application, from logging in and registering to entering patient checkup information into a database.

AWS Serverless Aurora, much like Lambda Functions, are fully managed by Amazon. However, instead of providing serverless computing, this managed service provides serverless database clusters that scale up and down automatically in response to application usage, which means that Amazon allocates more and less resources for the developer automatically based on current traffic. This option allows for developers to pay on a per-second basis for the database capacity in use ("Amazon Aurora Serverless," 2018). Like most other managed serverless resources, when the database is not receiving traffic, no charges are accrued.

Angular Web Framework

Angular is an open-source JavaScript web framework that was originally developed at Google. Angular applications are written in TypeScript, which is a superset of JavaScript, allowing developers to not have to learn a whole new language (Bodrov-Krukowski, 2018). Angular is also a complete web framework, allowing for the reduction of development time by using out-of-the-box functionalities that are baked into the framework. One such functionality is that it is mobile and desktop ready (Bodrov-Krukowski, 2018), which will allow the users of the system to view the application on mobile devices without developers having to write any additional code.

As mentioned before, TypeScript (and JavaScript) are client-side languages that will run in the user's browser, which will allow the browser code to trigger the application's serverless functions, without needing a server for the web application itself. Web applications developed with the Angular framework are Single Page Applications (SPA) . SPAs re-render the content of the website without making a request to the server to fetch new HTML (Sherman, 2018). This allows the web application to be hosted statically in an AWS S3 bucket, and all the JavaScript will be pulled into the browser with one single request, allowing no web server to be necessary to host the main web application.

Goals and Methods

The goal of my capstone project is to get a working prototype for the serverless web application as described before which will be hosted exclusively on AWS. AWS is HIPAA compliant, so all patient data will be as secure as any digital information can be. Once a working prototype of the basic functionalities of the system in in place, the goal is to apply machine learning algorithms to the data to analyze patient data and predict the outcome of their recovery from treatment.

III. STS Thesis

Big Data and Healthcare

With the rise of technology and amount of information being produced daily, data is being put to use to better our everyday lives. The data can be analyzed to produce insight that is otherwise invisible to a human by using machine learning and statistical models. The rise of these big data analysis techniques is starting to spread into the healthcare space. The goal of my STS thesis is to analyze the current status of big data in healthcare such as roadblocks, current uses, and future uses. I will use the Actor Network Theory to analyze how data, physicians, and patients interact with each other to drive the development and the use of big data in the healthcare industry. Through these relationships, I hope to gain insights on which actors in the network is responsible for pushing adoption, and which are responsible for the pushback. I will also use the Social Construction of Technology framework to analyze what factors of modern society drove the adoption of big data techniques. Technological factors like smartphones (as well as other "smart" devices), social media, and Internet of Things (IoT) devices will be analyzed to see if, and how, the existence of this data can be used to increase the quality of healthcare received.

Healthcare Information Technology

Healthcare Information Technology (HIT) refers to the electronic systems that physicians and patients use to store, share, and analyze health-related information ("Health IT: Advancing America's Health Care"). Such systems include Electronic Health Records (EHRs), Personal Health Records (PHRs), and Electronic Prescribing (E-Prescribing). HIT systems also increase the security and privacy of a patient's data. For example, data can be encrypted so only certain people can view a document, and those who do can be tracked. In contrast, a physical copy who can be accessed by anyone near the copy, and there is no way to tell who read it and when ("Health IT: Advancing America's Health Care"). These HIT solutions allow patients and physicians to focus on health instead of the hassles of managing healthcare information by providing faster and more accurate prescriptions, rapid information sharing, reduced paperwork, and reduction in unnecessary tests ("Health IT: Advancing America's Health Care").

In one study (Chaudhry et al., 2006), the digitalization of records and other patient information allowed for improved clinical care through large-scale screening and the aggregation of data in ways that was not feasible when the information was stored using the traditional paper based system. One study reported that the institutions were using an electronic health record to identify adverse health events, and design plans that would prevent them in the future. Identification of such events increased by 2.34%, and the adverse drug events themselves decreased by from 7.6% to 2.2% (Chaudhry et al., 2006). This is just one example of how HIT can increase the quality of care patients receive.

Roadblocks to Implementation

Smaller healthcare institutions face roadblocks that are much more difficult for them to overcome when compared to institutions of larger size. For one, these smaller institutions lack funding required for the upfront cost of implementing these systems, and the compounding costs of maintenance. The reality of the fact is that these practices are small, independent businesses that are having a declining income (Poon et al., 2006). One small practice reported that their budget for IT for a whole year was only $1,200 (Poon et al., 2006). Not only is there a lack of funding, but the rate of unsuccessful deployments are much higher than successful implementations. Huge associated risk and lack of funding make it impossible for small institutions to implement HIT systems into their environment, and they are not very lucrative. It really does not make sense to front a large sum of money for something that most likely will fail before it is even up and running.

Age seems to play a factor as well in the acceptance of HIT systems. In the study done by Bhattacherjee & Hikmet (2007), it was found that younger physicians tended to be more open to the acceptance of a Computerized Physician Order Entry (CPOE) system that was being tested at a facility, and the reasoning is believed to be some form of exposure to similar systems during schooling and general technology exposure growing up. On the other side, there was heavy resistance from the older physicians who did not have the same exposure recently, if ever. One younger doctor in particular loved the system, and would regularly log into the system from at home. He claimed that the system would help him optimize his daily rounds while also saving time on processing paperwork, which in turn allowed him to spend more time interacting with his patients (Bhattacherjee & Hikmet, 2007). As much as this one physician loved the system, there were plenty of those who despised it. Instead of learning the system and adapting, they would create intricate ways to avoid the system like smuggling paperwork. Some physicians would go as far as throwing tantrums in protest. In total, only 25% of the physicians used the systems for around 50% of their orders (Bhattacherjee & Hikmet, 2007).

Big Data

Big data consists of large amounts of ever growing complex data coming from new sources. So much data in fact, that new ways to analyze this amount of data had to be created. And thanks to this large amount of new data, insight can be drawn from such large amounts of information ("Oracle Big Data"). In the healthcare industry, there is enormous amounts of data that, if analyzed, could lead to increased quality for patients and lead to new innovations to detect, prevent, and predict diseases and illness.

Electronic Health Records (EHRs) are one area where big data could really revolutionize the medical field. Mining data from EHRs can be used to increase clinical knowledge and be beneficial in supporting clinical research (Andreu-Perez et al., 2015). Mining data from EHRs, to use for analysis, has proven to be effective on a wide range of challenges like: building models for predicting health risk assessments, enhancing knowledge about survival rates, and pharmacovigilance (the practice of monitoring the effects of medical drugs after they have been licensed for use) (Andreu-Perez et al., 2015).

Another area where big data analysis could improve healthcare is with chronic diseases. For example, some chronic diseases are identifiable by acute events that are very hard to predict solely on data collected during a doctor's visit (Andreu-Perez et al., 2015). There is a rare disease called thoracic aortic dissection which affects between 3-4 per 100,000 people per year. Individuals with connective tissue disorders are more susceptible to aortic aneurysms or tears. By using large-scale population screening for such diseases, at-risk patients could be identified and notified that they could be at-risk for such a rare disease, that otherwise the patient would be unaware of (Andreu-Perez et al., 2015).

IV. Conclusion

My Technical Report and my STS Thesis are addressing two different but increasing popular fields in the computing world: big data and the cloud computing. My Technical Project's goal is to design a prototype of a web application that will allow patients and physicians to enter patient data, track progress, and compare against other patients. I will make use of the cloud to design a serverless web architecture for the application. My STS Research Thesis has a goal of analyzing big data in the healthcare industry using STS frameworks. I will use the adoption of early forms of Healthcare Information Technology, and relate it to big data use, to try and gain some new insight as well of some of the issues that will face.

V. References

Amazon Aurora – Auto-Scaling Serverless Database Service - AWS. (2018). Retrieved from https://aws.amazon.com/rds/aurora/serverless/

Andreu-Perez, J., C, C., Poon, Y., Merrifield, R. D., Wong, S. T., & Yang, G. (2015). Big Data for Health. *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS,19*(4). Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7154395.

AWS Lambda – Product Features. (n.d.). Retrieved from https://aws.amazon.com/lambda/features/

Bhattacherjee, A., & Hikmet, N. (2007). Physicians resistance toward healthcare information technology: A theoretical model and empirical test. *European Journal of Information Systems,16*(6), 725-737. doi:10.1057/palgrave.ejis.3000717

Bodrov-Krukowski, I. (2018, April 23). Angular Introduction: What It Is, and Why You Should Use It. Retrieved from https://www.sitepoint.com/angular-introduction/

Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., . . . Shekelle, P. G. (2006). Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care. *Annals of Internal Medicine,144*(10), 742. doi:10.7326/0003-4819-144-10-200605160-00125

Chronic sinusitis. (2018, March 03). Retrieved from https://www.mayoclinic.org/diseases-conditions/chronic-sinusitis/symptoms-causes/syc-20351661

Health IT: Advancing America's Health Care(Rep.). (n.d.). The Office of the National Coordinator for Health Information Technology.

Oracle Big Data. (n.d.). Retrieved from https://www.oracle.com/big-data/guide/what-is-big-data.html

Poon, E. G., Jha, A. K., Christino, M., Honour, M. M., Fernandopulle, R., Middleton, B., . . . Kaushal, R. (2006). Assessing the level of healthcare information technology adoption in the United States: A snapshot. *BMC Medical Informatics and Decision Making,6*(1). doi:10.1186/1472-6947-6-1

Serverless Architectures. (n.d.). Retrieved from https://martinfowler.com/articles/serverless.html

Sherman, P. (2018, April 11). How Single-Page Applications Work – Paul Sherman – Medium. Retrieved from https://medium.com/@pshrmn/demystifying-single-page-applications-3068d0555d46