Section 2: Week 4: Theory to Practice: Health Care

Nate Bachmeier

TIM-7020: Database and Business Intelligence

December 29th, 2019

North Central University

Data Management in Healthcare

The global growth of data has increased from 4.4 to 44ZB between 2013 to 2020 and continues to double every two years (Mansouri, Nadjaran, & Buyya, 2017). This new era of Big Data impacts all industries, as they need to evolve data management strategies to align with the high volume, velocity, and variety of their data sources. The health care industry, in particular, stands to benefit significantly from this transformation, yet they remain one of the laggards due to regulation and compliance limitations. These limitations impact multiple participants in the health care ecosystem, ranging from patients to doctors and insurers. Until removing these constraints, the health care industry will be slow to innovate and encounter higher operational and capital expenditures (OPX/CAPX), relative to their peer industries.

# What is the current state of Health Care

Consider all of the different data sources that the health care system needs to consume and correlate today. These sources include low-resolution manual entry sources, such as schedules and visitation notes, along with high-resolution images, such as magnetic resonance imaging and ultra-sound video. Then factor the secondary layers that exist to support these systems, like insurers and pharmaceutical companies. Outside of the medical facility, IoT devices, such as personal fitness trackers and outpatient care monitors, stream continuous feeds of high volume unstructured data to disparent third-party servers. Each of these siloed systems could benefit from having insight into their peers. For instance, researchers cannot always access sufficient data to prove or disprove a treatment works. Doctors do not always have the full picture of their patients, leading to misinformed decisions. Patients need to manage multiple online profiles and cannot easily access a ‘single pane of glass’ that encompasses the entire state of their health.

# Why do these challenges exist

Regulation and compliance requirements force many health care systems to remain in private data centers and out of the public cloud. Applications within a private data center are less agile than those in the public cloud, precisely because of a lacking of instantaneous provisioning of elastic resources. For example, connecting two private data centers requires buying proprietary Virtual Private Network Appliances (VPN) and secondary hardware to support the scenario. Purchasing and installing these components could take several months, compared to the public cloud, which can provision a VPN in a matter of minutes. Similar experiences exist in data platform tooling, like though provided by Microsoft Azure and Amazon Web Service. These modern solutions can glue together multiple built-for-purpose data management technologies that address specialized scenarios. Consider the distinction between a surgeon performing open-heart surgery and machine learning practitioner seeking disease correlations. The prior is substantially more time-critical and could benefit in-memory databases versus the later where slower and cheaper batch processing stores are acceptable. However, there are scenarios where the inverse is also true, such as the practitioner needs interactive business intelligence. In these scenarios, the public cloud allows for costly high-performant stores to economically become hydrated for a matter of hours, then released when no longer needed. Private data centers also lack access to cloud-native solutions such as Storage as a Service (StaaS), Machine Learning as a Service (MLaaS), and other Software as a Service (SaaS) offerings. These technologies remove aspects of the learning curve, accelerate innovation by removing boiler-plate efforts, and allow the engineering teams to focus on the core competencies of the specific business goals.

A pivotal hindrance to adopting modernized platforms comes from the legal requirements of health care professionals. The Health Insurance Portability and Accountability Act (HIPAA) the Genetic Information Nondiscrimination Act of 2008, Health Information Technology for Economic and Clinical Health Act (HITECH), and similar laws seek to protect the patients from discrimination and secure their information against negligence (Hofstra University, 2019) (HSG, 2018). These health care administrators deliver these requirements through a combination of encryption and network segmentation strategies. However, these objectives are hard to accomplish in practice due to the many-to-many relationship between Electronic Health Records (EHR) and decentralized consumers of the data (Virtru, 2019).

Media sensations, such as Celebgate, where thirty-four celebrity Apple Cloud accounts became compromised (Owen, 2019), cause the public to distrust of cloud security. While members of the technology industry see weak passwords as the culprit, others assume that the core platform is at fault. These opinions discount facts such as Microsoft’s spending one billion dollars per year on Azure security (Patterson, 2018) and Amazon’s purchase harvest.ai, a cybersecurity artificial intelligence solution, for nineteen billion dollars (Business Insider, 2017). Few, if any, health care providers have security budgets that rival that scale. Their investments enable Virtual Private Clouds (VPC) to encapsulate their cloud resources within a bubble that enjoys multiple levels of both physical and virtual systems protection.