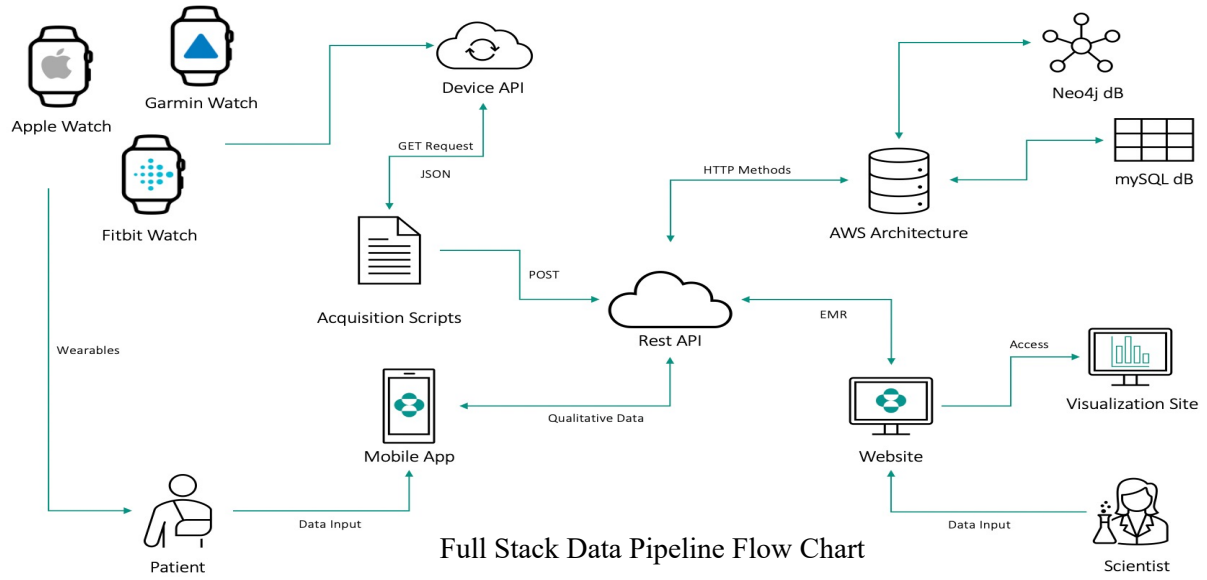
CLINICAL TRIAL DATA PIPELINING WITH BIOMETRIC WEARABLE DEVICES

The Purdue Data Mine and Merck Corporate Partnership

Problem Statement Overview

- Clinical trial data acquisition and pipelining is a messy and difficult process in the pharmaceutical industry
- Data is often lost, misplaced, unsecure, or incomplete offering trouble for analysis of drug efficacy
- These data pipeline shortcomings can delay the release of drug development and result in lost time and money for leaders in drug development
- Using biometric wearable technologies, including sensors, heart rate monitors, and watches, Merck's data management and acquisition of clinical trial studies can be streamlined
- This project specifically focuses on creating a clean and consistent data pipeline stemming from wearable fitness technologies (Fitbit Watch, Apple Watch, Garmin Watch)
- Efficiency, clean data, low costs, patient privacy, and user preferences are the basis of this data pipeline
- Our goal is to join and centralize patient information – thereby automating the collection of real time information on clinical trial patients and providing new insight into the effects of drugs on physical and mental performance



Full Stack Data Pipeline Flow Chart

Depth	MySQL Time (S)	Neo4J Time (S)
2	0.016	0.010
3	30.267	0.168
4	1543.505	1.359
5	> 1 hour	2.132

Vicknair, Chad, et al. "A Comparison of a Graph Database and a Relational Database." *Proceedings of the 48th Annual Southeast Regional Conference on - ACM SE '10*, 2010, doi:10.1145/1900008.1900067.

Data Storage

- We elected to store clinical trial biometric data using MySQL and Neo4J
- Data schemas and scripts were written for the Apple Watch, Garmin Watch, and Fitbit Watch. Our focus while constructing this project's data architecture was cost and timeliness
- The cost for MySQL is around $5,000 while Neo4J is around $36,000; however, it is priced around $65 an hour so prices can be controlled by the user.
- Our queries performed in 0.016 seconds in MySQL, and 0.010 seconds in Neo4J. Neo4j ran faster; however, MySQL can work with more data points at one time.

Visualization

- Visualizing patient clinical trial data is crucial for understanding the effects of new medication
- Data visualizations were made using R Shiny and accompanying packages like plotly and ggplot2
- Our development features two Shiny dashboards; one for visualizations focusing on the population of the study and the other depicting individual patient data
- Dashboards are connected to the database, allowing the graphs to update as new data is collected
- Built mobile application and website to aid clinical trial data acquisition and visualization

Data Acquisition

- Data acquisition is the first step in the data pipeline
- Making use of existing APIs and frameworks, data is collected directly from user's wearables accounts
- For Fitbit device data, data is pulled via a series of requests using Fitbit's open-sourced API and Selenium to run as a headless browser.
- User biometric data is saved both locally as a csv file and to the AWS database.
- Window's Task Scheduler was used to run this process daily.

Abhimanyu Agarwal, Pranav Anandarao, Xin Du, Denae Galloway, Allison Hill, Connor Koelsch, Joshua Kosnoff, Patricia M.M. Casaca, Riya Mogli, Karthik Ravishankar, Praveen Sentha, Priyanka Seth, Anav Sharma, Judy Si, Siddharth Srinivasan, Surya Suresh, Nick Rosenorn, Dr. Terri Bui, Kai Bode