

DataWrangler:

An Efficient Platform for Large-Scale Exercise Study Management

Team: Steven Sprouls, Andrea Caviglia, Cheyenne Clutter, Samantha Rodriguez, Jensen Roe

Faculty Mentor: Dr. Eck Doerry

Client: Dr. Kyle Winfree

Why is it Needed?

Cardiovascular diseases (CVD) are the leading cause of death in the United States, accounting for one in every four deaths nationwide. Research on the relationship between exercise and CVD is urgently needed, but these studies can be difficult to conduct:

- Need to monitor hundreds / thousands of participants 24/7 for weeks or months
- Lab grade activity trackers are expensive and impractical
- Data is difficult to recover efficiently

Our client, Dr. Kyle Winfree, has pioneered the use of cheap, durable Fitbit devices as effective activity trackers, but a tool for deploying them in real studies is needed.

Our Solution: DataWrangler

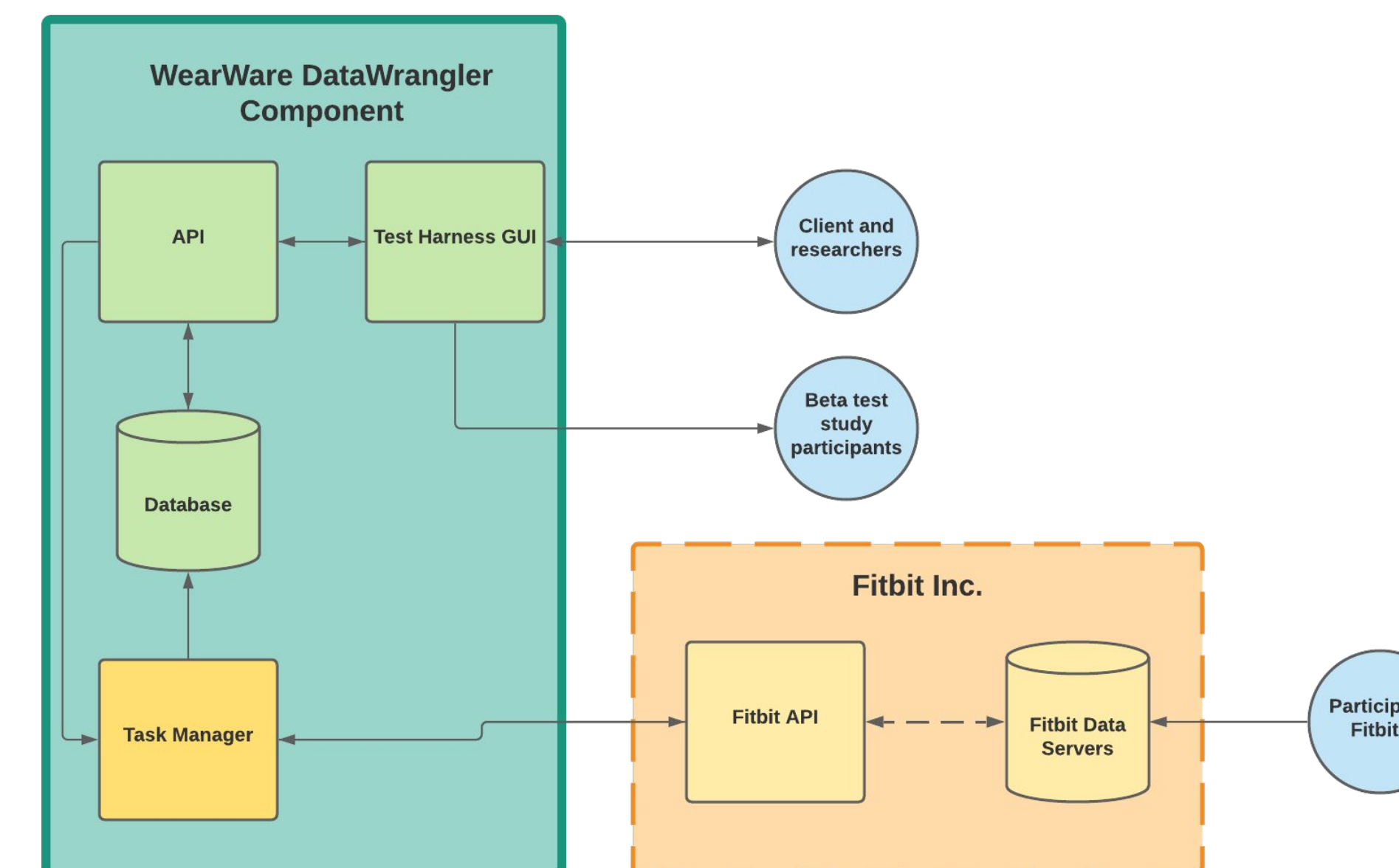
The WearWare system provides the informatics support for gathering raw activity and sleep data from study participants' Fitbit devices. In particular, the DataWrangler module is the core data collection and management backend piece that allows researchers to:

- Create and manage studies
- Enroll participants using a one-time signup link
- Access study participant Fitbit data
- Perform key data management functions through an API

System Architecture

DataWrangler is made up of several key components, and interacts closely with the Fitbit company's API to access device data.

- Fitbit Inc. gathers raw activity and sleep data from Fitbit devices and stores it on their own data servers
- Celery tasks request raw data through Fitbit API
- Data is stored on WearWare's own database on the cloud
- Researchers interact with the DataWrangler API to access stored data



Technologies Used

Used a variety of cutting edge web2.0 tech including:

- Amazon Web Services - cloud hosting service
- Django REST framework - Python based web toolkit for building APIs
- FitBit API - used to collect data from Fitbit devices
- PostgreSQL - relational database system
- Celery - asynchronous task queue system

How Does DataWrangler Work?

id	Active	Start_date	End_date	Name	Creation_time	Comment
1	True	2021-03-04	2021-03-26	TESTSTUDY	2021-03-18T22:56:11.628308Z	yes!
6	True	2021-03-16	2021-04-30	fake study test	2021-03-16T00:00:00Z	test study populated with fake data

1: A list of studies that a researcher can access and view participant data

The form is titled 'Create' and contains the following fields: First name (Test), Last name (Tester), Email (test100@gmail.com), Sex (Male), Gender (Male), Pairing token, and Active (checked). There are 'Close' and 'Save' buttons at the bottom right.

2: A form which creates a participant and stores them into the database

id	Second	Bpm	Timestamp	Device
100000	0	69	2021-03-12T00:00:00Z	1000100
100001	0	60	2021-03-12T00:01:00Z	1000100
100002	0	74	2021-03-12T00:02:00Z	1000100
100003	0	71	2021-03-12T00:03:00Z	1000100
100004	0	60	2021-03-12T00:04:00Z	1000100
100005	0	79	2021-03-12T00:05:00Z	1000100
100006	0	67	2021-03-12T00:06:00Z	1000100
100007	0	66	2021-03-12T00:07:00Z	1000100
100008	0	77	2021-03-12T00:08:00Z	1000100
100009	0	69	2021-03-12T00:09:00Z	1000100
101000	0	62	2021-03-12T00:00:00Z	1000101
101001	0	61	2021-03-12T00:01:00Z	1000101
101002	0	61	2021-03-12T00:02:00Z	1000101
101003	0	67	2021-03-12T00:03:00Z	1000101
101004	0	61	2021-03-12T00:04:00Z	1000101

3: The heart record of participants in a specific study. Researchers can ask for all data in a study, or filter for specific participants.

What's Left?

While the DataWrangler module provides the core functionality for the larger WearWare system vision, future extensions to improve usability and efficiency include:

- The StudyWrangler module
 - A more advanced user interface
 - A graphical data browser
- The Interactions module
 - Real time interactivity with study participants
 - Scans data for "events of interest"

The WearWare system is being deployed on our client's AWS server and will be used for several upcoming studies. This system will significantly reduce the cost and effort required to perform such studies, allowing for larger studies at a reasonable cost.