Crowdsourcing Smart Home Data

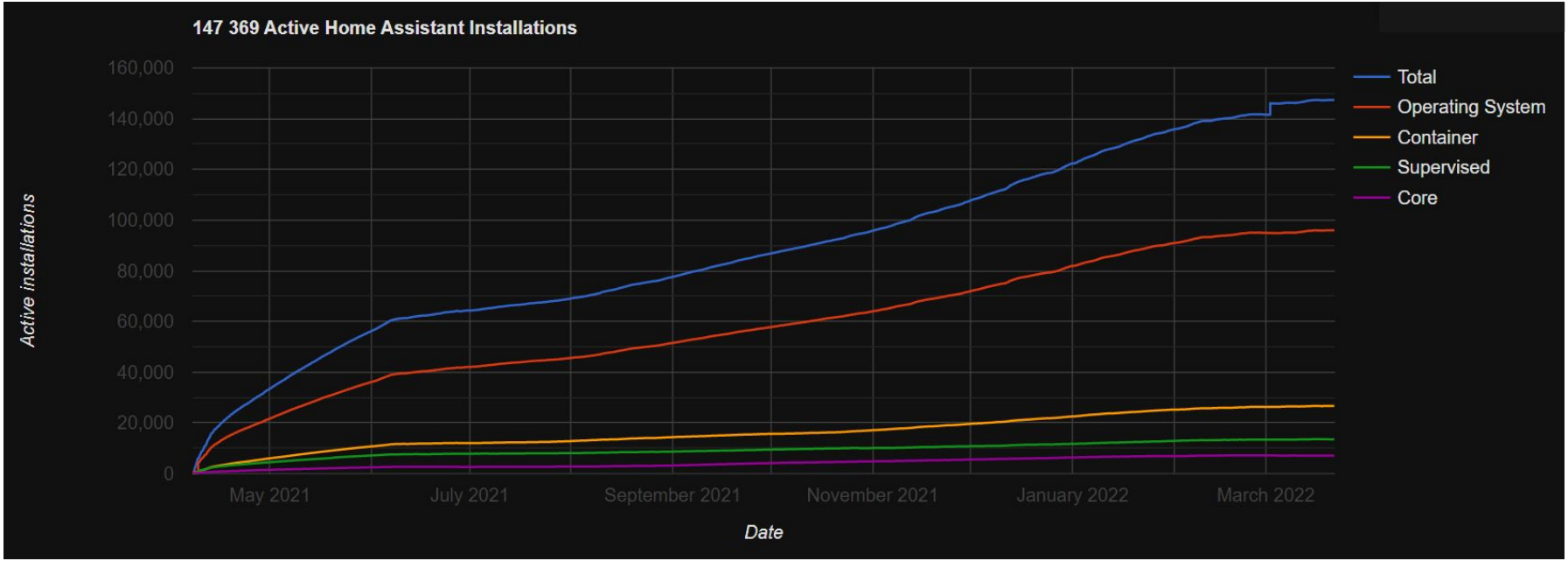
The Team

Team Leader	- Martinho Tavares
Frontend Dev	- Diogo Monteiro
Backend Dev	- Camila Fonseca
DevOps Master	- Rodrigo Lima

Coordinator	- Diogo Gomes

Context

Smart homes have been on the rise, and will continue to grow in the foreseeable future



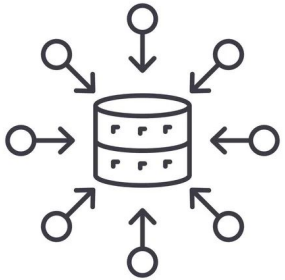
From https://analytics.home-assistant.io/

Problem

- ❏ Lack of real data about smart home usage (information about the house itself, rather than statistics)
- ❏ No central and automated way of gathering data from many sources
- ❏ Few datasets provide the information we're looking for

Goals

- ❑ Collect smart home usage data from volunteers
- ❑ Aggregate and store collected information
- ❑ Respect user privacy and anonymity
- ❑ Export data in CKAN compliant formats
- ❑ Visualize data in a web-based dashboard



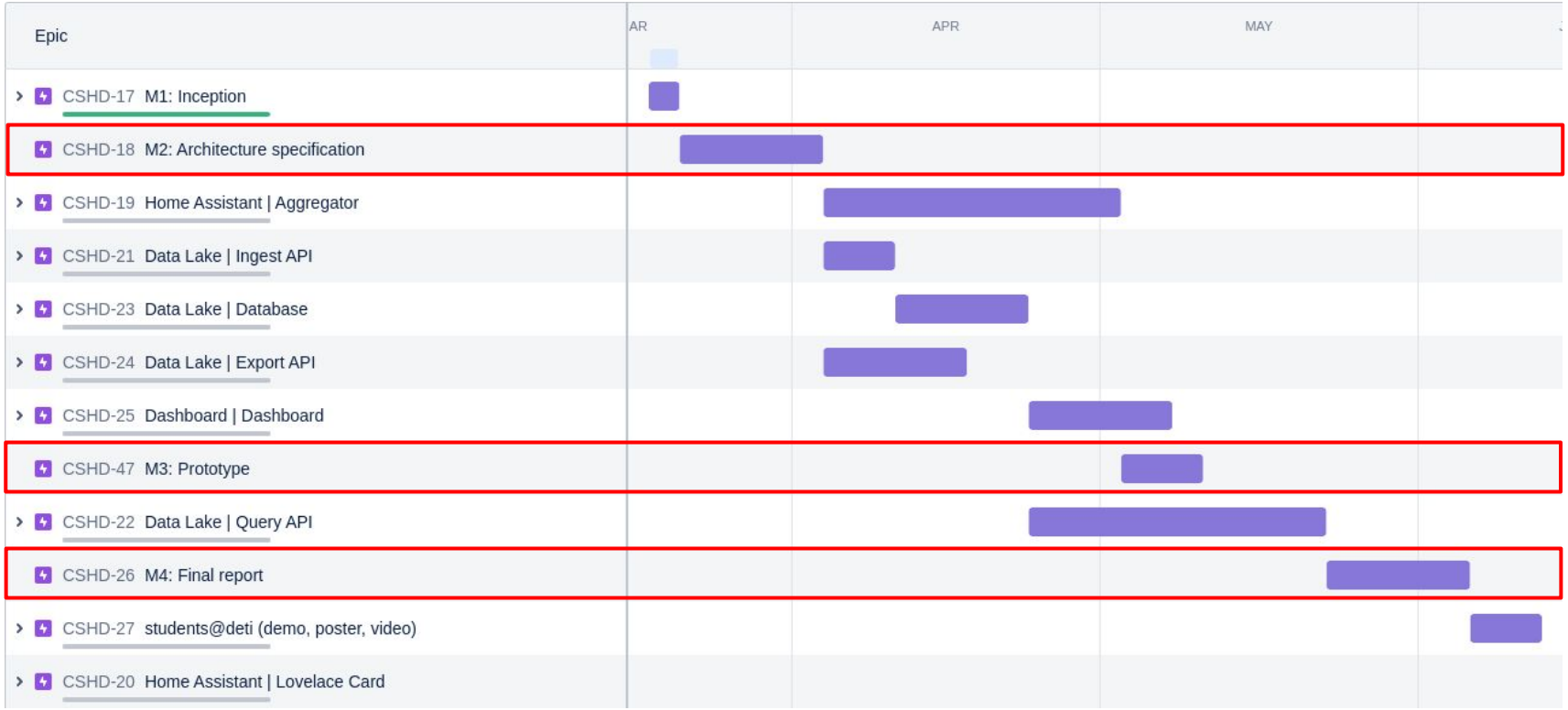
Expected Results

An easy-to-use platform that provides quality data for further research

We expect our solution to be provided in the following deliverables:

DATA LAKE	DASHBOARD	HOME ASSISTANT
Ingest API Query API Export API	Web Application	Home Assistant Aggregator

Calendar


Epic
AR
APR
MAY
CSHD-17 M1: Inception
CSHD-18 M2: Architecture specification
CSHD-19 Home Assistant | Aggregator
CSHD-21 Data Lake | Ingest API
CSHD-23 Data Lake | Database
CSHD-24 Data Lake | Export API
CSHD-25 Dashboard | Dashboard
CSHD-47 M3: Prototype
CSHD-22 Data Lake | Query API
CSHD-26 M4: Final report
CSHD-27 students@deti (demo, poster, video)
CSHD-20 Home Assistant | Lovelace Card

Tasks

Home Assistant Aggregator:

4 weeks (Camila & Diogo)

- ❏ Aggregation: develop custom component for data collecting, treatment, and sending
- ❏ Connection setup: compact information efficiently for sending and connect to the Ingest API
- ❏ Anonymity guarantee: ensure that all collected data is anonymous, and prevent any kind of identifiable info from being inferable

Tasks

Data Lake database:

2 week (Martinho & Rodrigo)

- ❏ Database setup: preparing the database to ingest large quantities of data with proper configurations

Tasks

Data Lake Ingest API:

1 weeks (Martinho)

- ❏ API setup: setup the public API, create the endpoints, storage of raw, anonymous data into the database

Tasks

Data Lake Export API:

2 week (Rodrigo)

- ❏ API setup: setup the API, create the endpoints, dumping of data in CKAN compliant formats

Tasks

Data Lake Query API:

4 weeks (Camila, Rodrigo & Martinho)

- ❏ API setup: setup the API, create the endpoints, real-time aggregation/transformation of data

Tasks

Dashboard:

2 weeks (Diogo & Rodrigo)

- ❏ Develop a web-based platform to view collected information, as well as data about the platform itself
- ❏ Display information in various forms, such as graphs
- ❏ Design a user friendly interface with responsive design
- ❏ Signup to participate in data collection and configure what data is collected, and right to be forgotten
- ❏ Integration of the dashboard with the Home Assistant Aggregator that will send the data to the data lake

Related Work

Open source data extraction script:
https://github.com/Anschke/ML_Home_Assistant

Dataset of smart home electricity consumption:
https://paperswithcode.com/dataset/uk-dale

Communication

Website: https://crowdsorcerer.github.io/crowdsource-smart-home-site/

Github page: https://github.com/CrowdSorcerer

Jira: https://martinhotav.atlassian.net/jira/software/projects/CSHD/boards/1