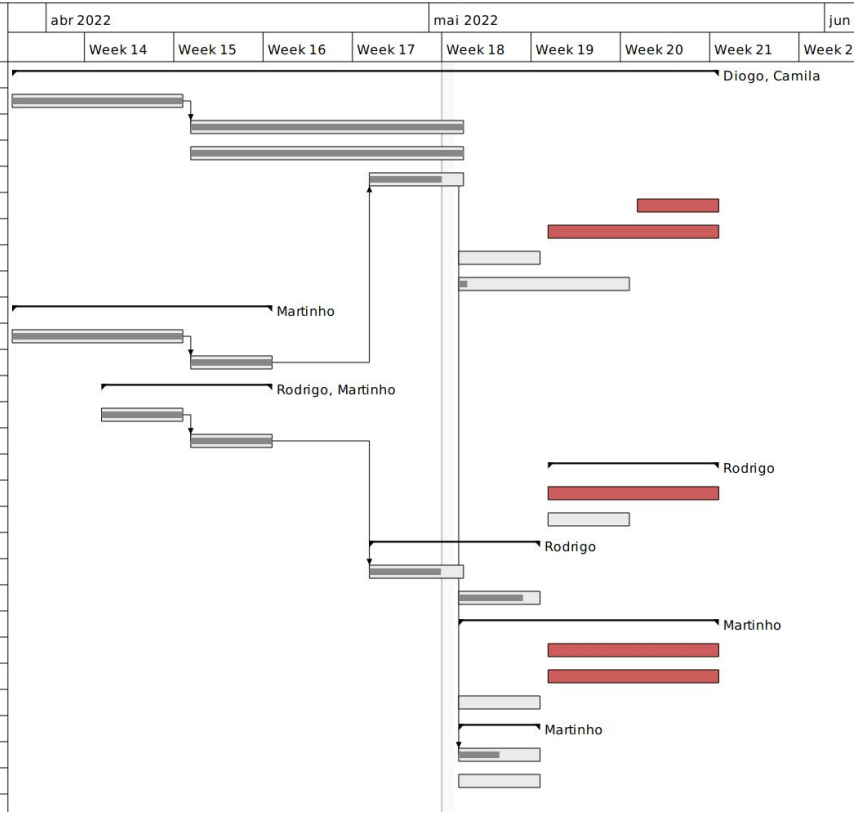
Crowdsourcing Smart Home Data

Prototype

Group 5

Prototype modules

Start	Finish	Name	Work
mar 29	**mai 23**	**Home Assistant | Aggregator**	**45d 2h**
mar 29	abr 11	Creation and successful setup/interaction	10d
abr 12	mai 3	Obtain data from the integrations	11d 4h
abr 12	mai 3	Aggregation setup	6d
abr 26	mai 3	Aggregator connection	1d
mai 17	mai 23	Compression and network	1d
mai 10	mai 23	Dashboard Card	1d
mai 3	mai 9	Generate UUID	1d
mai 3	mai 16	Anonymity setup	13d 6h
mar 29	**abr 18**	**Data Lake | Ingest API**	**14d 5h**
mar 29	abr 11	Ingest setup	9d 7h
abr 12	abr 18	Deploy and properly define endpoints	4d 6h
abr 5	**abr 18**	**Data Lake | Database setup**	**10d**
abr 5	abr 11	Create and configure database	5d
abr 12	abr 18	Deploy and test connectivity	5d
mai 10	**mai 23**	**Data Lake | Export API**	**7d 5h**
mai 10	mai 23	Dump in CKAN compliant formats	6d 5h
mai 10	mai 16	API setup with data extraction from data lake	1d
abr 26	**mai 9**	**Dashboard | Dashboard**	**2d**
abr 26	mai 3	Backend setup	1d
mai 3	mai 9	Frontend setup	1d
mai 3	**mai 23**	**Data Lake | Database optimizations**	**3d**
mai 10	mai 23	Data filling with AI	1d
mai 10	mai 23	Investigate and test possible optimizations	1d
mai 3	mai 9	Configure security	1d
mai 3	**mai 9**	**Data Lake | Ingest API optimizations**	**2d**
mai 3	mai 9	Set compression method on incoming data	1d
mai 3	mai 9	Rate-limit operations on each host	1d



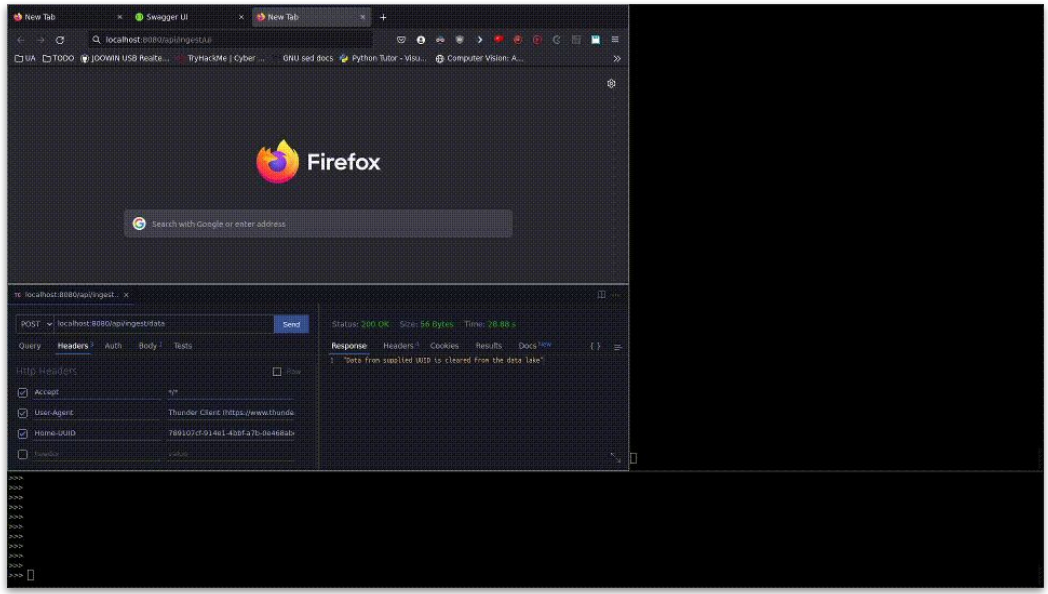
Ingest API

Features:

- Endpoint for uploading data relative to a home, without treatment of any kind
- Endpoint for the opt-out feature, deleting data from a home

TODO:

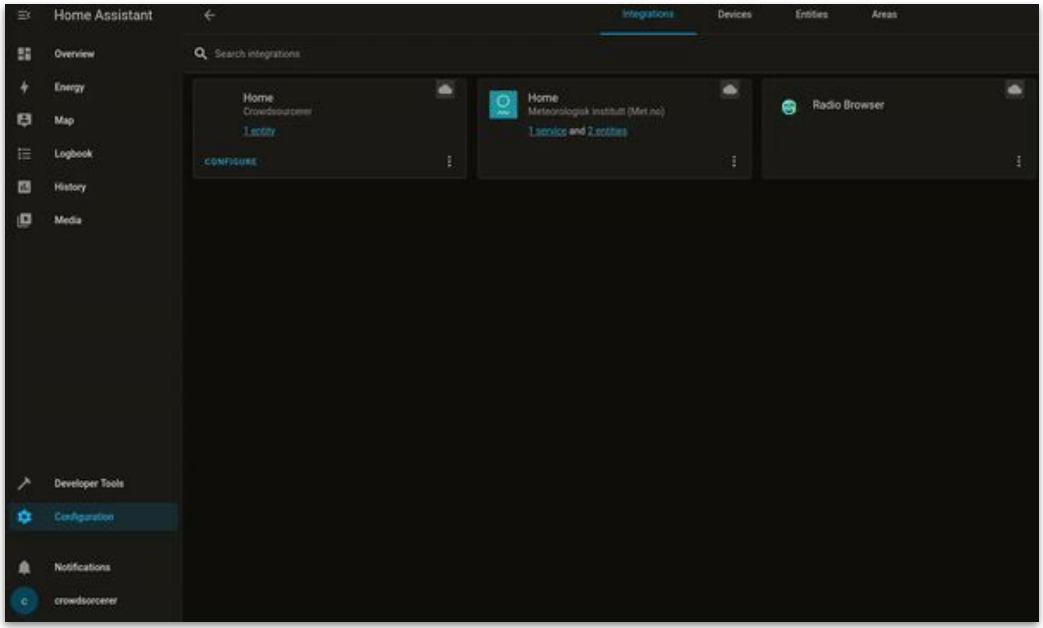
- Define the compression method used for the payload on data upload
- Rate-limit operations for each host



Home Assistant Aggregator

Features:

- Data retrieval from all sensors in the Home Assistant installation
- Customizable blacklist to allow user to select which sensors' data to send
  - Via Graphical Interface
- Send retrieved data to Ingest API



TODO:

- Dashboard Card
- Anonymity

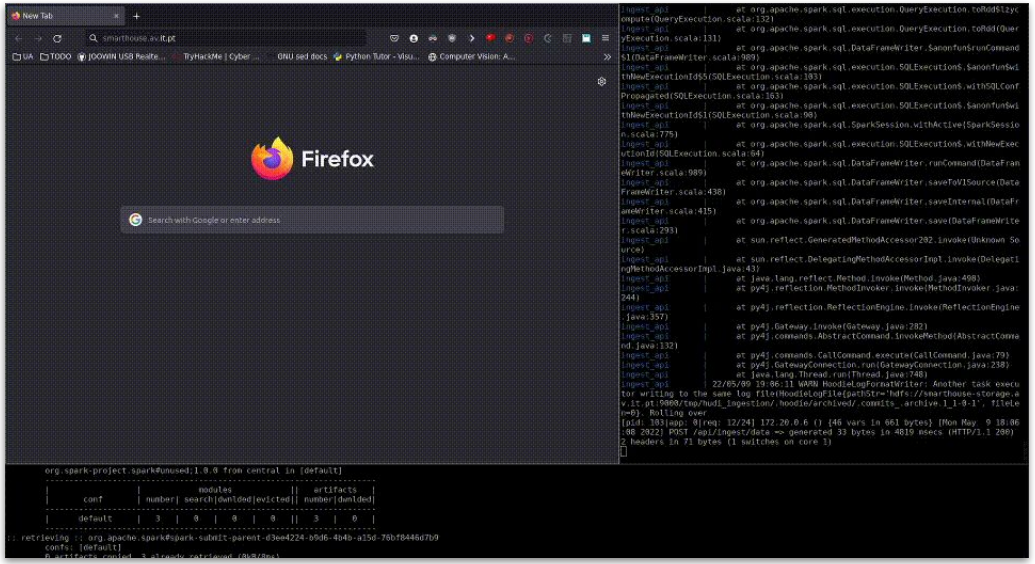
Dashboard

Features:

- Custom dashboards and panels given by user defined queries
- Periodic metric gathering with the help of prometheus
- Graphical data presentation

TODO:

- Add API endpoints health checks
- Improve Prometheus access privileges
- Define the role of the Query API in this module
- Add more relevant metrics (such as alerts on data upload that has been discontinued)



Backend (Data Lake)

Features:

- Hadoop Distributed File System (HDFS) with Spark on Yarn
- Periodic metric generation (every 15 minutes)
- Spark communication between the machines
- Service-level components (API, dashboard) fully dockerized

TODO:

- Improve security (Hadoop...)
- Check for misuses/misconfigurations
- Data filling with AI, for privacy

Next modules

Export API

- Create API using Swagger
- Export data in data lake into CKAN compliant formats

Thank you