Redback Operations - Deakin

Project 11 FIT File Handling and Data Pipeline

Mark Telley

Overview

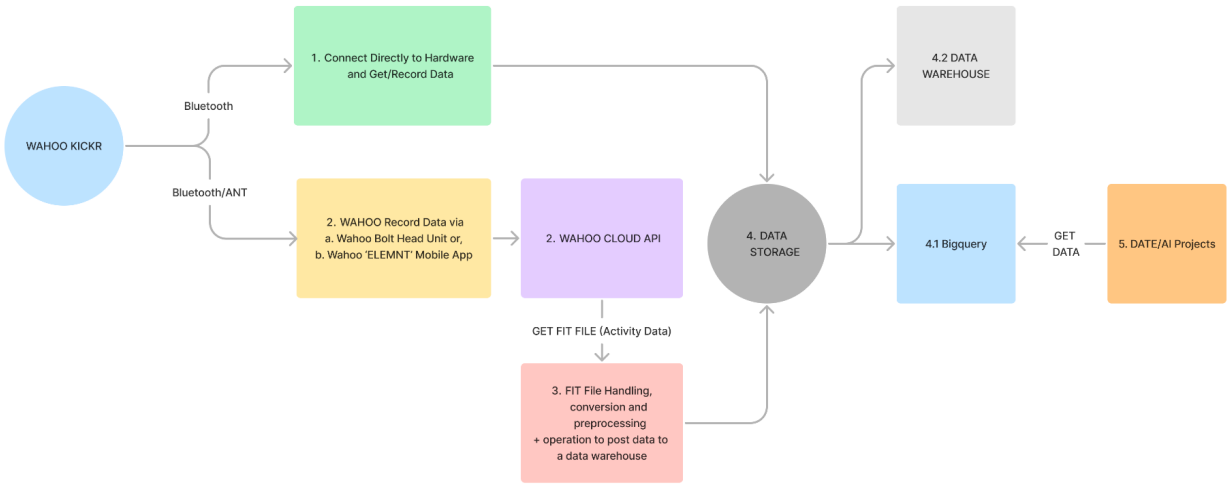
Project 11 ‘FIT File Handling and Data Pipeline’ aims to establish a prototype version of a data pipeline that processes activity data generated by the user using the Wahoo Kickr. This pipeline, comprising six primary stages, involves capturing data directly from the Wahoo Kickr or through other Wahoo devices, retrieving FIT files via the Wahoo Cloud API, preprocessing these files, storing the processed data in a data warehouse, and facilitating access to this data repository for Data/AI applications. The objective is to simplify and enhance data handling processes, thereby bolstering future Data and AI initiatives. It's important to note that this is the first iteration of the pipeline, and improvements and modifications are anticipated as we continue to refine and optimise the process.

Pipeline (Macro)

The pipeline is made of a number of elements, where data is generated via the Wahoo Kickr:

Click HERE to see the full pipeline or click the Figma:
https://www.figma.com/file/T8VS4o9cbdaOhu4HgAvCkM/Wahoo-FIT-FILE-Pipe?type=whiteboard&node-id=0%3A1&t=L0q8kkfBtjOwcUwc-1

1. Data recorded directly from the Wahoo Kickr.
2. Data record via Wahoo (FIT File)
3. Connection to Wahoo Cloud API to get the FIT File
4. Handling the FIT File and getting it ready for ingestion
5. Storage in a data warehouse
6. Access to the data warehouse for Data/AI purposes.



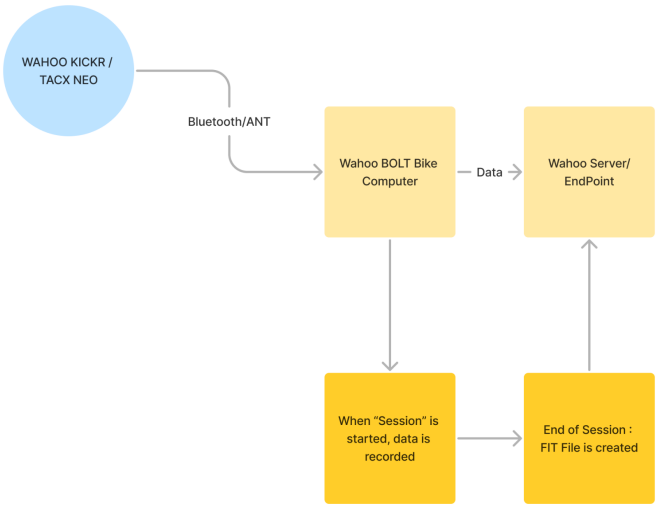
Pipeline (Detailed)

1. Data recorded directly from the Wahoo Kickr.

This component of the pipeline aims to access raw data generated by the Wahoo Kickr. This element of the pipeline was established by looking at different ways to retrieve and record data. The IoT team is working on being able to get the data directly (which should be modelled off the FIT File Data schema) and push activity or session data to a data warehouse.

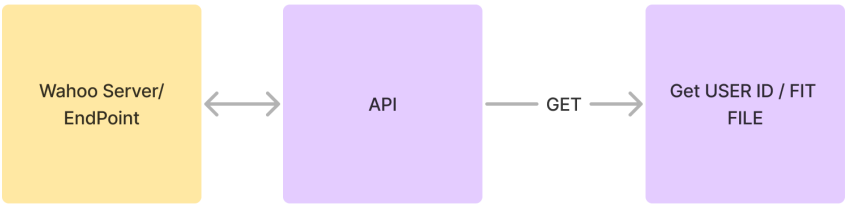
2. Data record via Wahoo (FIT File)

The alternative to getting data directly from the Wahoo Kickr is via other Wahoo devices. Wahoo has provided an excellent roadmap for the Redback team i.e. how data is captured and then organised i.e. A user creates a workout and then a workout summary/workout is created. This is completed using either the Wahoo ELEMNT app or a Wahoo Bolt head unit. Moreover, there are many other devices that can capture activity data and produce a FIT file.



3. Connection to Wahoo Cloud API to get the FIT File

A key part of the pipeline is being able to retrieve a FIT file from Wahoo. To do this the Wahoo Cloud API is used.



4. Handling the FIT File and getting it ready for ingestion

Once a FIT file is successfully retrieved from Wahoo, it needs to be handled. FIT Files are not in a human-readable format and need to be converted and handled in such a way that data is nicely organised.

5. Storage in a data warehouse

Assuming the FIT file is converted and handled correctly the final step is to ingest the data. To do this, a simple SQL query can be executed to either create or update a table for a specific user - at this point, the Data Warehouse Project then begins.

6. Access to the data warehouse for Data/AI purposes.

Initially, a Bigquery data warehouse will be used to store a range of data (refer to the GCP read me file). Moreover, the data located within the ‘Fitness Data’ dataset essentially mirrors the data that is derived from step 4. At this point, the DATA and AI team members are then able to point their focus to this location to conduct an array of different tasks re. ML analysis, or as means of supporting a particular project they are working on.

Pipeline Figma Documentation