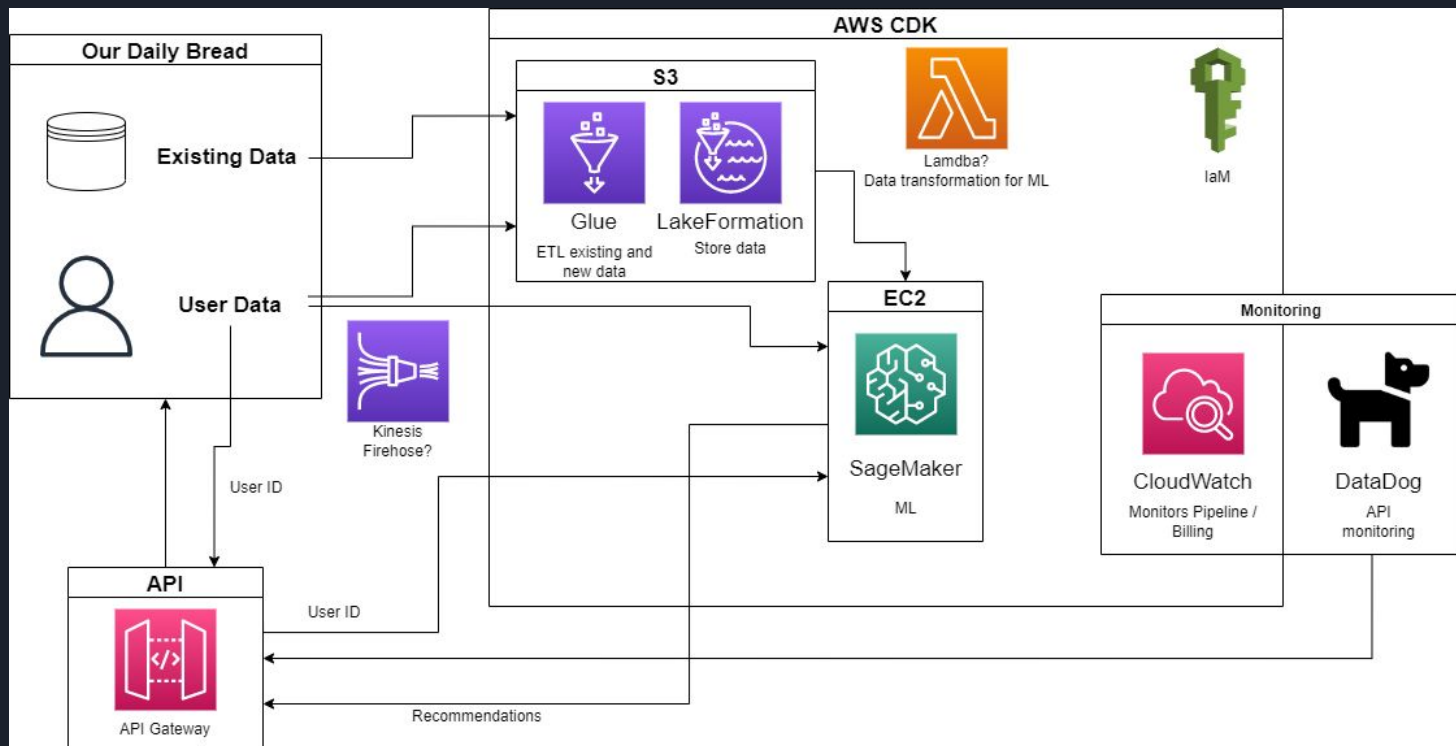# Team OBD

Michael Dykema, Emily Linderman,
Tom Lobbestael, Jacob Johnson

Project Summary

An AWS data pipeline that will provide users with suggested content based on previously viewed content on odb.org. This pipeline is made up of a data lake that receives incoming data, a recommendation engine that makes recommendations based on the data from the lake, and an api that will export recommendations to odb.

Architecture Overview

Issues

- Delayed access to ODB's AWS organization
    - Very limited access and red tape from ODB hierarchy

- Pull Request review and publishing done by ODB
    - Different than initially expected and slows progress

Intended and Actual Progress - Michael

Complete

- API Initialized
    - Setup with simple Lambda function for now
- Helped with Cloud Watch infrastructure
    - Connected to API to monitor errors and cost
- Helped with Lake Formation

Incomplete

- Connecting API to SageMaker/Recommendation engine

Intended and Actual Progress - Emily

Complete

- Get Access to AWS
- Research - Cloudwatch
    - Metric Data
    - Alarm Creation
    - Log Groups
- Implementation
    - Beginnings of Cloudwatch skeleton

Incomplete

- Research
    - Cloudwatch Events
    - Cloudwatch Dashboards

Intended and Actual Progress - Tom

Complete

- Get Access to AWS
- Research - SageMaker
  - Bucketing
  - Tensorflow
- Implementation
  - SageMaker Initial Implementation

Incomplete

- Research
  - SageMaker Pipelining
  - Further Improvement of Algorithm

Intended and Actual Progress - Jacob

Complete

- Get Access to AWS
- Research - Lake Formation
  - Crawlers
  - Data ingestion
  - Log Groups
- Implementation
  - Data lake skeleton

Incomplete

- Data lake deployment
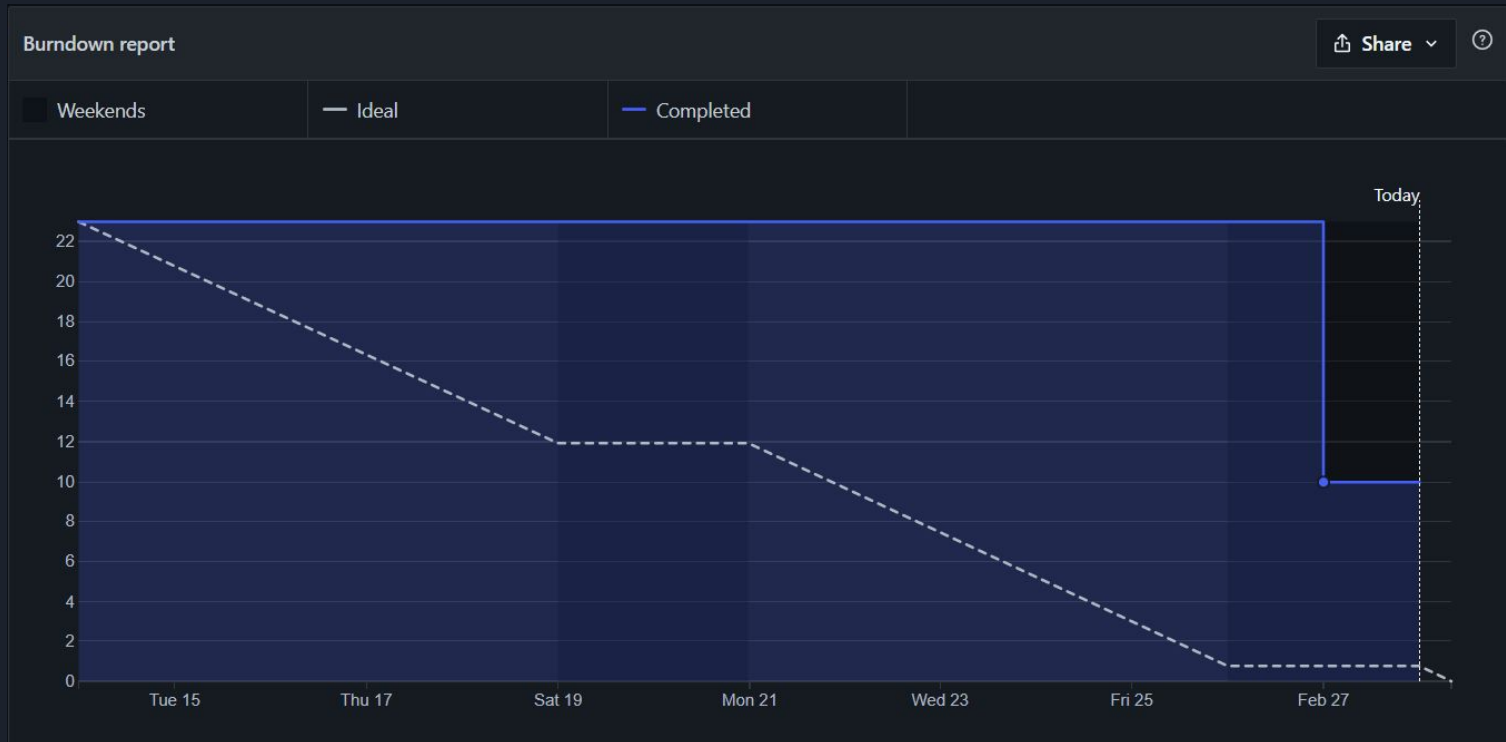- Existing data ingestion into data lake

Incomplete

- Import existing data into Data Lake - Waiting on ODB
- Stream Data into Data Lake - Waiting on ODB
- Connect API to recommendation engine

Burndown

Sprint #3 Project Progress

Overall Goals:

Data

- Import data
    - ODB responsibility
- Stream new data in

Recommendation

- Initialize SageMaker
- Train model

API

- Connect to recommendation engine

Monitoring

- Connect to other service besides API

Individual Goals:

Jacob

- Configure Glue for streaming data

Tom

- Initialize SageMaker

Michael

- Finish API

Emily

- Initialize Cloud Watch dashboard

Demo coming next presentation

Questions?