Status Report:
UBCO MDS Capstone - Urban Data Labs

WEEK 2

Connor Lee, Claudia Nikel, Eva Nguyen, Alex Tamm

Outline

- Project Background
- Overview of the Project
  - Proposal
  - Research questions
  - Statement of work
- Our Approach to the Problem
- Project Deliverables

Project Background

Urban Data Labs (UDL) Capstone Client Overview

Contacts are:

- Jiachen Wei (MDS alum)
- Mike Kennedy, Ph.D

Client Overview

- Formed in September 2019 to advance data analytics capabilities on UBC Vancouver campus to address campus-wide sustainability challenges

Data Overview

- UDL mirrored and stored live-streaming building energy datasets in InfluxDB and made it available to students and researchers
- The SKYSPARK database provides data recorded by the meters and smart devices of many UBC building
- The ION database provides higher granularity data on power, energy, water, and gas use of many UBC building

Overview of Project

Overview

Research Question

- Based on a building's sensor data, how can the data be grouped automatically into NRCan secondary end-use classifications?

Proposal

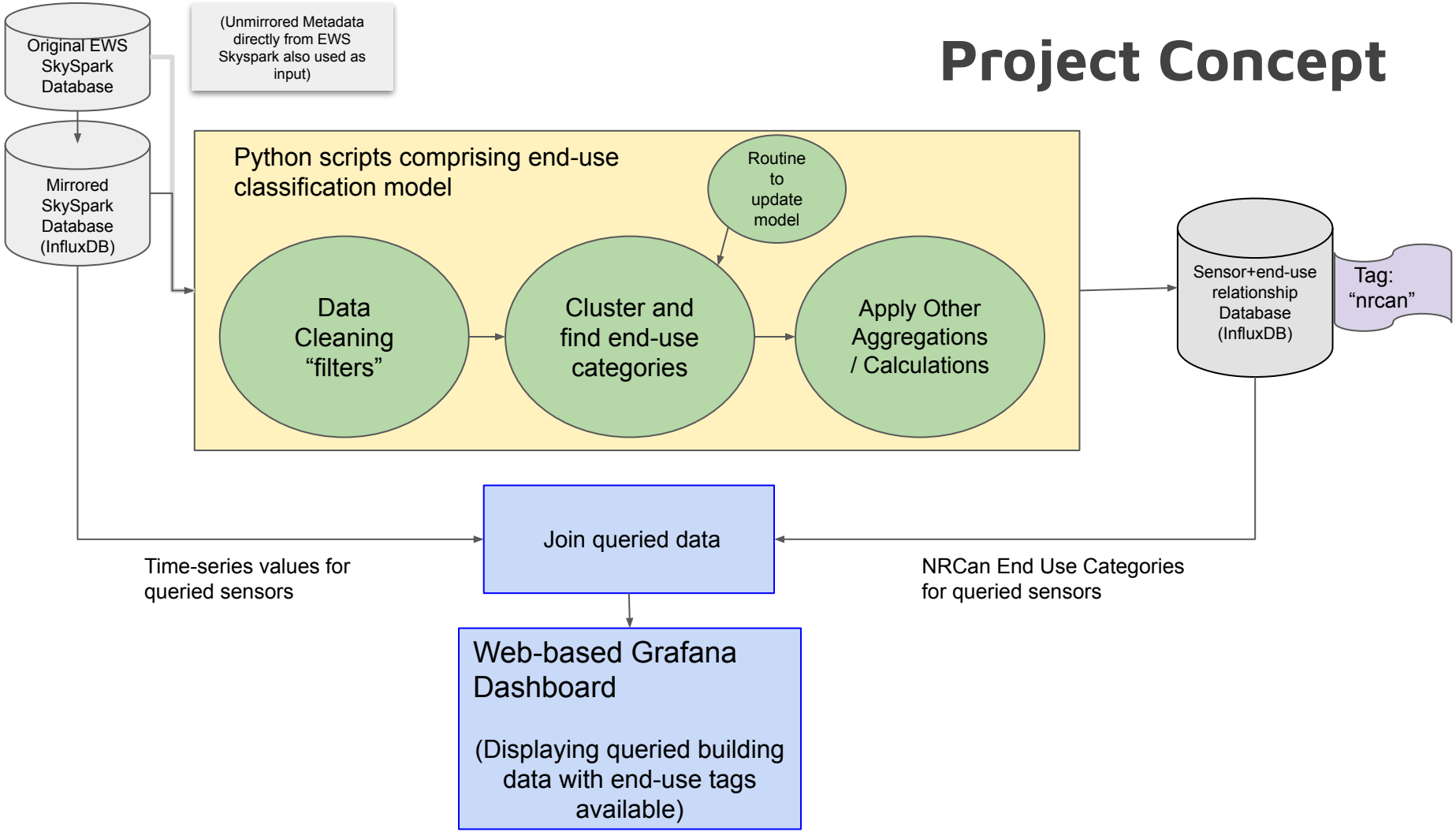
- Create a program that queries live streaming sensor data from the UDL SkySpark database, cleans and uses appropriate ML methods to apply NRCan Secondary End-Use Classifications to the data, and then passes the results to a Grafana visualization dashboard with a few simple example visualizations.

Statement of Work

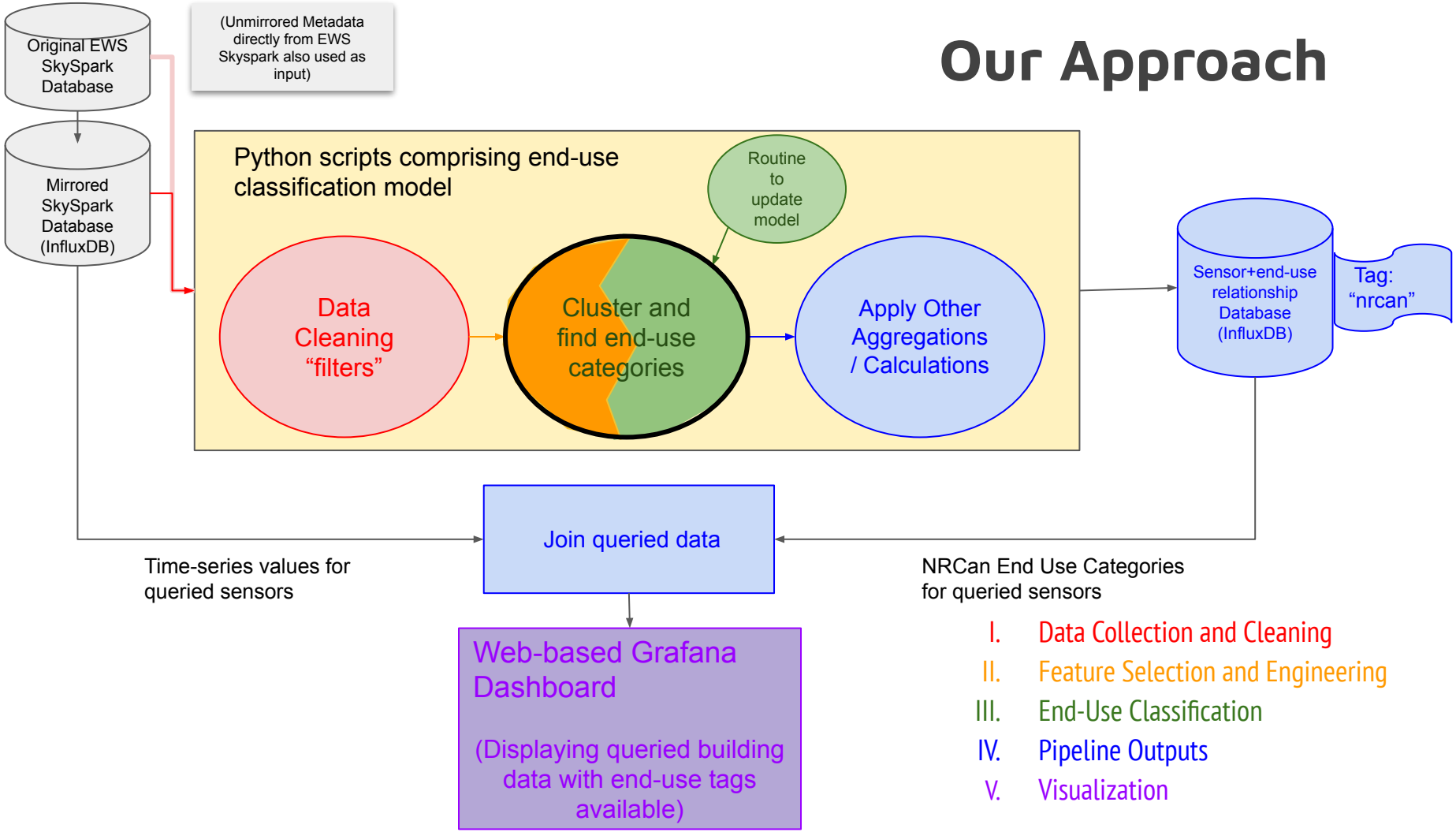
- In order to develop a viable end-use classification product the following steps need to be completed:
    1. Data Collection and Cleaning
    2. Feature Selection and Engineering
    3. End-Use Classification
    4. Pipeline Outputs
    5. Visualization

Problem Approach

Project Concept



Our Approach



I. Data Collection and Cleaning
II. Feature Selection and Engineering
III. End-Use Classification
IV. Pipeline Outputs
V. Visualization

Project Deliverables

What are the exact deliverables?

- A Python program that:
  - Queries and cleans the data required for classifying instrumentation by end-use for the Pharmacy building
  - Classifies instruments by end-use (NRCAN guidelines)
  - Time permitting, scale the model so that it can be used on buildings other than the Pharmacy building
- Web-based Grafana dashboard
- Final report & presentation to UDL

Thanks for listening-Questions?

Work Progress

Progress - Individual Work Logs

Connor

Edited final proposal + EDA + researched MLlib tools in Databricks + created initial code in python & Databricks (sklearn) + researched distance measures + compared clustering methods

Claudia

Edited final proposal + researched ML techniques + EDA + looked into sklearn for Databricks + added documents to GitHub

Alex

Edited final proposal + EDA + created pipeline flowchart + research into pipeline implementation + worked with chronograf & influxDB

Eva

Edited final proposal + researched distance measures + client communication for meetings + organized Jira + created & organized project tasks + reviewed & compared clustering methods + EDA

Progress - Team Work Logs

Accomplishments

- Finished Draft Proposal, got feedback from UDL on proposal and submitted final copy of Proposal
- Developed project overview chart and confirmed project objectives
- Created a GitHub workflow plan
- Initial research on project solutions (clustering methods, distance measures, MLlib, databricks)

Tasks for the Next Weekly Cycle

1. Identify Relevant Features
2. Research Data Flow Connections
3. Transform Data for ML Tasks
4. Develop a tool to identify missing information
5. Develop a tool to populate missing information
6. Identify NRCan's classifications