Status Report:
UBCO MDS Capstone - Urban Data Labs

WEEK 4

Connor Lee, Claudia Nikel, Eva Nguyen, Alex Tamm

Outline

- Progress made during previous week
  - Individual logs
  - Team logs
- Difficulties and roadblocks
- Updated project approach
- Plan for next cycle

Previous Week’s Progress

Progress - Individual Work Logs

Connor	Create flowchart for machine learning aspect + created code for clustering + researched semi-supervised methods + coded for implementing Gower's distance
Claudia	Research encoding for categorical data + started preparing data for feature selection + preliminary data labelling + code review + researched semi-supervised methods
Alex	Queried data for labelling + preliminary data labelling + update model concept flowchart + research restructuring metadata format
Eva	Started preparing data for feature selection + created list of tasks for next sprint + researched encoding for feature selection + joined metadata with training data for feature selection

Progress - Team Work Logs

Accomplishments

- Research approach to populating missing data
- Modified current project objectives to a new report
- Confirmed details for project objective
- Identified relevant metadata fields
- Preliminary data labeling
- Started data prep for feature selection

Difficulties & Roadblocks

Project Schedule

THEMES	WEEKS	DATES	GOALS
Investigation and Data Prep	3	27 April - 14 May	Identify project objectives and key data features + understand data dictionaries + transform data for machine learning tasks.
Model	1	15 - 21 May	Develop a classification model to apply group tags to end-uses for the Pharmacy building.
Model	1	22 - 28 May	Validate and evaluate models.
Scale + Analysis	2	29 May - 11 June	Expand the model to other UBC buildings (if time permits) + complete user-acceptance testing of model + identify conclusions + create visualizations of results + complete user-acceptance testing of dashboards + UBC mid-term presentation
Wrap-Up	2	12 - 26 June	Final report + package final code + UDL final presentation + UBCO final presentation
Total Weeks	**9**		

(OLD DELETE IF REPLACEMENTS ARE OK) Difficulties/Roadblocks

Difficulty

- We feel like we are a bit behind schedule
  - Training and test set not finalized
  - Feature selection and engineering not complete
  - Model development not complete

Roadblock

- Waiting for response from UDL for data labelling in order to create a training set

Difficulties

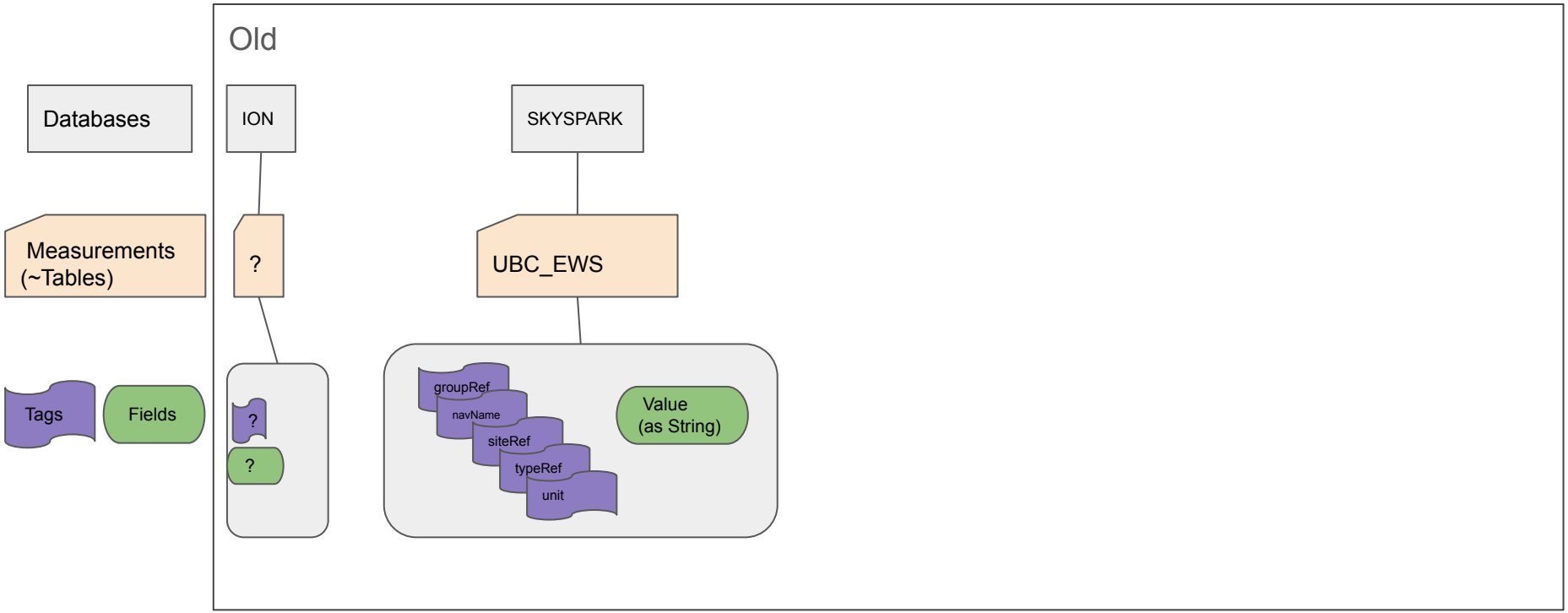
Difficulty	Complete	Working on
Training and test set not finalized	• Initial attempt at hand labeling	• Clarifying labeling • Finalizing labeling
Feature selection and engineering not complete	• Identified common fields between buildings • Encoding • Scaling	• Aggregation • Finalizing selected features • Integrating non-energy consumption sensors as predictors for NRCan classifications
Model development not complete	• Researched alternatives • Found Python packages for alternatives	• Identifying which model(s) to implement

Roadblocks

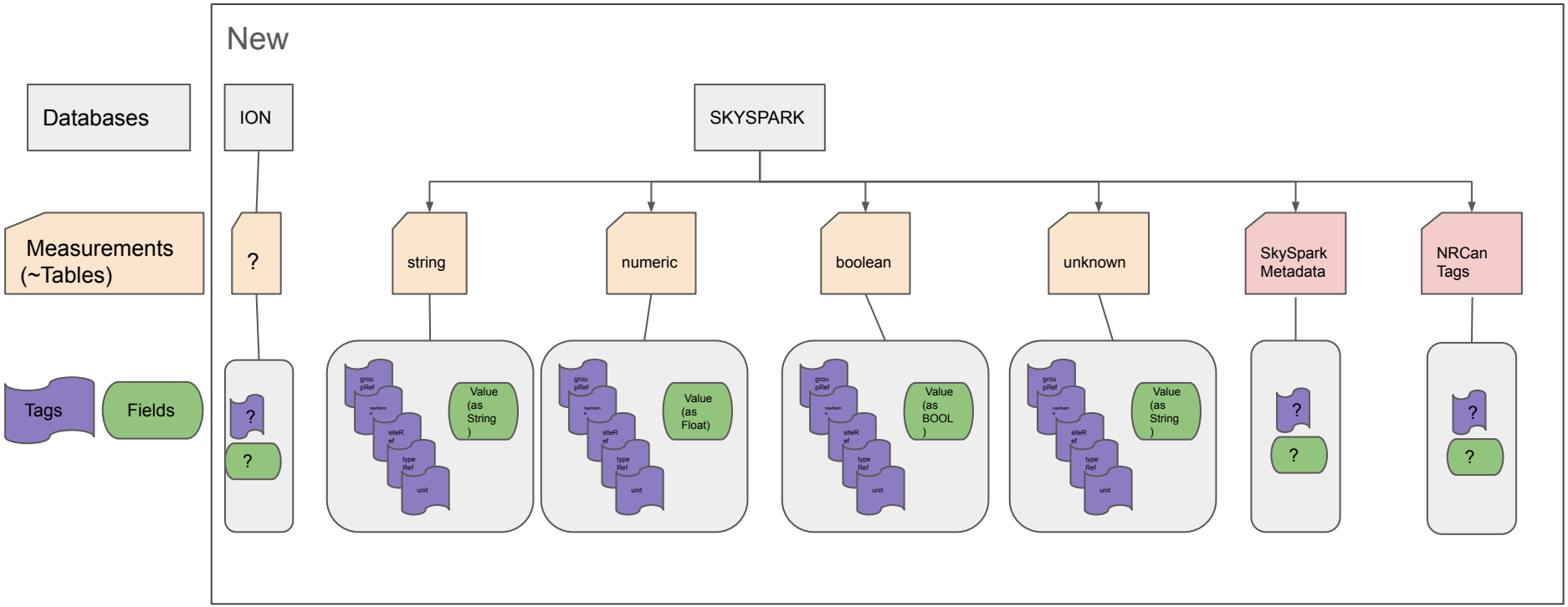
- Waiting for response from UDL for data labelling in order to create a training set
- Timing of database changes may be an issue

Updated Project Approach

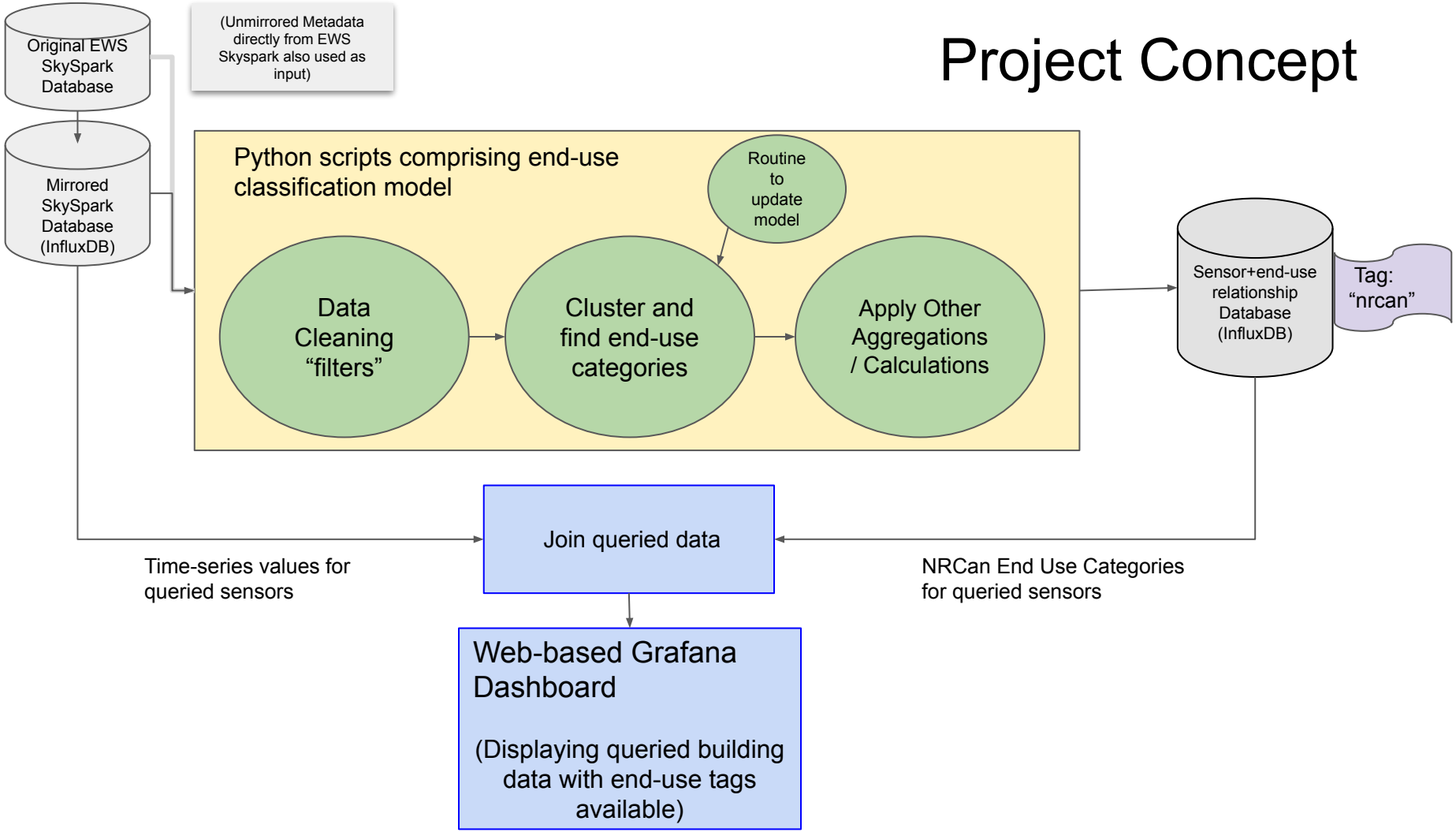
Changes to InfluxDB (slide 1 of 2)



Changes to InfluxDB (slide 2 of 2)

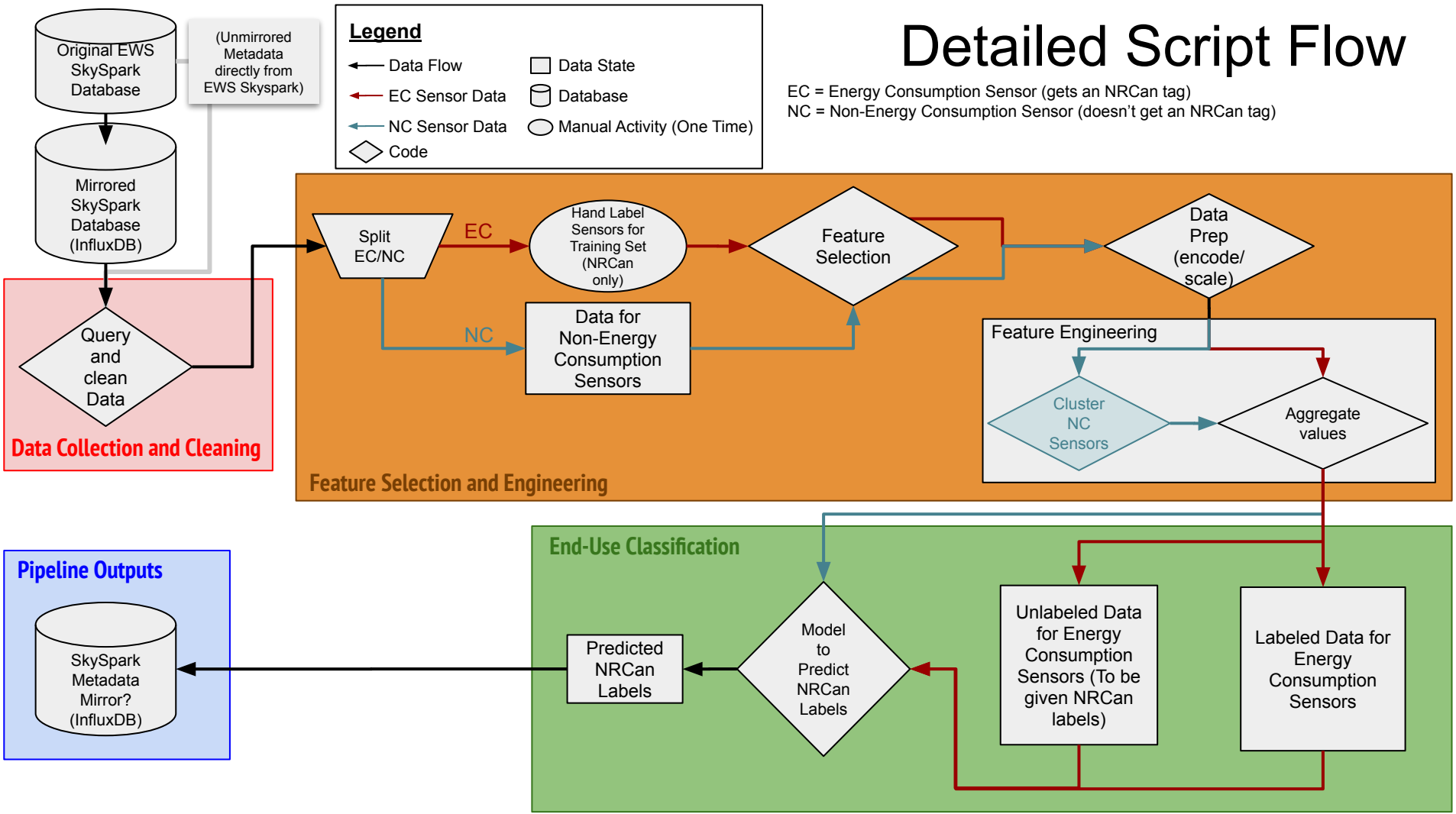


Project Concept



Detailed Script Flow

EC = Energy Consumption Sensor (gets an NRCan tag)
NC = Non-Energy Consumption Sensor (doesn't get an NRCan tag)



Tasks for Next Cycle

Tasks for the Next Weekly Cycle

1. Finalize training & testing data
2. Implement feature selection/engineering techniques & identify relevant features
3. Test & compare different machine learning models

Questions