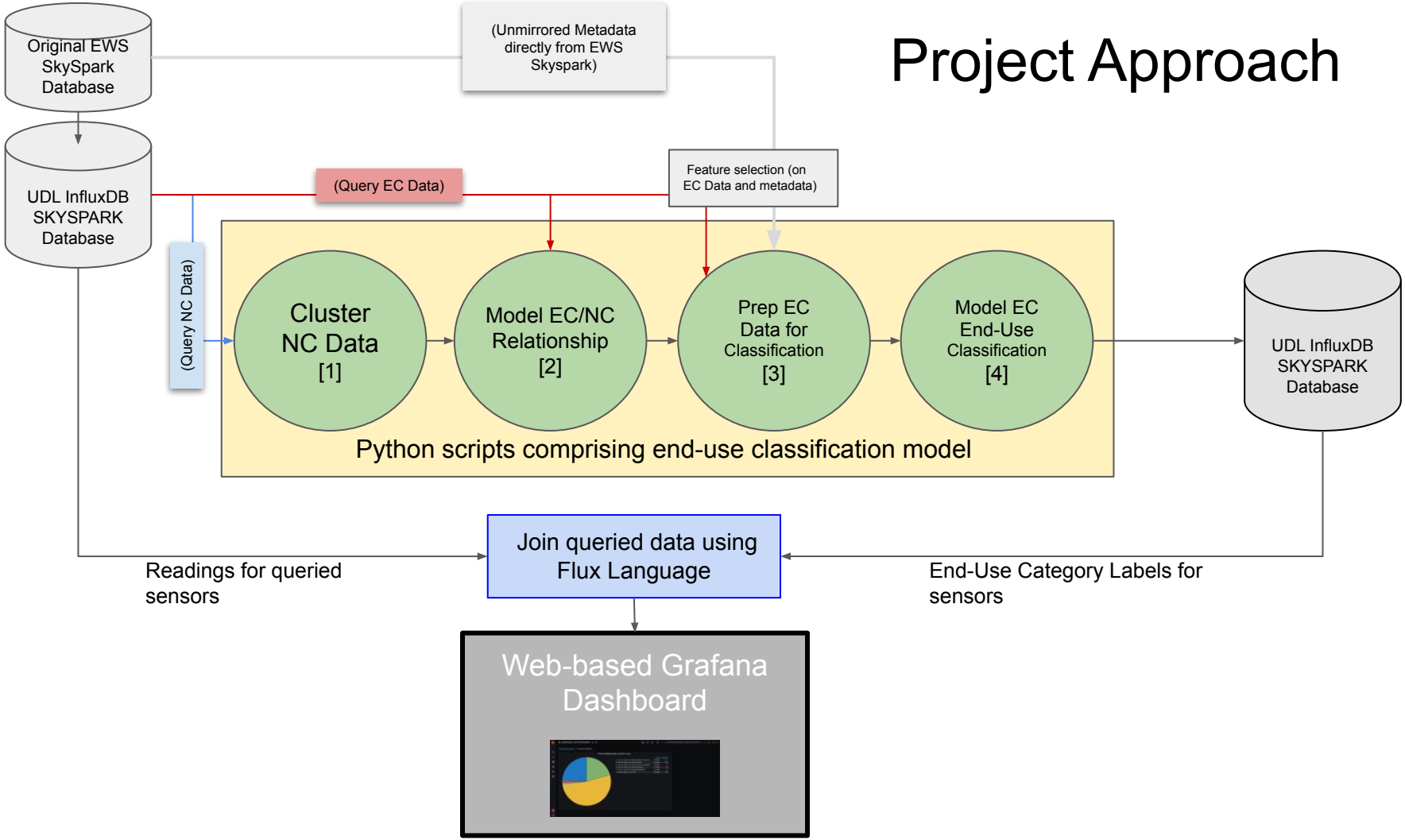
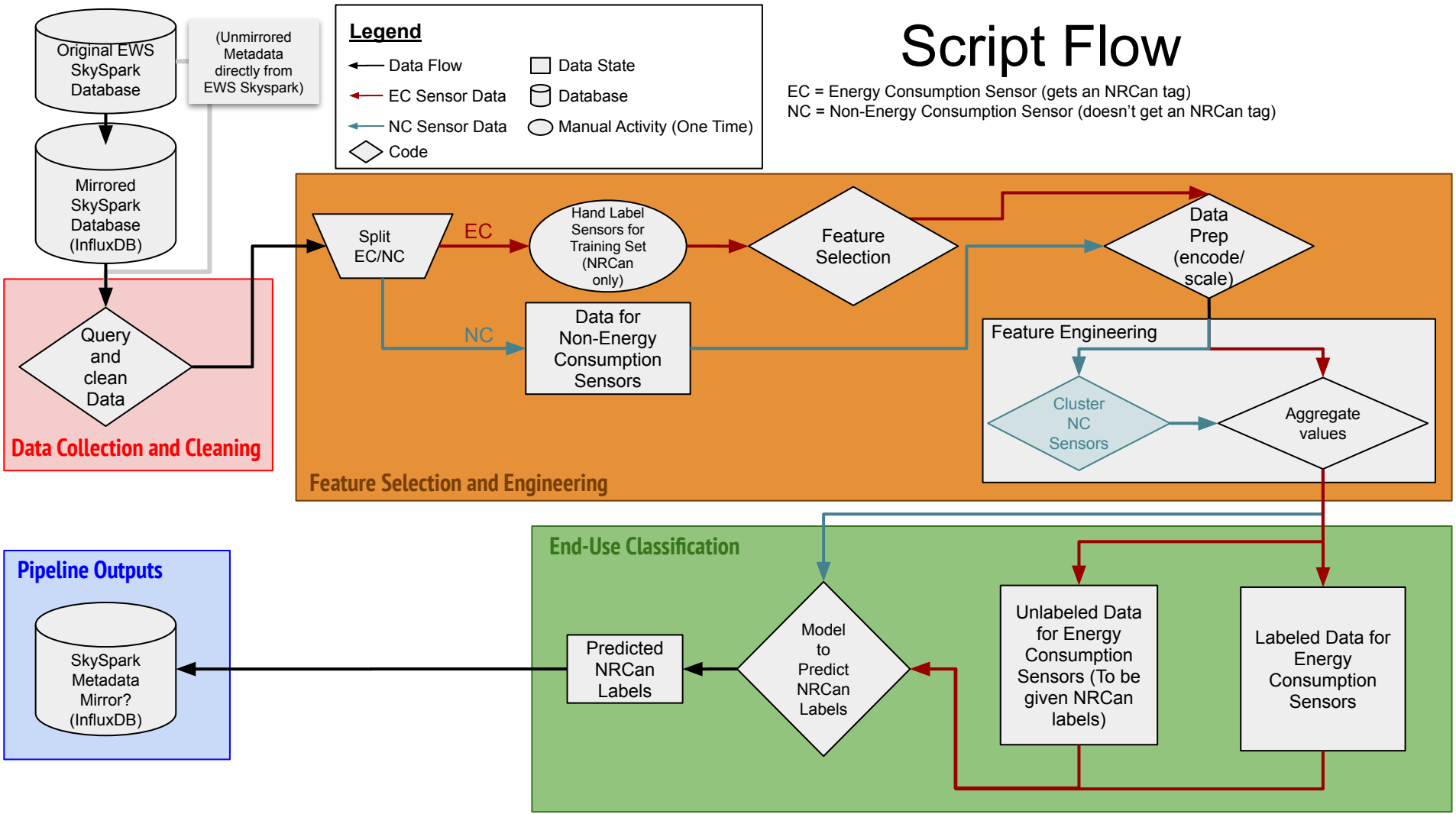
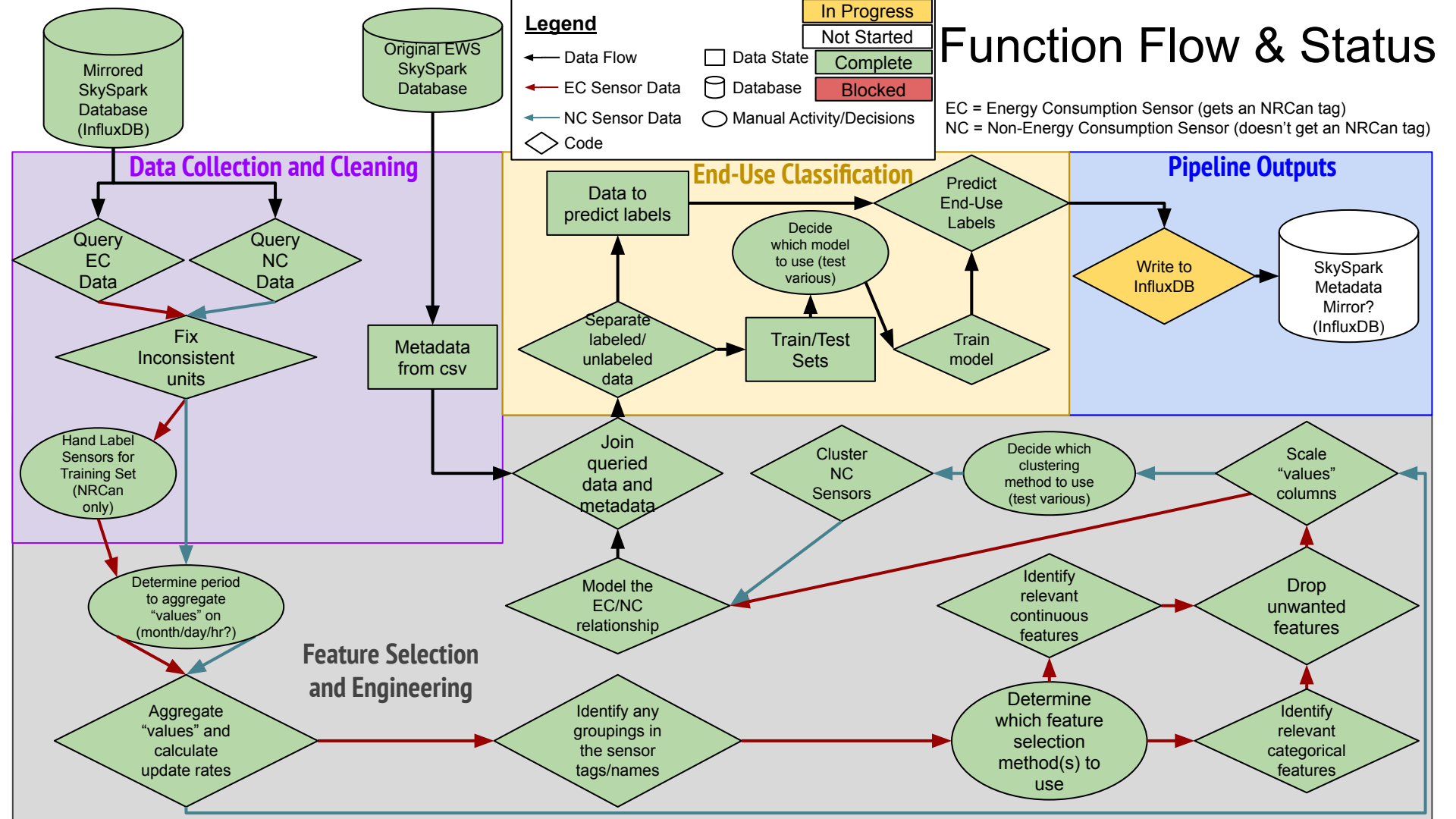
Project Approach



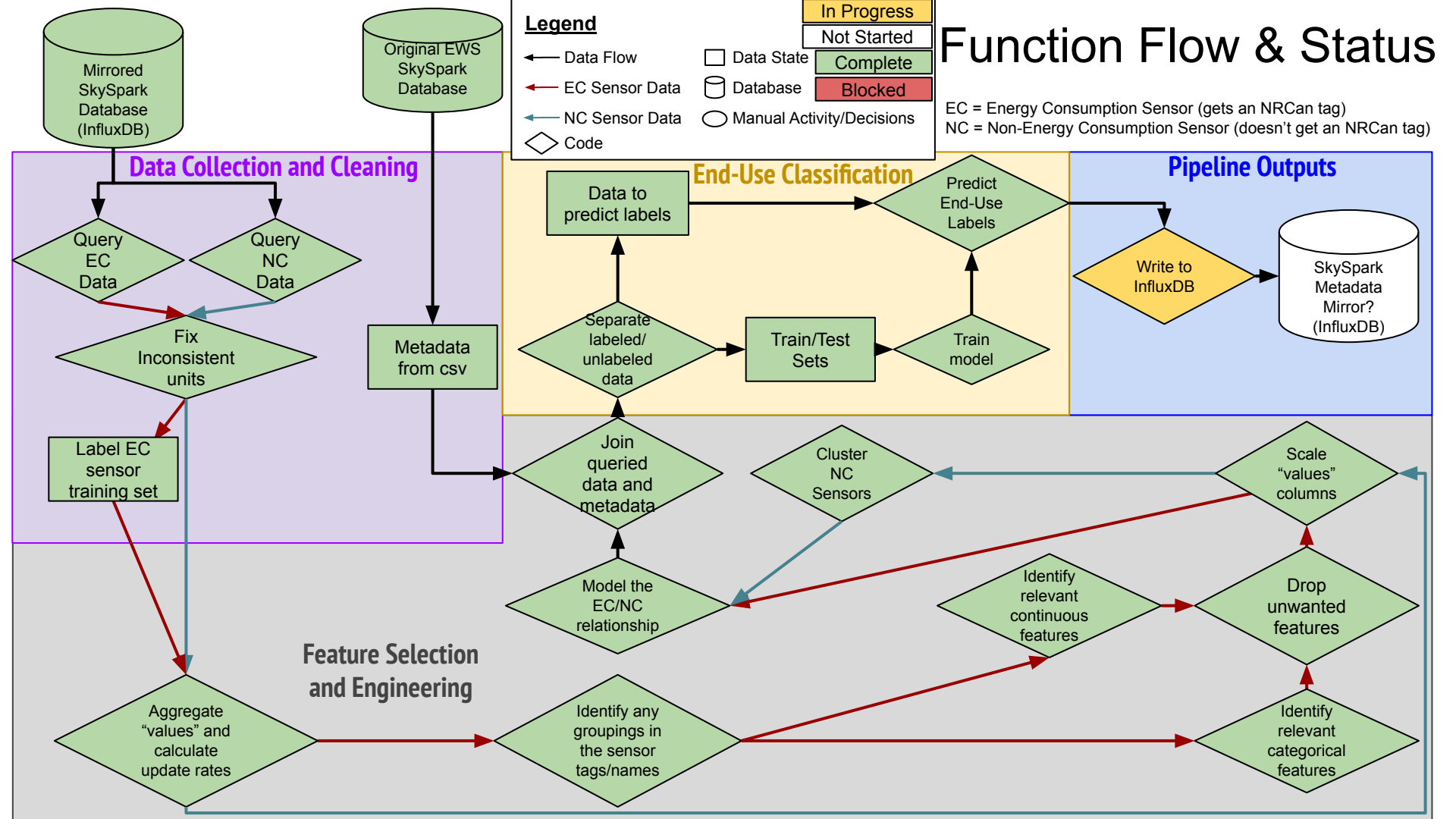
Script Flow

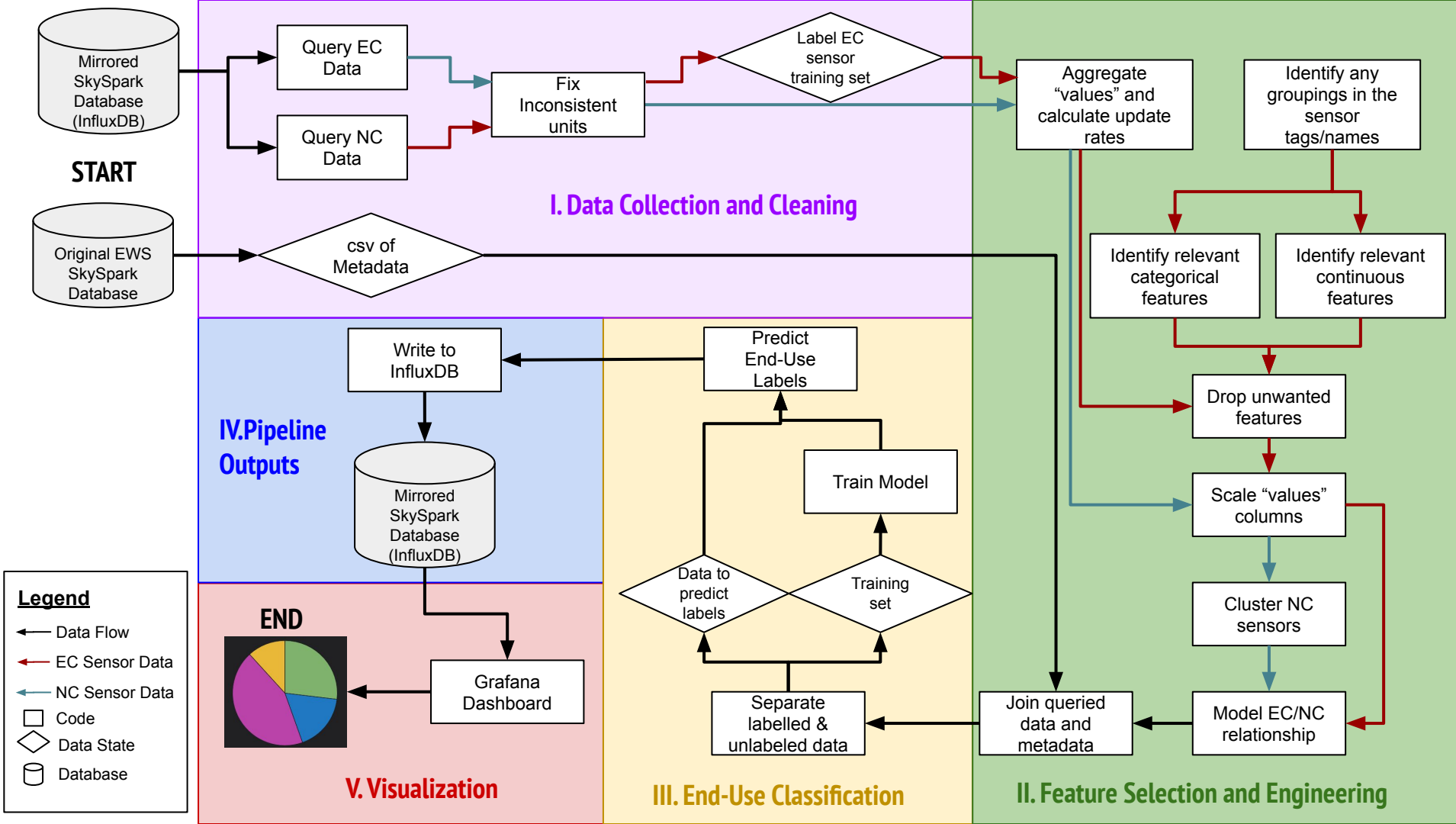
EC = Energy Consumption Sensor (gets an NRCan tag)
NC = Non-Energy Consumption Sensor (doesn't get an NRCan tag)

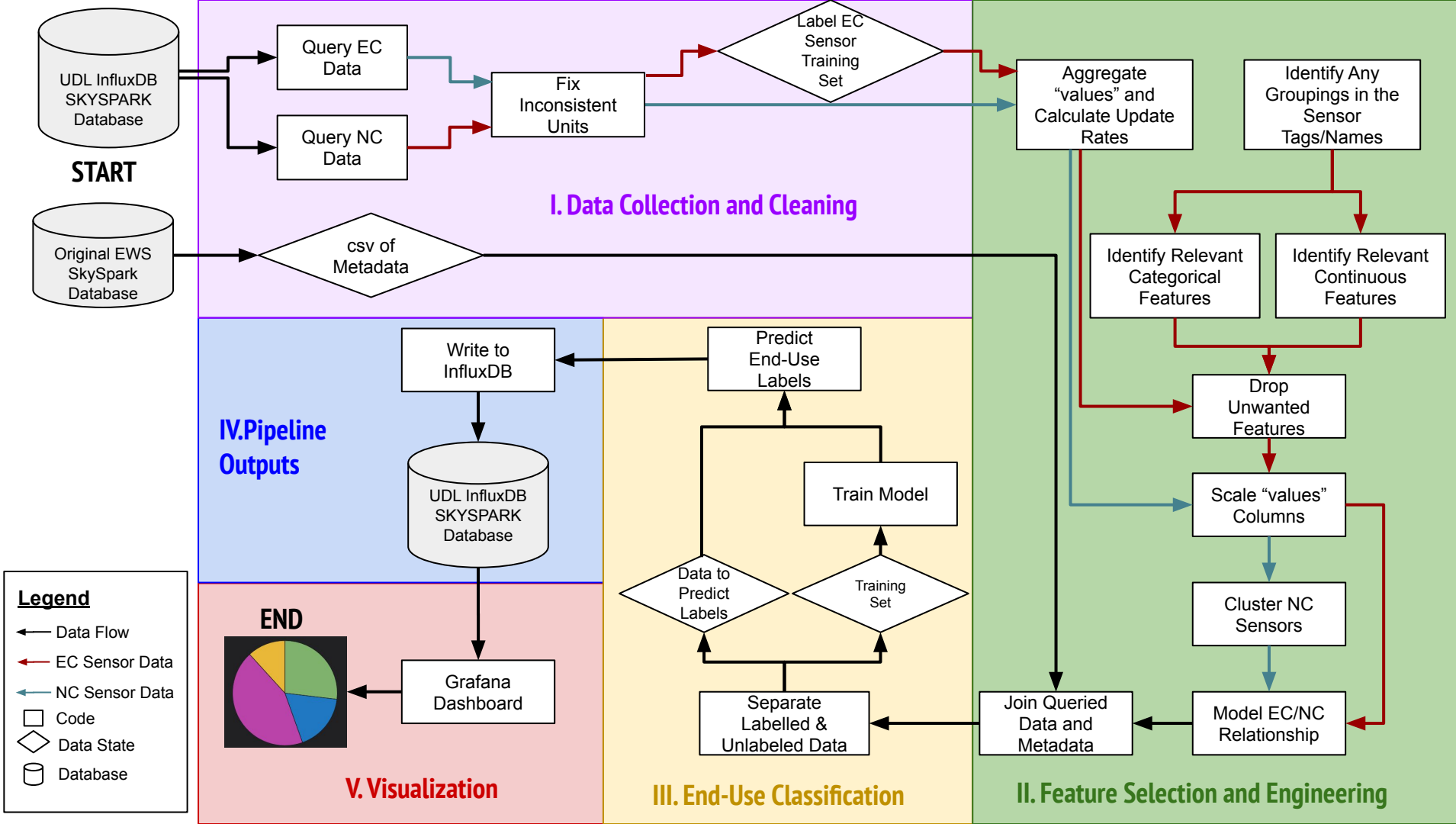



Function Flow & Status
Legend
Data Flow
EC Sensor Data
NC Sensor Data
Code
Data State
Database
Manual Activity/Decisions
In Progress
Not Started
Complete
Blocked
EC = Energy Consumption Sensor (gets an NRCan tag)
NC = Non-Energy Consumption Sensor (doesn't get an NRCan tag)
Mirrored SkySpark Database (InfluxDB)
Original EWS SkySpark Database
Data Collection and Cleaning
Query EC Data
Query NC Data
Fix Inconsistent units
Hand Label Sensors for Training Set (NRCan only)
Metadata from csv
End-Use Classification
Data to predict labels
Predict End-Use Labels
Decide which model to use (test various)
Separate labeled/ unlabeled data
Train/Test Sets
Train model
Pipeline Outputs
Write to InfluxDB
SkySpark Metadata Mirror? (InfluxDB)
Join queried data and metadata
Cluster NC Sensors
Decide which clustering method to use (test various)
Scale "values" columns
Model the EC/NC relationship
Identify relevant continuous features
Drop unwanted features
Determine period to aggregate "values" on (month/day/hr?)
Feature Selection and Engineering
Aggregate "values" and calculate update rates
Identify any groupings in the sensor tags/names
Determine which feature selection method(s) to use
Identify relevant categorical features

Function Flow & Status




Mirrored SkySpark Database (InfluxDB)
START
Original EWS SkySpark Database
Query EC Data
Query NC Data
Fix Inconsistent units
Label EC sensor training set
csv of Metadata
I. Data Collection and Cleaning
Aggregate "values" and calculate update rates
Identify any groupings in the sensor tags/names
Identify relevant categorical features
Identify relevant continuous features
Drop unwanted features
Scale "values" columns
Cluster NC sensors
Model EC/NC relationship
Join queried data and metadata
II. Feature Selection and Engineering
Separate labelled & unlabeled data
Data to predict labels
Training set
Train Model
Predict End-Use Labels
III. End-Use Classification
IV.Pipeline Outputs
Write to InfluxDB
Mirrored SkySpark Database (InfluxDB)
END
Grafana Dashboard
V. Visualization
Legend
Data Flow
EC Sensor Data
NC Sensor Data
Code
Data State
Database


UDL InfluxDB SKYSPARK Database
START
Query EC Data
Query NC Data
Fix Inconsistent Units
Label EC Sensor Training Set
Original EWS SkySpark Database
csv of Metadata
I. Data Collection and Cleaning
Aggregate "values" and Calculate Update Rates
Identify Any Groupings in the Sensor Tags/Names
Identify Relevant Categorical Features
Identify Relevant Continuous Features
Drop Unwanted Features
Scale "values" Columns
Cluster NC Sensors
Model EC/NC Relationship
Join Queried Data and Metadata
II. Feature Selection and Engineering
Separate Labelled & Unlabeled Data
Data to Predict Labels
Training Set
Train Model
Predict End-Use Labels
III. End-Use Classification
IV.Pipeline Outputs
Write to InfluxDB
UDL InfluxDB SKYSPARK Database
END
Grafana Dashboard
V. Visualization
Legend
Data Flow
EC Sensor Data
NC Sensor Data
Code
Data State
Database

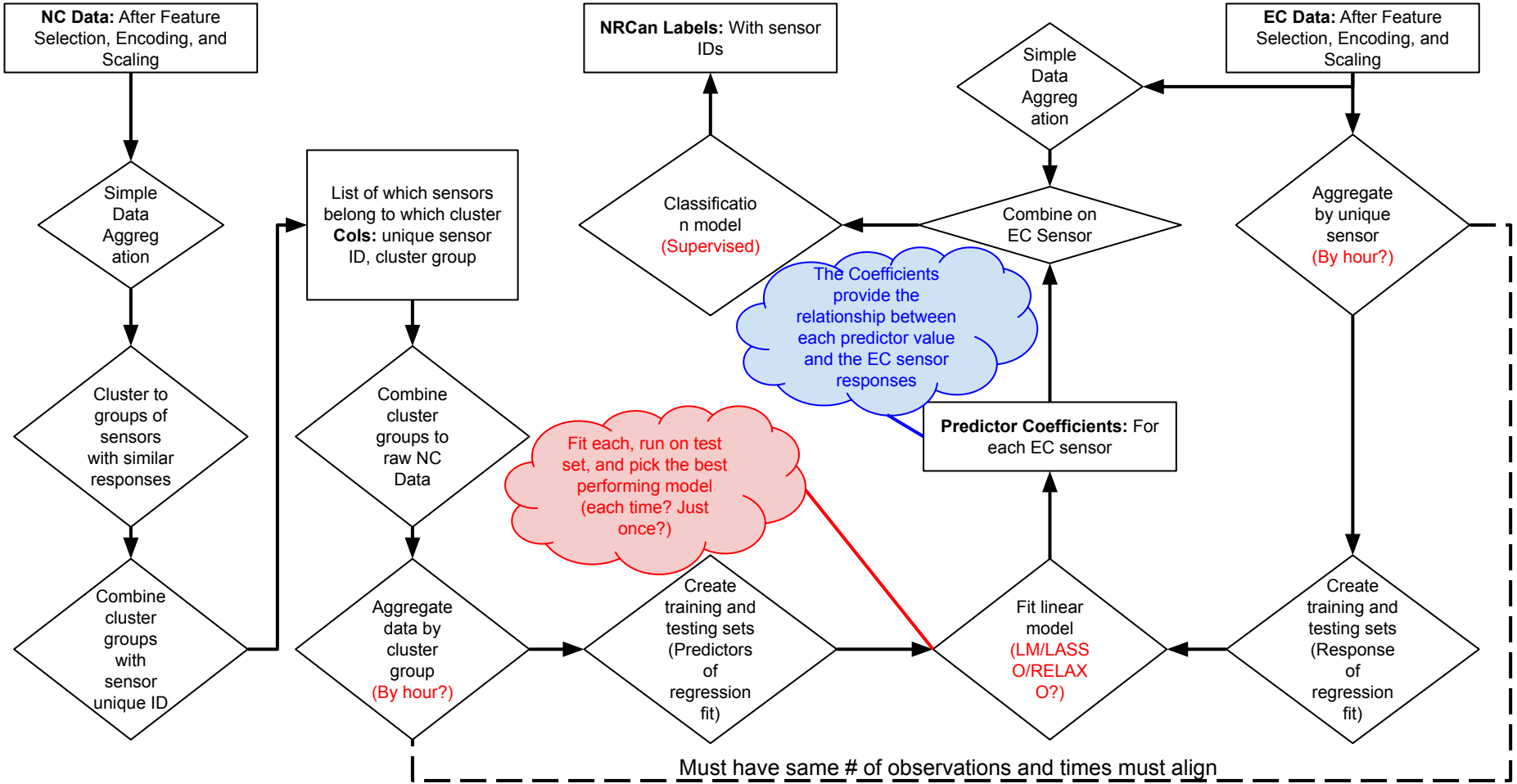
Write to InfluxDB Diagram (to copy/paste as picture into final presentation)

time	uniqueID	endUseLabel
2020-01-01	HW Submeters FM-3 Pharmacy Utilities Energy MV...	00_HEATING_SPACE_AND_WATER
2020-01-01	HW Submeters FM-4 Pharmacy Utilities Energy FM...	00_HEATING_SPACE_AND_WATER
2020-01-01	HW Submeters FM-6 Pharmacy Utilities Energy FM...	00_HEATING_SPACE_AND_WATER

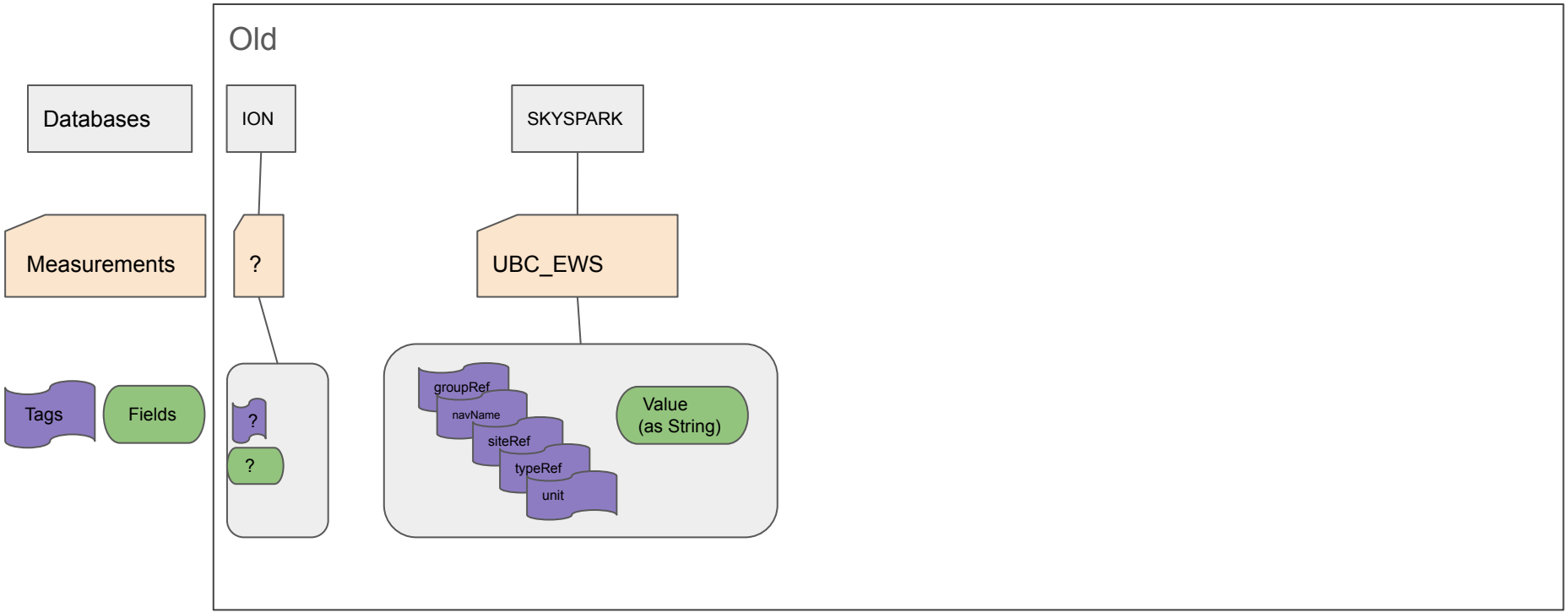
<Write with influxdb-python>

```
name: END_USE
tags: uniqueID=AHU-01 SF Air Systems Energy AHU1_SF_VFD_PWR(kWh)
time                endUseLabel
----                -----------
1577836800000000000 02_HEATING_COOLING_COMBINED

name: END_USE
tags: uniqueID=AHU-01 SF Pharmacy Air Systems Energy AHU1_SF_VFD_PWR(kWh)
time                endUseLabel
----                -----------
1577836800000000000 02_HEATING_COOLING_COMBINED

name: END_USE
tags: uniqueID=AHU-02 SF Air Systems Energy AHU2_SF_VFD_PWR(kWh)
time                endUseLabel
----                -----------
1577836800000000000 02_HEATING_COOLING_COMBINED
```

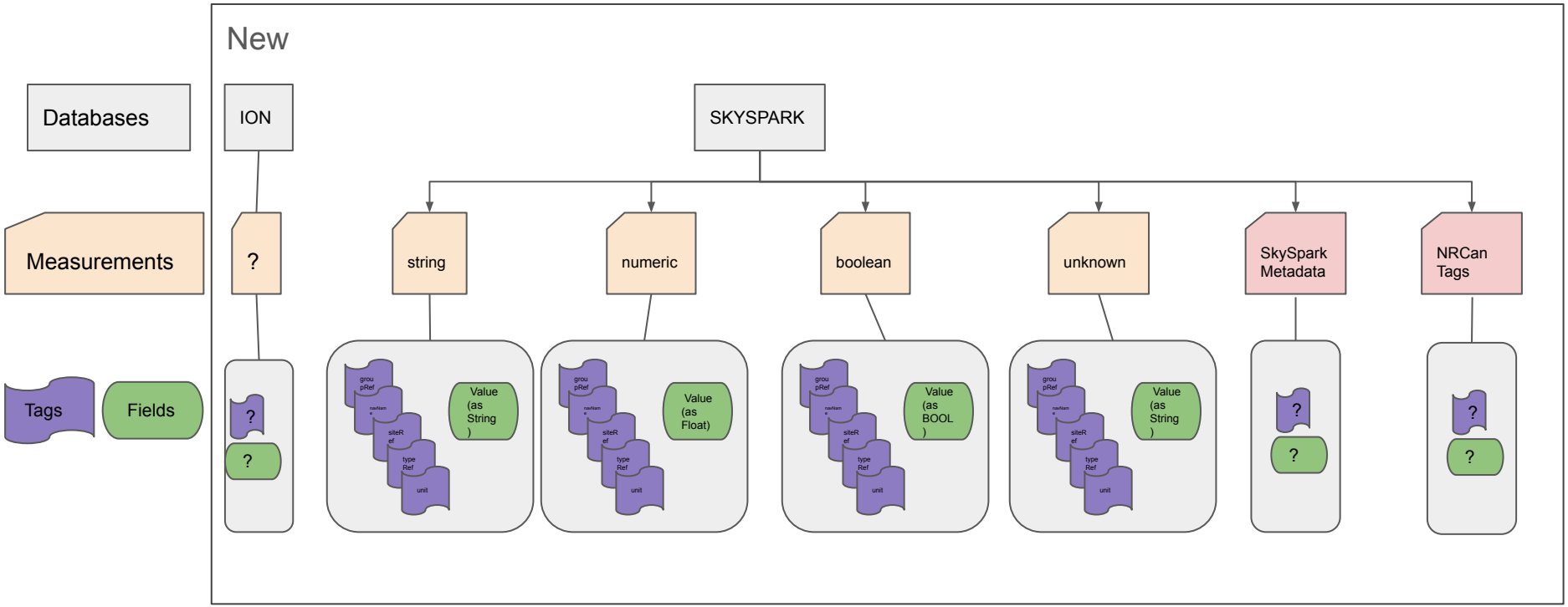
Concept for relating NC sensor changes to EC sensor values



Changes to InfluxDB (slide 1 of 2)

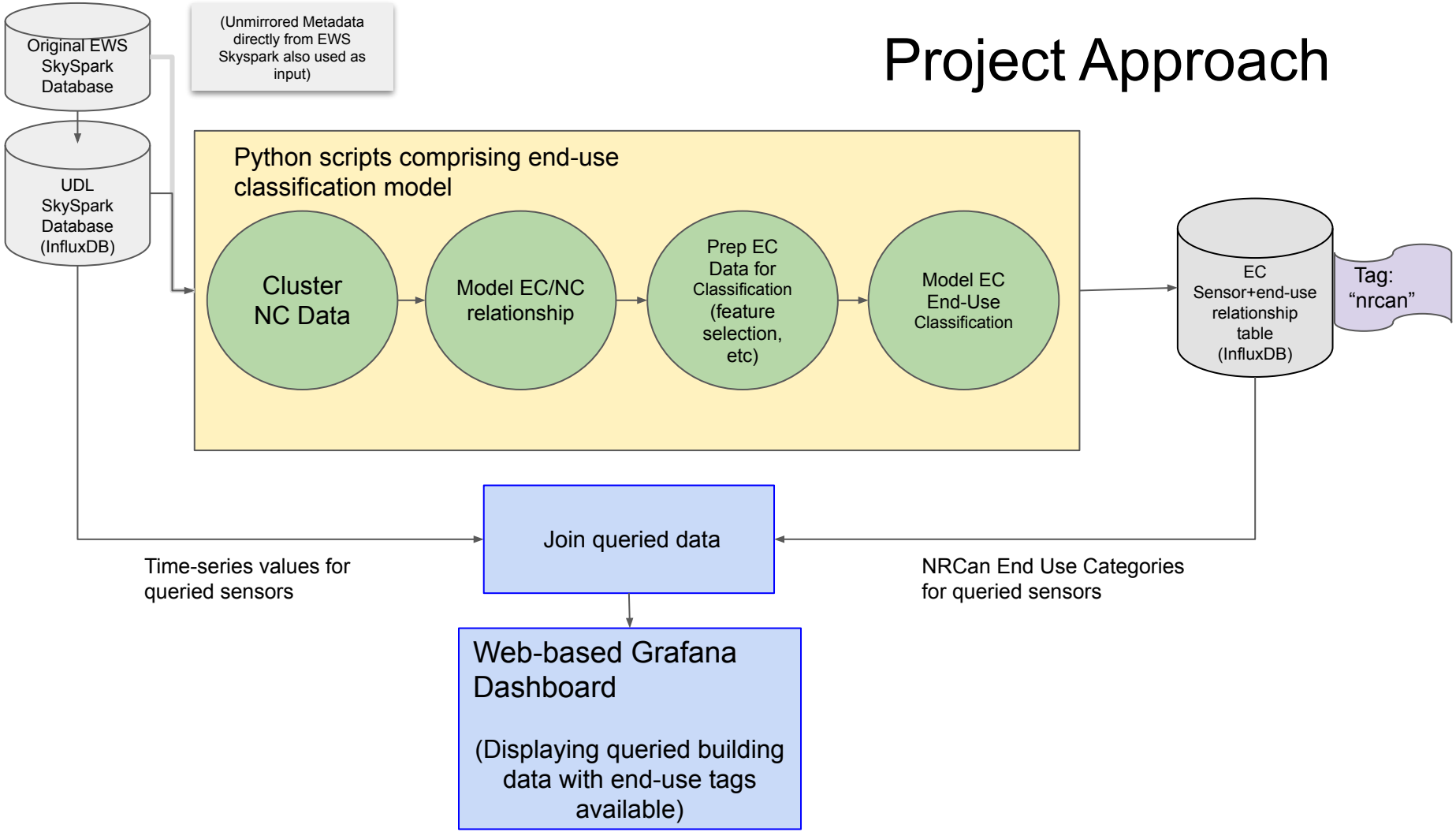


Changes to InfluxDB (slide 2 of 2)



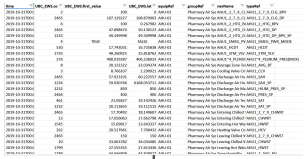
[Old/outdated slides follow]

Project Approach



Labeling

(why/what are we labelling?)



All ~5400 Pharmacy Building “sensors” from influxDB

Energy Consumption

Not Energy Consumption

Unknown + Not-Useful

This sub-label probably matters for greenhouse gas calcs

Gas

Elec

X number of categories that will group sensors that have similar relationships to each end-use power usage (Done with clustering? Manually labeled?).

Model

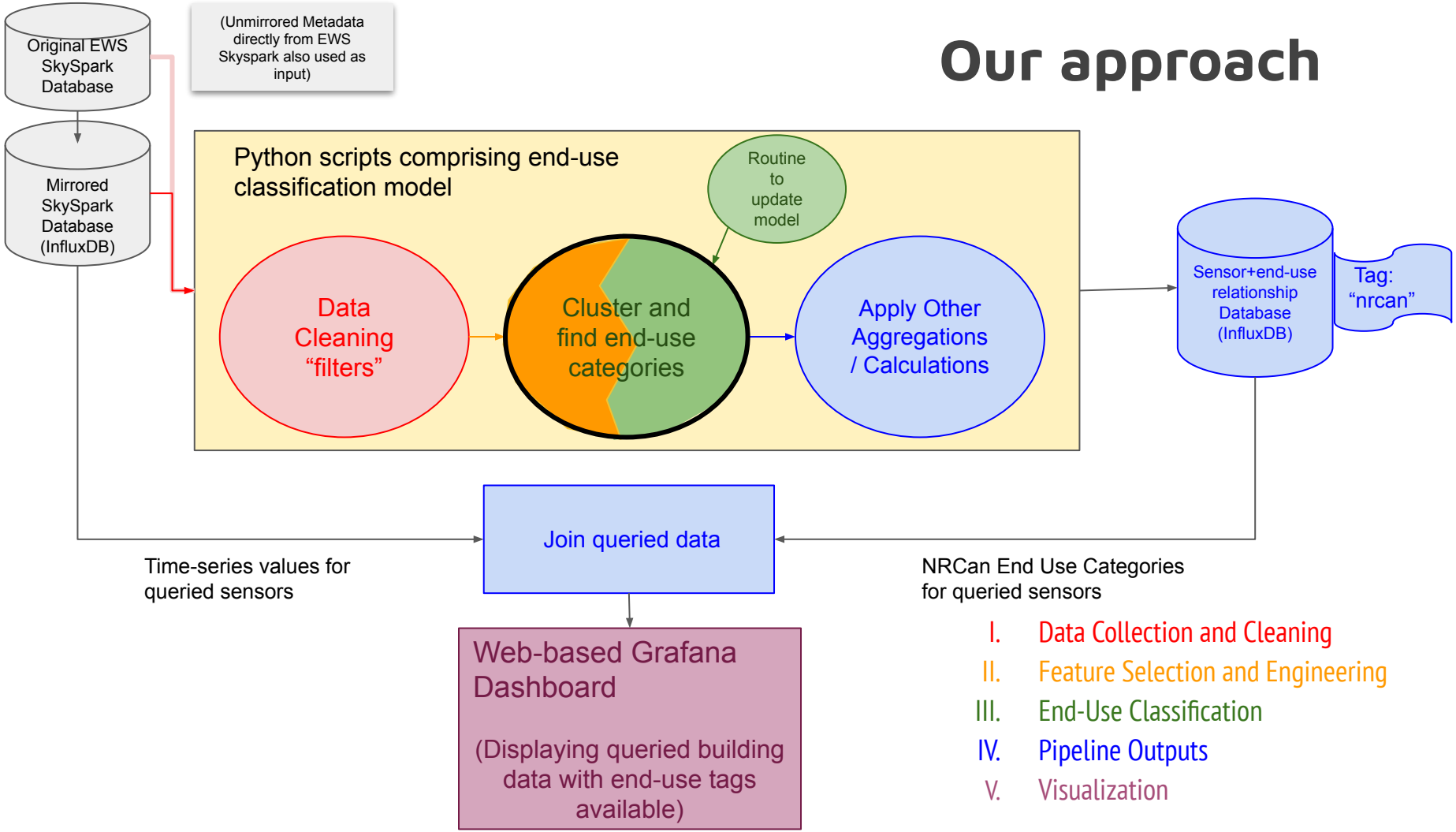
Known NRCan Tag

Unknown NRCan Tag

Table with <u>all</u> energy consumption sensor ID’s and their NRCan tags

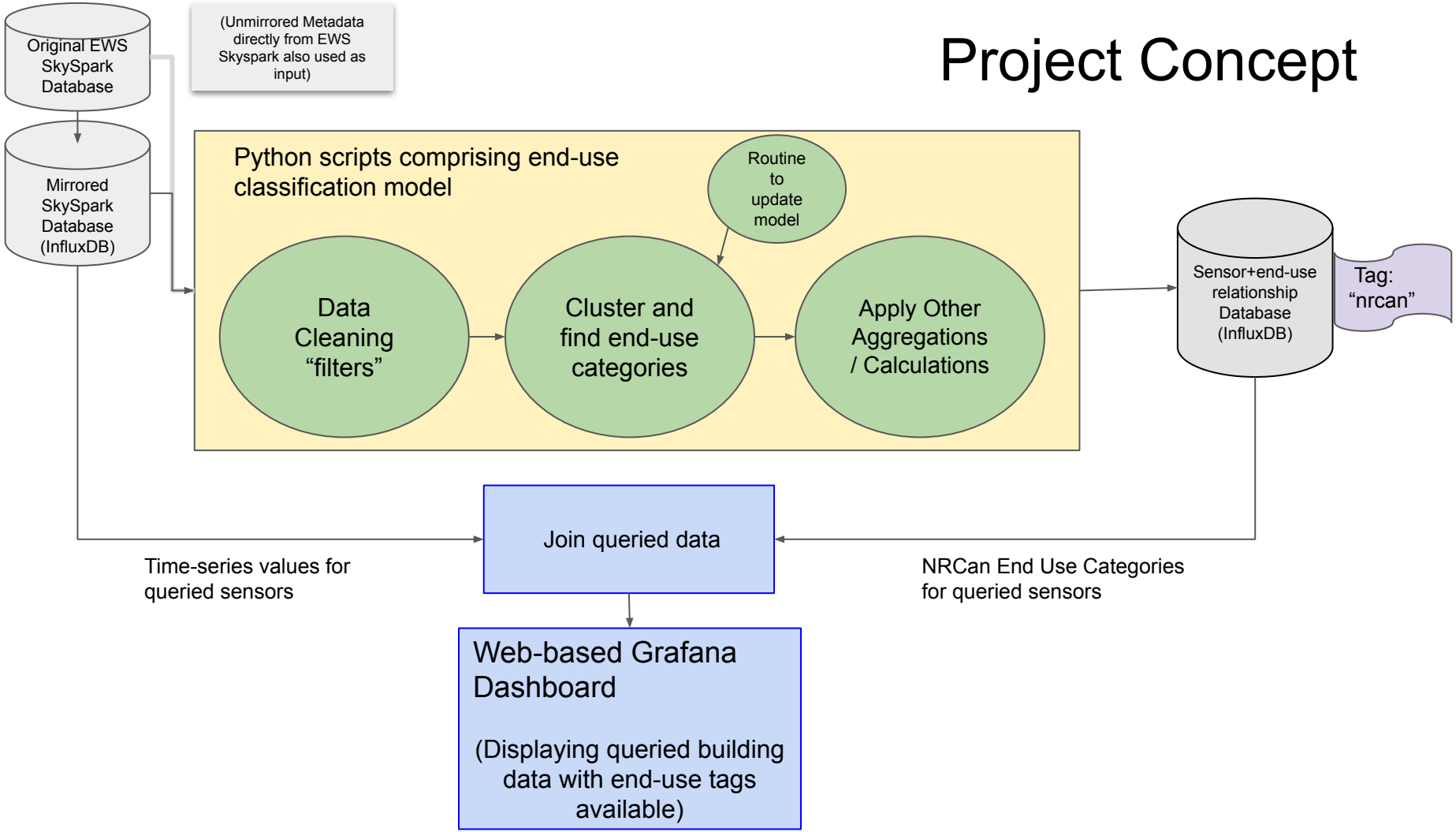
To visualization

Our approach



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

Project Concept



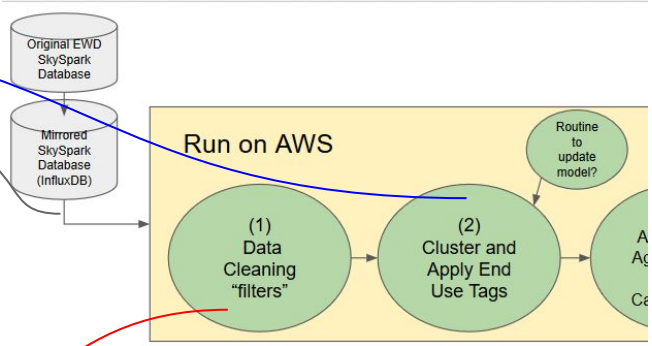
It might be useful to have a diagram kinda like this.

0. Query Data From UDL Skyspark database

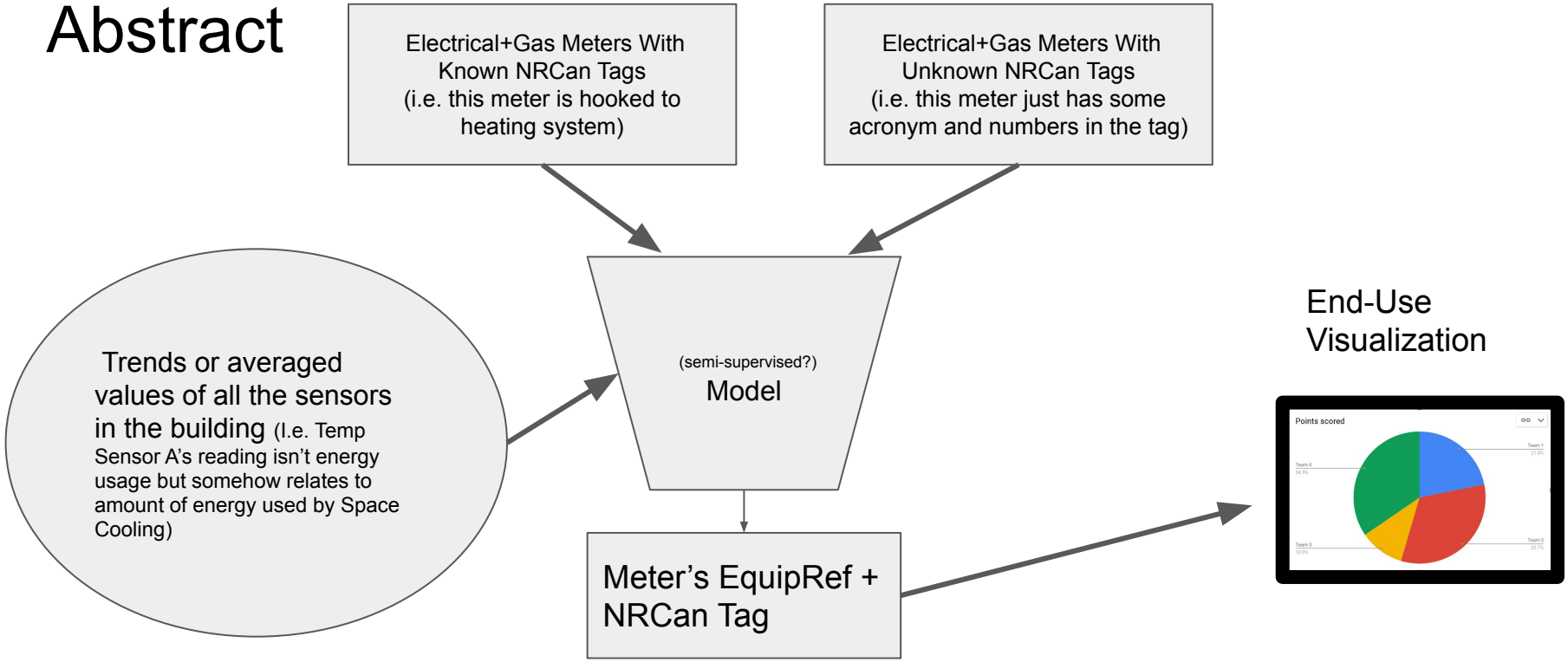
2. Apply Secondary End Use (SEU) tags (using available information such as typeRef,unit,value)

	equipRef	groupRef	navName	siteRef	typeRef	unit	value	nrcanSEU
2020-05-03 12:47:46-07:00	AHU-09	Pharmacy Air Systems	AHU9_RA_AFLW	Pharmacy	AHU9_RA_AFLW	ft/min	119.896248	001_ventilation
2020-05-03 12:47:46-07:00	Heating Plant SBLR-2	Pharmacy Hydronic Systems	SB2_FIRE_RATE	Pharmacy	SB2_FIRE_RATE	%	15.991011	002_heating
2020-05-03 12:47:46-07:00	Rm 3122 VAV-3S001	Pharmacy Floor 3	Zone Temp Effective Sp	Pharmacy	VAV_3S001_RT_SP	°C	21	002_heating
2020-05-03 12:47:46-07:00	Rm 3315 RAD-3-11	Pharmacy Floor 3	Zone Temp	Pharmacy	RZ3_11_SLAB_T1	°C	-	
2020-05-03 12:47:46-07:00	Rm 6321 VAV-6S074	Pharmacy Floor 6	Discharge Air Flow	Pharmacy	VAV_6S074_FLW	L/s	-194.79935	
2020-05-03 12:47:47-07:00	Heating Plant SBLR-1	Pharmacy Hydronic Systems	SB1_2_FGAS_INLET_T	Pharmacy	SB1_2_FGAS_INLET_T	°C	273.521576	002_heating
2020-05-03 12:47:47-07:00	Heating Plant SBLR-1	Pharmacy Hydronic Systems	SB1_2_FWT_T	Pharmacy	SB1_2_FWT_T	°C	45.976299	002_heating
2020-05-03 12:47:47-07:00	LEF-3	Pharmacy Air Systems	Exhaust Air Flow	Pharmacy	EF3_AFLW	ft/min	450.231384	001_ventilation

1.Clean Data (remove missing& incorrect data)

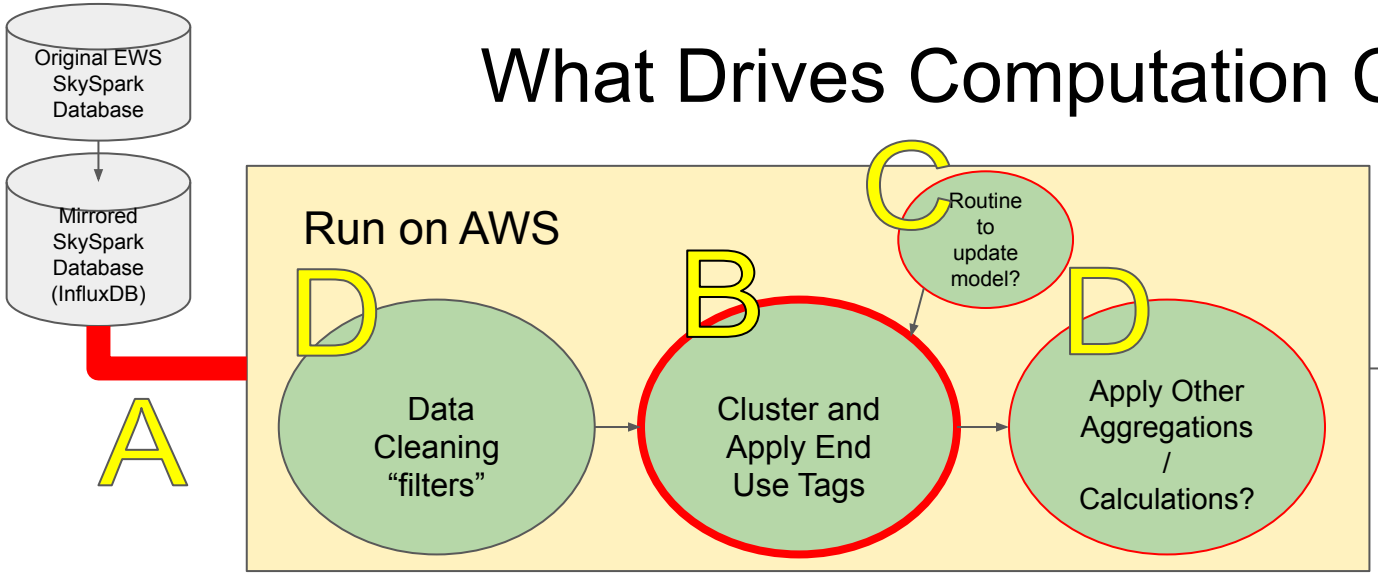


Abstract



Using the sensors as predictors for NRCan tags of meters is hard enough… but what happens if an electric meter with an unknown tag actually has part of its power used for space heating, part for space cooling, and part for water heating? What if the split between those categories is dynamic?

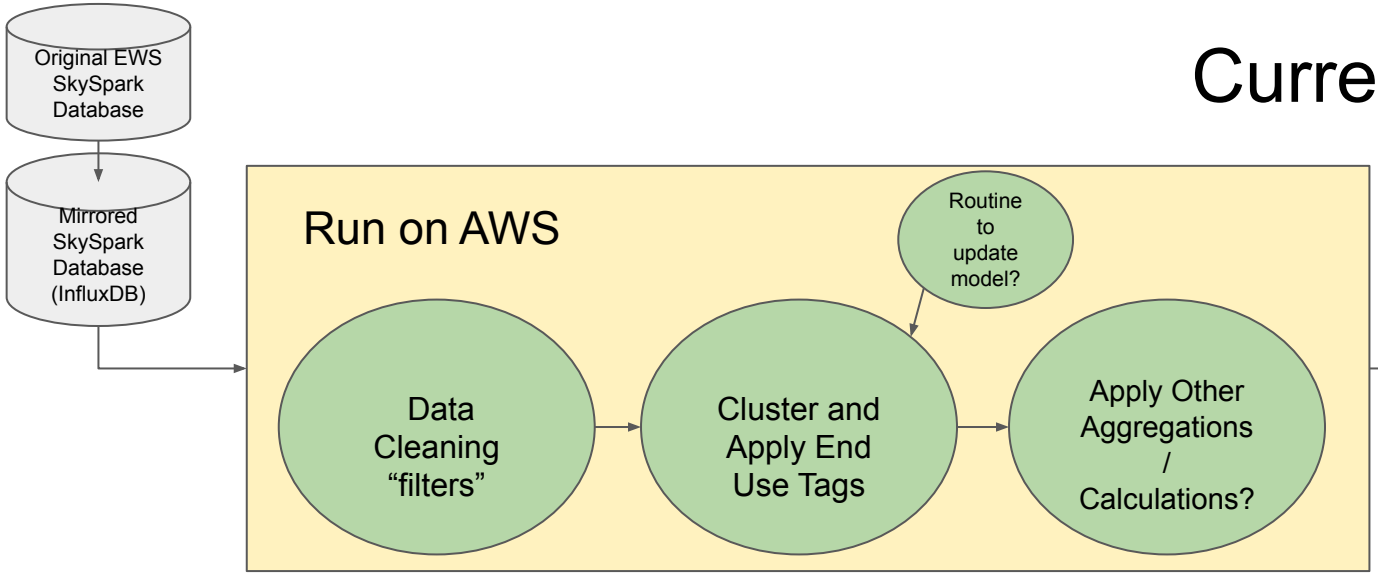
What Drives Computation Complexity?



A. What's the maximum query size we have to deal with (how many buildings? How many sensors? Does time frame matter?)
B. How complex is the clustering/end-use algorithm?
C. Likely only need to run occasionally and maybe even completely separate compute instance from the "streaming" process to Grafana.
D. Probably low impact but may need to be distributed (i.e. Spark) if (A) is large.

Possible Pipeline Designs

Current Concept



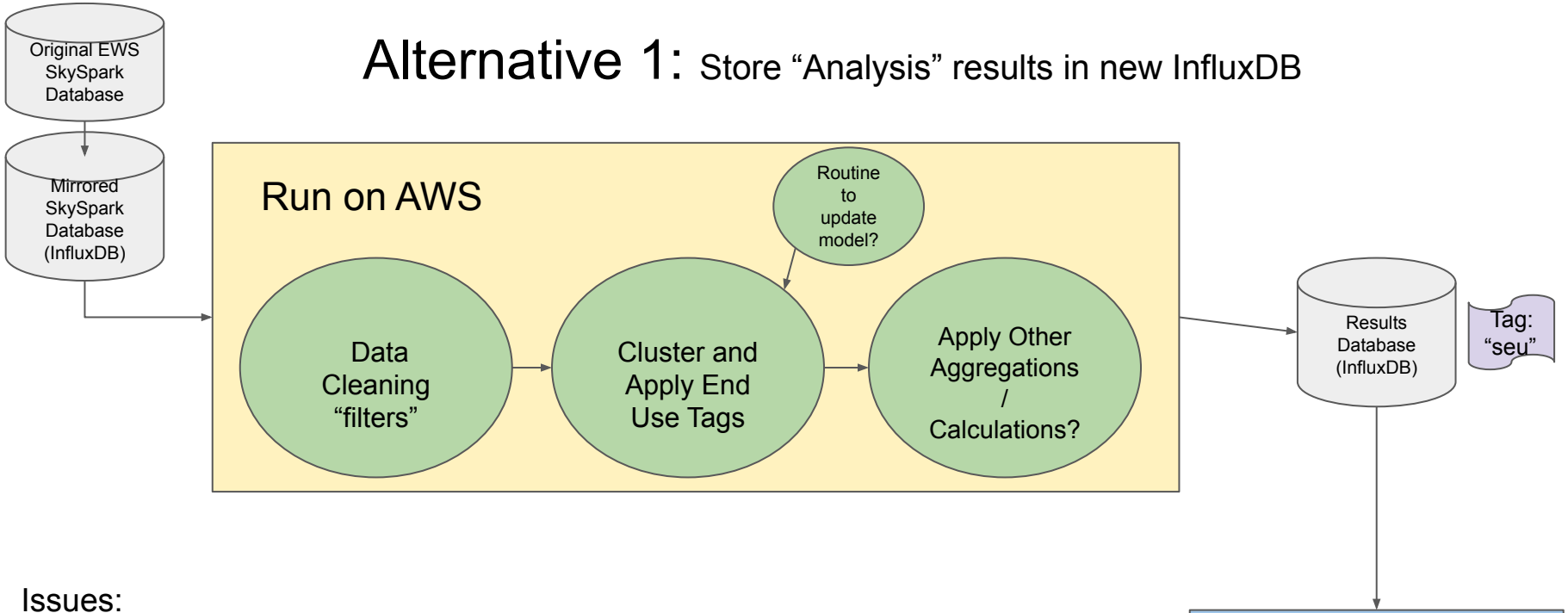
Issues:

1. Inefficient: Why recalculate SEU categories on data when it doesn't change?The specifics of the query run for visualization have no impact on the actual SEU tag values. Querying SensorA for 1 week will yield the same SEU for SensorA as it would if you query 2 months of data. Querying SensorA+SensorB will yield the same SEU for SensorA as when only querying SensorA.
2. Haven't found example of this pipeline connecting Spark directly to Grafana

Web-based Grafana Dashboard

(Displaying queried building data with end-use grouping available)

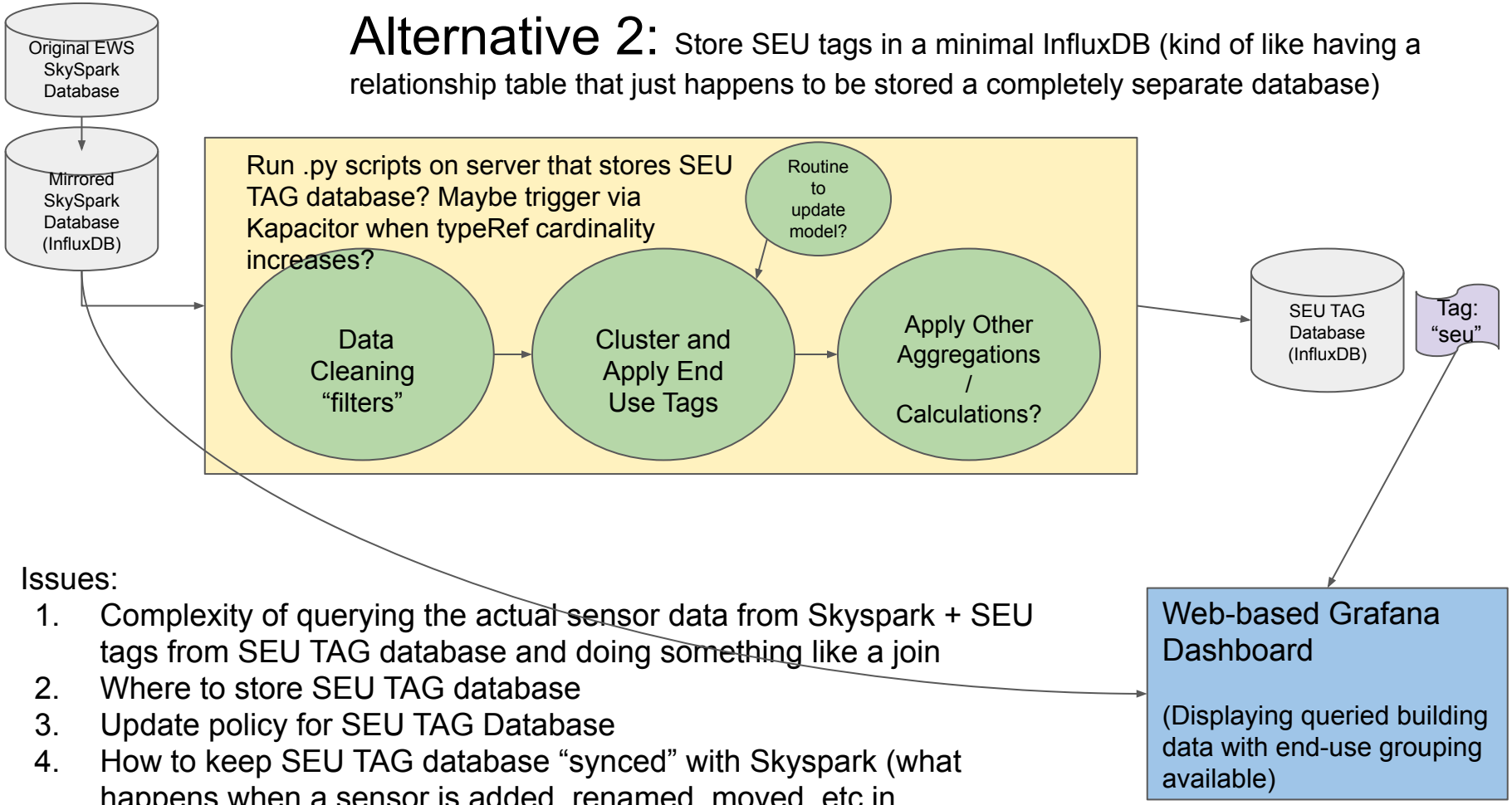
Alternative 1: Store “Analysis” results in new InfluxDB



Issues:

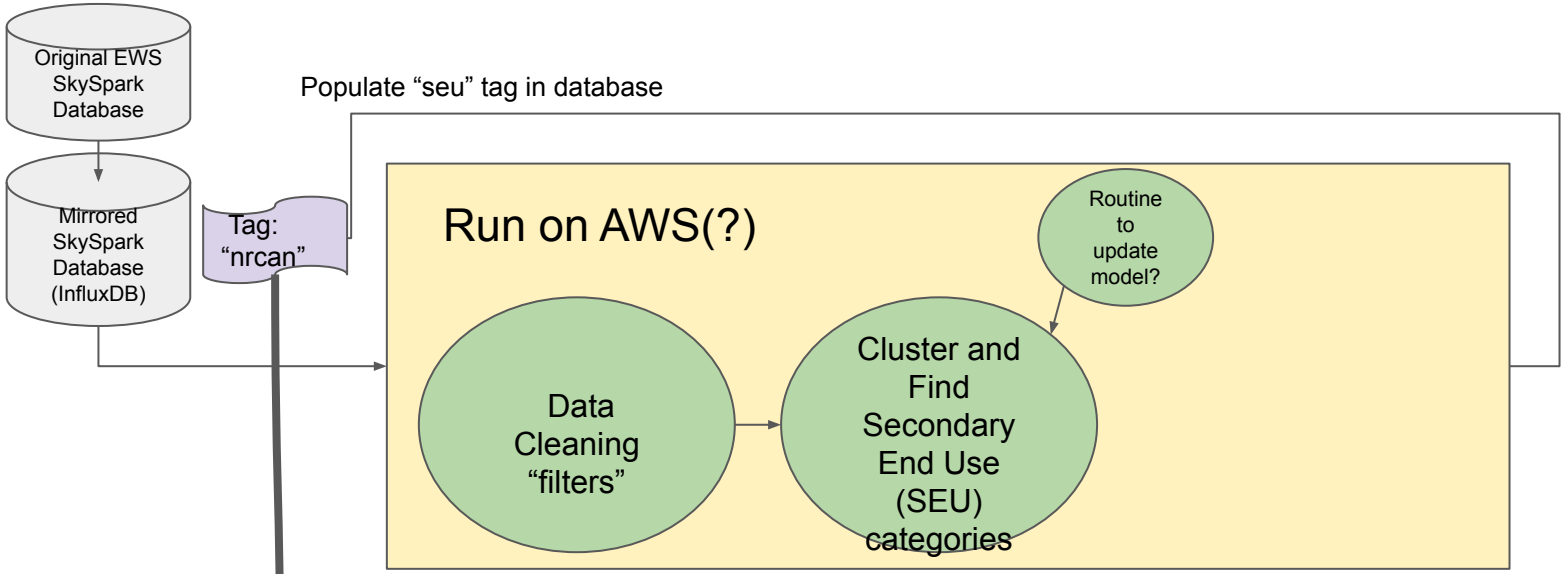
1. Where to store results database?
2. Results Database will basically need to be a mirror of Skyspark Database? Or will retention policy only store the last few queries of info?
3. Update Policy for data (including SEU tag) in Results Database?
4. If two queries are run on the same data with overlapping time windows...what happens? Lots of duplicated information in Results Database?

Alternative 2: Store SEU tags in a minimal InfluxDB (kind of like having a relationship table that just happens to be stored a completely separate database)



Issues:

1. Complexity of querying the actual sensor data from Skyspark + SEU tags from SEU TAG database and doing something like a join
2. Where to store SEU TAG database
3. Update policy for SEU TAG Database
4. How to keep SEU TAG database "synced" with Skyspark (what happens when a sensor is added, renamed, moved, etc in Skyspark?)

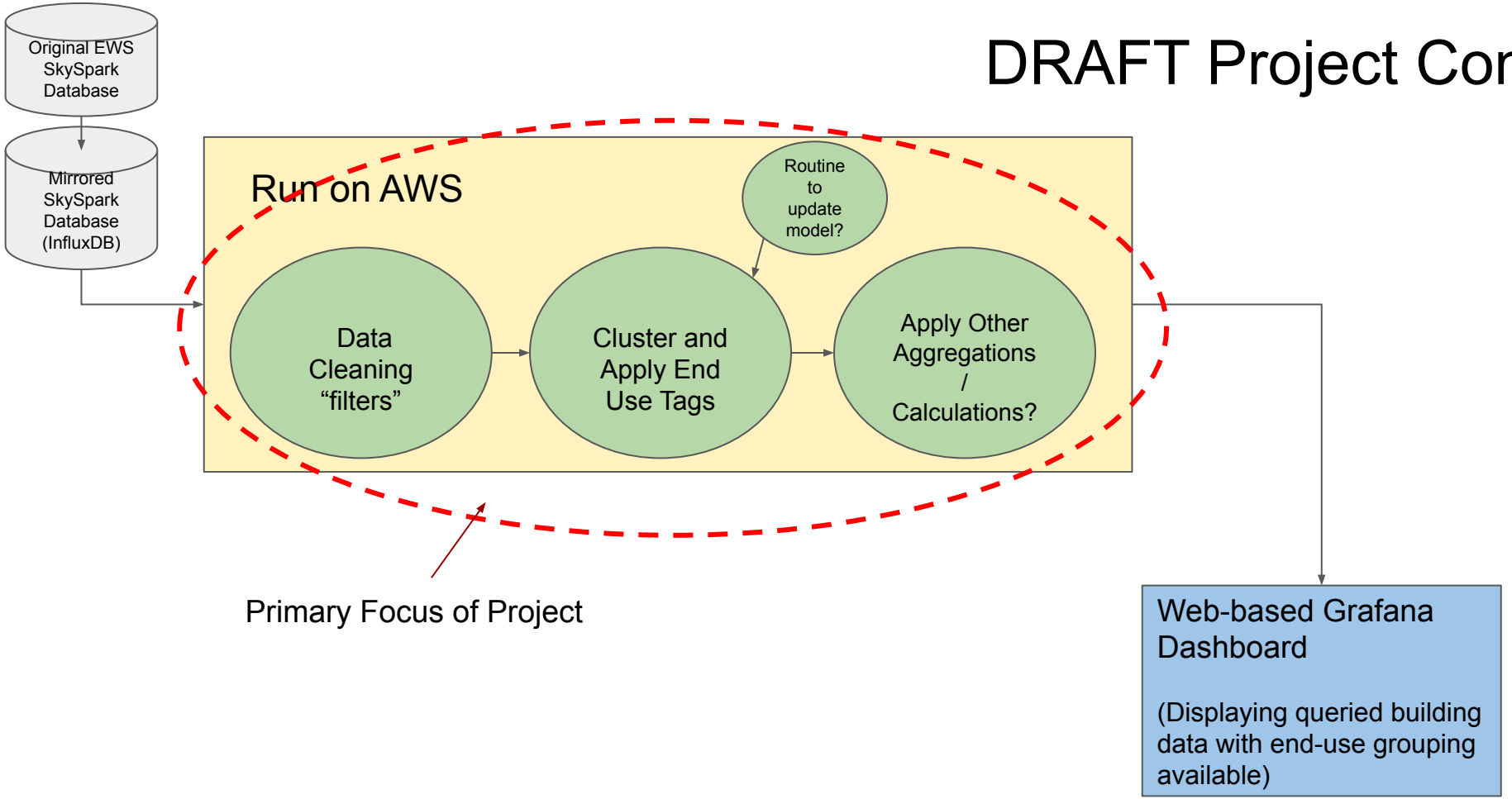


Issues:

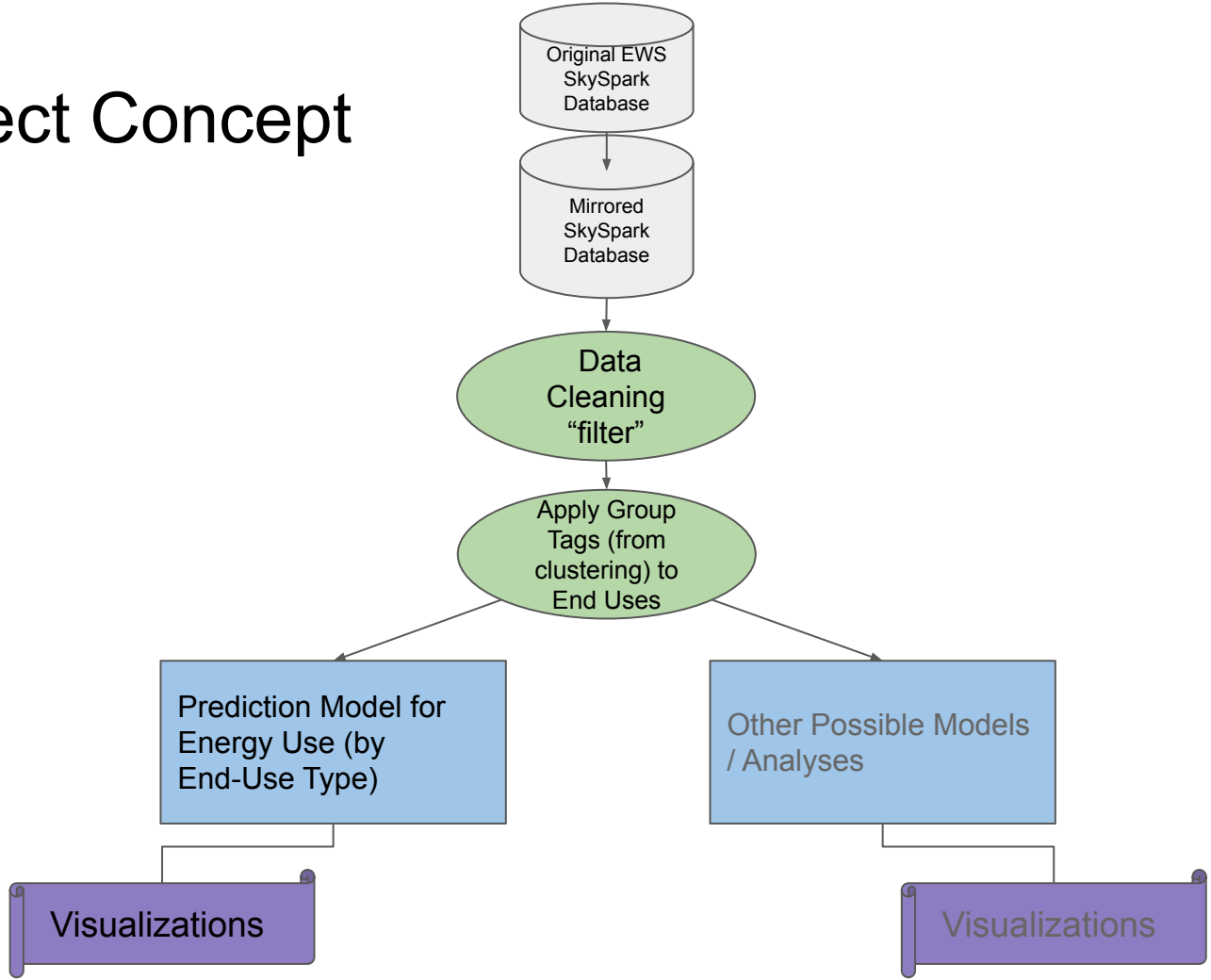
1. Need to be able to write results to Skyspark mirrored database
2. Need to consider how to handle updating the SEU tags when model is improved or new sensors added.
3. Will expand size of Skyspark Database

Alternative 3: The "cleanest" option

DRAFT Project Conc



Project Concept



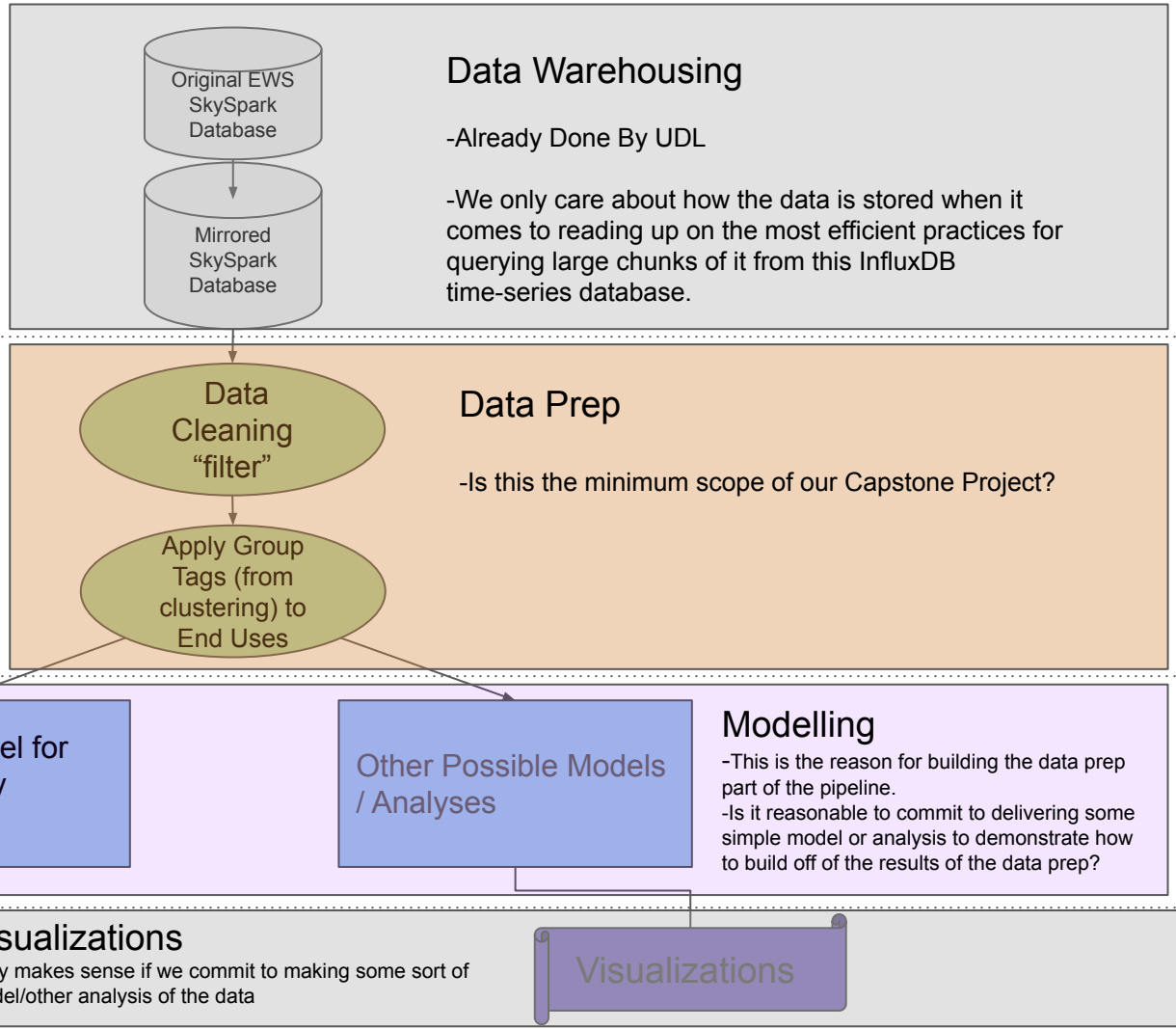
Project Scope

Already Done

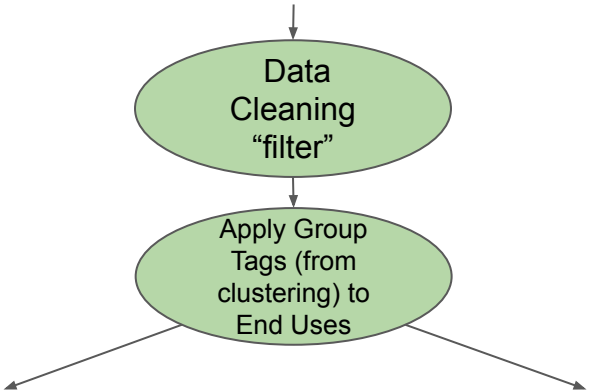
Minimum Requirement?

Reasonable to Include?

Bonus?



Thoughts on Data Prep ‘layer’



Possible way to develop this:

1. **Initially develop cleaning in Python+Pandas on a smaller dataset**
   a. Data cleaning steps written in a jupyter notebook - come up with generalized methods we can apply to data from any building.
2. **Use results to develop classification model AND start working on more scalable method.**
   a. Python notebook again for developing classification model using cleaned data from pt1.
   b. At the same time can be working to scale up methods developed in pt1 to apply to much larger datasets (i.e. using Spark).

<u>If we then want to do an analysis/create a model, when could we start working on that?</u>
Example: predicting energy use by end use type and outside temperature.

Could someone be working on that as soon as 1A is done and just create fake end use type groups as placeholders in their input data?

Feature Selection