

# AutoML COMP Ranker

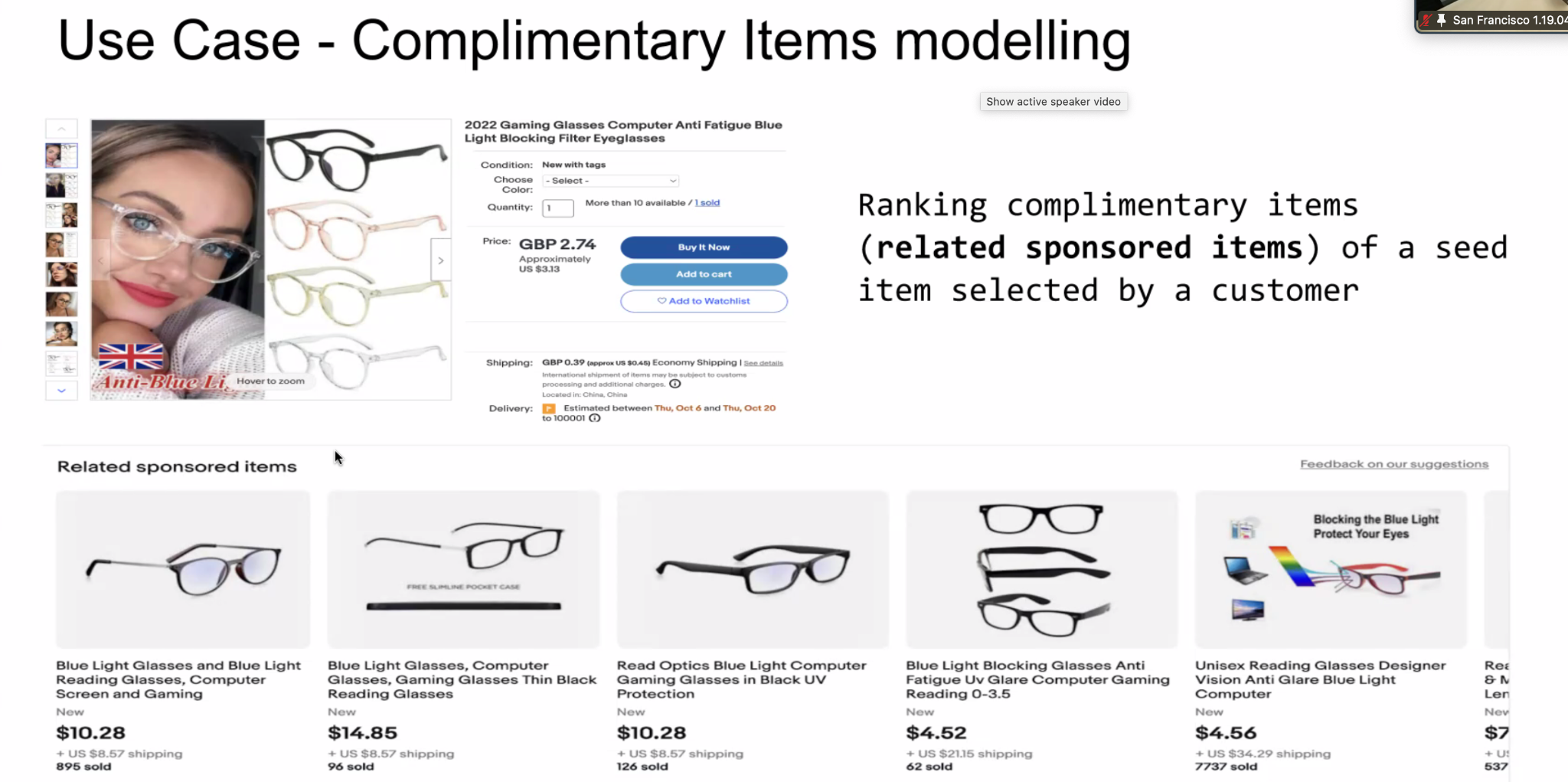
Ads Cloud Team

Merch public cloud team. Discover scalable opportunities.

M1 - focus on merchandising use cases

AutoML Tabular

Complimentary items modeling



Dataset

Training: 22M

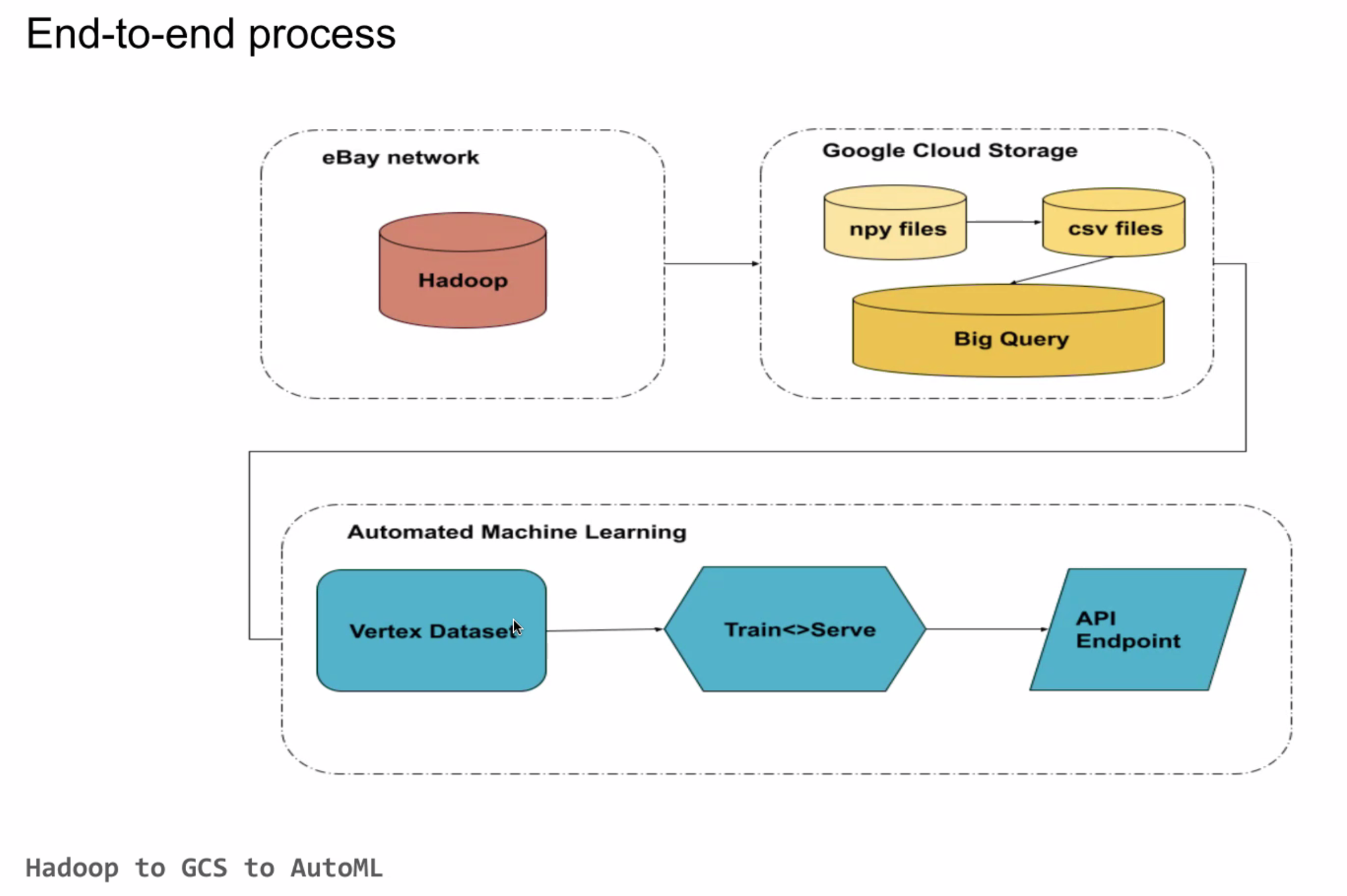
Validation: 5M

Test: 3M

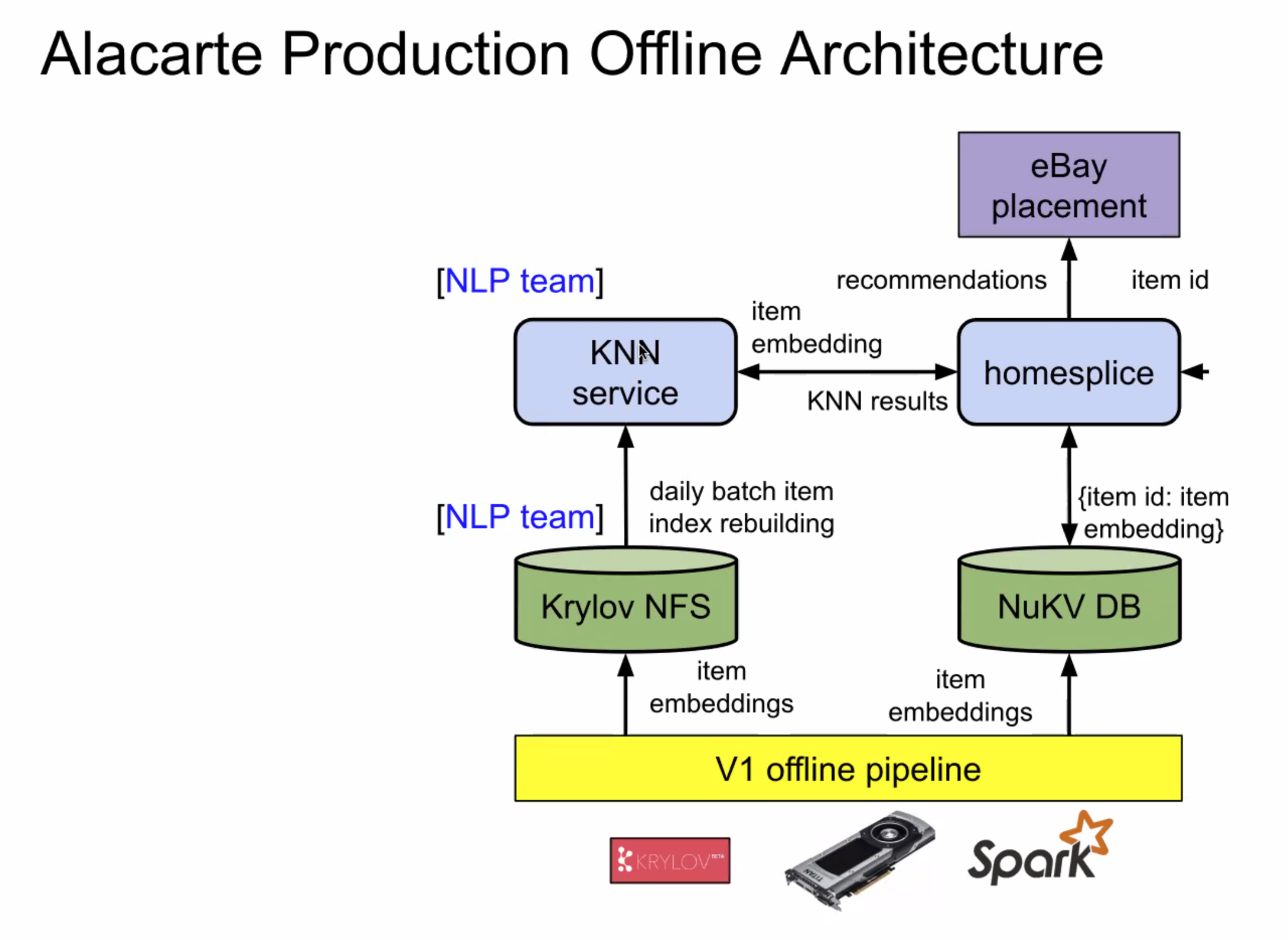
No. of model features: 77

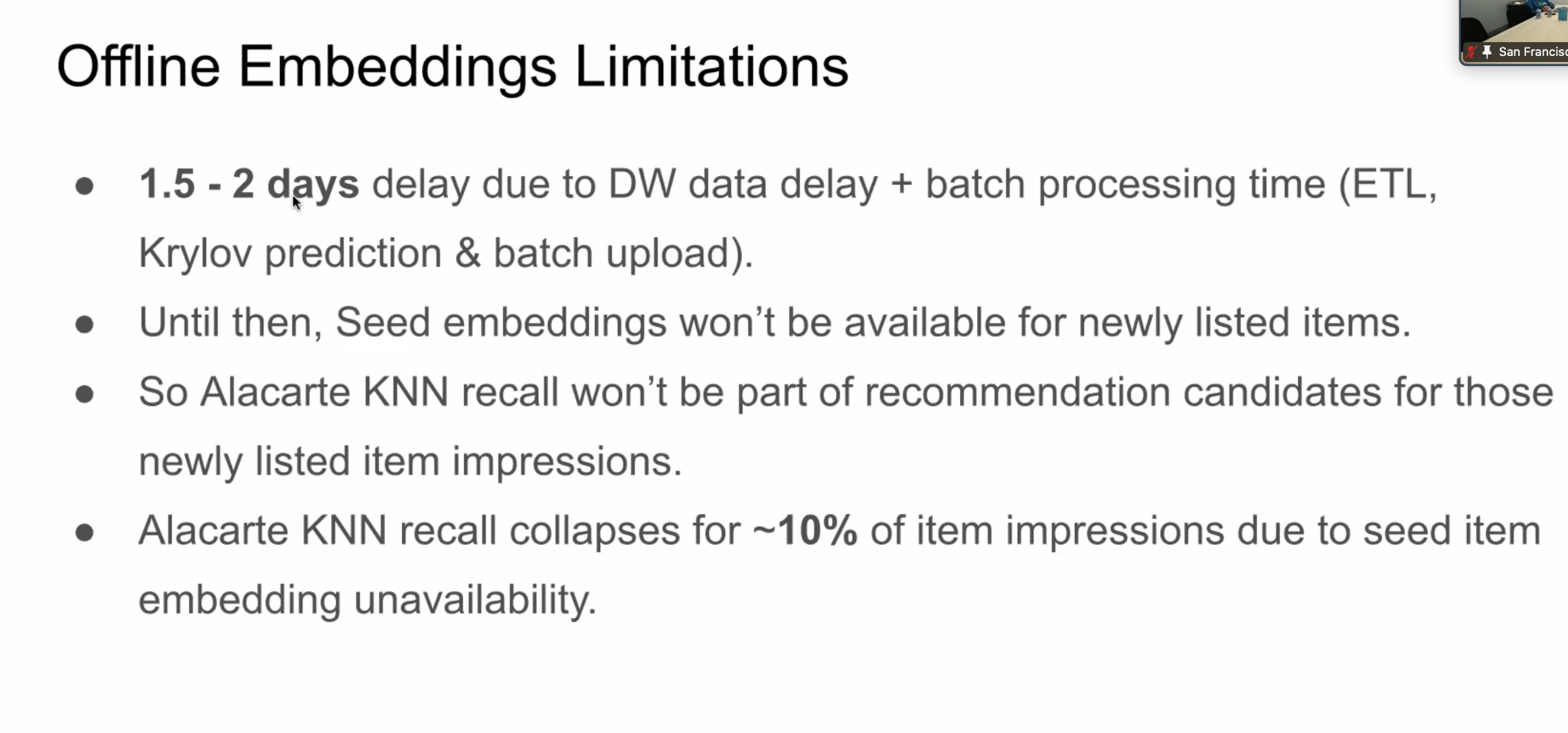
labels : no-click (0), click (1), purchase (2)

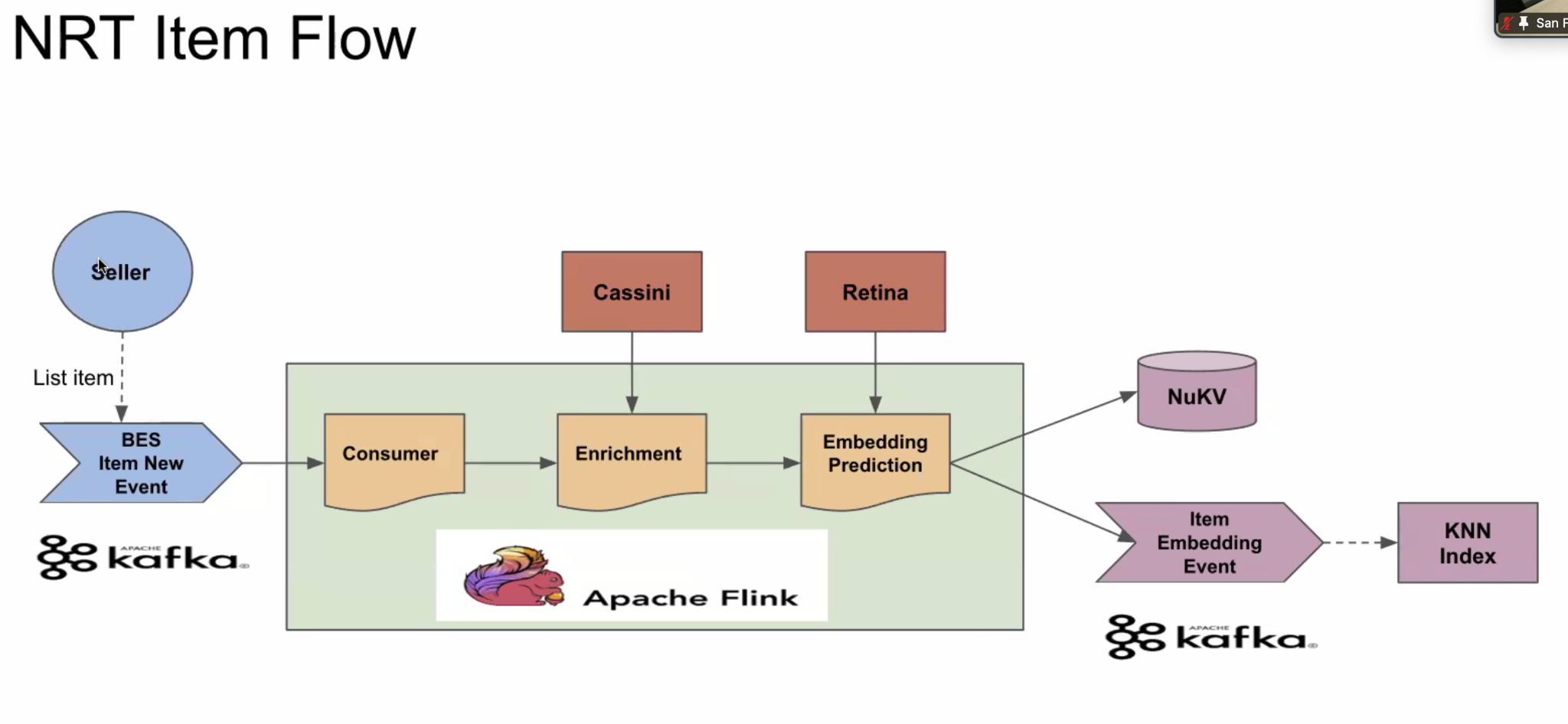
* Distribution of labels
  + 4K rows of class 2
  + 3M rows of class 1
  + 18M rows of class 0



# NRT Item Flow







# NRT monitor

Not creating business. But trying to reduce business loss. Mostly due to a site impacting change.

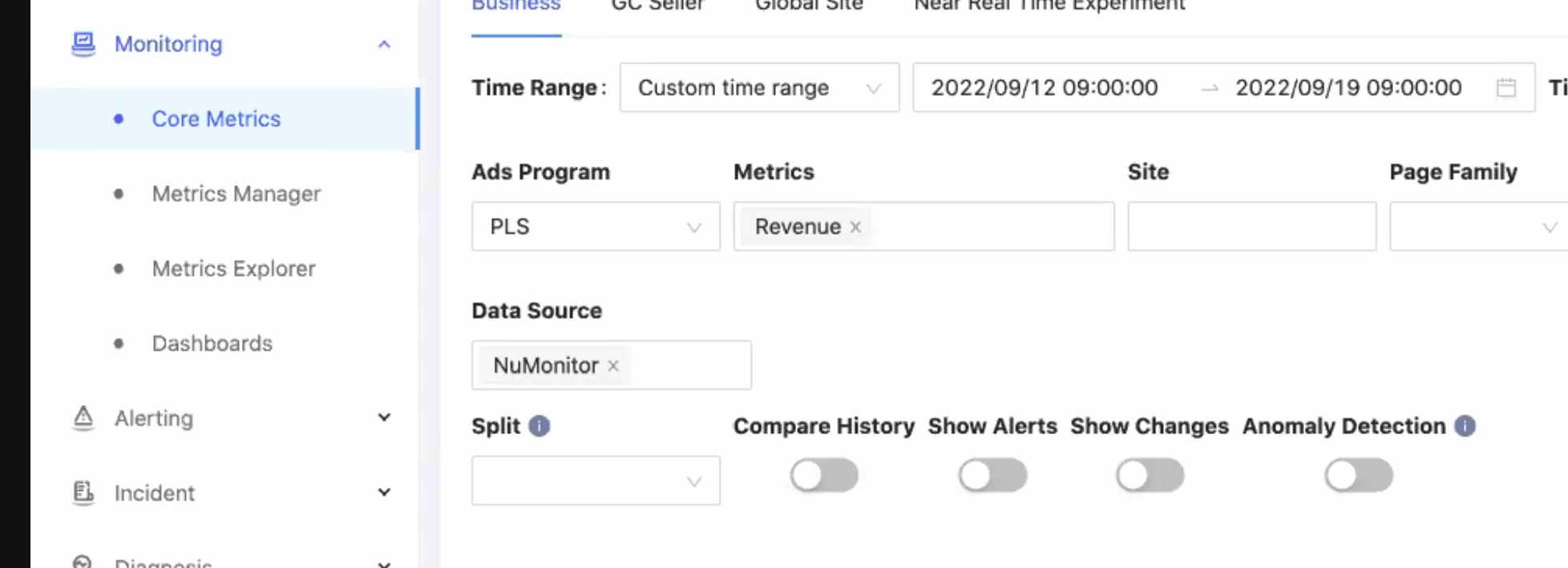
Dimension is first step to narrow down.

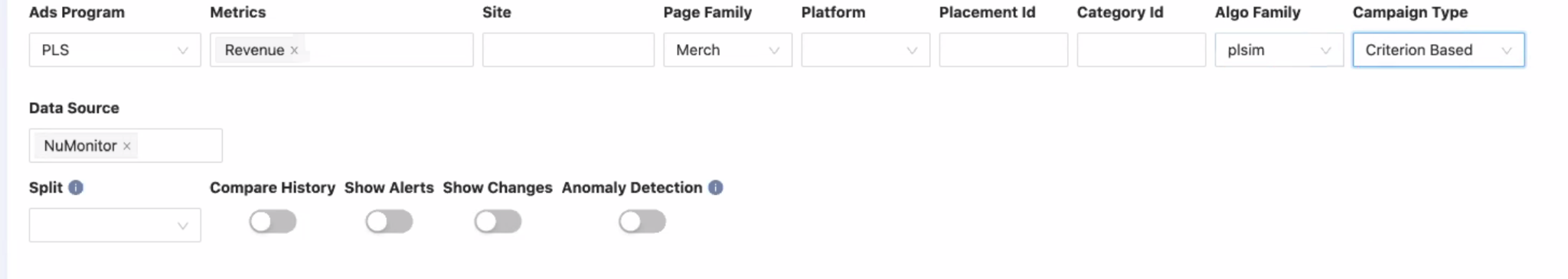
If we add new columns in sellerhub, it will cause long joins and take forever. If we add to sherlock metrics, it will cause cardinality problems.

Hence, we added nukv

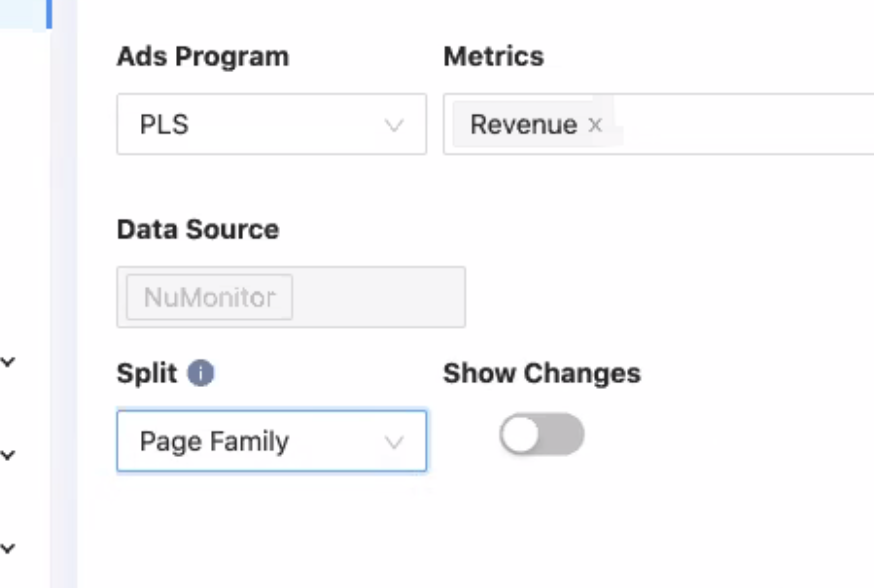
NuMonitor data source

NRT + PL Sales event. Completely decoupled data. No impact on sellerhub performance.





Compare for page family performance - in real time



Drill down - breaks down delta by dimension

Julie Cheng to Everyone (11:52 AM)

overall question for control center is that some metrics being shared by analytics are some time not aligned. could we get an diff inventory?

so we can close the gap?

Zhiyu Ding to Everyone (11:55 AM)

Julie - yes we always cross check with MADCLD and data used by the Analytics team to ensure data quality

Julie Cheng to Everyone (11:56 AM)

thanks. recently we discovered that rev calc using different currency conversions. just wondering if we should do another one time deep dive to discover any other discrepancies so not addressing them as escalations.

Weiyu Zhou to Everyone (12:00 PM)

@Julie, the main difference is caused by the exchange rate. Almost every where else, we are using Planed rate based on the analytics/business team except CPC for the ClickHouse which is using the real time exchange rate. We can chat more offline.

Weiyu Zhou to Everyone (12:00 PM)

@Julie, the main difference is caused by the exchange rate. Almost every where else, we are using Planed rate based on the analytics/business team except CPC for the ClickHouse which is using the real time exchange rate. We can chat more offline.

Pavel Liu to Everyone (12:03 PM)

we're implementing micro services to get SellerHub - NuMonitor parity, we can add Analytics (DSS) for sure

Pavel Liu to Everyone (12:11 PM)

SHEPHERD-1675 parity check between Anaytics data (such as MADCDL) and NuMonitor

filed @Julie

existing parity check SHEPHERD-1257 Create metrics for data quality/parity between SellerHub Clickhouse and NuMonitor

already done but need to support more "dimensions" to narrow down, say diff regarding site/placement/AlgoFamily

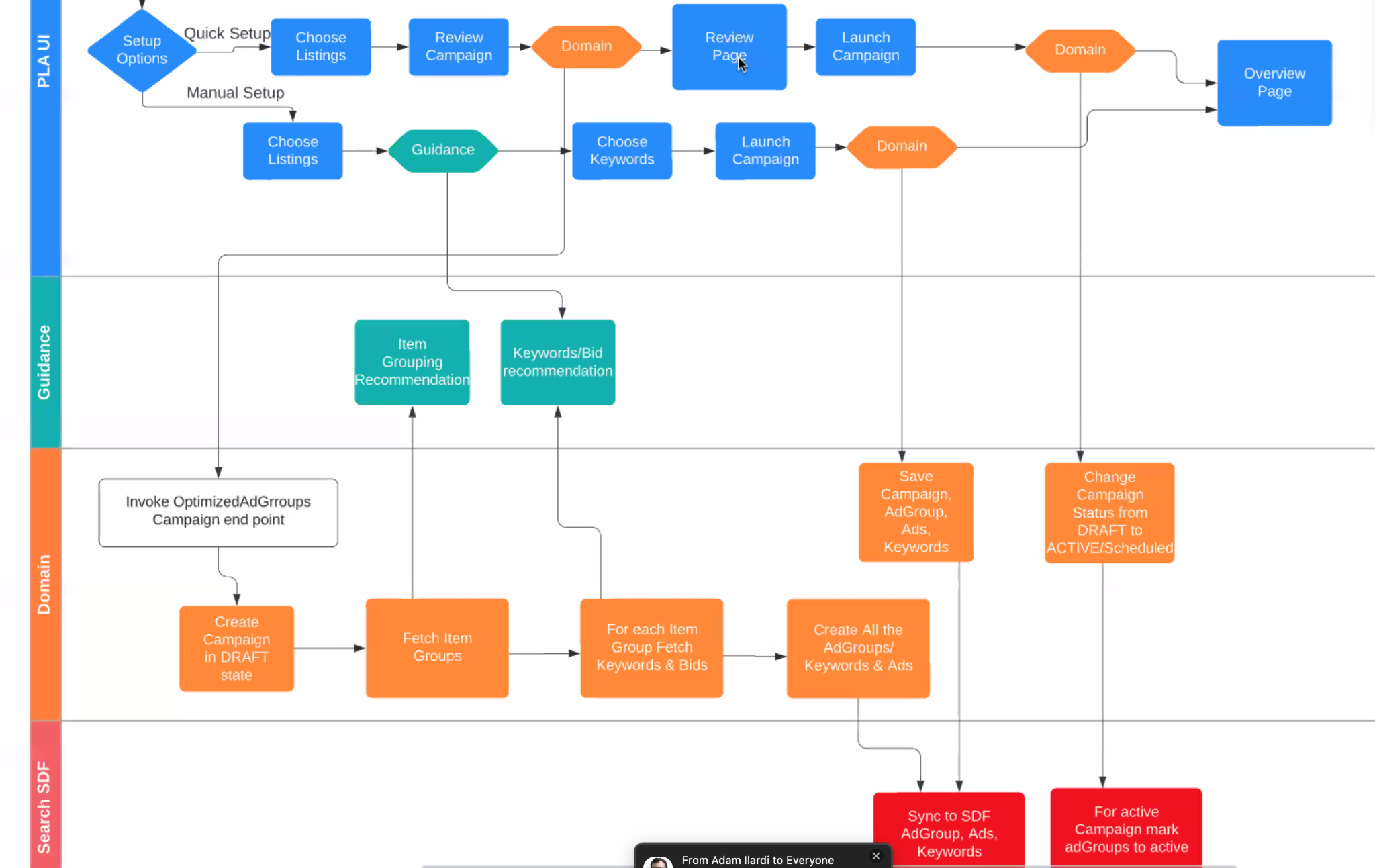
x over x can go back month, and year as well?

Manohara - yes we plan to retain the data for months and even over a year if capacity is not a problem. That will help with seasonality analyses.

Xt tag works for whole event level

“Search” placement not yet available. Planned.

# Ad Group Optimizer - launch’’



# CCOE Sprint Demo

Adam Ilardi to Everyone (5:07 PM)

where are you storing the data from the buyer interactions?

Lili Weng to Everyone (5:07 PM)

Will store the user dimension interactions data in K-V database

Marshall Hansi Wu to Everyone (5:08 PM)

We plan to store in it NuKV, in the future it would be on top of the NRT feature platform

Adam Ilardi to Everyone (5:08 PM)

This must be a ton of data?

Lili Weng to Everyone (5:08 PM)

Estimated data size will be ~500G.

Can refer to design doc https://docs.google.com/document/d/1mItK-kdyzDDBCMAqH09timuuF4171F3JoVWRHBqr2C0/edit#

Adam Ilardi to Everyone (5:08 PM)

cool

Manohara Shankar to Everyone (5:09 PM)

will we lose the diversity of recommendation..?

Julie Cheng to Everyone (5:10 PM)

can you zoom in?

Adam Ilardi to Everyone (5:10 PM)

You might

You can try training a model with “N" seeds each new seed would be the item you just clicked from merch

Marshall Hansi Wu to Everyone (5:11 PM)

If the buyer is already not interested in that product, we'll bring more product for him. In fact the buyer is going to see more items thus more diversity.

Julie Cheng to Everyone (5:13 PM)

How do you account for changes in forecast as more sellers opt into more aggressive ad rates?

Manohara Shankar to Everyone (5:13 PM)

i mean.. i want to buy tent.. but then may be i see a pair shoes which i am curious about.. now my recommendations will start skewing towards shoes within that session right?

Jianian Jin to Everyone (5:13 PM)

Yup agree with marshall that with more exposure on low ranked items, diversity should be kind of better. Also we got a separate controller (based on MMR) to enhance diversity as well