Data Council - Austin 2022

The Case for Declarative Machine Learning

Tristan Zajonc

Built one of first enterprise data science platforms (Sense.io)

Built the machine learning platform for Cloudera

Currently cofounder of Continual


KubeCon
CloudNativeCon
North America 2018
Enterprise Machine Learning on K8s: Lessons Learned and the Road Ahead
Timothy Chen & Tristan Zajonc, Cloudera

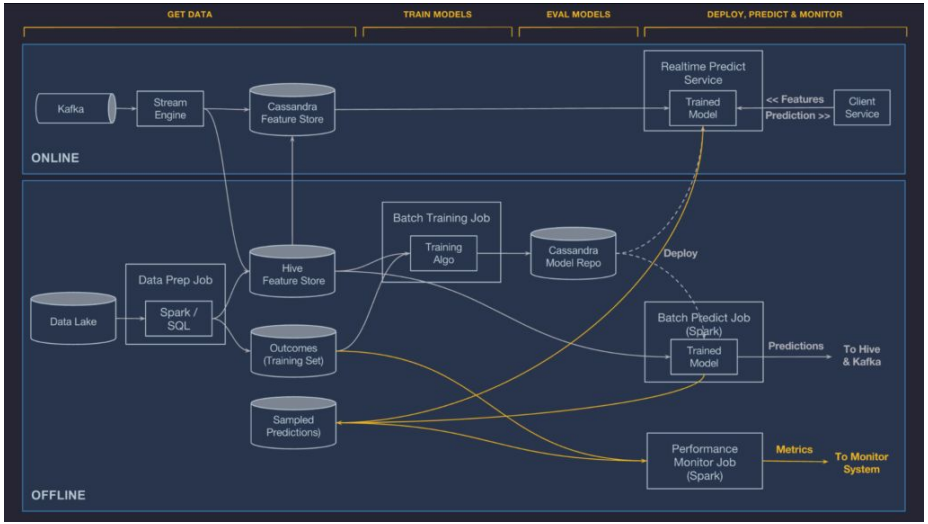
MLOps is drowning in complexity.

We’ve spent the last five years copying Uber

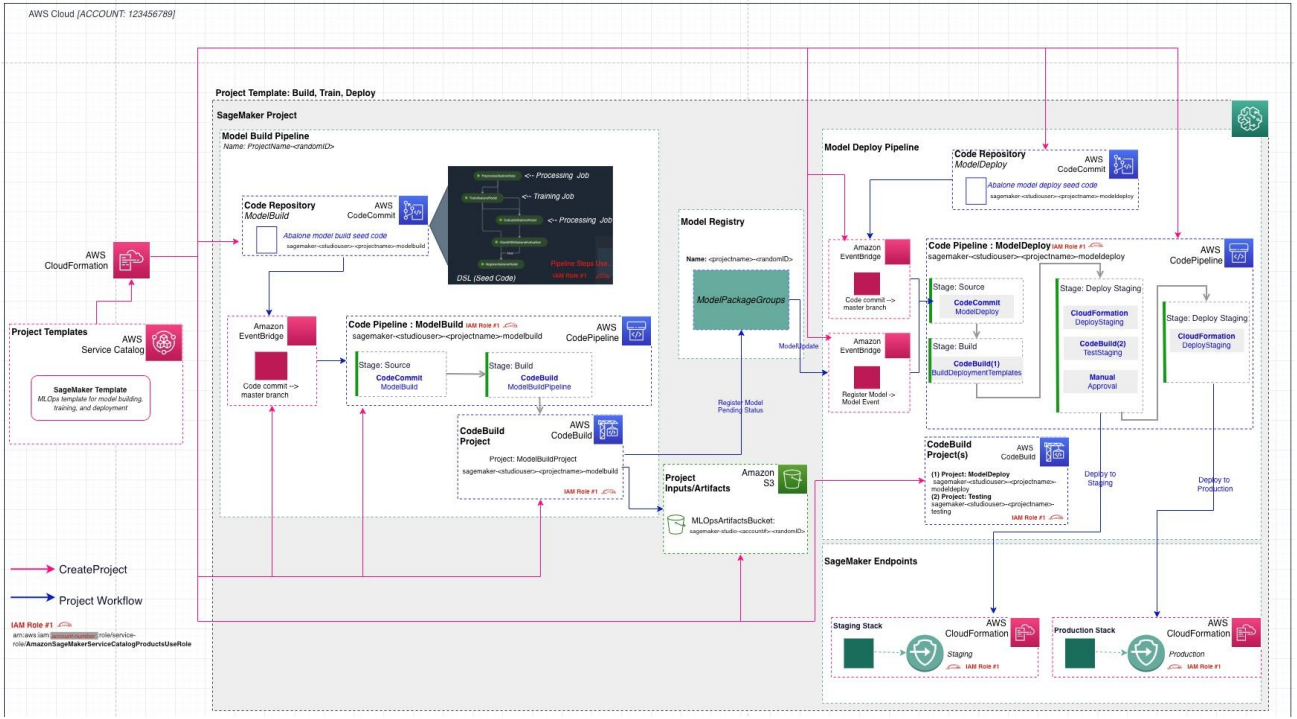
`predict(arrival)`

What the business needs

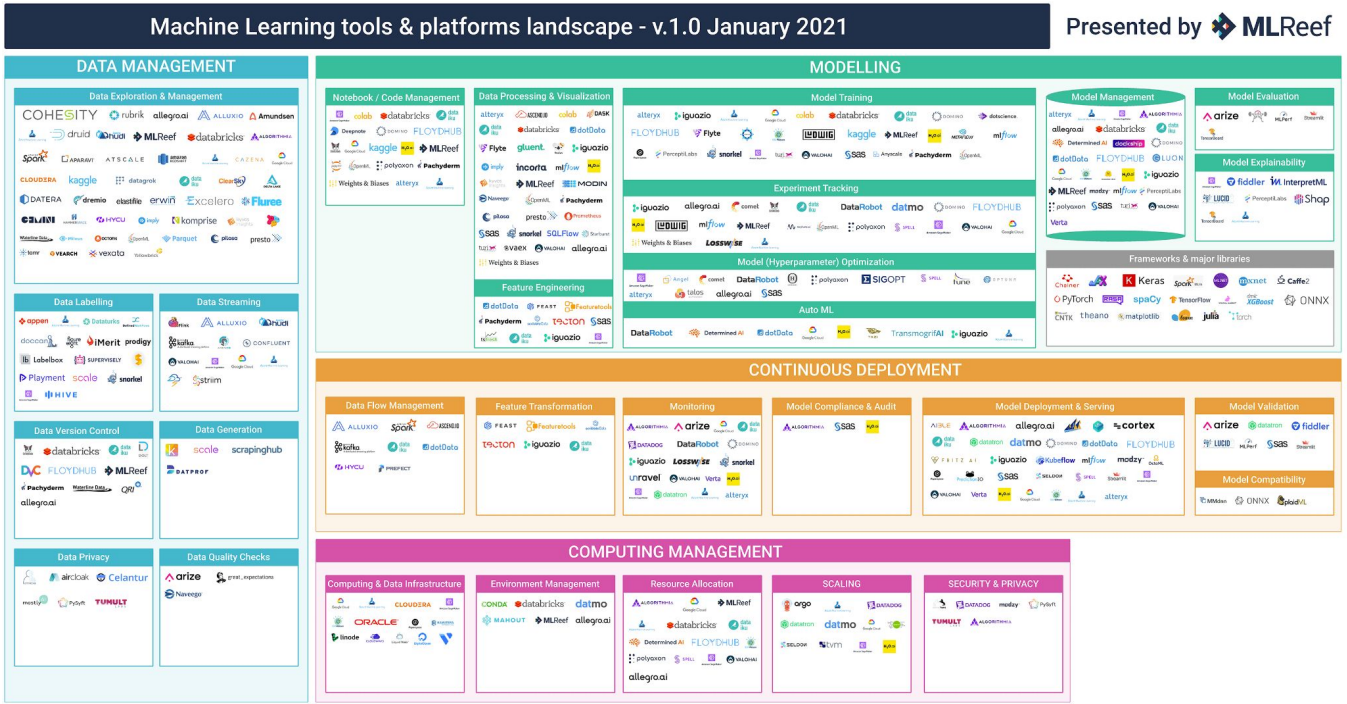
What the engineer must build and maintain →



Or trying to do the same thing in the cloud



Or by stitching together our favorite ML products



What’s the end result?

Infrastructure Challenges

AI infrastructure is complex, expensive, and siloed from data and analytics stack.

Team Challenges

Every new uses cases requires significant time investment from highly skilled teams.

Operational Challenges

No simple, scalable, and governed way to deploy ML use cases across a business.

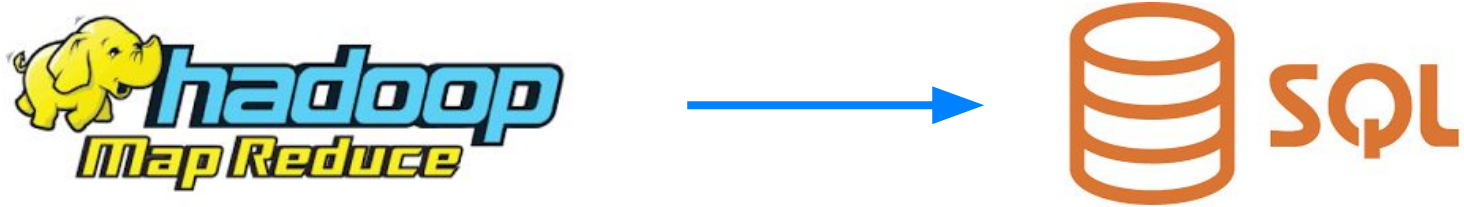
It’s just less fun.

Is there another way?

We’ve seen this sort of problem before and fixed it.

Infrastructure

Data



UI

The solution has been higher-level declarative abstractions.

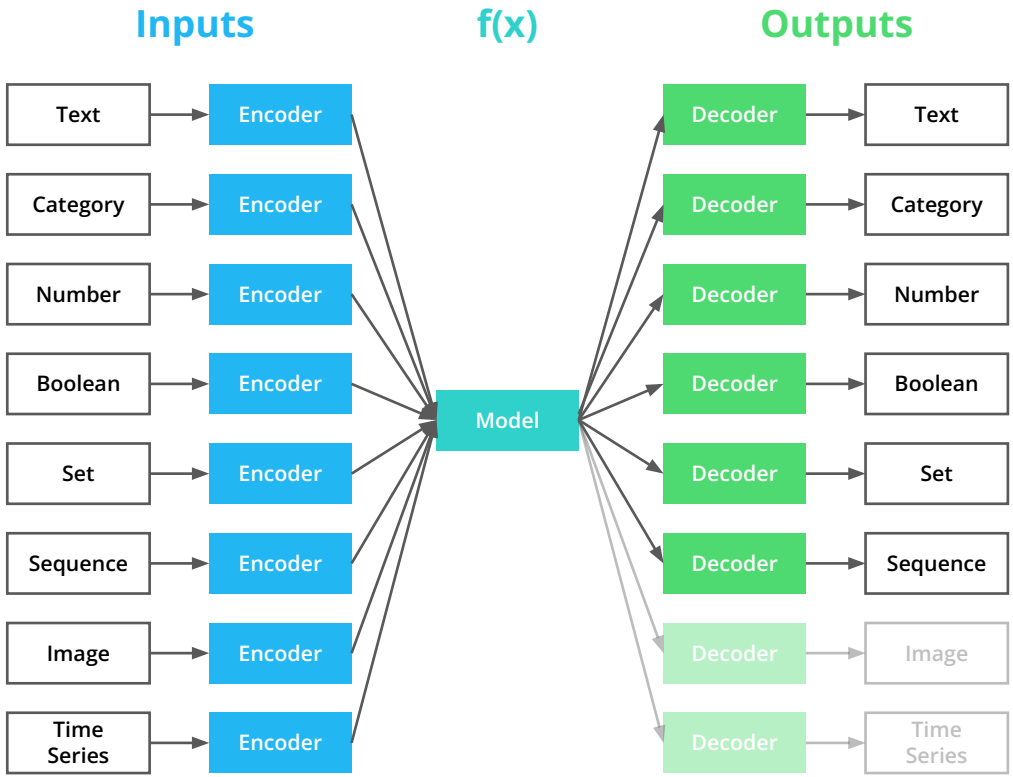
Can we do the same for operational AI?

What’s is operational AI?

PIPELINES

~~PIPELINES~~

Task


Inputs
f(x)
Outputs
Text
Category
Number
Boolean
Set
Sequence
Image
Time Series
Encoder
Encoder
Encoder
Encoder
Encoder
Encoder
Encoder
Encoder
Model
Decoder
Decoder
Decoder
Decoder
Decoder
Decoder
Decoder
Text
Category
Number
Boolean
Set
Sequence
Image
Time Series

Andrej Karpathy
@karpathy

The ongoing consolidation in AI is incredible. Thread:
➡️ When I started ~decade ago vision, speech, natural language, reinforcement learning, etc. were completely separate; You couldn't read papers across areas - the approaches were completely different, often not even ML based.

6:03 PM · Dec 7, 2021 · Twitter Web App

1,447 Retweets **210** Quote Tweets **6,504** Likes

Policies

Training

Schedule

Manual	Daily	Weekly	Monthly	Cron Schedule
Run manually.	Run once a day, midnight.	Run once a week, midnight between Sat/Sun.	Run once a month, midnight, first of month.	ex: 0 2 * * *

↓ Show Advanced Options

Promotion

Policy

Latest default	Best	Manual
The latest model versions are automatically promoted.	Model versions with the best performance are promoted.	Model versions must be promoted manually.

Prediction

Schedule

Manual	Hourly	Daily	Weekly	Cron Schedule
Run manually.	Run once an hour, beginning of hour.	Run once a day, midnight.	Run once a week, midnight between Sat/Sun.	ex: 0 2 * * *

Continual - Operational AI for the Modern Data Stack

Data (Warehouse) Centric

Connects bidirectionally to your cloud data warehouse.

Declarative Workflow

Provides a declarative workflow that's accessible to anybody.

Operationally Focused

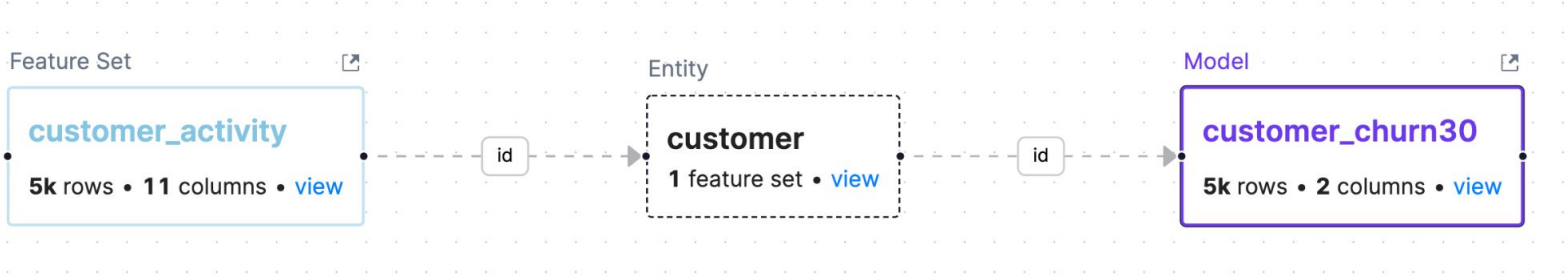
Continually maintains and monitors models in production.

Feature Set

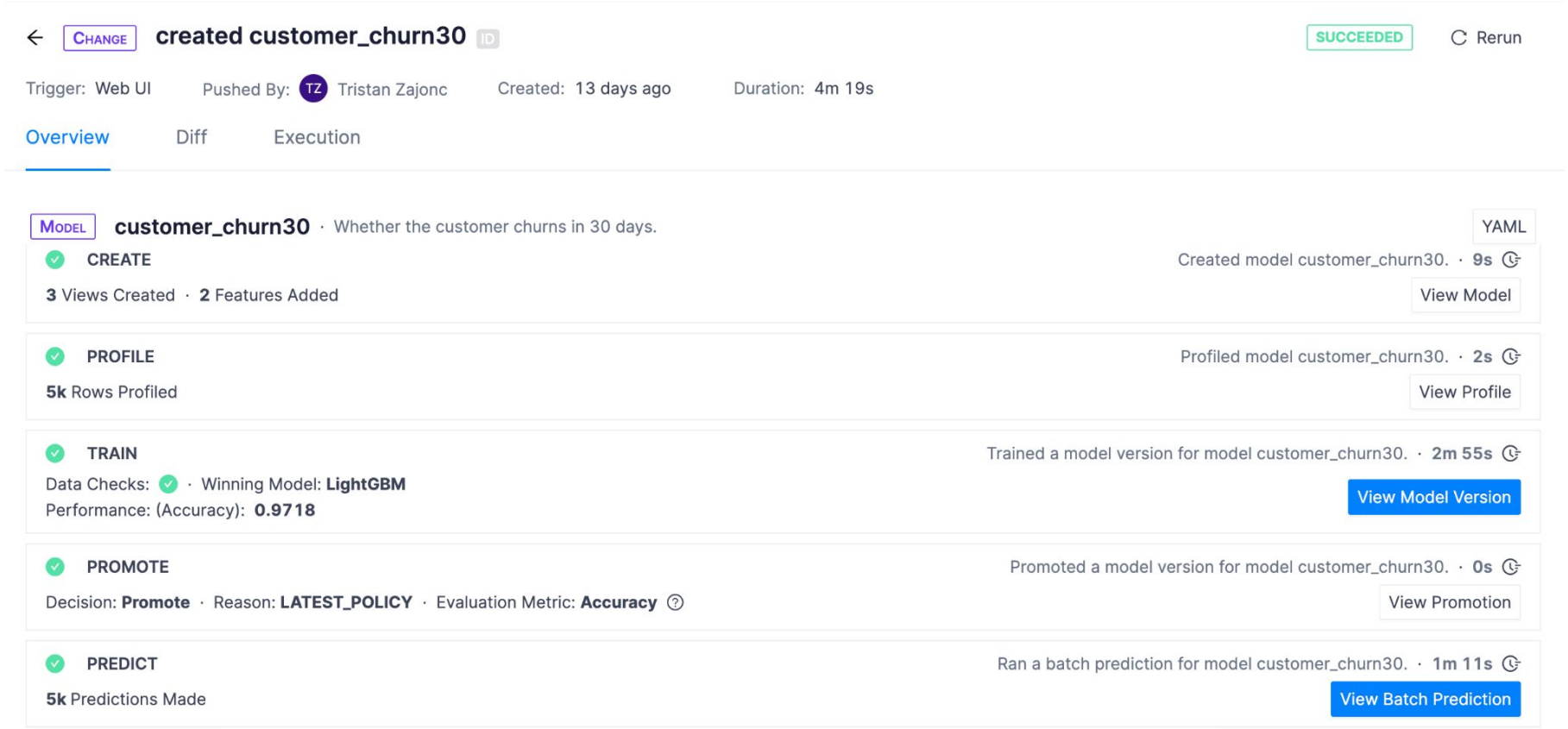
```
{{
  config(
    meta = {
      "continual": {
        "type": "FeatureSet",
        "entity": "customer"
        "index": "customer_id",
        "time_index": "month",
      }
    }
  )
}}
SELECT
  customer_id,
  month,
  feature1,
  feature2
FROM customer_activity;
```

Model

```
{{
  config(
    meta = {
      "continual": {
        "type": "Model",
        "index": "customer_id",
        "time_index": "month",
        "target": "churn_one_month",
        "columns": [
          {"name": "customer_id", "entity": "customer"}
        ],
        "train": {
            "schedule": "@monthly"
        },
        "predict": {
            "schedule": "@daily"
        }
      }
    }
  )
}}
SELECT
  customer_id,
  month,
  churn_one_month
FROM customer_churn_history;
```


Feature Set
customer_activity
5k rows • 11 columns • view
id
Entity
customer
1 feature set • view
id
Model
customer_churn30
5k rows • 2 columns • view

```
dbt run
continual run
```


CHANGE
created customer_churn30
ID
SUCCEEDED
Rerun
Trigger: Web UI
Pushed By: TZ Tristan Zajonc
Created: 13 days ago
Duration: 4m 19s
Overview
Diff
Execution
MODEL
customer_churn30 · Whether the customer churns in 30 days.
YAML
CREATE
Created model customer_churn30. · 9s
3 Views Created · 2 Features Added
View Model
PROFILE
Profiled model customer_churn30. · 2s
5k Rows Profiled
View Profile
TRAIN
Trained a model version for model customer_churn30. · 2m 55s
Data Checks: · Winning Model: LightGBM
Performance: (Accuracy): 0.9718
View Model Version
PROMOTE
Promoted a model version for model customer_churn30. · 0s
Decision: Promote · Reason: LATEST_POLICY · Evaluation Metric: Accuracy
View Promotion
PREDICT
Ran a batch prediction for model customer_churn30. · 1m 11s
5k Predictions Made
View Batch Prediction

Performance

Train ○ Validation ○ Test ●

LightGBM	0.97 ▲ 0.0%	0.95 ▲ 0.0%	0.22 ▲ 0.0%	0.89 ▲ 0.0%	425 ▲ 0.0%	2m 48s ▲ 0.0%
Winning Experiment	Accuracy	ROC AUC	Log Loss	F1	Test Rows	Training Time

Experiments

Train ○ Validation ● Test ○

Running	Failed	Cancelled	Total	Total Duration	Average Duration
0	0	0	11	2m 48s	1s

Rank	Experiment Name	Accuracy	ROC AUC	Log Loss	F1	Duration	State
1	LightGBM	0.958	0.92	0.24	0.82	1s	SUCCEEDED
2	Catboost	0.958	0.93	0.37	0.82	4s	SUCCEEDED
3	Weighted Ensemble	0.958	0.92	0.24	0.82	1s	SUCCEEDED
4	XGBoost	0.955	0.93	0.49	0.82	2s	SUCCEEDED

MODEL **customer_churn30** **boolean**

Last Training: 13 days ago | Last Promotion: 13 days ago | Last Prediction: 21 hours ago

Train | Predict

Training Schedule: @monthly | Prediction Schedule: @daily | Created: 13 days ago | Updated: 13 days ago

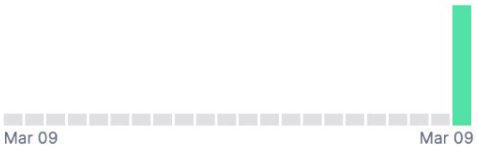
Overview | Schema | Data Preview | Versions | Compare | Batch Predictions | Promotions | Activity

Current Performance details

Train | Validation | Test

LightGBM	0.97 ▲0.0%	0.95 ▲0.0%	0.22 ▲0.0%	0.89 ▲0.0%	425 ▲0.0%	2m 48s ▲0.0%
Winning Experiment	Accuracy	ROC AUC	Log Loss	F1	Test Rows	Training Time

Training History



2m Duration | 100 % Reliability

Prediction History

18s Duration | 100 % Reliability

Historical Performance

Train | Validation | Test | Accuracy | Last year

Let’s build more declarative ML systems!

Benefits:

- Empowers your whole data team.
- Reduces the operational burden of AI/ML.
- Accelerate business impact from AI/ML.

You can try Continual for free at https://continual.ai.

Or scan the QR code to claim your Continual swag.

SCAN ME