INTUIT turbotax creditkarma quickbooks mailchimp

Towards Zero Change Incidents

Intuit's Strategy for Implementing AI-Driven Progressive Delivery

Avik Basu, Staff Data Scientist
Saravanan Balasubramanian, Senior Staff Software Engineer

Technology @ Intuit

Intuit is leading the way in building an AI-native development platform using cloud native open source technology. We're committed to building tools that scale and giving back to the open source community.

INTUIT

KubeCon CloudNativeCon North America 2024

INTUIT turbotax creditkarma quickbooks mailchimp

97M
customers

107B
tax refunds

$2T+
invoices managed

18M
workers paid via QB payroll

88B
requests during peak season

AI-native development platform

INTUIT

KubeCon CloudNativeCon North America 2024

AI-powered App Experiences

810M

AI-driven customer interactions last year

AI-assisted development: coding, testing, debugging

8x

Developer velocity increase in past four years

AI-powered app centric runtime

65B

Machine learning predictions per day

Smart Operations using AIOps

40M+

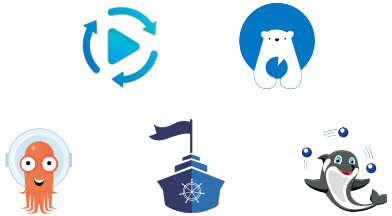
AIOps inferences/day

We believe in open source and open collaboration

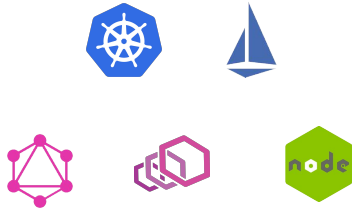
INTUIT

KubeCon CloudNativeCon North America 2024

Recipient of the End User Award in 2019 & 2022



Created, open-sourced, used, and maintained by Intuit



End user of Cloud Native and mobile open source tech

bit.ly/intuit-oss

INTUIT

KubeCon CloudNativeCon North America 2024

What is Progressive Delivery?

Progressive Delivery

- Gradual release of new version
- Reduces the risk of bugs or failures
- Quick rollbacks
- e.g. Blue Green, Canary, Feature Flags
- Argo Rollouts
    - Progressive delivery for Kubernetes

INTUIT

KubeCon CloudNativeCon North America 2024

Why is there a need for AIOps based Progressive Delivery?

Change-Induced Incidents

- 1/3rd of P0/P1 incidents at Intuit were caused by changes
- Changes can be
  - New features
  - Bug fixes
  - Simple dependency updates
- Can be avoided/reduction of impact if detected & resolved early

Static Thresholding based rollbacks

- Set a hard threshold for every metric, e.g.
  - 4% error rate
  - 400 ms of latency
- If any of the metric templates fail, then rollback

Argo Rollout Example

```yaml
apiVersion: argoproj.io/v1alpha1
kind: AnalysisTemplate
metadata:
  name: success-rate
spec:
  args:
  - name: service-name
  - name: prometheus-port
    value: 9090
  metrics:
  - name: success-rate
    successCondition: result[0] >= 0.95
    provider:
      prometheus:
        address: "http://prometheus.example.com:{{args.prometheus-port}}"
        query: |
          sum(irate(
            istio_requests_total{reporter="source",destination_service=~"{{args.se
          )) /
          sum(irate(
            istio_requests_total{reporter="source",destination_service=~"{{args.se
          ))
```

Drawbacks of Static Thresholding

- Not all anomalies are Global
- Many time series metrics are seasonal
  - Daily and/or weekly
  - Contextual anomalies
- Multiple metrics collectively determine system health
  - Collective anomalies
  - Different weightage of each metric
- Every service is unique
  - Different thresholds
  - Different metrics that makes sense
  - Non operational metrics

AIOps journey at Intuit

2022

- Univariate anomaly detection on error rate

2023

- Introduced static thresholding based ensemble score

2024

- Multivariate anomaly detection

INTUIT

KubeCon CloudNativeCon North America 2024

Multivariate Anomaly Detection

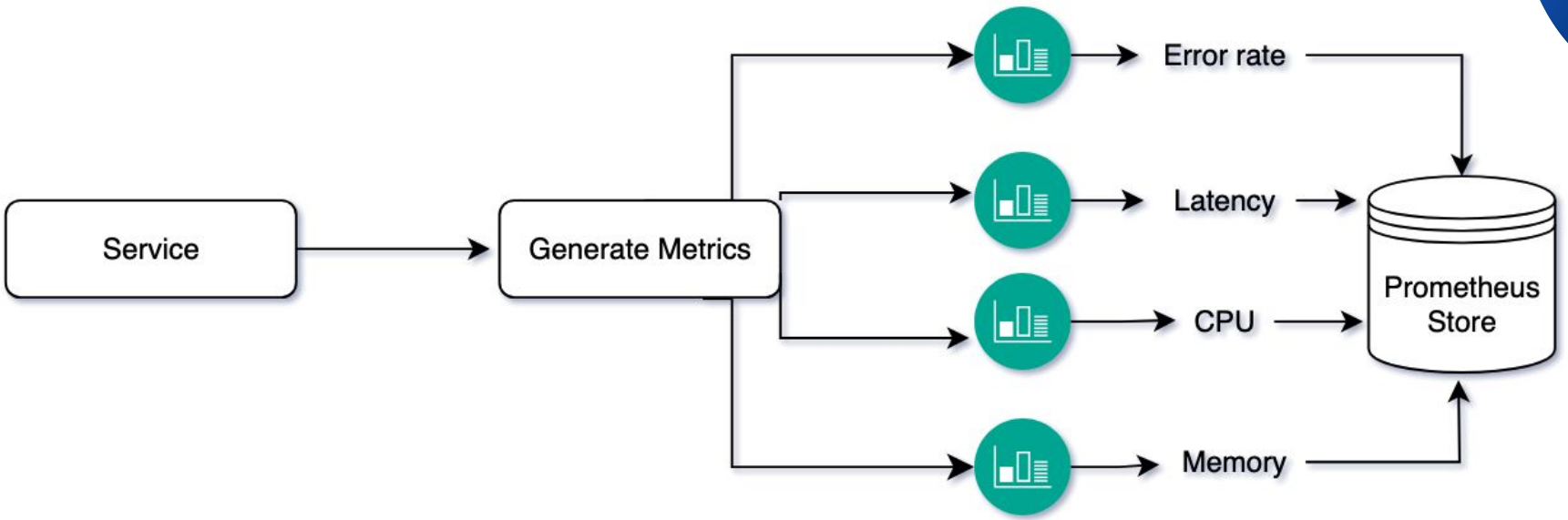
Machine Learning Requirements

- Completely unsupervised
- Able to handle multiple features
- Understand the underlying structure of the timeseries
- Fairly quick to train
- Need not more than 8 days worth of data for training
- Interpretable anomaly scores
- Auto Model Life Cycle management

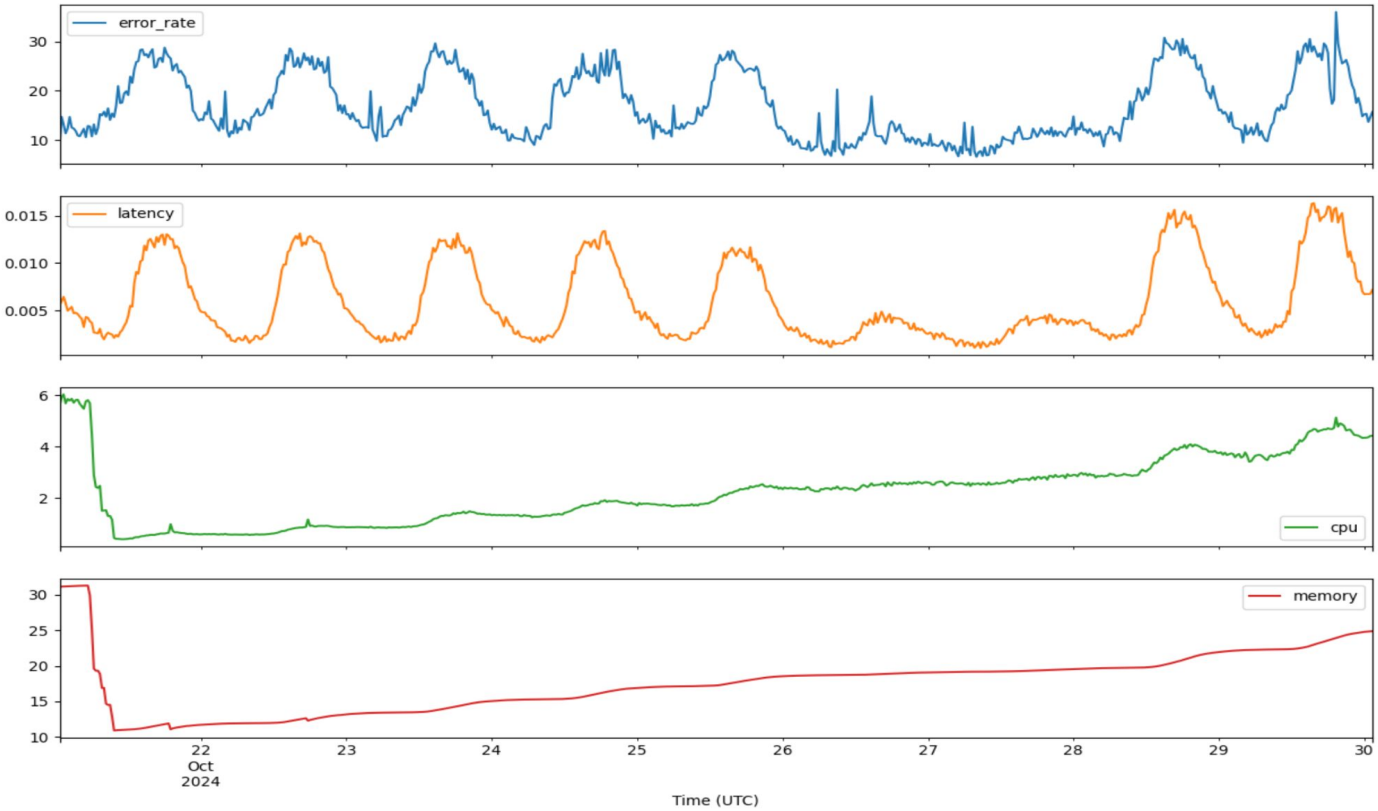
Engineering Requirements

- Stream data processing system
- Support custom sources and sinks
- Sliding window aggregation support
- Lightweight pipeline
- Easy to deploy to multiple clusters
- Right tool for progressive delivery

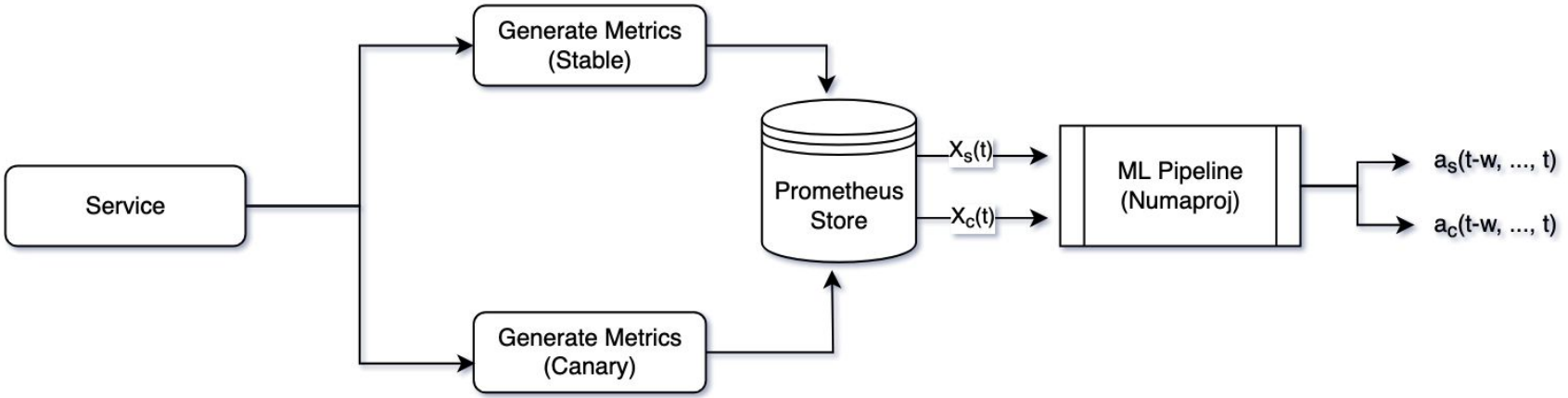
Concept


Service
Generate Metrics
Error rate
Latency
CPU
Memory
Prometheus Store

Multivariate Metrics

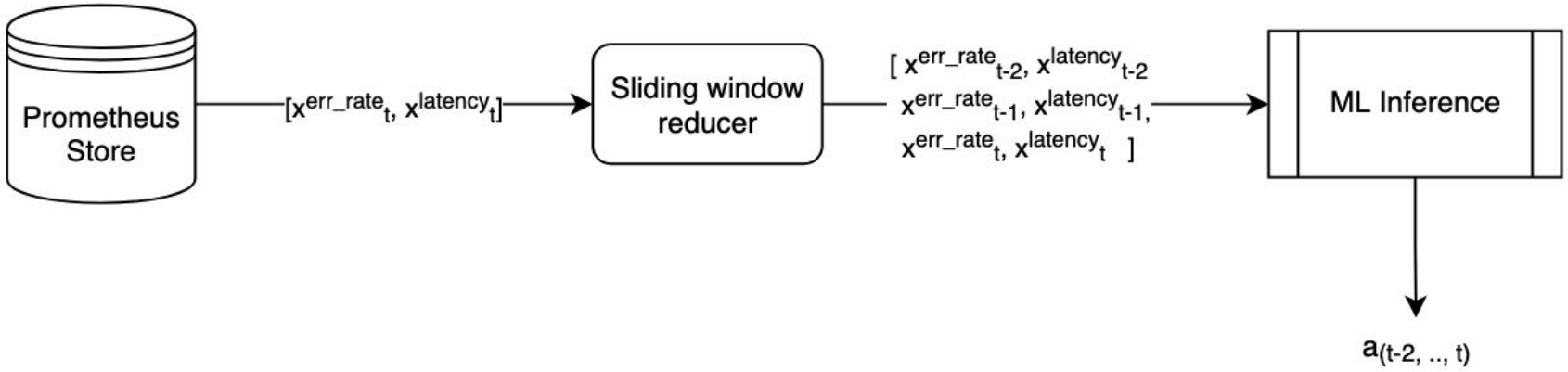


During Progressive Delivery

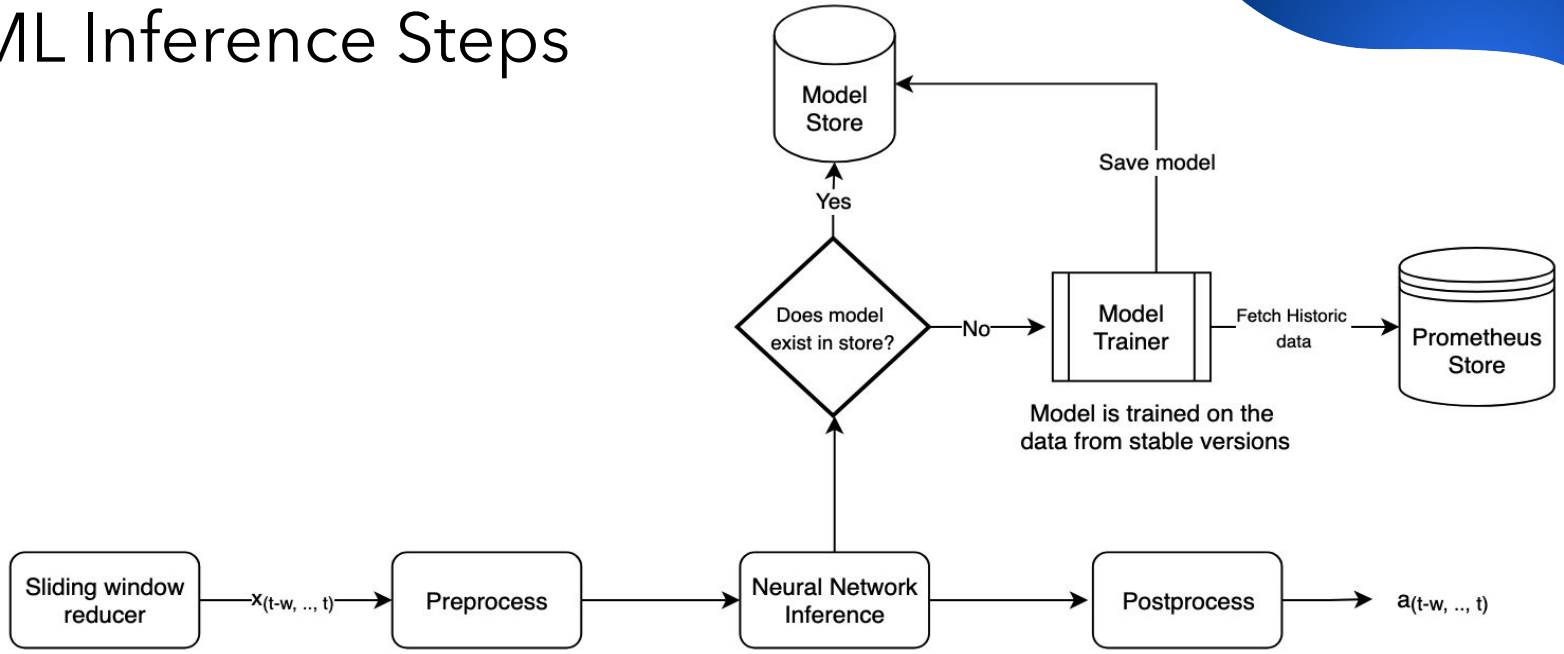


Input Data Processing

- Assume a window size of 3
- Assume 2 multivariate metrics to be processed
- Stable and Canary come in different payloads



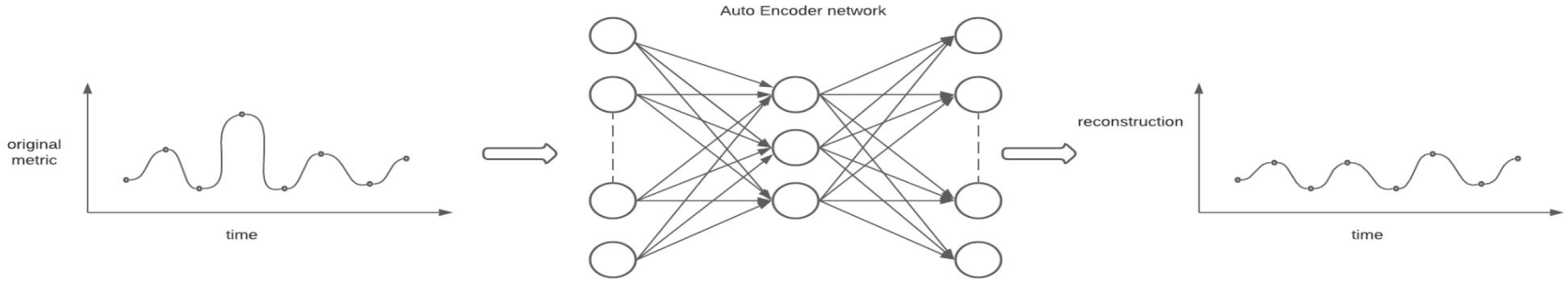
ML Inference Steps



Model Details

numalogic

- CNN, RNN based autoencoder networks
- Quick to train even without GPUs
- Robust to anomalies in the training data
- Feature/Metric weighting capability
- Interpretable anomaly scores
  - Unified
  - Per Metric



Output example

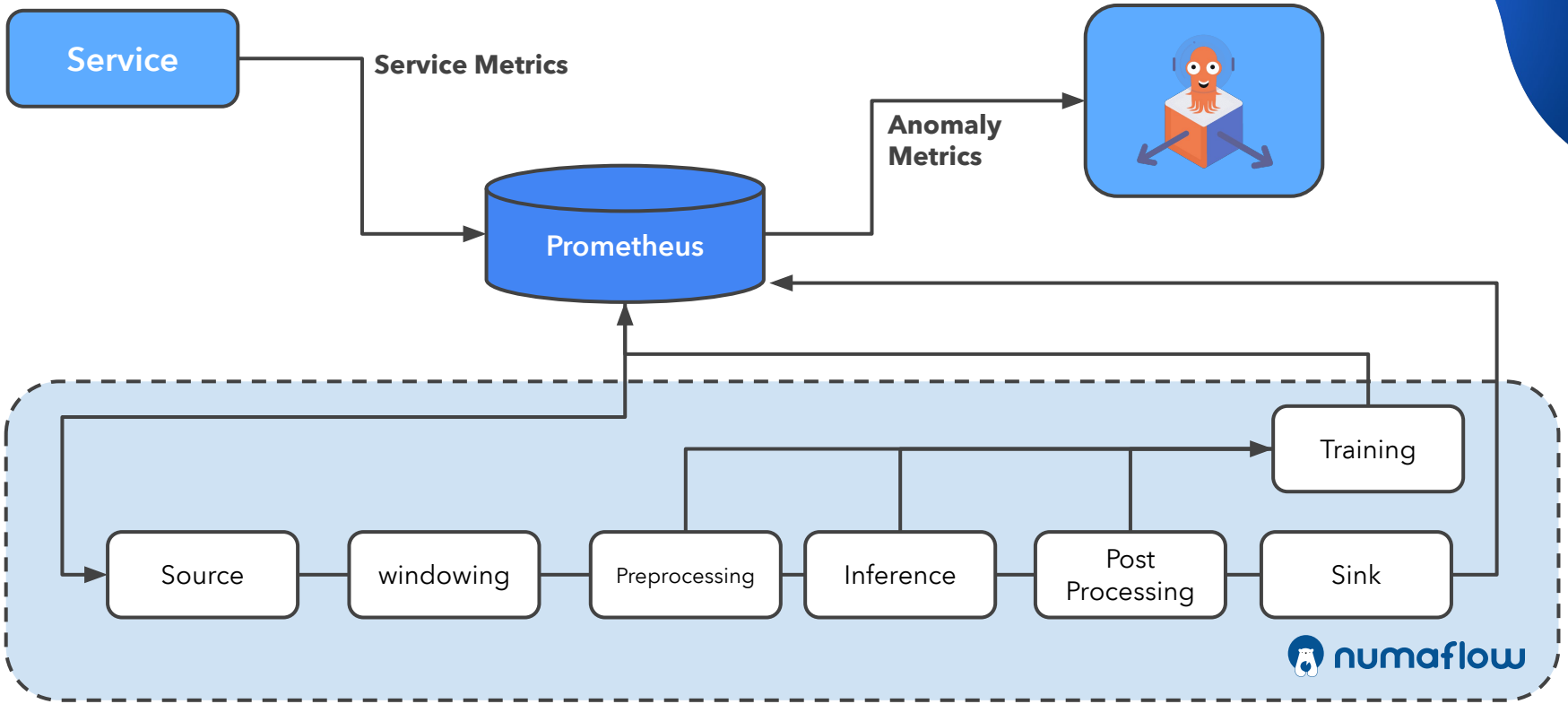
```
{
    "app": "some-service",
    "uuid": "c19d0bb770b2469eb1d8bbfe05f311a4-s",
    "role": "stable",
    "start_ts": 1729194630,
    "end_ts": 1729194690,
    "feature_scores": {
        "latency": 4.36,     // 40%
        "cpu": 1.53,         // 10%
        "error_rate": 0.0,   // 30%
        "memory": 1.23       // 20%
    },
   "unified_score": 2.14,   // weighted_average(ML_scores)
}
```

INTUIT

KubeCon CloudNativeCon North America 2024

How Intuit Implemented AI-driven Progressive Delivery?

Architecture



numaflow

K8s native, **serverless platform for running scalable and reliable event processing**

K8s native event processing

K8s native lightweight event processing with fully featured stream processing semantics

Versatile and can seamlessly operate on the edge, on-prem or in the cloud

Language agnostic framework

SDKs in Java, Python, Golang, Rust.

In-built source/sink connectors. Easy to write sources, functions and sinks

Scalable and Cost efficient

Automatically scales from 0 to X, handling backpressure, while being lightweight and cost-efficient.

Capable of running on edge with a low resource footprint

INTUIT

KubeCon CloudNativeCon North America 2024

Demo

INTUIT

KubeCon CloudNativeCon North America 2024

Thank You

Numalogic Team

- Avik Basu
- Saravanan Balasubramanian
- Kushal Batra
- Nandita Koppisetty

Stay in the loop

INTUIT

KubeCon CloudNativeCon North America 2024

FOLLOW

Intuit Open Source

Don’t miss on exciting OSS events, activities & news

Scan or visit
bit.ly/intuit-oss

Visit our Booth

Get some exciting OSS swag - while supplies last

Let's keep the conversation going

Check out

Numalogic

https://github.com/numaproj/numalogic