box

Beam Summit 2025

Real-time Threat Detection at Box with Apache Beam

Mark Chin

Elango Prasad Logeswar

Agenda

Current threat landscape

Architecture overview

Challenges and solutions

Dataflow orchestration at Box

How bad is ransomware?

In a world of evolving threats and cybersecurity challenges, ransomware has remained on top

▲37%

Increase in the presence of ransomware in observed breaches[1]

$115K

Median ransom paid for a ransomware attack[1]

50%

of leaders that said they were prepared for attack still fell victim[2]

1. Verizon 2025 Data Breach Investigations Report
2. Fortinet - Global Threat Landscape Report

Ransomware leads to major issues

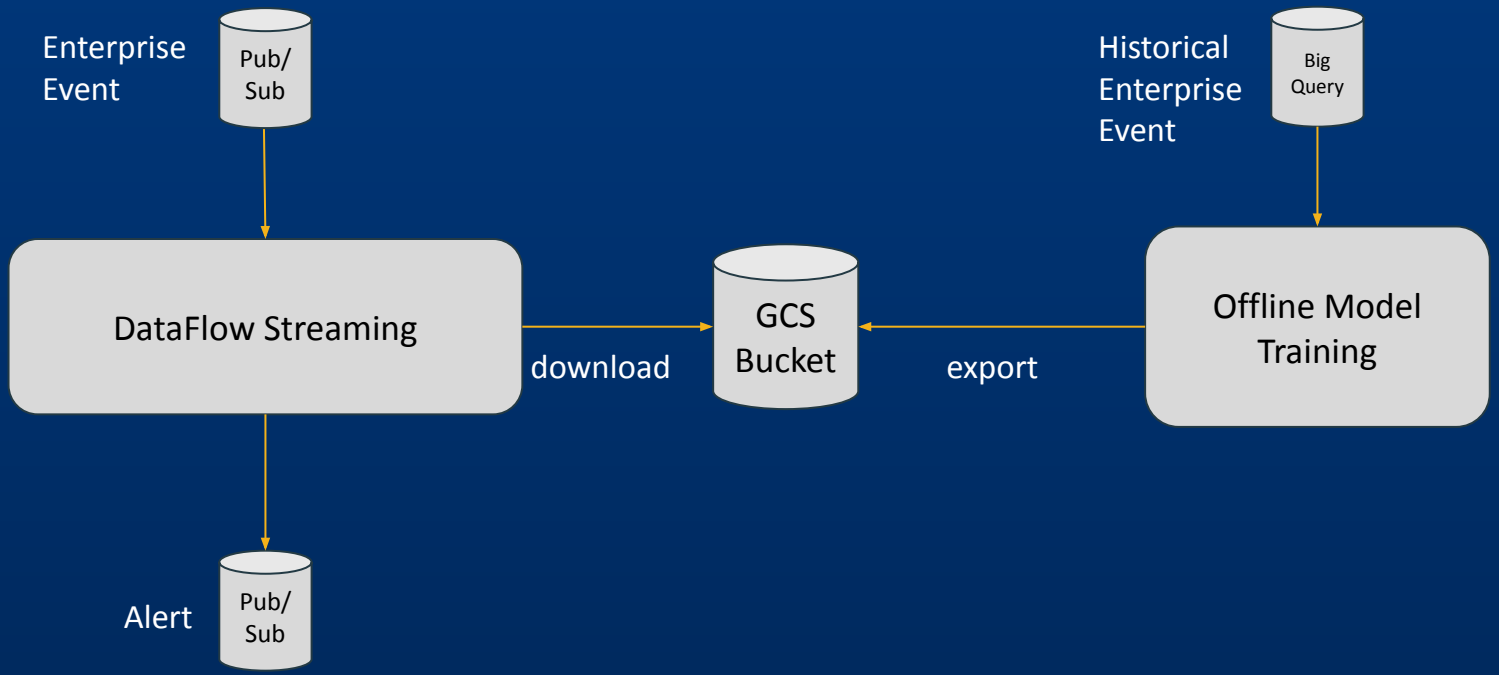
A ransomware attack can create big problems for organizations that get hit

Disruption of business

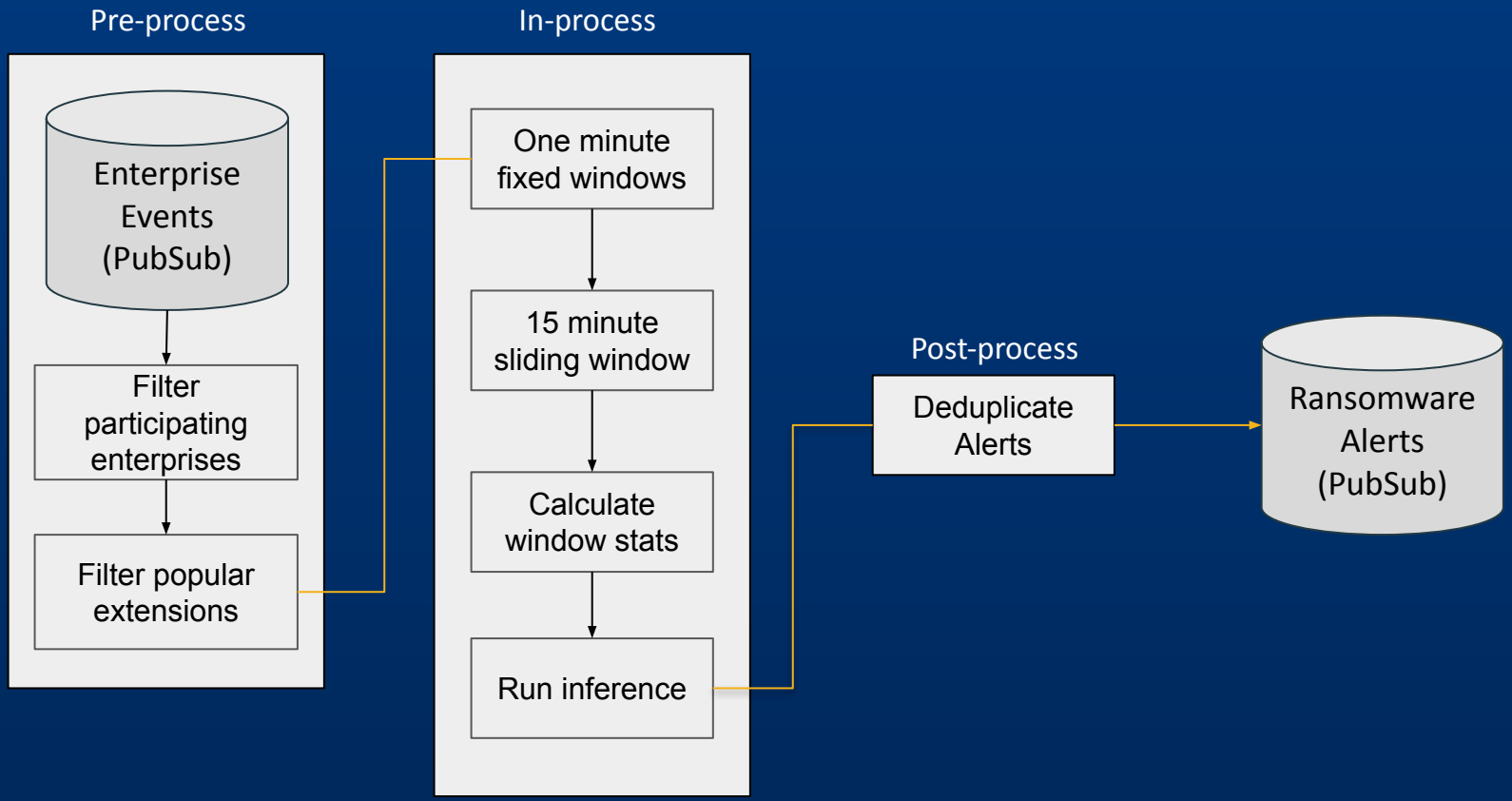
Compromised personal data

Expensive ransoms

Real-Time Ransomware Detector



Streaming Pipeline



Core Challenges

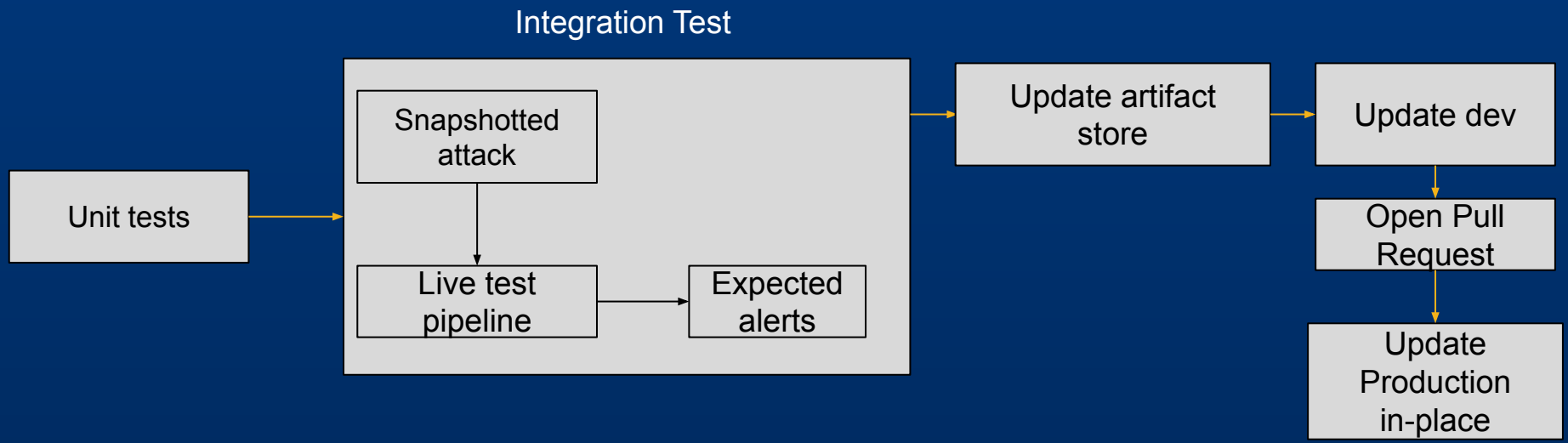
Testing on a live pipeline

Ensuring 100% uptime

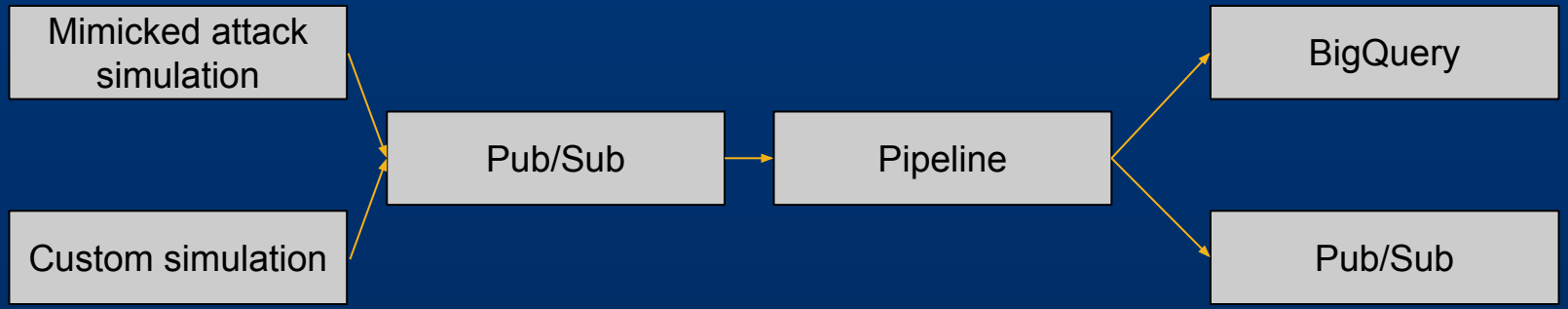
Scaling to production volume

Alert handling

Change Deployment



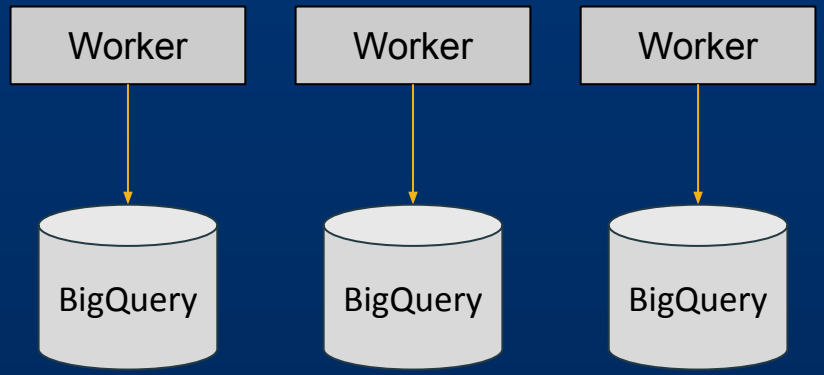
Real-time Continuous Testing



Querying Auxiliary Data within Dataflow

Expensive call to BigQuery

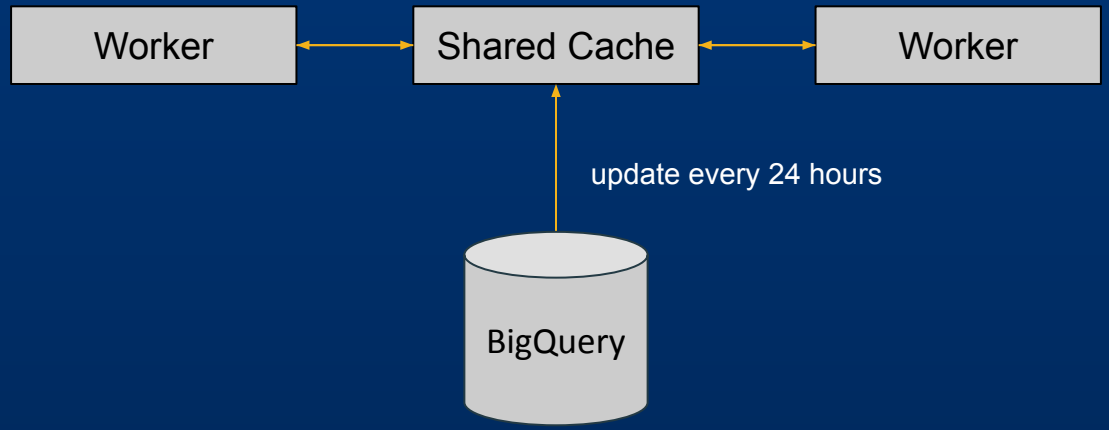
Frequent reloading of data to prevent stale data



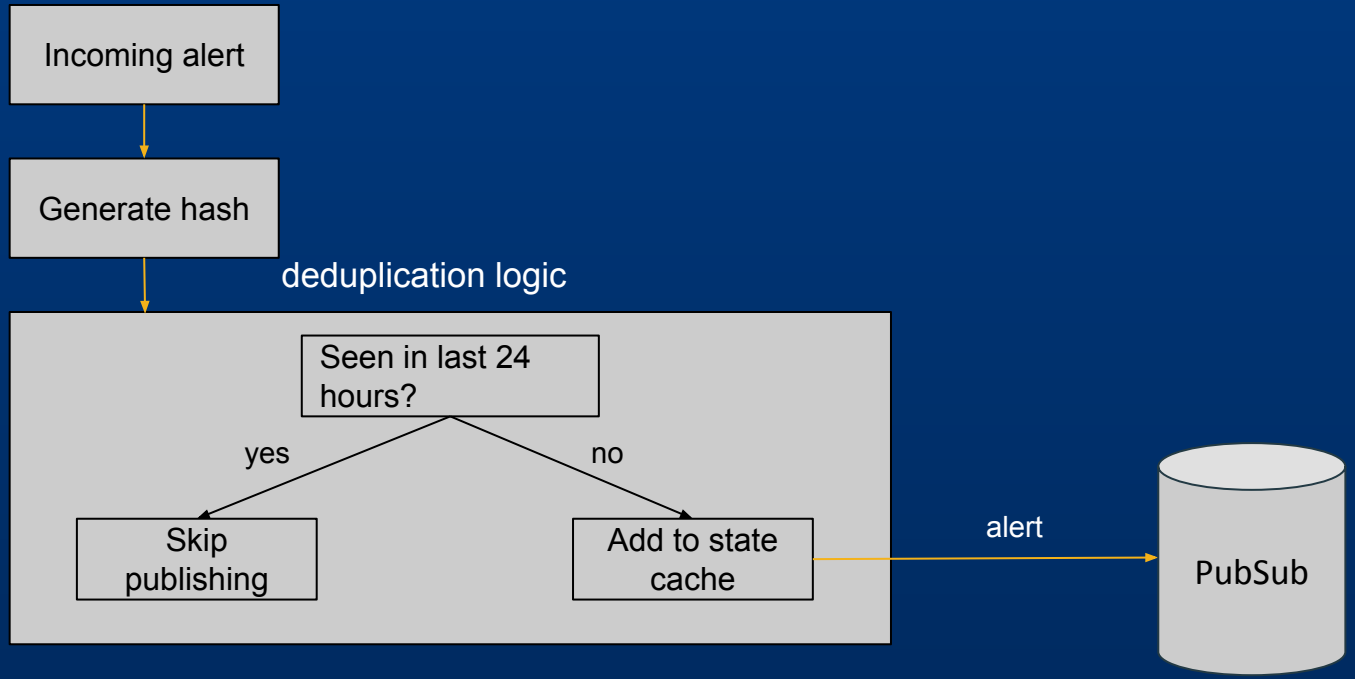
Sharing Resources

Shared cache implementation

Workers only need to call BigQuery based on stale timer



Alert Deduplication



Production Readiness

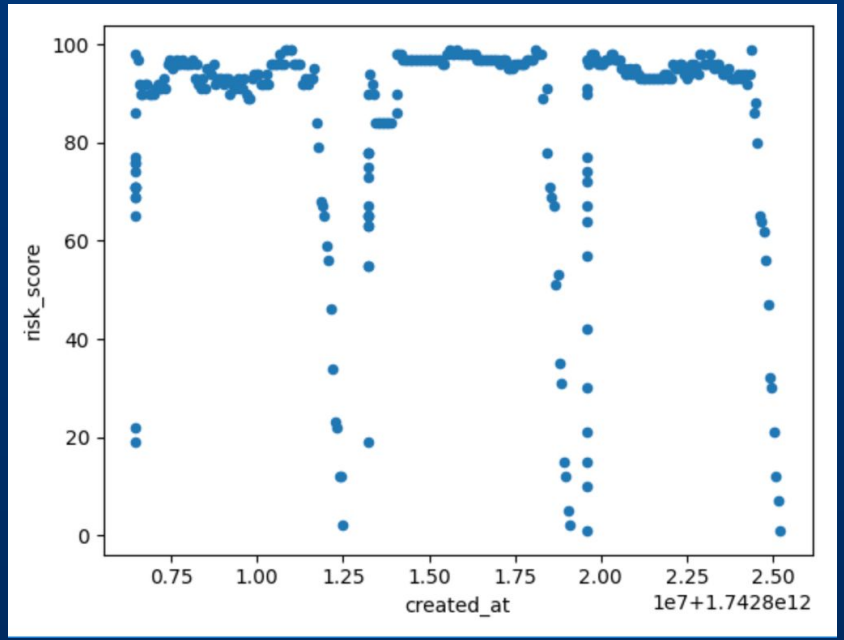
Autoscaling

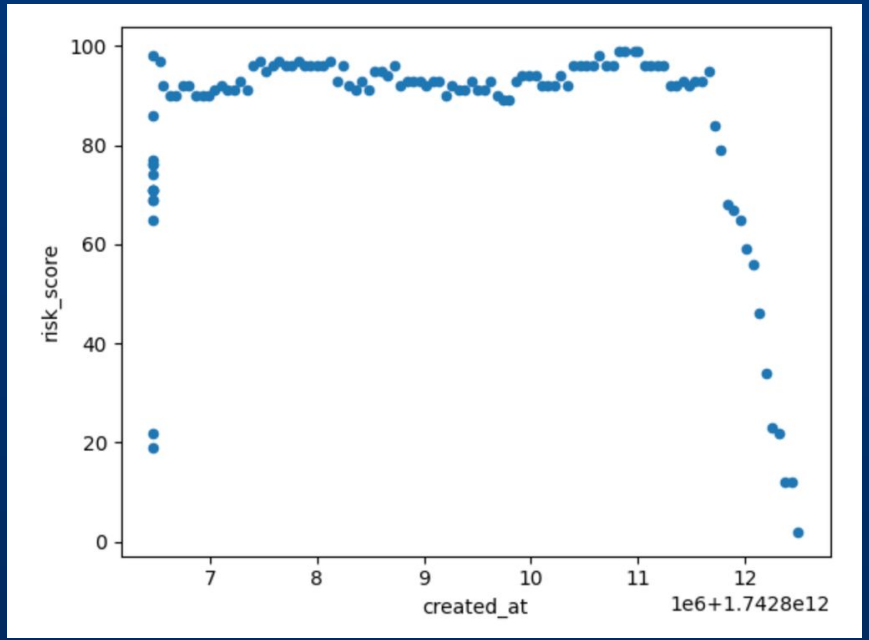
Data sampling

Fault tolerance

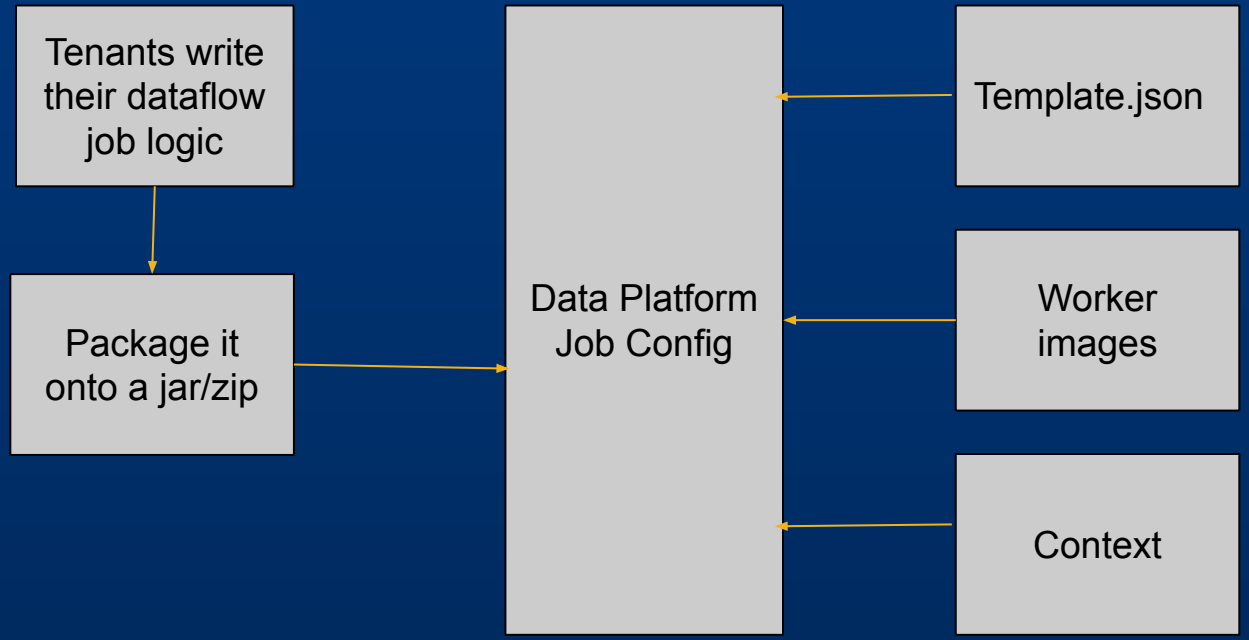
Efficient memory management

The result?

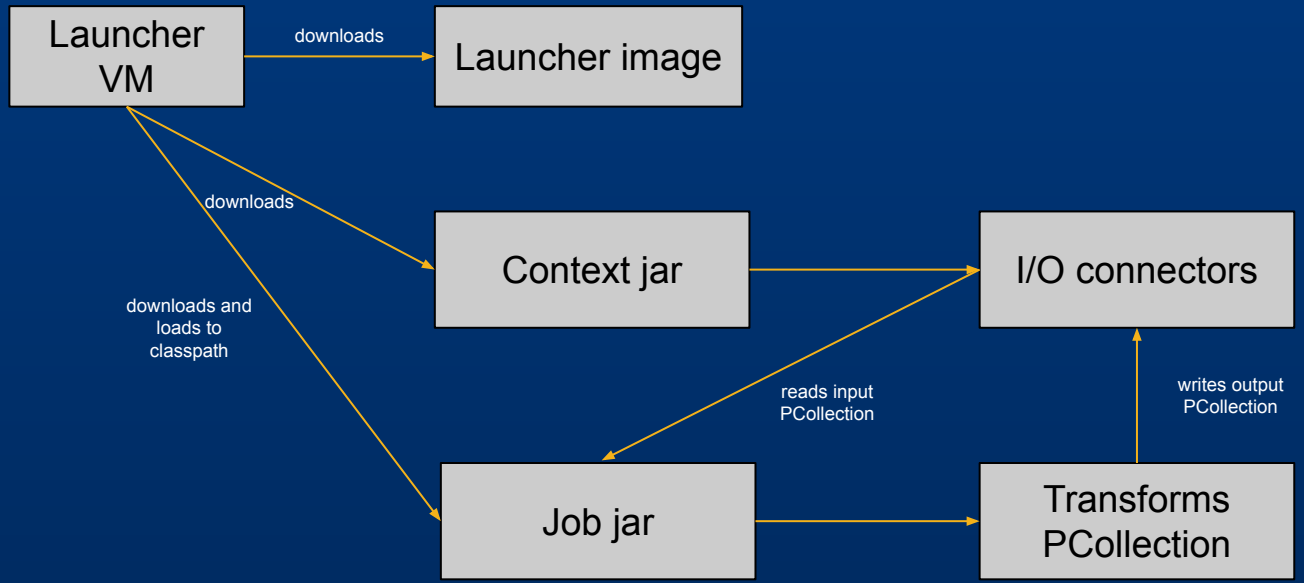

risk_score
100
80
60
40
20
0
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
created_at
1e7+1.7428e12


risk_score
100
80
60
40
20
0
7
8
9
10
11
12
created_at
1e6+1.7428e12

Dataflow at Box



Dataflow Orchestration



Mark Chin
Elango Prasad Logeswar

QUESTIONS?

https://www.linkedin.com/in/elangoprasad96/
https://www.linkedin.com/in/mark-c-02a429151/