

Empowering Real-Time Fraud Prevention with Apache Beam

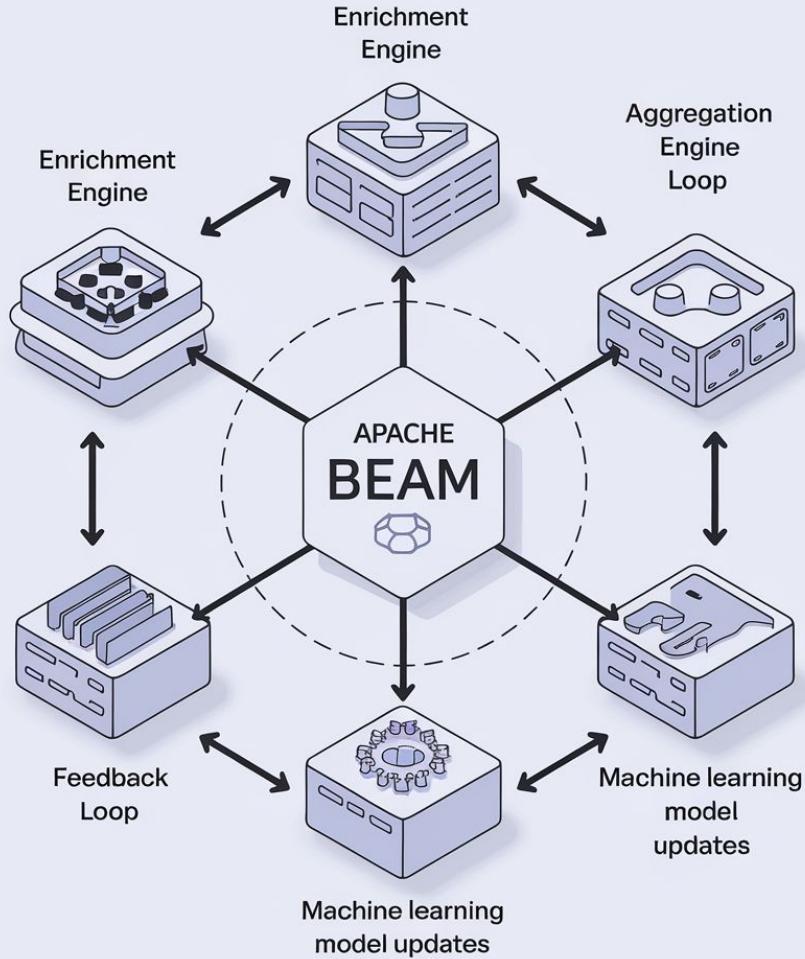
Hai Sadon



September 4-5, 2024
Sunnyvale, CA. USA

Real-Time Fraud Prevention with Apache Beam

A Deep Dive into Our Modular, High-Performance Fraud Detection System



About me

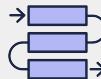
- Leading the Data Platform Group at Transmit Security
- Previously managed a team overseeing Apache Flink infrastructure for Microsoft engineering teams



BEAM
SUMMIT



Transmit Security Mosaic Platform



Orchestration

Simplified integrations, policy & decisioning, journey workflow



Identity Management

User profiles, authorization, SSO



Authentication

Passkeys, passwordless MFA, magic links



Identity Verification

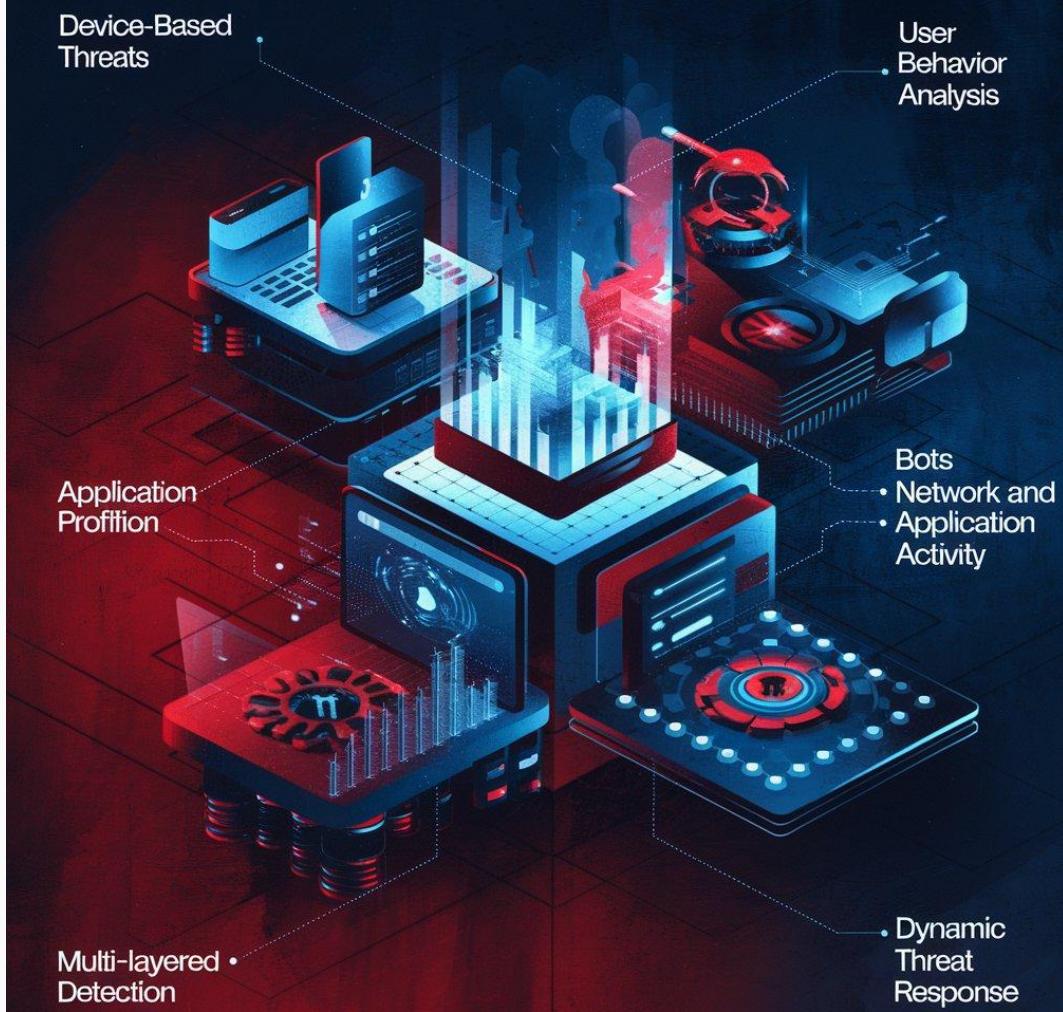
Documentation and database checks, liveliness and selfie analysis, embedded fraud intelligence



Detection and Response Service

Multi-mode, real-time and post detection, ML and AI driven

The Challenge



The Challenge of Real-Time Fraud Prevention

Device-Based Threats

Detecting compromised or spoofed devices used for fraud.

Identifying device anomalies in real time.

Network and Application Activity

Identifying coordinated attacks across networks and apps.

Real-time monitoring for unusual patterns.

User Behavior Analysis

Differentiating between normal and suspicious activities.

Continuous monitoring and adaptive detection.

User Profiling

Differentiating legitimate users from fraudsters.

Maintaining accurate, dynamic user profiles.

Bot Activity

Automated attacks mimicking human behavior.

Advanced detection of sophisticated bot activities.

Multi-Layered Detection

Integrating various detection methods like device fingerprinting, biometrics, and network analysis.

Creating a unified system to prevent fragmented decision-making.



BEAM
SUMMIT

Our Fraud Detection System Architecture



Data Capture



Enrichment
Engine



Aggregation
Engine



Feature Engine



Machine Learning
Models Engine



Rule Engine



Feedback Loop

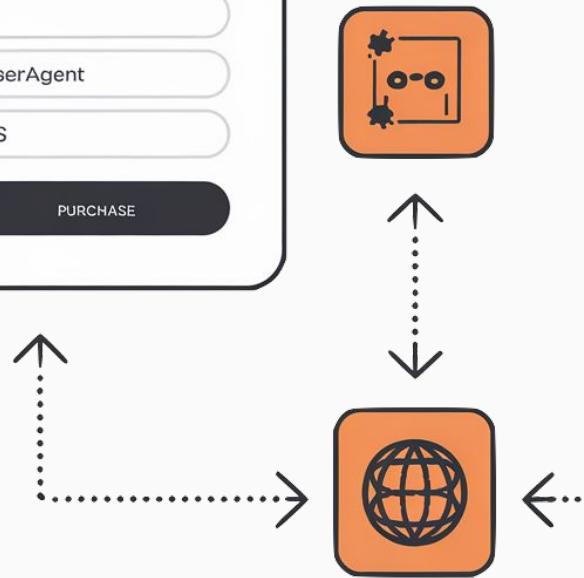
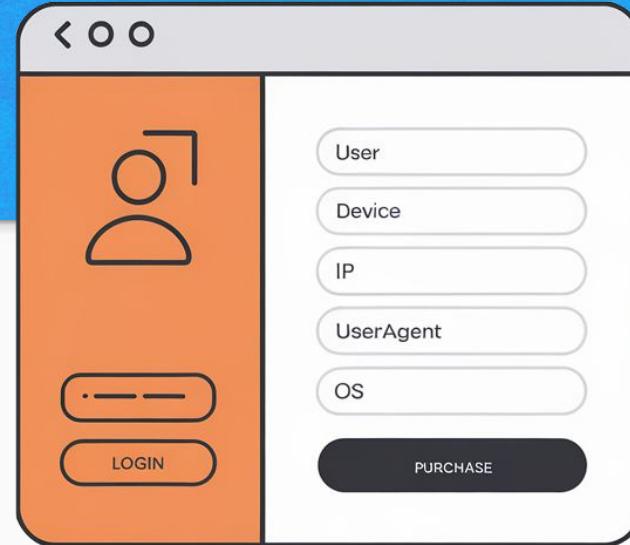
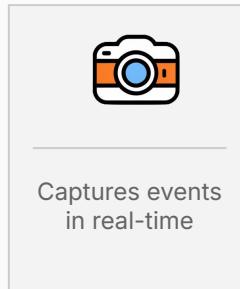


Data Skew
Management



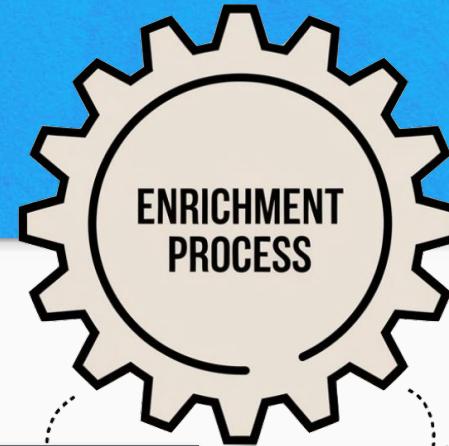
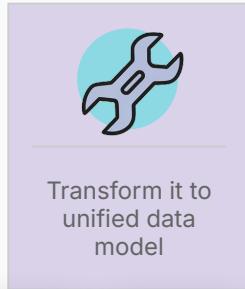
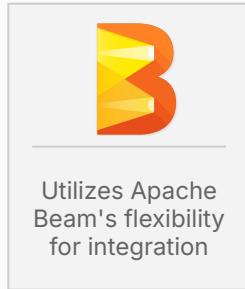
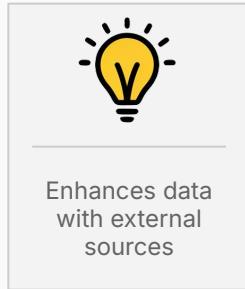
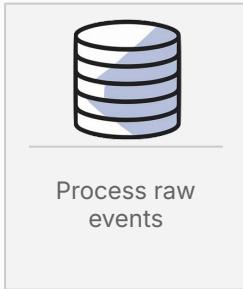
ΞΔM
SUMMIT

The Data Capture



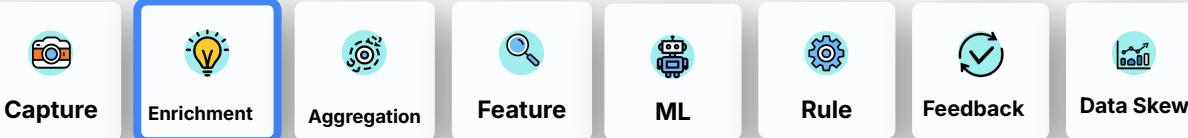
3EΔM
S U M M I T

The Enrichment Engine



```
{  
  "actionType": "login",  
  "ip": "192.168.1.1",  
  "userId": "user123",  
  "deviceId": "device456"  
}
```

```
{  
  "actionType": "login",  
  "ip": "192.168.1.1",  
  "userId": "user123",  
  "deviceId": "device456",  
  "geoLocation": "New York, USA",  
  "asn": "AS12345",  
  "knownProxies": [  
    "proxy1",  
    "proxy2"  
  ],  
  "hostedServices": [  
    "service1",  
    "service2"  
  ]  
}
```



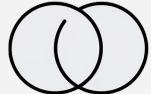
BEAM
SUMMIT

Split and Merge Approach



THE CONCEPT

When dealing with multiple stages or processes that can operate independently, this approach can be applied to parallelize processing.



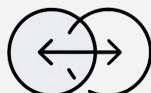
SPLIT

The event is cloned for each key, allowing parallel processing in different stages or processors.



PROCESS INDEPENDENTLY

Each processor handles its assigned key without interference from other stages.

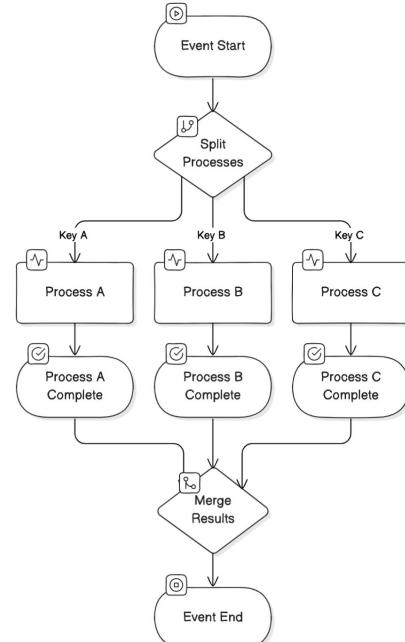


MERGE

Once all the stages are completed, the results are merged back into a single event, integrating the outputs from each stage.

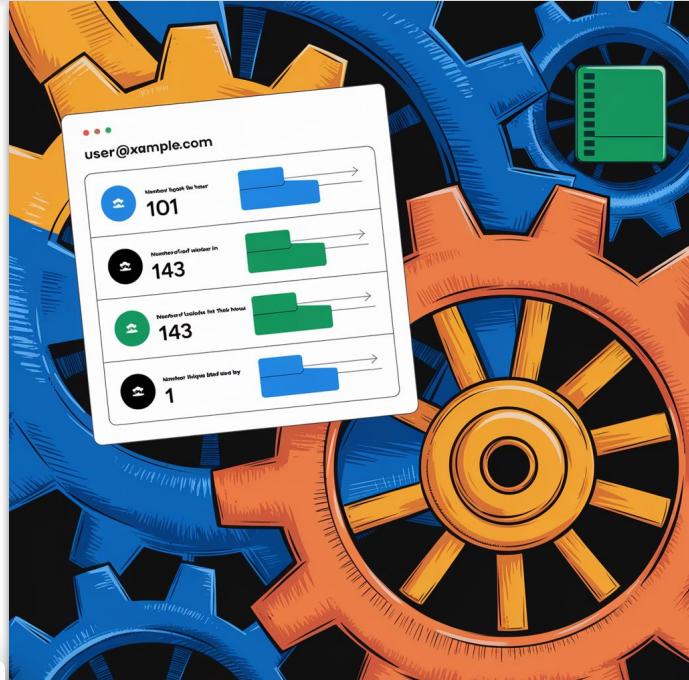
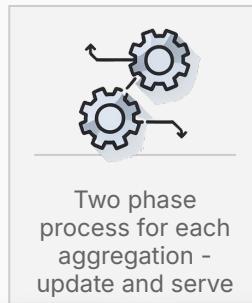
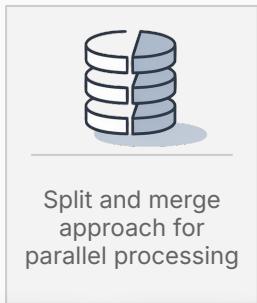
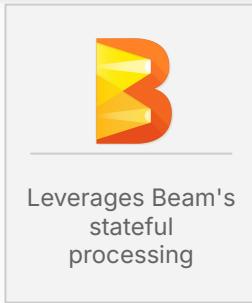
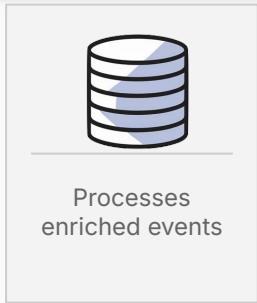


ParDo



BEAM
SUMMIT

The Aggregation Engine



Capture



Enrichment



Aggregation



Feature



ML



Rule



Feedback

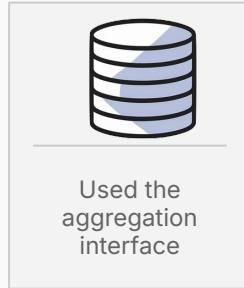


Data Skew



BEAM
SUMMIT

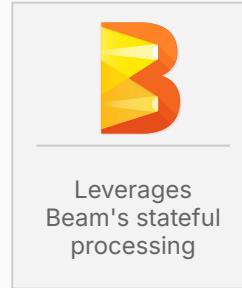
Virtual Aggregations



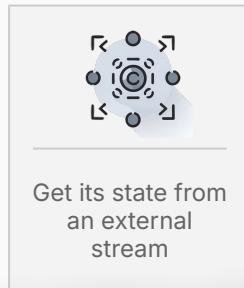
Used the aggregation interface



Splits events by aggregation key



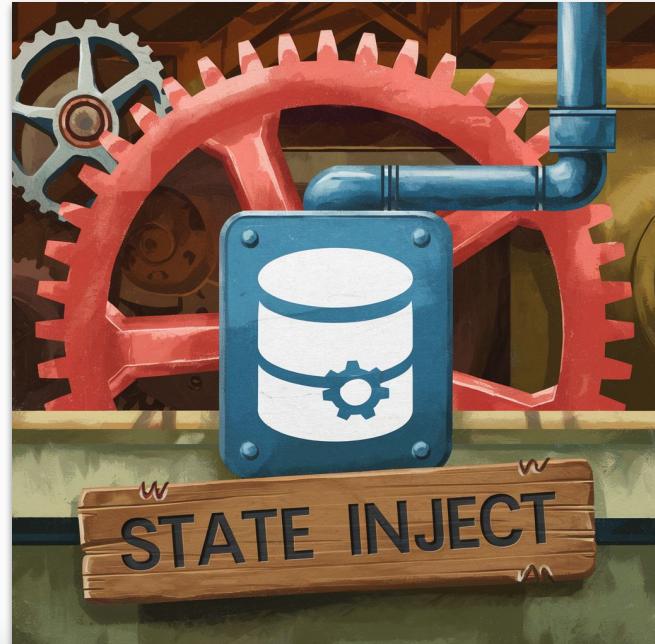
Leverages Beam's stateful processing



Get its state from an external stream



An integral part of the aggregation engine



Capture



Enrichment



Aggregation



Feature



ML



Rule



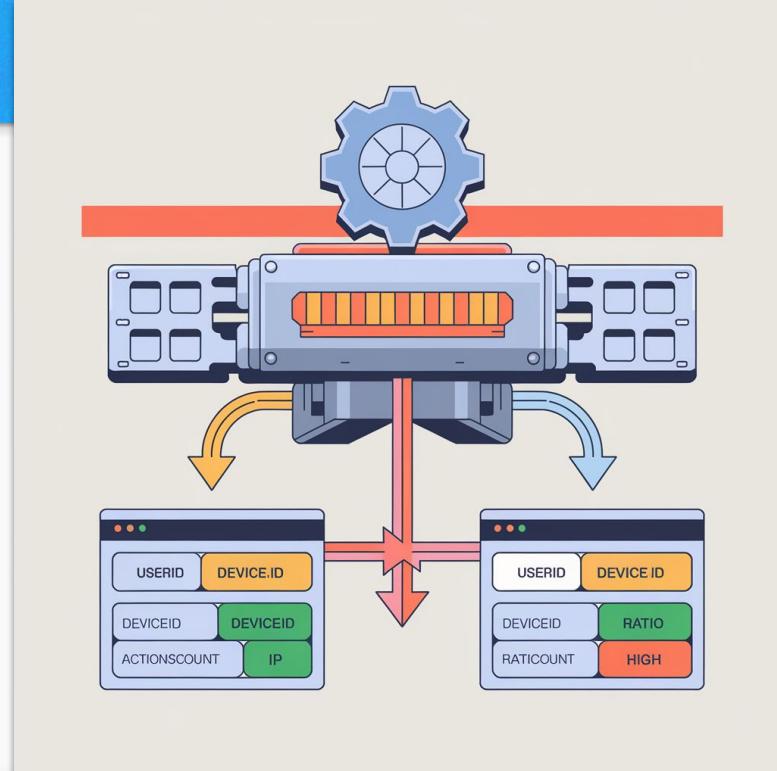
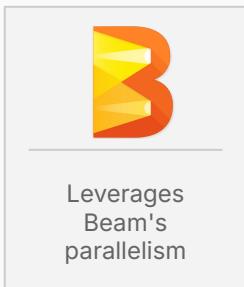
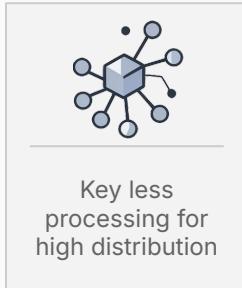
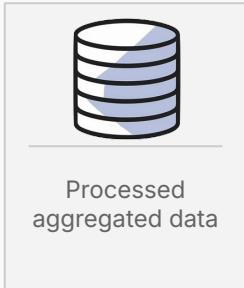
Feedback



Data Skew



The Feature Engine



Capture



Enrichment



Aggregation



Feature



ML



Rule



Feedback



Data Skew

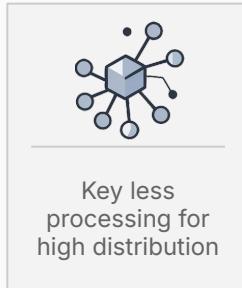


BEAM
SUMMIT

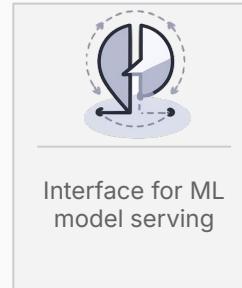
Machine Learning Models



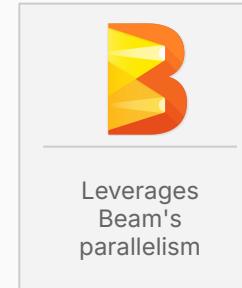
Processed
featured data



Key less
processing for
high distribution



Interface for ML
model serving



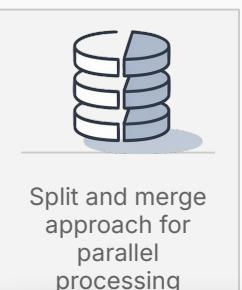
Leverages
Beam's
parallelism



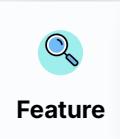
Ability to run
multiple versions
of models and
compare results



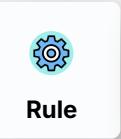
Maintained by the
data science team



Split and merge
approach for
parallel
processing



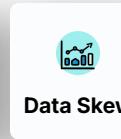
ML



Rule



Feedback

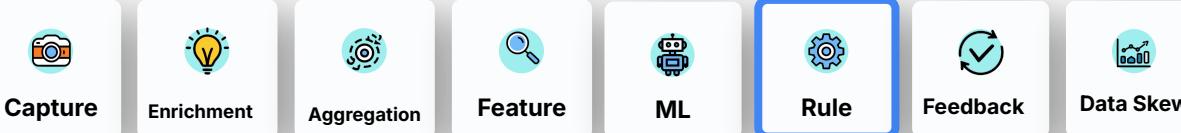
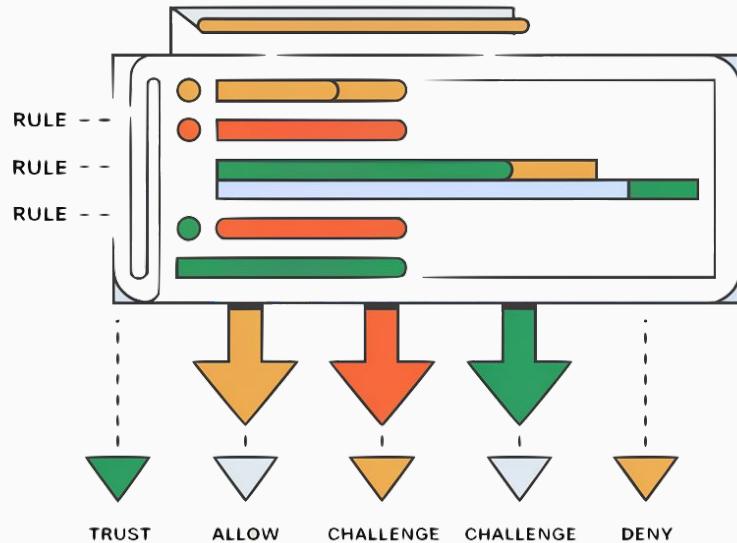
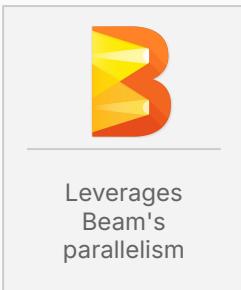
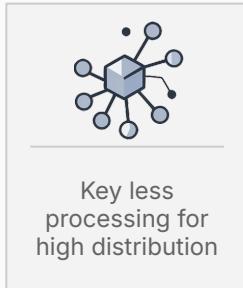
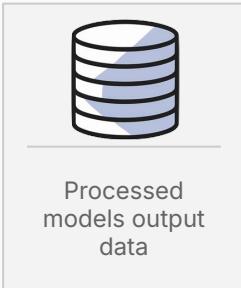


Data Skew



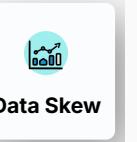
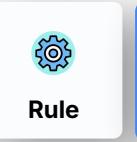
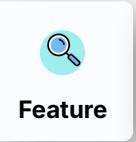
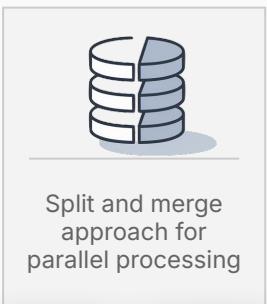
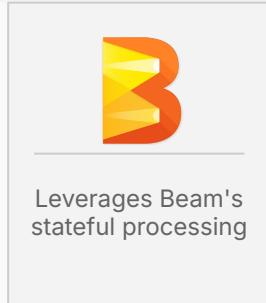
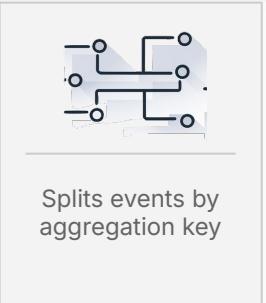
BEAM
SUMMIT

Rule Engine

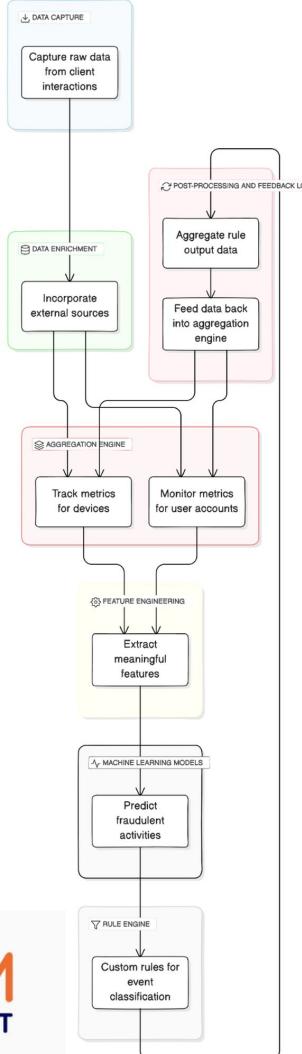


BEAM
SUMMIT

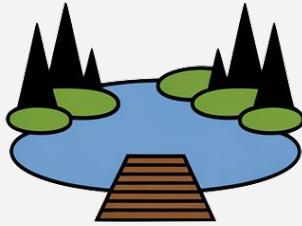
Post Processing Aggregations



BEAM
SUMMIT

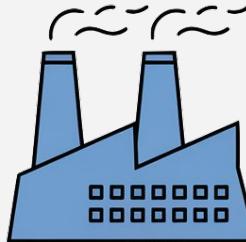


Data Skew Management



Store the data for batch processing

DATA LAKE



Process the data offline

BATCH PROCESS



Ingest back aggregation results

INGESTION



Serve the skewed keys (virtual aggregations)

AGGREGATIONS SERVING



Capture



Enrichment



Aggregation



Feature



ML



Rule



Feedback



Data Skew



Combiners



BEAM
SUMMIT

Challenges and Workarounds

Lack of structure object support in state <p>Like dictionaries. Also states are must be declared and are not dynamic.</p> <p>Used naive serializer</p>	No built-in state TTL mechanism <p>No offering</p> <p>Using timer for TTL</p>	Key iterations is not possible <p>No offering</p> <p>Create a key base store</p>
Cannot query the state outside the application <p>No offering</p> <p>Pushed the state into BigTable</p>	Ordering <p>There is no low latency solution for ordering.</p> <p>Custom logic (no real solution)</p>	Cannot clear window state <p>No offering</p> <p>Create job from snapshot from time to time</p>



Collaboration and Modularity

MODULAR

Modular design supporting cross-team collaboration

INDEPENDENT

Independent development of pipeline stages

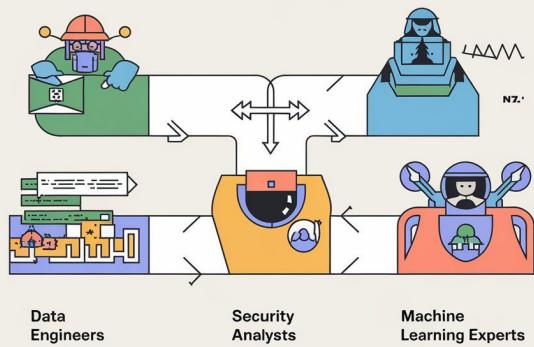
MODULAR EXECUTION

Ability to run stages independently or end-to-end

FLEXIBILITY

Once all the stages are completed, the results are merged back into a single event, integrating the outputs from each stage.

APACHE BEAM ENABLES MODULARITY.



BEAM
SUMMIT

Conclusion

Recap of challenges and solutions

Importance of Apache Beam in our success

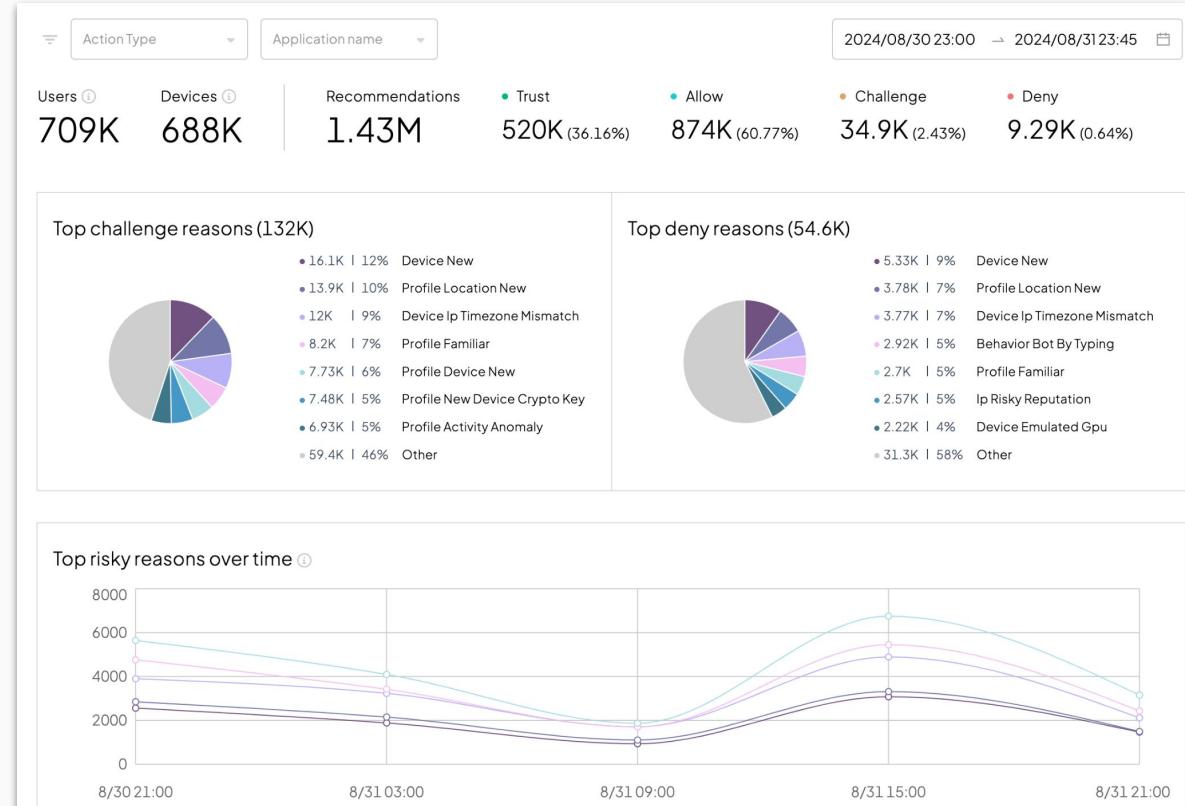
System adaptability and power

Inspiration for leveraging Beam in other projects



BEAM
SUMMIT

Transmit Security DRS product



3EΔM
SUMMIT

Transmit Security DRS product

Recommendations

Gain insights into your threat landscape by exploring recommendations and reasons. [Learn more](#)

30/08/2024|23:00 ~ 31/08/2024|23:50

Filters Action type Recommendation Deny IP country More filters Clear filters

Timeline

00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 23:50

User actions

Action type	Recommendation	Network	Browser	OS	Date
Authenticated	Deny	129.22.84.138	Chrome	Windows 15.0.0	8/31/2024, 11:51:32 PM
Password reset	Deny	70.213.59.215	Chrome	Android 13.0.0	8/31/2024, 11:51:13 PM
Register	Deny	107.91.56.235	Chrome	Windows 10.0.0	8/31/2024, 11:50:07 PM
Login	Deny	156.238.122.7	Chrome	Windows 15.0.0	8/31/2024, 11:49:39 PM
Authenticated	Deny	149.40.50.59	Safari	iOS 17	8/31/2024, 11:49:02 PM
Authenticated	Deny	31.146.200.148	Safari	iOS 17	8/31/2024, 11:48:01 PM
Authenticated	Deny	119.41.50.59	Safari	iOS 17	8/31/2024, 11:47:55 PM
Password reset	Deny	45.36.227.104	Chrome	Android 14.0.0	8/31/2024, 11:47:09 PM
Password reset	Deny	24.54.200.205	Edge	Windows 15.0.0	8/31/2024, 11:46:41 PM
Authenticated	Deny	172.91.0.11.182	Safari	iOS 17 X 1	8/31/2024, 11:46:57 PM

9,999 items

Show timeline Details Deny

Action ID 751c4ceb46401640fc7ea6034...	Risk score 97.8
Application name default	App URI https://www.exa...
User ID 2e69ca399e4890c031084504...	Number of logins 9
Earliest seen 11/18/2023	Last seen 8/31/2024
Device public key ⓘ 2a2a9f546f8698158ba2e4eb58fec45a27b9723c025...	
Device ID ⓘ ...rSeP6Oh0jl8pmIlsNtDF..._Oaleu4B0jIGNYsCAcg...gY	
Device fingerprint ⓘ 1128ca8bbf4c40d0a506536d92cff7677924d015bfa...	
Earliest seen ⓘ 8/31/2024	Last seen ⓘ 8/31/2024
Top recommendation reasons	
Suspected Bot	
New Device	2
Device First Seen	
New Location	
Familiar Behavior	



3EΔM
SUMMIT

Thank you!

Questions?

Hai Saadon

<https://www.linkedin.com/in/hai-saadon-61a34a74/>



BEAM
SUMMIT