AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

OPN307

OpenSearch: Building the future of search together

Jules Graybill (he/him)
Director – OpenSearch
AWS

Saurabh Mehta (he/him)
Sr. Engineering Manager – Observability
Airbnb

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

- What and why OpenSearch?
- Recap of this year
- Deep dive on the roadmap
- How you can contribute
- OpenSearch at Airbnb

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What's OpenSearch

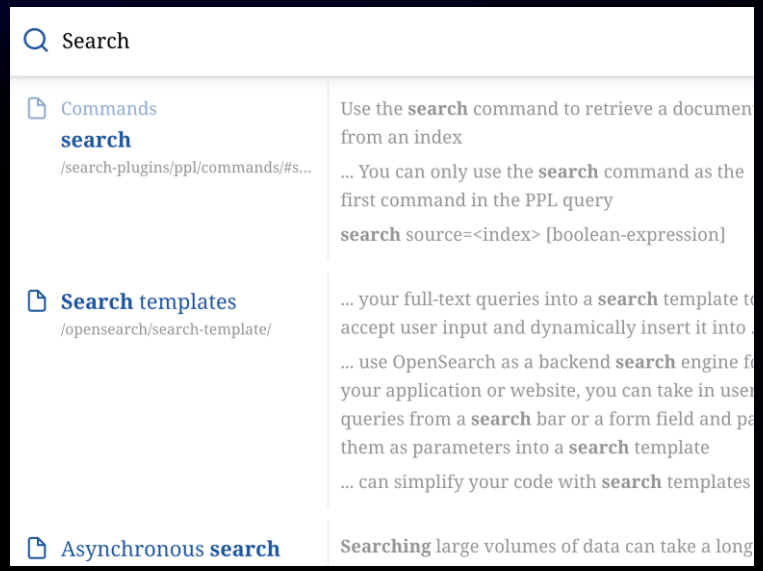


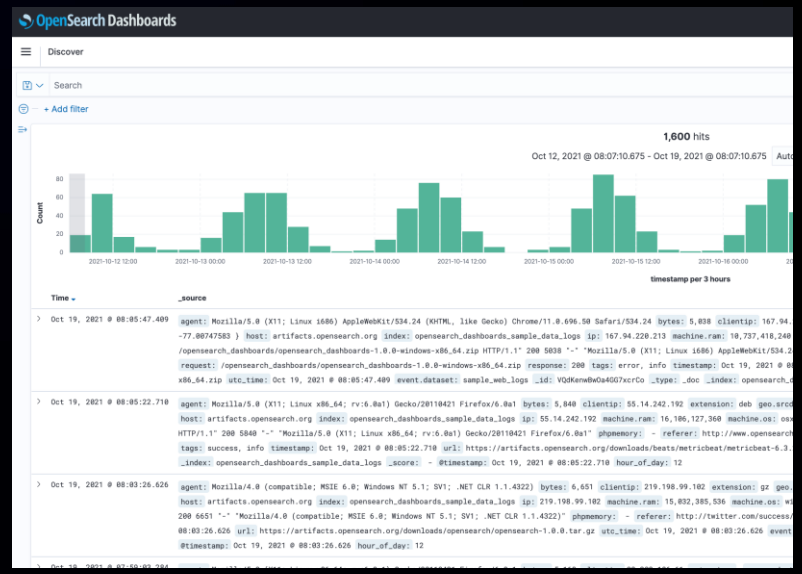
Community-driven, open-source search and analytics suite derived from Apache 2.0 licensed Elasticsearch 7.10.2 & Kibana 7.10.2.

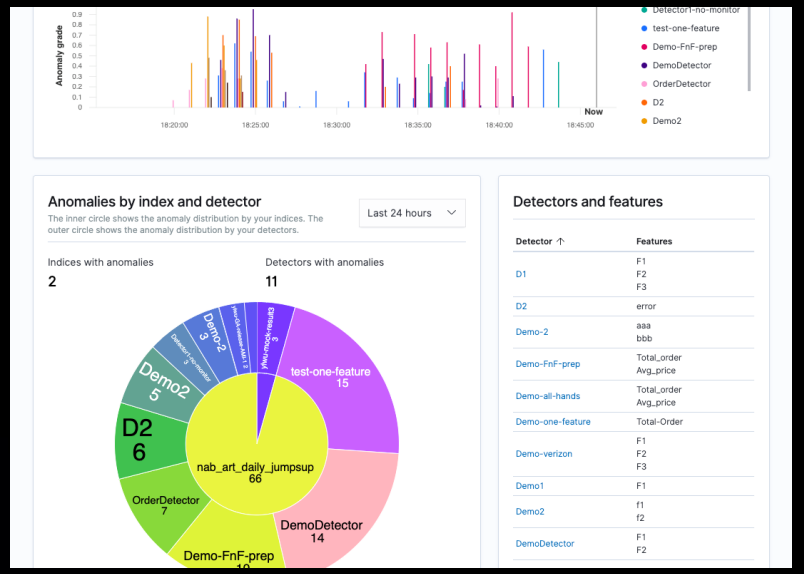
Consists of a search engine, OpenSearch, a visualization user interface, OpenSearch Dashboards, plus supporting tools and plugins.

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Used for search, analytics, and more







Text search

Natural language

Boolean queries

Relevance

Streaming data

High-volume ingest

Near real time

Distributed storage

Analysis

Time-based visualizations

Nestable statistics

Time series tools

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Search, analyze, and visualize log data



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why OpenSearch?

Community-Driven

AWS is the steward

Made with your input

Public roadmap

Welcome contributions

Co-maintainers

100% Open Source

Apache License, Version 2

Use, modify, monetize

Built in the open

No contributor license agreement

Full Featured

Wire-compatible with Elasticsearch 7.10.2

Full text search

Analytics

Fast ingest

Machine learning

Security

... and more

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What does it mean for you?



A path forward: Open-source Elasticsearch users can keep advancing

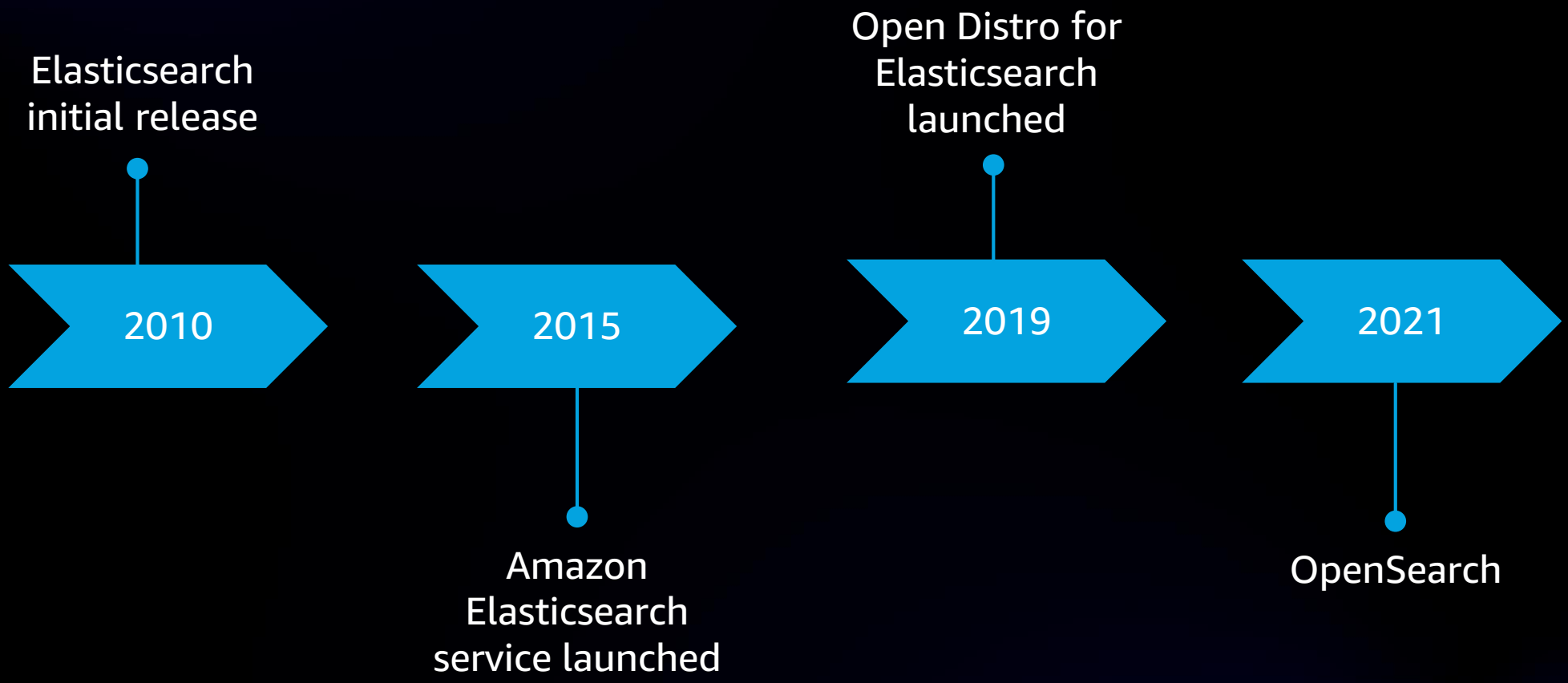
No lock in: The same software you get from us can be run on-premises or in any other cloud

Compatible: No code changes required – including new universal language specific clients

Innovative: Dozens of new features in just first few releases including: *Cross-cluster replication, shard-level backpressure, notebooks, data streams, index transforms, bucket level alerting, and more*

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

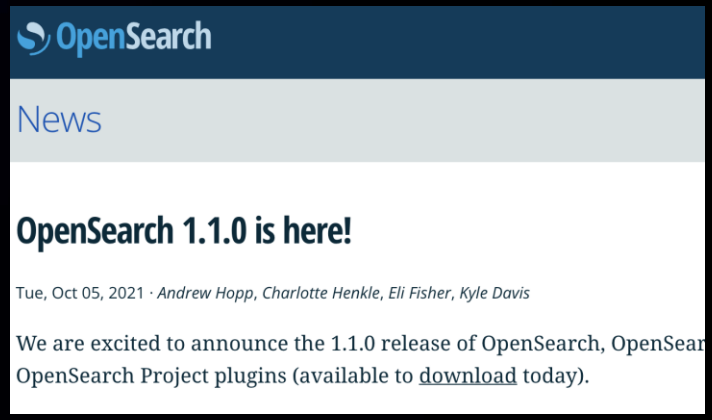
History of OpenSearch at AWS

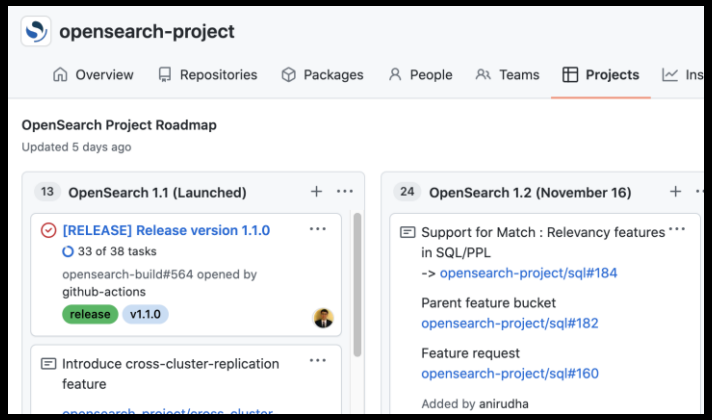


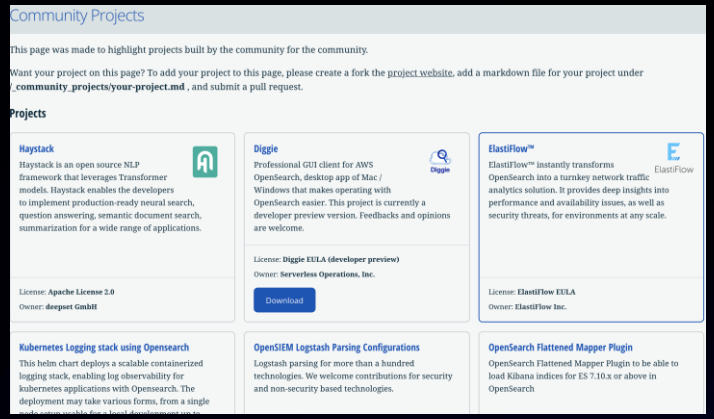
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

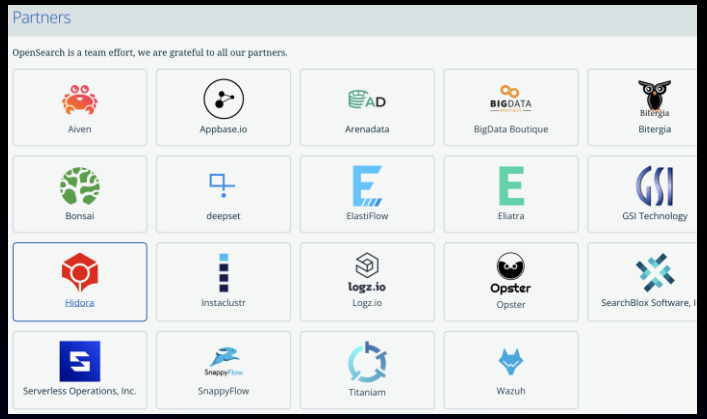
What happened since announcing?

- New versions and clients
- Public roadmap
- Community projects
- OpenSearch partners









© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

And dozens of new features…

Cross-cluster replication

Data streams

Lucene 8.9

Data Prepper

Random cut forest 2

JavaScript client

Bucket-level alerting

Unified Anomaly Detection UI

Go Client

Python Client

Upgrade Tool

Index Management

Granular Security

ARM 64

Output plugin for Logstash

Log Event Explorer

Piped Query Language Auto Complete

Docker Images

Relevancy features in SQL and PPL

Log Explorer Charts

Unified Notifications

Custom Dashboard Branding

Shard-level indexing backpressure

Alerting on HTTP Inputs

Overallocated shard prevention

k-NN with FAISS

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

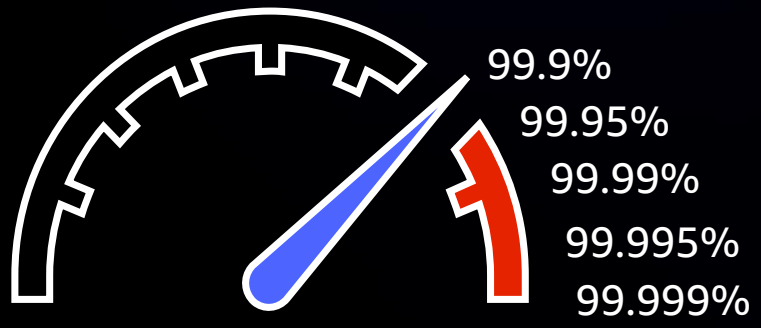
Let’s dive into two

High availability with cross-cluster replication

Real-time Anomaly Detection

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Achieving high availability



Critical apps need 24x7 operation

Each step in availability is increasingly harder

Sometimes you need more than one cluster

- Dealing with disasters
- Getting closer to your users

Synchronizing can be a challenge

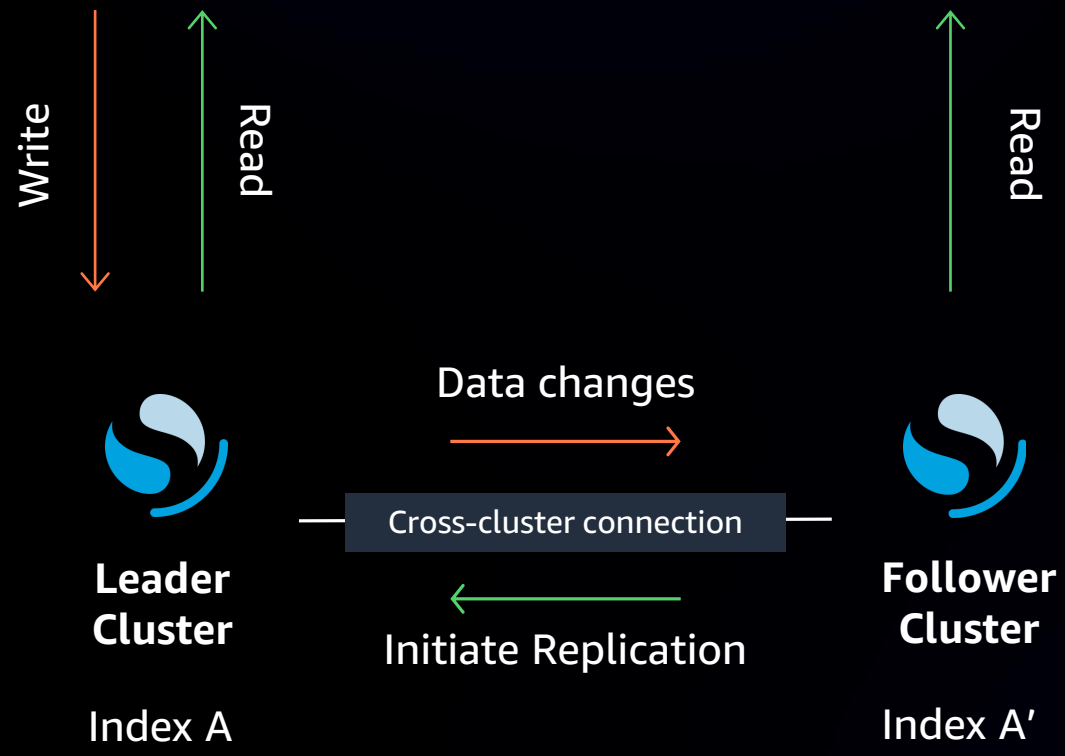
Availability %	Downtime per year	Downtime per month
99.9% ("three nines")	8.77 hours	**43.83 minutes**
99.95% ("three and ½ nines")	4.38 hours	21.92 minutes
99.99% ("four nines")	52.60 minutes	4.38 minutes
99.995% ("four ½ nines")	26.30 minutes	2.19 minutes
99.999% (“five nines”)	5.26 minutes	26.30 seconds

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cross-cluster replication

High Availability, Disaster Recovery, and Data Proximity



OpenSearch 1.1

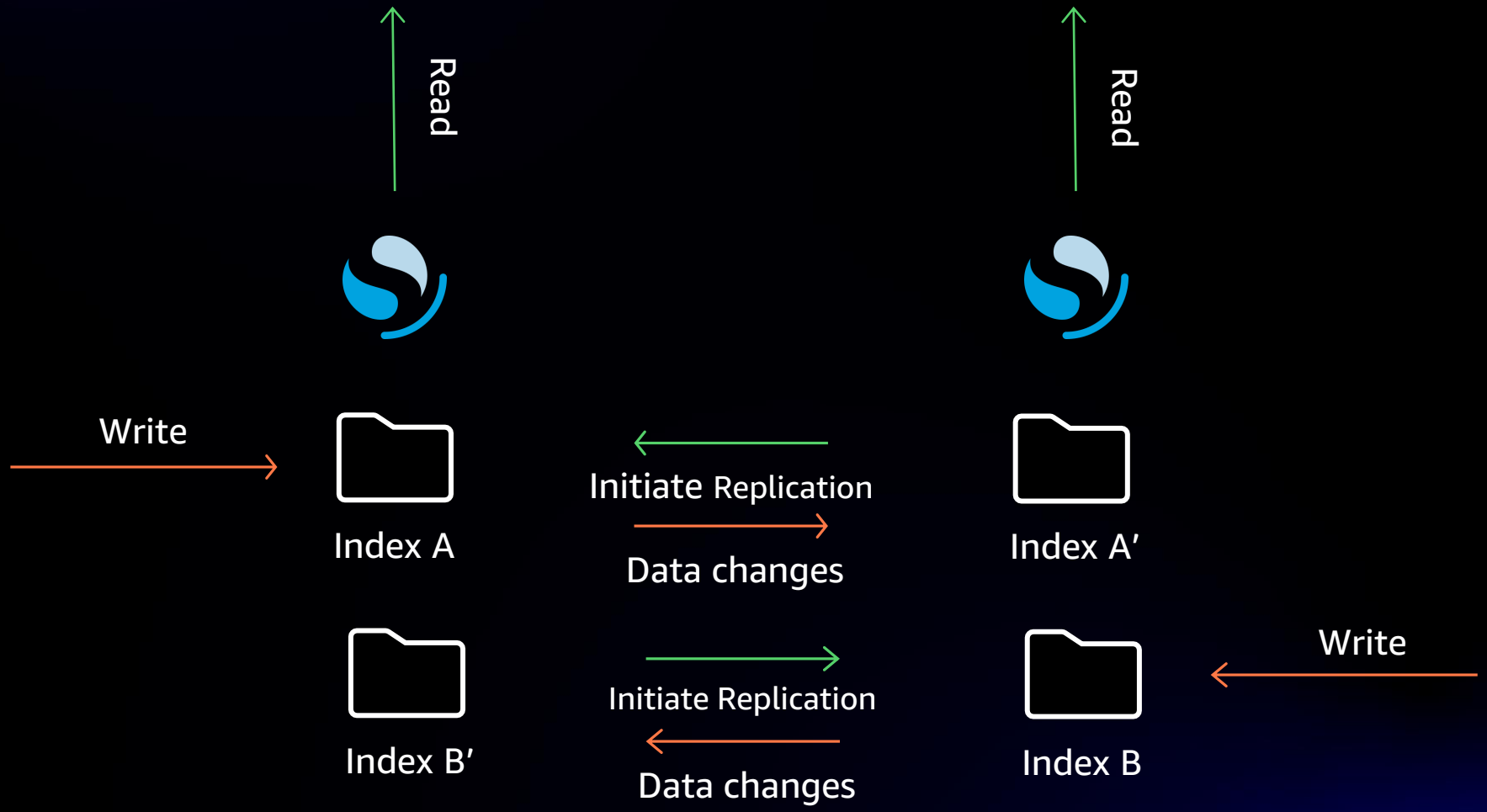
Active-passive replication

Sequential consistency and sub-minute global replication

Granular control at index level — set up many configurations

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cross-cluster replication topologies



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cross-cluster replication topologies



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Anomaly Detection in OpenSearch

Detect and mitigate issues faster with Anomaly Detection

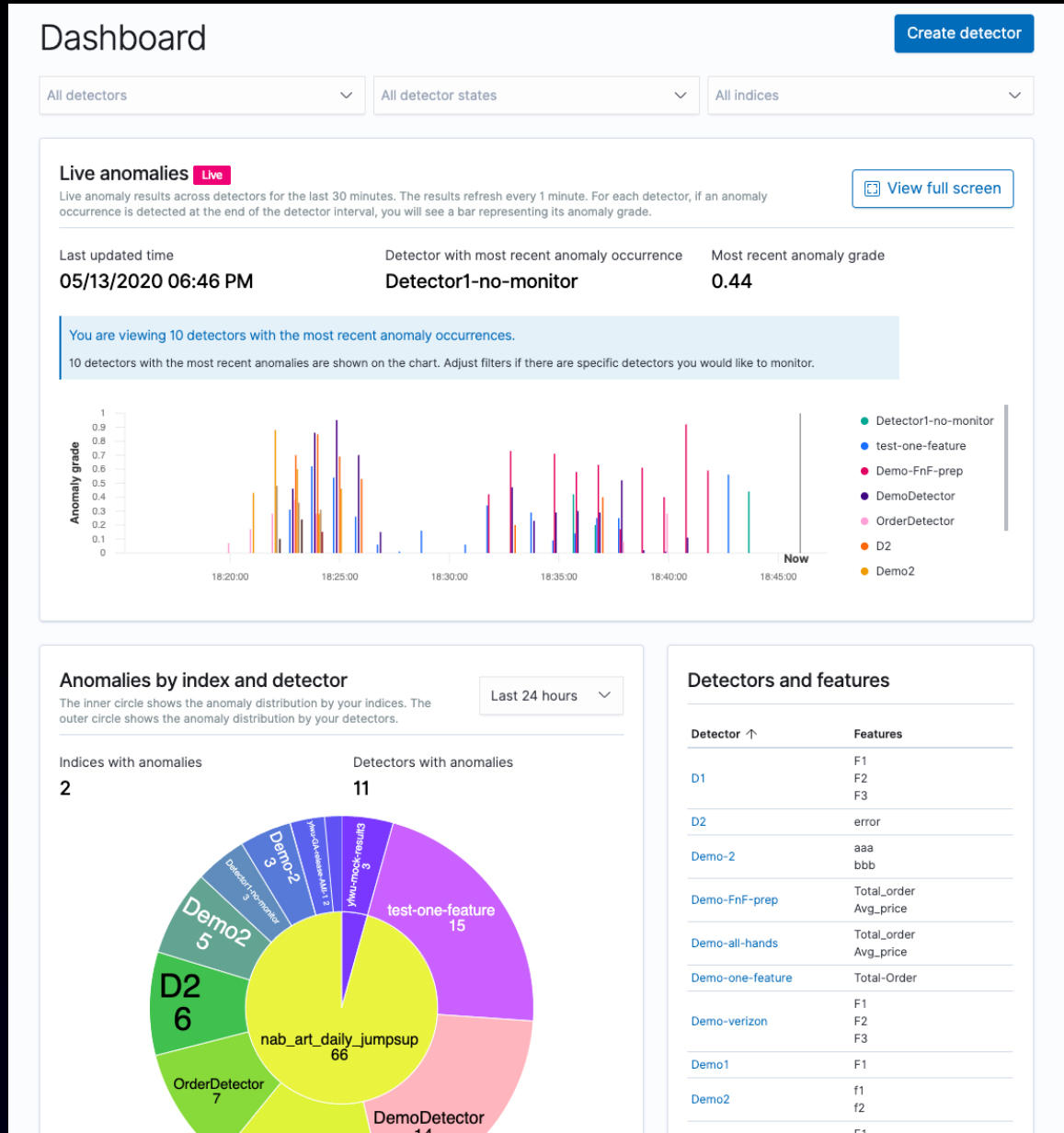
Performant at scale

- Machine learning models are distributed and processed across nodes

Easy to use

- No machine learning expertise required

Based on Random Cut Forest (RCF):
[Guha, Mishra, Roy, Schrijvers ICML 2016]



Anomaly Detection at Mobilewalla

“We integrate with hundreds of our customers using Amazon S3 API for data exchange, and we monitor their traffic patterns by analyzing S3 access logs in Amazon OpenSearch Service. **Anomaly Detection** has helped us detect changes in traffic patterns without configuring complex rules for each customer. This has enabled us to respond to those changes in a timely manner, ensuring security and reliability of our platform.”

Harikrishnan Chandrasekaran, Principal Cloud Architect



mobilewalla

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Real-time Anomaly Detection

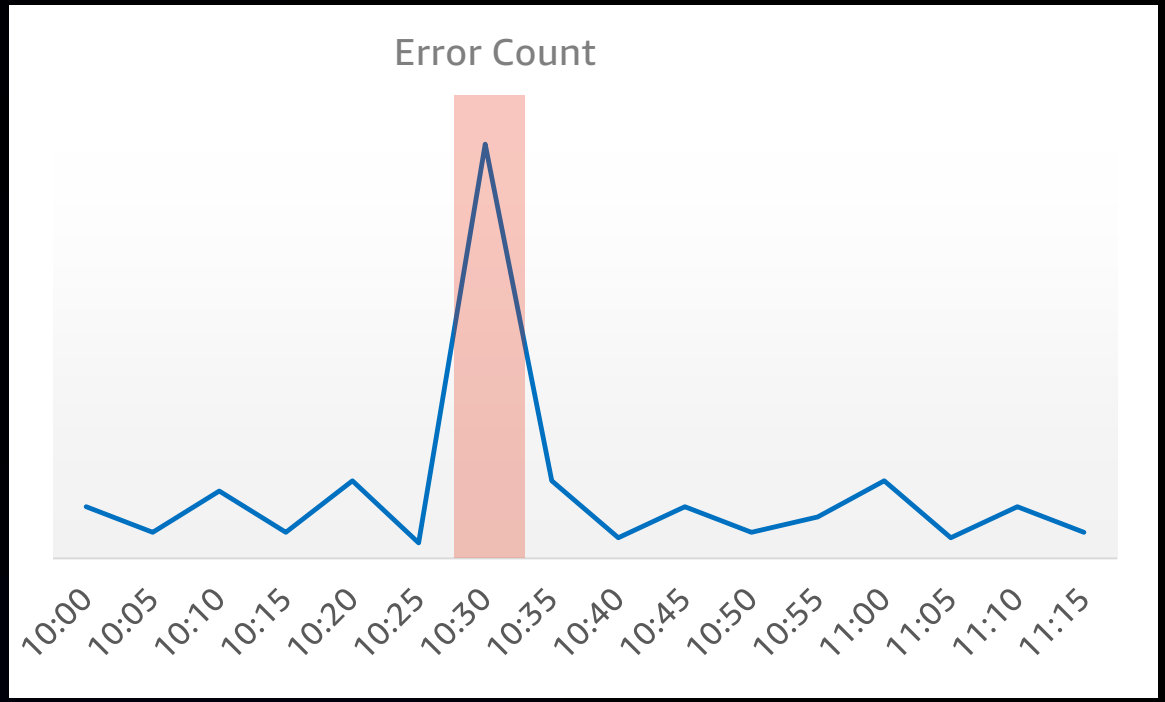
Released April 2020 in Open Distro for Elasticsearch

Dynamically detects anomalies in near real-time on streaming data

Detector supports up to five dimensions
For example: CPU, memory, disk storage, latency

Integrated with OpenSearch Alerting, providing real-time notifications on anomalies

Use cases: increases in error rates, unusual data downloads, spikes in CPU and IO usage, increases in network traffic rejections



aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

High Cardinality Anomaly Detection

Released November 2020 in Open Distro for Elasticsearch

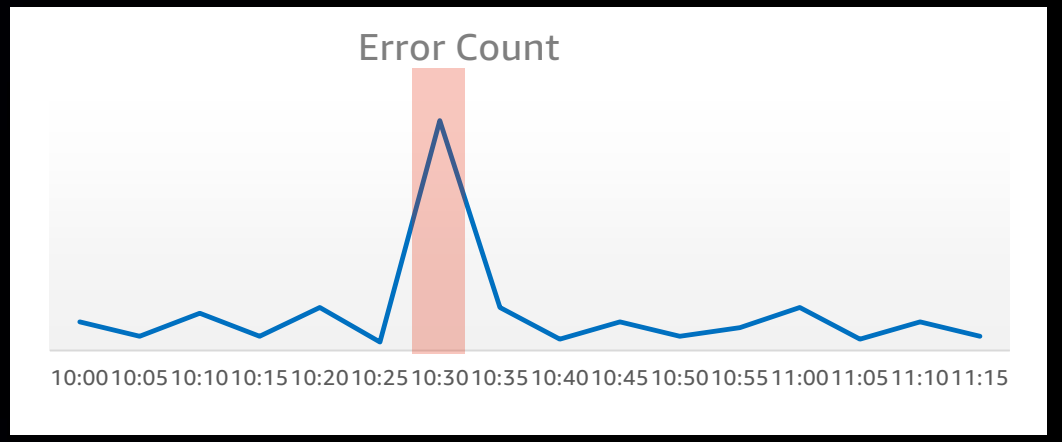
Detects anomalies on entities by ‘group-by’ category

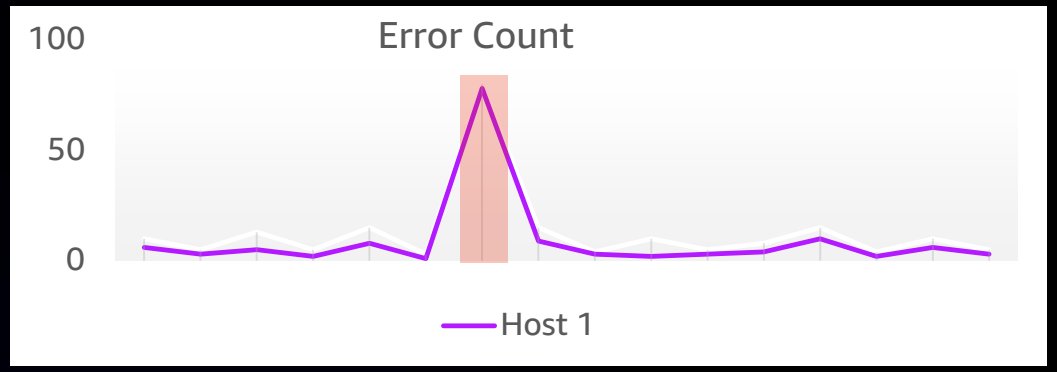
One model per entity
(e.g., host id where host id = host1, host2)

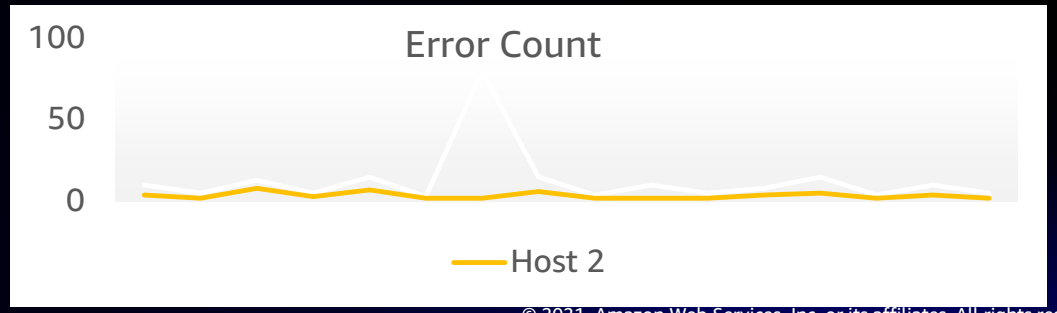
Surfaces individual entities with anomalous patterns

Use cases: identifying individual hosts or services with high error rates, isolating suspicious users, and detecting outliers in sales by Region

Only scaled to support ~1000 unique entities







© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

High Cardinality Anomaly Detection

OpenSearch 1.1

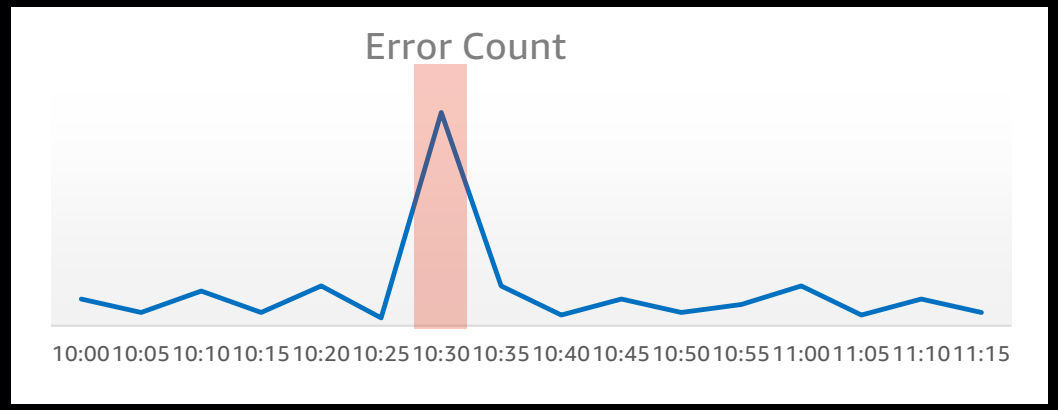
Support for 1 million entities

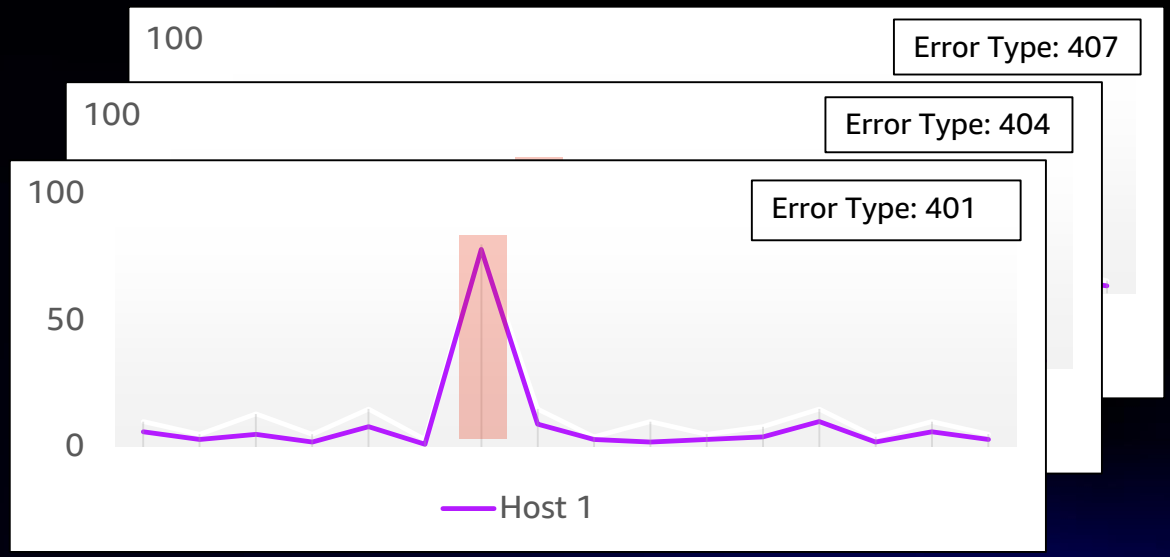
Underpinned by improved RCF algorithm (RCF 2.0)

Supports an additional ‘group-by’ category

One model per entity
(e.g., entity = host id * error type)

Use cases: identifying individual hosts with high error count by error types, isolating suspicious users accessing sensitive information by S3 buckets





© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What’s next?

- Enhancing the core
- Ease of use
- Doubling down on log analytics

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Enhancing core OpenSearch

Higher availability and resiliency

e.g., cross-cluster replication, index-level throttle, skew avoidance during failures

Improve performance

e.g., k-NN FAISS, RCF 2.0, Lucene 8.9

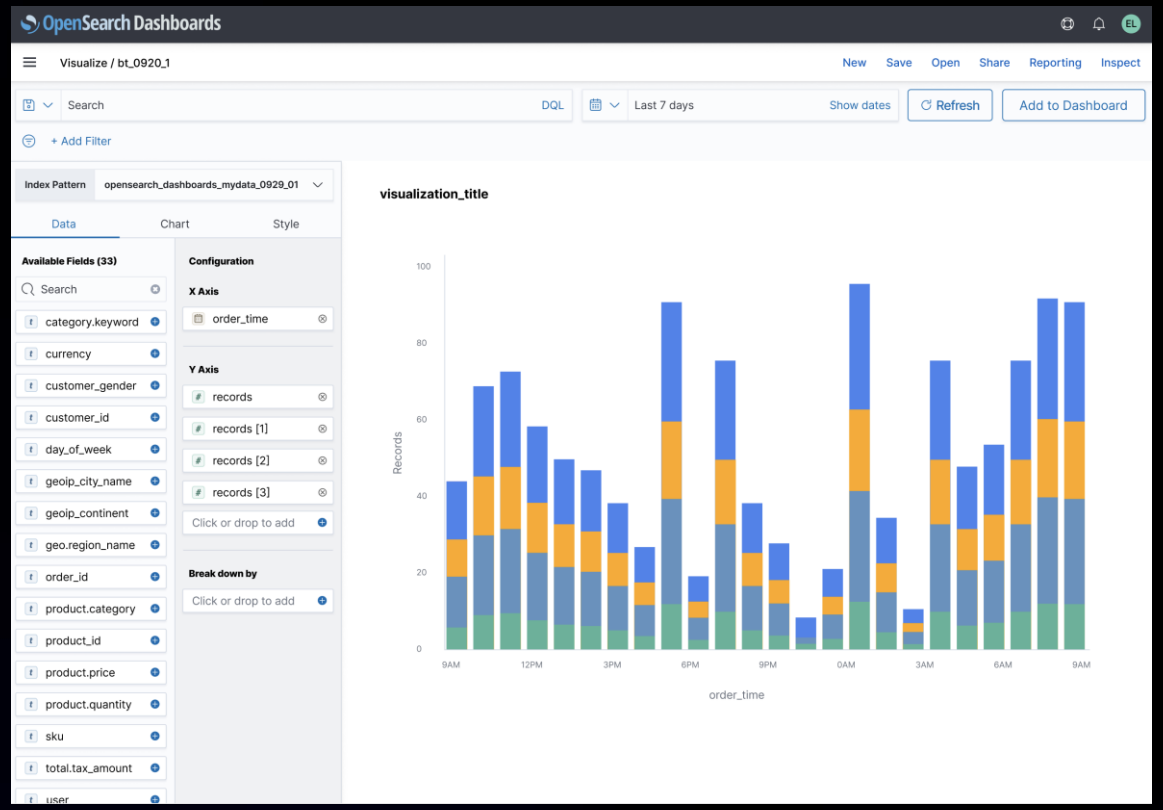
Drive down cost



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Ease of use

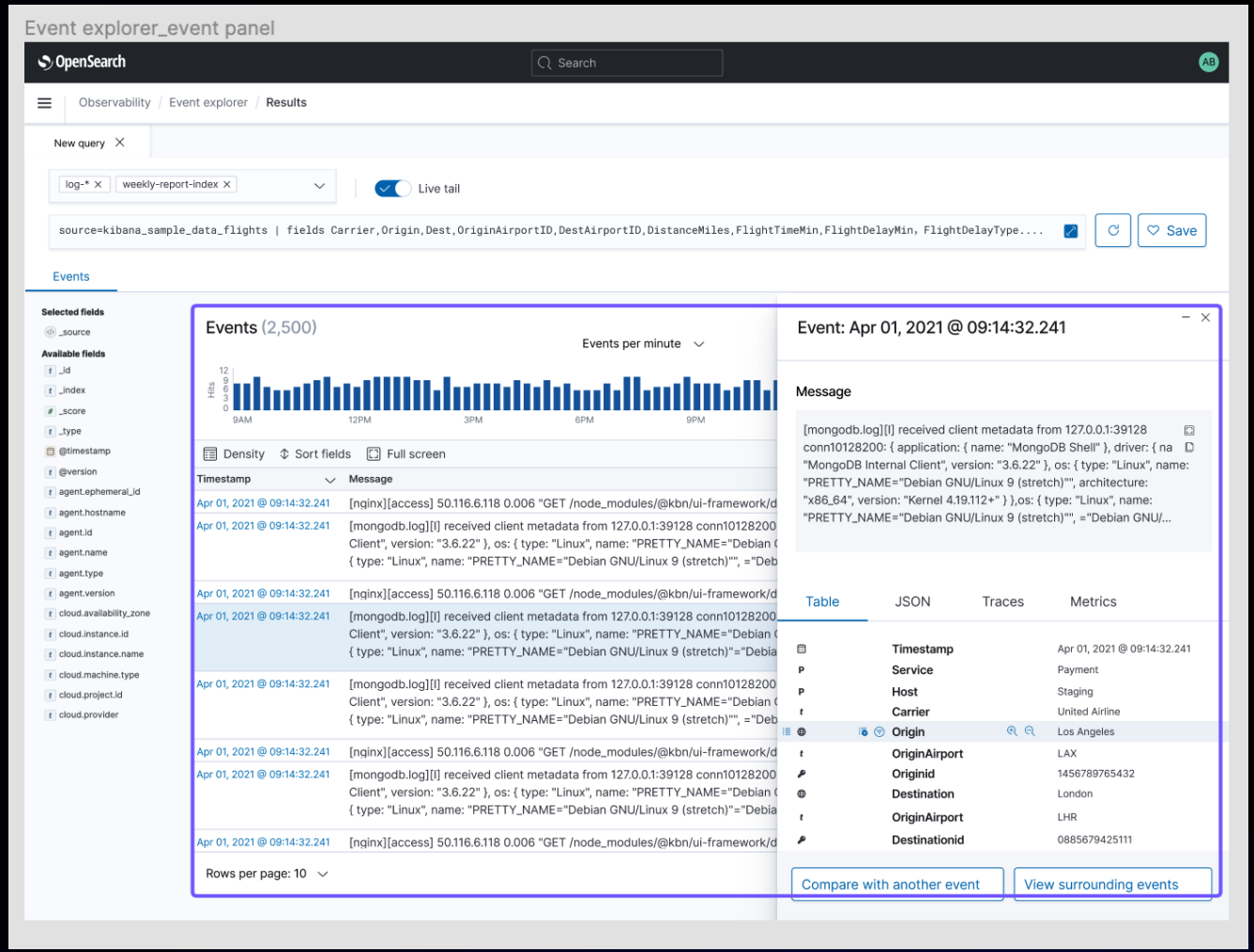
- Drag and Drop visualizations
- More release artifacts
- Anomaly Detection feature attribution
- Unified notifications



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

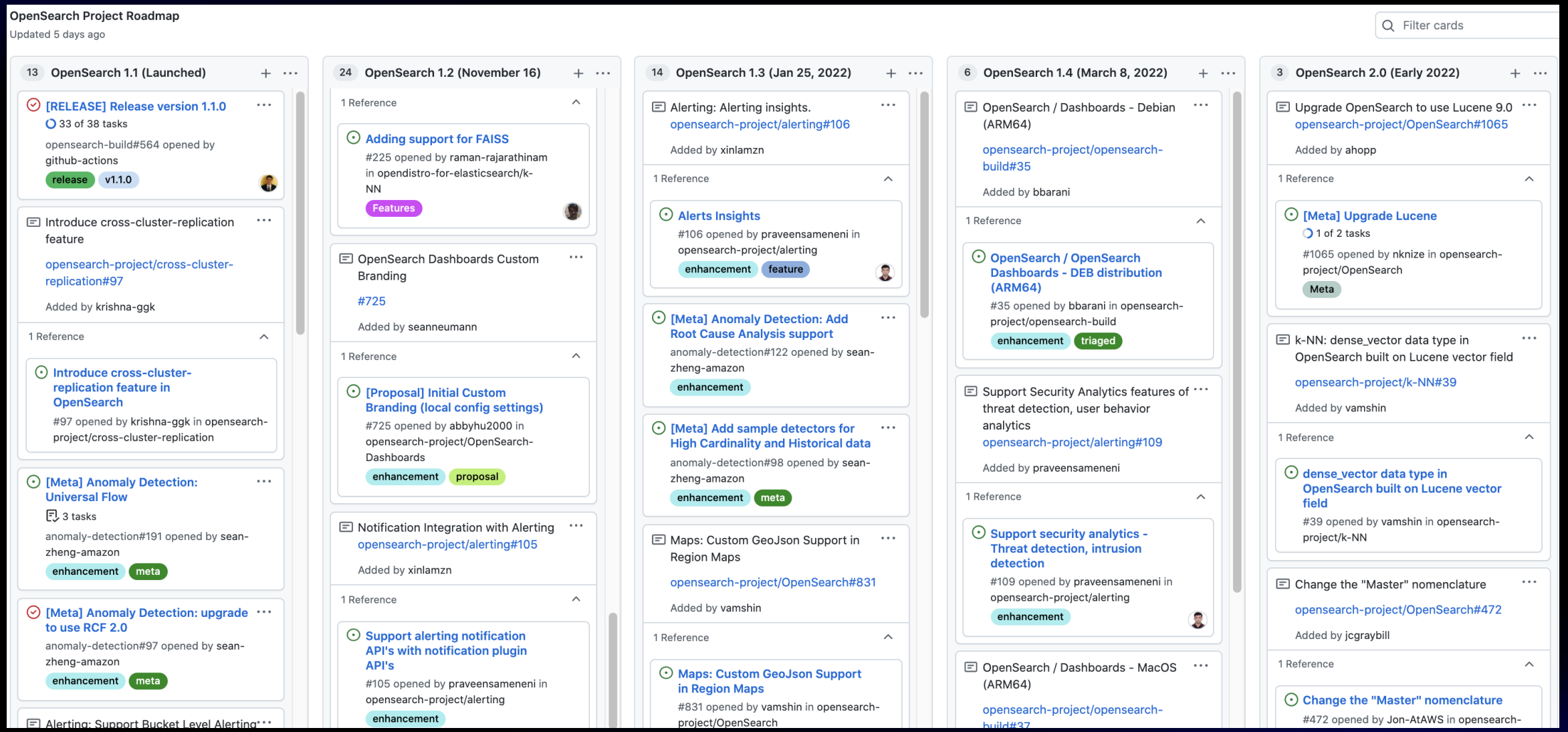
Doubling down on log analytics

- Event viewer
- Field Analysis
- Lifecycle management
- Integrated ML



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The roadmap



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How can you contribute?

Multiple ways to participate

- Community forum
- Blog posts
- Documentation
- Code

Some specific ideas

- Co-maintain a language client
- Dashboard plugin template
- Kubernetes operator
- Install artifacts and tools (Windows, brew)

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Saurabh Mehta

Sr. Engineering Manager – Observability

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

OpenSearch at Airbnb

OpenSearch for Observability

Challenges

Opportunities

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

OpenSearch at Airbnb



Observability

Logging

Tracing



Customer facing

Payments

Reservations

Reviews

Airbnb help, guidebooks

Promotions and communications

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

OpenSearch for Observability



Scalability

Used by every engineer at Airbnb

Scales at 100s of terabytes/day with 10s of clusters



Criticality

Help reduce the debugging and mitigation time for outages

Helpful in prevention and actively used during the development



Flexibility

Cost efficiency through storage tiering

Self-managed quotas for teams to manage their data efficiently

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Challenges

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Three steps



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Collection

- Multiple Data protocols
- Lack of standards for Native and Web
- Legacy codebase
- External tools (3rd party, vendors)
- Multiple side cars



Storage

- Disjoint storage (telemetry vs. logs vs. traces)
- Varying performance (write and read)
- Supporting new signal types

Access

- Multiple query language
- Multiple tools
- Lack of alerting and correlation

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Opportunities

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Collection

OpenTelemetry SDK

OpenTelemetry Collector

Drop-in instrumentation

Open source



Storage

OpenSearch (Logs, Traces)

Prometheus/Cortex (Metrics)

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What’s next



Prevention

Open standards adoption for logs, traces and metrics

Expanding the coverage for native and web (Frontend)



Detection

Signal correlation

Actionable insights

Reduce outage times



Automate

Closed-loop systems

Anomaly Detection

GitOps

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Jules Graybill

linkedin.com/in/jules-graybill/
twitter.com/JulesGraybill

Saurabh Mehta

linkedin.com/in/saumehta/
twitter.com/iSaurabhMehta

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.