
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

ANT206

What's new in Amazon OpenSearch Service

Mukul Karnik (he/him)

GM, Search Services
AWS

Rajiv Taori (he/him)

Principal PMT, Search Services
AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

Introduction to Amazon OpenSearch Service

Innovations in OpenSearch—search, log analytics/observability

What's new in Amazon OpenSearch Service—cost optimizations, security, and automation at scale

Road ahead

Q&A

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer use cases and momentum



Get started quickly



Increased performance



Event monitoring and alerting



Availability

Automate administrative tasks with cloud-first solutions

Search services

Cost, scale & security

Machine learning

Observability

Smarter results with high availability and stability



Fully open-source licensing



ML-driven search results



Stability and availability



Optimizing storage costs



Isolate critical events



Collect, transform and route data

2012 | 2015 | 2019 | 2021 | 2022

1.4 million downloads

100s of thousands of managed clusters

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

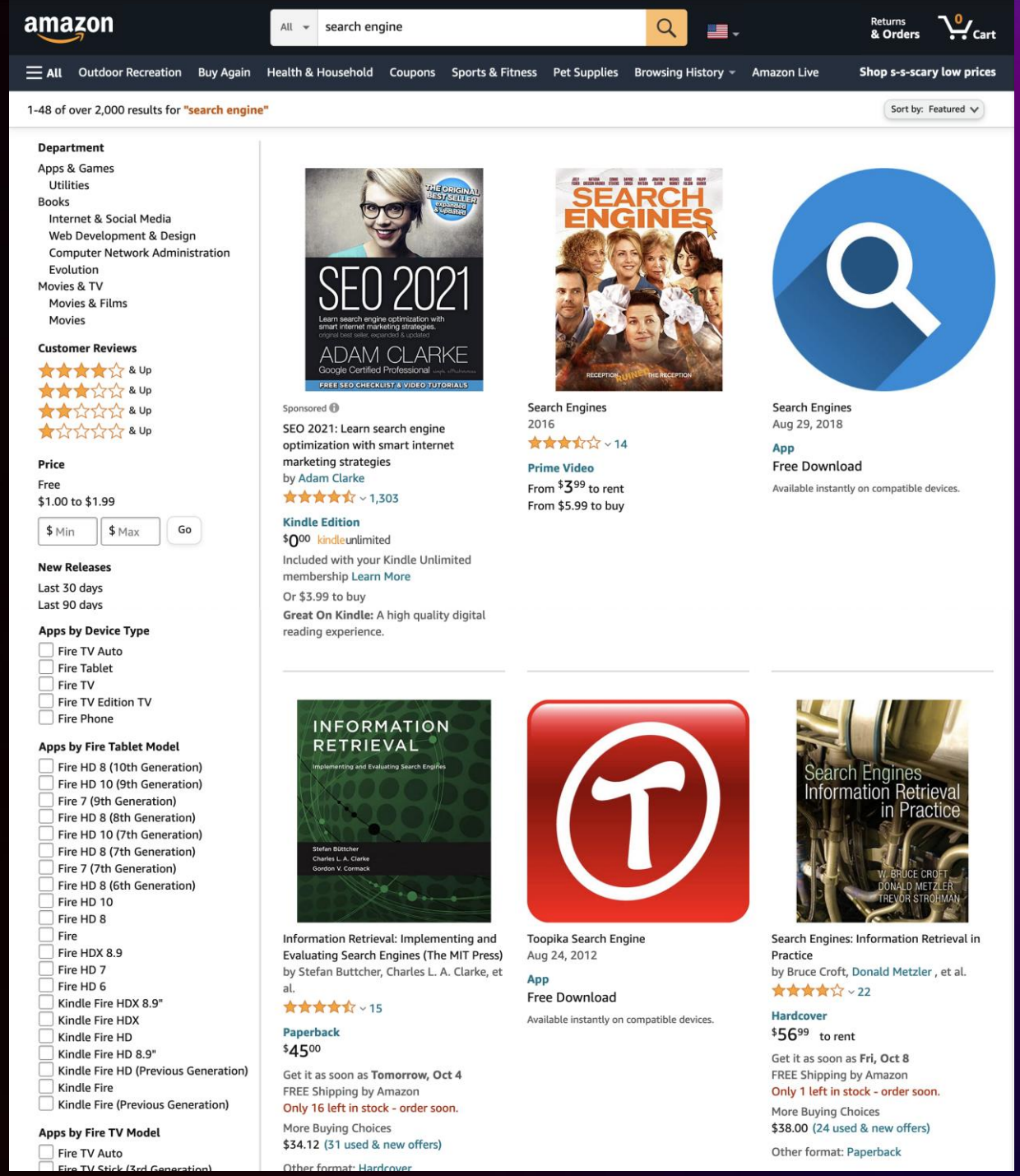
What is search all about?



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

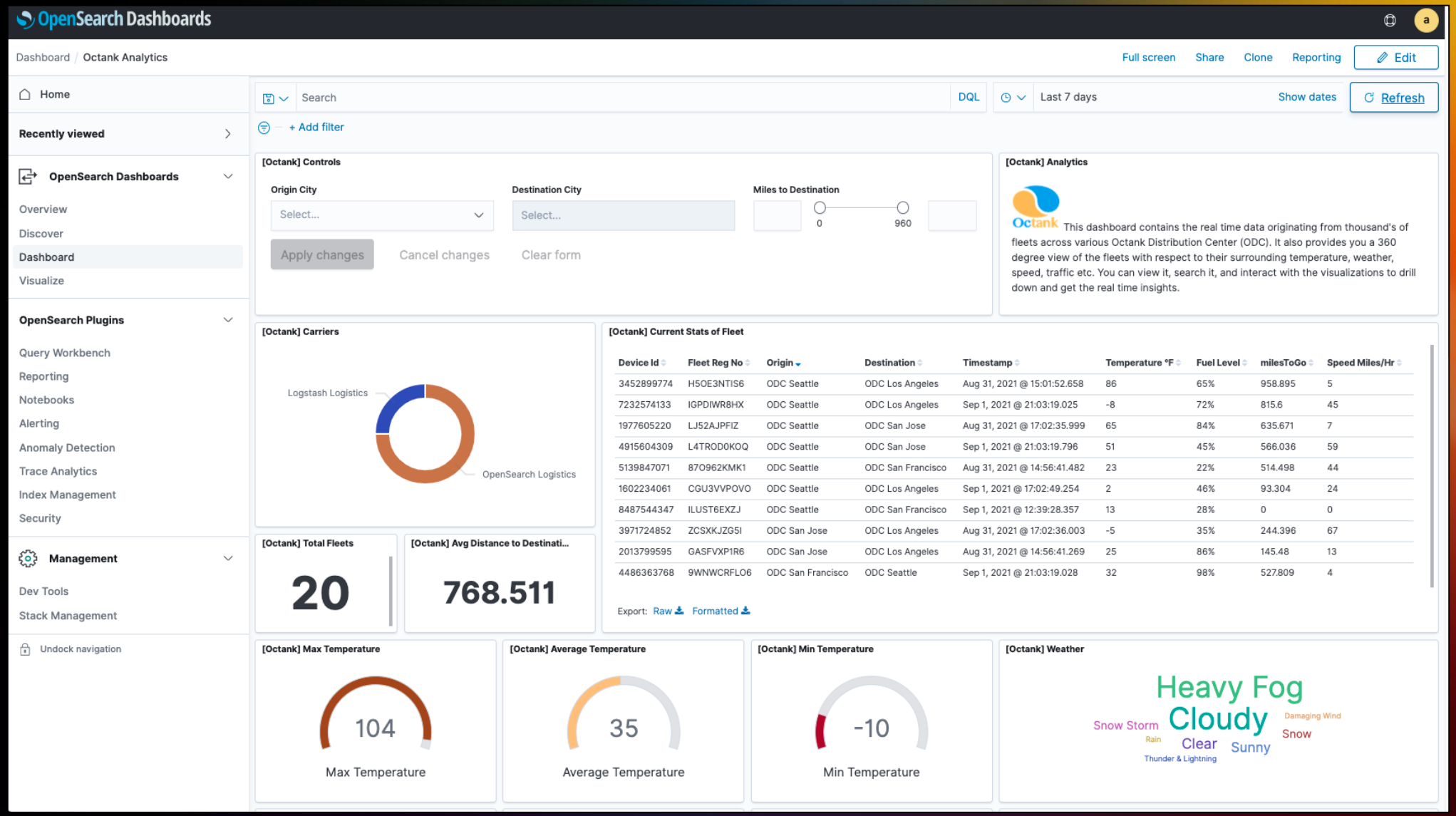
Searching

- Search engines match terms to objects in a catalog
- Match free text or structured data
- Relevance determines sort order
- Facets enable users to drill in to the search results to narrow by attribute value



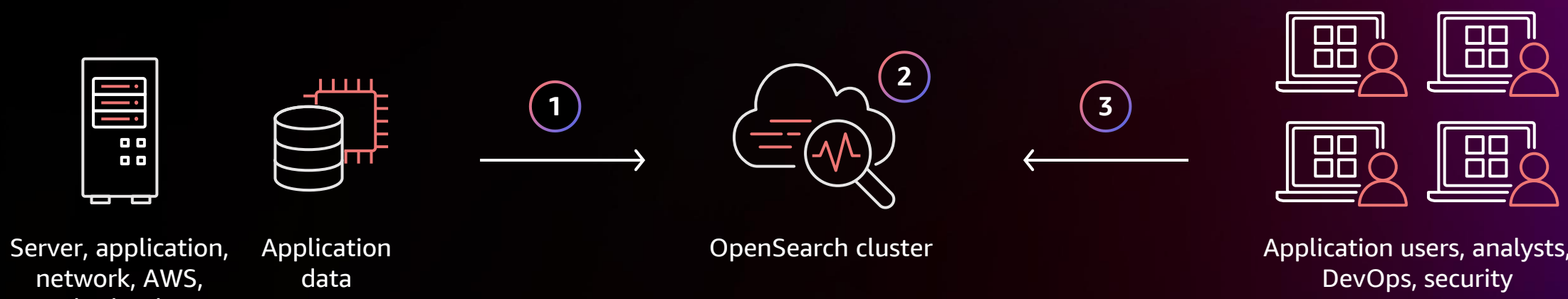
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Analyzing streaming data (logs)



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How does it work?



1. **Send data as JSON via REST APIs**
2. **Data is indexed—all fields searchable, including nested JSON**
3. **REST APIs, for fielded matching, Boolean expressions, sorting, and analysis**

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



OpenSearch project growth

>100MM
OpenSearch project downloads since launch

Top 4 search engine
DB-Engines ranking

40+
Partners and growing

10k+ pull request merged
200%+ growth

Multiple service providers
AWS, Oracle, Aiven-Azure, Bonsai-GCP

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Amazon OpenSearch Service

Amazon OpenSearch Service securely unlocks real-time search, monitoring, and analysis of operational data.



Managed: Increase operational excellence by using a popular open-source solution



Secure: Audit and secure your data with a data center and network architecture and built-in certifications



Cost conscious: Optimize time and resources for strategic work



Observability: Detect, analyze, and remediate to system issues through an open-source solution for machine learning, alerting, and visualization

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon OpenSearch Service customers

Software and Internet

Adobe, IAC, AUTODESK

Education technology

McGraw Hill Education, INSTRUCTURE, Blackboard

BioTech and pharma

MONSANTO, Bristol-Myers Squibb

Financial services

THOMSON REUTERS, stripe, KICKSTARTER

Media and entertainment

NETFLIX, FOX

Social media

cookpad, ancestry, Nextdoor

Telecommunications

COMCAST, T-Mobile

Travel and transportation

lyft, Expedia, UBER

Real estate

Zillow, move

Logistics and operations

here, elementum, infor

Publishing

FT FINANCIAL TIMES, ELSEVIER, The Washington Post

Other

Canon, SAMSUNG, British Gas

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

"Amazon OpenSearch Service enables **2 dedicated engineers** to serve **550 users and 80 applications**, which generate over **100 TB** of monthly data across 5 separate environments. We used reserved instances to realize big savings. On a per document basis, the **TCO is infinitesimally low.**"

Fred Kirwin

Sr. Software Engineer, 3M

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Innovations in OpenSearch

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

OpenSearch: 10+ releases, 100+ features

OpenSearch 1.2
- Observability interface
- Feature attribution in anomaly detection
- Shard-level indexing back-pressure
- More efficient k-NN dense vectors

Data Prepper 1.2
- Compatibility with Logstash
- Fluent Bit
- Grok processor

OpenSearch 2.0
- Lucene 9.1 performance optimizations
- Document-level alerting
- RPM package manager

OpenSearch 2.2
- Logistic regression
- Lucene HNSW implementation
- Custom GeoJSON
- RCFSummarize algorithm
- Support in region maps

Data Prepper 1.5
- S3/SQS source processor

OpenSearch 2.4
- Security analytics
- Windows support
- Log patterns
- Metrics from logs
- Federation with Prometheus metric store
- Point-in-time search
- Geo Hex Grid Aggregation
- KNN Pre-filter support

Q1'22 Q2'22 Q3'22 Q4'22

OpenSearch 1.3
- App analytics, Trace ID Correlation, and live tail in observability
- PPL runtime support
- K-means and Random Cut Forest algorithm support

Data Prepper 1.3
- Aggregate, drop, date, mutate, KV processors
- Rolling index setup with date and time patterns in index names

OpenSearch 2.1
- Automate snapshots with snapshot management
- Dedicated resources for ML workloads
- **Multi-terms aggregation**

Data Prepper 1.4
- Metric processor with OTEL collector as source

OpenSearch 2.3
- Segment replication
- Remote-backed storage
- Drag-and-drop visualization
- OpenSearch playground

Data Prepper 2.0
- Multi-node aggregations
- Conditional routing
- JSON parser
- CSV parser

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Innovations in OpenSearch: search

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

ML innovations for Search

New algorithms added to support search and analytics use cases

- k-NN with FAISS library for efficient similarity search and clustering
- Similarity search via Lucene 9.1 (OpenSearch 2.x) based approximate k-NN
  - Up to 30% improvement in index throughput for high-dimensional vectors
  - Up to 10% faster nearest neighbor searches on high-dimensional vectors

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Improving search relevance

Semantic Search
(experimental feature in OpenSearch 2.4)

- Semantic search using neural models
- Increase relevance of searches by combining neural models with standard free-text search (BM25)
- Language model fine-tuning based on the document set

Search result comparison tool

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Demo: OpenSearch semantic search

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Innovations in OpenSearch: log analytics/observability

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Downtime costs money (and stress)



87 hours per year

Of downtime per year*

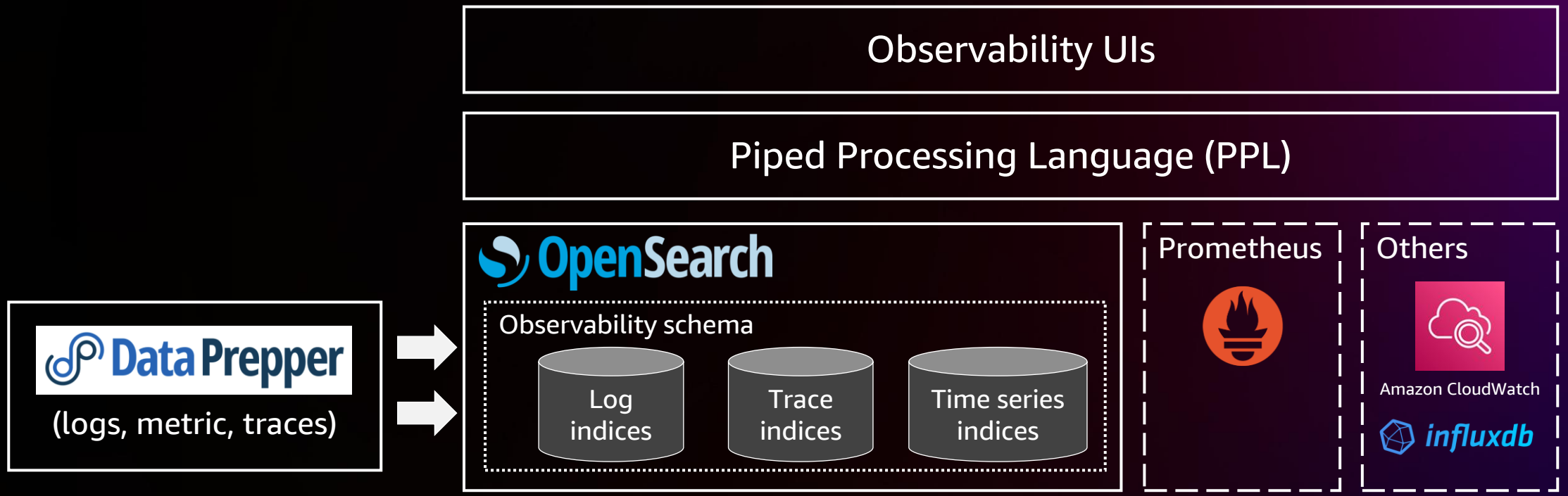
- $42K per hour avg hourly cost
- Opportunity loss, cost of not doing business
- Penalties of not meeting SLAs
- Brand alienation
- Interruptions to business

*Gartner

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

OpenSearch observability

Open source | Logs | Metrics | Traces



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

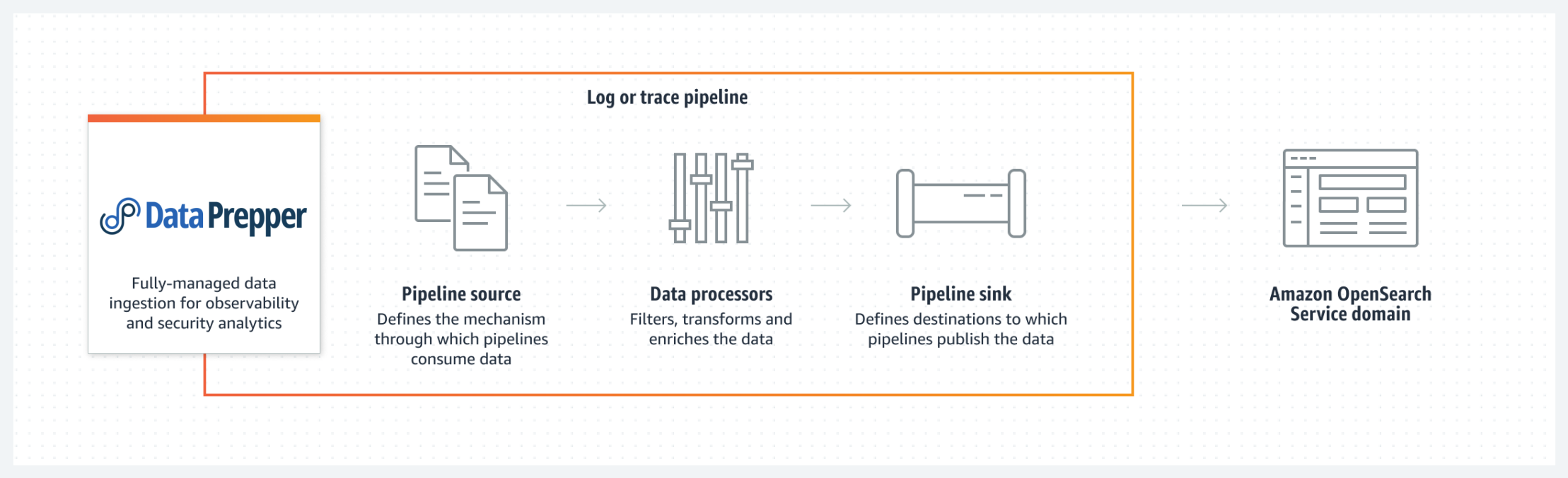
Why use data pipelines?

Benefits	Capabilities
Cost	Data deduplication, filtering, sampling, routing to lower cost storage
Data quality	Enforce standards, normalize data, adopt schemas
Privacy	Redact and obfuscate sensitive data; control routing
Streaming analytics	Enrich, inference processing

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

OpenSearch Data Prepper



Source plugins

- OpenTelemetry
- Fluent Bit
- S3/SQS
- JSON, CSV parser

Processor plugins

- Trace
- Service map
- Grok
- Drop
- Date
- Mutate
- Key-value
- Aggregate
- Metric

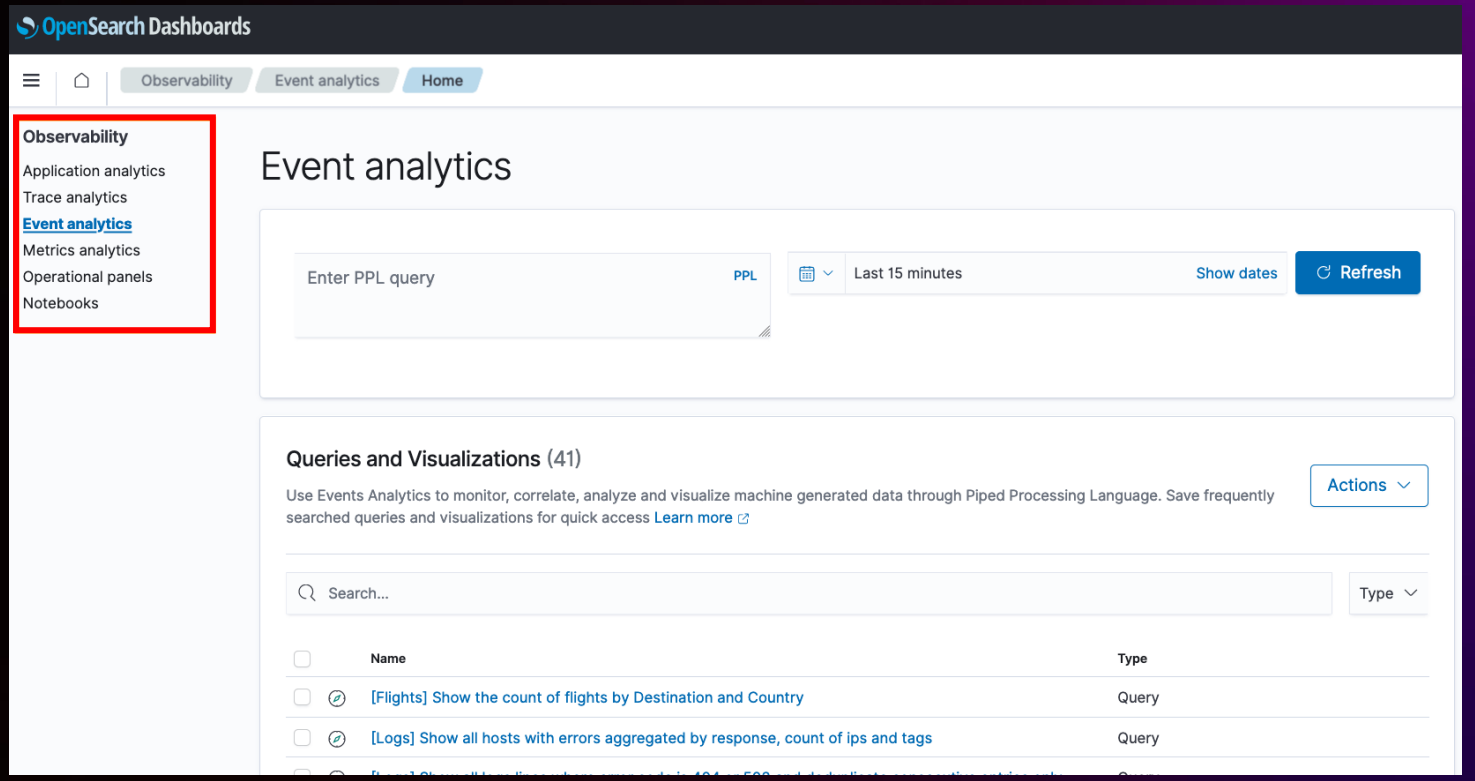
Sink plugins

- OpenSearch
- Conditional routing
- Rolling index config

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

OpenSearch observability

- Distributed tracing — trace analytics
- APM — application analytics to correlate logs and traces
- Log monitoring, event analytics — expanded log analytics capabilities
- Metric analytics, metrics from logs, Prometheus metrics *(OpenSearch 2.4)*

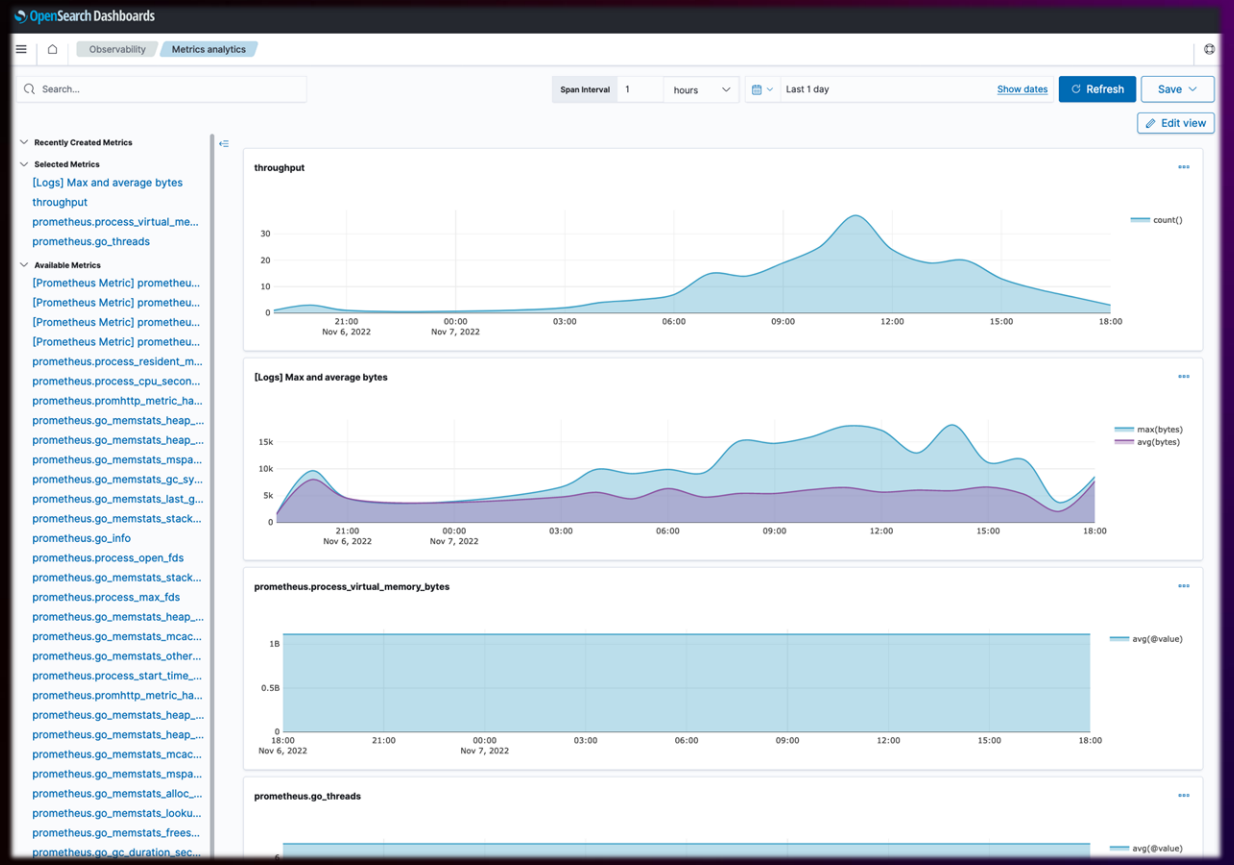


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Metric analytics

COMPREHENSIVE ANALYSIS WITH CORRELATED METRICS, TRACES, AND LOGS

- Metrics generated from aggregating log data
- Metrics from external sources like Prometheus *(OpenSearch 2.4)*
- OpenTelemetry-compliant schema ensures interoperability
- Drive low MTTRs and conclusive RCAs

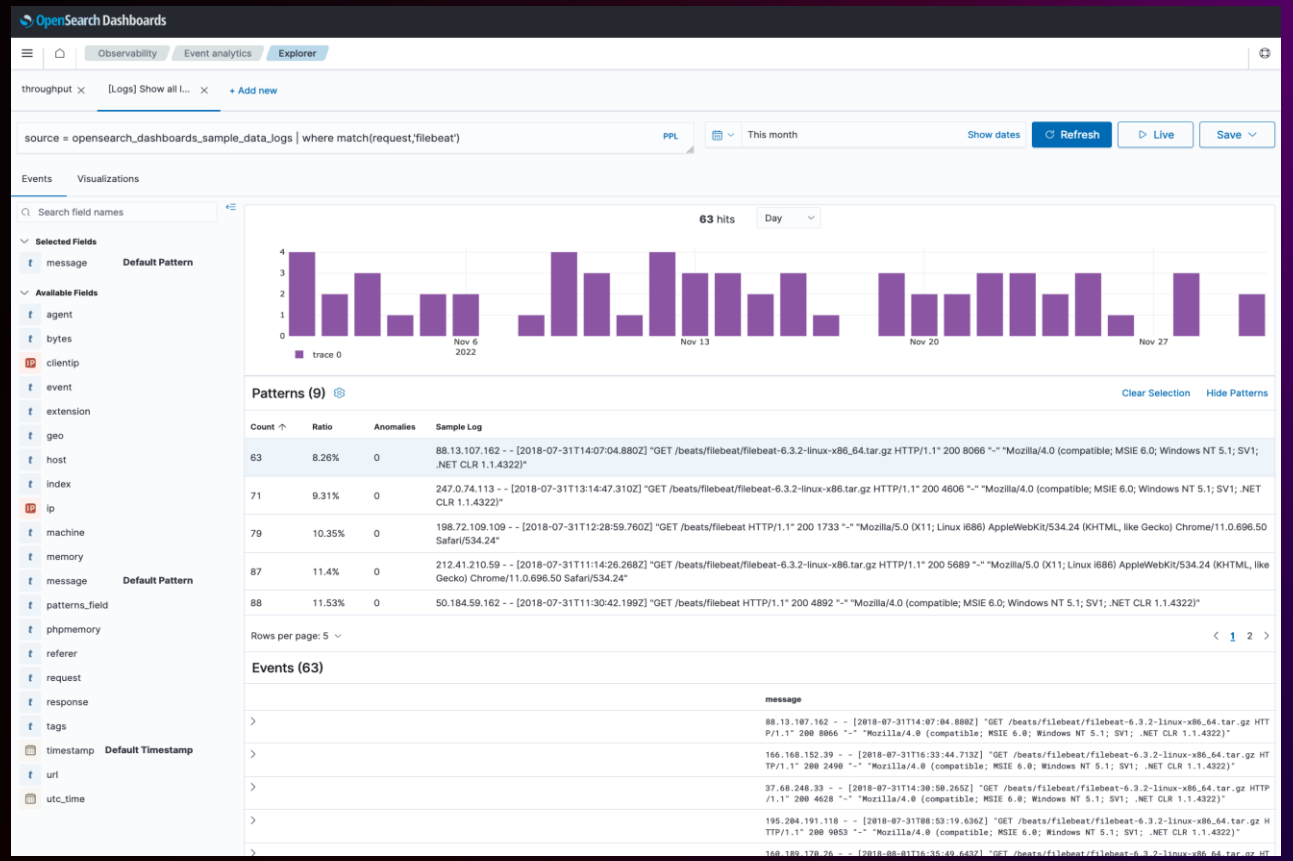


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Log patterns

GET AUTOMATIC INSIGHTS INTO MOUNTAINS OF LOG DATA

- Detect unexpected outliers
- Detect common “non-interesting” patterns for drop filters
- Patterns based on recurrence of event signatures
- Builds on log tailing, log surround, and log monitoring capabilities

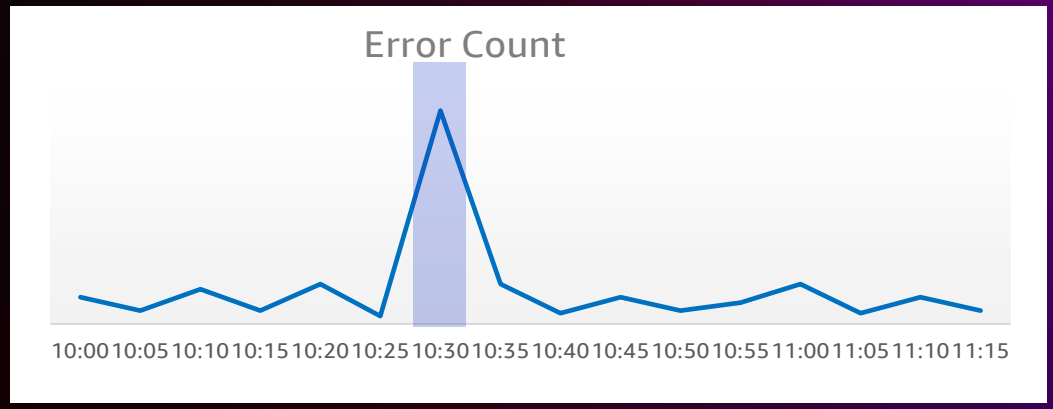


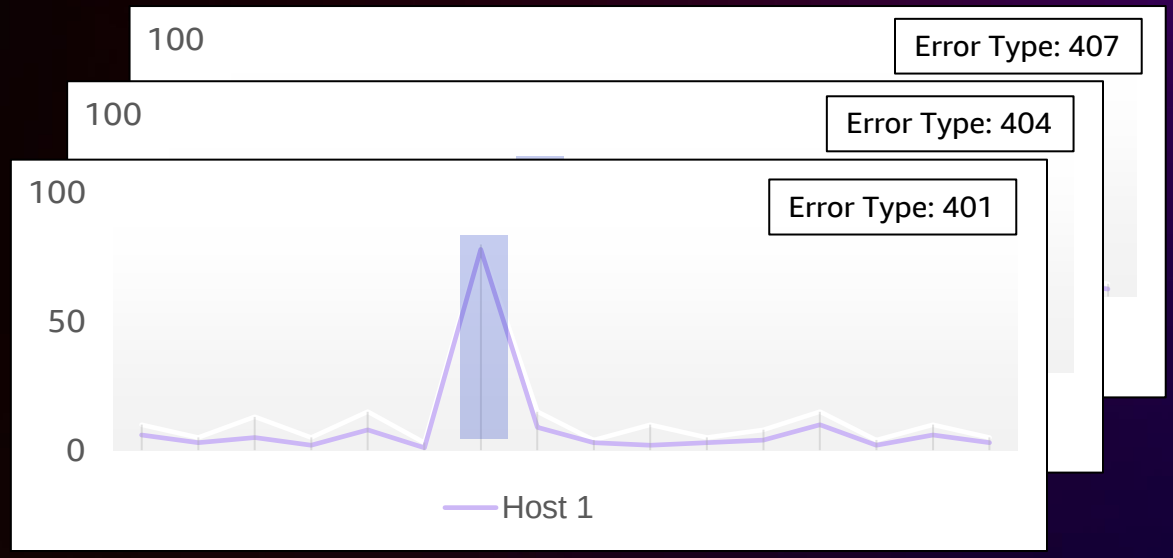
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Anomaly detection enhancements

Support for 1 million entities in 1 minute

- One model per entity (e.g., entity = host id * error type)
- *“Shorten the interval to one minute for one million entities, without changing the model output or increasing the number of nodes.”*





aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

"At Dow Jones, observability is critical when developing and managing applications—especially as we adopt more microservices and service mesh techniques. We use Amazon OpenSearch Service trace analytics and log analytics to optimize service performance and troubleshoot issues, making our services more reliable. The open-source implementation and support for open standards like OpenTelemetry is vital to our strategy. We were able to grow our logging and trace data collection 10x while keeping costs flat."

Chris Nelligan

Senior Vice President, PIB and Platform Engineering, Dow Jones

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What’s new in Amazon OpenSearch Service

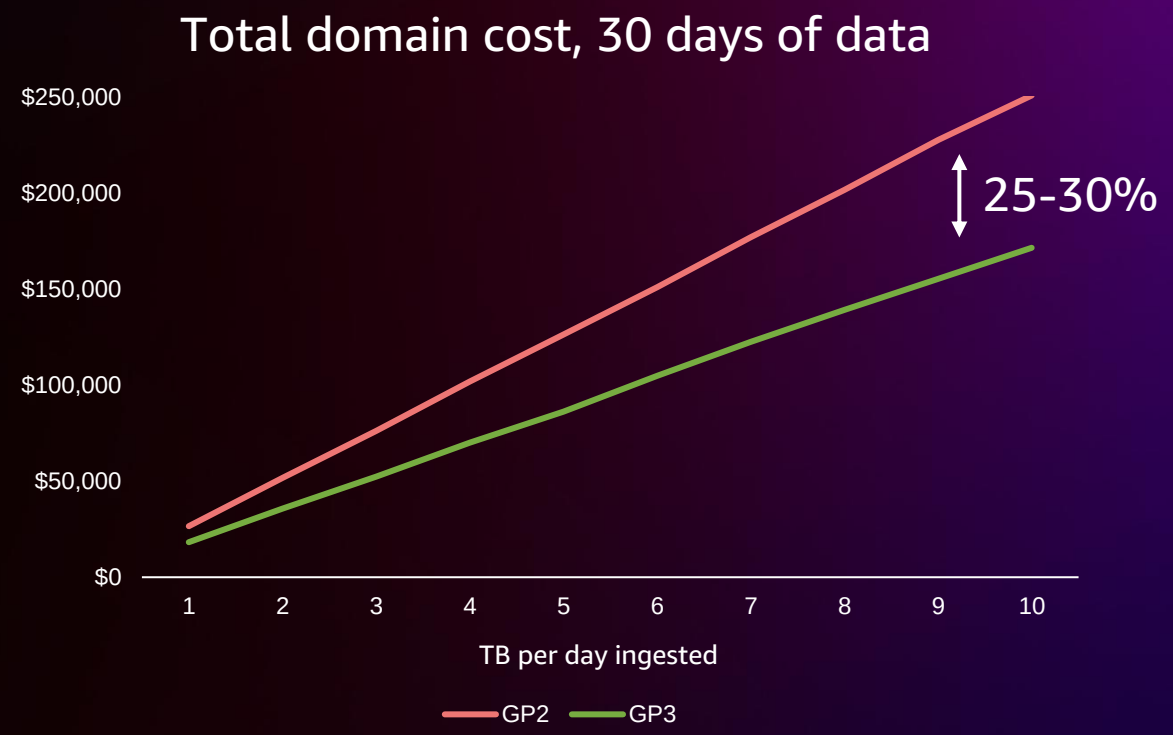
Cost optimizations, security, and automation at scale

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Support for GP3 (General Purpose SSD) volume type

Optimize Costs at Scale with GP3

- 2x denser storage on R5, R6g, M5, M6g instance families
- Available at a 9.6% lower price point per GB than existing gp2 volumes
- Provision IOPS independent of storage capacity
- Customers can migrate from their existing gp2 volumes, by changing their domain configuration volume type to gp3

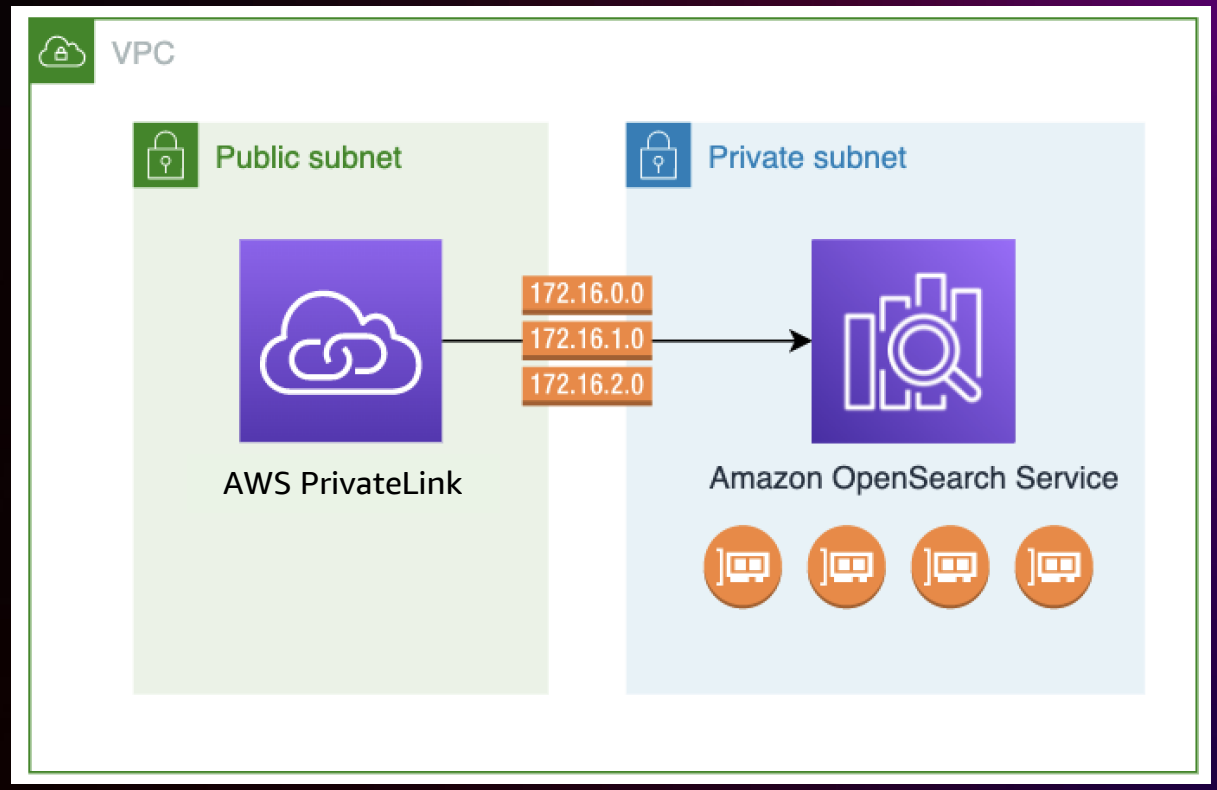


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

New security feature highlights

- AWS PrivateLink support for simplified network management for secure communications

- Fine-grained access control for existing domains
  - Including support for SAML, audit logging, document and field level security, security for plugins and cross-cluster search



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Simplifying operations at scale



Config change validations

Blue/Green deployment check when making config changes

Blue/Green progress visibility

EventBridge/CloudWatch events and console notifications for blue/greens



Operations automation

Admission control for Auto-Tune

Shard-level indexing backpressure

Notifications: Red cluster, snapshot restore, low disk space, high shard count, EBS burst balance, throughput throttling

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Road ahead

OpenSearch



Amazon
OpenSearch Service

https://github.com/orgs/opensearch-project/projects/1

- **Log analytics/observability**
  (e.g., trace data metrics, CW metric and CW log integration, alerting and anomaly detection integrations)
- **Search**
  (e.g., semantic search, external rerankers, search comparison tool)
- **Durability and reliability**
  (e.g., monitoring and operational visibility, search backpressure)
- **Extensibility**
  (e.g., new plug-in framework, extensions catalog)
- **Performance, cost, and scale**
  (e.g., segment replication, remote storage)
- **Machine Learning**
  (e.g., anomaly detection on streaming data)

Introducing Security Analytics (experimental feature in OpenSearch 2.4)

SECURE YOUR BUSINESS DATA AND RAPIDLY DETECT POTENTIAL THREATS



- **Sigma rules**: Out of the box support for 2000+ opensource Sigma security rules
- **Log types**: Out of the box support for multiple log sources including Windows, Netflow, AWS CloudTrail, DNS, AD/LDAP and more
- **Security findings**: Customize or use the default Sigma rules to detect potential security threats by filtering through the security findings
- **Security alerts**: Customize or use default Sigma rules to rapidly detect potential security threats and send alerts to a pre-selected destination
- **Rules editor**: Using an intuitive rules builder, copy default rules or create your own rules for your custom application logs
- **Insights dashboards**: See a bird's eye view of insights by log source

Benefits

- Enable your SecOps teams to detect potential threats quickly while having the tools to help with security investigations, at a low data retention cost

https://github.com/opensearch-project/security-analytics/issues/7

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What to do next



Learn more

Building security operations with Amazon OpenSearch Service (ANT311)

Dow Jones and 3M: Observability with Amazon OpenSearch Service (ANT219)



Hands-on Workshops and Builders (bring a laptop)

Get started with OpenSearch (OPN203)

Improve search relevance with ML in Amazon OpenSearch Service (ANT307)

Beyond monitoring: Observability with operational analytics data (ANT306)

How to build dashboards for machine-generated data (ANT316)



Ask questions at Chalk Talks

Ingest machine-generated data at scale with Amazon OpenSearch Service with Syte (ANT401)

Observing logs and traces with OpenSearch (OPN401)

Optimizing Amazon OpenSearch Service domains for scale and cost (ANT315)

Why and how to migrate to Amazon OpenSearch Service (ANT314)

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What to do next

Subscribe to the YouTube playlist

Find a partner to help

Sign up for migration assessment and assistance

Read more customer success stories

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.