AWS re:Invent

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

OPN302-R2

Upgrade to OpenSearch

Jon Handler (he/him)
Sr. Principal Solutions Architect
AWS

Andrew Hopp (he/him)
Sr. Product Manager – Technical
AWS

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

The what and why of OpenSearch

Preparing for the upgrade

Types of upgrades

Validation

Best practices

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

The what and why of OpenSearch

Preparing for the upgrade

Types of upgrades

Validation

Best practices

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

What's OpenSearch?

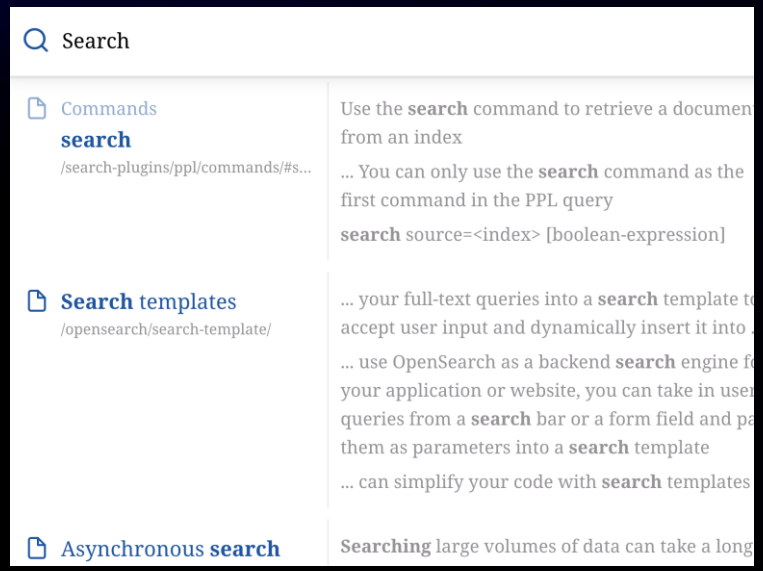


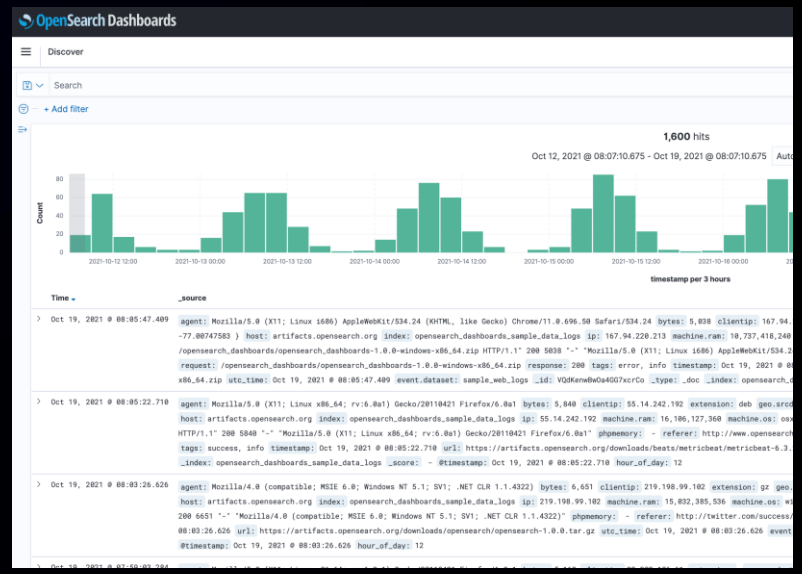
Community-driven, open-source search and analytics suite derived from Apache 2.0 licensed Elasticsearch 7.10.2 & Kibana 7.10.2

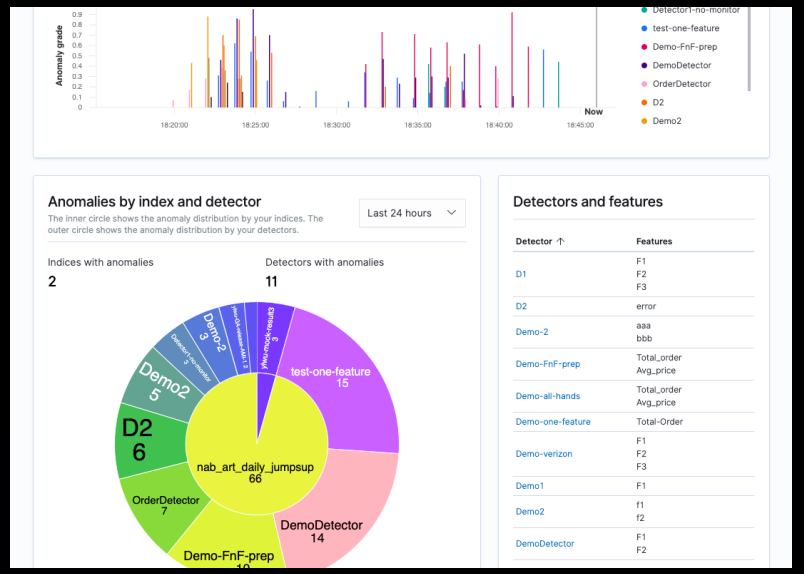
Consists of a search engine, OpenSearch; a visualization user interface, OpenSearch Dashboards; as well as a series functionality-adding tools and plugins

© 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.

Why OpenSearch?

100% open source

- Use, modify, monetize
- Built in the open
- No contributor license agreement

Community-driven

- Made with your input
- Public roadmap
- Welcome contributions

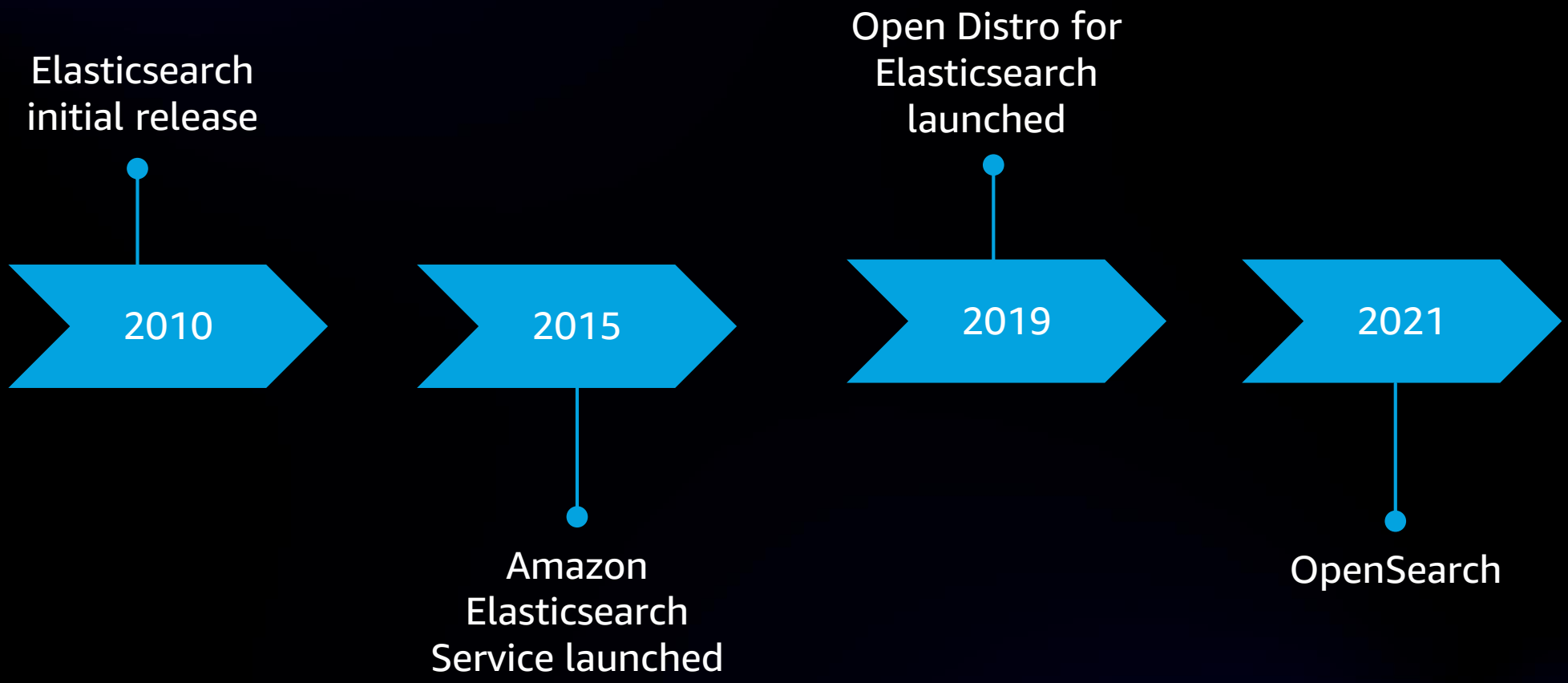
Full-featured

- Full text search
- Analytics
- Fast ingest
- Machine learning
- Security
- and more…

aws

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

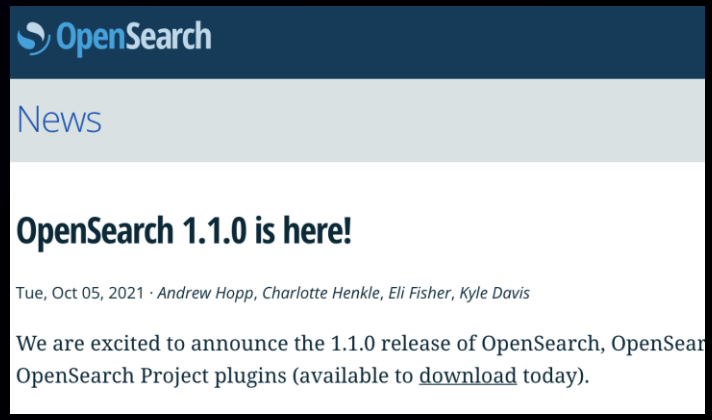
History of OpenSearch at AWS

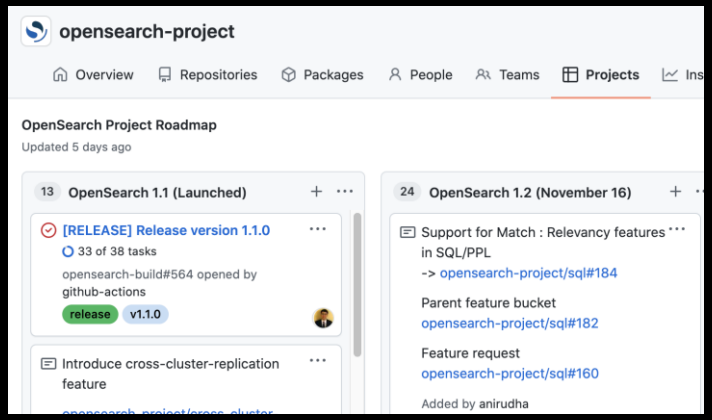


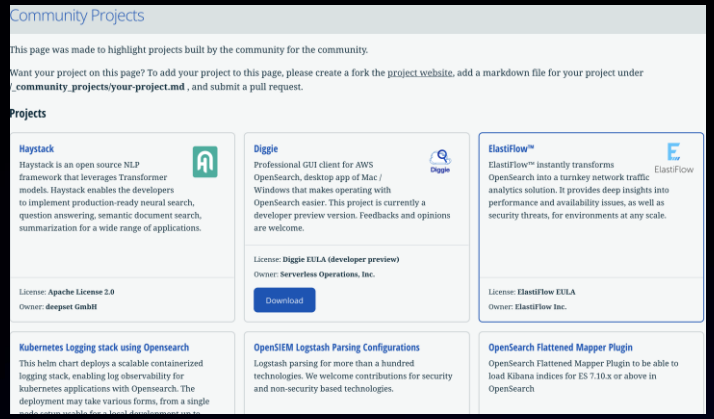
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

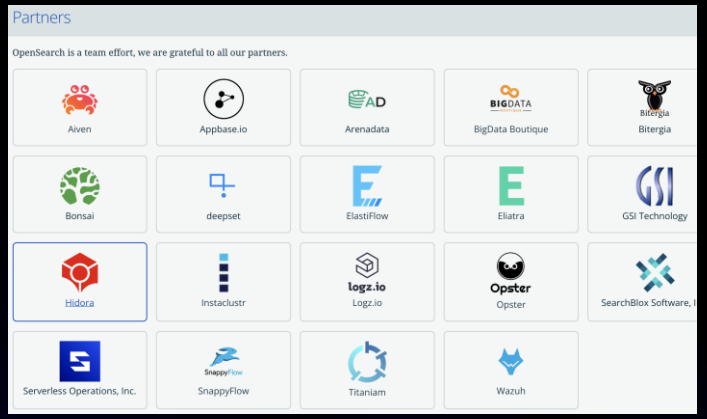
What has happened since the announcement?

- 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

Over-allocated shard prevention

k-NN with FAISS

aws

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

The what and why of OpenSearch

Preparing for the upgrade

Types of upgrades

Validation

Best practices

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

Preparing for the upgrade

1. Take a snapshot

2. Verify the version of your existing cluster

3. Download the latest version of OpenSearch

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

The what and why of OpenSearch

Preparing for the upgrade

Types of upgrades

Validation

Best practices

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

Types of upgrades

We are going to provide instructions on 3 types of upgrades

- Snapshot
- Restart
- Rolling

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

Snapshot upgrade

What?
The first approach to upgrading to OpenSearch is to take a snapshot of your data, create an OpenSearch cluster, restore the snapshot on the new cluster, and point your clients to the new host

Why?
This is typically the safest approach if you can spin up multiple clusters and move your customers to the new endpoint after the upgrade

We recommend this approach for beginners

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

Restart upgrade

What?
The second approach is restarting your whole cluster while upgrading it to OpenSearch; this method incurs downtime, and traffic from clients should be stalled until the new cluster is up

Why?
This is a moderately difficult approach, as it takes down the whole cluster; validation is performed at the very end

We recommend this approach for intermediate users

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

Restart upgrade

1. Verify that the existing cluster is green and healthy
2. Verify your snapshot (just in case!)
3. Disable shard allocation – during upgrade, we do not want any shard movement, as the cluster will be taken down
4. Stop indexing and perform a flush
5. Shut down all the nodes and upgrade all the nodes to OpenSearch
6. Verify that the existing cluster is still green and healthy
7. Start each upgraded node
8. Re-enable shard allocation
9. Verify that the indexed data is now searchable and indexable in OpenSearch

aws

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

Rolling upgrade

What?

This last approach upgrades an existing cluster node by node to OpenSearch while keeping the cluster active during the upgrade

Why?

This is the most complex approach, but requires no downtime and provides cluster health feedback at every step of the process; this method takes down one node at a time for an upgrade

We recommend this approach for advanced users

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

Rolling upgrade

1. Verify that the existing cluster is green and healthy
2. Disable shard allocation
3. Stop indexing and perform a flush
4. Shut down a single node
5. Upgrade the node that was shut down
6. Verify that the existing cluster is still green and healthy
7. Wait for the node to recover
8. Re-enable shard allocation
9. Verify that the indexed data is now searchable and indexable in OpenSearch
10. Repeat steps 4–7 until all the nodes are upgraded (data nodes first)

aws

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

The what and why of OpenSearch

Preparing for the upgrade

Types of upgrades

Validation

Best practices

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

Validation

Once you've completed the upgrade, it's always good practice to validate the new version

- Check cluster health
- Check indices
- Check settings

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

Validation

You can verify via these APIs, which are just a few of many different ways

```
# get cluster health
GET _cluster/health

# get nodes in a cluster
GET _cat/nodes

# get node health
GET _cat/health?v=true
```

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

The what and why of OpenSearch

Preparing for the upgrade

Types of upgrades

Validation

Best practices

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

Best practices

Testing in an isolated environment is always a good idea

Back up, back up, back up

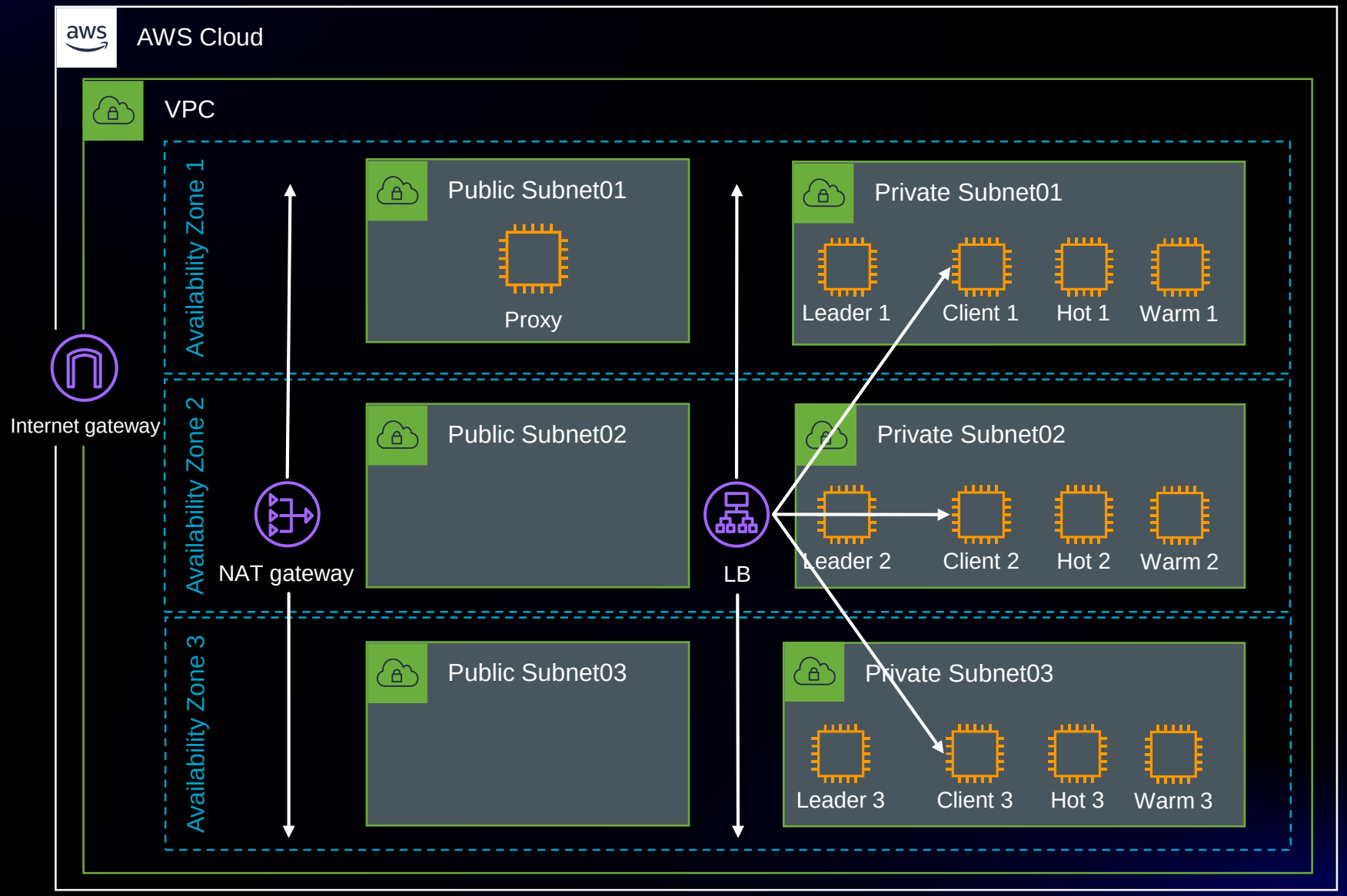
Use the OpenSearch upgrade tool

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

Deployment and sizing

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

Deployment architecture (at scale)



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

Capacity planning

Gather information

- Data per day
- Number of indexes
- Concurrency
- RPS (Docs, Qs)
- Index/query complexity
- Durability need

Plan for storage

- Ephemeral/ Amazon EBS
- Replication
- Retention
- UltraWarm
- Cold storage

Storage is a solid starting point

Plan for compute

- Concurrency – number of readers/writers
- Active indexes/shards
- Total request counts and latencies

Active shards < CPUs

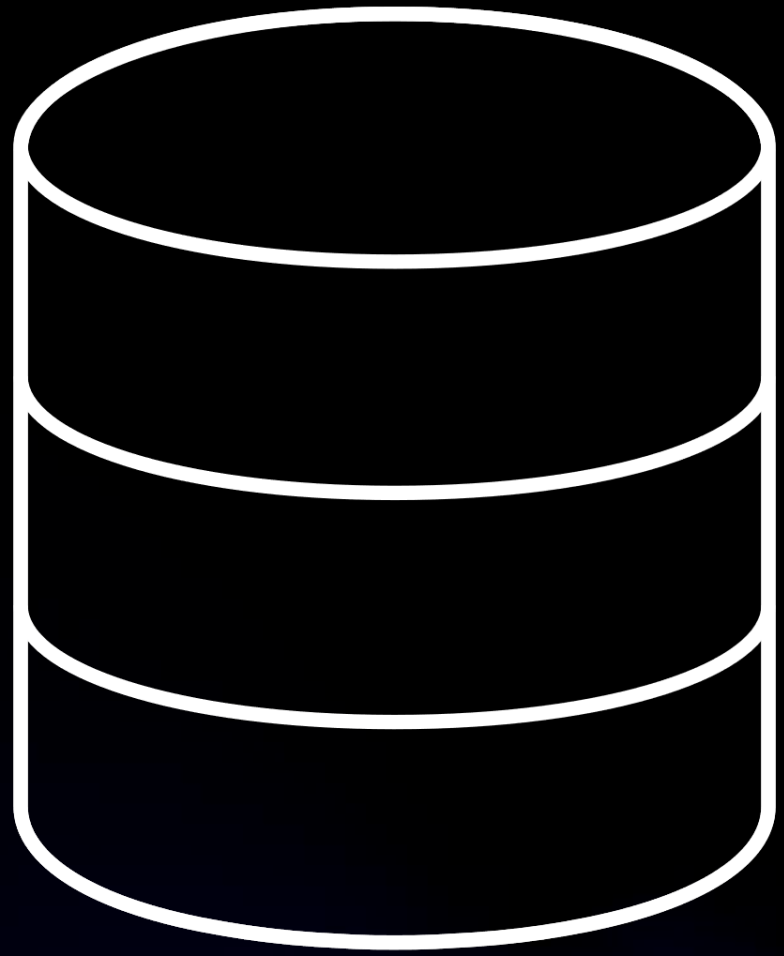
aws

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

Storage needed

Storage needed = source * 1.1 * 2 * 7 * 1.15

- On disk, an index is 1.1 x source
- Each replica requires that much additional storage
- For streaming data, multiply by retention (ignore for single indexes)
- Add 15% for overhead
- Deploy instances based on storage supplied



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

Scale CPUs based on usage

Write heavy – Active CPUs = total number of primaries and replicas for all indexes receiving write traffic

Read heavy – Active CPUs = total number of primary shards for all indexes receiving read traffic

CPUs = 1.5 * active shards

Example: Weblogs and syslogs; 1 TB per day ingest; use 50 GB shard size, yields ~50 shards (p and r); use 75 CPUs as a starting scale point

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

Features

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

Security

Encryption

Keep your data secure when in transit

Authentication

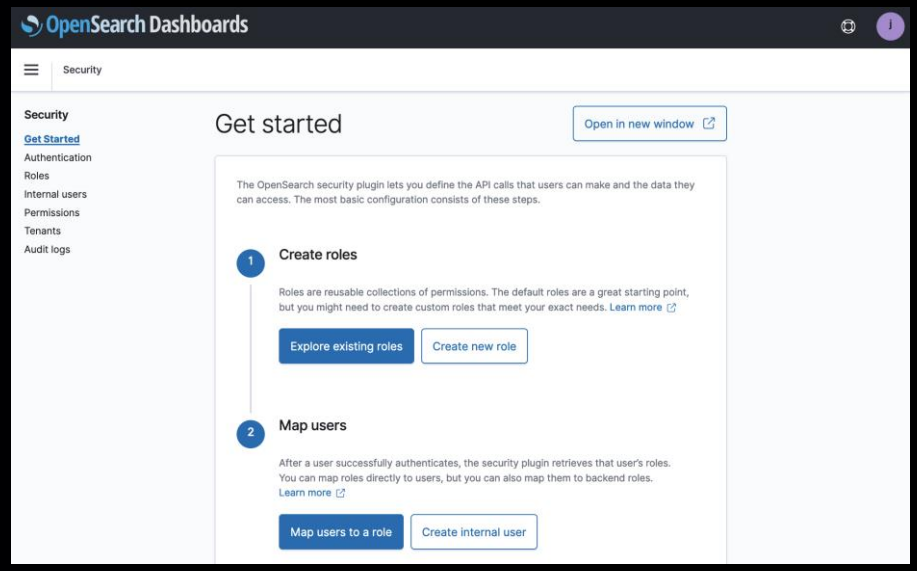
Leverage your existing authentication infrastructure

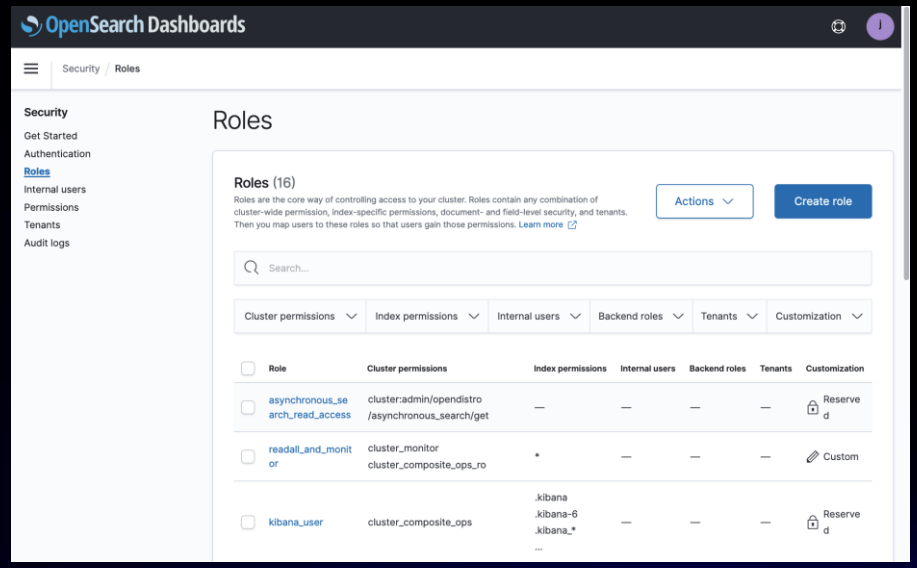
RBAC

Granular access control over user actions on your cluster

Audit logging

Track and record all user actions and support HIPAA and PCI compliance

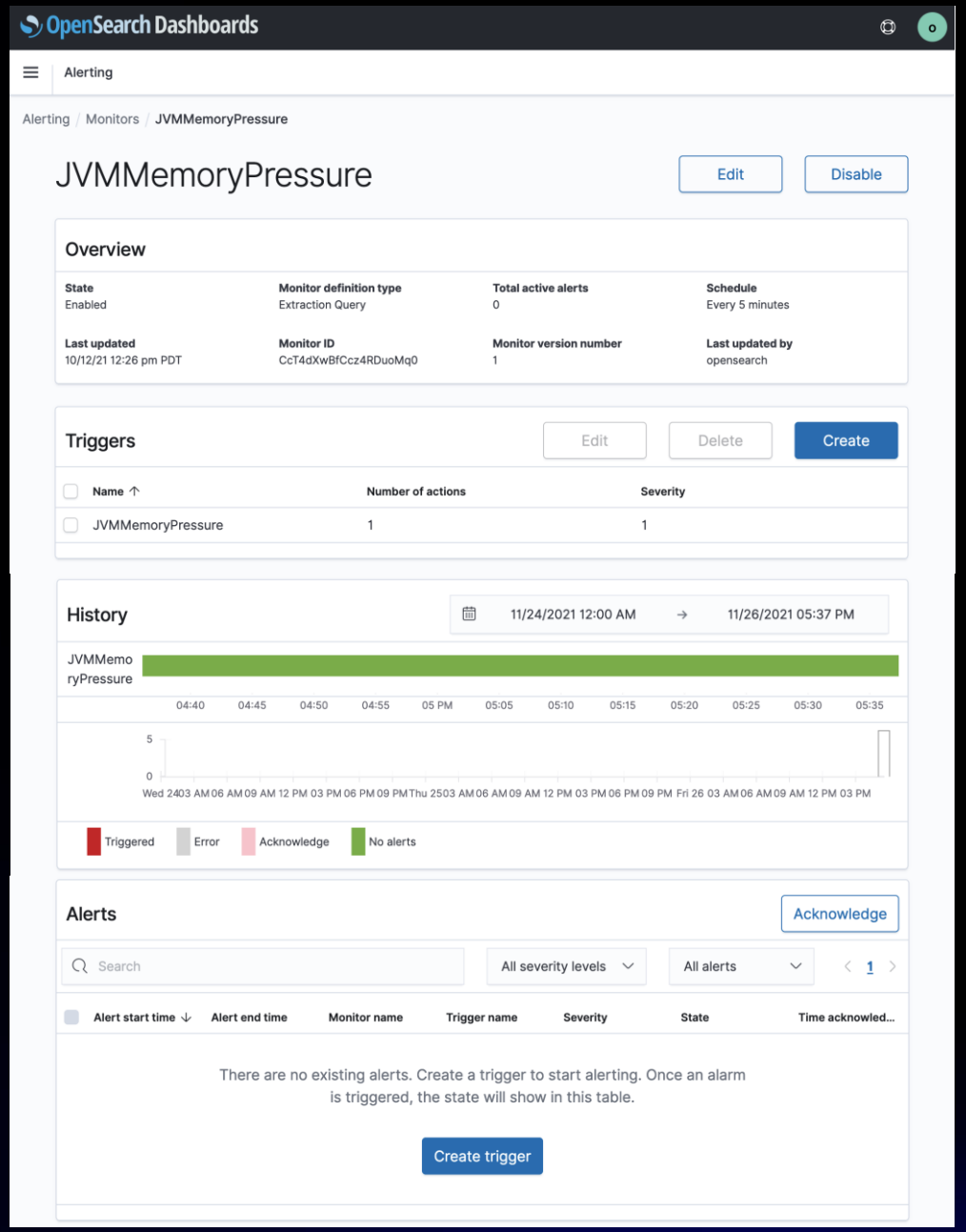




© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Alerting

- Create monitors for your data
- Customize alert conditions
- Define alerting threshold and severity for multiple trigger conditions
- View alerts; get notifications
- Built-in integrations for webhook and Slack to get notified on the channels you use
- All alert executions are indexed for easy tracking and visualization



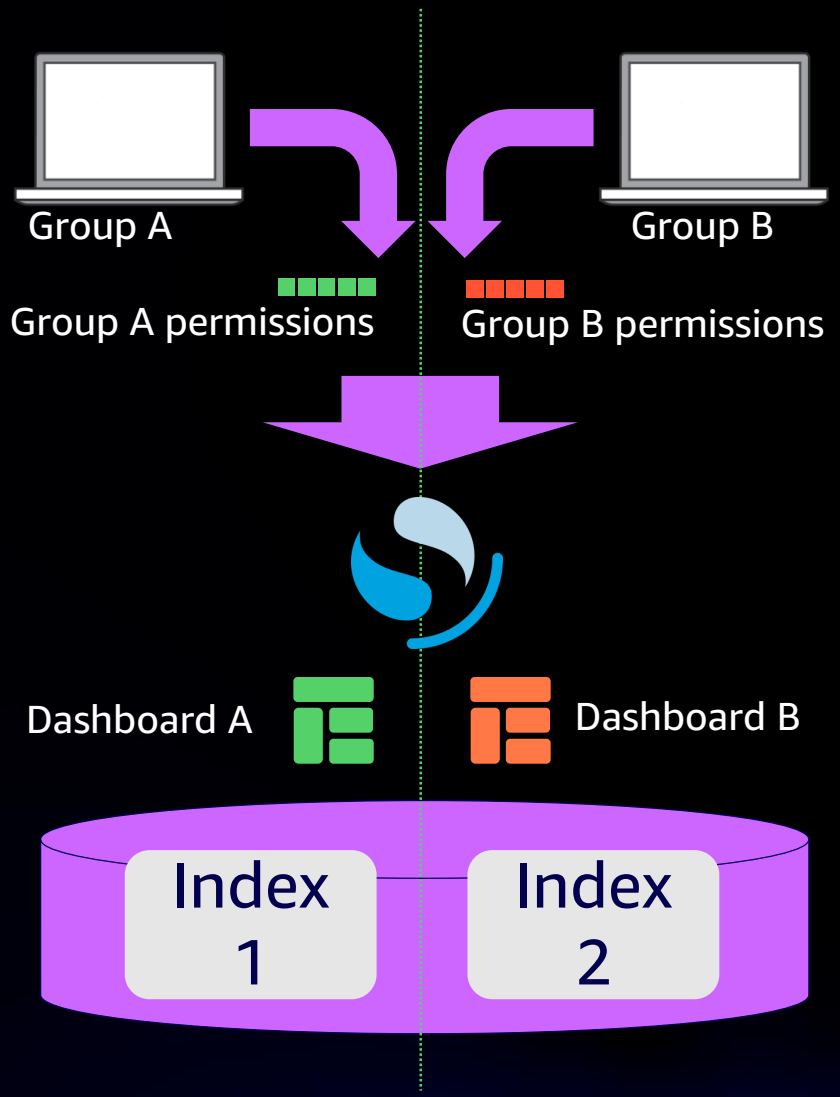
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Kibana multi-tenancy

Users select a tenant, and all work is local to that tenant

There is a pre-configured global tenant that everyone can use

Access control is via OpenSearch Security: tenants must have access to the underlying indices to view the data



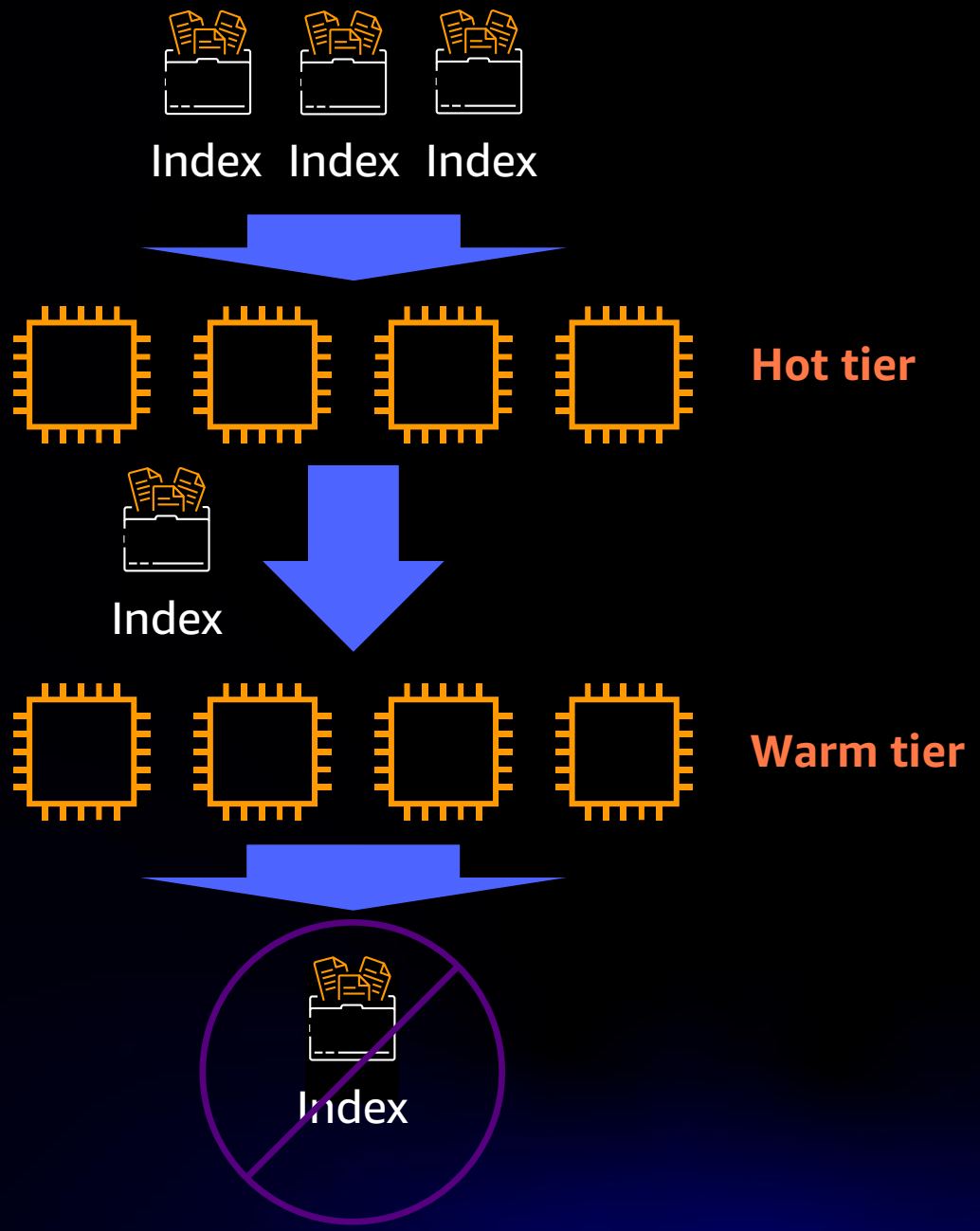
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Index management

Manage data lifecycle by policy

Roll up jobs to reduce overall storage

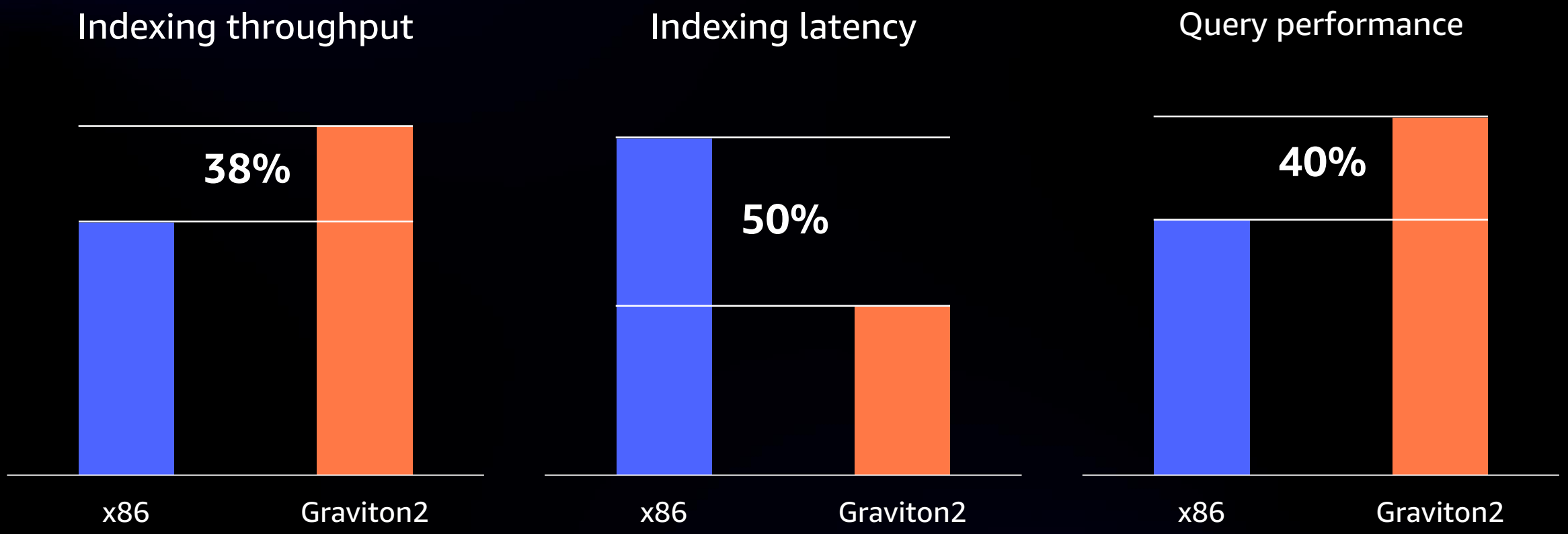
Transforms



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

AWS Graviton2 for OpenSearch

CUSTOM-BUILT SILICON WITH NEXT-GENERATION PERFORMANCE IMPROVEMENT



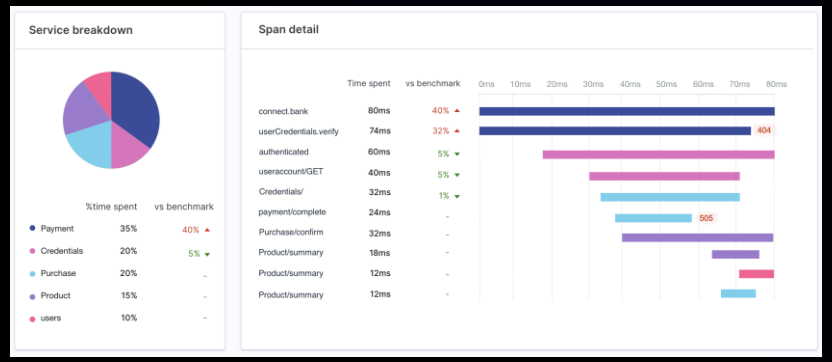
44% better price to performance compared to fifth-generation instances

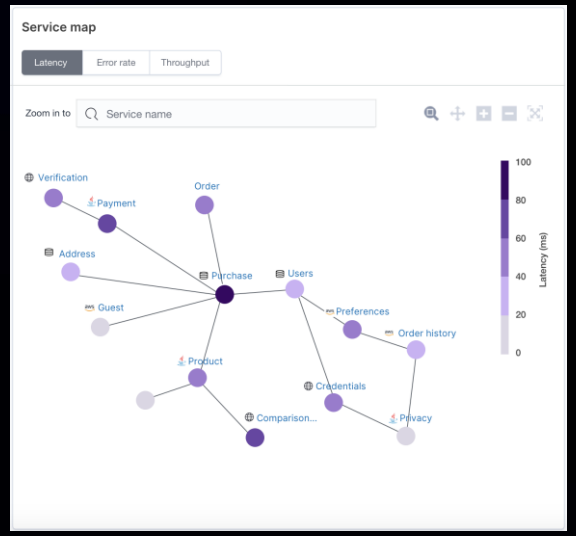
aws

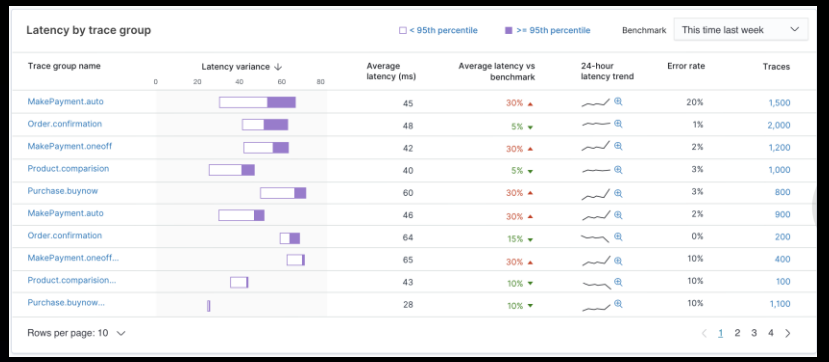
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is Trace Analytics?

SUPPORTS OPENTELEMETRY (JAEGER, ZIPKIN, AWS X-RAY SDKS)







Trace-Span Details

Single request performance

Diagnose root cause

Service Maps

End-to-end view

Isolate issues to services

Trace Groups

Monitor performance

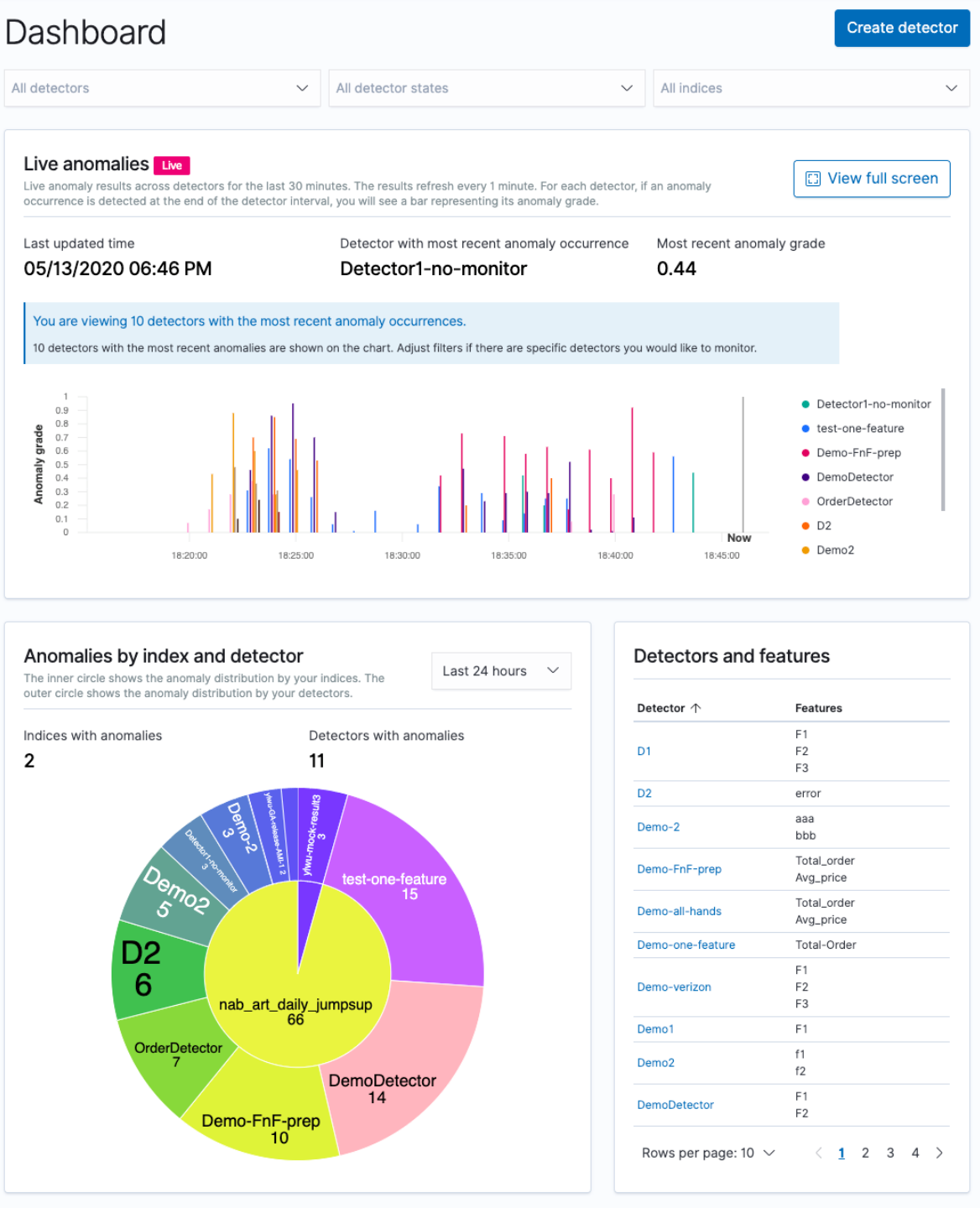
Identify issues early

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

Streaming anomaly detection

Performant at scale; machine learning models are distributed and processed across nodes

Easy to use; machine learning expertise is not required



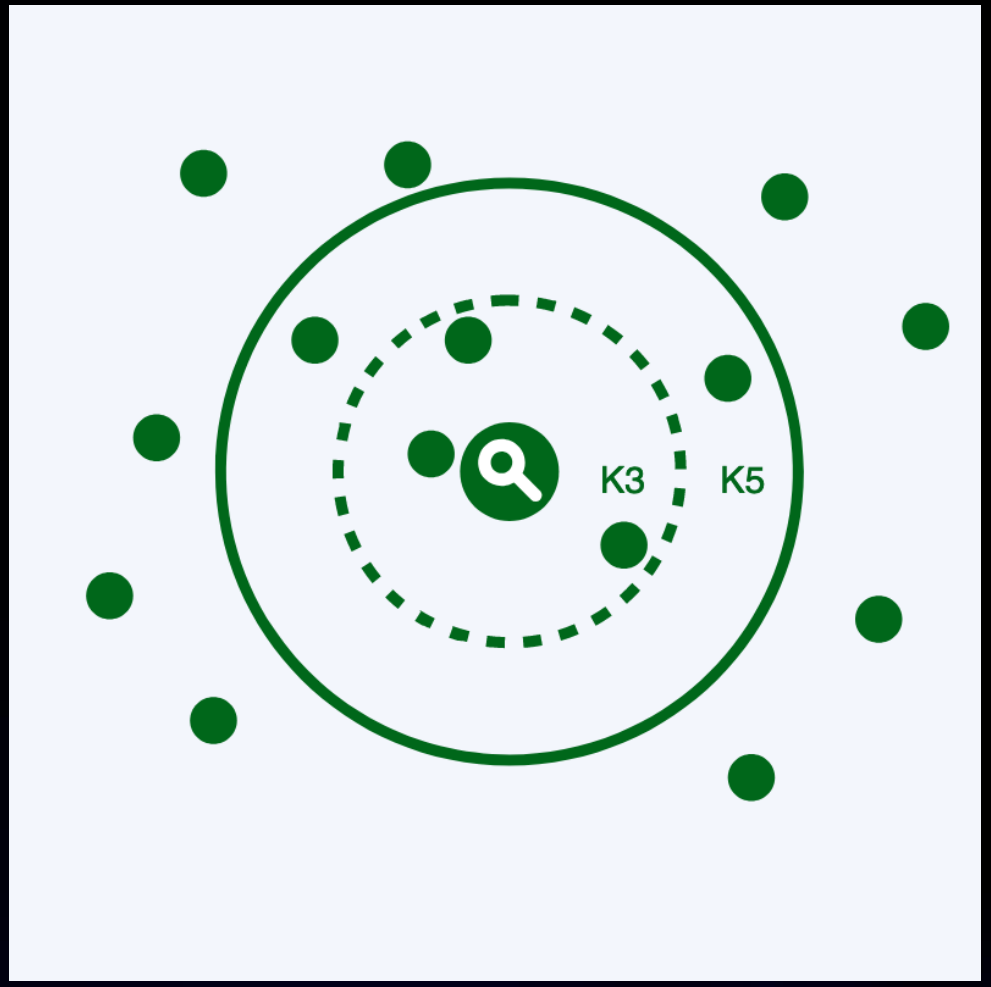
K-NN search

Vector-based search algorithm

Use feature extraction or other to create feature vectors

Query with a vector to retrieve the geometrically closest points

Geometric or cosine similarity

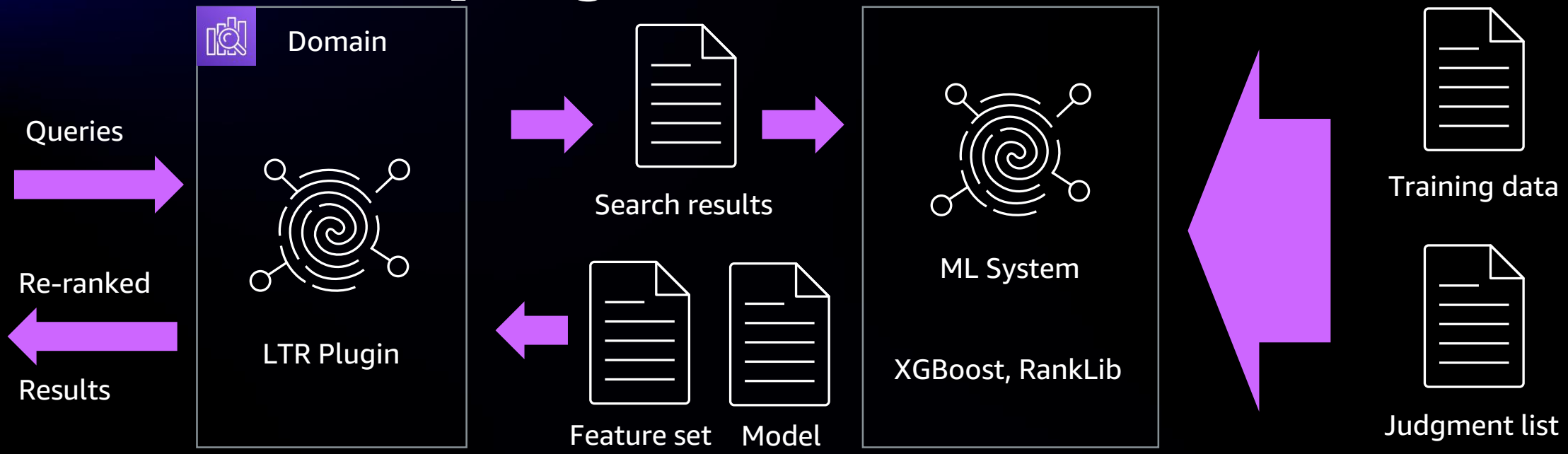


Provides a means of mixing machine-learned demographics, preference clusters, etc. to matching

aws

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

Learn to rank plugin



Machine-learned models dial in the ranking of your search results

Use features like recency, popularity ratings, and preferences

Feed the model to the LTR plugin in Elasticsearch

The LTR plugin applies the model, re-ranking search results

Use in a loop, feeding back search results and click data to improve the model

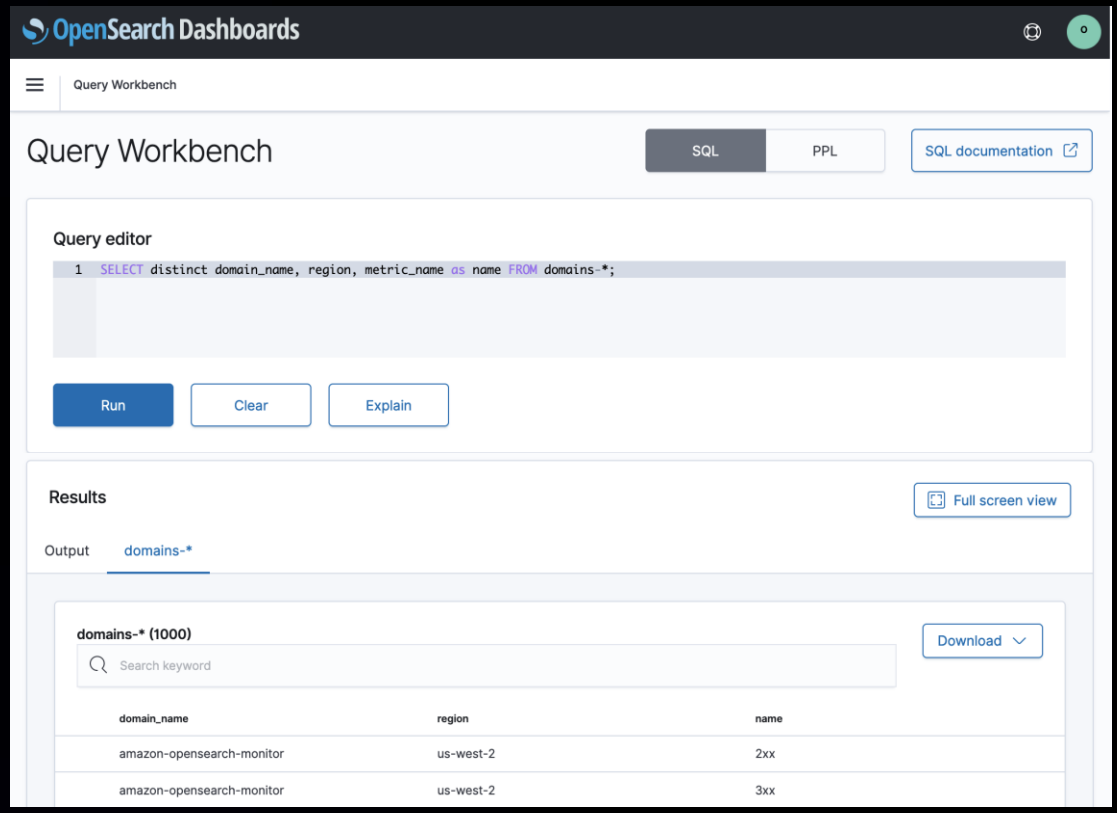
aws

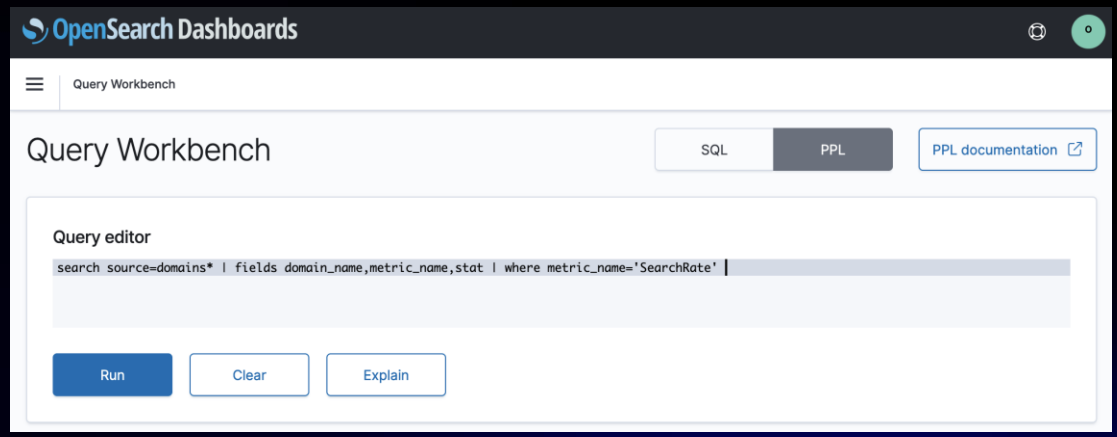
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Query workbench

Use SQL queries – supports ~40 SQL primitives

Piped Processing Language (PPL) – flexible language for applying commands sequentially





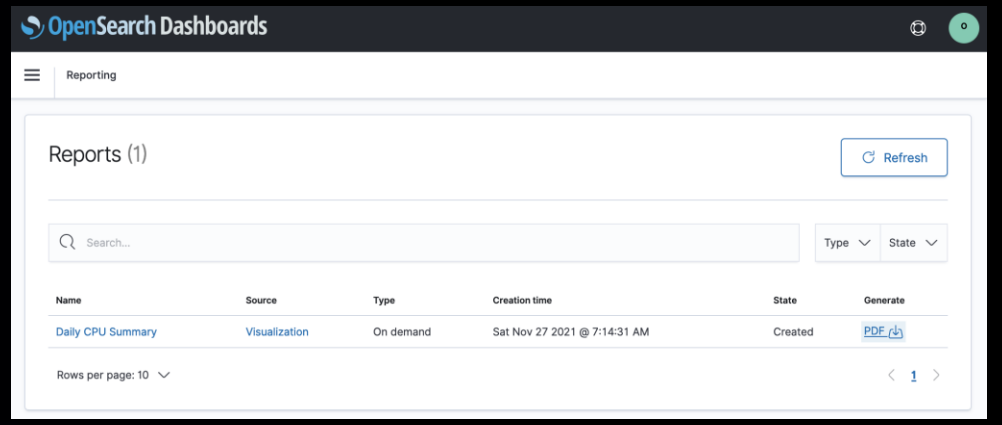
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

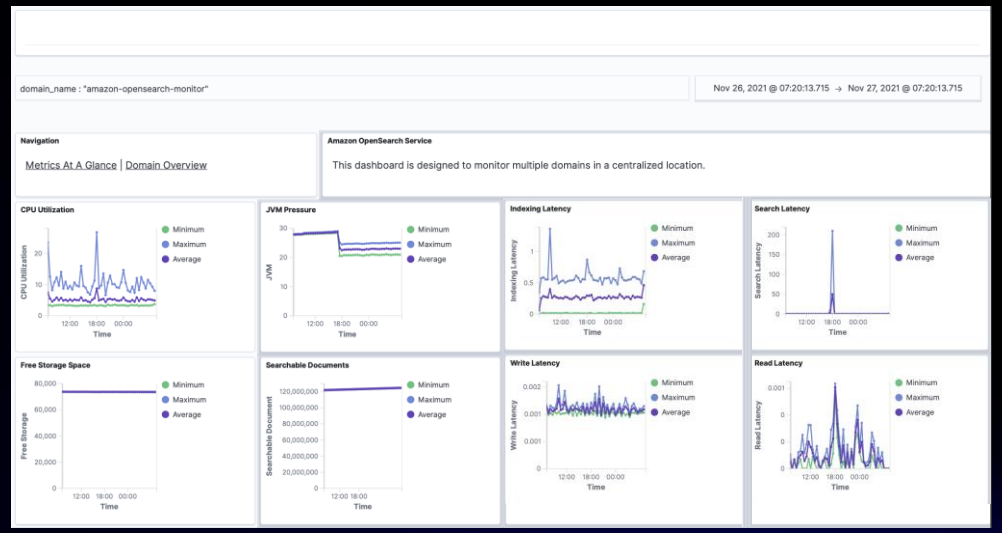
Reporting

Generate reports on demand or on a schedule

Source from a dashboard, visualization, Notebook, or saved search

Download as a PDF or PNG





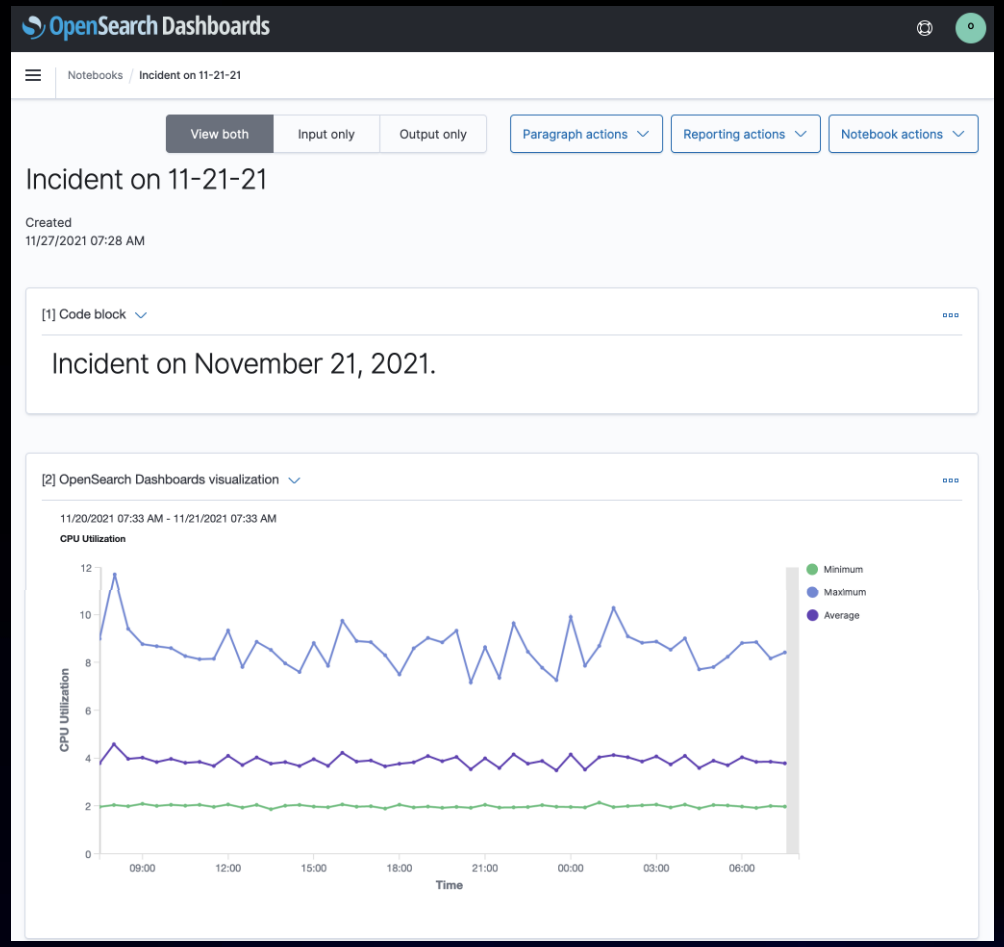
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Notebooks

Build out detailed narratives around past events.

Visualizations are set in time; you can update as well

Run arbitrary markdown, SQL, and PPL



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

Thank you!

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