Amazon OpenSearch Service

Arun Lakshmanan
Solutions Architect

Agenda

Challenges customers face

What is Amazon OpenSearch Service

Why Amazon OpenSearch Service

Leading use cases

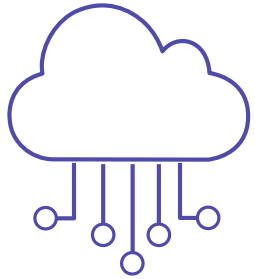
New capabilities

Getting started

Challenges customers face

Machine-generated data is exploding

Applications and infrastructure



Services/micro-services
Web Applications
Business Applications
APIs

IT and DevOps

Databases
Load balancers
Networking
Servers

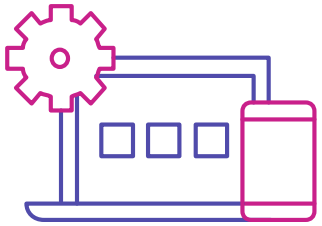
IoT and wireless

Automotive
Home devices
Manufacturing
Mobile Applications

```
199.72.81.55 - - [01/Jul/1995:00:00:01 -0400] "GET /history/apollo/ HTTP/1.0" 200 6245
unicomp6.unicomp.net - - [01/Jul/1995:00:00:06 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
199.120.110.21 - - [01/Jul/1995:00:00:09 -0400] "GET /shuttle/missions/sts-73/mission-sts-73.html HTTP/1.0" 200 4085
burger.letters.com - - [01/Jul/1995:00:00:11 -0400] "GET /shuttle/countdown/liftoff.html HTTP/1.0" 304 0
199.120.110.21 - - [01/Jul/1995:00:00:11 -0400] "GET /shuttle/missions/sts-73/sts-73-patch-small.gif HTTP/1.0" 200 4179
burger.letters.com - - [01/Jul/1995:00:00:12 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 304 0
burger.letters.com - - [01/Jul/1995:00:00:12 -0400] "GET /shuttle/countdown/video/livevideo.gif HTTP/1.0" 200 0
205.212.115.106 - - [01/Jul/1995:00:00:12 -0400] "GET /shuttle/countdown/countdown.html HTTP/1.0" 200 3985
d104.aa.net - - [01/Jul/1995:00:00:13 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
129.94.144.152 - - [01/Jul/1995:00:00:13 -0400] "GET / HTTP/1.0" 200 7074
unicomp6.unicomp.net - - [01/Jul/1995:00:00:14 -0400] "GET /shuttle/countdown/count.gif HTTP/1.0" 200 40310
unicomp6.unicomp.net - - [01/Jul/1995:00:00:14 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 200 786
unicomp6.unicomp.net - - [01/Jul/1995:00:00:14 -0400] "GET /images/KSC-logosmall.gif HTTP/1.0" 200 1204
d104.aa.net - - [01/Jul/1995:00:00:15 -0400] "GET /shuttle/countdown/count.gif HTTP/1.0" 200 40310
d104.aa.net - - [01/Jul/1995:00:00:15 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 200 786
d104.aa.net - - [01/Jul/1995:00:00:15 -0400] "GET /images/KSC-logosmall.gif HTTP/1.0" 200 1204
129.94.144.152 - - [01/Jul/1995:00:00:17 -0400] "GET /images/ksclogo-medium.gif HTTP/1.0" 304 0
199.120.110.21 - - [01/Jul/1995:00:00:17 -0400] "GET /images/launch-logo.gif HTTP/1.0" 200 1713
ppptky391.asahi-net.or.jp - - [01/Jul/1995:00:00:18 -0400] "GET /facts/about_ksc.html HTTP/1.0" 200 3977
net-1-141.eden.com - - [01/Jul/1995:00:00:19 -0400] "GET /shuttle/missions/sts-71/images/KSC-95EC-0916.jpg HTTP/1.0" 200 34029
ppptky391.asahi-net.or.jp - - [01/Jul/1995:00:00:19 -0400] "GET /images/launchpalms-small.gif HTTP/1.0" 200 11473
205.189.154.54 - - [01/Jul/1995:00:00:24 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
waters-gw.starway.net.au - - [01/Jul/1995:00:00:25 -0400] "GET /shuttle/missions/51-l/mission-51-l.html HTTP/1.0" 200 6723
ppp-mia-30.shadow.net - - [01/Jul/1995:00:00:27 -0400] "GET / HTTP/1.0" 200 7074
205.189.154.54 - - [01/Jul/1995:00:00:29 -0400] "GET /shuttle/countdown/count.gif HTTP/1.0" 200 40310
alyssa.prodigy.com - - [01/Jul/1995:00:00:33 -0400] "GET /shuttle/missions/sts-71/sts-71-patch-small.gif HTTP/1.0" 200 12054
ppp-mia-30.shadow.net - - [01/Jul/1995:00:00:35 -0400] "GET /images/ksclogo-medium.gif HTTP/1.0" 200 5866
dial22.lloyd.com - - [01/Jul/1995:00:00:37 -0400] "GET /shuttle/missions/sts-71/images/KSC-95EC-0613.jpg HTTP/1.0" 200 61716
smyth-pc.moorecap.com - - [01/Jul/1995:00:00:38 -0400] "GET /history/apollo/apollo-13/images/70HC314.GIF HTTP/1.0" 200 101267
205.189.154.54 - - [01/Jul/1995:00:00:40 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 200 786
ix-orl2-01.ix.netcom.com - - [01/Jul/1995:00:00:41 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
```

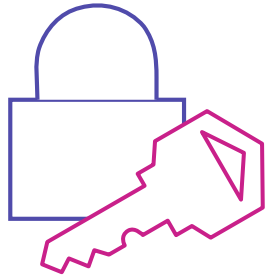
Machine-generated data is exploding

Applications



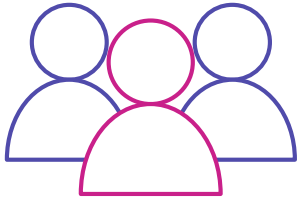
Is my infrastructure working?
What is the latency and error rate?
What caused my application issue?

Security



Is there any suspicious authentication activity?
What data was accessed by this IP address?
Are there instances of fraud?

Business insights



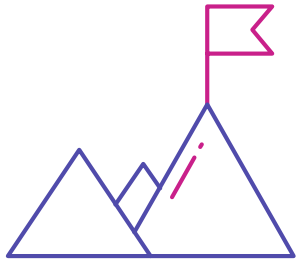
What content/products are my users interested in?

Which features are used most or least?

What users are most active and why?

There’s a need for real-time search at scale

Customers want to **find the right product, service, document, or answer to their problem as quickly** as possible. Their searches will be across both semi-structured and unstructured data, and across different facets and attributes.



In today’s world, **search results have to be relevant and delivered in real time. Building and maintaining** a system that achieves this **cost-effectively, securely, and at scale is challenging.**

Where is there a need for real-time search at scale?

If you have an **e-commerce platform,** you want customers to find the product they are looking for quickly.

If you offer a **document portal** with documents including, but not limited to, scientific research articles, investment analyses, or health records, you want to enable a speedy and relevant search experience for your users

You may want to increase user engagement on your platform by **delivering personalized recommendations,** like a weekly music playlist or food recipes.

Beyond these examples, you may have other parts of your tech stack where you want to add an **easy to use and snappy search experience,** especially with the option to integrate machine learning capabilities to power a personalized experience.

Amazon OpenSearch Service

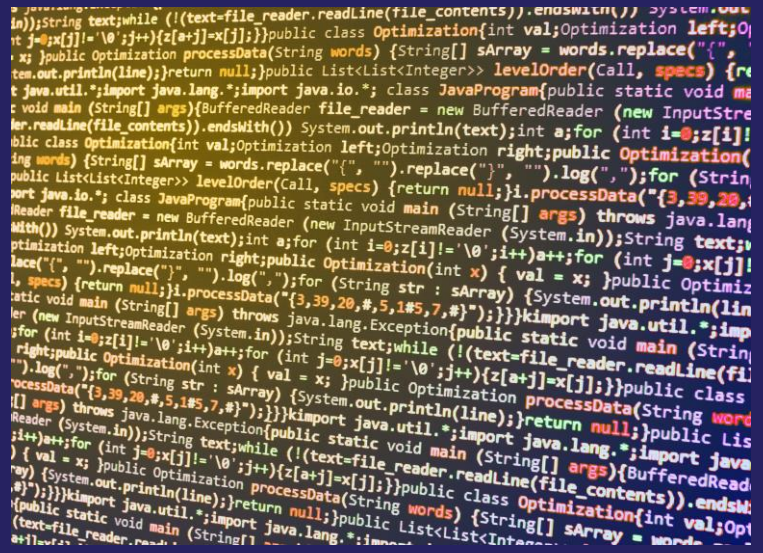
Amazon OpenSearch Service

Amazon OpenSearch Service makes it easy for you to perform interactive log analytics, real-time application monitoring, website search, and more. OpenSearch is an open source, distributed search and analytics suite derived from Elasticsearch. Amazon OpenSearch Service offers the latest versions of OpenSearch, support for 19 versions of Elasticsearch (1.5 to 7.10 versions), and visualization capabilities powered by OpenSearch Dashboards and Kibana (1.5 to 7.10 versions).

OpenSearch

- OpenSearch is a **community-driven,** open source search and analytics suite derived from Apache 2.0 licensed Elasticsearch 7.10.2
- The OpenSearch project consists of a distributed search engine powered by Apache Lucene, *OpenSearch,* and a data visualization and user interface, *OpenSearch Dashboards*
- OpenSearch also includes all of the advanced functionality ported over from Open Distro for Elasticsearch

OpenSearch is a powerful analytics engine



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

How does it work?

1

Send data as JSON via REST APIs

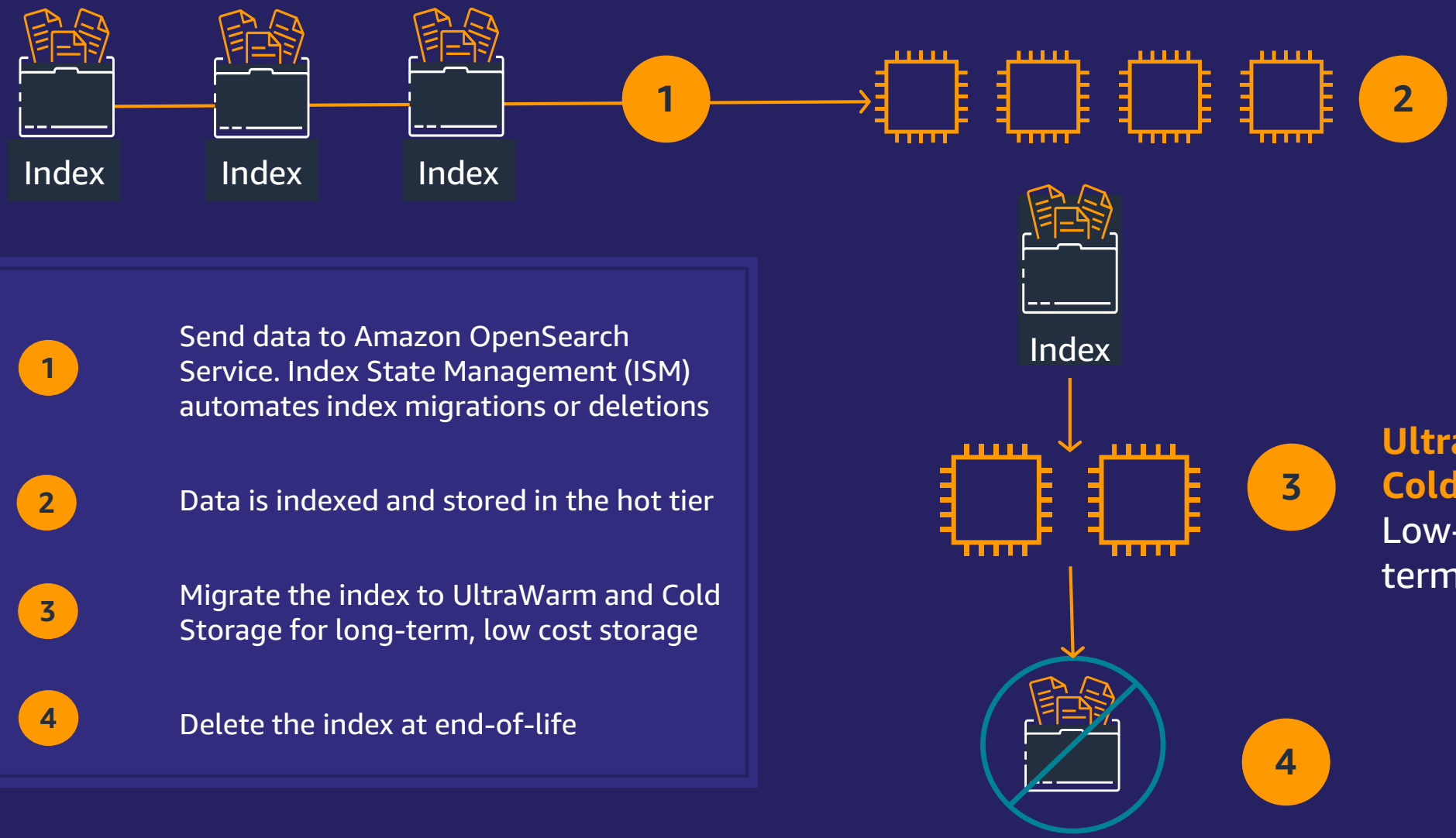
2

Data is indexed - all fields searchable, including nested JSON

3

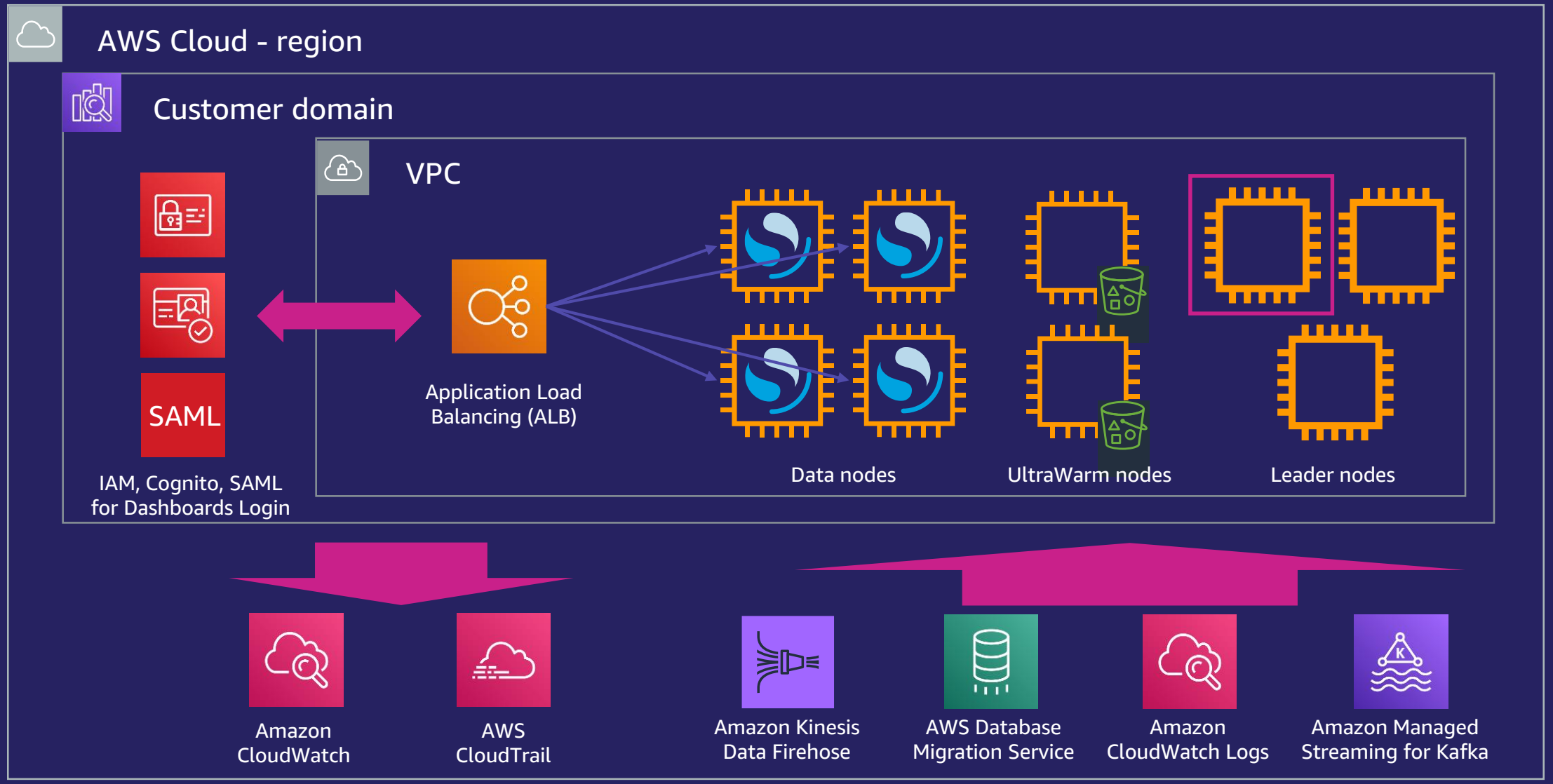
REST APIs, for fielded matching, Boolean expressions, sorting, and analysis

Data lifecycle in Amazon OpenSearch Service

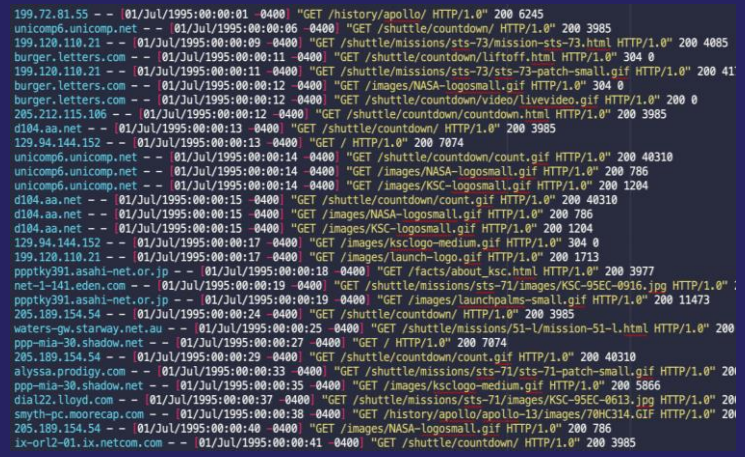


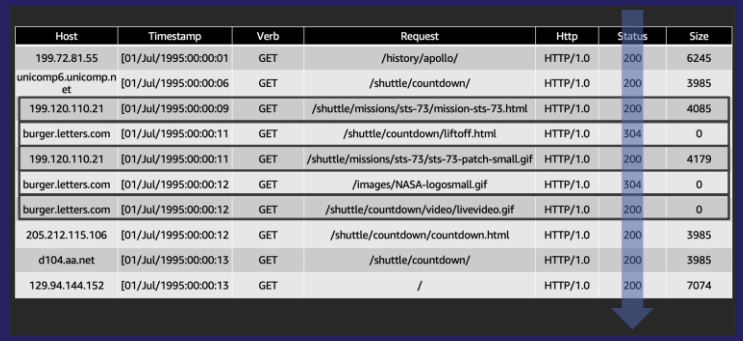
1. Send data to Amazon OpenSearch Service. Index State Management (ISM) automates index migrations or deletions
2. Data is indexed and stored in the hot tier
3. Migrate the index to UltraWarm and Cold Storage for long-term, low cost storage
4. Delete the index at end-of-life

Amazon ES Deployment Architecture

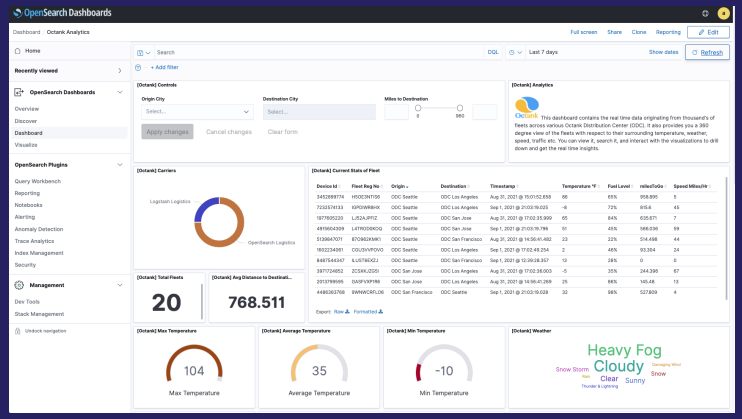


Getting insights from your data



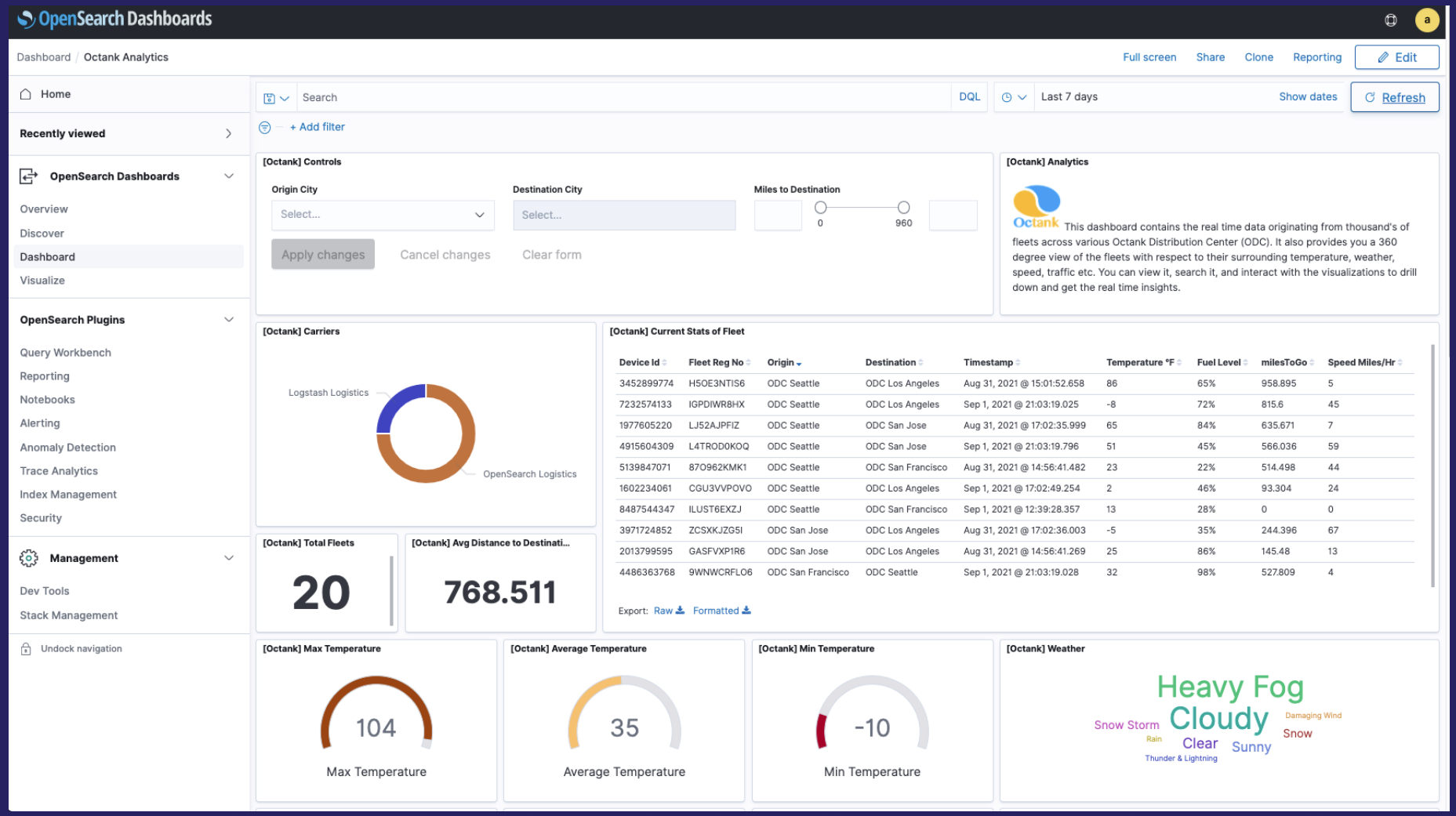


Host	Timestamp	Verb	Request	Http	Status	Size
199.72.81.55	[01/Jul/1995:00:00:01	GET	/history/apollo/	HTTP/1.0	200	6245
unicomp6.unicomp.net	[01/Jul/1995:00:00:06	GET	/shuttle/countdown/	HTTP/1.0	200	3985
199.120.110.21	[01/Jul/1995:00:00:09	GET	/shuttle/missions/sts-73/mission-sts-73.html	HTTP/1.0	200	4085
burger.letters.com	[01/Jul/1995:00:00:11	GET	/shuttle/countdown/liftoff.html	HTTP/1.0	304	0
199.120.110.21	[01/Jul/1995:00:00:11	GET	/shuttle/missions/sts-73/sts-73-patch-small.gif	HTTP/1.0	200	4179
burger.letters.com	[01/Jul/1995:00:00:12	GET	/images/NASA-logosmall.gif	HTTP/1.0	304	0
burger.letters.com	[01/Jul/1995:00:00:12	GET	/shuttle/countdown/video/livevideo.gif	HTTP/1.0	200	0
205.212.115.106	[01/Jul/1995:00:00:12	GET	/shuttle/countdown/countdown.html	HTTP/1.0	200	3985
d104.aa.net	[01/Jul/1995:00:00:13	GET	/shuttle/countdown/	HTTP/1.0	200	3985
129.94.144.152	[01/Jul/1995:00:00:13	GET	/	HTTP/1.0	200	7074



OpenSearch delivers near real-time insights

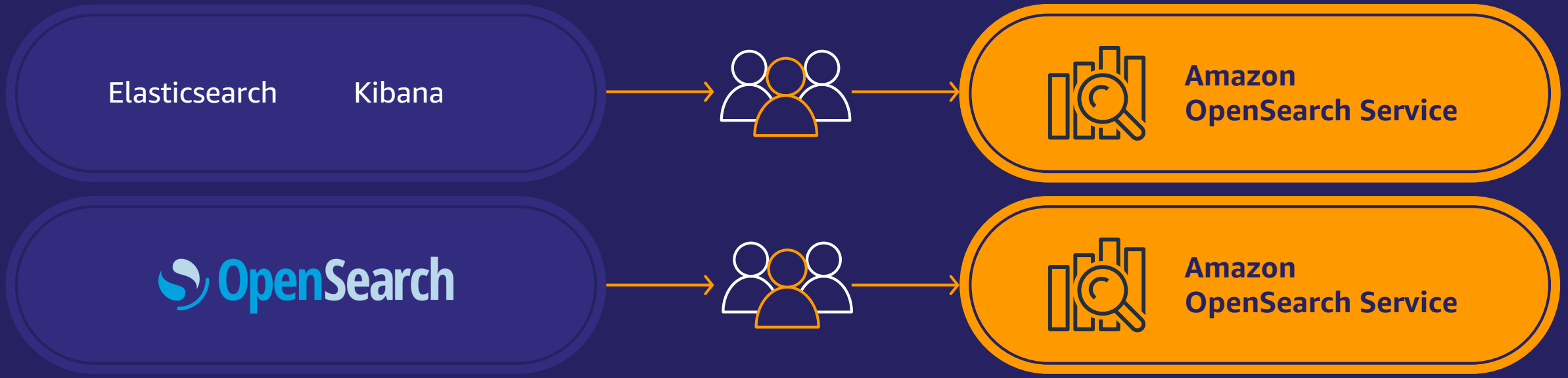
OpenSearch Dashboards is a lightweight, real-time visualization tool



Why Amazon OpenSearch Service

Moving from self-managing open-source solutions to Amazon OpenSearch Service

- Managing and scaling requires dedicated expertise, driving up the total cost of ownership
- Customers need to build or pay for advanced security, alerting, and other features
- Customers need to purchase and manage their own infrastructure



Self-managed vs. Amazon OpenSearch Service

Self-Managed		Managed Service
On-Premises	**Amazon EC2**	**Amazon OpenSearch Service**
App dev/optimization	App dev/optimization	App dev/optimization
Hot/warm storage tiers	Hot/warm storage tiers	Hot/warm storage tiers
Plugins (additional cost)*	Plugins (additional cost)*	Plugins
24x7 monitoring & repair	24x7 monitoring & repair	24x7 monitoring & repair
In-place upgrades/patches	In-place upgrades/patches	In-place upgrades/patches
Cluster scaling	Cluster scaling	Cluster scaling
Cross-AZ data transfer cost	Cross-AZ data transfer cost	Cross-AZ data transfer cost
Backups	Backups	Backups
High availability	High availability	High availability
Security (FGAC, Auth)	Security (FGAC, Auth)	Security (FGAC, Auth)
Hardware & OS maintenance	Hardware & OS maintenance	Hardware & OS maintenance
Hardware lifecycle	Hardware lifecycle	Hardware lifecycle
Power/network/HVAC	Power/network/HVAC	Power/network/HVAC

* SQL querying, Real-time Alerting, Index State Management, Anomaly Detection, Machine Learning

Moving from licensed solutions to Amazon OpenSearch Service

- Other, more packaged solutions can drive excessive cost as data volumes grow
- Database solutions and some packaged solutions have lower limits on capacity and higher latency
- Amazon OpenSearch Service is a very flexible tool, supporting search—for application data, but also for logging data. This enables many customers to use Amazon OpenSearch Service for issue debugging and repair

Benefits of Amazon OpenSearch Service

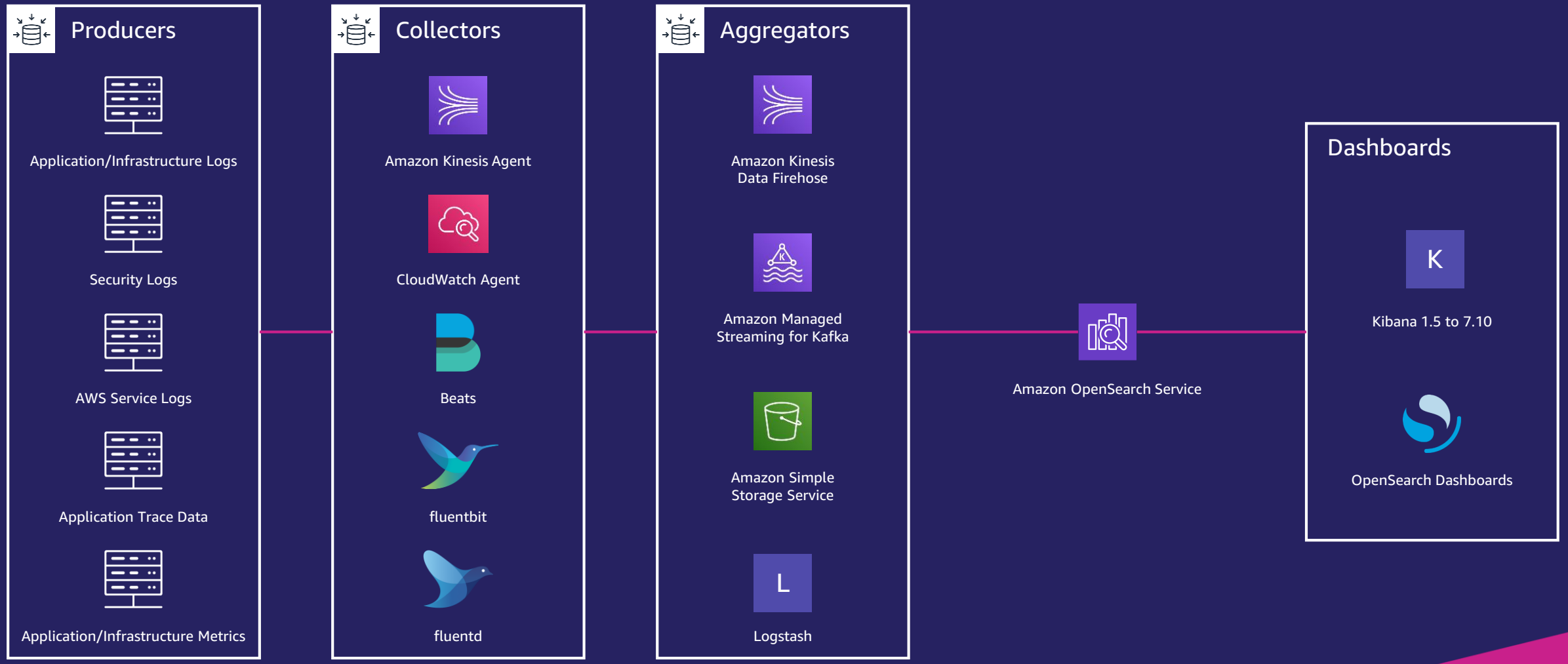
Operationalize OpenSearch with the leading contributor of the community-driven, open-source software

Quickly search and analyze your unstructured and semi-structured data to easily find what you need

Use machine learning to detect anomalies in real time, autotune your clusters, and personalize your search results

Eliminate operational overhead and reduce cost with automated provisioning, software installation, patching, storage tiering, and more

Amazon OpenSearch Service data ingestion flow



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

mlbam · NETFLIX · FOX

Social Media

cookpad · ancestry · Nextdoor

Telecommunications

Jio · 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

Use cases

Use case #1

Observability

Pinterest case study

Why observability?

You need to correlate logs, metrics, and traces to gain insights into application health and performance and resolve issues across the business.

How Amazon OpenSearch Service can help

Centralizes log analytics to identify or predict performance problems across your business. With cross-cluster search, you can analyze and query all of your log data via a single OpenSearch Dashboards interface.

Use case #2

Application & infrastructure monitoring

Autodesk case study

Why application and infrastructure monitoring?

You need to proactively monitor your applications and infrastructure log data to find performance issues faster and improve operational health.

How Amazon OpenSearch Service can help

Provides real-time search and log analytics capabilities to identify or predict performance problems and enable your teams to do real-time root cause and forensic analysis, therefore reducing Mean Time to Detect (MTTD) and Mean Time to Resolve (MTTR) issues.

Use case #3
Search

COMPASS

Compass blog

Why search?

You need a fast search experience for your applications, websites, and data lake catalogs, allowing your users to quickly find relevant data.

How Amazon OpenSearch Service can help

Delivers high-quality and personalized search results to customers. You get access to all of Elasticsearch's search APIs, supporting natural language search, auto-completion, faceted search, adjustable ranking, and location-aware search.

Use case #4

Security monitoring

Pearson case study

Why security monitoring?

You need to keep your data safe, preventing security threats such as data breaches, unauthorized login attempts, DoS attacks, and fraud.

How Amazon OpenSearch Service can help

Accelerate security incident detection, forensic analysis, and response by being able to quickly analyze logs from disparate applications and systems across your network.

Key features

Key Amazon OpenSearch Service functionality

Currency & scalability

Improve search quality and relevance with K-Nearest Neighbor (K-NN) and Learning to Rank (LTR) models

Update search accuracy on the fly with custom dictionaries and hot-reload of synonym files

Easy to use

Secure your domain at every level with Fine Grained Access Control and Audit Logging

Troubleshoot performance and availability issues in your distributed applications with Trace Analytics

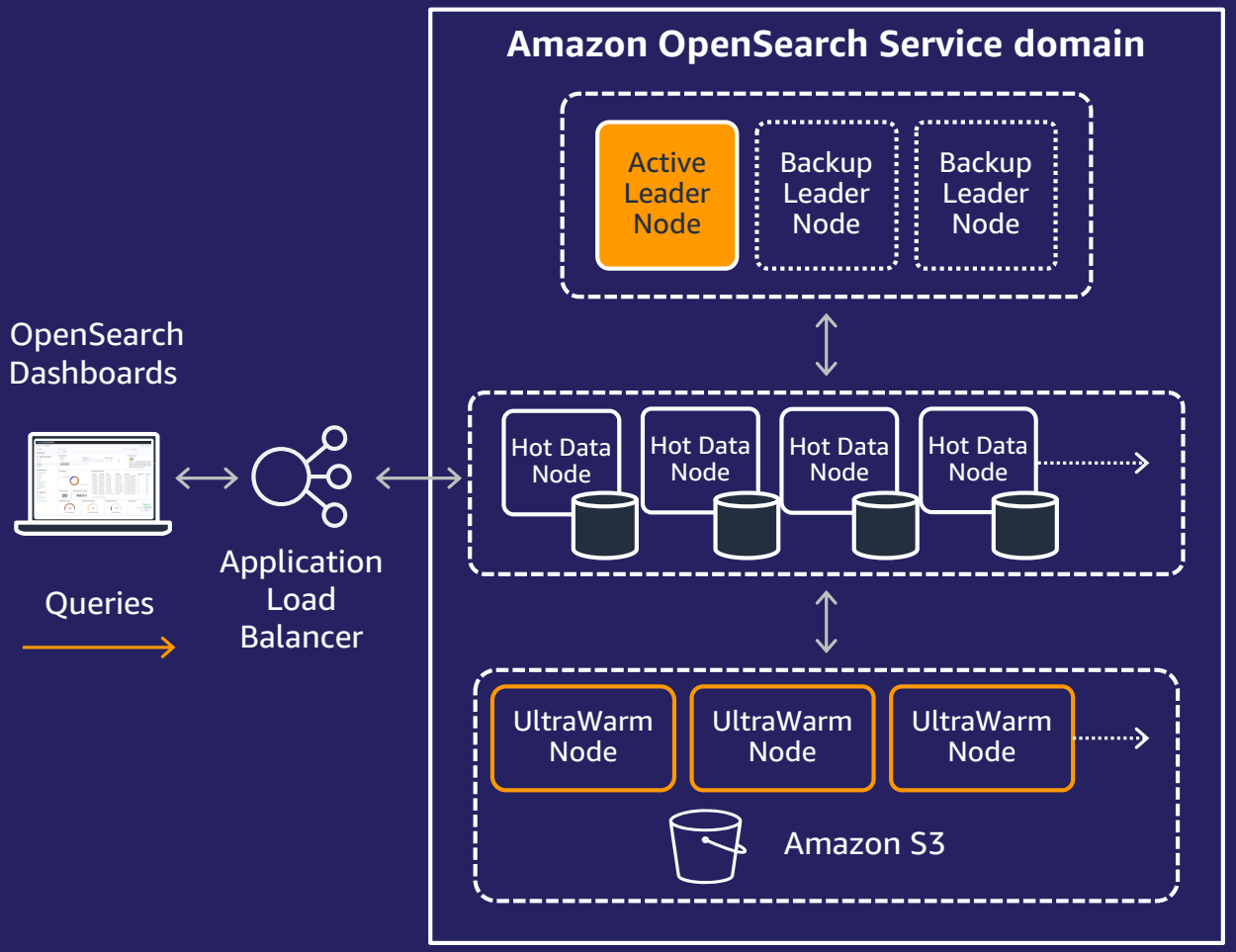
Security & availability

Lower your storage costs and extend your data retention with UltraWarm and Cold Storage

Self-healing nodes and **automatically optimize memory resources** with Auto-tune

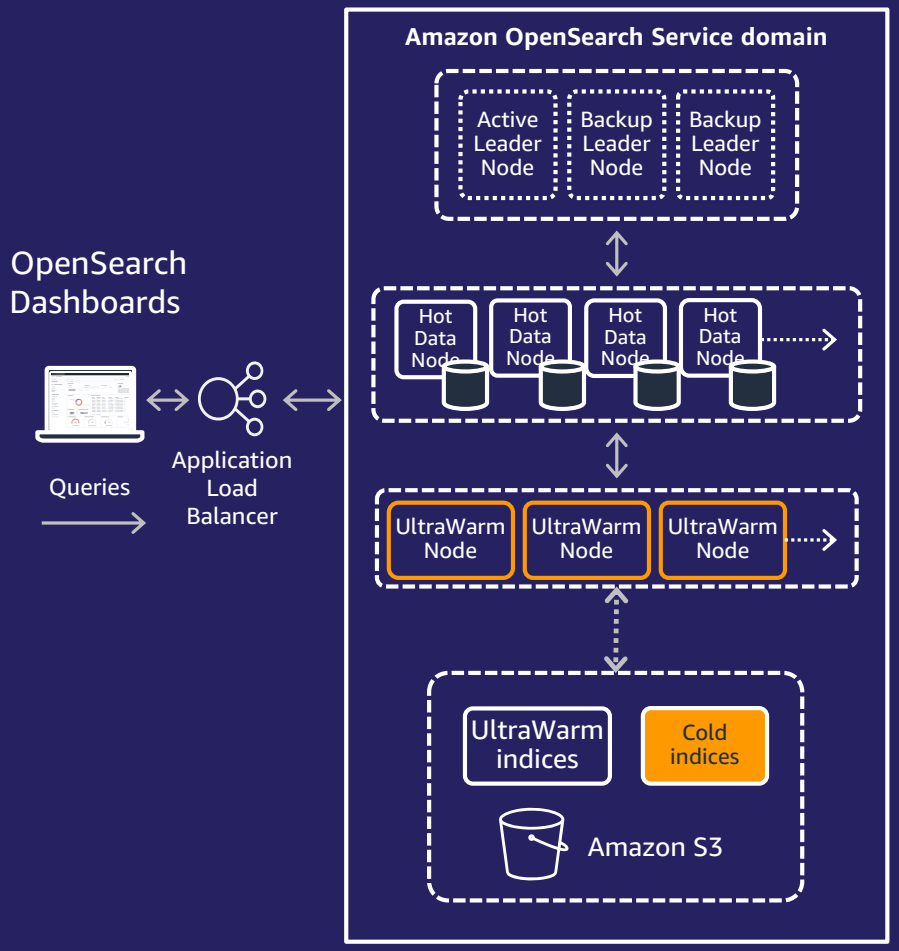
UltraWarm for Amazon OpenSearch Service

A low-cost storage tier for Amazon OpenSearch Service



- Up to 90% storage cost reduction vs. hot storage tier
- Scale up to 3 PB per domain
- Analyze years of operational data
- Interactive log analytics and visualization

Cold storage



A fully-managed lowest cost storage tier to durably store infrequently accessed or historical logs at near Amazon S3 prices.

No limits: You can keep as much data you need in cold storage.

Keep your data: Don't throw away older data. Generate valuable business insights from historical long tail data.

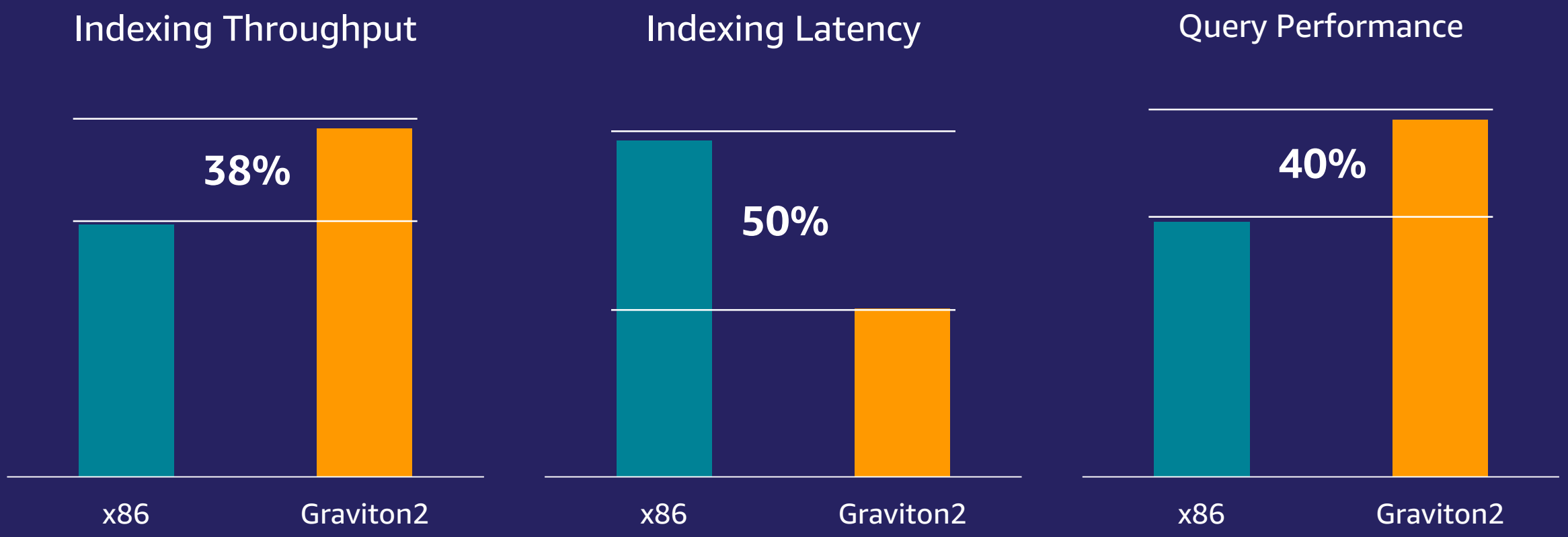
Lower cost: Decouple compute from storage. No longer pay for compute for data that you only need to analyze infrequently.

Easy: Simple discovery features gives your users secure access to their data via self-service.

On-demand access: Migrate the data you need to analyze from cold storage to your UltraWarm nodes in seconds.

AWS Graviton2 on Amazon OpenSearch Service

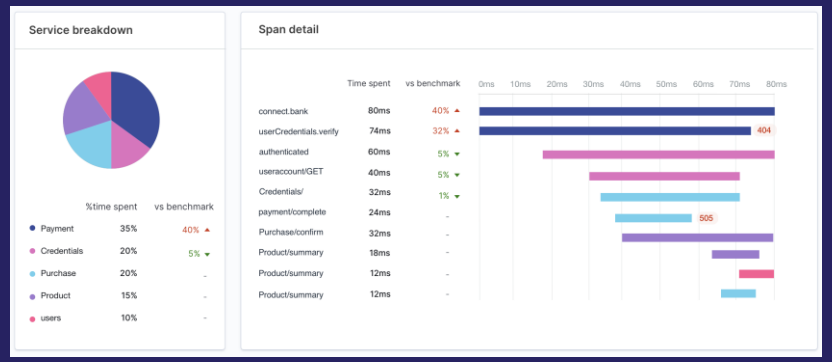
Custom-built silicon with next-generation performance improvement

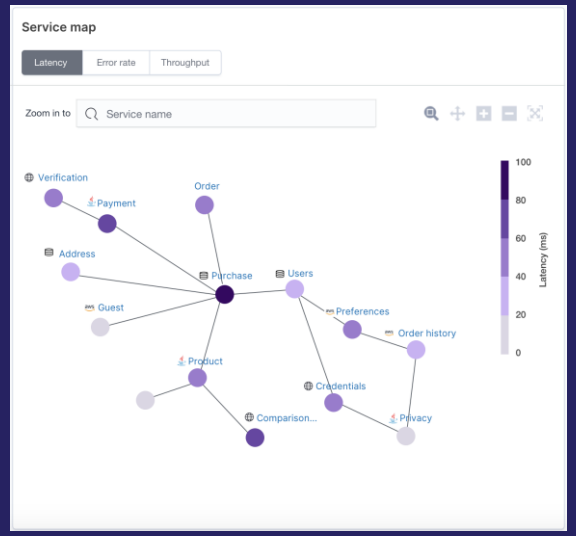


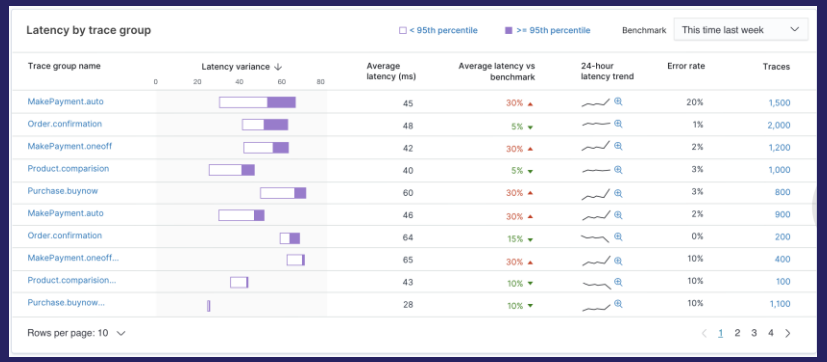
44% better price to performance compared to fifth-generation instances

What is Trace Analytics?

Supports OpenTelemetry (Jaeger, Zipkin, 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

Cross-cluster search for Amazon OpenSearch Service

Increase scalability, efficiency, and availability, by separating distinct workloads

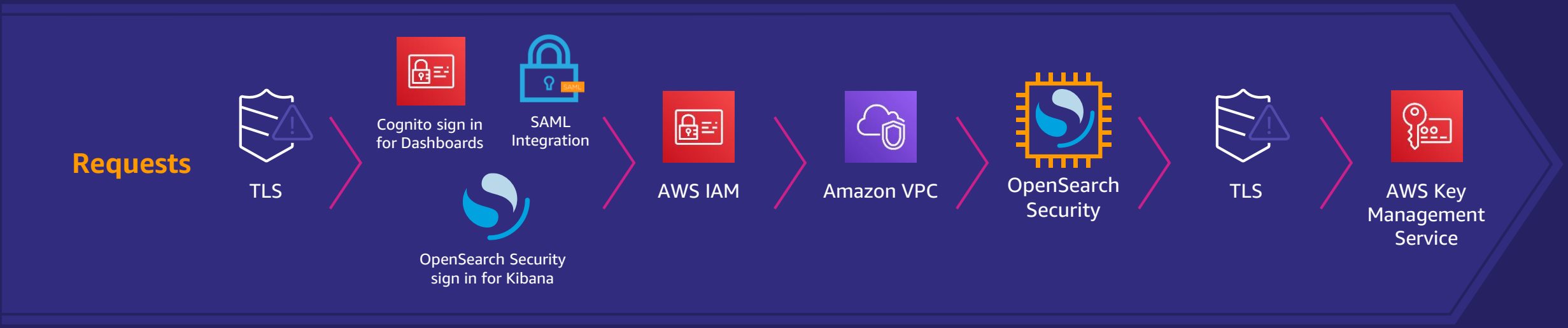
Search and visualize data across multiple domains from a single OpenSearch Dashboards interface

Increase efficiency— optimize domain resources for specific workloads

Improve availability— isolate failures to specific workloads

Secured through Fine Grained Access Control (FGAC) policies

Multi-layer security with Amazon OpenSearch Service



Integrate with SAML and Cognito for OpenSearch Dashboards login

IAM to control access to the endpoint

Use a private endpoint to deploy into your VPC and security groups for traffic control.

Use OpenSearch fine-grained access control to secure your data and dashboards

Encrypt your data, in flight and at rest

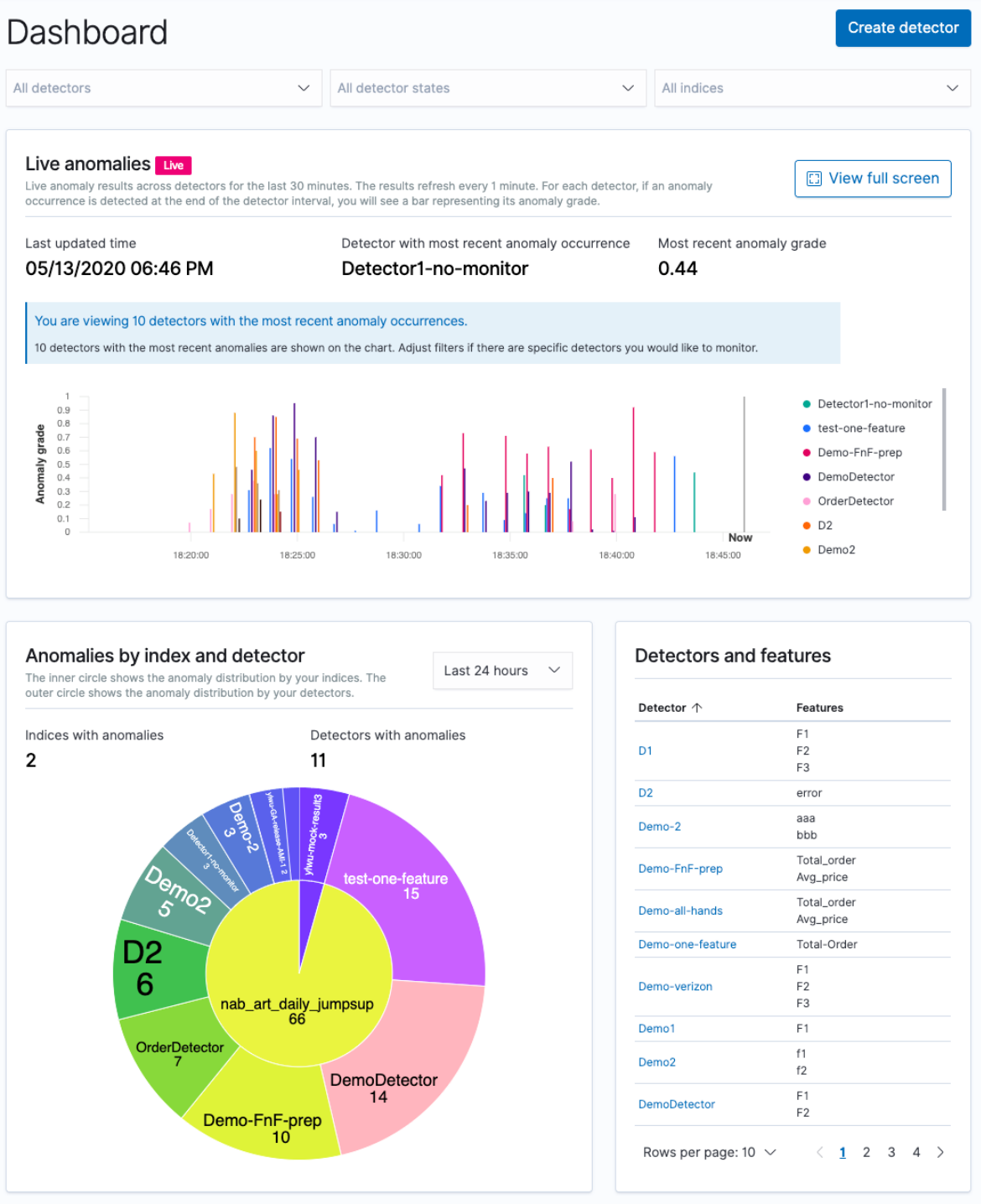
Machine learning in Amazon OpenSearch Service

Mitigate issues faster with anomaly detection in streaming data

Improve search quality and relevance with K-Nearest Neighbor (K-NN) and Learning to Rank models

Performant at scale. Machine learning models are distributed and processed across nodes

Easy to use. Machine learning expertise is not required to leverage the service



Getting started

Resources

OpenSearch project
opensearch.org

Amazon OpenSearch Service Immersion Days

Provides a deep dive into Amazon OpenSearch Service through a mix of online trainings and hands-on labs led by AWS Solutions Architects. You will learn all the key concepts to leverage the service along with the operational best practices.

Interested in scheduling Immersion Days?

Contact us
searchservices-ww-gtm@amazon.com

New releases

What's New

Documentation

Developer Guide

Blogs

Moving to managed: The case for the Amazon OpenSearch Service

Best practices for configuring your Amazon OpenSearch Service domain

Simple to get started…

1

Create an AWS Free Tier account

2

Launch an Amazon OpenSearch Service Cluster in minutes

3

Follow the Getting Started Tutorial to build a log analytics solution

Thank you!