
AWS
re:Invent

ANT320

What's new in Amazon Elasticsearch Service

Jon Handler

Principal Specialist SA, Search Services

AWS

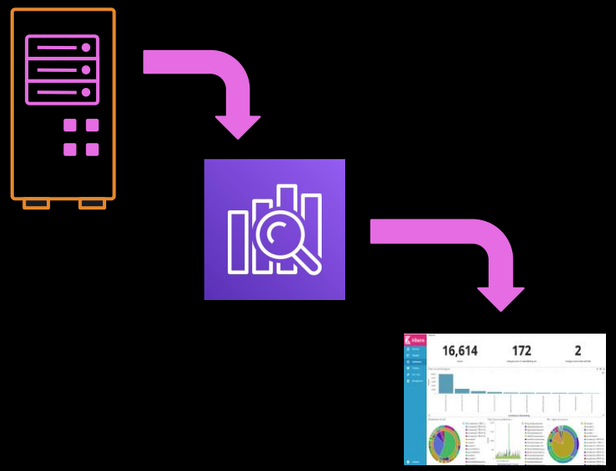
AWS re:Invent

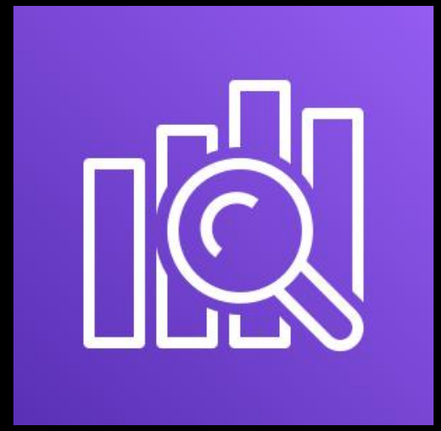
© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Elasticsearch Service (Amazon ES)

STORE, SEARCH, AND ANALYZE ALL YOUR LOGS







Challenge: Operational and Security data is distributed across logs and infrastructure. You need a high-level view for monitoring and responding to incidents

Use Amazon Elasticsearch Service and the bundled Kibana to analyze log data in near real time to monitor your infrastructure, application, security, IoT, and AWS services

Amazon Elasticsearch Service is a fully managed service that makes it simple to deploy, scale, monitor, and secure Elasticsearch in the AWS cloud

Amazon Elasticsearch Service

FULLY MANAGED, SCALABLE, AND SECURE ELASTICSEARCH SERVICE



Fully managed



Access to all data



Cost effective



Secure and compliant



Highly scalable and available



Tightly integrated with AWS ecosystem

Self-managed vs. Amazon Elasticsearch Service

Self managed		Amazon ES
On-Premise	**Amazon EC2**	**Amazon Elasticsearch Service**
App dev / optimization	App dev / optimization	App dev / optimization
Hot / warm storage tiers	Hot / warm storage tiers	UltraWarm storage tier
Plugins (additional cost)*	Plugins (additional cost)*	Plugins*
In-place upgrades / patches	In-place upgrades / patches	In-place upgrades / patches
In-place upgrades	In-place upgrades	In-place upgrades
Cluster scaling	Cluster scaling	Cluster scaling
Cross AZ data transfer cost	Cross AZ data transfer cost	No cross AZ data transfer cost
Backups	Backups	Hourly backups
HA, 24x7 monitoring & repair	HA, 24x7 monitoring & repair	HA, 24x7 monitoring & repair
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

USE CASE #1

Search

COMPASS

Re:Invent Session

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 Elasticsearch 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 #2

Application and 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 Elasticsearch 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

Overview and best practices



Best practices blog post

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How does Elasticsearch work

1

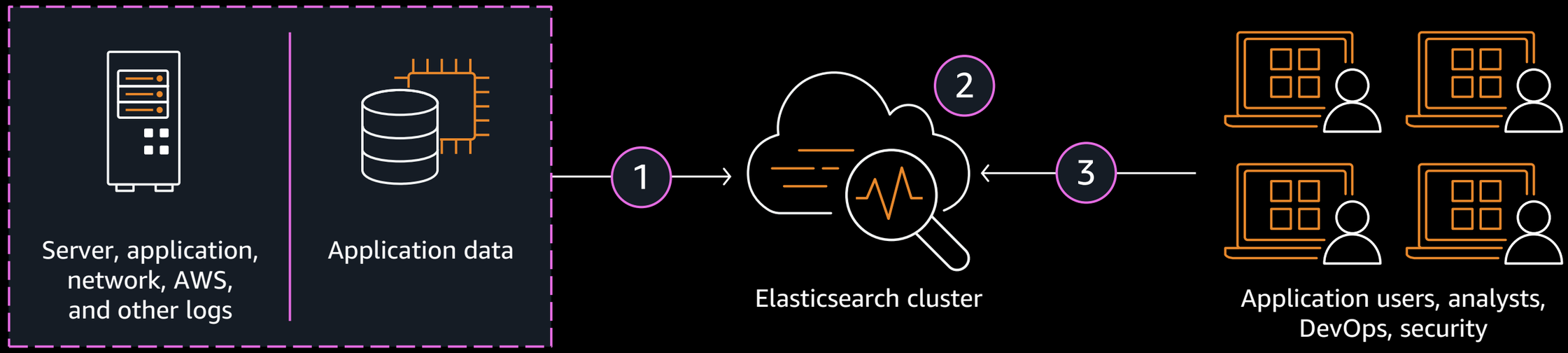
Send data as JSON via REST APIs

2

Data is indexed—all fields searchable, including nested JSON

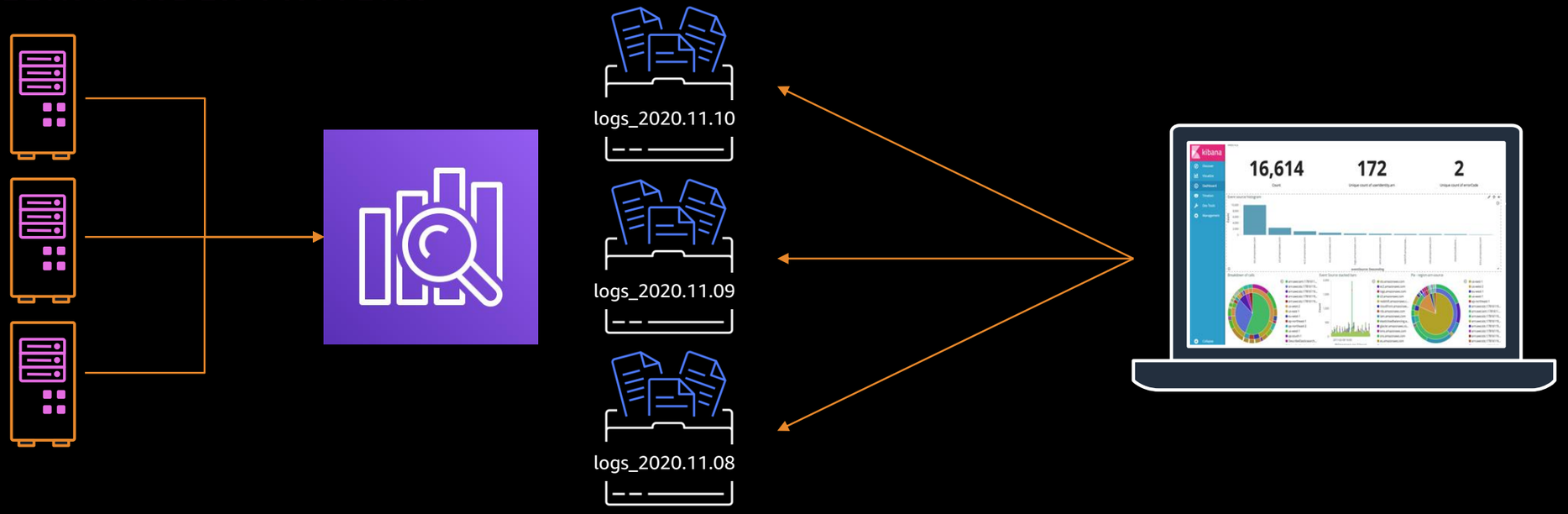
3

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



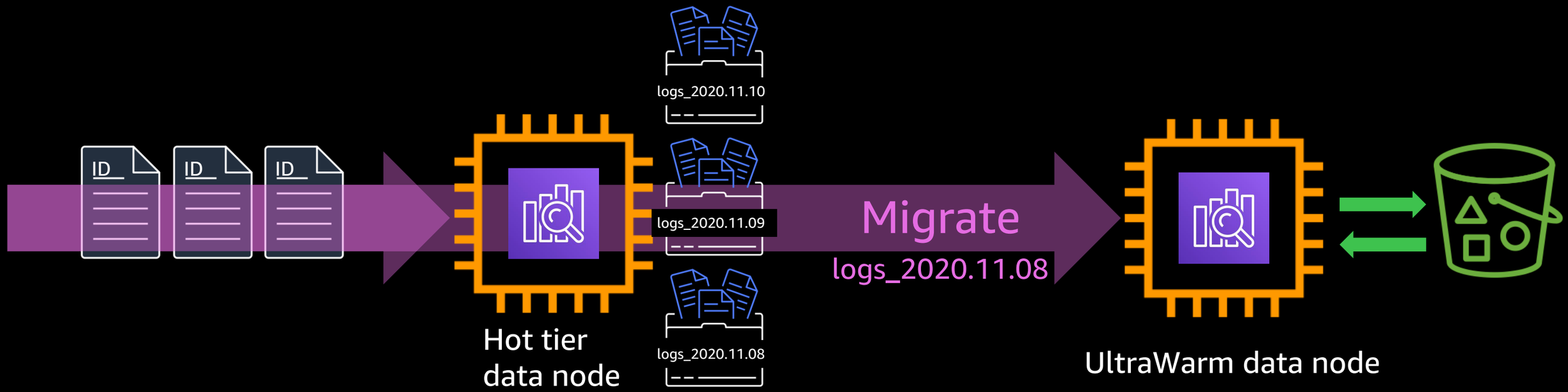
Best practice data lifecycle

CREATE A ROLLING INDEX PATTERN



You send log data to Amazon Elasticsearch Service
Elasticsearch uses indexes to organize searchable/analyzable data
You create indexes on time (or size) boundaries
Kibana and REST API access the data based on the collection of indices
You delete the oldest index at the end of its lifecycle

Use UltraWarm for long-term, low-cost storage

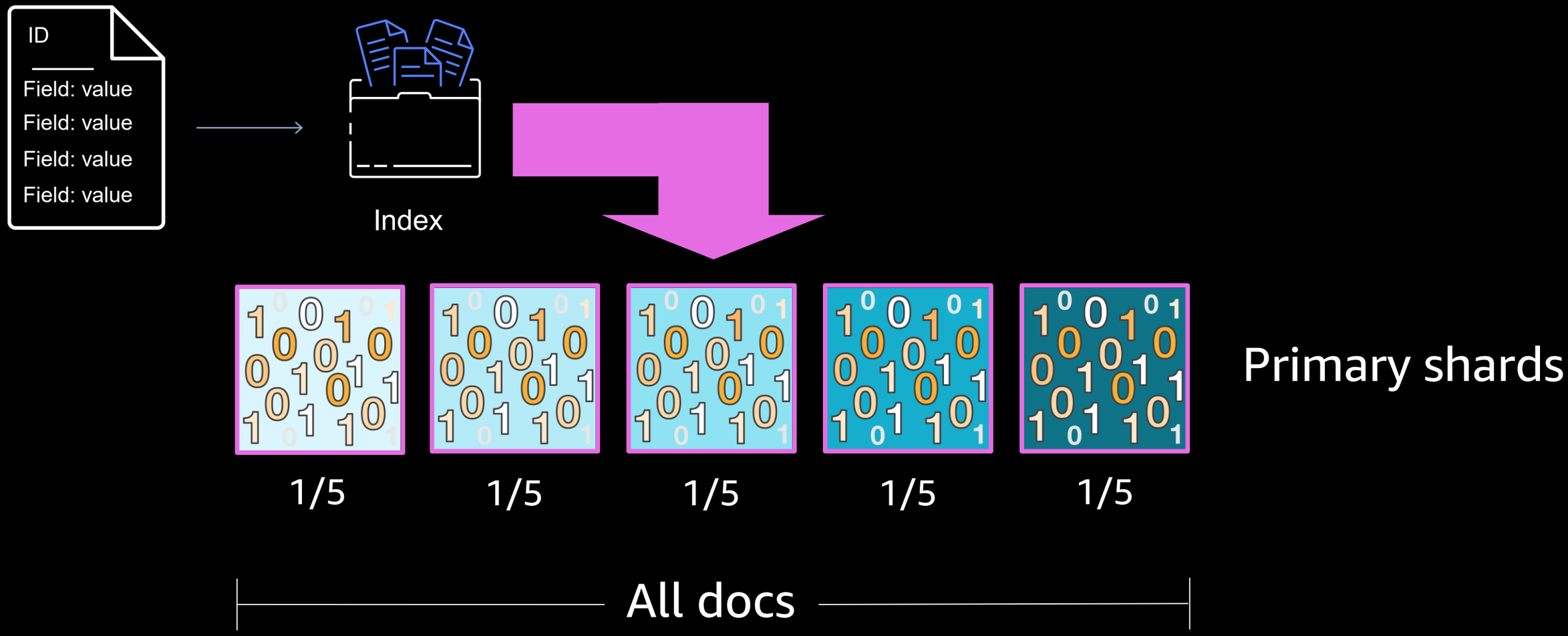


Data is indexed and stored in the hot tier, with a rolling index pattern

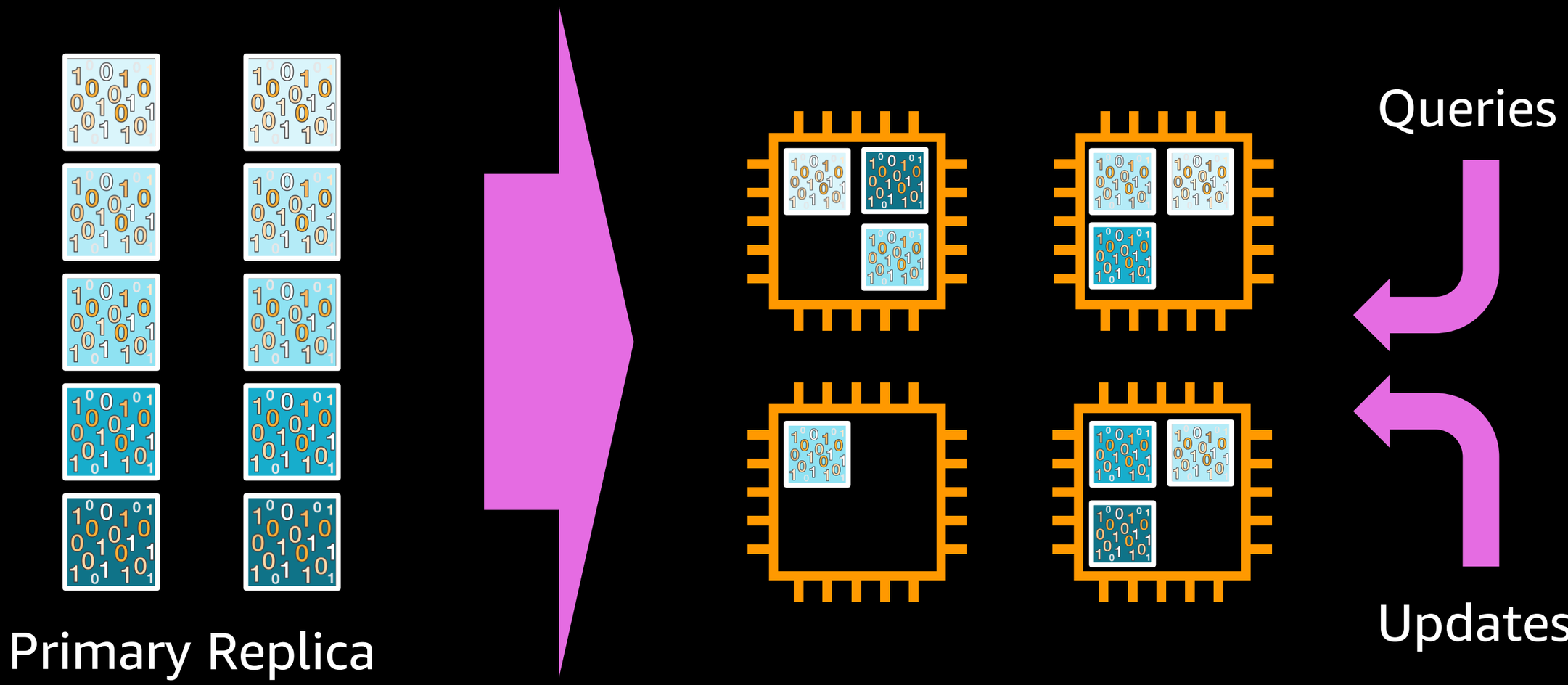
You migrate a single index via API or Index State Management

Force merge, drop replicas, snapshot during migration

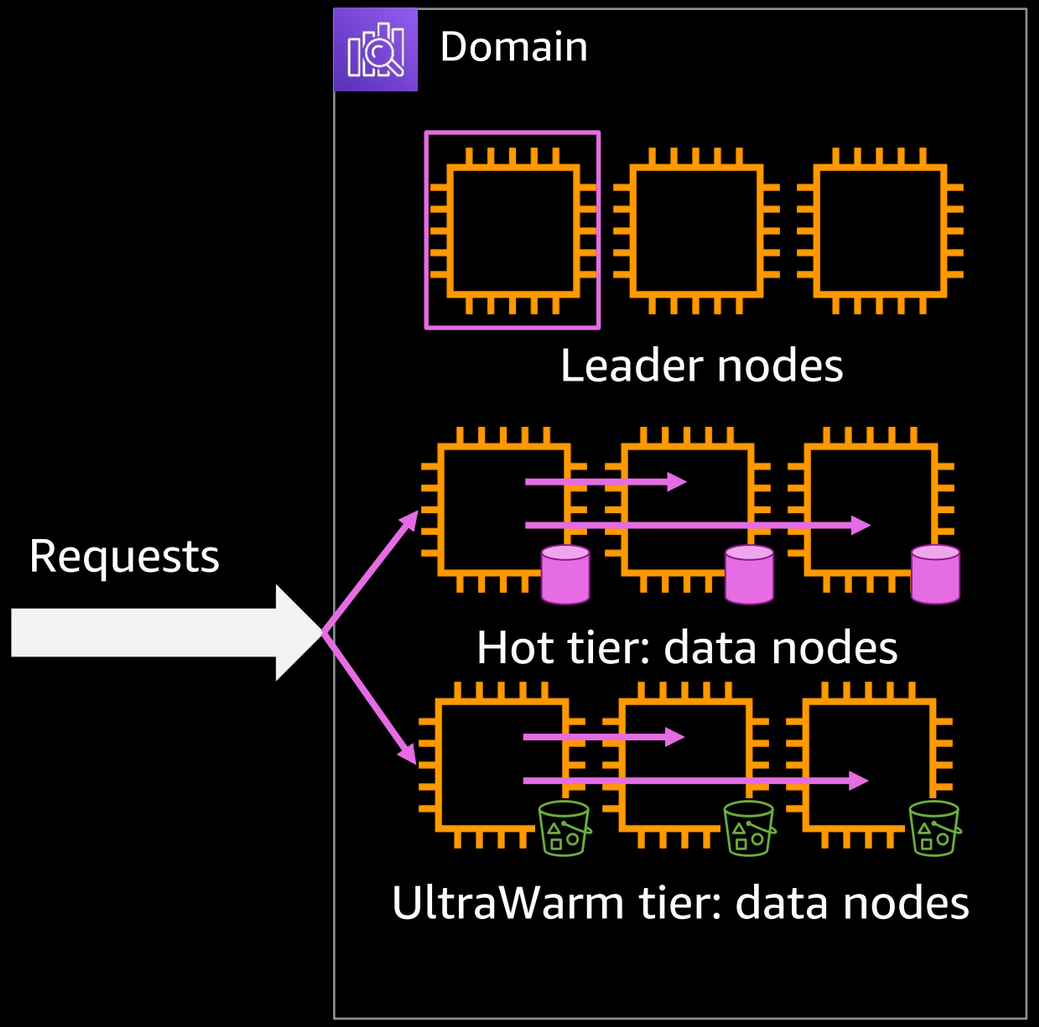
Indexes are comprised of shards



Elasticsearch distributes shards to data nodes



Processing requests



- Amazon ES creates an Elasticsearch cluster, with EC2 instances
- Dedicated leader nodes maintain the cluster state
- Data nodes index and analyze data
  - Hot data nodes use instance storage
  - UltraWarm data nodes use S3 for storage
- Elasticsearch routes requests seamlessly

Size first by storage then by CPU

Best practices blog post

Storage

On disk an index with 1 replica is ~2.2-2.5 times the source data size. Deploy instances with that storage plus 25% overhead

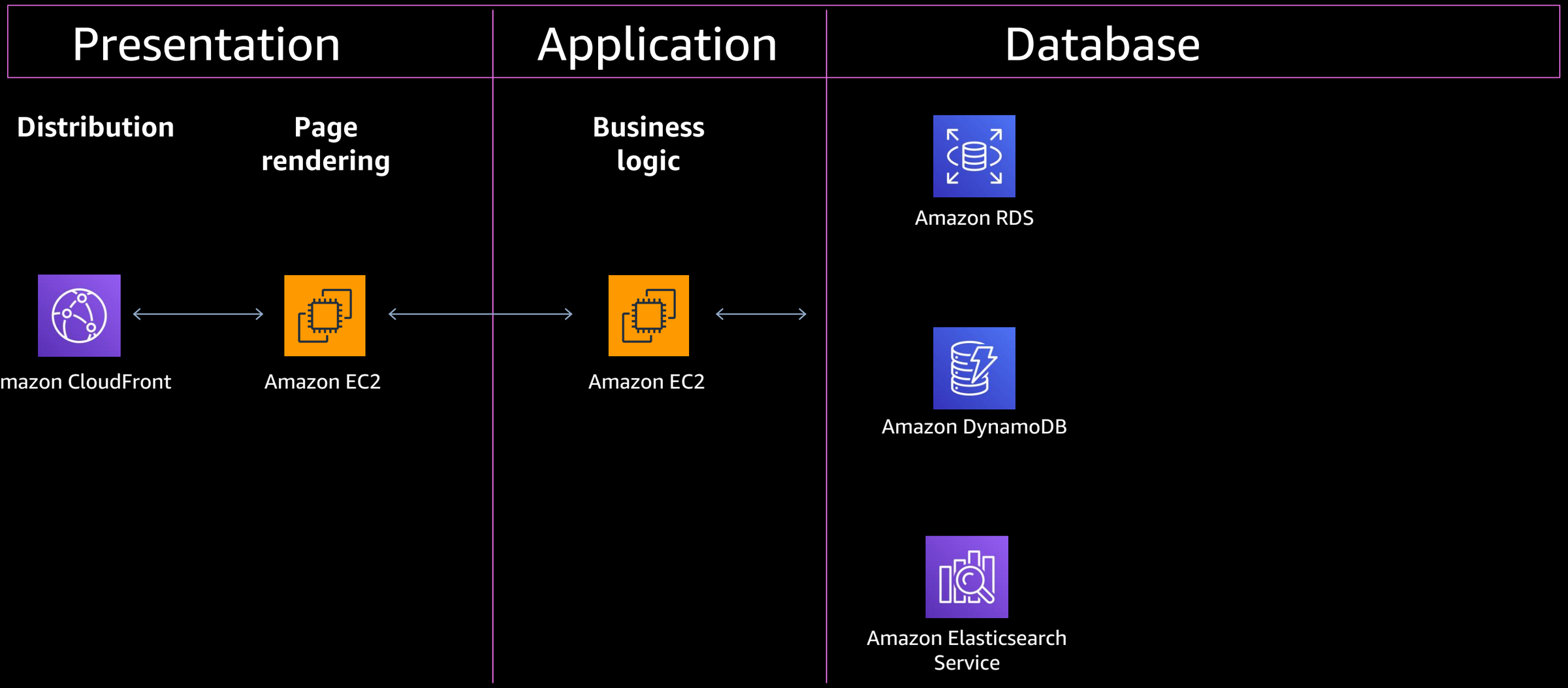
CPU

Each shard uses a CPU to process a request. Deploy instances with vCPUs equal to 1.5 times the number of shards that receive substantial traffic

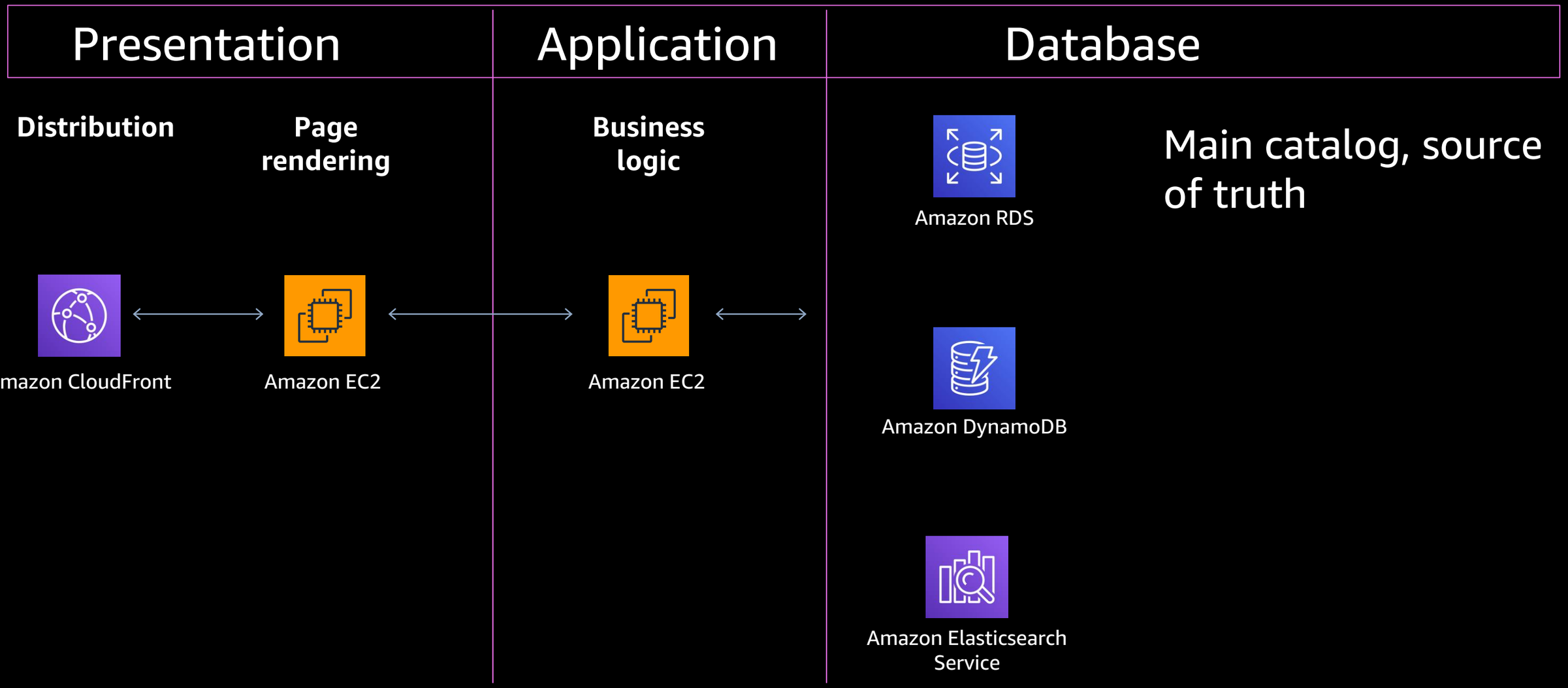
Instance Types

R5/I3 are currently the best performing instance types for data nodes. Use C5 for dedicated leader nodes

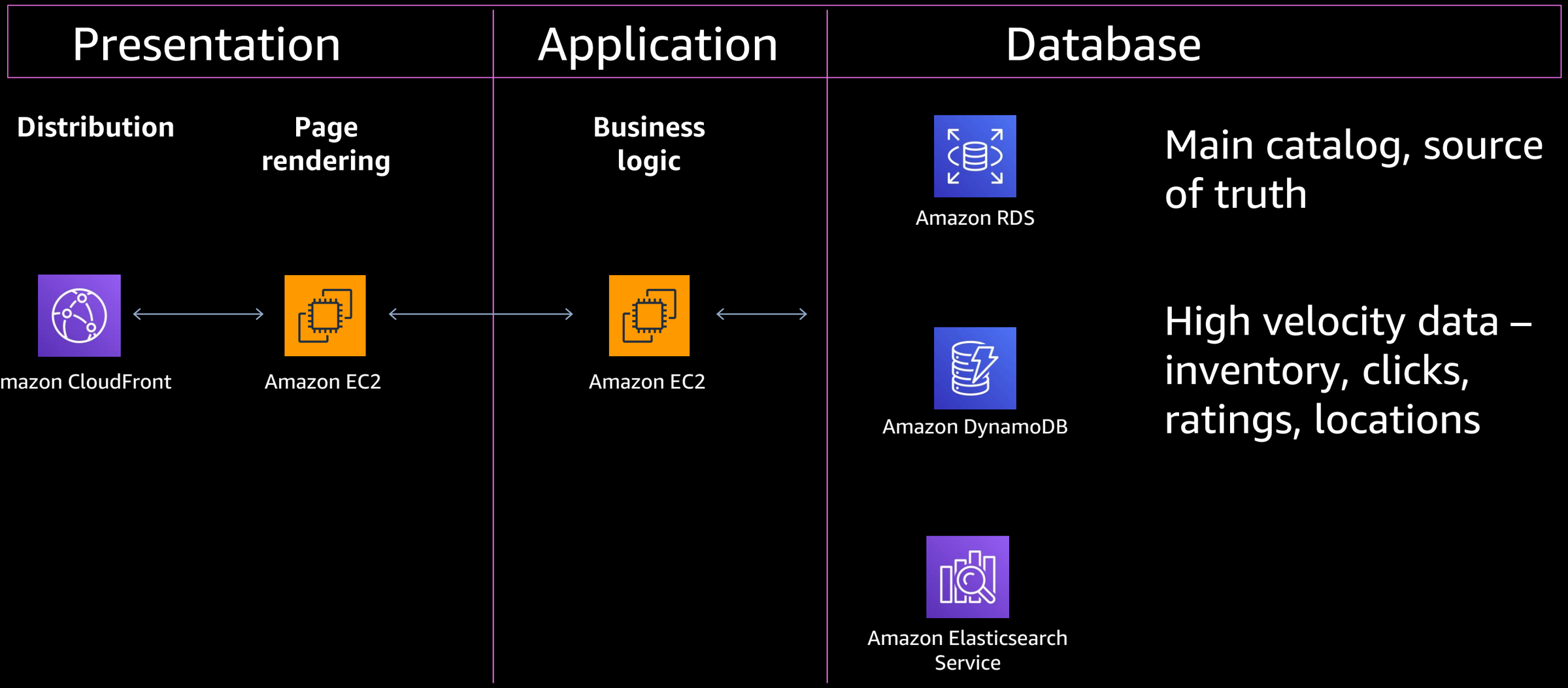
Search components and data flow



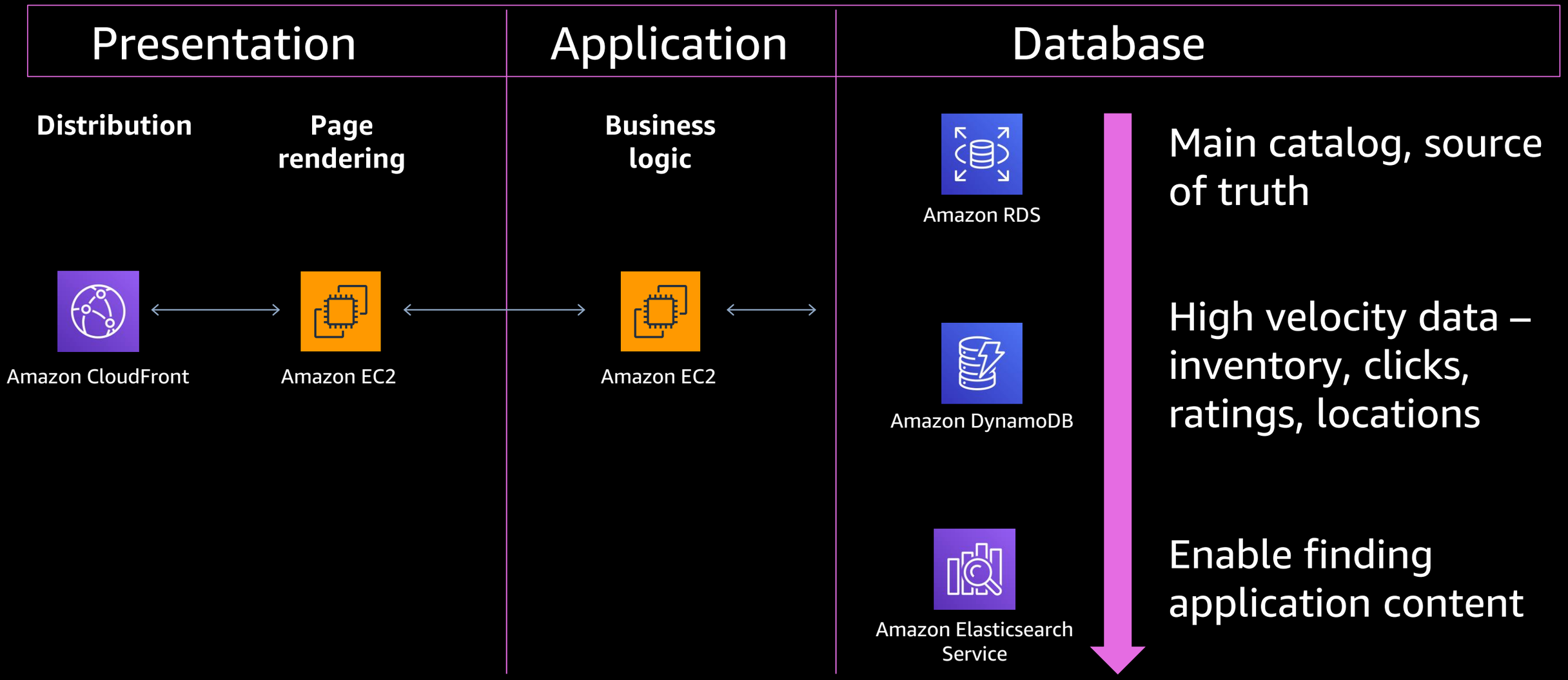
Search components and data flow



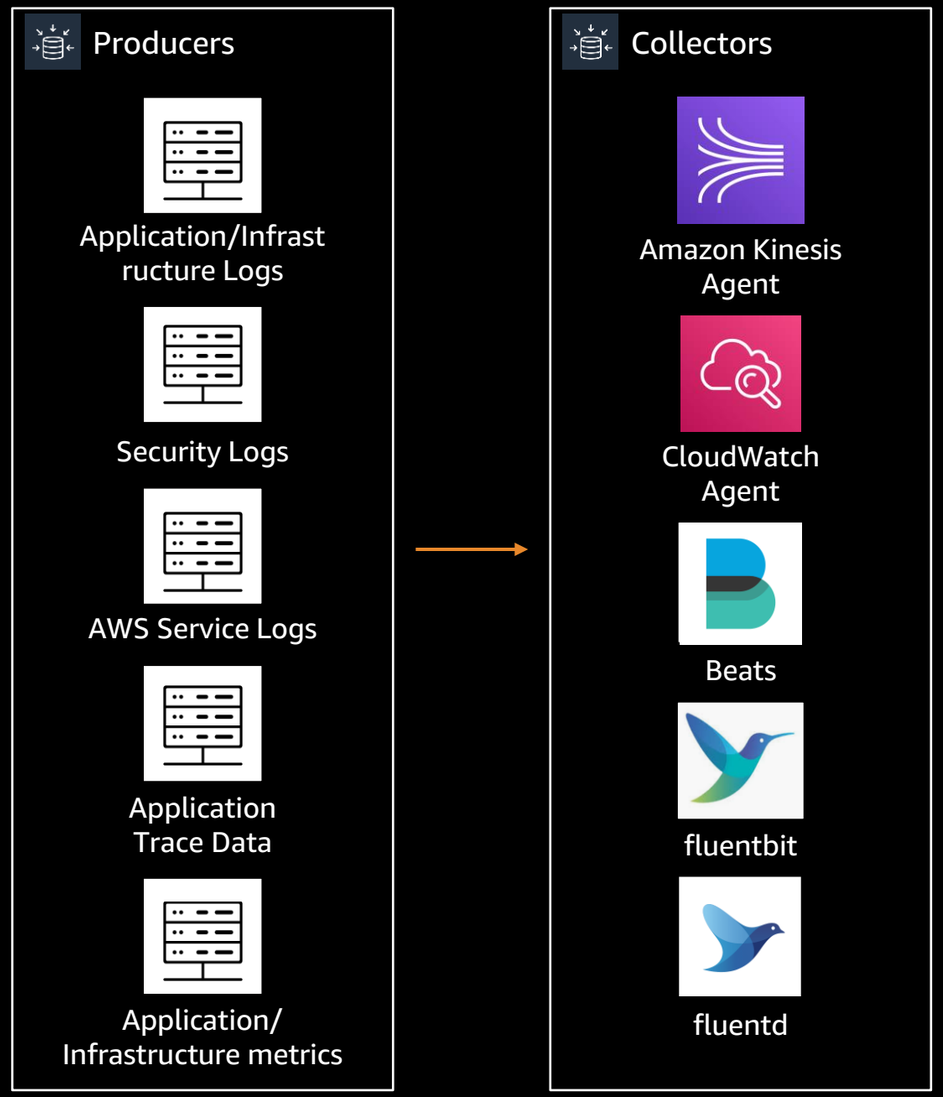
Search components and data flow



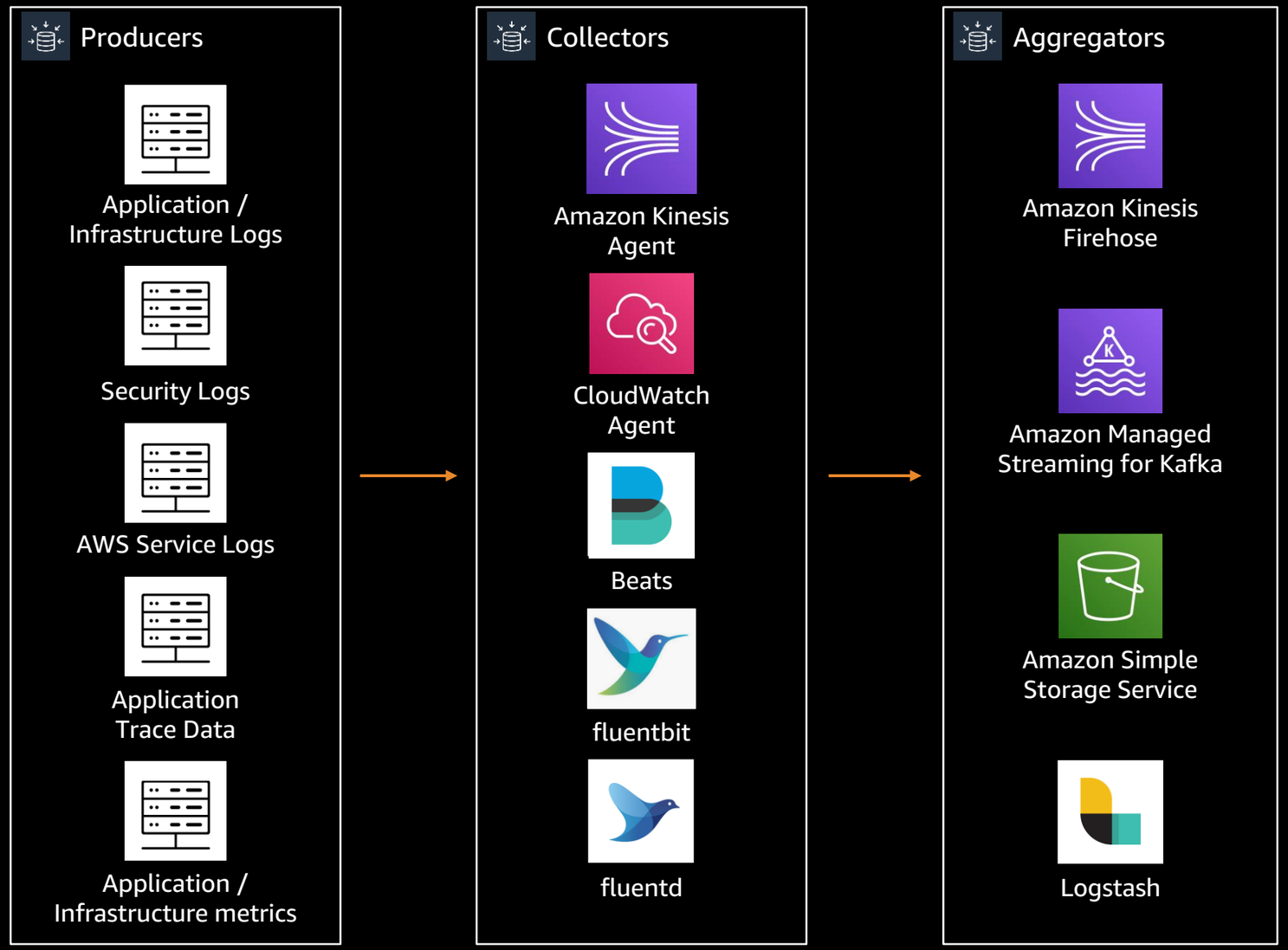
Search components and data flow



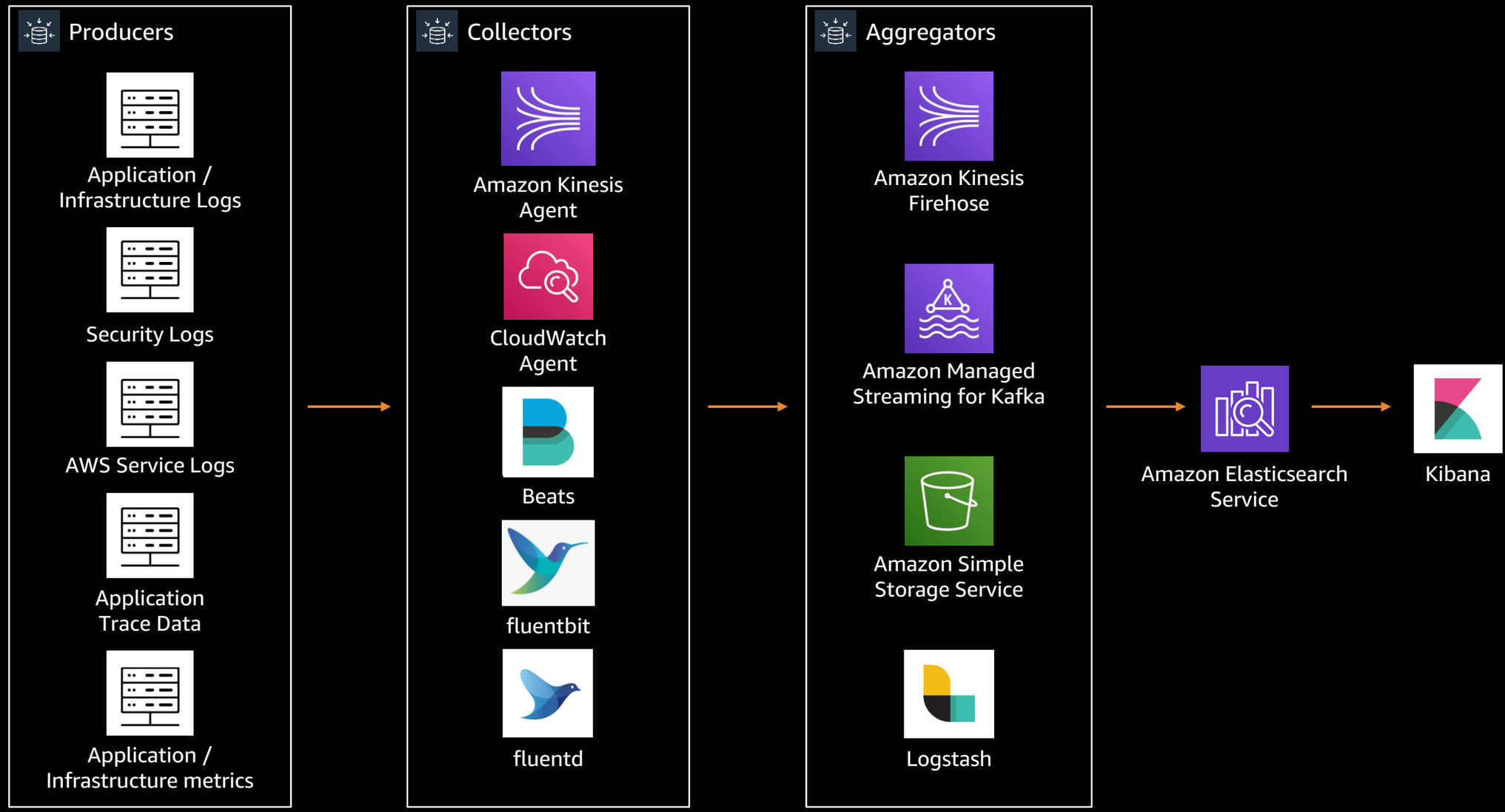
Log Analytics components and data flow



Log Analytics components and data flow



Log Analytics components and data flow



Security



Security online tech talk

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-layer security with Amazon ES

Requests



TLS

Encrypted from end to end—in flight with Transport Layer Security (TLS), at rest with AWS Key Management Service (KMS).

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

Multi-layer security with Amazon ES

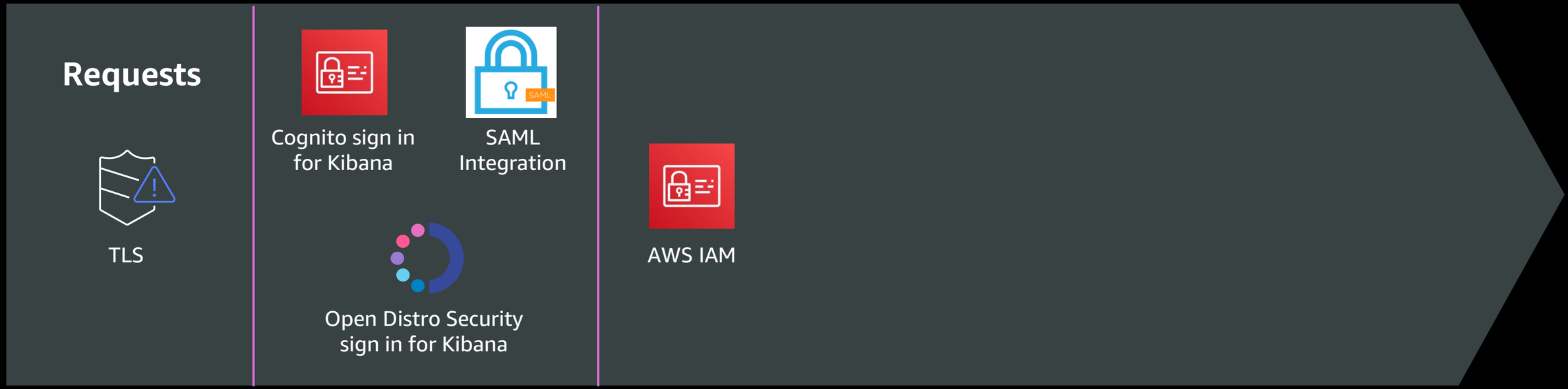


Encrypted from end to end—in flight with Transport Layer Security (TLS), at rest with AWS Key Management Service (KMS).

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

Includes Kibana login via Cognito integration, or native with Open Distro Security.

Multi-layer security with Amazon ES



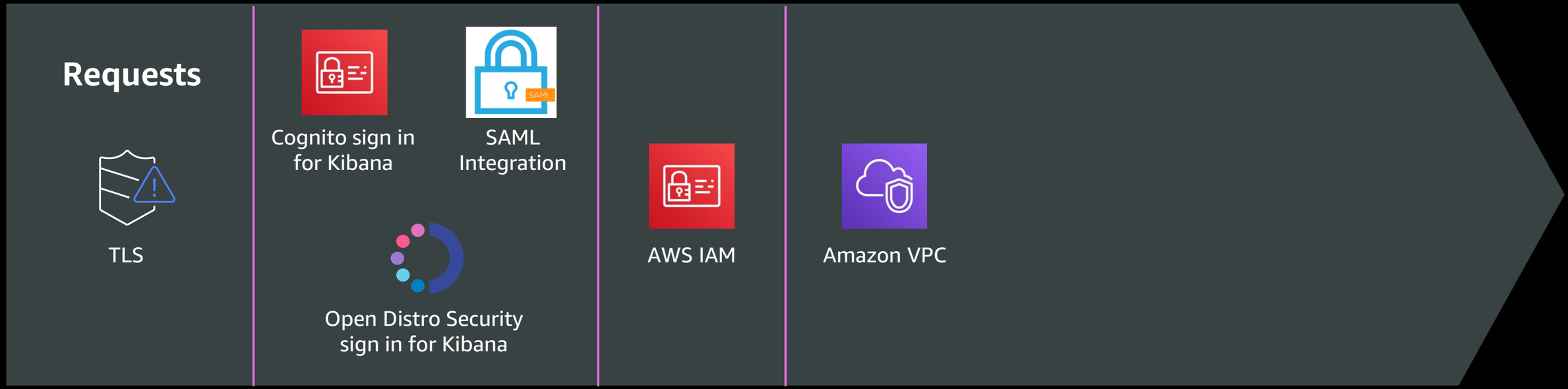
Encrypted from end to end—in flight with Transport Layer Security (TLS), at rest with AWS Key Management Service (KMS).

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

Includes Kibana login via Cognito integration, or native with Open Distro Security.

Coarse-grained access control with AWS Identity and Access Management (IAM) policies.

Multi-layer security with Amazon ES



Encrypted from end to end—in flight with Transport Layer Security (TLS), at rest with AWS Key Management Service (KMS).

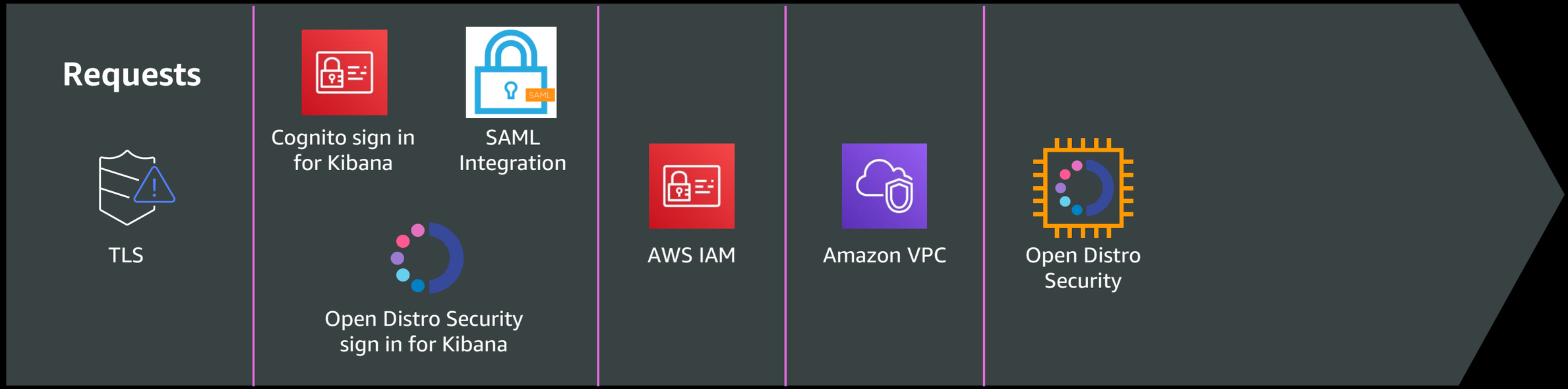
Use a private endpoint to deploy into your VPC and security groups for traffic control.

Includes Kibana login via Cognito integration, or native with Open Distro Security.

Coarse-grained access control with AWS Identity and Access Management (IAM) policies.

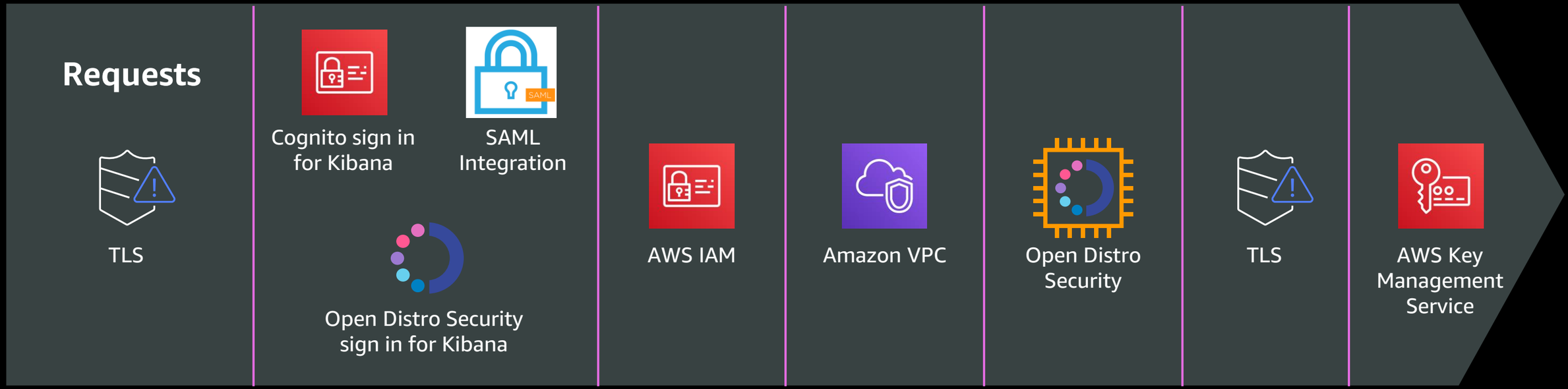
Use a private endpoint to deploy into your VPC and security groups for traffic control.

Multi-layer security with Amazon ES



Fine-grained access control for tighter control over your data.

Multi-layer security with Amazon ES



Fine-grained access control for tighter control over your data.

TLS end to end and integration with AWS KMS for at-rest encryption.

Have your front end and monitor it too!

Visit this hands-on lab to explore a search application with logging and secure access

Hands-on lab

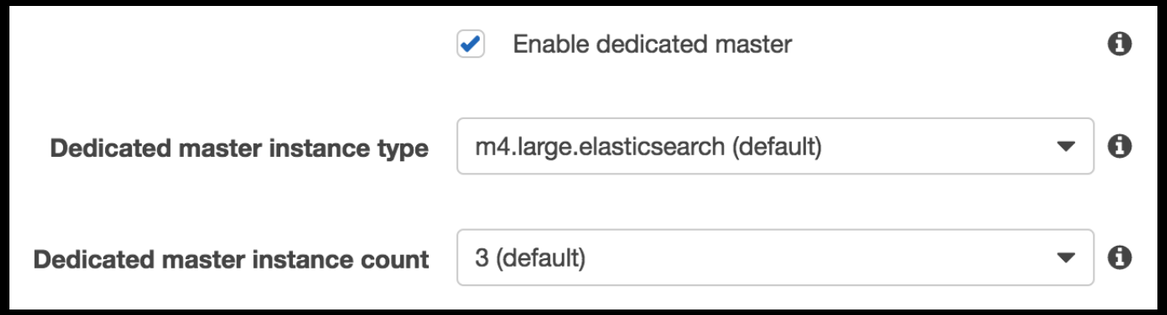
Reliability



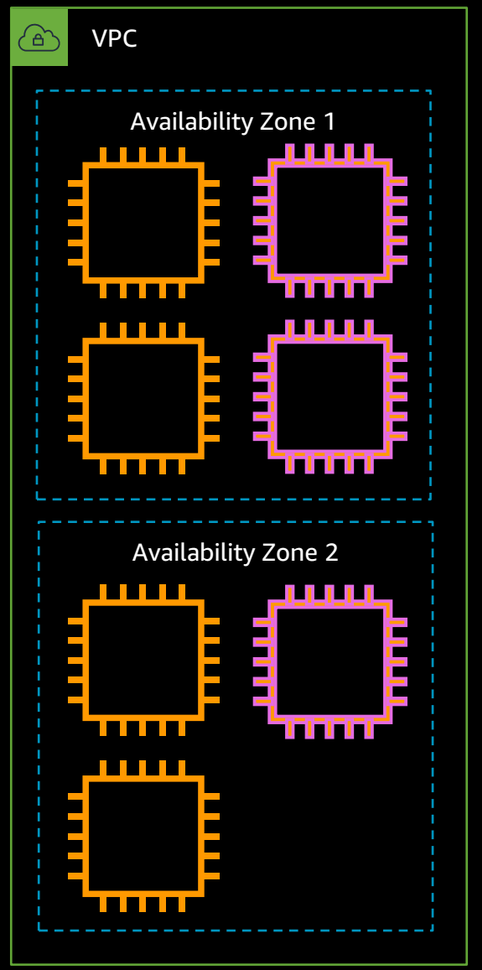
AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Easily deploy dedicated leader nodes



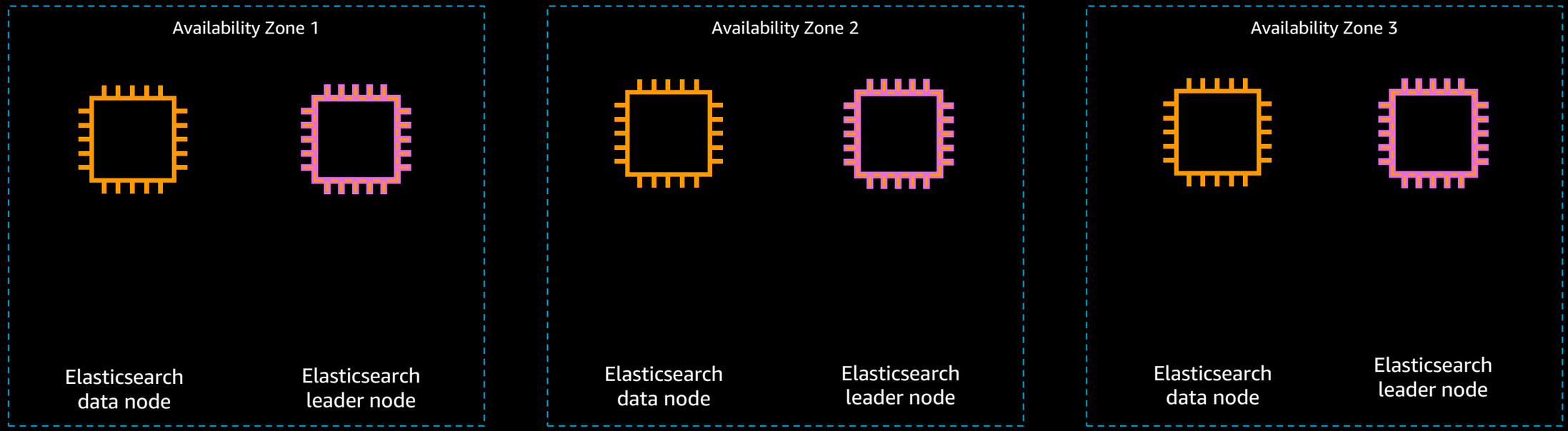
- Leader nodes orchestrate and make your cluster more stable
- Use 3 leader nodes
- Leader nodes are deployed to 1 or 3 AZs



Amazon ES Domain

Blog post

One click for a higher availability solution



For best practice, highest availability:

- Use 3 zones, 2 replicas, and 300% capacity
- Take manual snapshots to your own S3 bucket
- Use dual-write to a second domain as a hot or warm standby

Blog post

A best practices mini-checklist

- [x] Deploy, monitor, adjust
- [x] Use the _bulk API
- [x] Use 3 Dedicated leader nodes
- [x] Select 3 zones with Multi-AZ
- [x] Set a template and use 2 replicas
- [x] Free storage >= 25%
- [x] Shards:vCPU = 1:1.5
- [x] Shard size = 50GB (logs), less for search
- [x] Max shards per node = 25 * JVM heap size
- [x] Total shard count < 30K

A best practices mini-checklist

- √ Deploy, monitor, adjust
- √ Use the _bulk API
- √ Use 3 Dedicated leader nodes
- √ Select 3 zones with Multi-AZ
- √ Set a template and use 2 replicas
- √ Free storage >= 25%
- √ Shards:vCPU = 1:1.5
- √ Shard size = 50GB (logs), less for search
- √ Max shards per node = 25 * JVM heap size
- √ Total shard count < 30K

A best practices mini-checklist

- √ Deploy, monitor, adjust
- √ Use the _bulk API
- √ Use 3 Dedicated leader nodes
- √ Select 3 zones with Multi-AZ
- √ Set a template and use 2 replicas
- √ Free storage >= 25%
- √ Shards:vCPU = 1:1.5
- √ Shard size = 50GB (logs), less for search
- √ Max shards per node = 25 * JVM heap size
- √ Total shard count < 30K

A best practices mini-checklist

- √ Deploy, monitor, adjust
- √ Use the _bulk API
- √ Use 3 Dedicated leader nodes
- √ Select 3 zones with Multi-AZ
- √ Set a template and use 2 replicas
- √ Free storage >= 25%
- √ Shards:vCPU = 1:1.5
- √ Shard size = 50GB (logs), less for search
- √ Max shards per node = 25 * JVM heap size
- √ Total shard count < 30K

A best practices mini-checklist

- √ Deploy, monitor, adjust
- √ Use the _bulk API
- √ Use 3 Dedicated leader nodes
- √ Select 3 zones with Multi-AZ
- √ Set a template and use 2 replicas
- √ Free storage >= 25%
- √ Shards:vCPU = 1:1.5
- √ Shard size = 50GB (logs), less for search
- √ Max shards per node = 25 * JVM heap size
- √ Total shard count < 30K

A best practices mini-checklist

- √ Deploy, monitor, adjust
- √ Use the _bulk API
- √ Use 3 Dedicated leader nodes
- √ Select 3 zones with Multi-AZ
- √ Set a template and use 2 replicas
- √ Free storage >= 25%
- √ Shards:vCPU = 1:1.5
- √ Shard size = 50GB (logs), less for search
- √ Max shards per node = 25 * JVM heap size
- √ Total shard count < 30K

A best practices mini-checklist

- ☑ Deploy, monitor, adjust
- ☑ Use the _bulk API
- ☑ Use 3 Dedicated leader nodes
- ☑ Select 3 zones with Multi-AZ
- ☑ Set a template and use 2 replicas
- ☑ Free storage >= 25%
- ☑ Shards:vCPU = 1:1.5
- ☑ Shard size = 50GB (logs), less for search
- ☑ Max shards per node = 25 * JVM heap size
- ☑ Total shard count < 30K

A best practices mini-checklist

- √ Deploy, monitor, adjust
- √ Use the _bulk API
- √ Use 3 Dedicated leader nodes
- √ Select 3 zones with Multi-AZ
- √ Set a template and use 2 replicas
- √ Free storage >= 25%
- √ Shards:vCPU = 1:1.5
- √ Shard size = 50GB (logs), less for search
- √ Max shards per node = 25 * JVM heap size
- √ Total shard count < 30K

A best practices mini-checklist

- √ Deploy, monitor, adjust
- √ Use the _bulk API
- √ Use 3 Dedicated leader nodes
- √ Select 3 zones with Multi-AZ
- √ Set a template and use 2 replicas
- √ Free storage >= 25%
- √ Shards:vCPU = 1:1.5
- √ Shard size = 50GB (logs), less for search
- √ Max shards per node = 25 * JVM heap size
- √ Total shard count < 30K

A best practices mini-checklist

- √ Deploy, monitor, adjust
- √ Use the _bulk API
- √ Use 3 Dedicated leader nodes
- √ Select 3 zones with Multi-AZ
- √ Set a template and use 2 replicas
- √ Free storage >= 25%
- √ Shards:vCPU = 1:1.5
- √ Shard size = 50GB (logs), less for search
- √ Max shards per node = 25 * JVM heap size
- √ Total shard count < 30K

What we've been up to



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Open Distro for Elasticsearch

Open Distro Page





Amazon ES

Amazon ES Overview





Open Distro Plugins

Security, Alerting, SQL, kNN, Index State Management, Streaming Anomaly Detection, Audit Logging, Learn to Rank, SAML Integration

Performance Analyzer, Piped Processing Language SQL WorkBench



Service Software

Open Distro Plugins

Security, Alerting, SQL, kNN, Index State Management, Streaming Anomaly Detection, Audit Logging, Learn to Rank, SAML Integration

Amazon Elasticsearch Service – last 18 months

What's new page

Currency and scalability

- New Elasticsearch versions: 6.5, 6.7, 6.8, 7.1, 7.4, 7.7, 7.8
- 3PB, 200 node clusters
- New instances: M5, R5, C5 & T3
- UltraWarm
- Cross cluster search

Easy to use

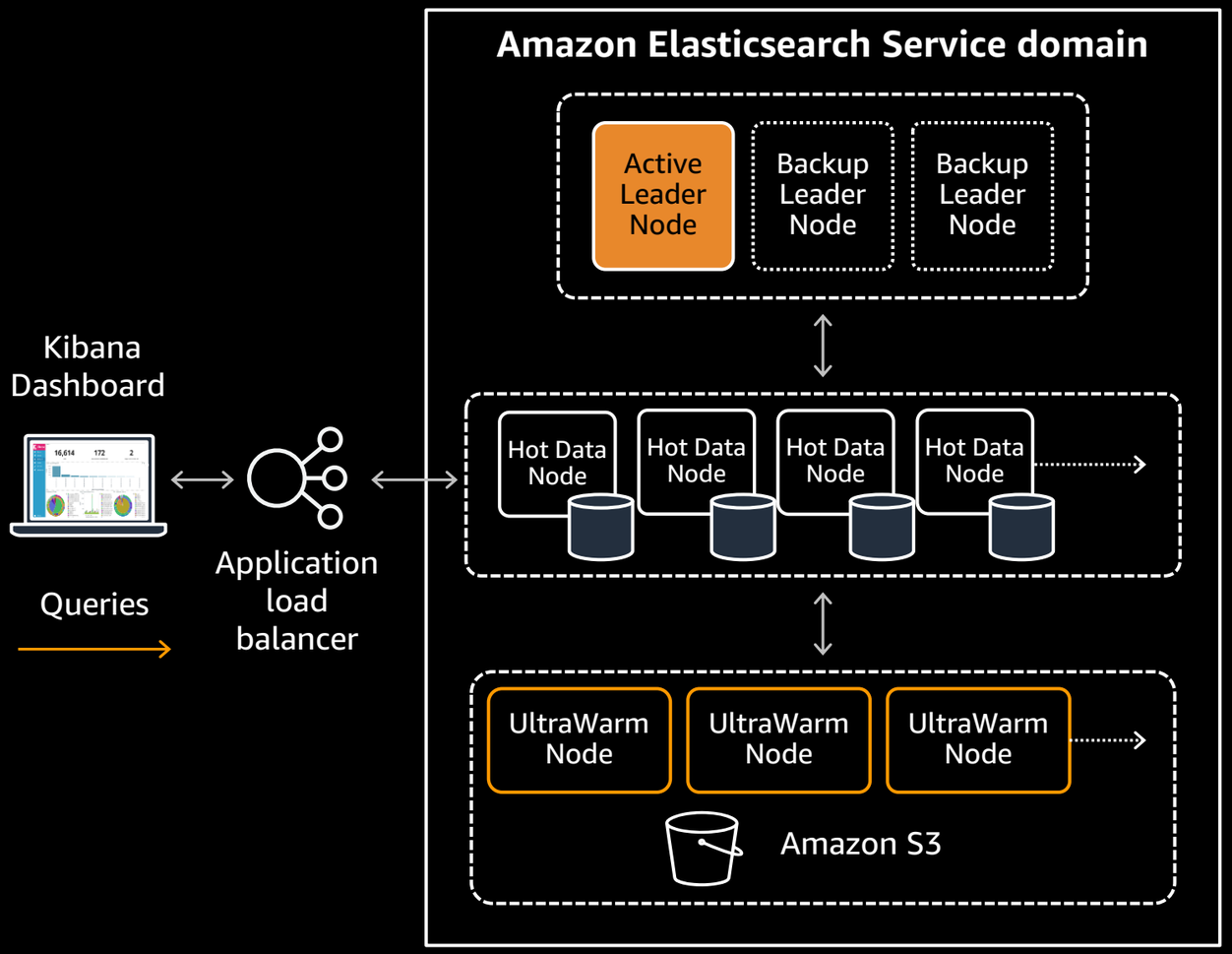
- Alerting
- SQL
- Index state management
- k-Nearest Neighbor Search
- Custom dictionaries
- Anomaly detection
- HTTP compression
- Learn to rank
- Custom domain names
- Tableau Integration

Security and availability

- Enforce HTTPS w/ TLS 1.2
- SOC2 Compliance and FedRAMP
- Fine grained access control
- Multi-tenancy support
- 3 AZ support
- Instance count updates without blue / green
- Hourly snapshots
- SAML Integration
- Audit logging

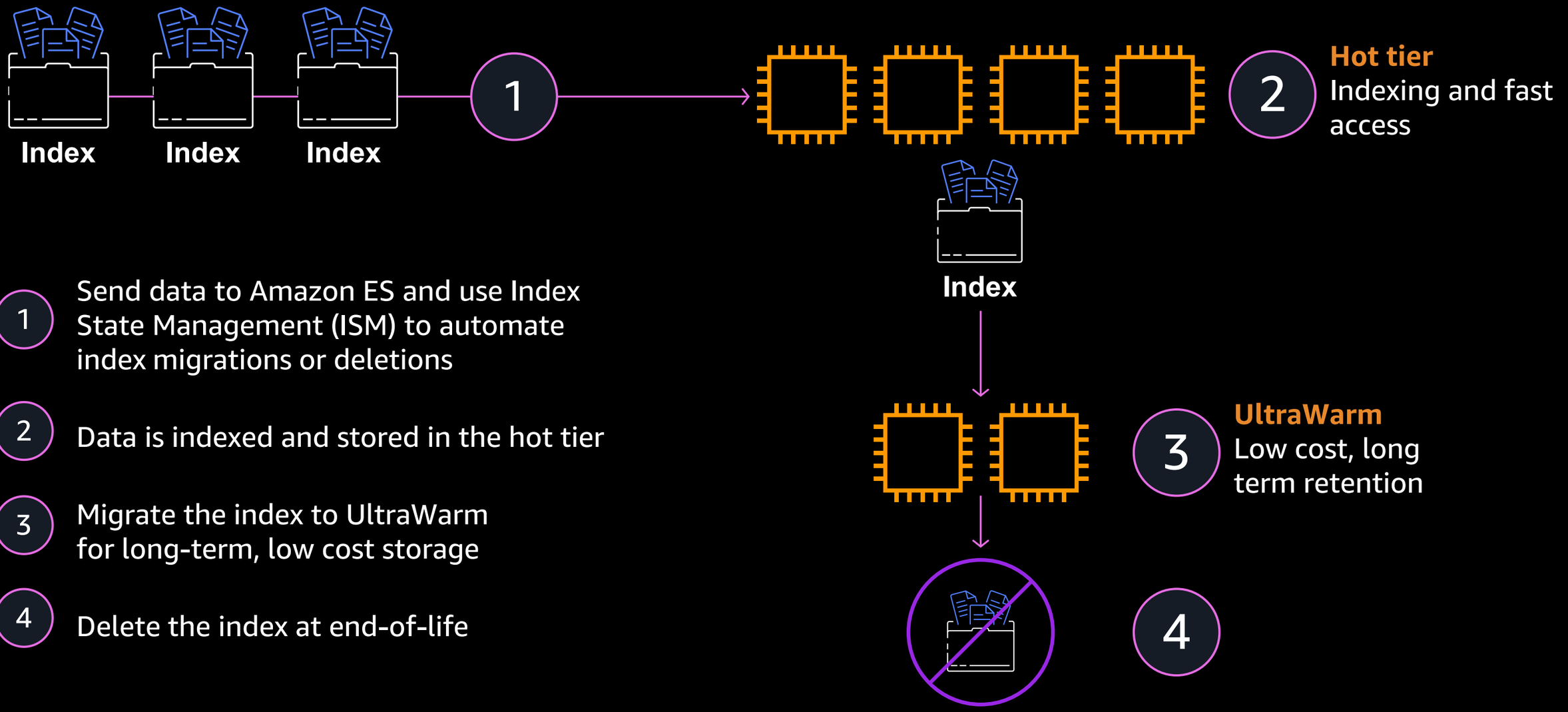
UltraWarm for Amazon Elasticsearch Service

A NEW, LOW-COST STORAGE TIER FOR AMAZON ELASTICSEARCH SERVICE



- Up to 90% cost reduction vs. hot storage tier
- Scale up to 3 PB per domain
- Interactive log analytics and visualization
- 2x performance vs. OS Hot/Warm/Cold architectures

Index State Management (ISM) for data lifecycle



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

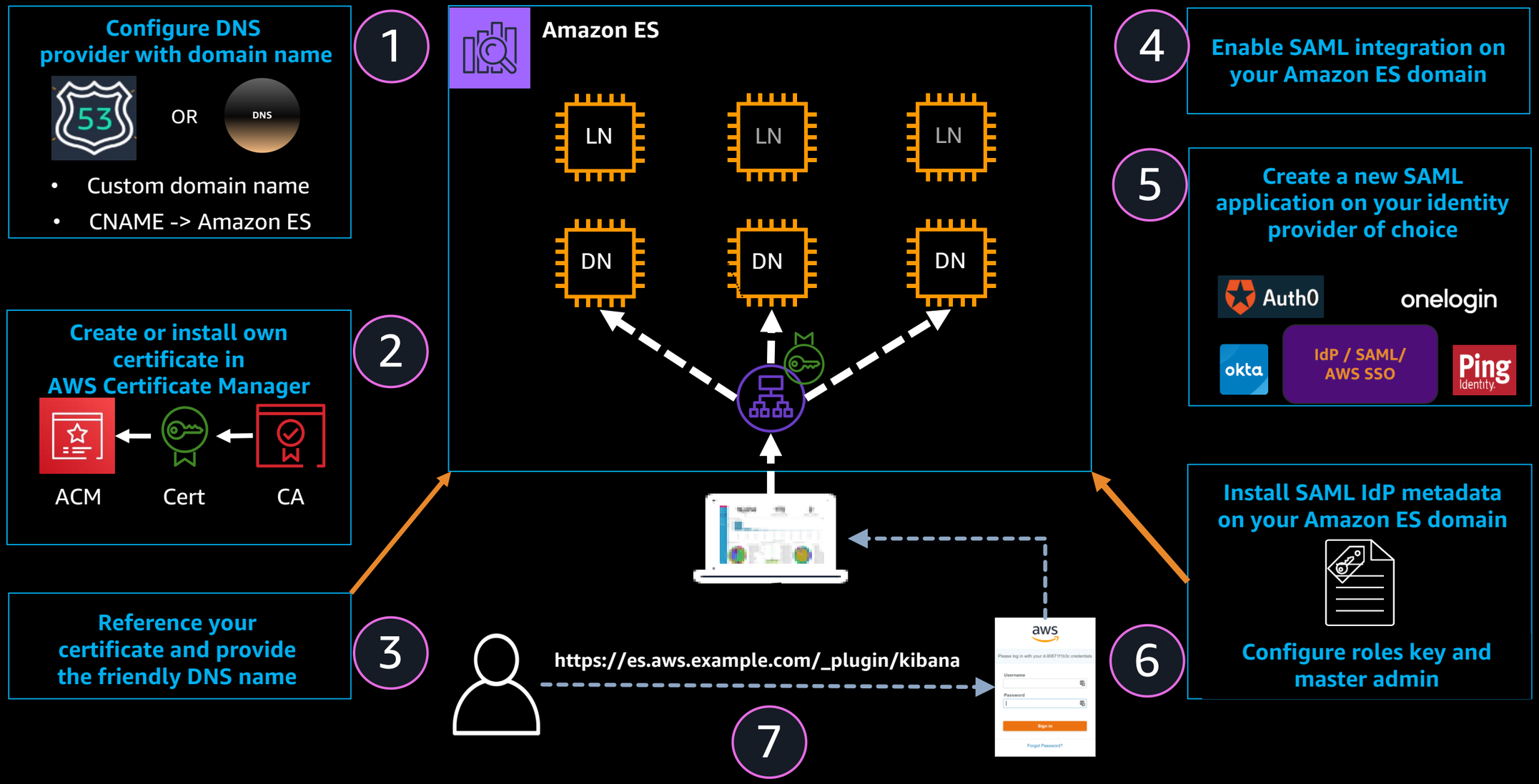
Cross-cluster search for Amazon ES

INCREASE SCALABILITY, EFFICIENCY & AVAILABILITY, BY SEPARATING DISTINCT WORKLOADS



- Single Kibana interface to search across all included domains
- Tune domain resources for specific workloads
- Isolate failures to specific workloads

Custom domain names and SAML integration



K-NN search

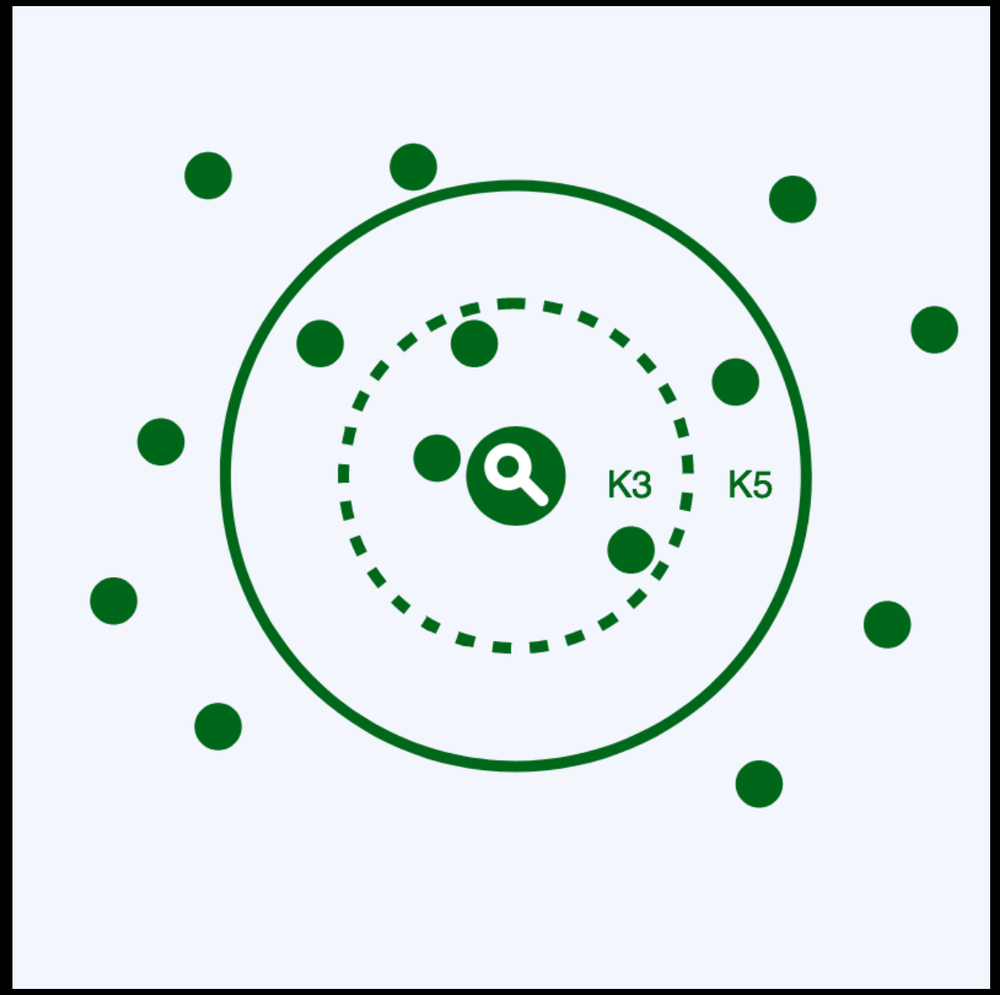
Vector-based search algorithm

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

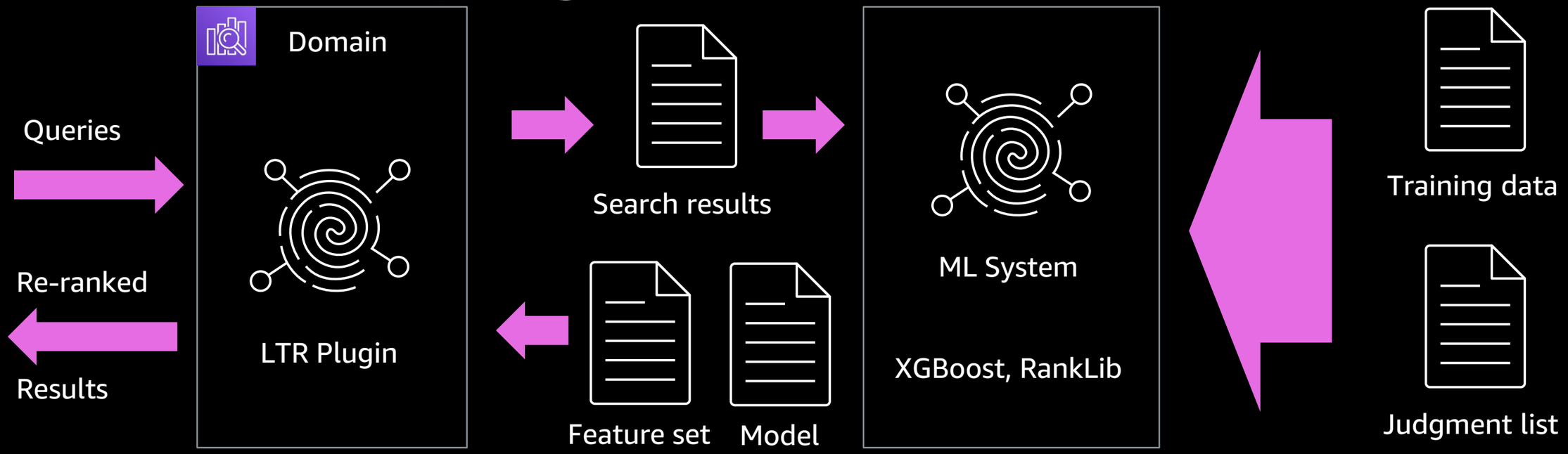


Anomaly Detection in Amazon ES

- Learns what is normal
- Doesn’t require prior knowledge
- Deviations from normal are anomalies
- Works in near real time on time series data
- Uses the Random Cut Forest (RCF) algorithm



Learn to rank plugin



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

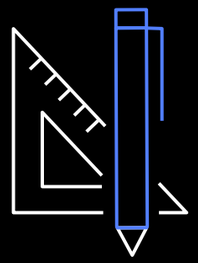
Use features like recency, popularity ratings, and preferences

Feed the model to the LTR plugin in Elasticsearch

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

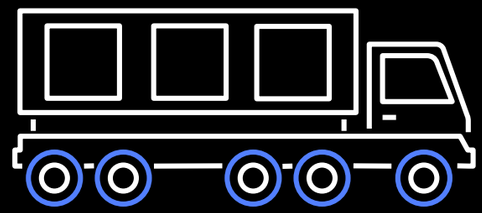
Use in a loop, feeding back search results and click data to improve the model

Advanced querying



Sql WorkBench

Kibana level integration with Query Workbench makes it easy to use Amazon ES's SQL support to work with your data



Piped Processing Language

With PPL you can chain together multiple commands, with one result serving as the input to the next, to explore your data even further

Thank you!

Jon Handler

Principal Solutions Architect, Search Services
AWS
handler@amazon.com

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.