Modernize Your Log Analytics with the Amazon Elasticsearch Service

Kevin Fallis – Senior Search Specialist Solutions Architect

Log analytics involves ***searching, analyzing, and visualizing*** machine data generated by your IT systems and technology infrastructure to gain operational insights.

Where does this data come from?

IoT & Wireless

- Automotive
- Home
- Tools
- Manufacturing
- Mobile applications

IT & DevOps

- Databases
- Load balancers
- Networking
- Deployment tools
- Servers

Applications & Cloud

- Access monitoring
- Environment change
- Web applications
- Business applications
- Container frameworks

aws

Actionable insights come from proper tools

- Most databases cant scale horizontally and have finite resources

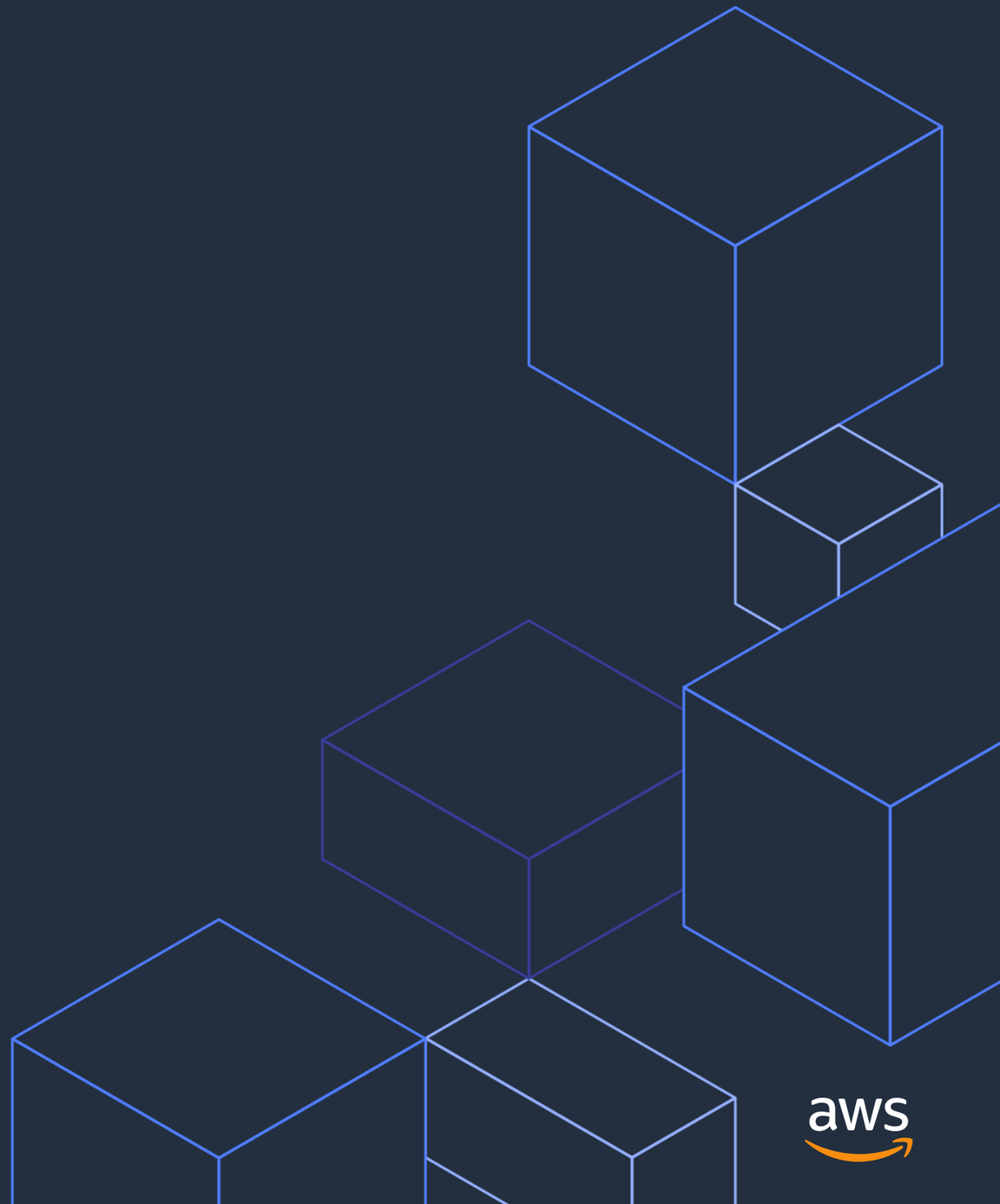
- Data warehouses can scale horizontally but suffer due to lack of indexes

- Manual interrogation of text files bottleneck at the user

aws

Traditional data analytics tools ***are simply not built*** to handle the variety and volume of rapidly proliferating machine and human generated data.

What is Elasticsearch?



Elasticsearch - Distributed search and analytics engine

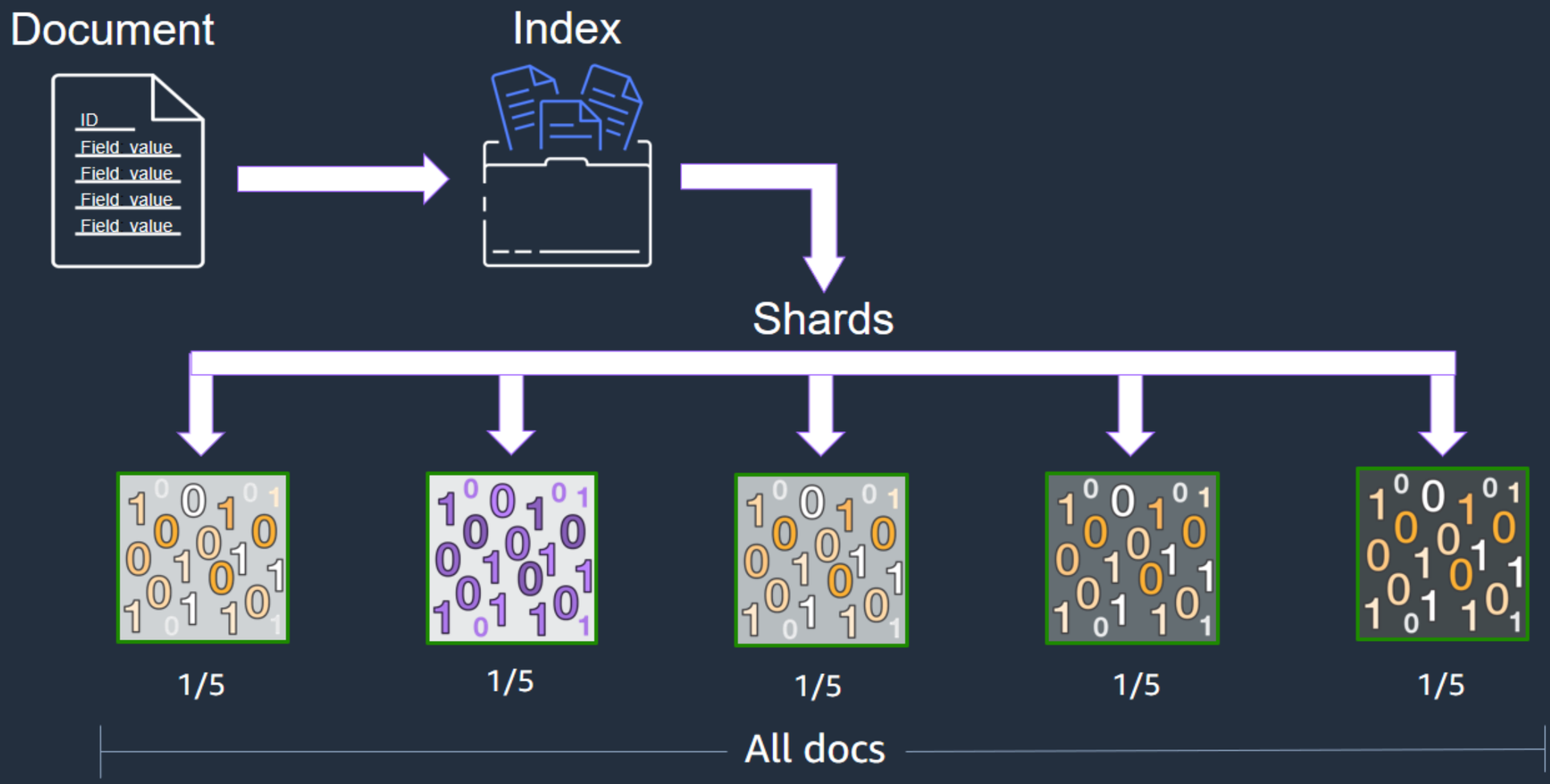
- Open source
- Built on Apache Lucene
- Specializes in full text search
- Scalable & distributed
- Near real-time search engine
- Handles variant structures
- IP and geospatial data types
- Can search every field in a document
- Supports multiple criteria

Rank			DBMS	Database Model	Score		
May 2020	Apr 2020	May 2019			May 2020	Apr 2020	May 2019
1.	1.	1.	Oracle	Relational, Multi-model	1345.44	+0.02	+59.89
2.	2.	2.	MySQL	Relational, Multi-model	1282.64	+14.29	+63.67
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	1078.30	-5.12	+6.12
4.	4.	4.	PostgreSQL	Relational, Multi-model	514.80	+4.95	+35.91
5.	5.	5.	MongoDB	Document, Multi-model	438.99	+0.57	+30.92
6.	6.	6.	IBM Db2	Relational, Multi-model	162.64	-2.99	-11.80
7.	7.	7.	Elasticsearch	Search engine, Multi-model	149.13	+0.22	+0.51
8.	8.	8.	Redis	Key-value, Multi-model	143.48	-1.33	-4.93
9.	9.	↑ 11.	SQLite	Relational	123.03	+0.84	+0.14
10.	10.	↓ 9.	Microsoft Access	Relational	119.90	-2.02	-23.88
11.	11.	↓ 10.	Cassandra	Wide column	119.16	-0.91	-6.57
12.	12.	12.	MariaDB	Relational, Multi-model	90.09	+0.19	+3.57
13.	13.	13.	Splunk	Search engine	87.75	-0.33	+2.51
14.	14.	14.	Hive	Relational	81.54	-2.51	+3.64
15.	15.	15.	Teradata	Relational, Multi-model	73.89	-2.70	-2.15

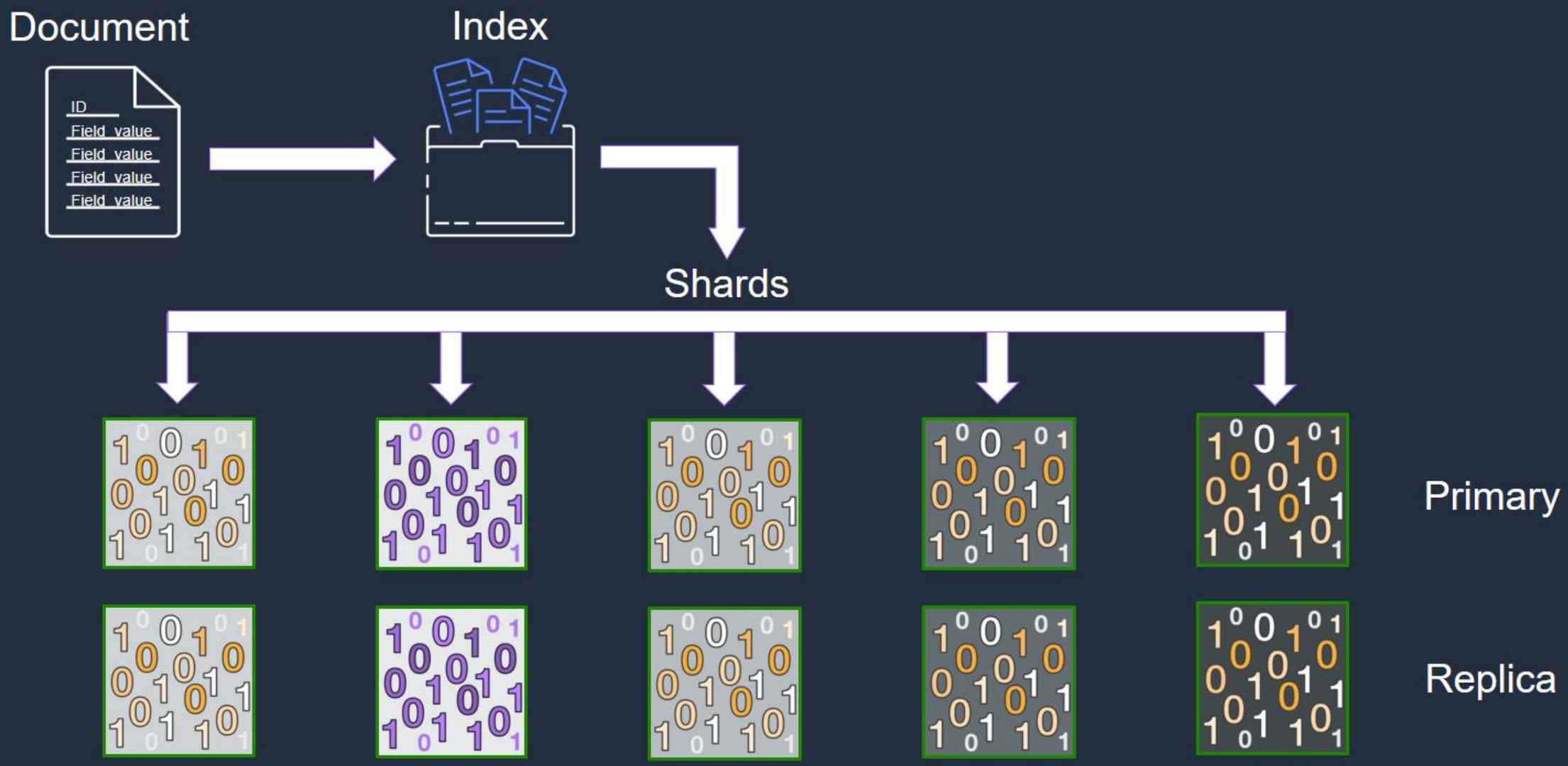
https://db-engines.com/en/ranking

aws

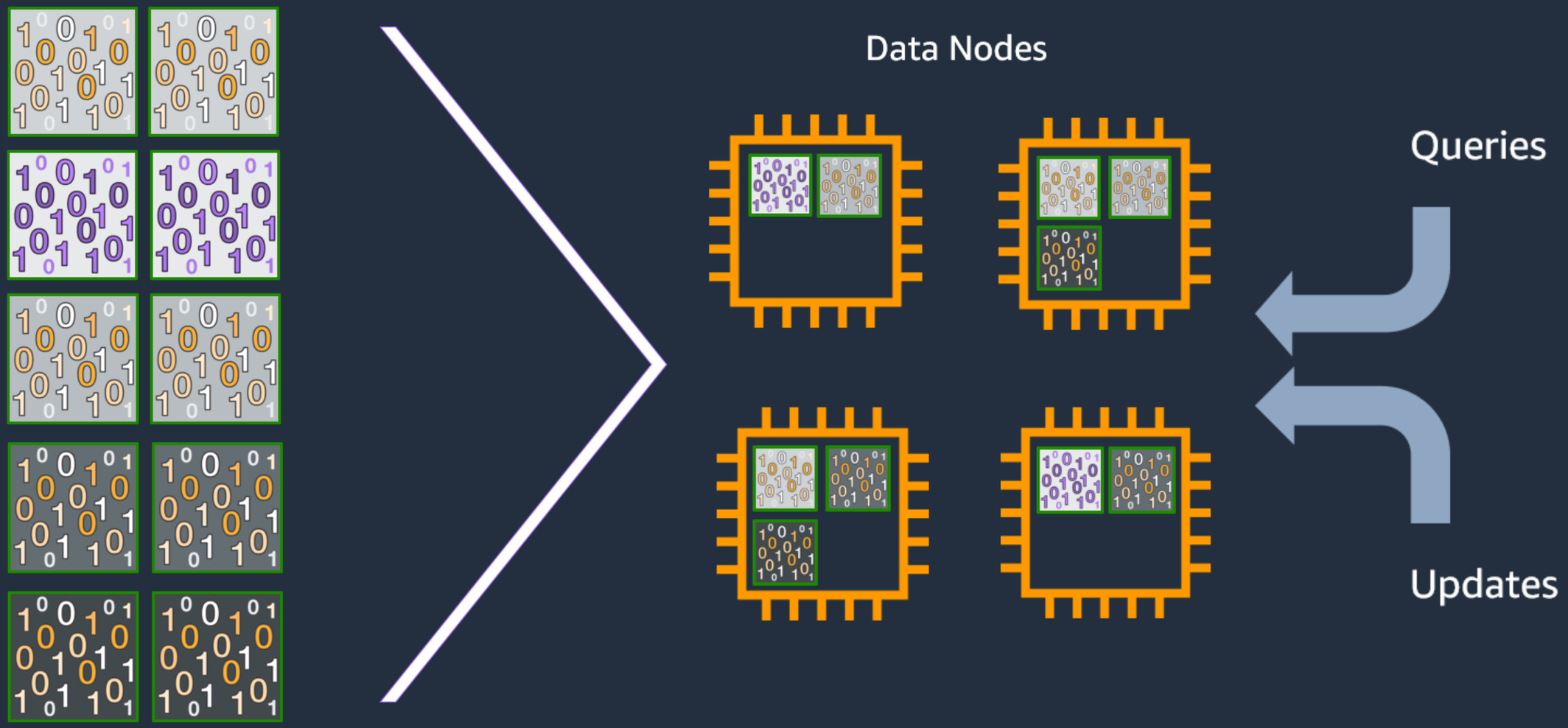
Elasticsearch – how data is organized



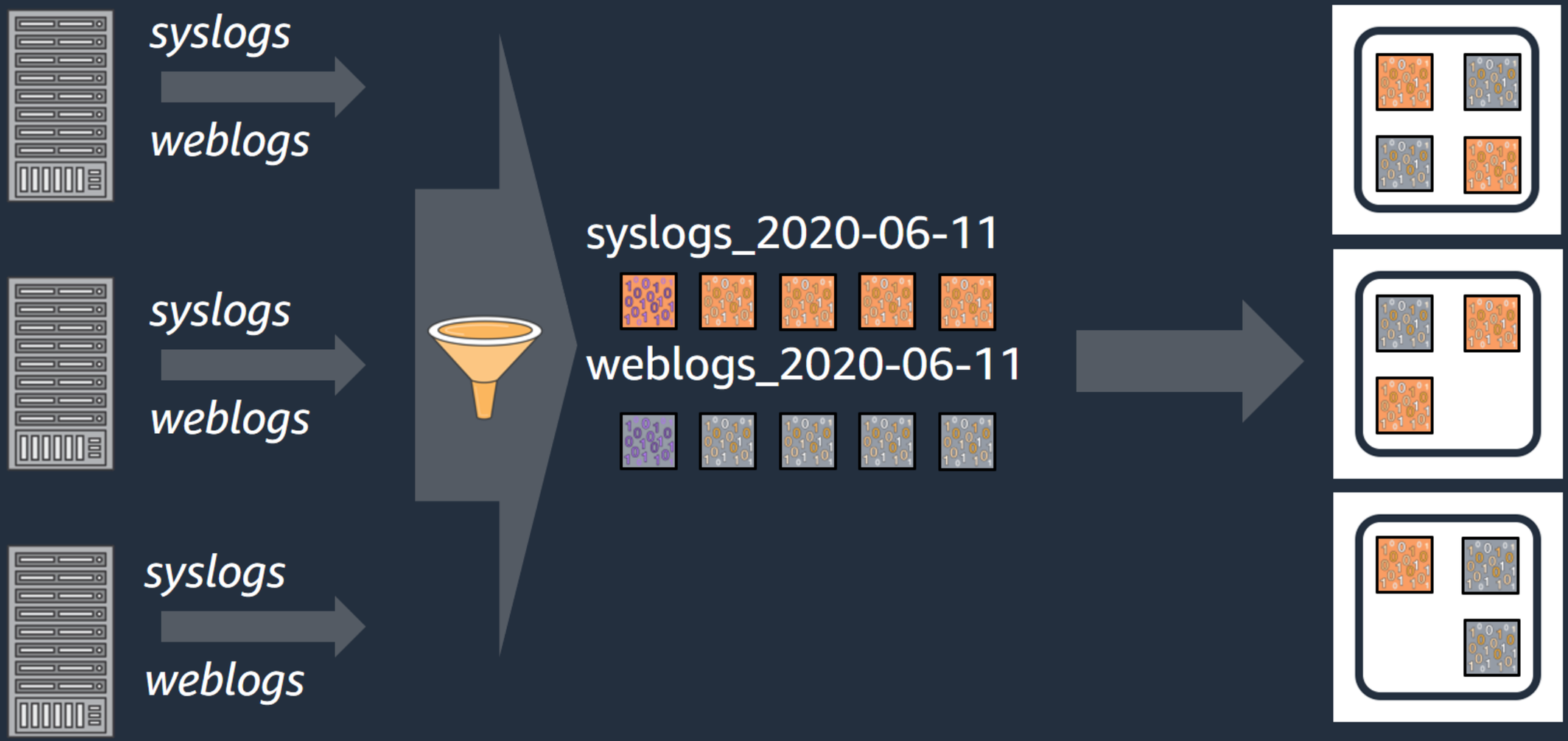
Elasticsearch – durability and performance levers



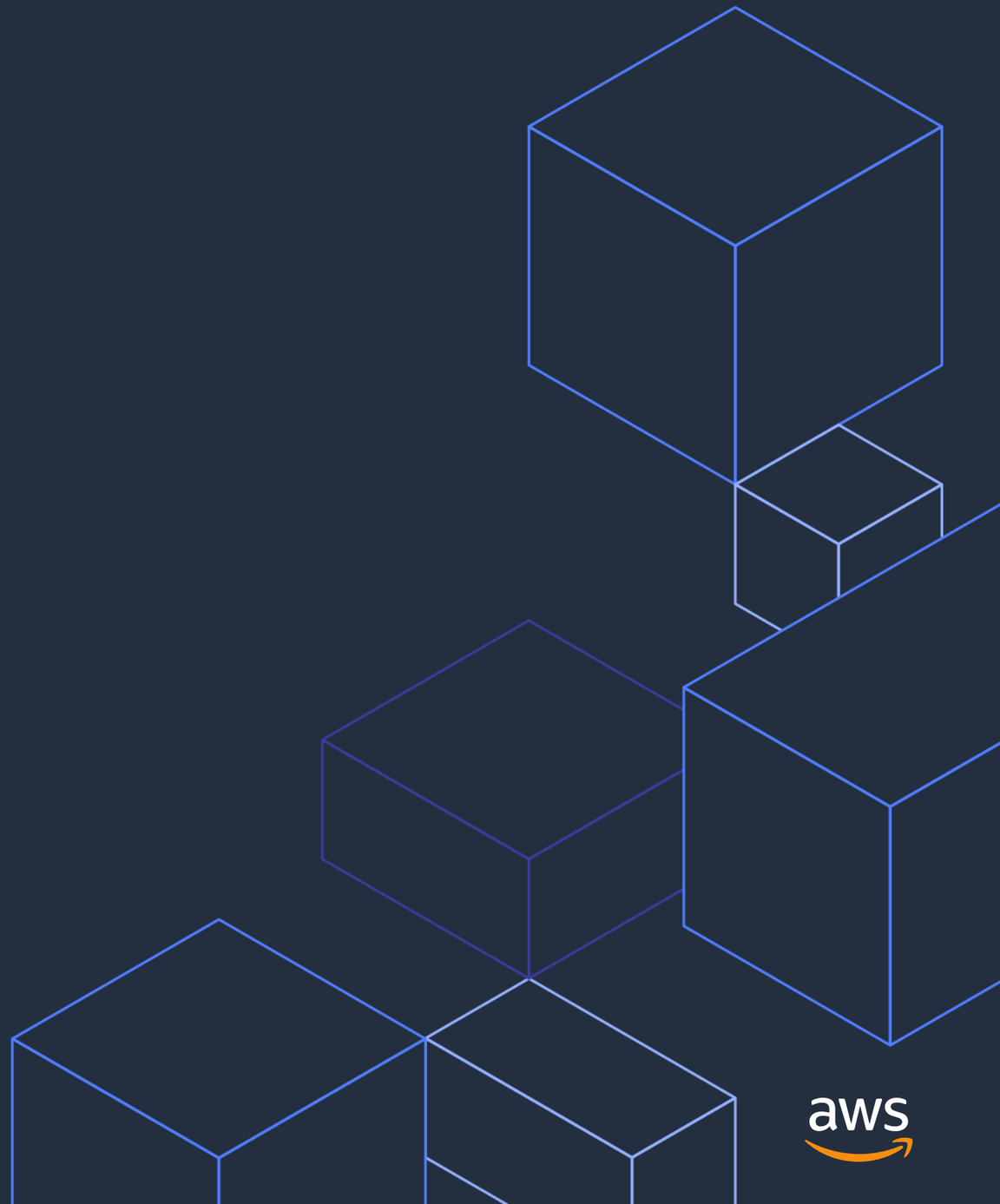
Elasticsearch – compute distribution and scaling



Elasticsearch – handles multiple concurrent streams



What is Kibana?



Kibana – visualizes your Elasticsearch data

- Open source
- Rich visualizations
- Dashboards
- Plugin based ecosystem
- Query and work with your data in real time
- Tools for developers
- Line charts
- Pie charts
- Scatter plots
- Heatmaps
- And more

Elasticsearch and Kibana – putting it together

Kibana

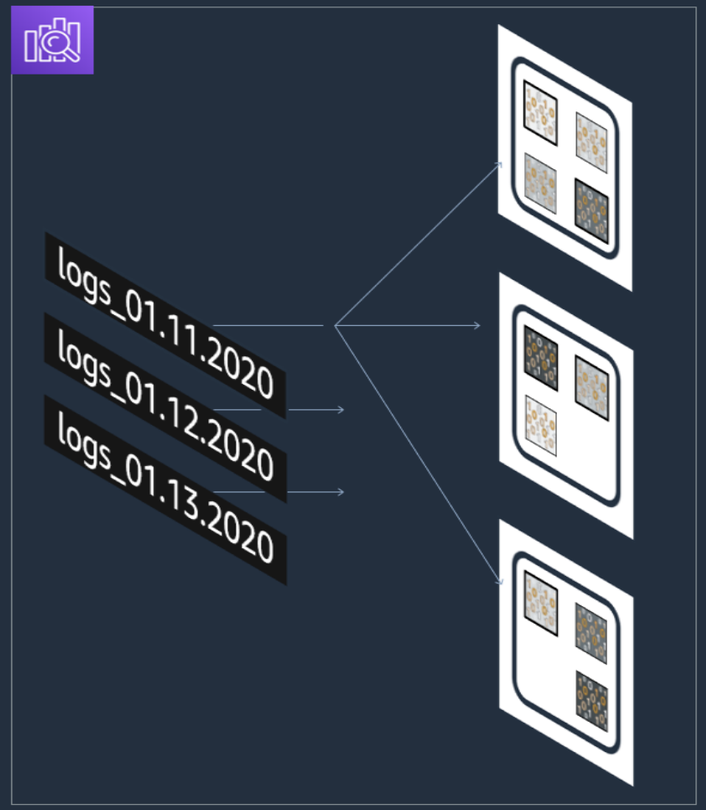
Create visualizations

Query DSL

```
"query": {
  "bool": {
    "must": [ { "match_all": {} },
      {
        "range": {
          "@timestamp": {
            "gte": 1555340004616,
            "lte": 1555340904616,
            "format": "epoch_millis"
          }
        }
      }
    ]
  }
}
"aggs": {
  "2": {
    "terms": {
      "field": "eventSource.keyword",
      "size": 5,
      "order": {
        "_count": "desc"
      }
    }
  }
}
```

Generate query

Elasticsearch



Process query

aws

Elasticsearch and Kibana – perfect for log analytics

Text search

Natural language

Boolean queries

Relevance

Streaming data

High-volume ingest

Near real time

Distributed storage

Rich analysis

Time-based visualizations

Nest-able statistics

Time series tools

aws

Amazon Elasticsearch Service

What is the Amazon Elasticsearch Service (Amazon ES)?

Amazon ES is a fully managed service that makes it easy to deploy, operate, and scale Elasticsearch clusters with Kibana securely and cost-effectively in the AWS Cloud.

aws

Amazon ES – capabilities and benefits

- fully managed, provisions all the resources for your Elasticsearch cluster in a secure and compliant environment and launches it in any Region of your choice within minutes

- deploys Kibana and offers direct access to the Elasticsearch APIs, which makes your existing code and applications using Elasticsearch work seamlessly with the service

- cost effective with pay as you go pricing with features that are free of any licensing fees and cost savings options for long term retention of valuable data

- can scale your cluster horizontally or vertically, up to 3 PB of data, with zero downtime through a single API call or a few clicks on the AWS Management Console

- HIPAA eligible with PCI DSS, SOC, FedRamp and ISO compliances

aws

Centralized log analytics security monitoring and alerting system

The challenge

- Needle in a haystack problem, need to find meaningful events in billions of documents
- Optimize engineering resources
- Time series data requiring efficient aggregations in real time
- User friendly interface

The solution

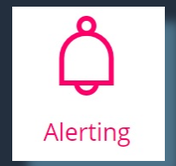
- Ingest thousands of endpoints into Amazon ES
- Real-time visualizations using Kibana
- Full text search for ad-hoc queries
- Build visual drilldowns for threats
    - Focus on the event window

"get up and running in minutes"

Dennis Griffin, Senior Director of Engineering, Managed Threat Response

Amazon ES plugin ecosystem

Recent plugin review – Amazon ES



- Security – fine grained access control down to document and field level – granular entitlements
- Alerting – look for patterns in your data and send automated notifications via popular services
- SQL – query Elasticsearch using structured query language known to relational database users
- k-NN – k-nearest neighbors algorithm for similarity matching in vectors
- Index Management – perform schedule maintenance on your indexes without the need of scheduled AWS Lambda

aws

Open source innovation leveraged by Amazon ES

Open Distro for Elasticsearch

100% open source Providing you the freedoms, so you can freely view, use, change, and distribute the code

Enterprise-grade
Delivering security and advanced capabilities such as alerting, SQL, and cluster diagnostics

Community-driven
Providing individuals and organizations the freedom to easily contribute changes to the distro

aws

UltraWarm for cost savings

UltraWarm – a new storage tier for Amazon ES

Store massive amounts of data

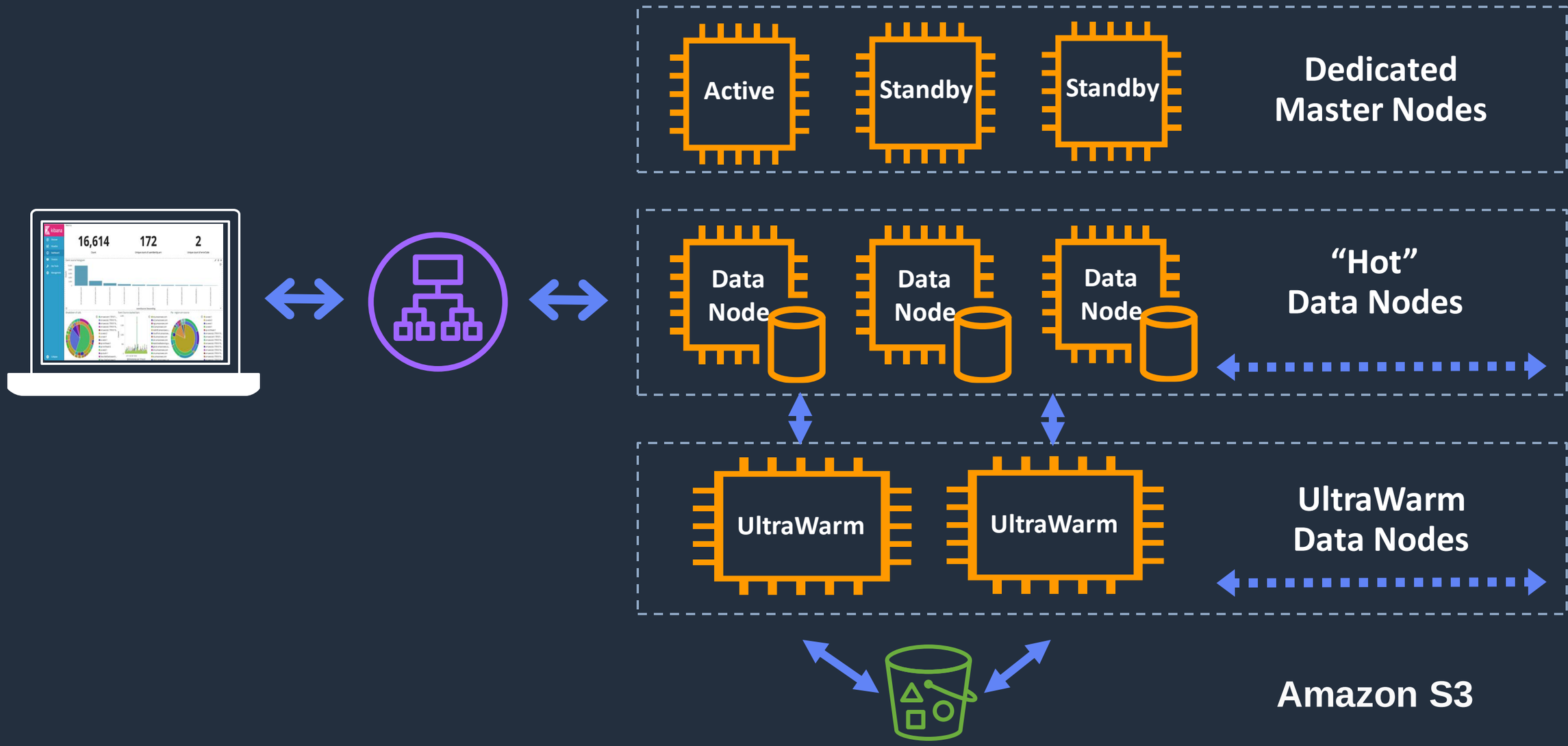
Run interactive log analytics and visualizations

Higher performance and durability

Achieve up to 90% cost savings

aws

UltraWarm architecture



Centralized log analytics platform

The challenge

- Consolidate security, application, and web logs
- Must be secure, may contain PII
- Small operational team
- Existing proprietary solution too expensive

The solution

- Amazon Kinesis Data Streams coupled with AWS Lambda deliver data to Amazon ES
- Kibana for real time visualization and dashboards
- Tested for 100,000 log lines a second with 500TB on a single cluster
- Was a “fraction of the cost” of their proprietary log analytics solution

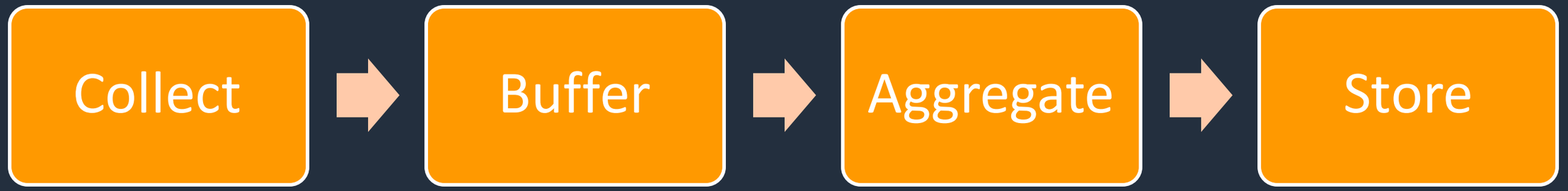
”My job is not to manage Elasticsearch”

“Benefit of a centralized logging solution without having to buy into a whole big ecosystem”

Josh Pavel, Senior DevSecOps Engineer, Pearson

Delivering data to Amazon ES

Amazon ES – Ingestion workflow



Open source log delivery ecosystem

aws

AWS native and managed service ingestion ecosystem

Amazon
CloudWatch agent

Amazon
CloudWatch

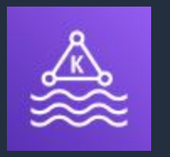
Amazon
Simple Storage
Service

Amazon
Simple Queue
Service

Amazon
Kinesis agent

Amazon
Kinesis Data
Streams

Amazon
Kinesis Data
Firehose



Amazon
Managed Streaming
for Kafka

AWS IoT

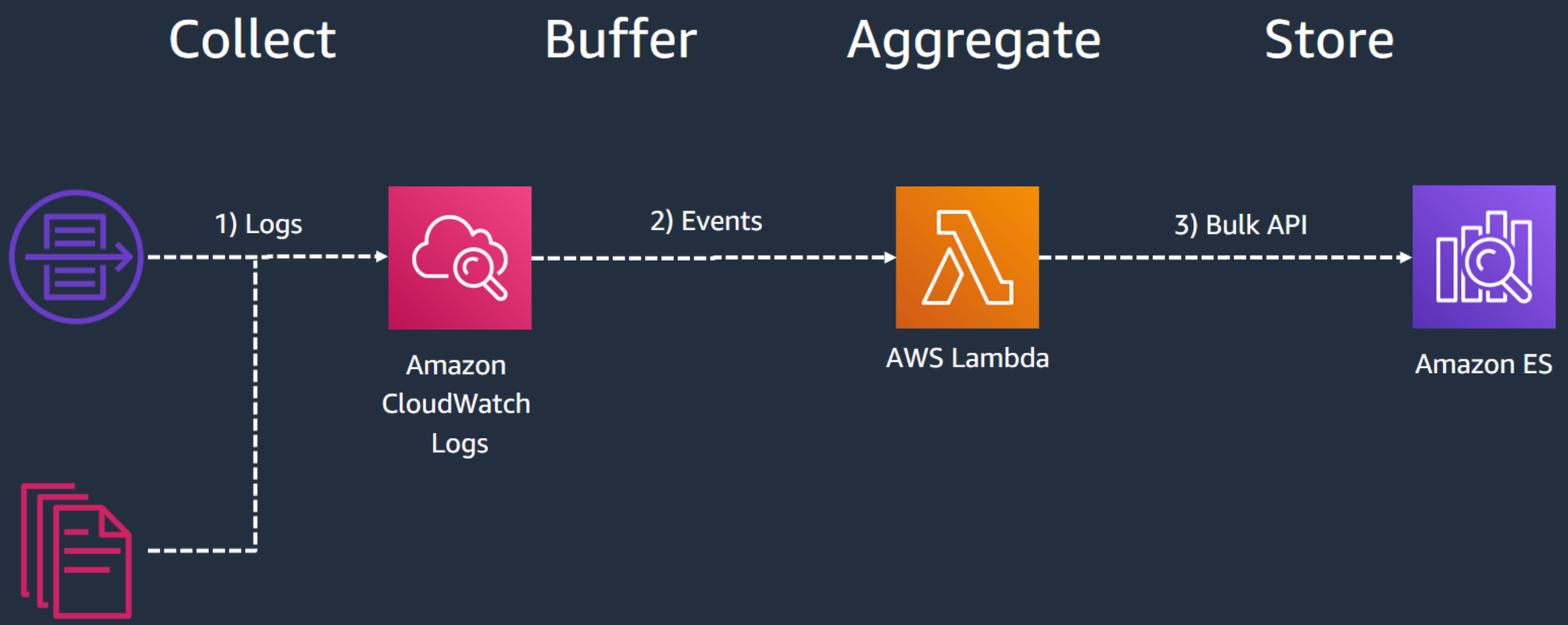
Amazon ElastiCache
for Redis

AWS Lambda

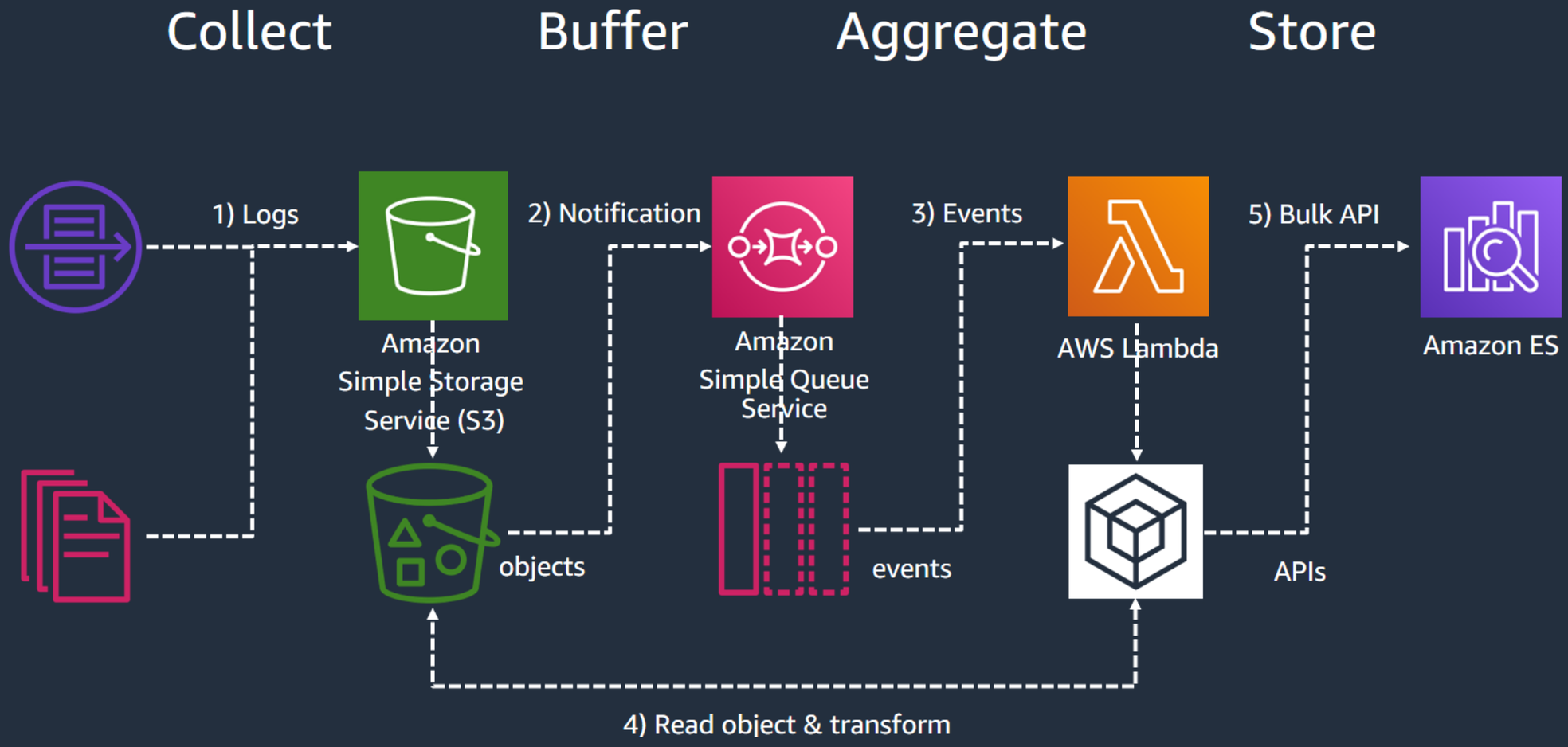
aws

Popular ingestion patterns

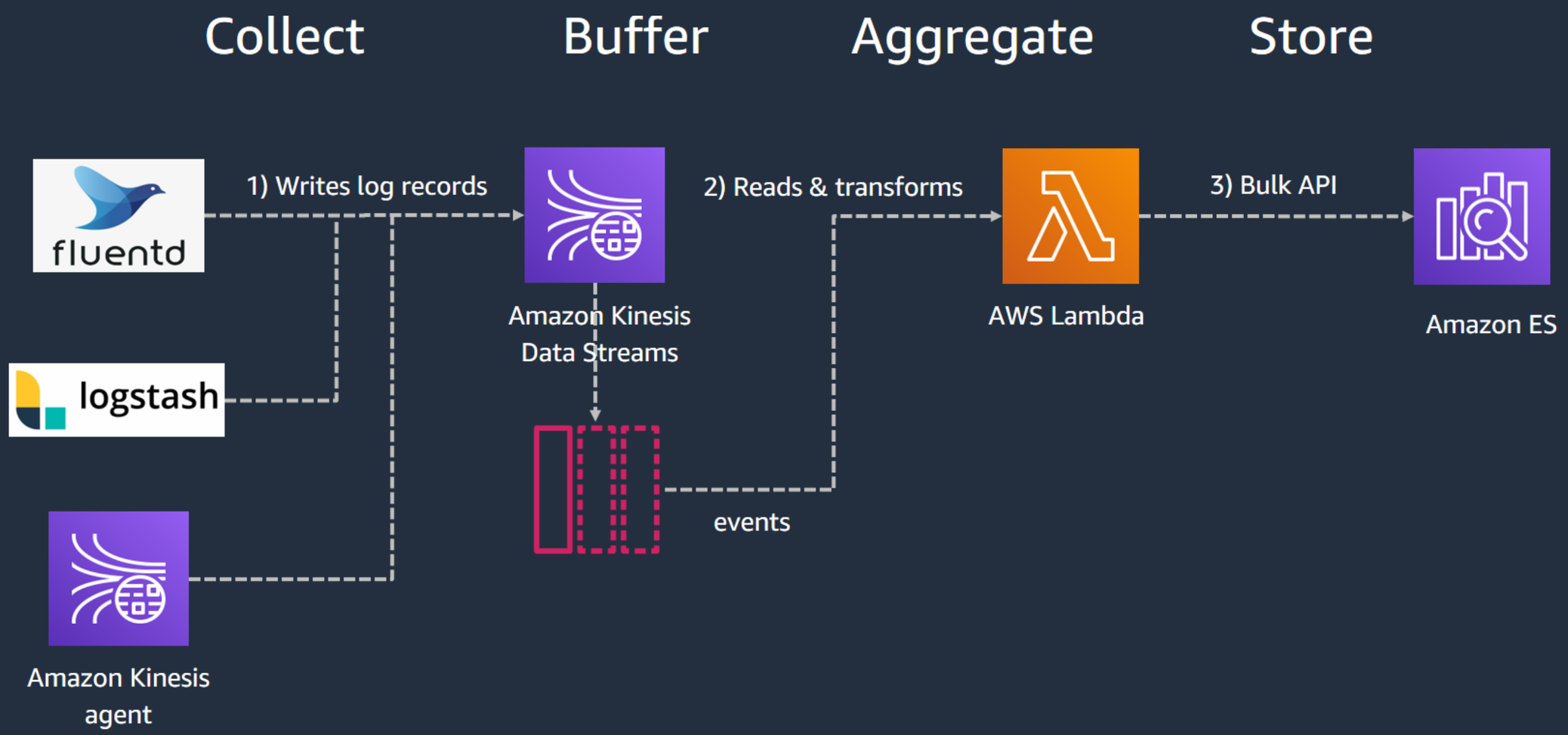
Amazon CloudWatch Logs pattern



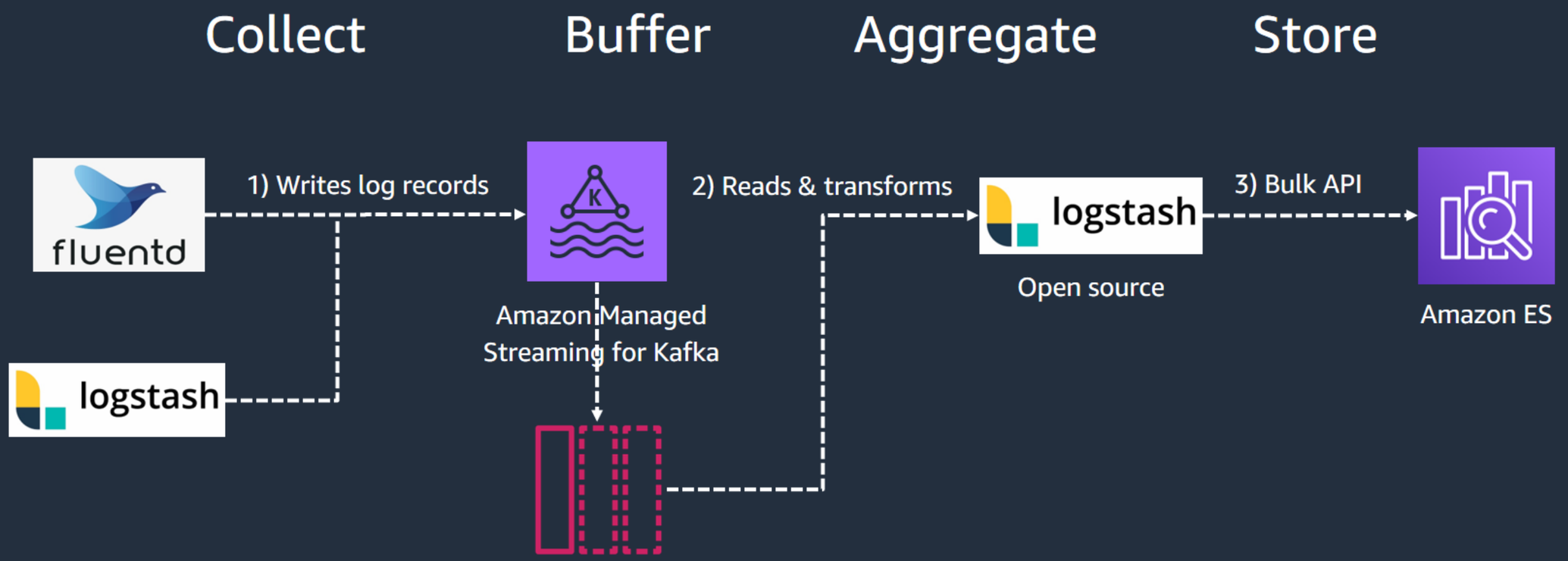
Amazon Simple Storage Service pattern



Amazon Kinesis Data Streams and AWS Lambda pattern



Amazon Managed Streaming for Kafka pattern

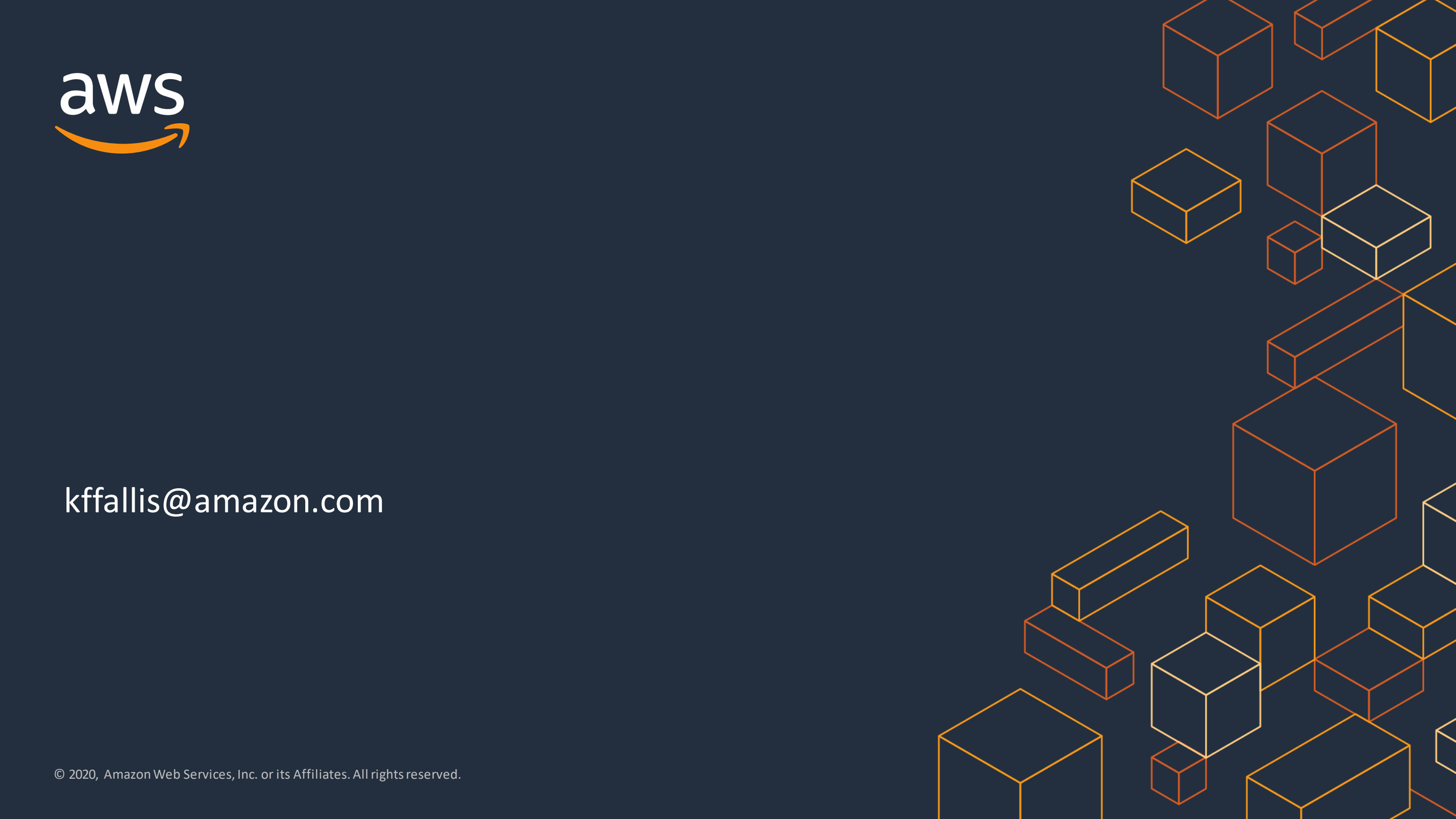


Demo – VPC Flow Logs

Q&A

Kevin Fallis

Senior Search Specialist Solutions Architect



kffallis@amazon.com