

The background of the image features a dark blue gradient on the left, transitioning into a large, vibrant, abstract shape on the right. This shape is composed of overlapping curved segments in shades of orange, pink, and purple, creating a dynamic, modern aesthetic.

AWS re:Invent

NOV. 27 – DEC. 1, 2023 | LAS VEGAS, NV

DAT339

Advanced integration patterns with Amazon DynamoDB

Jon Handler

he/him

Senior Principal Solutions Architect,
OpenSearch
AWS

Jason Hunter

he/him

Principal Solutions Architect,
DynamoDB
AWS



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon DynamoDB

FAST, FULLY MANAGED & SERVERLESS KEY-VALUE DATABASE



Performance at Scale

Consistent latency at any scale
Unlimited throughput
Unlimited storage



Secure & Reliable

Data encryption
Highly available (99.999% SLA)
Active-Active global replication
Highly durable storage
Continuous backups



Serverless

Nothing to provision*
Automated scaling with no availability impact
Pay for what you use
No maintenance windows or version updates

*using On-Demand Capacity



Built-in Integration with AWS Services

AWS Lambda
AWS Identity and Access Management (IAM)
Amazon CloudWatch
Amazon Kinesis Data Streams
Amazon S3
Amazon Cognito
AWS Backup
AWS CloudTrail
AWS Step Functions
& **more**



Example Data — Product Questions and Answers

- Amazon Standard Identification Number (ASIN) of the product
- A unique question_id
- Question text
- Answer(s)
- Product details
- And then we've simulated metadata about the answerer: a rating, name, age, gender, and location

<https://registry.opendata.aws/amazon-pqa/>

```
{
  "asin": "B003YFHQI2",
  "question_id": "Tx3UX24HTBF20DP",
  "brand_name": "Belle Epoque",
  "item_name": "Belle Epoque 420 Full Thread Count Sheet Set with Hemstitch, Taupe",
  "question_text": "are these sheets 100% cotton?",
  "answers": [
    {
      "answer_text": "Yes, but the cotton is a bit rough. I've washed the sheets about six times and they are finally softening."
    },
    {
      "answer_text": "Yes!"
    }
  ],
  "bullet_point1": "Incredible luxury: Unbelievable soft hand comes from Belle Epoque processing formula combined with highest quality compact combed cotton yarns",
  "bullet_point2": "Soft mirror finish, Special processing ensures that these sheets will be satin-soft and pill-free after many wash and dry cycles",
  "bullet_point3": "Oversized sheets, Flat sheet: 1 foot longer than standard flat sheet with 6-inch hem at the top and 1-inch mitered hems on the sides and foot of the sheet, 6-inch hem features meticulous hemstitch for added beauty, Extra length provides additional turn",
  "bullet_point4": "Fitted sheet: Deep pocket fits up to a 20-inch mattress, Elastic is sewn in all around for secure fit",
  "bullet_point5": "Pillowcases: Oversized with 6-Inch cuff and hemstitch detailing to match flat sheet Full Sheet Set",
  "product_description": "",
  "question_type": "yes-no",
  "answer_aggregated": "yes"
}
```

DynamoDB access patterns

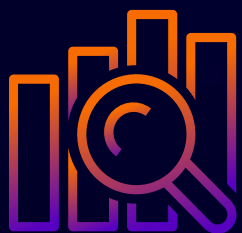
1. Retrieve all questions and answers for an ASIN
2. Add a question to an ASIN
3. Add an answer to a question

DynamoDB table layout

Primary key		Attributes						
Partition key: pk	Sort key: sk							
B07BFVT2PN	metadata	item_name						
		EMONIA Queen Sheets Set -6 Pieces Bed Sheets- Microfiber Super Soft 1800 Series						
	question#Tx 1E43I11RHT WG8#Q	asin	question_id	question_text				
		B07BFVT2PN	Tx1E43I11RH TWG8	Are these sheets noisy?				
	question#Tx 1E43I11RHT WG8#answe r00001	asin	question_text	answer_text	gender	age	user_lat	user_lon
		B07BFVT2PN	Are these sheets noisy?	"Hi, Mine have no noise at all. I wash them and use Downy before I put them on the bed with no noise! Hope that helps"	male	97	-17.182330 5	80.307699

Additional access patterns

1. Find questions or answers containing certain words or phrases
2. Sort the results using a relevance algorithm
3. Generate a histogram chart of search results
4. Identify the most frequent answerer for certain topics
5. Limit results to answers from certain age ranges, genders
6. Limit results to those from a geographic area
7. Match using semantic meaning, not just words and phrases



Amazon OpenSearch Service

Intelligent search and log analytics
solution to help you get the most out of
your data



Search: Improve the relevancy of your search results in near real time with cutting edge search innovations



Analytics: Securely, easily, and efficiently visualize and analyze your operational data



Vector Database: Ingest, store, and query vectors. Hybrid search and efficient filtering



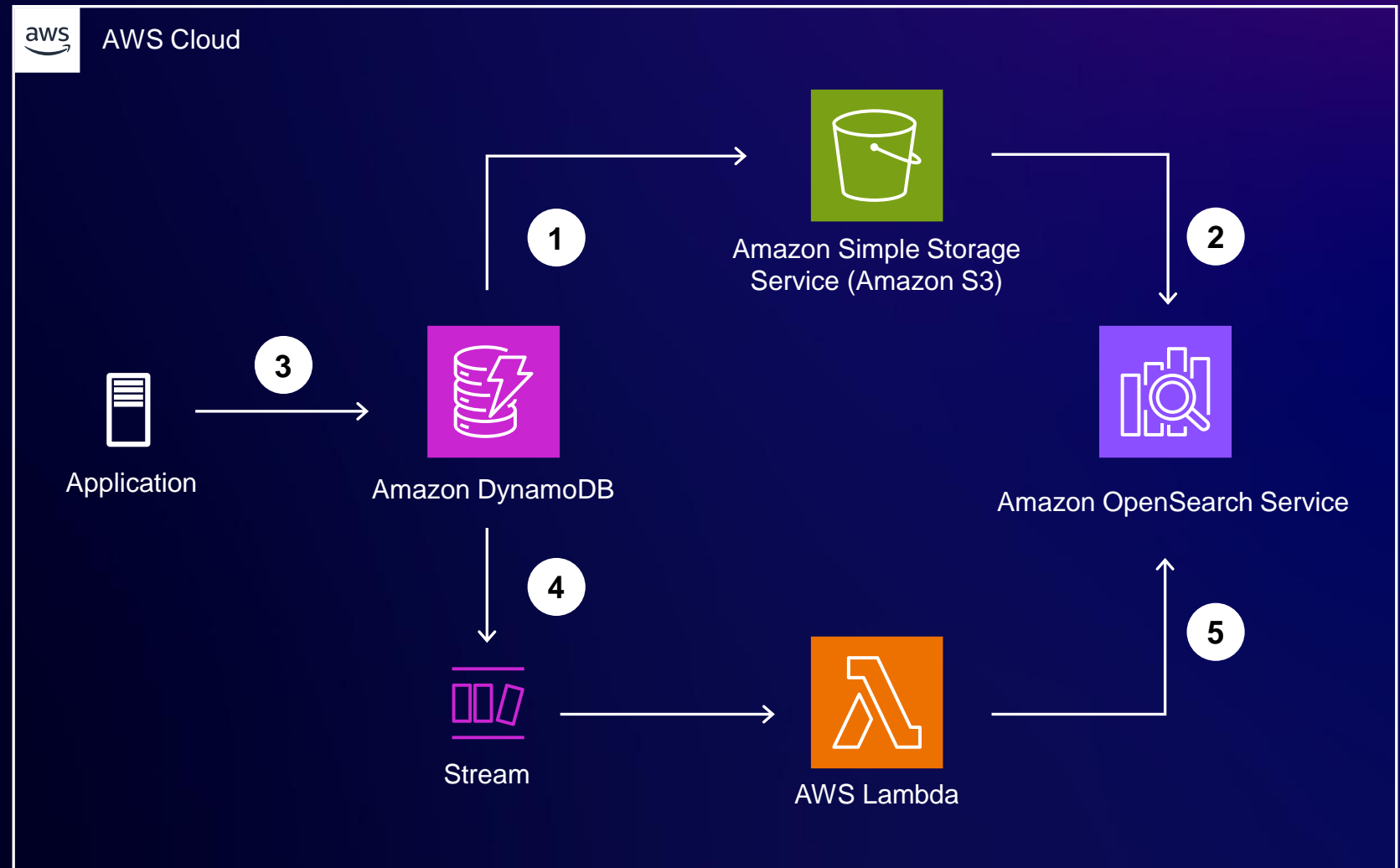
Cost effective: Optimize time and resources for strategic work



Deployment options: Serverless simplicity or managed control

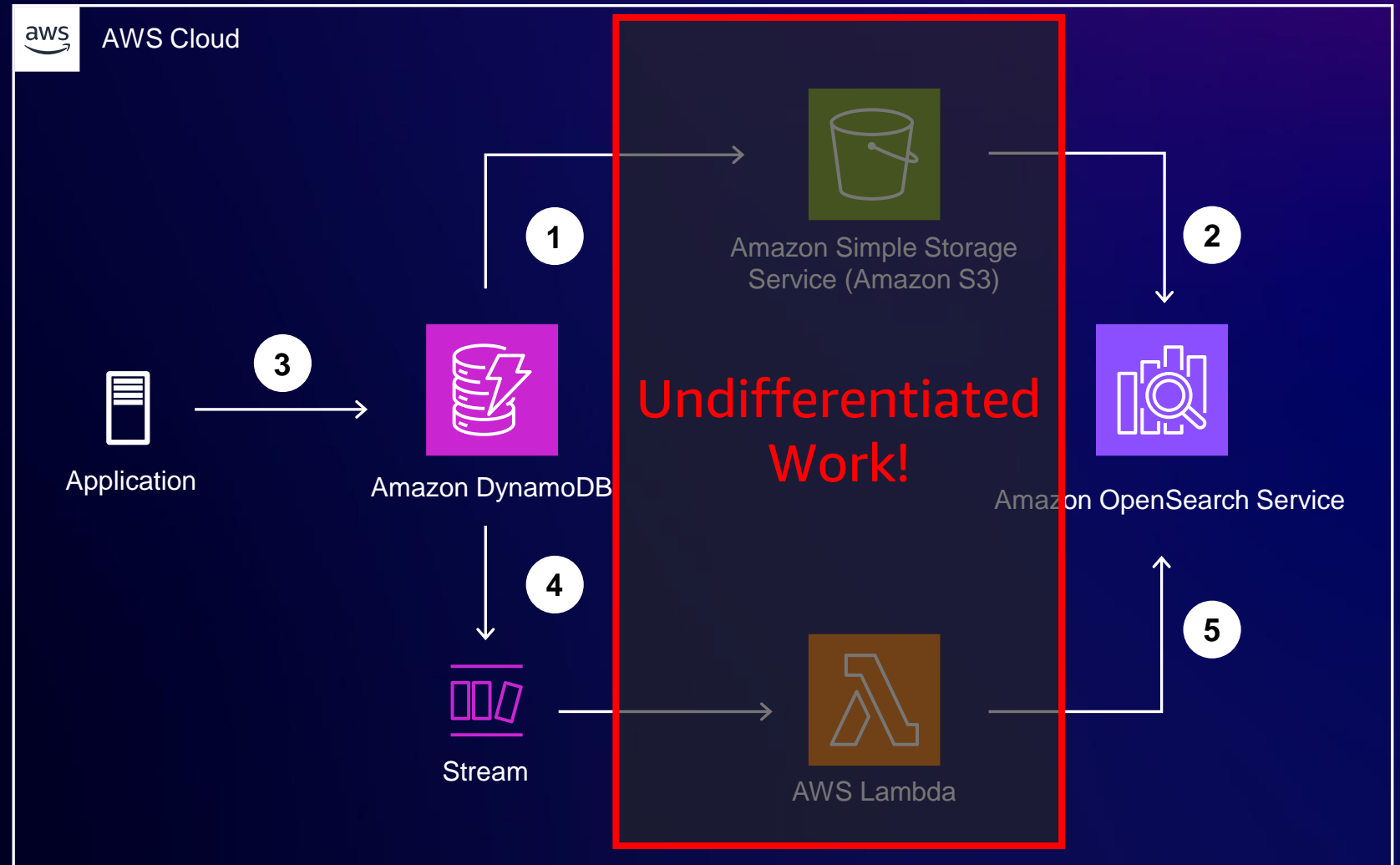
Data replication pipeline with near real time

1. Export a snapshot to S3 with ExportToPointInTime
2. Ingest the snapshot from S3 to the OpenSearch Service
3. Application continues to send updates to DynamoDB while 1 & 2 are happening
4. Updates to DynamoDB appear on the Stream
5. Process the stream with AWS Lambda and push into OpenSearch



Data replication pipeline with near real time

1. Export a snapshot to S3 with ExportToPointInTime
2. Ingest the snapshot from S3 to the OpenSearch Service
3. Application continues to send updates to DynamoDB while 1 & 2 are happening
4. Updates to DynamoDB appear on the Stream
5. Process the stream with AWS Lambda and push into OpenSearch



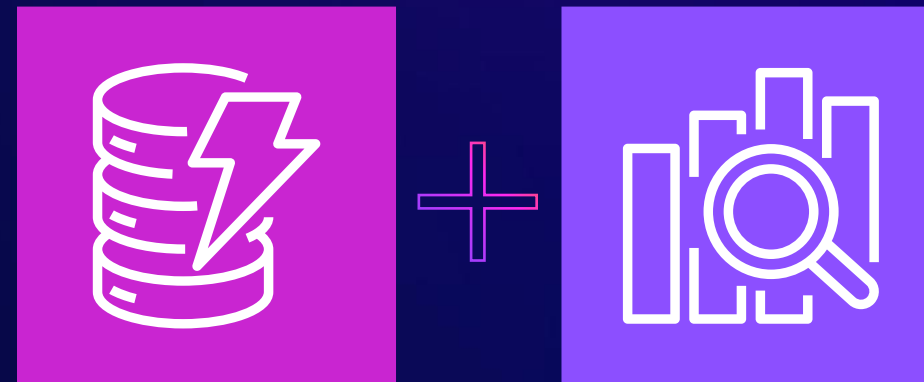
Announcing! Amazon DynamoDB Zero ETL integration with Amazon OpenSearch Service

Near real time – Bootstraps an S3 export then replicates data with Amazon DynamoDB Streams within seconds of changes

Autoscaling – Automatically scales to the demands of your application

No code required – Routing, mapping, & transforms are defined via configuration

Powerful – Built on Amazon OpenSearch Ingestion



OpenSearch Ingestion is powered by



Serverless

Easy to administer with console or APIs. Automatic application of patches and security upgrades



Secure

Redact and obfuscate sensitive information. Route data for compliance



Cost Efficient

Automatically scales to fit workload demands



Open Source

Part of the Apache 2.0-licensed OpenSearch project

Data Prepper has a rich set of Processors

Manipulate data values

Rename keys, combine data into new entries

Routing

Route different data into different indexes

Sub-pipelines

Send data through multiple pipelines for tight control

Aggregate

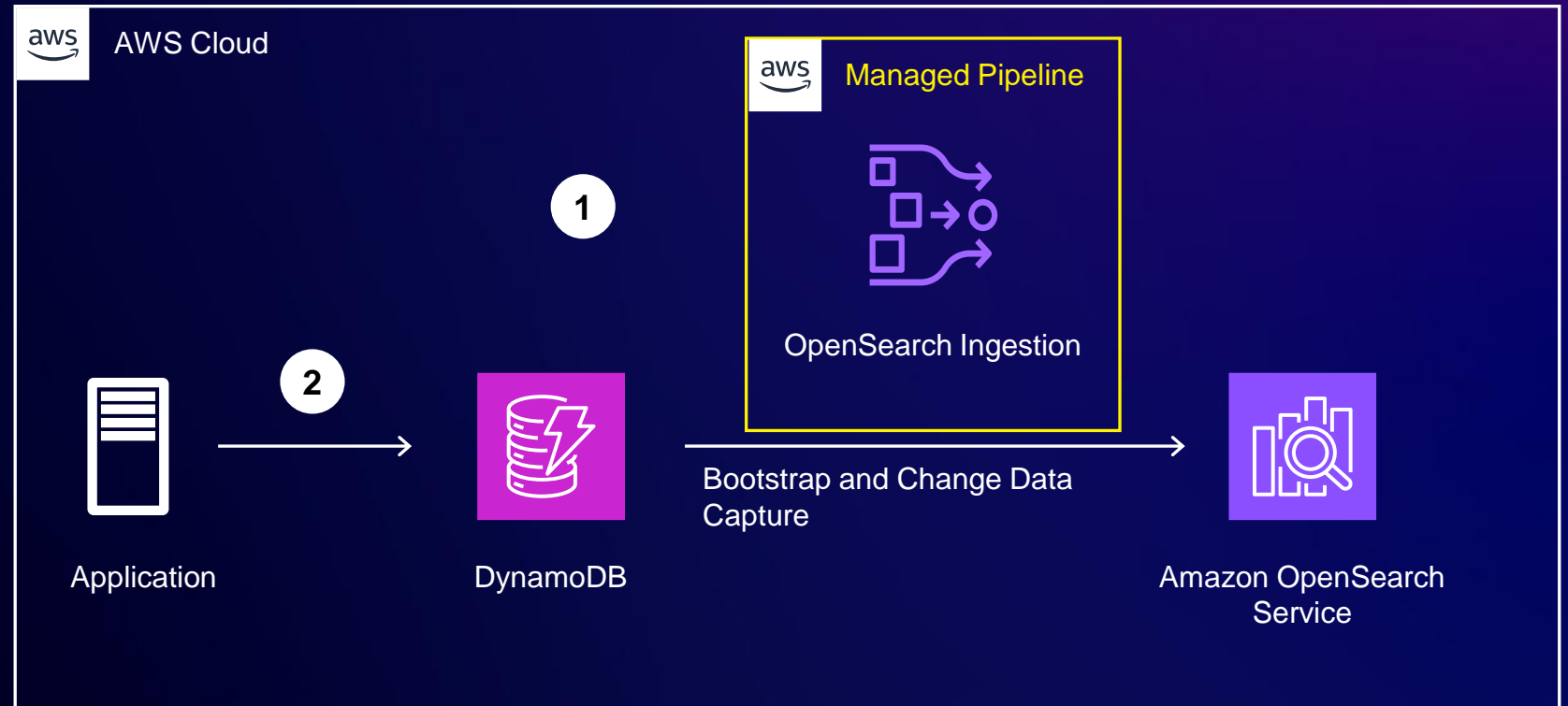
Combine, reduce, sample, and aggregate

Parsers

Native understanding of CSV and JSON

Rethinking integration

1. Setup a pipeline between DynamoDB and OpenSearch Service
2. Application continues to send updates to DynamoDB, which are synced across



Creating a pipeline – DynamoDB console

The screenshot displays the AWS DynamoDB console interface. On the left, the navigation sidebar is visible with the 'Integrations' menu item highlighted in a red box. The main content area shows the 'tblSales' table selected in the 'Tables (6)' list, also highlighted with a red box. To the right, the 'tblSales' integration page is shown, featuring a 'Create pipeline' button in the top right corner, which is highlighted with a red box. Below this, a message states 'No Opensearch integrations are created for this table' with another 'Create pipeline' button centered below it.

DynamoDB ✕

- Dashboard
- Tables
 - Update settings
 - Explore items
 - Integrations New**
 - PartiQL editor
 - Backups
 - Exports to S3
 - Imports from S3
 - Reserved capacity
 - Settings
- ▼ DAX
 - Clusters
 - Subnet groups
 - Parameter groups
 - Events

DynamoDB > Integrations > TblSales

Tables (6) ✕

Find tables by table name

Any table tag ▼

< 1 > ⚙️

- ☐ Table1
- ☒ **tblSales**
- ☐ Table3
- ☐ Table4
- ☐ Table5
- ☐ Table6

tblSales

Opensearch (0) Info

Create Zero-ETL integration with Amazon Opensearch

Find integrations

< 1 > ⚙️

Integration name	Status	Created on	Destination
No Opensearch integrations are created for this table			

Create pipeline ↗️

Creating a pipeline – configure pipeline

[Amazon OpenSearch Service](#) > [Ingestion: Pipelines](#) > Create pipeline

Step 1

Pipeline settings and
configuration

Step 2

Review and create

Pipeline settings and configuration

Pipeline configuration [Info](#)

Upload file

Configuration blueprints ▼

Browse resources

```
1 version: "2"
2 dynamodb-pipeline:
3   source:
4     dynamodb:
5       acknowledgments: true
6     tables:
7       - table_arn: "arn:aws:dynamodb:us-east-1:123456789012:table/MyTable"
8     stream:
9       start_position: "LATEST"
10    export:
11      s3_bucket: "<<my-bucket>>"
12      s3_prefix: "<<export-folder/>>"
13    aws:
14      sts_role_arn: "<<arn:aws:iam::123456789012:role/Example-Role>>"
15      region: "us-east-1"
16  sink:
17    - opensearch:
18      hosts: [ "<<https://search-mydomain-1a2a3a4a5a6a7a8a9a0a9a8a7a.us-east-1.es.amazonaws.com>>" ]
19      index: "table-index"
20      document_id: "${getMetadata(\"primary_key\")}"
21      action: "${getMetadata(\"opensearch_action\")}"
22      document_version: "${getMetadata(\"dynamodb_item_version\")}"
23      document_version_type: "external"
24    aws:
25      sts_role_arn: "<<arn:aws:iam::123456789012:role/Example-Role>>"
26      region: "<<us-east-1>>"
```

Pipeline fundamentals: source

```
source:
  dynamodb:
    tables:
      - table_arn: "arn:aws:dynamodb:us-west-2:123456789012:table/QandA"
        stream:
          start_position: "LATEST"
        export:
          s3_bucket: "osisdynamodbexports"
          s3_prefix: "/ddbexports"
    aws:
      region: "us-west-2"
      sts_role_arn: "arn:aws:iam::123456789012:role/osis-source-role"
```

Source – The DynamoDB table to pull from
Can be 1 or many

Pipeline fundamentals: processor

```
processor:
```

Processor – A hook to manipulate the data

```
  - add_entries:
```

```
    - key: "location"
```

```
      format: "${user_lat},${user_long}"
```

```
      overwrite_if_key_exists: true
```

Pipeline fundamentals: sink

sink:

Sink – The target OpenSearch index
Can be 1 or many

- opensearch:

- hosts: ["https://qanda.us-west-2.es.amazonaws.com"]

- index: "amazon-qanda"

- document_id: "\${getMetadata(\"primary_key\")}"

- aws:

- sts_role_arn: "arn:aws:iam::123456789012:role/osis-sink-role"

- region: "us-west-2"

Amazon OpenSearch Service



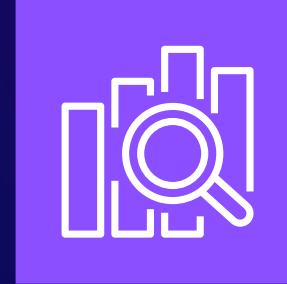
To each technology, its purpose



Amazon DynamoDB

Use DynamoDB for consistent low latency, durability, and flexibility

Most likely your “system of record”

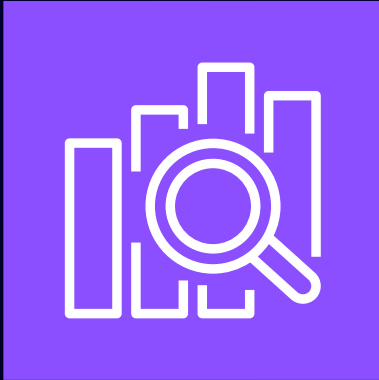


Amazon OpenSearch Service

Use OpenSearch Service to provide rich search capabilities, and relevant results

Index what you want to search then use the primary key to retrieve the items from DynamoDB

Deployment options



Managed domains

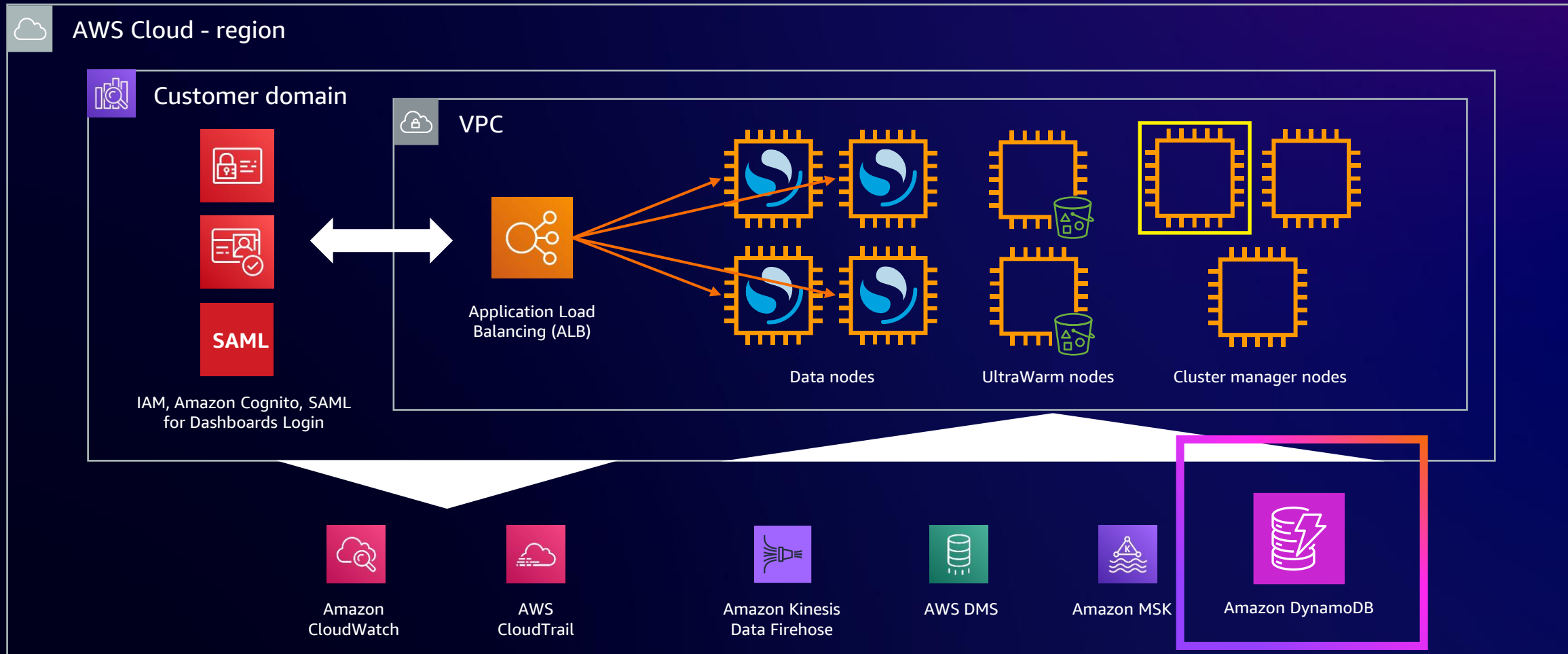
Provides fine-grained control over instance types for better resource and cost optimization, and larger scale



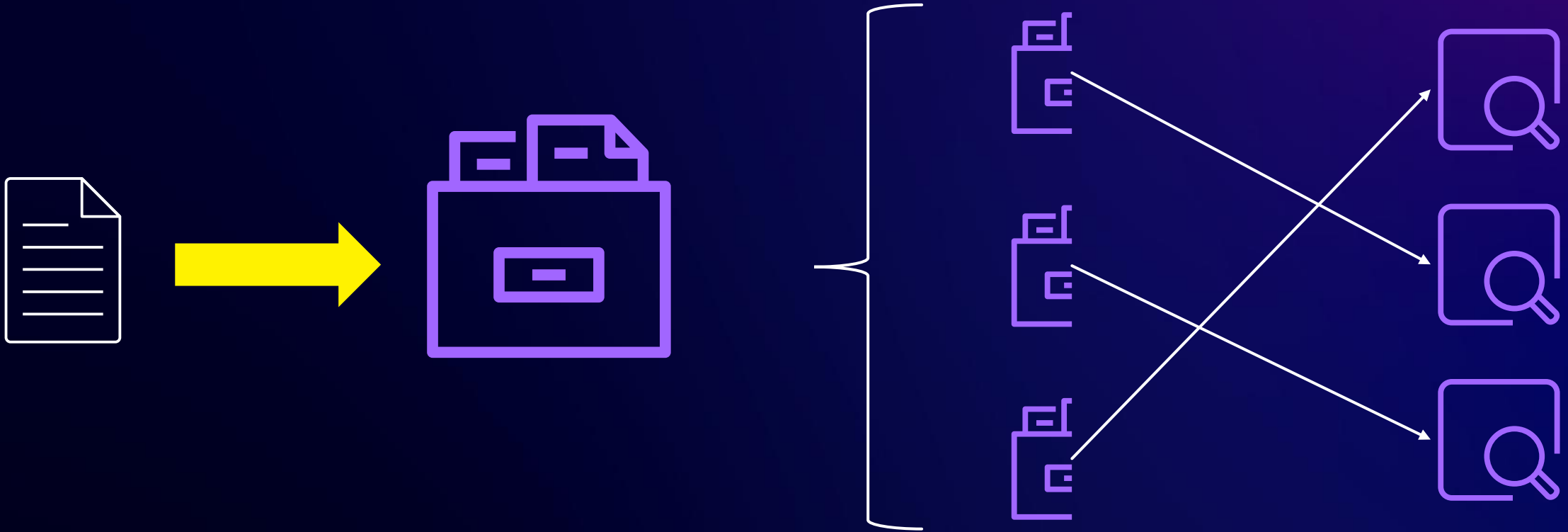
Serverless

Auto-scaled and simple to manage.
Serverless collections manage
OpenSearch indices

OpenSearch Service is a distributed solution



OpenSearch Service is a distributed database



Send documents to an index

An index is comprised
of shards

Shards are
distributed to
data nodes

Data layout

DynamoDB

Table

OpenSearch Service

Index

Data layout

DynamoDB



OpenSearch Service



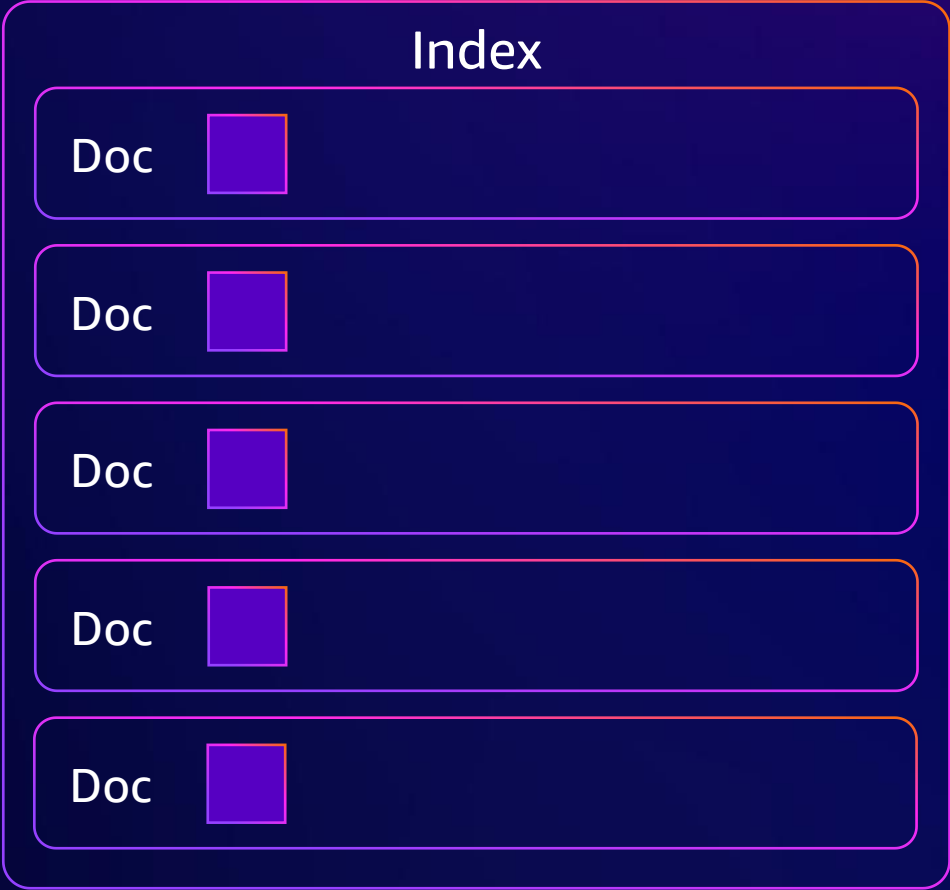
Data layout

DynamoDB



 Partition  Sort

OpenSearch Service



 _id



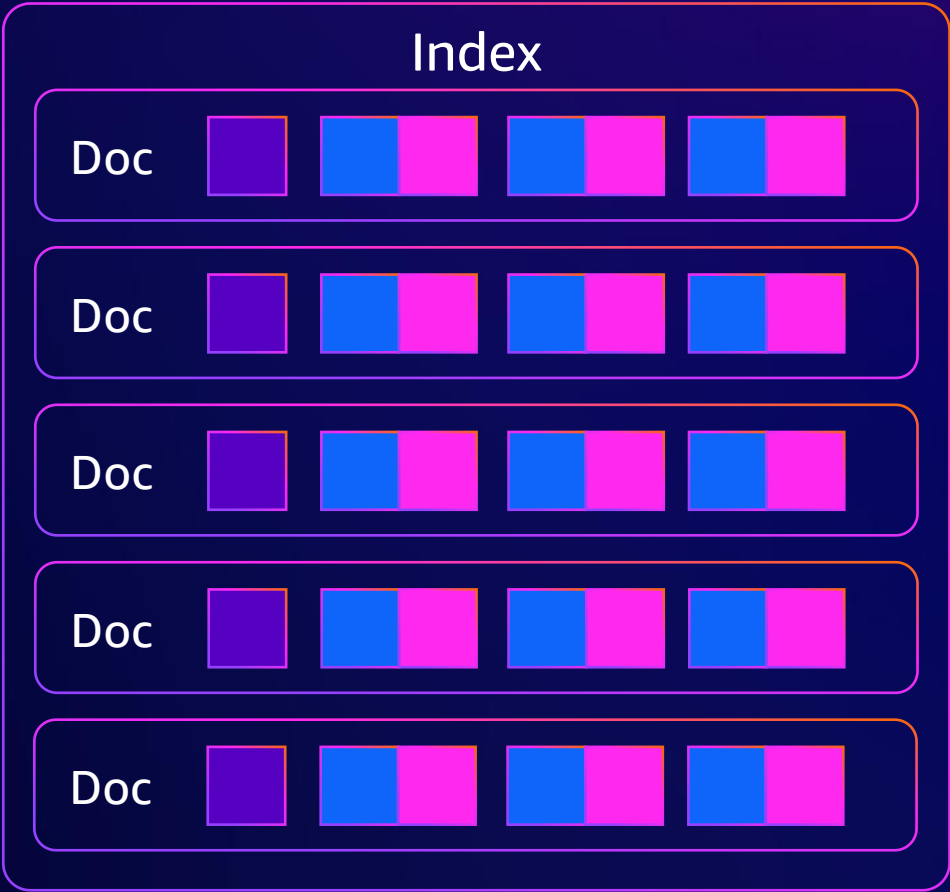
Data layout

DynamoDB



 Partition  Sort  Attribute

OpenSearch Service



 _id  Field – typed, JSON field



OpenSearch Service data layout

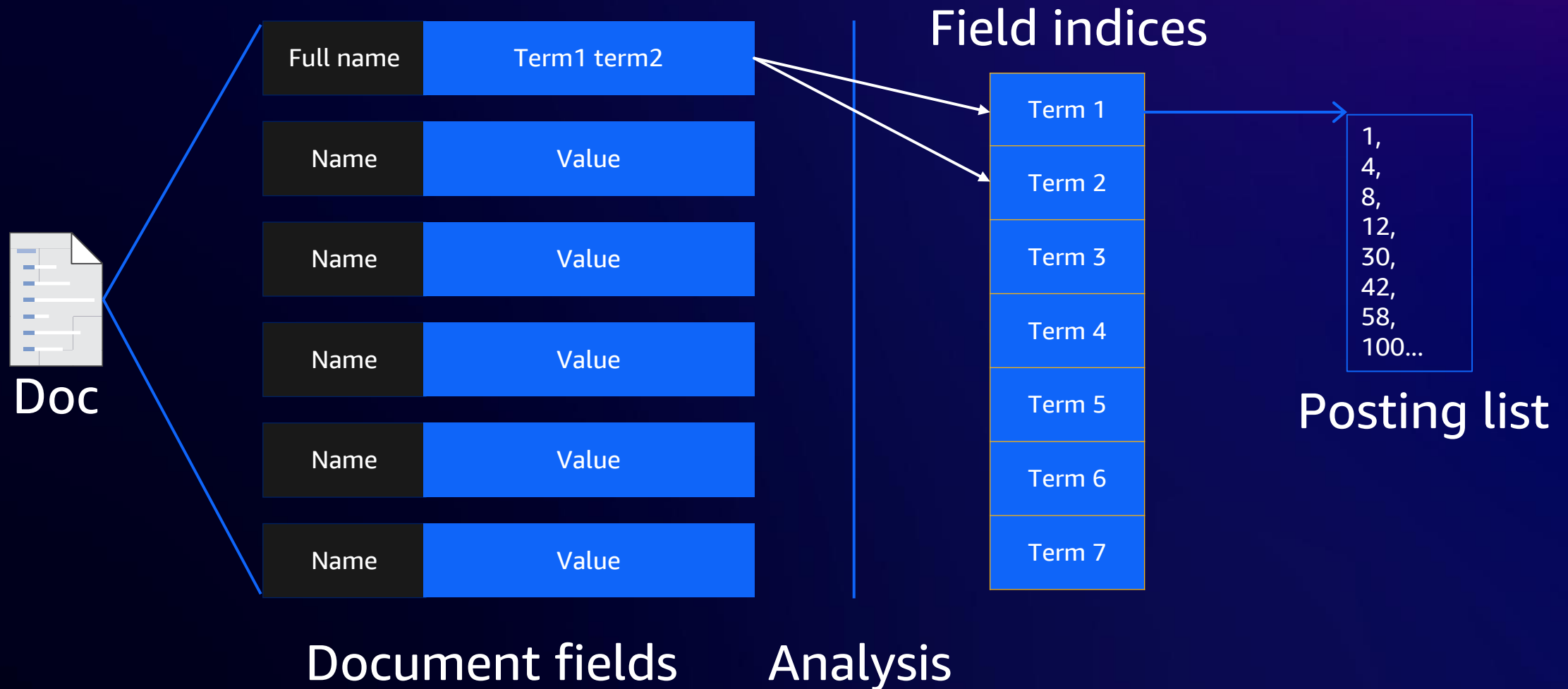
```
{
  "_index": "amazon-data2",
  "_id": "B00J2HW6VI|metadata",
  "_score": 2.4748108,
  "_source": {
    "category_name": "wrist watches",
    "sk": "metadata",
    "item_name": "Nixon Men's A4051888  
Sentry Chrono Leather Watch",
    "brand_name": "Nixon",
    "location": null,
    "doc_type": "asin",
    "pk": "B00J2HW6VI",
    "product_description": "",
    "bullets": "Round watch featuring  
stick markers, three chronograph  
subdials, and date window at 4  
o'clock Stainless steel case with  
mineral dial window Japanese  
quartz movement with analog  
display Genuine leather band with  
buckle closure Water resistant to  
100 m (330 ft): In general,  
suitable for swimming and  
snorkeling, but not scuba diving"
  }
},
```

```
{
  "_index": "amazon-data2",
  "_id": "B01C5QY7DW|question#Tx1TBTIZRCFN7N#0",
  "_score": 1.1967057,
  "_source": {
    "question_text": "The packaging  
looks like standard x-treme vision  
bulbs. Are these standard or +100%  
x-treme vision bulbs?",
    "question_type": "WH",
    "sk": "question#Tx1TBTIZRCFN7N#Q",
    "location": null,
    "pk": "B01C5QY7DW",
    "doc_type": "question",
    "embedding_chunk": null,
    "question_id": "Tx1TBTIZRCFN7N"
  }
},
```

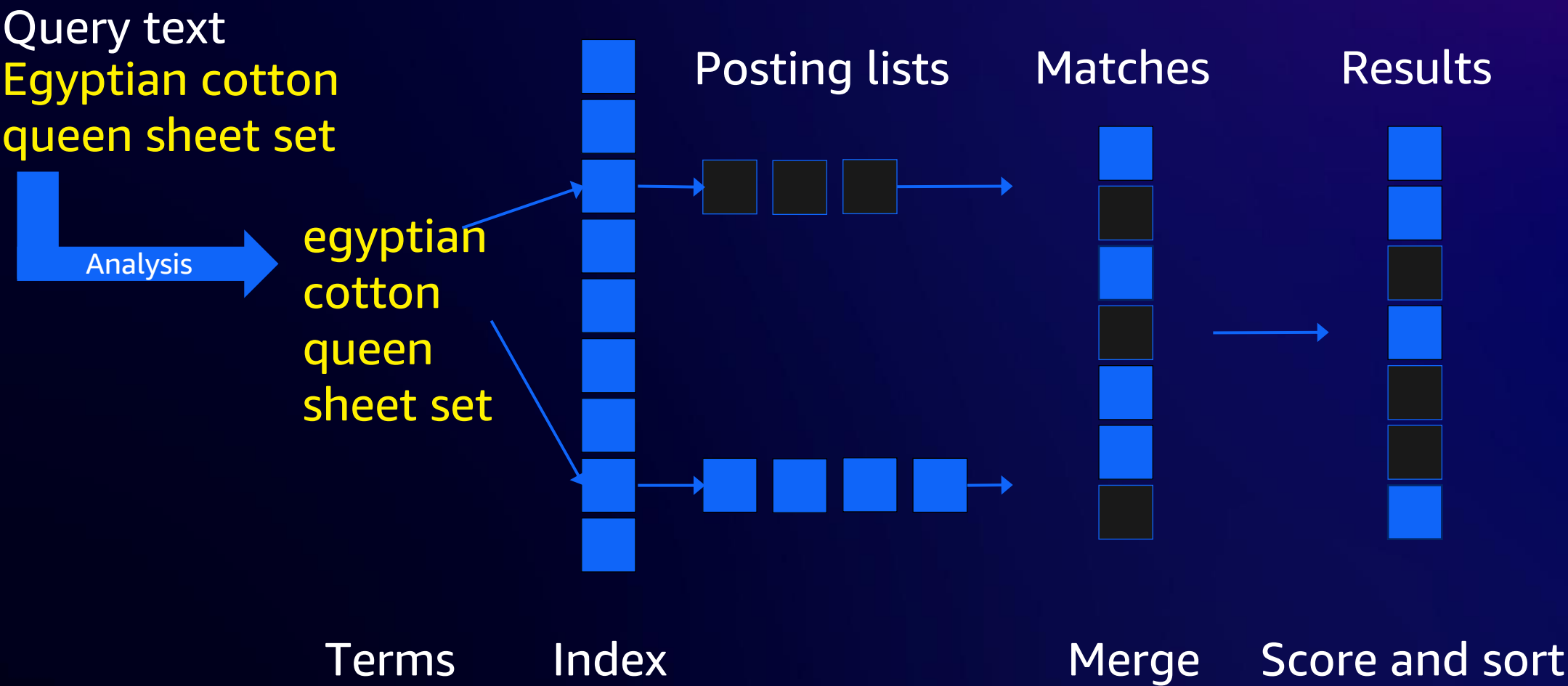
```
{
  "_index": "amazon-data2",
  "_id": "B01C5QY7DW|question#Tx1S2LI4ZDBS3NI#answer000004",
  "_score": 0.49161357,
  "_source": {
    "answer_text": "High and low beam!",
    "gender": "female",
    "doc_type": "answer",
    "sk": "question#Tx1S2LI4ZDBS3NI#answer000004",
    "name": "Jodi Herrera",
    "location": null,
    "pk": "B01C5QY7DW",
    "embedding_chunk": null,
    "product_rating": 2,
    "age": 45
  }
},
```

OpenSearch core algorithms

Lucene indices provide the core search functionality



Query processing



Text analysis

Source text

Name: 1100TC Egyptian Cotton Sateen Brown & Yellow Stripe Sheet Set – Queen

Description: 1100TC Egyptian Sateen Sheet Set Includes: - 1 Flat Sheet: Brown, Yellow & Sky blue stripes on white on front (gray stripe pattern on back) - 1 Fitted Sheet: Gray stripe pattern - 2 Sham Covers: Gray stripe pattern (Available Size: Queen / King)

Standard analyzer

Name: 1100tc egyptian cotton sateen brown yellow stripe sheet set queen

Description: 1100tc egyptian sateen sheet set includes 1 flat sheet brown yellow sky blue stripes on white on front gray stripe pattern on back 1 fitted sheet gray stripe pattern 2 sham covers gray stripe pattern available size queen king

English analyzer

Name: 1100tc egyptian cotton sateen brown yellow stripe sheet set queen

Description: 1100tc egyptian sateen sheet set includ 1 flat sheet brown yellow sky blue stripe white front grai stripe pattern back 1 fit sheet grai stripe pattern 2 sham cover grai stripe pattern avail size queen king

Text-text matching example

Query text

Egyptian cotton queen
sheet set



Analyzed

egyptian cotton queen
sheet set



Name: 1100t egyptian
cotton sateen brown yellow
stripe sheet set queen

Description: 1100t egyptian
sateen sheet set includ 1 flat
sheet brown yellow sky blue
stripe white front grai stripe
pattern back 1 fit sheet grai
stripe pattern 2 sham cover
grai stripe pattern avail size
queen king

Searching scenarios

DynamoDB table layout

Primary key		Attributes						
Partition key: pk	Sort key: sk							
B07BFVT2PN	metadata	item_name						
		EMONIA Queen Sheets Set -6 Pieces Bed Sheets- Microfiber Super Soft 1800 Series						
	question#Tx 1E43I11RHT WG8#Q	asin	question_id	question_text				
		B07BFVT2PN	Tx1E43I11RH TWG8	Are these sheets noisy?				
	question#Tx 1E43I11RHT WG8#answe r00001	asin	question_text	answer_text	gender	age	user_lat	user_lon
		B07BFVT2PN	Are these sheets noisy?	"Hi, Mine have no noise at all. I wash them and use Downy before I put them on the bed with no noise! Hope that helps"	male	97	-17.182330 5	80.307699

OpenSearch mapping defines the schema

OpenSearch has a single mapping per index

The mapping controls the storage and retrieval of the data

Unmapped fields are dynamically mapped

Caution: dynamic mapping detects strings for all

Caution: item-name collision

```
{
  "amazon-data": {
    "mappings": {
      "properties": {
        "pk": {
          "type": "text",
          "fields": { "keyword": { "type": "keyword",
                                   "ignore_above": 256 } }
        },
        "sk": {
          "type": "text",
          "fields": { "keyword": { "type": "keyword",
                                   "ignore_above": 256 } }
        },
        "doc_type": { "type": "..."},
        "location": { "type": "..."},
        "bullets": { "type": "..."},
        "category_name": { "type": "..."},
        "gender": { "type": "..."},
        "item_name": { "type": "..."},
        "name": { "type": "..."},
        "answer_text": { "type": "..."},
        "product_description": { "type": "..."},
        "question_id": { "type": "..."},
        "question_text": { "type": "..."},
        "question_type": { "type": "..."},
        "brand_name": { "type": "..."},
        "age": { "type": "long" },
        "product_rating": { "type": "long" }
      }
    }
  }
}
```

What kinds of queries can I run?

Text

Full-text and single term

Numeric

Exact match and range queries

Geo

Geohash, geo-point, geo-polygon, and bounding box

Vector

Exact and approximate nearest-neighbor search supporting vector matching and semantic search

Special

Percolate, distance, span, script, sparse, neural

You can sort results by most field types

Scenario: Text Search

Matching full text: "Egyptian cotton queen sheet set"

```
GET amazon-qanda/_search
{
  "query": {
    "simple_query_string": {
      "query": "Egyptian cotton queen sheet set",
      "fields": ["item_name^2",
        "product_description"]
    }
  },
  "_source": {"excludes": "*"},
  "highlight": {"fields": {"item_name" : {},
    "product_description": {}}}
}
```

```
{
  "_index": "amazon-qanda",
  "_id": "B00D60KQFA|metadata",
  "_score": 15.895072,
  "_source": {},
  "highlight": {
    "item_name": [
      "1100TC <em>Egyptian</em> <em>Cotton</em> Sateen
        Brown & Yellow Stripe <em>Sheet</em> <em>Set
          </em> - <em>Queen</em>"
    ],
    "product_description": [
      "100TC <em>Egyptian</em> Sateen <em>Sheet</em>
        <em>Set</em> Includes: - 1 Flat <em>Sheet</em>
          >: Brown, Yellow & Sky blue stripes on white
            on",
      "front (gray stripe pattern on back) - 1 Fitted
        <em>Sheet</em>: Gray stripe pattern - 2 Sham
          Covers: Gray stripe",
      "pattern (Available Size: <em>Queen</em> / King
        )"
    ]
  ]
},
```

Complex queries; text, range, filters

```
GET amazon-qanda/_search
{
  "query": {
    "bool": {
      "filter": [
        { "term": {
          "doc_type": "question"
        } }
      ],
      "must": [
        { "range": {
          "age": { "gte": 18, "lte": 24 }
        } },
        { "match": {
          "product_description": "Cotton"
        } }
      ]
    }
  },
  "_source": ["item_name", "name", "age"]
}
```

```
"hits": [
  {
    "_index": "amazon-qanda",
    "_id": "B06XL15KDN|question#Tx1N0H5I0HG0BLV#answer00001",
    "_score": 2.3822958,
    "_source": {
      "name": "Aaron Hicks",
      "item_name": "Bibb Home Paisley Printed 1200 Thread Count 6 Piece Cotton Sheet Set with Bonus Pillowcases - 8 Colors (Cal King, Gray)",
      "age": 23
    }
  },
  {
    "_index": "amazon-qanda",
    "_id": "B06XL15KDN|question#Tx1N0H5I0HG0BLV#answer00002",
    "_score": 2.3571477,
    "_source": {
      "name": "Vicki May",
      "item_name": "Bibb Home Paisley Printed 1200 Thread Count 6 Piece Cotton Sheet Set with Bonus Pillowcases - 8 Colors (Cal King, Gray)",
      "age": 18
    }
  },
]
```

Scenario: Analytics

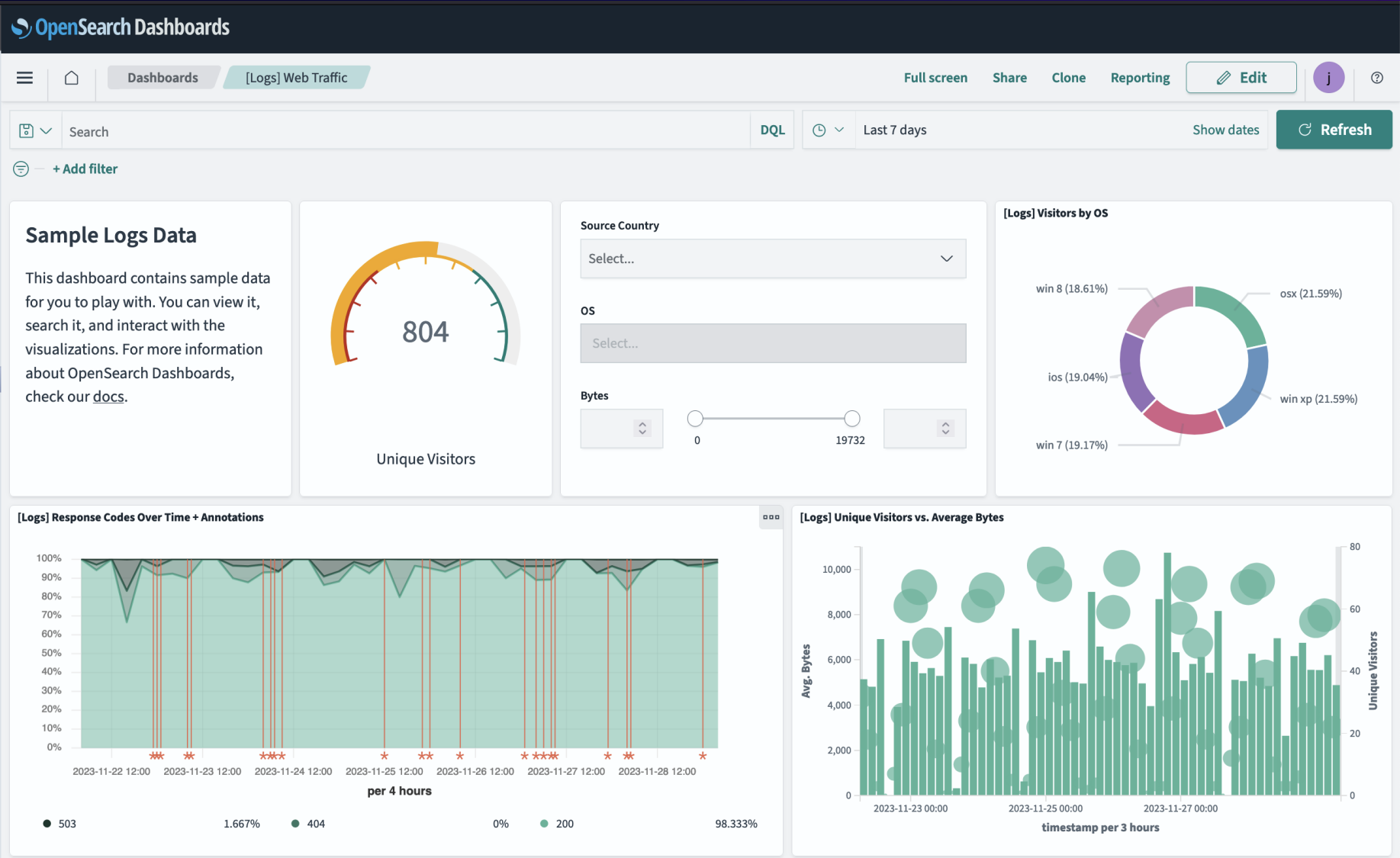


Count questions and answers by ASIN

```
GET amazon-data2/_search
{
  "query": {"match_all": {}},
  "size": 0,
  "aggs": {
    "Asin": {
      "terms": {
        "field": "pk",
        "size": 10
      },
      "aggs": {
        "Type": {
          "terms": {
            "field": "doc_type",
            "size": 10
          }
        }
      }
    }
  }
}
```

```
"aggregations": {
  "Asin": {
    "doc_count_error_upper_bound": 353,
    "sum_other_doc_count": 120475,
    "buckets": [
      {
        "key": "B01N32NCPM",
        "doc_count": 18050,
        "Type": {
          "doc_count_error_upper_bound": 0,
          "sum_other_doc_count": 0,
          "buckets": [
            {
              "key": "answer",
              "doc_count": 13123
            },
            {
              "key": "question",
              "doc_count": 4926
            }
          ]
        }
      },
      {
        "key": "asin",
        "doc_count": 1
      }
    ]
  }
},
```

OpenSearch Dashboards for visualizations



Scenario: Geospatial search

Mutate data and set a mapping

OpenSearch Ingestion has a rich templating capability for mapping

Add a template to your OpenSearch Ingestion config

OpenSearch Ingestion sends the template on export, and pipeline start/stop

To map geo types, use field type *geo_point*

```
processor:
  - add_entries:
      entries:
        - key: "location"
          format: "${user_lat}, ${user_lon}"
          overwrite_if_key_exists: true
```

```
template_type: "index-template"
template_content: |
{
  "template": {
    "mappings": {
      "properties": {
        "pk": { "type": "keyword", "ignore_above": 256 },
        "sk": { "type": "keyword", "ignore_above": 256 },
        "doc_type": { "type": "keyword", "ignore_above": 256 },
        "bullets": { "type": "text" },
        "category_name": { "type": "keyword", "ignore_above": 256 },
        "gender": { "type": "keyword", "ignore_above": 256 },
        "item_name": { "type": "text", "fields": { "keyword": { "type": "keyword", "ignore_above": 256 } } },
        "name": { "type": "text", "fields": { "keyword": { "type": "keyword", "ignore_above": 256 } } },
        "answer_text": { "type": "text" },
        "product_description": { "type": "text" },
        "question_id": { "type": "keyword", "ignore_above": 256 },
        "question_text": { "type": "text" },
        "question_type": { "type": "keyword", "ignore_above": 256 },
        "brand_name": { "type": "text", "fields": { "keyword": { "type": "keyword", "ignore_above": 256 } } },
        "location": { "type": "geo_point" },
        "age": { "type": "integer" },
        "product_rating": { "type": "integer" }
      }
    }
  }
}
```

Find me answers from women in Nevada

```
GET amazon-qanda/_search
{
  "query": {
    "bool": {
      "must": [
        {
          "term": {
            "gender": "female"
          }
        }
      ],
      "filter": [
        {
          "geo_bounding_box": {
            "location": {
              "top_left": {
                "lat": 42.0019276110661,
                "lon": -120.00654287832877
              },
              "bottom_right": {
                "lat": 35.00145019239192,
                "lon": -114.04113626206261
              }
            }
          }
        }
      ]
    }
  },
  "_source": ["name", "age", "gender", "location"]
}
```

```
{
  "_index": "amazon-qanda",
  "_id": "B072C22R8C|question#TxKGY0R6KXE7K6#answer00002",
  "_score": 0.7558831,
  "_source": {
    "gender": "female",
    "name": "Tiffany Hanson",
    "location": "36.671596, -117.24538",
    "age": 96
  }
},
{
  "_index": "amazon-qanda",
  "_id": "B003FGD8FU|question#Tx36DIIGL0DM9FW#answer00001",
  "_score": 0.7492434,
  "_source": {
    "gender": "female",
    "name": "David Alvarez",
    "location": "37.0252585, -118.310287",
    "age": 80
  }
}
```


Scenario: LLM-Backed, Semantic Search

Artificial intelligence / machine learning boom

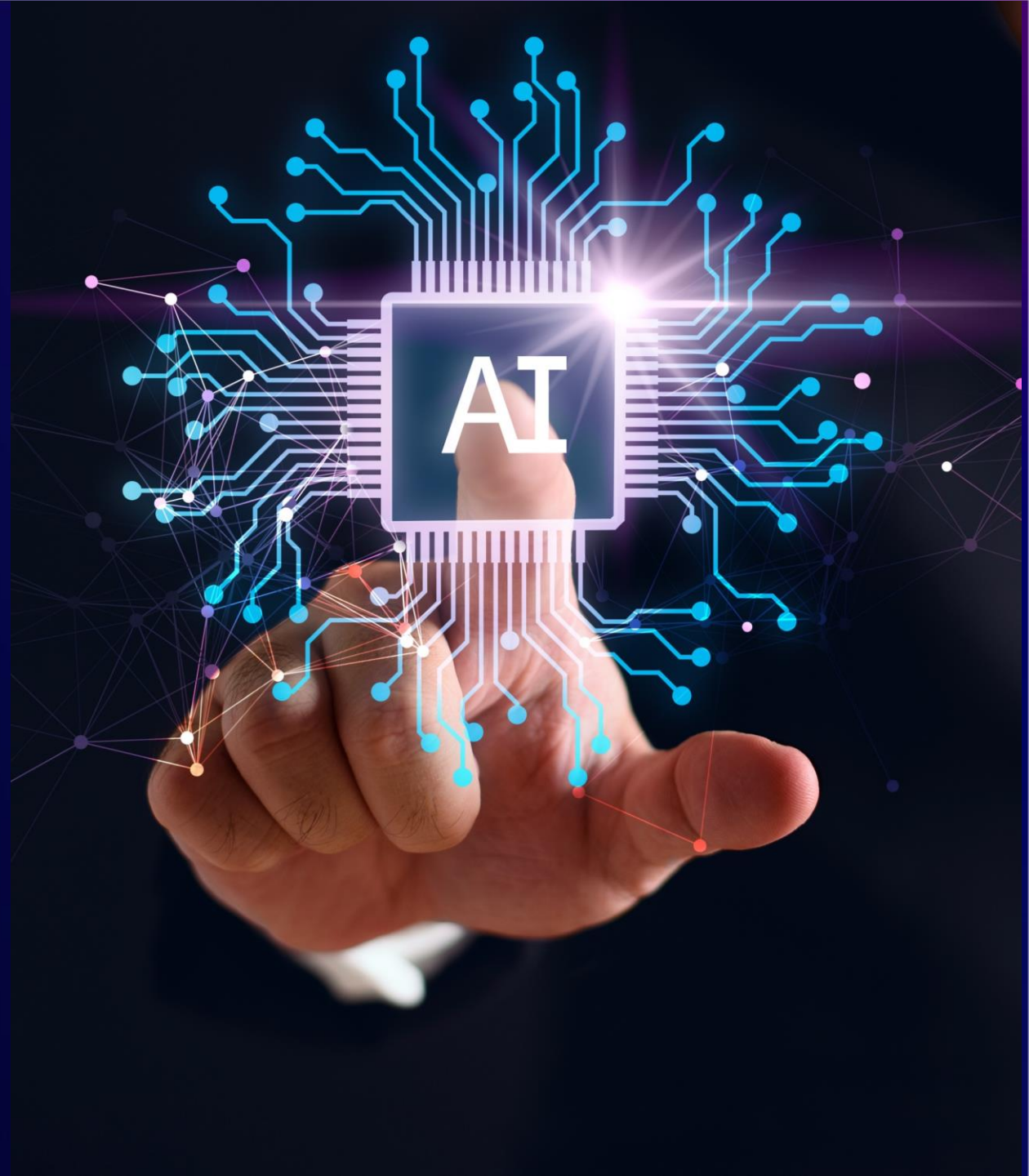
Large language models providing good baseline NLP

ChatBots hit the public's awareness

"Semantic" capabilities are used in search workloads

Other AI/ML techniques are used in search as well

Amazon OpenSearch Service enables working with vector embeddings



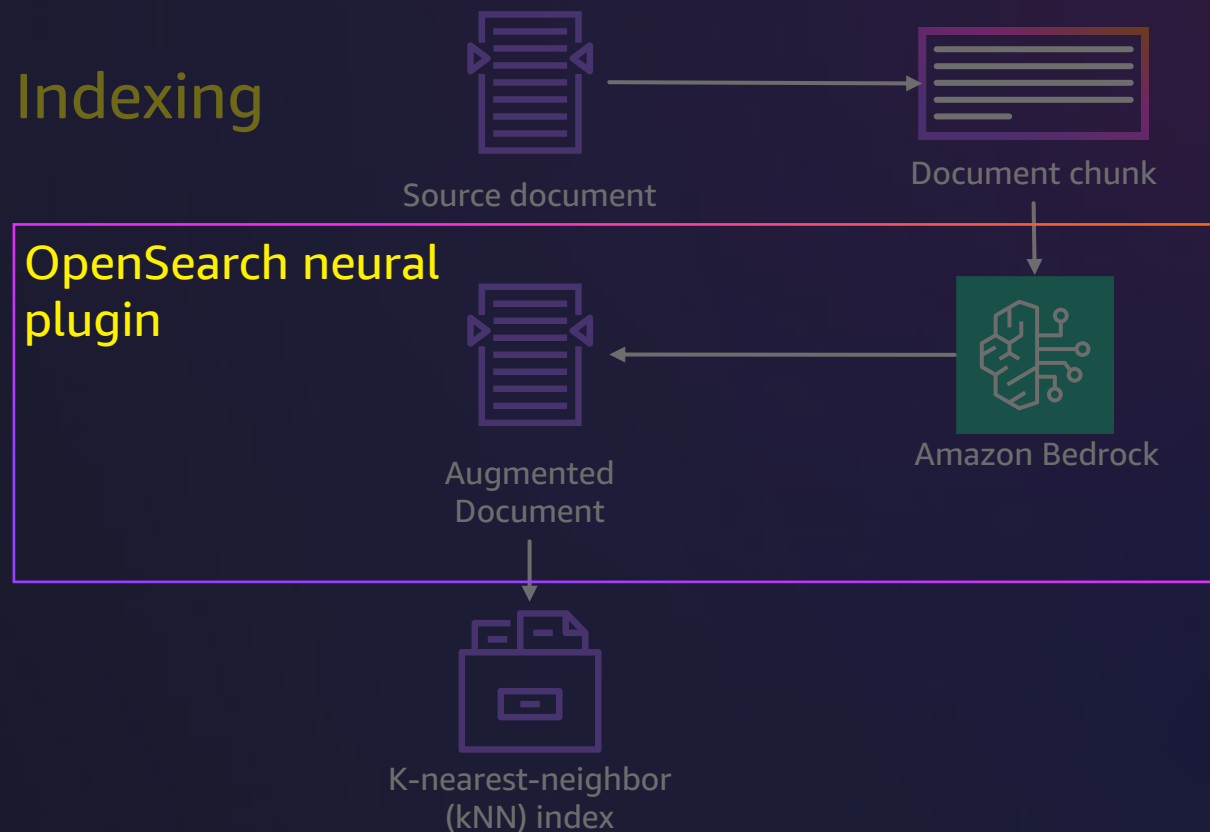
Working with embeddings

OpenSearch Service provides connectors to 3P model-hosting systems – e.g. Amazon Bedrock

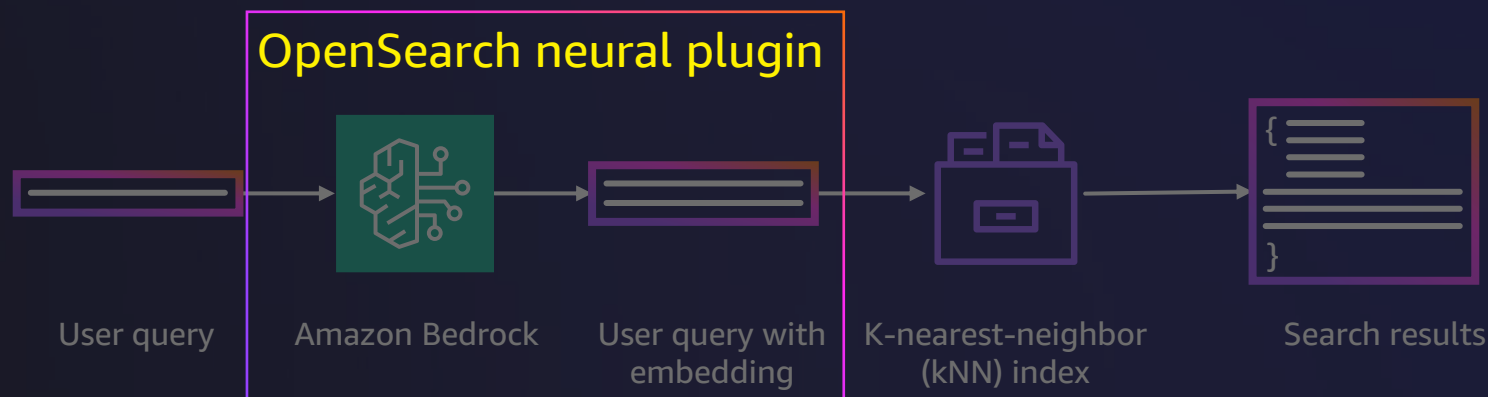
Indexing: Select chunks from the source document, send to the 3P system for embeddings

Search: Create embedding for the query then find nearest neighbors

Indexing



Search



OpenSearch Ingestion config

Add an entry containing
relevant portions of text

Add a vector field to the
mapping

OpenSearch Service's neural
plugin does the rest

```
processor:
  - add_entries:
      entries:
        - key: "embedding_chunk"
          format: "${item_name} ${product_description} ${question_text} ${answer_text}"
          overwrite_if_key_exists: true
template_type: "index-template"
template_content: |
{
  "template": {
    "settings": {
      "index.knn": true,
      "default_pipeline": "nlp_pipeline",
      "number_of_shards": 2,
      "number_of_replicas": 0
    }
    "mappings": {
      "properties": {
        "pk": { "type": "keyword", "ignore_above": 256 },
        "sk": { "type": "keyword", "ignore_above": 256 },
        "doc_type": { "type": "keyword", "ignore_above": 256 },
        "embedding": {
          "type": "knn_vector",
          "dimension": 1536,
          "method": {
            "engine": "nmslib",
            "name": "hnsw",
            "space_type": "l2"
          }
        }
      }
    }
  },
}
```

What sheets should I use if I have a cat?

Top 5 questions are about dogs!

Top 5 answers mention fur

```
GET amazon-data2/_search
{
  "query": {
    "neural": {
      "embedding": {
        "query_text": "What sheets should I use if I have a cat",
        "k": 100,
        "model_id": "AqSM64sBIPi9Y3qmJmnG"
      }
    }
  },
  "_source": ["item_name", "answer_text", "question_text"]
}
```

```
"answer_text": "I have a long-haired cat (a Maine Coon whose hair sticks to everything, but I find a lot less on these sheets than on any of the other sheets I use. This is my first set of microfiber sheets.
```

```
the others are cotton and bamboo. I even notice a big difference between the pillowcases... Since this set comes with only two cases, I supplement with cotton ones. My cat has taken over one of the pillows with a case from this set. She leaves less hair on the pillow she sleeps on every night than I find on my cotton pillowcases that she just walks past and brushes up against! I've been really impressed with how little hair they attract. (In fact, I'm on this page to order another set!) Plus they are unbelievably soft and comfortable.",
```

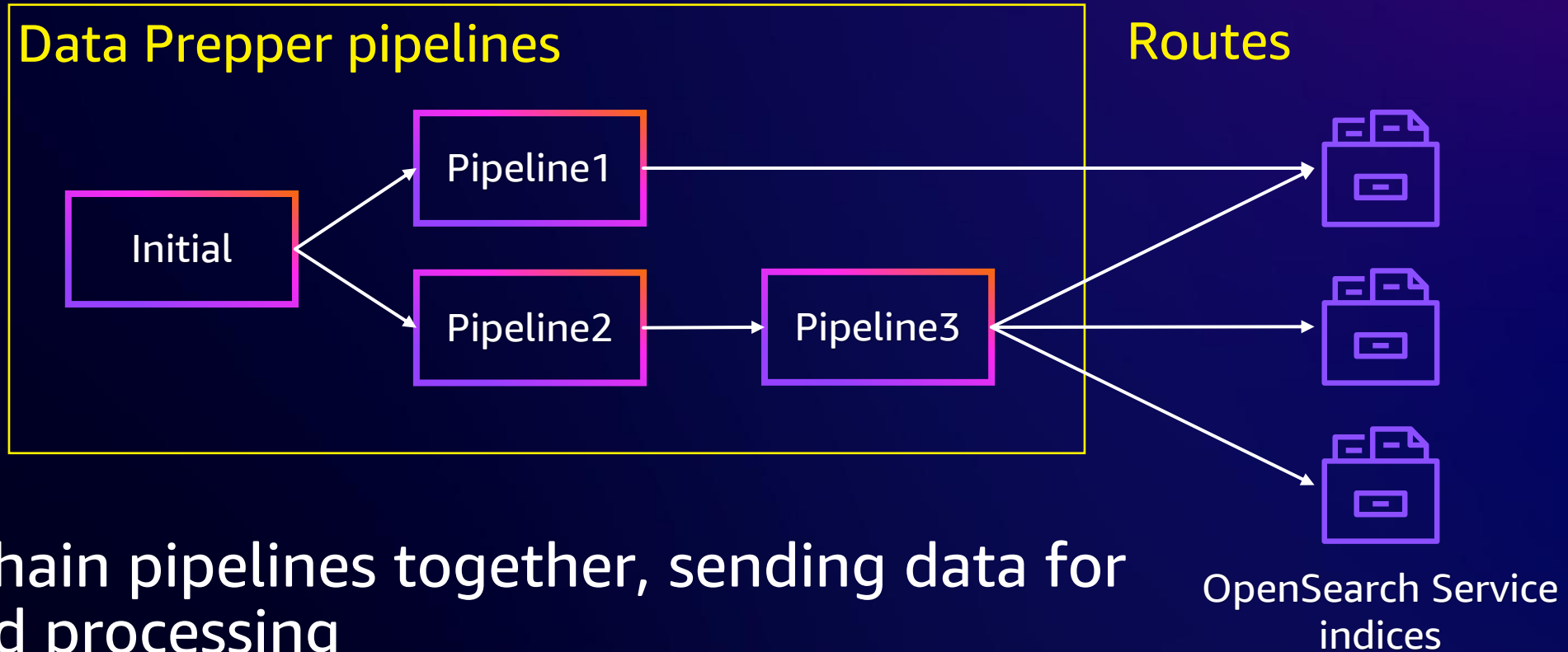
```
"question_text": "I have dogs that shed, will this stick to the bed sheets??",
```

```
"item_name": "BLC Bed Sheet Set, Hypoallergenic Microfiber 3-piece sheets with 14-Inch Deep Pocket (Twin, Black)"
```

How-to and Best Practices



Advanced pipeline architectures



You can chain pipelines together, sending data for specialized processing

Head/tail sampling, and data reduction

Route different range keys to different indices

How-to develop, best practices: getting started

Create the mapping(s) in OpenSearch Service and create your index first

Establish the pipeline with your configuration, and routing

Use the AWS console to create an item(s) in DynamoDB to ensure that the mapping is correct and the connection is working

Use OpenSearch Dashboards “Dev Tools” tab to make sure you can run your queries and your results are correct

When your mapping is correct and the queries are right, move the mapping into your OpenSearch Ingestion configuration

How-to develop, best practices: troubleshooting

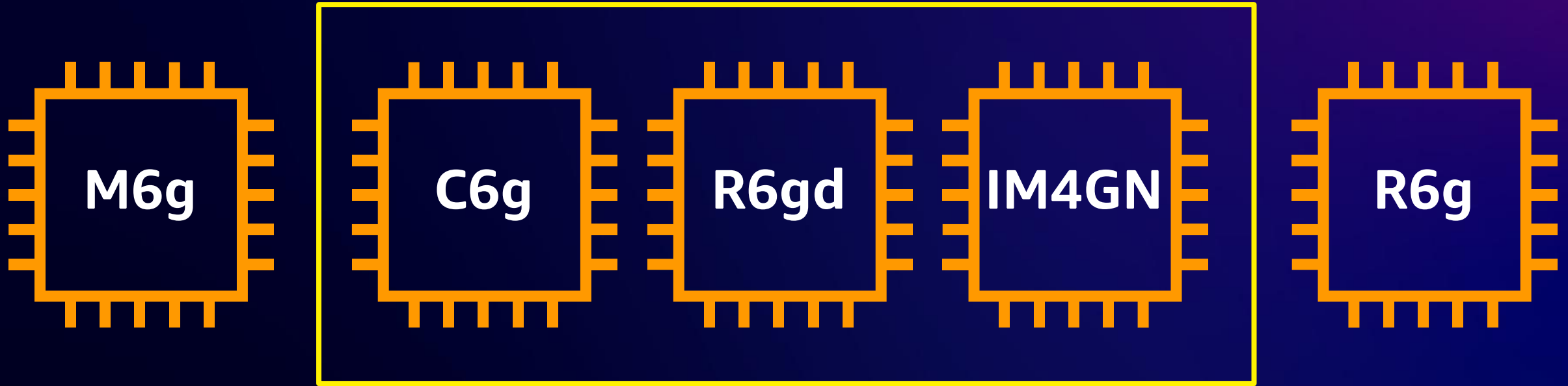
Enable CloudWatch logs for your Amazon OpenSearch Service domain, and for your OpenSearch Ingestion pipeline

Always use a “dead letter queue”

Stop the pipeline, delete and recreate your index, then start the Ingestion pipeline to re-flow data from your export

Add a second sink to Amazon S3 to capture transformed documents and find problems

OpenSearch Service managed domain sizing



Storage needed =
 $\text{source} * 1.1 * 2 * 1.15$

Target 30/50 GiB shards

Or use Amazon OpenSearch Serverless!

$\text{vCPUs} = 1.5 * \text{shards}$

assumes each shard occupies a
vCPU 100%

OCUs

You can set min/max OCUs for your OpenSearch Ingestion pipeline to 1/96 – autoscaling will optimize; limit for cost control

Throughput per OCU for normal cases is ~5 MiB/s*

1 OCU per 5,000 WCUs as a generic guideline*

*Note: actual sizing depends on data size, data velocity, pipeline manipulation, and OpenSearch Service domain size and index/shard strategy

Wrap-up

DynamoDB offers features that make integrating and replicating with other data stores relatively easy

OpenSearch offers best-of-breed search

OpenSearch Ingestion makes it easy to integrate DynamoDB and OpenSearch with zero code

OpenSearch indices are powerful and flexible when combined with DynamoDB to solve scenarios like full-text search, geo search, neural, and vector search

Available in 15 regions today!

Thank you!



Please complete the session
survey in the mobile app