101: Intro to OpenSearch: Logs, Search, Analytics

Francesco Tisiot & Ben Gamble

Agenda

- What is OpenSearch
- Overview and main concepts
- Data discovery
- OpenSearch for AI
- Data pipelines for OpenSearch

What is OpenSearch?

OpenSearch is a **scalable**, **flexible**, and **extensible open-source** software suite for **search**, **analytics**, and **observability** applications licensed under Apache 2.0

Use cases

- Full-text search
- Data aggregations
- Logs, metrics and trace analysis
- Data discovery and visualisation of data
- Vector search

When to use OpenSearch

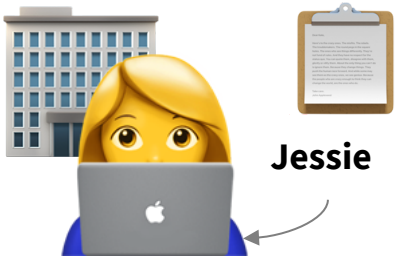
When to choose OpenSearch?

- **Fast full-text** search that takes into account:
  - Normalisations
  - Synonyms
  - Stemming
  - Ranking by relevance
- Fast **data aggregations** and analytics
- **Offload** the computational load **from main database**
- Fast search across **big number of properties**
- You’re searching in JSON objects

When NOT to use OpenSearch

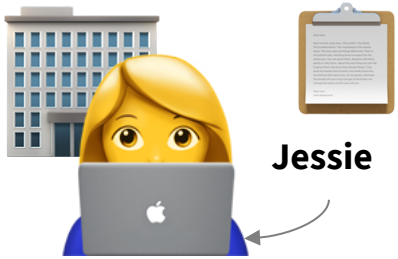
- ACID compliant requirements
  - Eventual consistency
- Mutation-heavy workload
  - Strong preference for append-only data

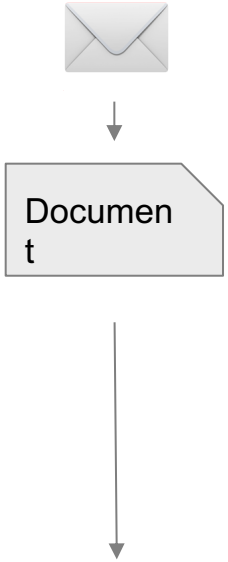
Overview and main concepts



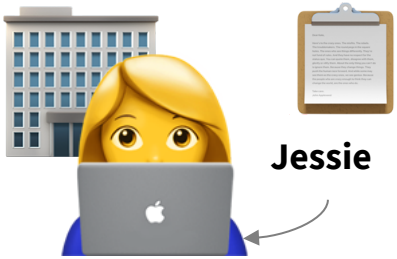
OpenSearch

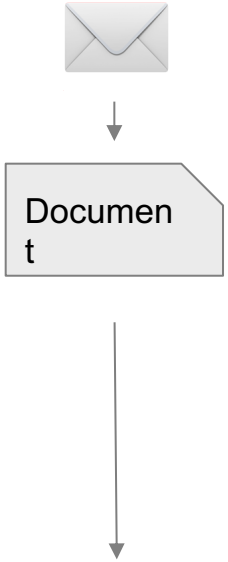
A document - single unit of data (JSON object)





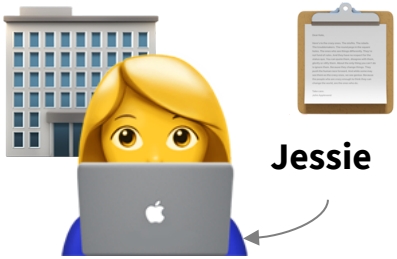
OpenSearch

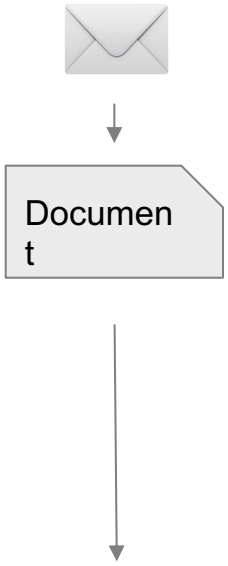




A document - single unit of data (JSON object)
An index - collection of documents stored across shards

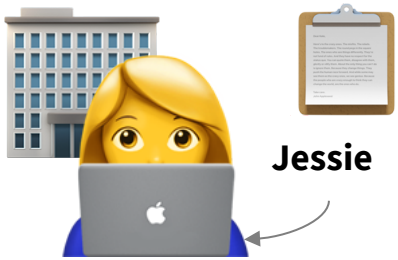
OpenSearch

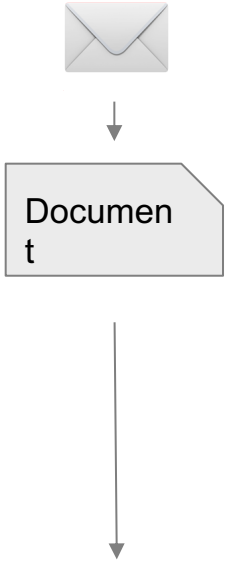




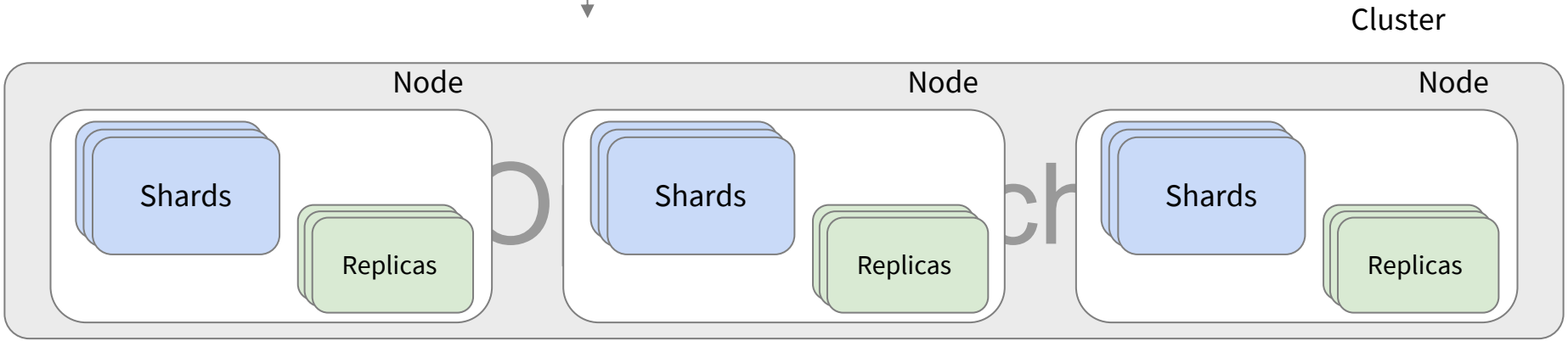
A document - single unit of data (JSON object)
An index - collection of documents stored across shards
Mapping - structure of the document data

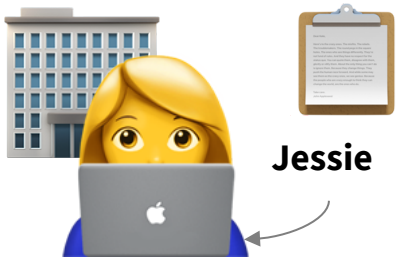
OpenSearch

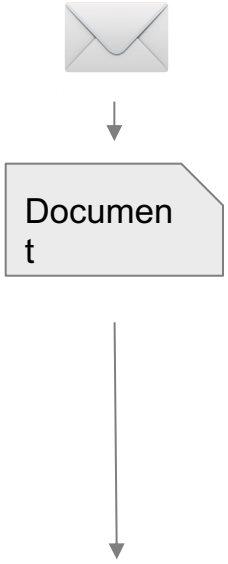




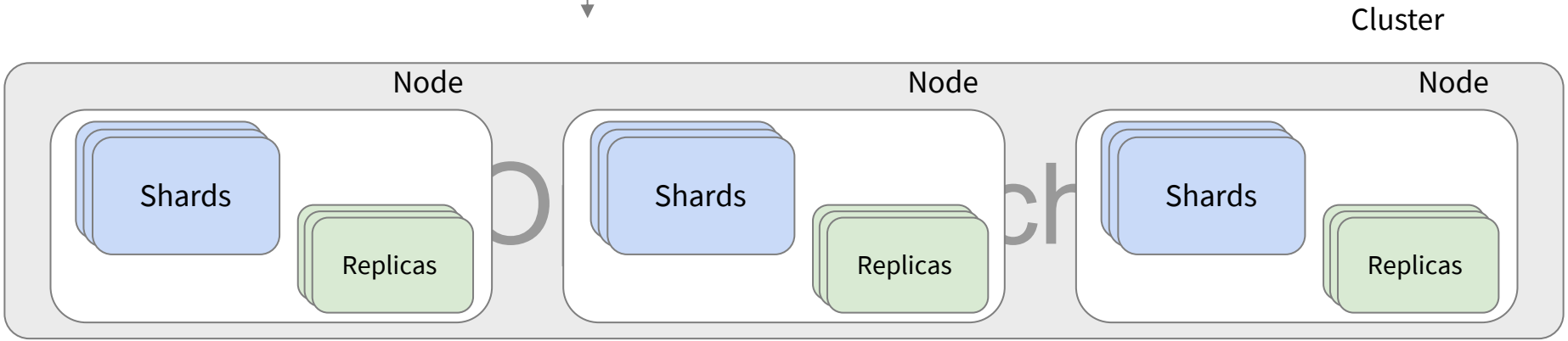
A document - single unit of data (JSON object)
An index - collection of documents stored across shards
Mapping - structure of the document data
A shard - a unit to distribute data across the cluster

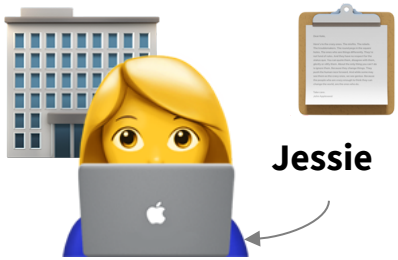


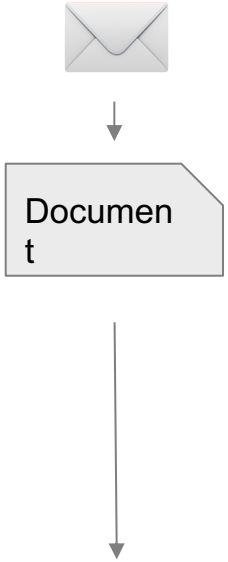




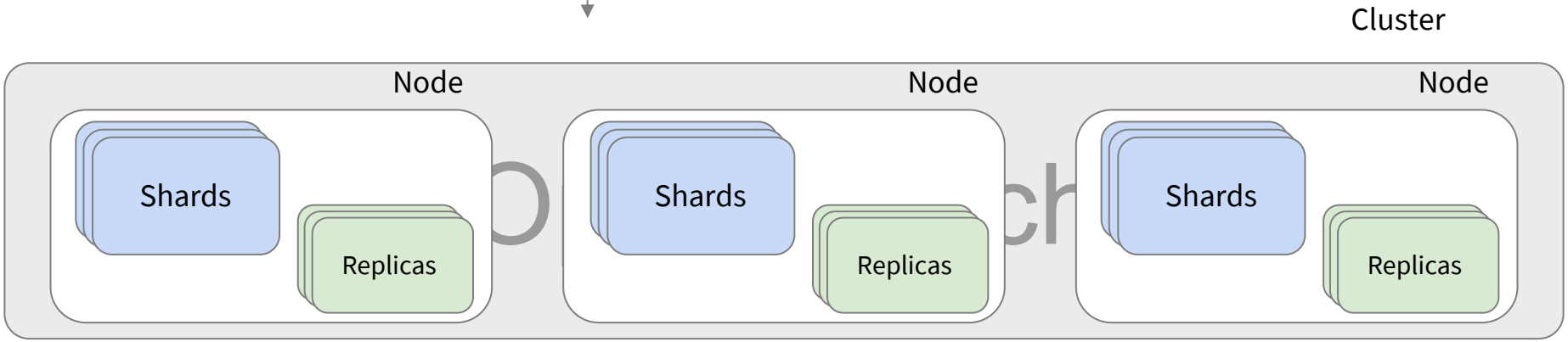
A document - single unit of data (JSON object)
An index - collection of documents stored across shards
Mapping - structure of the document data
A shard - a unit to distribute data across the cluster
A replica - copy of a shard

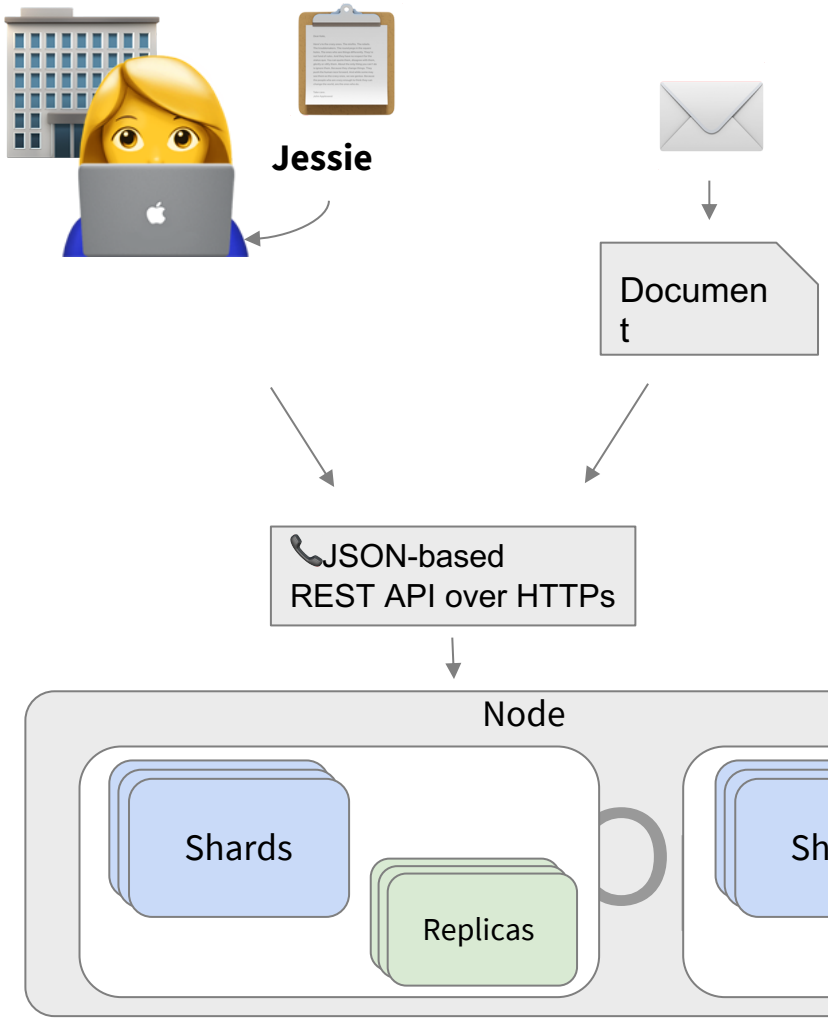






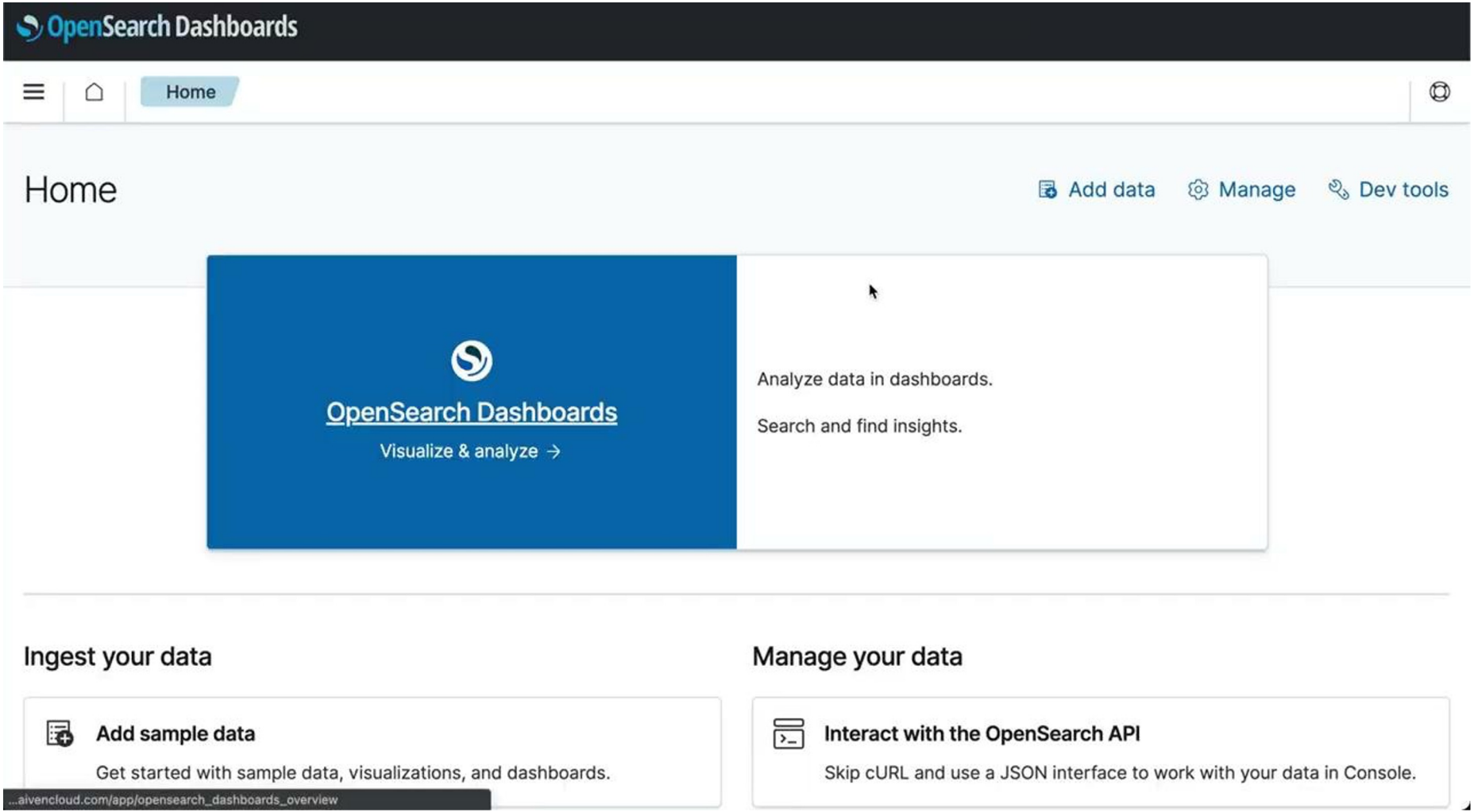
A document - single unit of data (JSON object)
An index - collection of documents stored across shards
Mapping - structure of the document data
A shard - a unit to distribute data across the cluster
A replica - copy of a shard
A cluster - a group of nodes

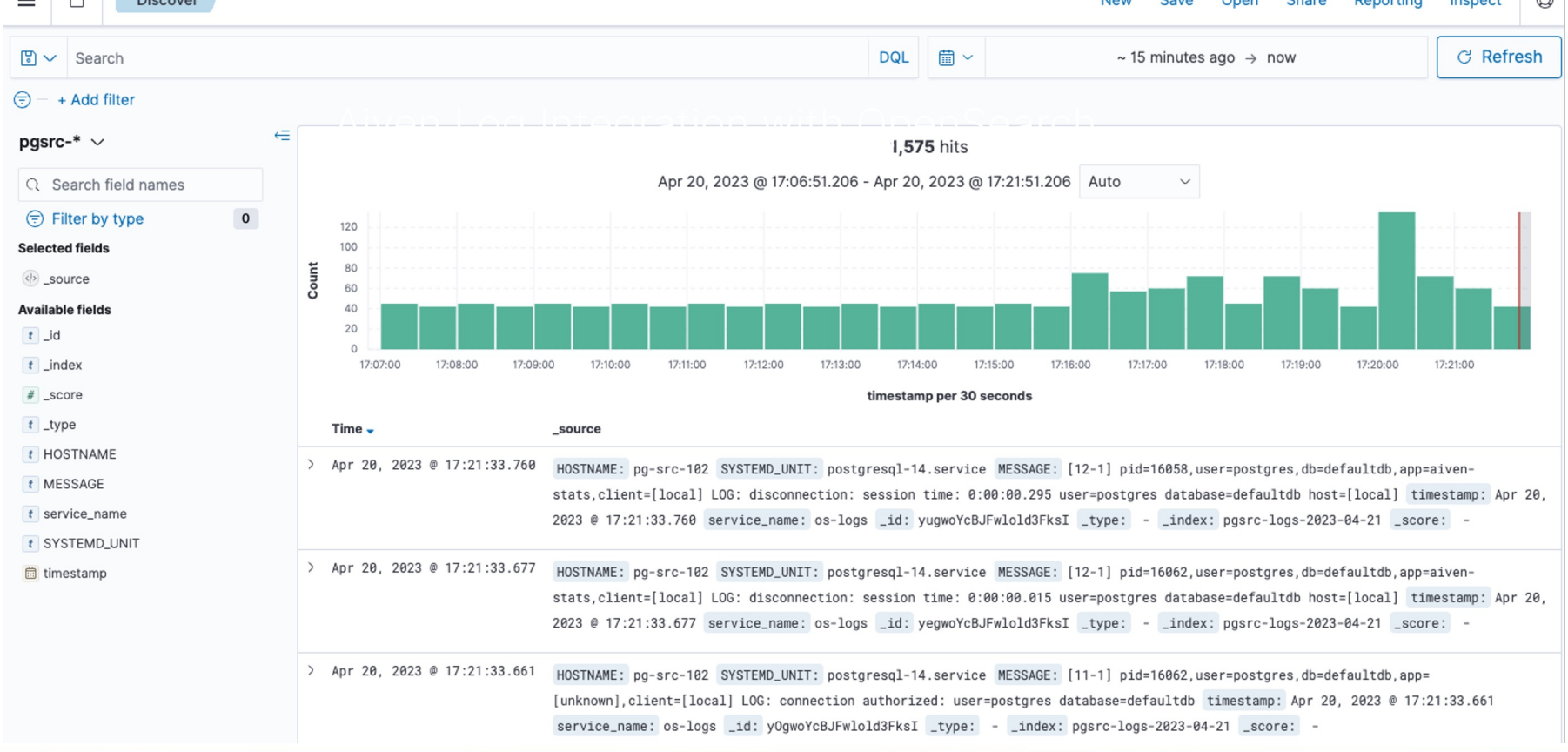




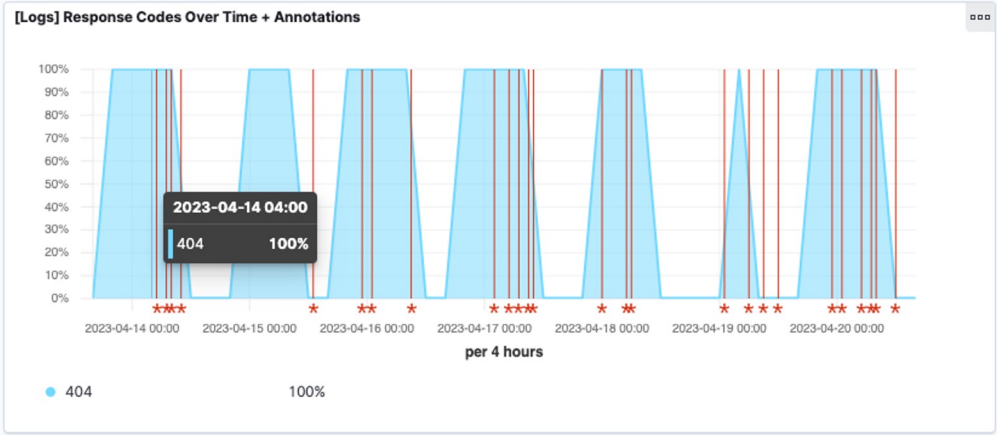
A document - single unit of data (JSON object)
Indexing - storing data and building inverted index
Mapping - structure of the document data
An index - collection of documents stored across shards
A shard - a unit to distribute data across the cluster
A replica - copy of a shard
A cluster - a group of nodes

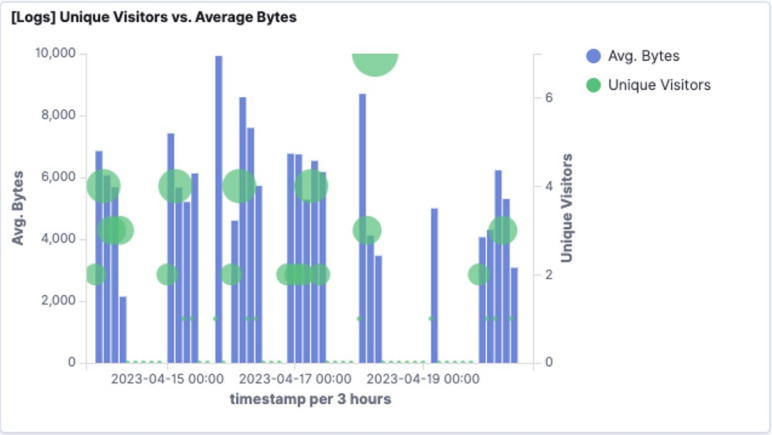
Data discovery with OpenSearch

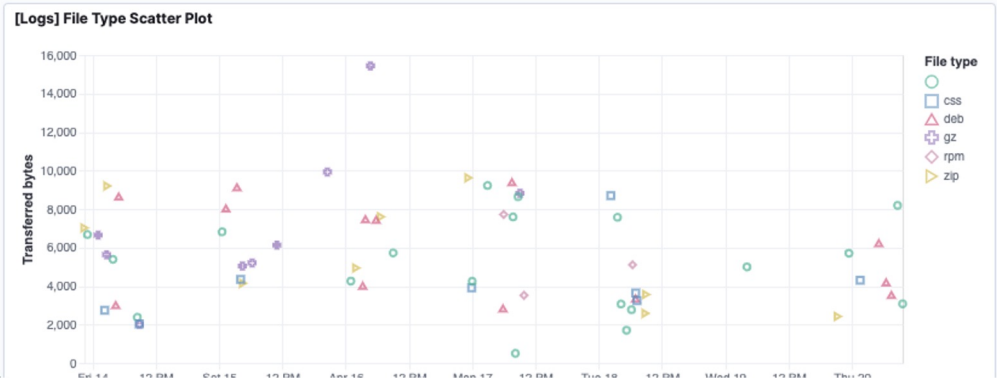

OpenSearch Dashboards
Home
Home
Add data
Manage
Dev tools
OpenSearch Dashboards
Visualize & analyze →
Analyze data in dashboards.
Search and find insights.
Ingest your data
Add sample data
Get started with sample data, visualizations, and dashboards.
Manage your data
Interact with the OpenSearch API
Skip cURL and use a JSON interface to work with your data in Console.
...aivencloud.com/app/opensearch_dashboards_overview



Aiven Log Integration with OpenSearch



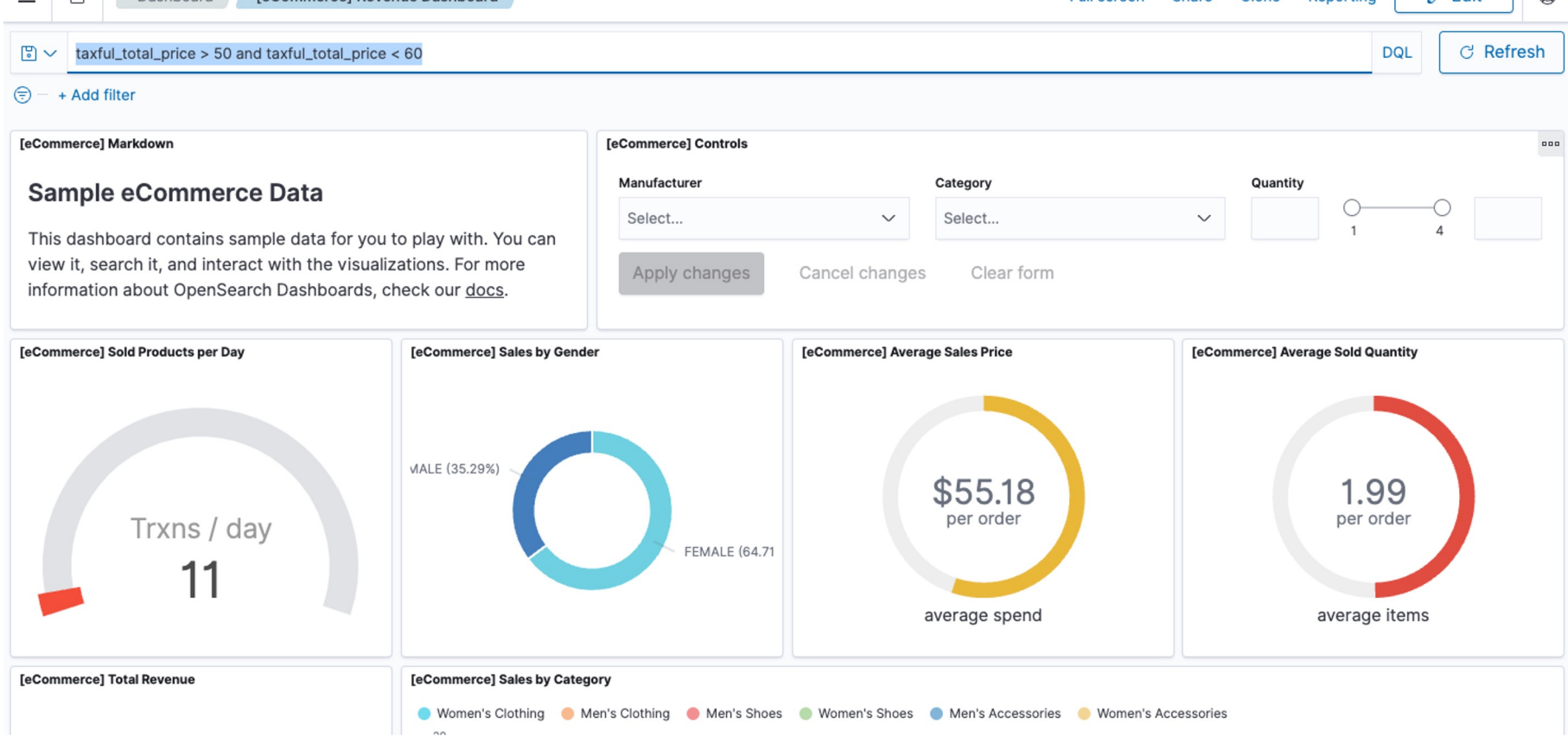




[Logs] Host, Visits and Bytes Table

Type	Bytes (Total)	Bytes (Last Hour)	Unique Visits (Total)	Unique Visits (Last Hour)
	96.2KB	0B	19	0
deb	75.4KB	0B	13	0
gz	63.3KB	0B	9	0
zip	49.9KB	0B	9	0
css	32.1KB	0B	8	0
rpm	16KB	0B	3	0

Web Traffic Sample



eCommerce Sample

OpenSearch for AI

OpenSearch as Vector Database

- Storage of Embeddings
- k-nearest neighbors search
- Use cases
  - Image/Semantic/Multimodal search
  - Recommendation engine
  - Pattern data matching

Data pipelines for OpenSearch


APACHE
kafka
MySQL
OpenSearch

One data platform for your cloud needs

Event streaming	Event stream processing	Relational databases	Key-value database	Wide column database	Data warehouse	Time series database	Search engine	Data visualization
Aiven for **Apache Kafka®** and **Kafka® Connect**	Aiven for **Apache Flink®**	Aiven for **PostgreSQL®**, Aiven for **MySQL**	Aiven for **Redis®**	Aiven for **Apache Cassandra®**	Aiven for **ClickHouse®**	Aiven for **M3**	Aiven for **OpenSearch®**	Aiven for **Grafana®**

STREAM | **STORE** | **ANALYZE**

Host

aws, Google Cloud, DigitalOcean, Microsoft Azure, Bring your own cloud

Deploy

Terraform, Kubernetes

API REST API, CLI Aiven CLI, Aiven Console

Integrate

Datadog, Prometheus, AWS CloudWatch, GCP Monitoring, MongoDB, AWS S3, GCP BigQuery

Couchbase, Snowflake, Splunk, Sumologic, Debezium, GCP Pub/Sub, GCP Storage

Customers


okta


snyk


DOORDASH


priceline


fiverr.


Norauto


DECATHLON


GTL


ACTIVISION
BLIZZARD


MIRAKL


GOV.UK


goto financial


spare


Schibsted


TOYOTA


paf


CONRAD


adeo


ometria


WÄRTSILÄ

Try it out!

https://go.aiven.io/reinvent23-os

300$ - 1 Month!