A D B 2 0 1

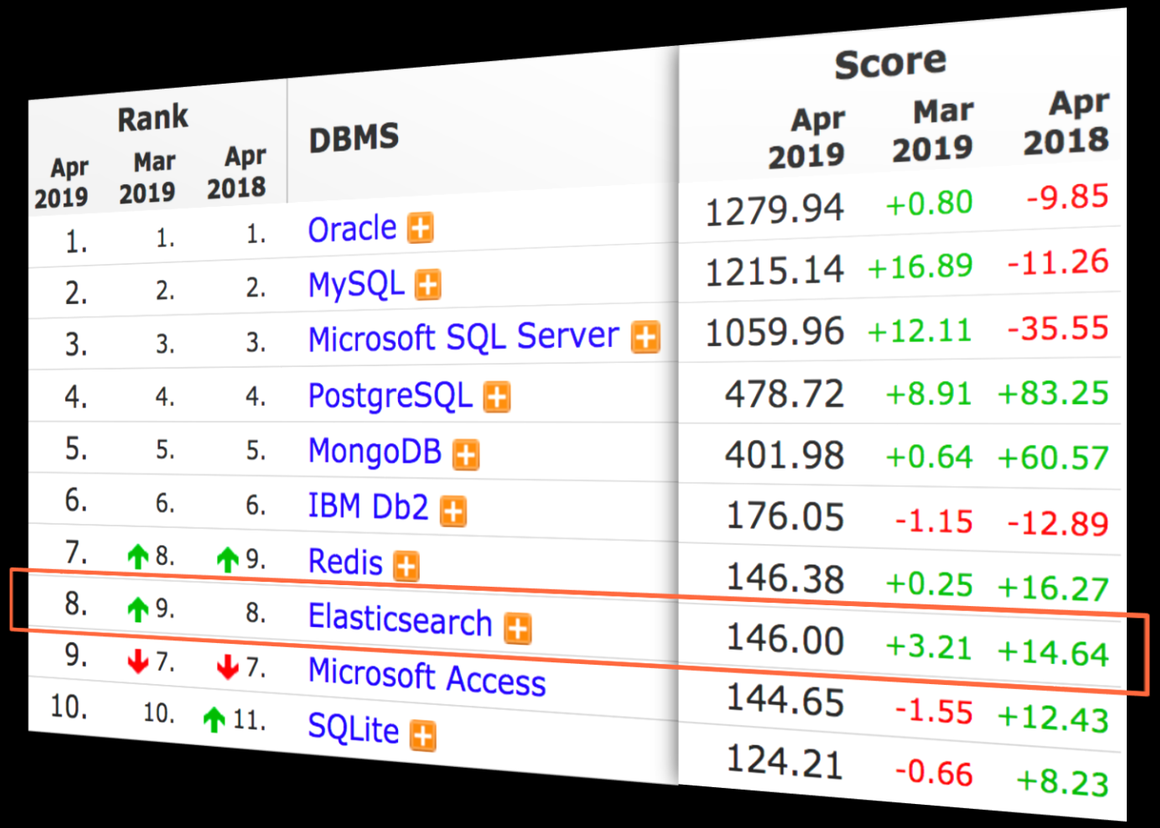
Introducing Open Distro for Elasticsearch

Carl Meadows
Principal Product Manager
AWS – Search Services

aws SUMMIT

What is Elasticsearch

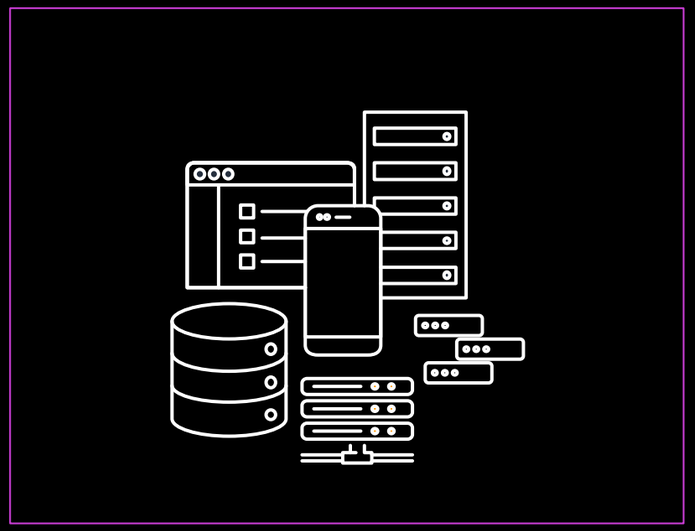
- Sometimes referred to as the “ELK Stack” – Elasticsearch, Logstash & Kibana
- Distributed search and analytics engine built on Apache Lucene
- Easy ingestion and visualization
- Developed in Java



Rank			DBMS	Score		
Apr 2019	Mar 2019	Apr 2018		Apr 2019	Mar 2019	Apr 2018
1.	1.	1.	Oracle	1279.94	+0.80	-9.85
2.	2.	2.	MySQL	1215.14	+16.89	-11.26
3.	3.	3.	Microsoft SQL Server	1059.96	+12.11	-35.55
4.	4.	4.	PostgreSQL	478.72	+8.91	+83.25
5.	5.	5.	MongoDB	401.98	+0.64	+60.57
6.	6.	6.	IBM Db2	176.05	-1.15	-12.89
7.	↑ 8.	↑ 9.	Redis	146.38	+0.25	+16.27
8.	↑ 9.	8.	Elasticsearch	146.00	+3.21	+14.64
9.	↓ 7.	↓ 7.	Microsoft Access	144.65	-1.55	+12.43
10.	10.	↑ 11.	SQLite	124.21	-0.66	+8.23

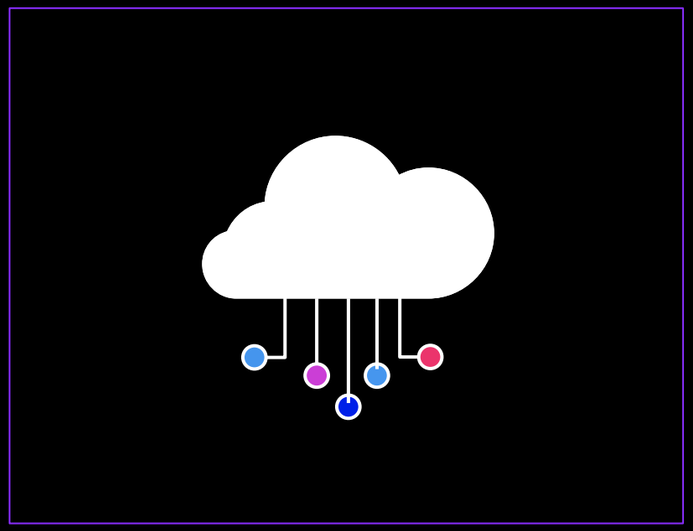
Machine data driving Elasticsearch growth

Machine-generated data is growing **10x faster** than business data Logs, logs, and more logs



IT & DevOps: Databases, servers, storage, networking

Increase in IoT and mobile devices: Gaming, sensors, web content



Cloud-based architectures

Source: insideBigData, “The Exponential Growth of Data,” February 16, 2017

Open Distro
for Elasticsearch

An Apache 2.0-licensed distribution of Elasticsearch enhanced with enterprise-grade security, alerting, SQL, and more

How we think about OSS licensing and distribution?

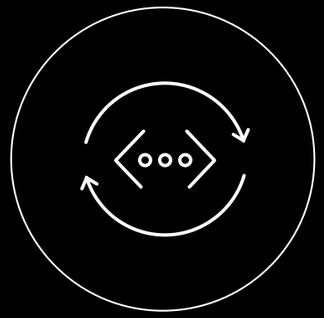
Base OSS free of proprietary code

Keep commercial software on top of OSS separate

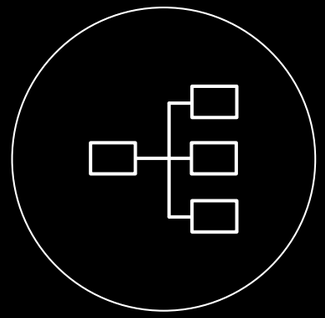
Allow anyone to innovate on OSS

Don’t change licensing or distribution midstream

Benefits of 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

Open Distro for Elasticsearch – Features

Security

Achieve encryption in flight, fine-grained access control, audit logging, and compliance

Alerting

Monitor your data and send automatic alerts on any changes in your data

SQL

Easily interact with your Elasticsearch cluster and extract insights using the familiar SQL query syntax

Performance Analyzer

Get deep visibility into system bottlenecks even when your Elasticsearch cluster is under duress.

Security

Keep your data secure

Encryption

Keep your data secure when in transit

Authentication

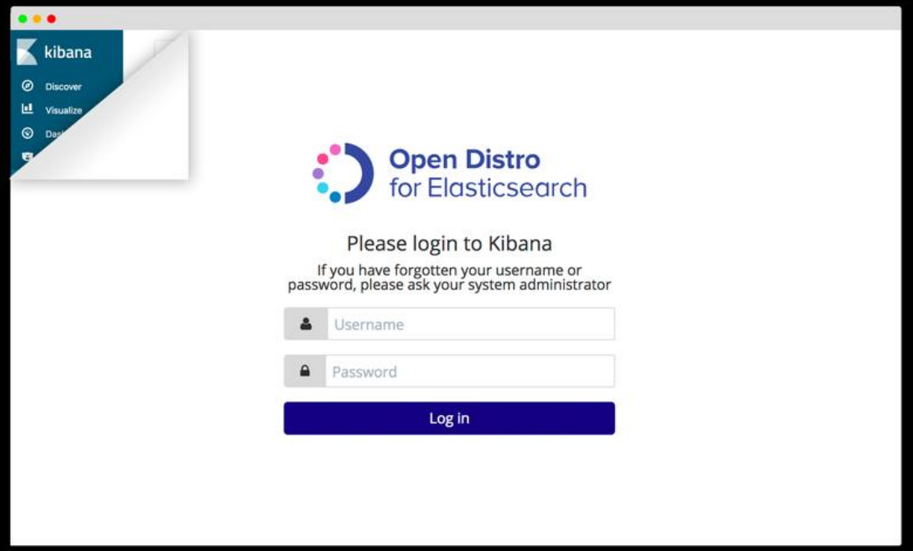
Leverage your existing authentication infrastructure

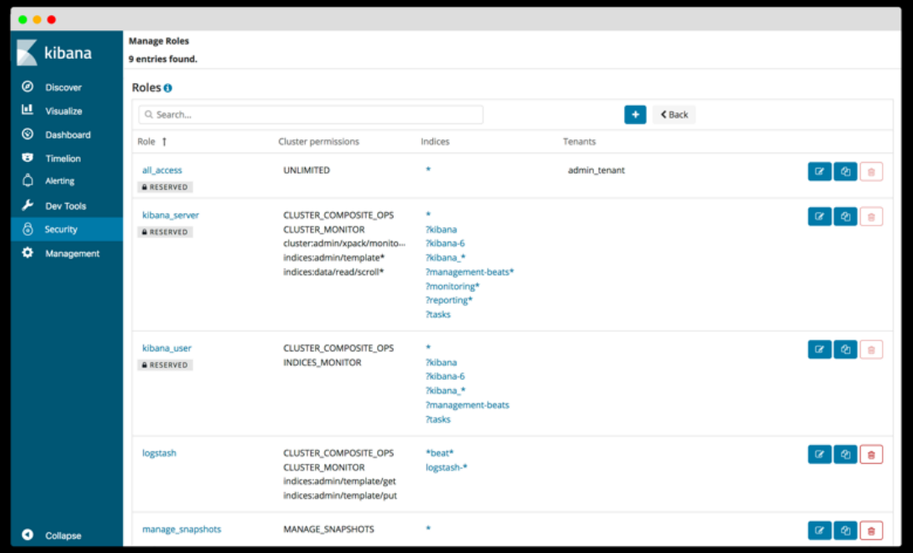
RBAC

Granular access control over user actions on your cluster

Audit logging

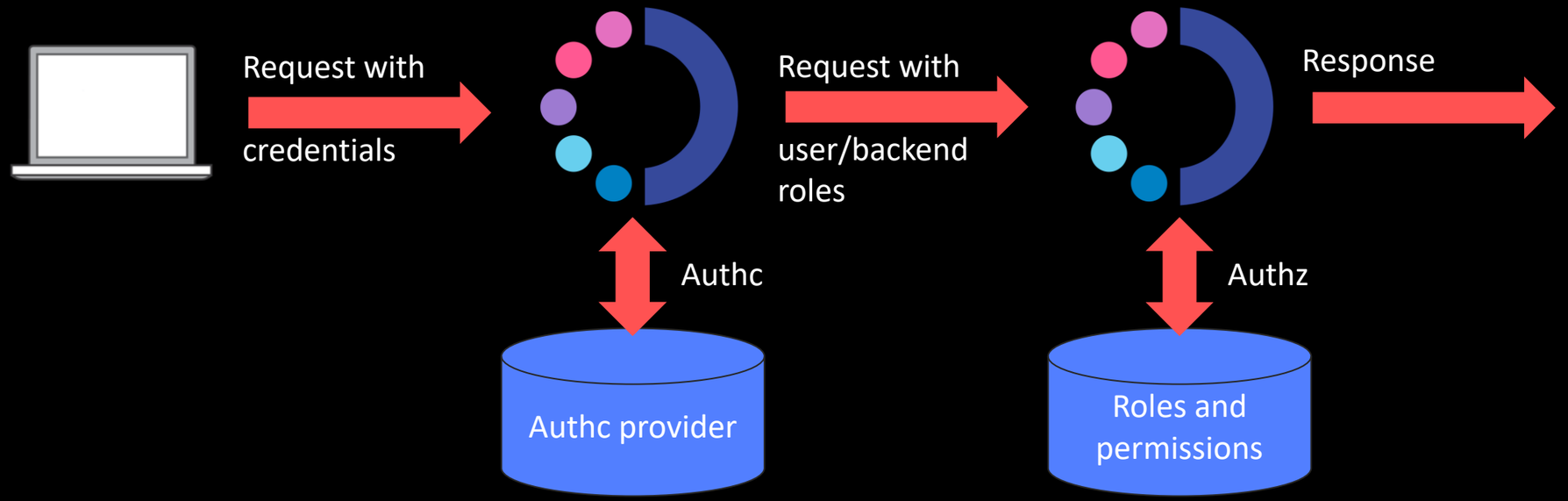
Track and record all user actions and meet HIPAA and PCI compliance





aws SUMMIT

Access control flow for RBAC



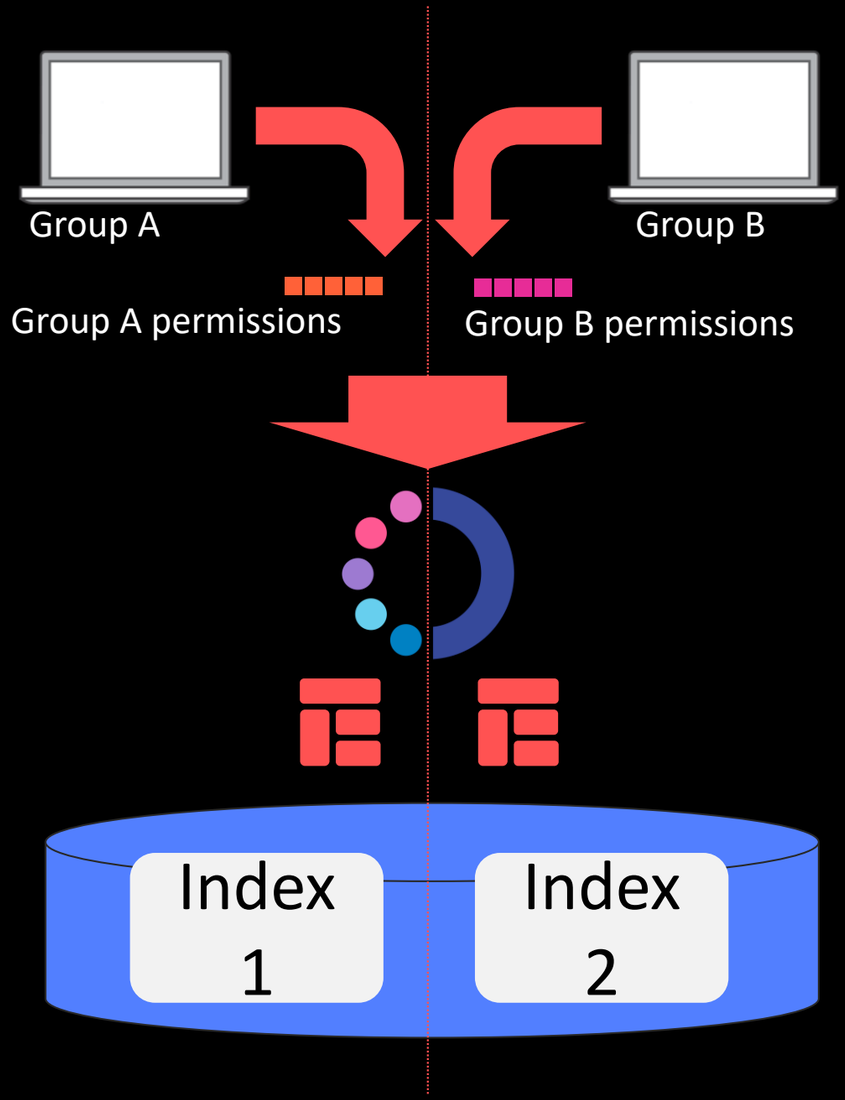
Authc – Via basic HTTP auth, LDAP, AD, SAML, web tokens, SSL

Authz – Backend identities mapped to Open Distro roles

Permissions – Allow a role to perform an action against a cluster/index/document/field

Action groups – Groups of permissions

Kibana multi-tenancy



Alerting

Receive alerts on your data

Create monitors

Query the data you want to and receive alerts on it

Customize alert conditions

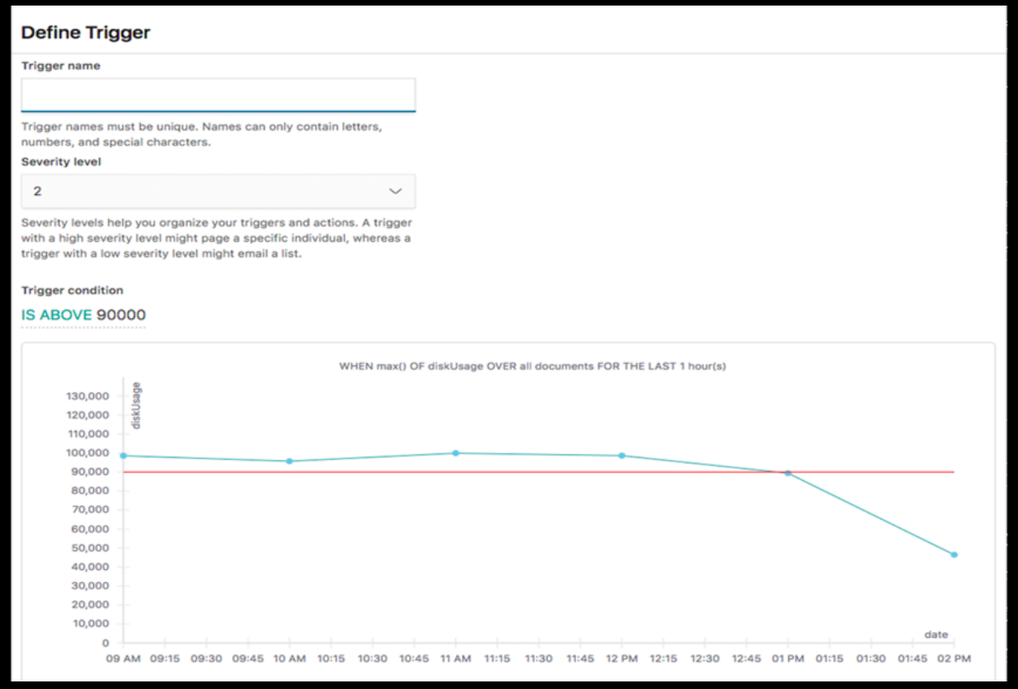
Define alerting threshold and severity for multiple trigger conditions

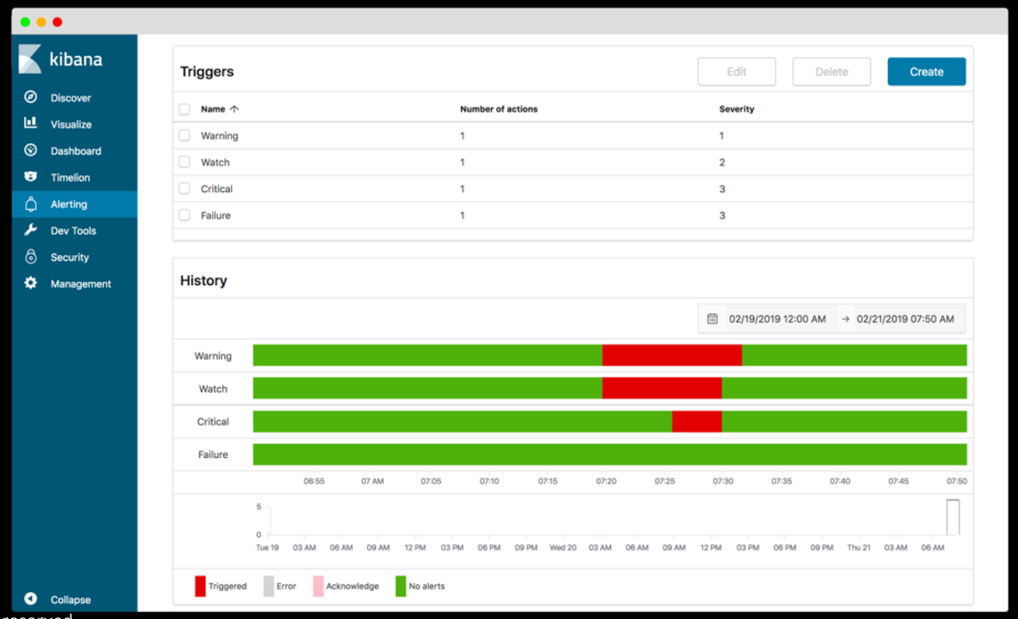
Get notifications

Built-in integrations for webhook and Slack to get notified on the channels you use

View alerts

All alert executions are indexed for easy tracking and visualization





aws SUMMIT

SQL Support

Query data with SQL

Comprehensive SQL support

Supports over 40 functions, data types, and commands, including join support

Translate SQL to JSON

Create JSON using SQL to configure sophisticated access control policies

Use existing tools

Provides a JDBC driver so you can use a variety of business intelligence, analytics, and ETL tools

```
# Get all accounts with JSON document response
GET _sql
{
  "query": "SELECT * FROM accounts"
}

# Get all accounts with CSV response
GET _sql?format=csv
{
  "query": "SELECT * FROM accounts"
}

# Get average age of employees
GET _sql?format=csv
{
  "query": "SELECT AVG(age) as avg, employer,
state, city FROM accounts GROUP BY employer
.keyword, state.keyword, city.keyword"
}

# Compase SQL query to Elasticsearch query DSL
GET _sql/_explain
{
  "query": "SELECT AVG(age) as avg, employer,
state, city FROM accounts GROUP BY employer
.keyword, state.keyword, city.keyword"
}
```

index	id	score	account_number	balance	firstname	lastname	age	gen
accounts	25	1	25	$ 40,540.00	Virginia	Ayala	39	F
accounts	44	1	44	$ 34,487.00	Aurelia	Harding	37	M
accounts	99	1	99	$ 47,159.00	Ratliff	Heath	39	F
accounts	119	1	119	$ 49,222.00	Laverne	Johnson	28	F
accounts	126	1	126	$ 3,607.00	Effie	Gates	39	F
accounts	145	1	145	$ 47,406.00	Rowena	Wilkinson	32	M
accounts	183	1	183	$ 14,223.00	Hudson	English	26	F
accounts	190	1	190	$ 3,150.00	Blake	Davidson	30	F
accounts	208	1	208	$ 40,760.00	Garcia	Hess	26	F
accounts	222	1	222	$ 14,764.00	Rachelle	Rice	36	M
accounts	227	1	227	$ 19,780.00	Coleman	Berg	22	M
accounts	253	1	253	$ 20,240.00	Melissa	Gould	31	M
accounts	260	1	260	$ 2,726.00	Kari	Skinner	30	F
accounts	265	1	265	$ 46,910.00	Marion	Schneider	26	F
accounts	335	1	335	$ 35,433.00	Vera	Hansen	24	M
accounts	366	1	366	$ 42,368.00	Lydia	Cooke	31	M
accounts	385	1	385	$ 11,022.00	Rosalinda	Valencia	22	M
accounts	397	1	397	$ 37,418.00	Leonard	Gray	36	F
accounts	400	1	400	$ 20,685.00	Kane	King	21	F
accounts	450	1	450	$ 2,643.00	Bradford	Nielsen	25	M
accounts	486	1	486	$ 35,902.00	Dixie	Fuentes	22	F

```
SELECT Avg(age) AS avg,
       employer,
       state,
       city
FROM   accounts
GROUP  BY employer.keyword,
          state.keyword,
          city.keyword
```

```
{
  "from": 0,
  "size": 0,
  "_source": {
    "includes": [
      "AVG",
      "employer",
      "state",
      "city"
    ],
    "excludes": []
  },
  "stored_fields": [
    "employer",
    "state",
    "city"
  ],
  "aggregations": {
    "employer.keyword": {
      "terms": {
        "field": "employer.keyword",
        "size": 200,
        "min_doc_count": 1,
```

aws SUMMIT

Performance Analyzer

Get deep diagnostic insights into your cluster

Identify bottlenecks across the stack

Provides a powerful REST API for querying Elasticsearch metrics to diagnose issues across stack

Runs independent of your cluster

Perform diagnostics even if the cluster is under duress

Analyze hundreds of data points

Supports over 60 metrics across 10 dimensions for instrumentation of your cluster health

aws SUMMIT

PerfTop CLI

- Provides pre-configured dashboards for analyzing cluster, node, and shard performance
- Custom JSON templates to create the dashboards to diagnose your cluster performance

Flexible deployment options

- Docker
- RPM
- Debian

Simple to get started

1

Visit the website

2

Download the Elasticsearch and Kibana packages

3

Load and query data

Community and contributions

Open Distro for Elasticsearch's success is driven by the community's participation, contributions, and innovation to the project.

You can follow project discussions, engage with fellow community members, contribute PRs, file bugs, or request a feature at:

Discussion forums
https://discuss.opendistrocommunity.dev/

Community
https://github.com/opendistro-for-elasticsearch/community/issues

Useful links

Project website and technical documentation
https://opendistro.github.io/for-elasticsearch/

Source Code
https://github.com/opendistro-for-elasticsearch

Thank you!

Carl Meadows
carlmead@amazon.com

aws SUMMIT