OpenSearch: The Open Source Path to Search and Observability

Dotan Horovits
@horovits

Anirudha Jadhav
Linkedin/in/ianirudha

Pratik Shenoy
Linkedin/in/pshenoy36

AGENDA

Tell them what you're gonna tell them

- **What is OpenSearch**
- **Search and Vectors**
- **Opensearch Observability-Stack**
- **Demos !!!**
  - **Agentic AI Observability**
  - **Native** prometheus **support – PromQL in action**
  - **Application performance monitoring, Tracing**
  - **OpenSearch Observability and OpenTelemetry**
- **Roadmap**

OpenSearch by the Numbers

130+ GitHub repositories

More than **90+** Project partners

28 New releases since launch

More than **1,300,000,000** Project downloads

1,400,000 Monthly page views for opensearch.org

4,000+ Members of public Slack workspace

7,000+ Members of User Forums

3,000+ Active contributors to project code

400+ Organizations actively contributing

600,000+ forum views per month

OpenSearch
SOFTWARE FOUNDATION

THE
LINUX
FOUNDATION

OpenSearch

Premier Members

- aws
- IBM®
- SAP®
- Uber

General Members

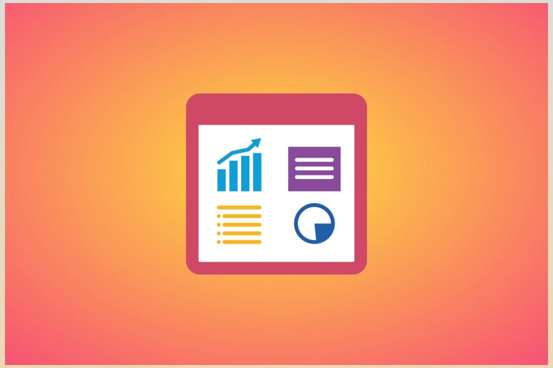
- aiven
- Aryn
- ByteDance
- Canonical
- DigitalOcean
- DTEX
-
- graylog
- NetApp Instaclustr
- OpenSource Connections
- portal26
- reSolve
- seacom
- BIGDATA BOUTIQUE

The OpenSearch Platform

OpenSearch Core

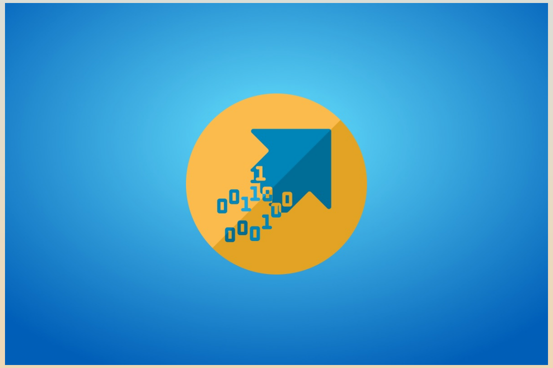
Powerful distributed search and analytics engine with integrated vector database

OpenSearch Dashboards



Open-source data visualization toolset and UI for OpenSearch

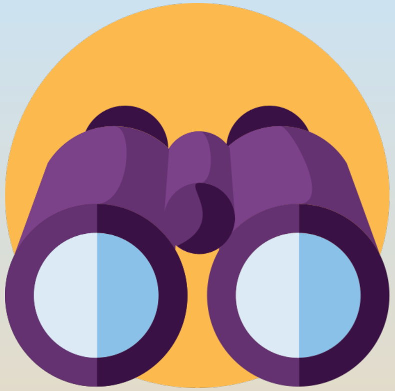
Data Prepper



Light-weight data collector for enriching, transforming, and aggregating data

How do people use OpenSearch?

Search



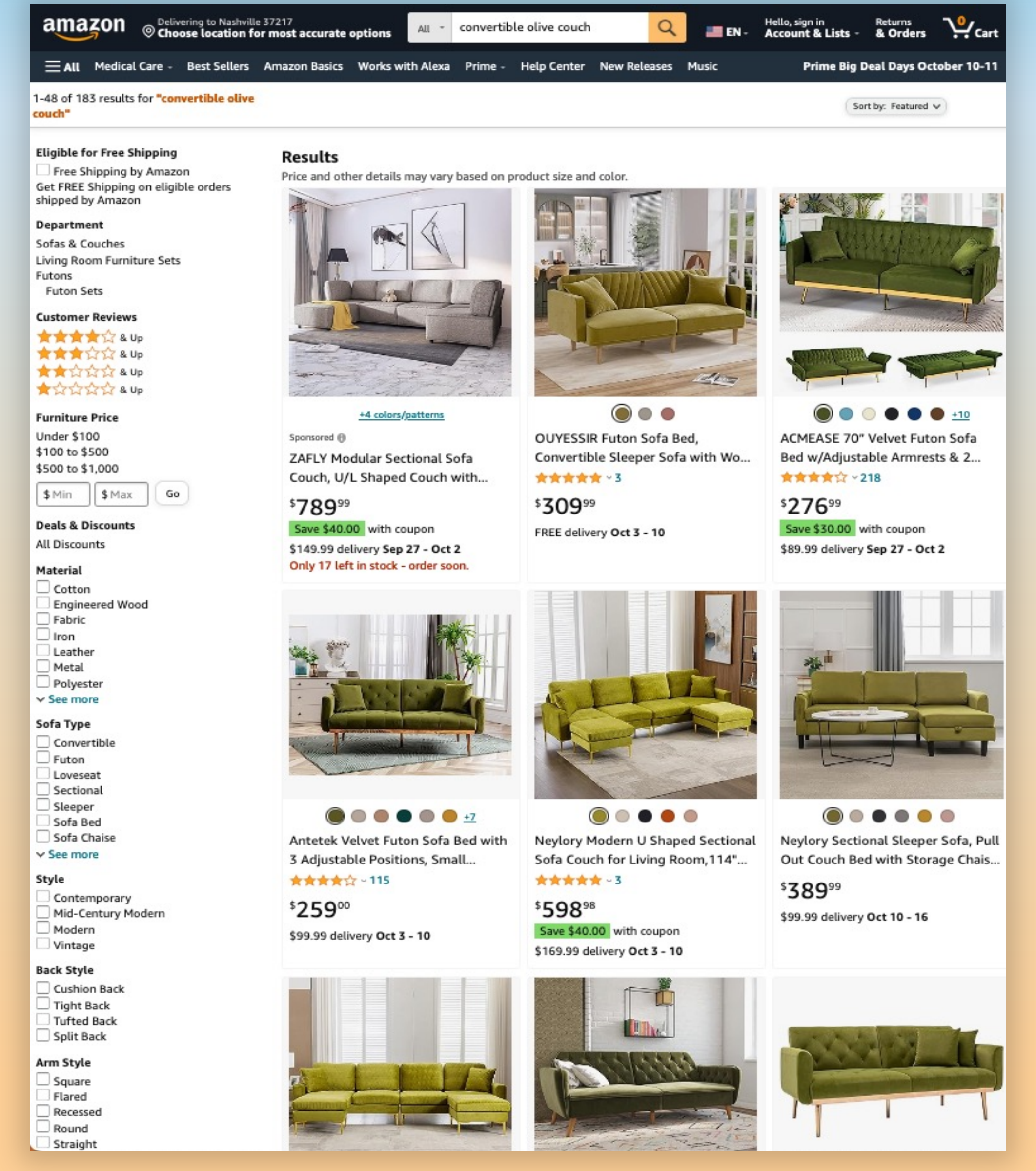
Observability

AI/ML applications

Security analytics

Search

- Free text search (unstructured)
- Faceting
- Geospatial
- Auto-complete
- Fuzzy matching
- Lexical and vector search



convertible olive couch

1-48 of 183 results for "convertible olive couch"

Eligible for Free Shipping

□ Free Shipping by Amazon

Get FREE Shipping on eligible orders shipped by Amazon

Department

Sofas & Couches
Living Room Furniture Sets
Futons
Futon Sets

Customer Reviews

★★★★☆ & Up
★★★☆☆ & Up
★★☆☆☆ & Up
★☆☆☆☆ & Up

Furniture Price

Under $100
$100 to $500
$500 to $1,000

Deals & Discounts

All Discounts

Material

□ Cotton
□ Engineered Wood
□ Fabric
□ Iron
□ Leather
□ Metal
□ Polyester

Sofa Type

□ Convertible
□ Futon
□ Loveseat
□ Sectional
□ Sleeper
□ Sofa Bed
□ Sofa Chaise

Style

□ Contemporary
□ Mid-Century Modern
□ Modern
□ Vintage

Back Style

□ Cushion Back
□ Tight Back
□ Tufted Back
□ Split Back

Arm Style

□ Square
□ Flared
□ Recessed
□ Round
□ Straight

Results

Price and other details may vary based on product size and color.

Sponsored
ZAFLY Modular Sectional Sofa Couch, U/L Shaped Couch with...
$\$789^{99}$
Save $40.00 with coupon
$149.99 delivery **Sep 27 - Oct 2**
Only 17 left in stock - order soon.

OUYESSIR Futon Sofa Bed, Convertible Sleeper Sofa with Wo...
★★★★★ 3
$\$309^{99}$
FREE delivery **Oct 3 - 10**

ACMEASE 70" Velvet Futon Sofa Bed w/Adjustable Armrests & 2...
★★★★☆ 218
$\$276^{99}$
Save $30.00 with coupon
$89.99 delivery **Sep 27 - Oct 2**

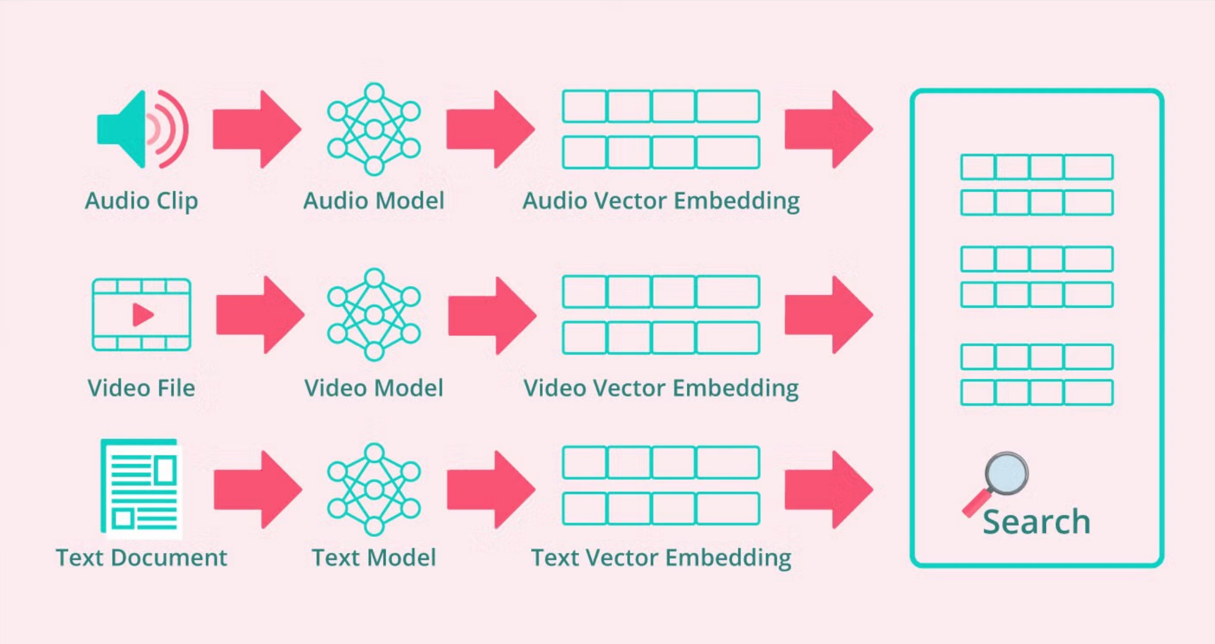
Antetek Velvet Futon Sofa Bed with 3 Adjustable Positions, Small...
★★★★☆ 115
$\$259^{00}$
$99.99 delivery **Oct 3 - 10**

Neylory Modern U Shaped Sectional Sofa Couch for Living Room,114"...
★★★★★ 3
$\$598^{98}$
Save $40.00 with coupon
$169.99 delivery **Oct 3 - 10**

Neylory Sectional Sleeper Sofa, Pull Out Couch Bed with Storage Chais...
$\$389^{99}$
$99.99 delivery **Oct 10 - 16**

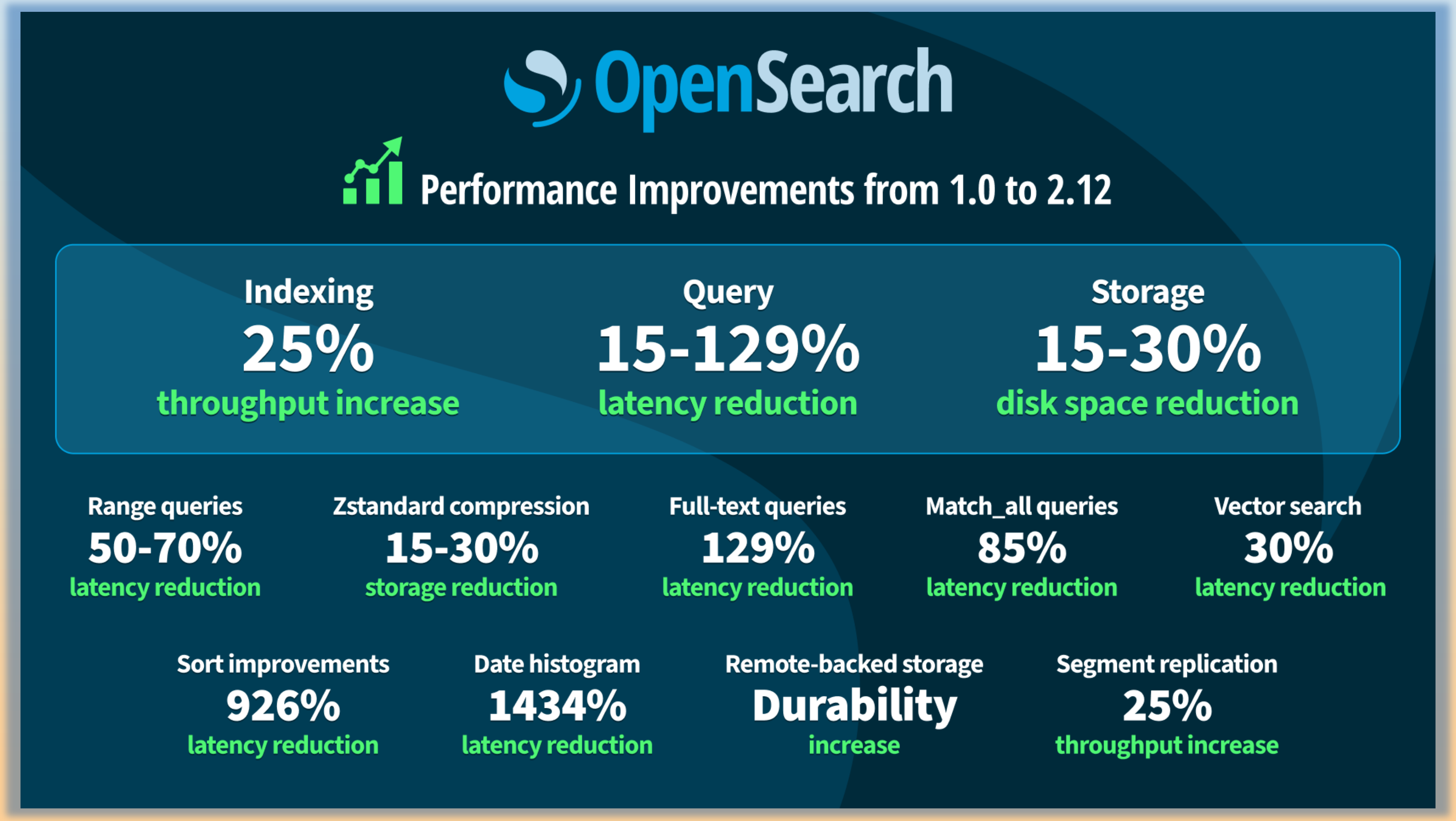
AI / ML Applications

- Full-fledged vector database
- Similarity search with exact or approximate k-NN algorithms (HNSW, IVF)
- Multimodal search – text, image, video
- Neural sparse search
- Bring your own LLM
- MCP support for agentic AI



OpenSearch

Performance improvements



OpenSearch

Performance Improvements from 1.0 to 2.12

Indexing
25%
throughput increase

Query
15-129%
latency reduction

Storage
15-30%
disk space reduction

Range queries
50-70%
latency reduction

Zstandard compression
15-30%
storage reduction

Full-text queries
129%
latency reduction

Match_all queries
85%
latency reduction

Vector search
30%
latency reduction

Sort improvements
926%
latency reduction

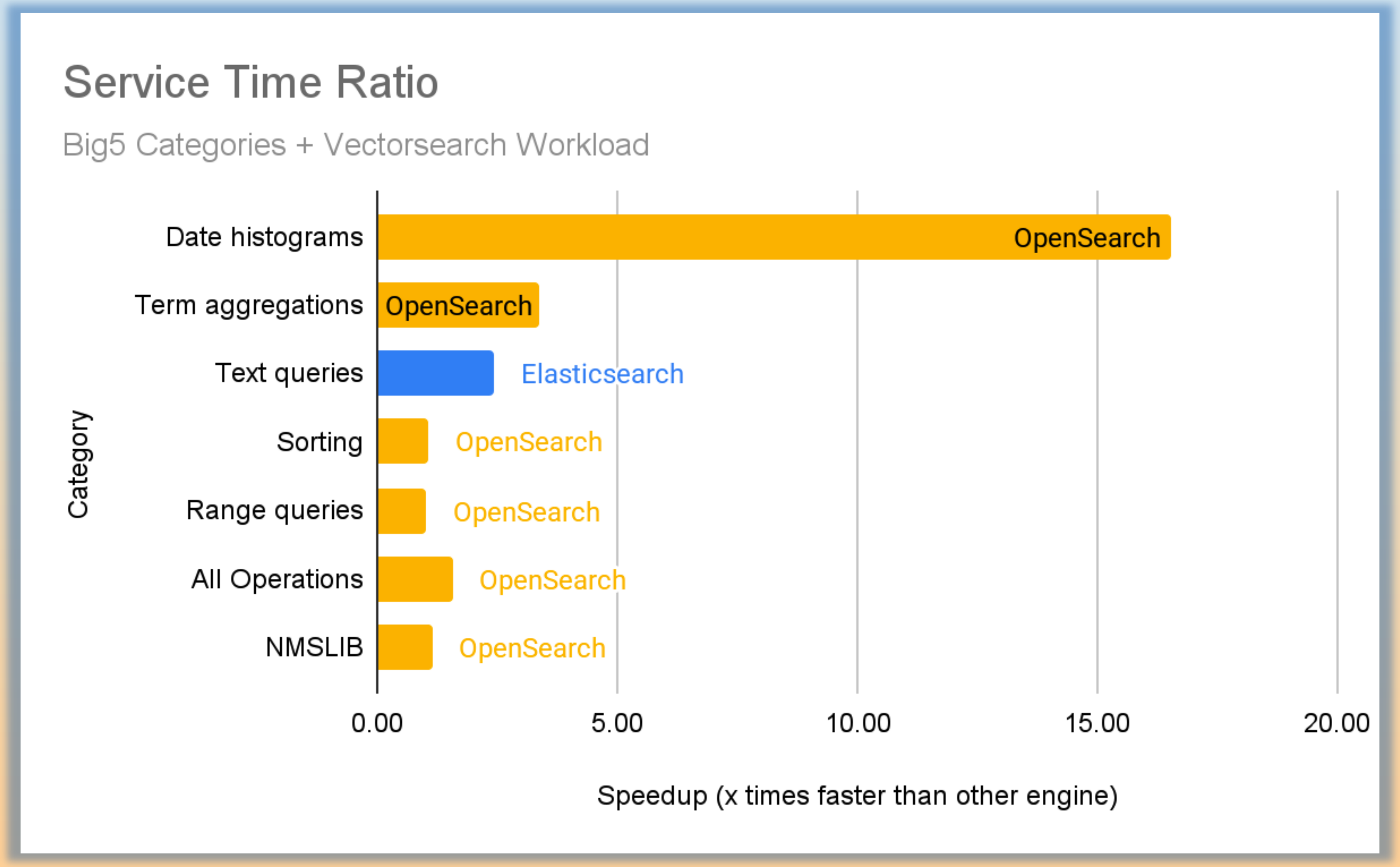
Date histogram
1434%
latency reduction

Remote-backed storage
Durability
increase

Segment replication
25%
throughput increase

Performance improvements

"During our assessment, we observed that OpenSearch v2.17.1 is **1.6x faster** on the Big5 workload and **11% faster** on the Vectorsearch workload…"



https://blog.trailofbits.com/2025/03/06/benchmarking-opensearch-and-elasticsearch/

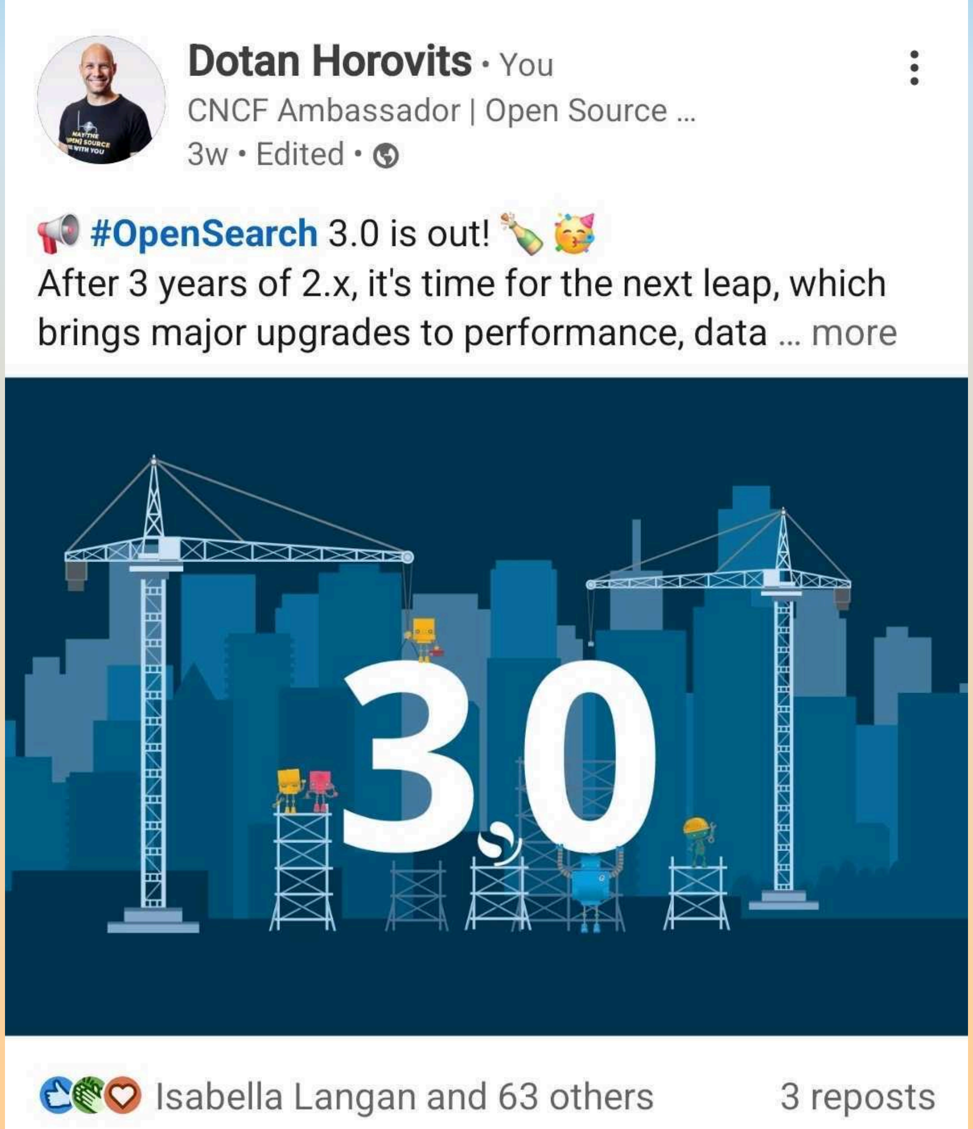
OpenSearch

UPDATES

OpenSearch v3

May ’25 v3.0 was released

- First major release since 2022



Dotan Horovits • You

CNCF Ambassador | Open Source ...

3w • Edited •

📢 #OpenSearch 3.0 is out! 🍾 🥳

After 3 years of 2.x, it's time for the next leap, which brings major upgrades to performance, data ... more

Isabella Langan and 63 others

3 reposts

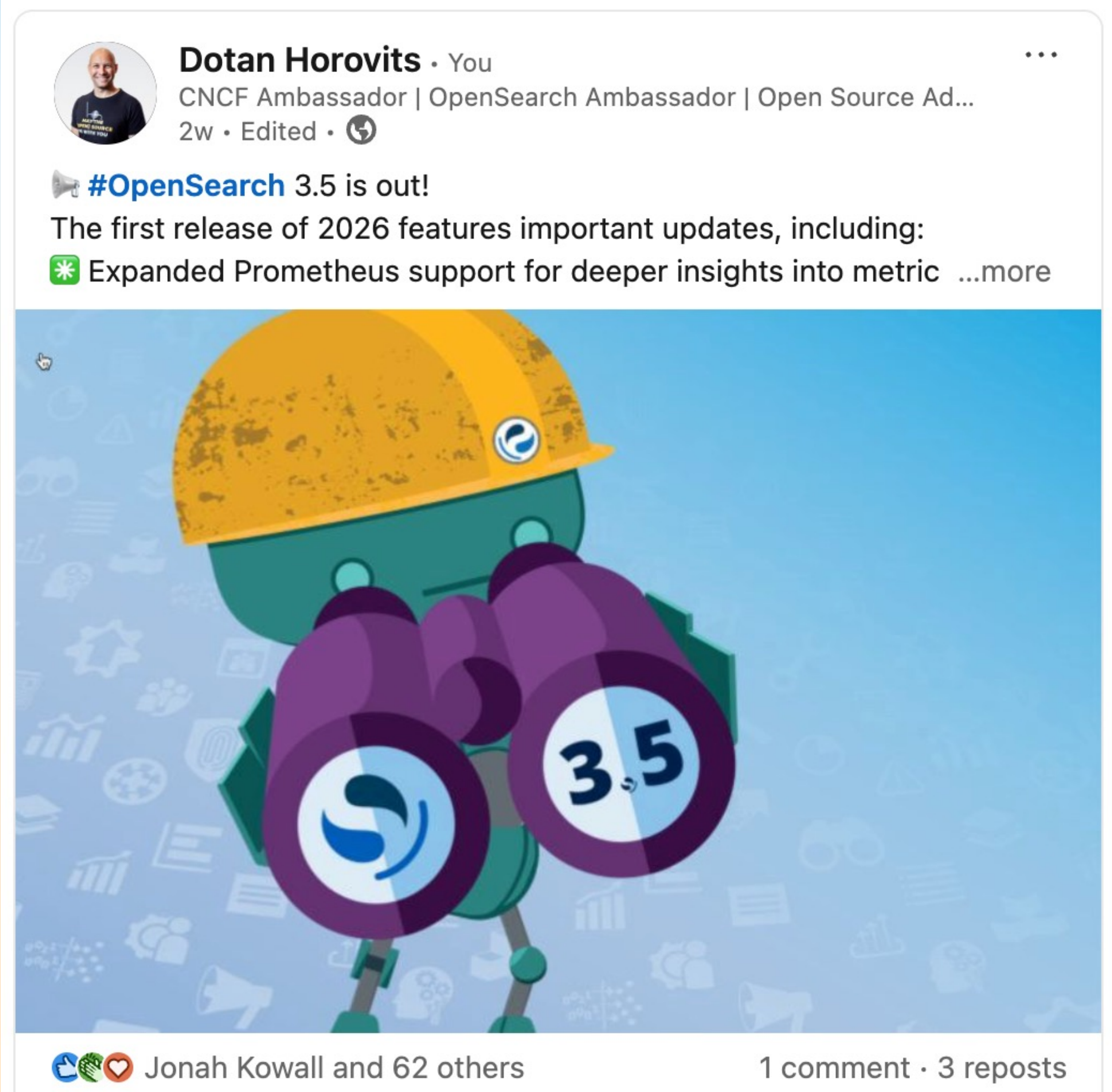
OpenSearch v3

May ’25 v3.0 was released

- First major release since 2022

A few weeks ago (Feb 10, ‘26) we’ve released v3.5

And they’re big!



Dotan Horovits • You
CNCF Ambassador | OpenSearch Ambassador | Open Source Ad...
2w • Edited •

📣 #OpenSearch 3.5 is out!
The first release of 2026 features important updates, including:
✳️ Expanded Prometheus support for deeper insights into metric ...more

Jonah Kowall and 62 others
1 comment • 3 reposts

OpenSearch v3

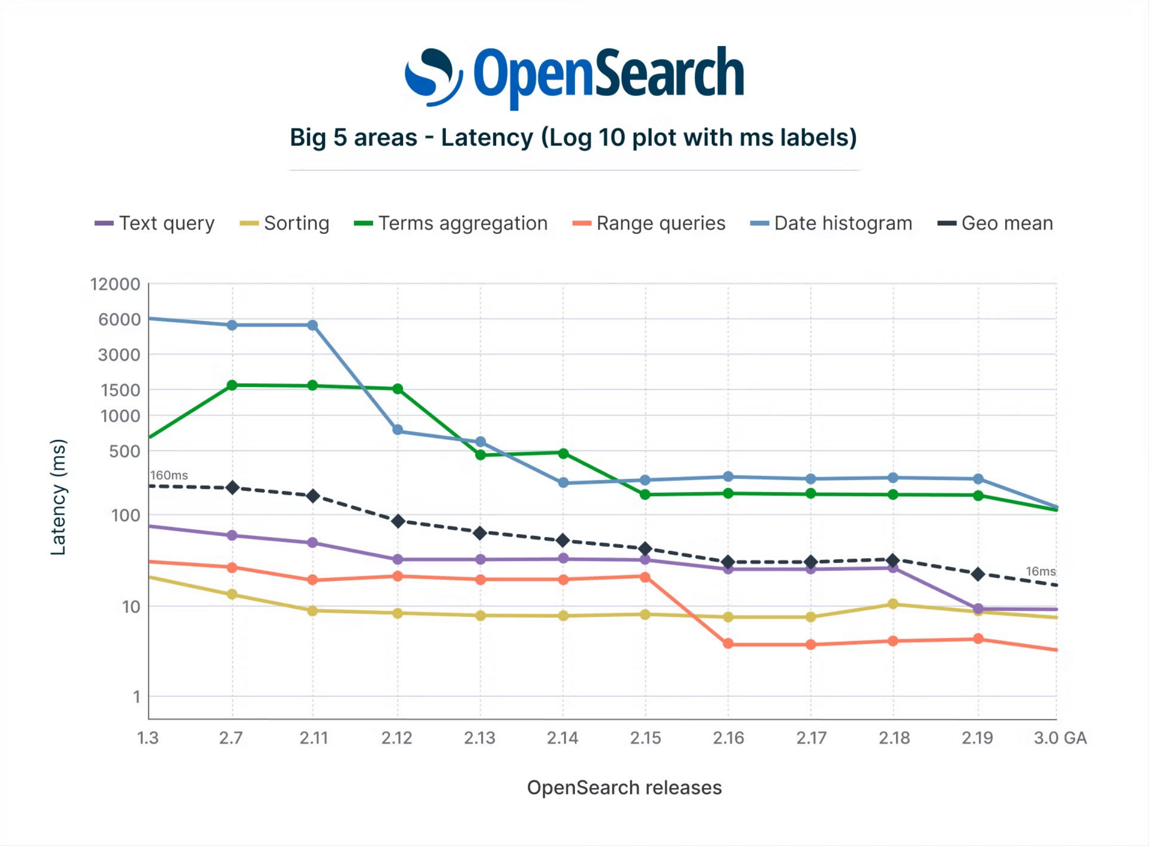
Unlocking new levels of performance and capability

- Advanced data management
    - Reader / Writer separation for improved resource utilization
    - gRPC / Protobuf for faster data transport
    - Pull-based ingestion for more resilient data pipelines

OpenSearch v3

Unlocking new levels of performance and capability

- Accelerated performance
  - **90% reduction** in query latency
  - **9.5x faster** for key query types
  - **9.3x** faster index builds

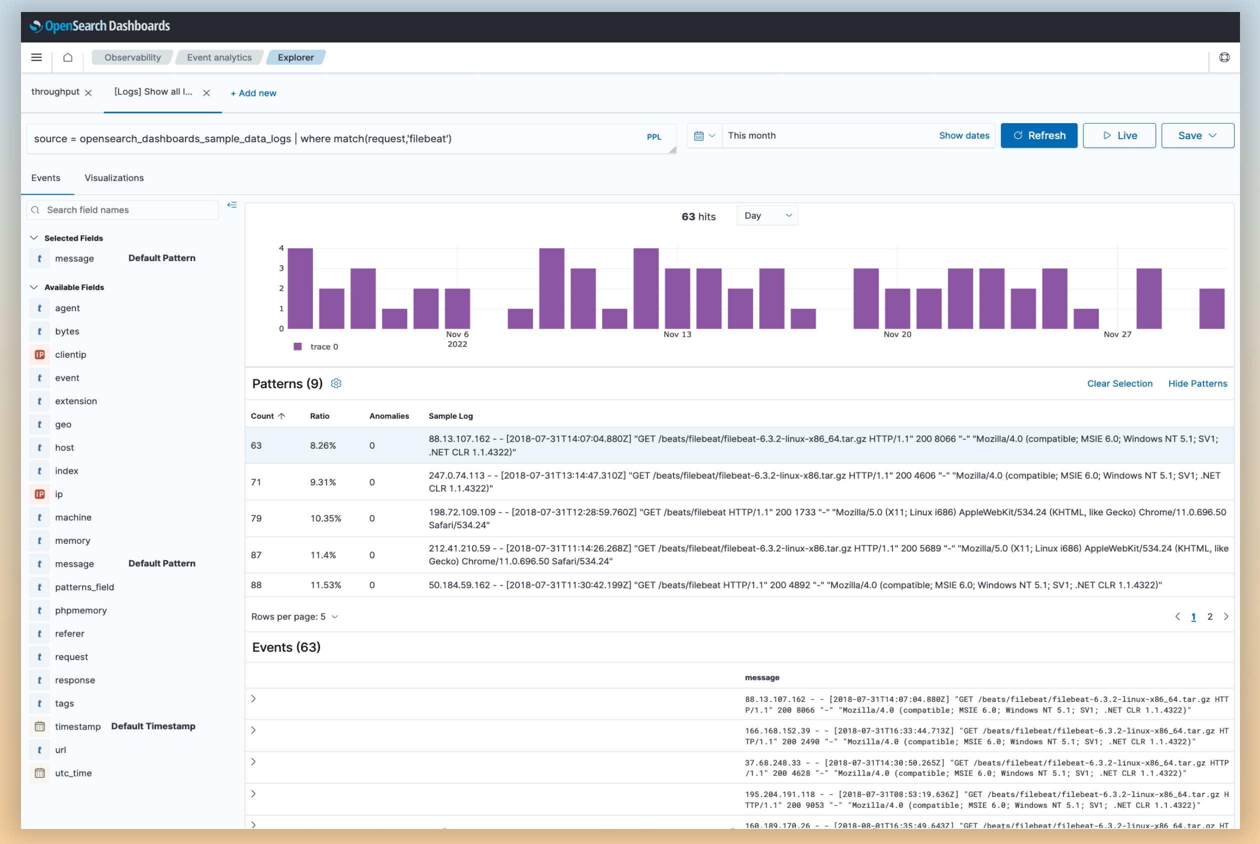


Performance of 3.0 compared to OpenSearch 1.3

OpenSearch

Observability

- Log analytics, Metrics, Traces
- Lucene & PPL query languages
- Natural language → PPL translator
- Support for *OpenTelemetry* data
- Extend tracing with *Jaeger*
- Query federation to *Prometheus*



```
source = opensearch_dashboards_sample_data_logs | where match(request,'filebeat')
```

Selected Fields

message	Default Pattern

Available Fields

agent	
bytes	
clientip	
event	
extension	
geo	
host	
index	
ip	
machine	
memory	
message	Default Pattern
patterns_field	
phpmemory	
referer	
request	
response	
tags	
timestamp	Default Timestamp
url	
utc_time	

Patterns (9)

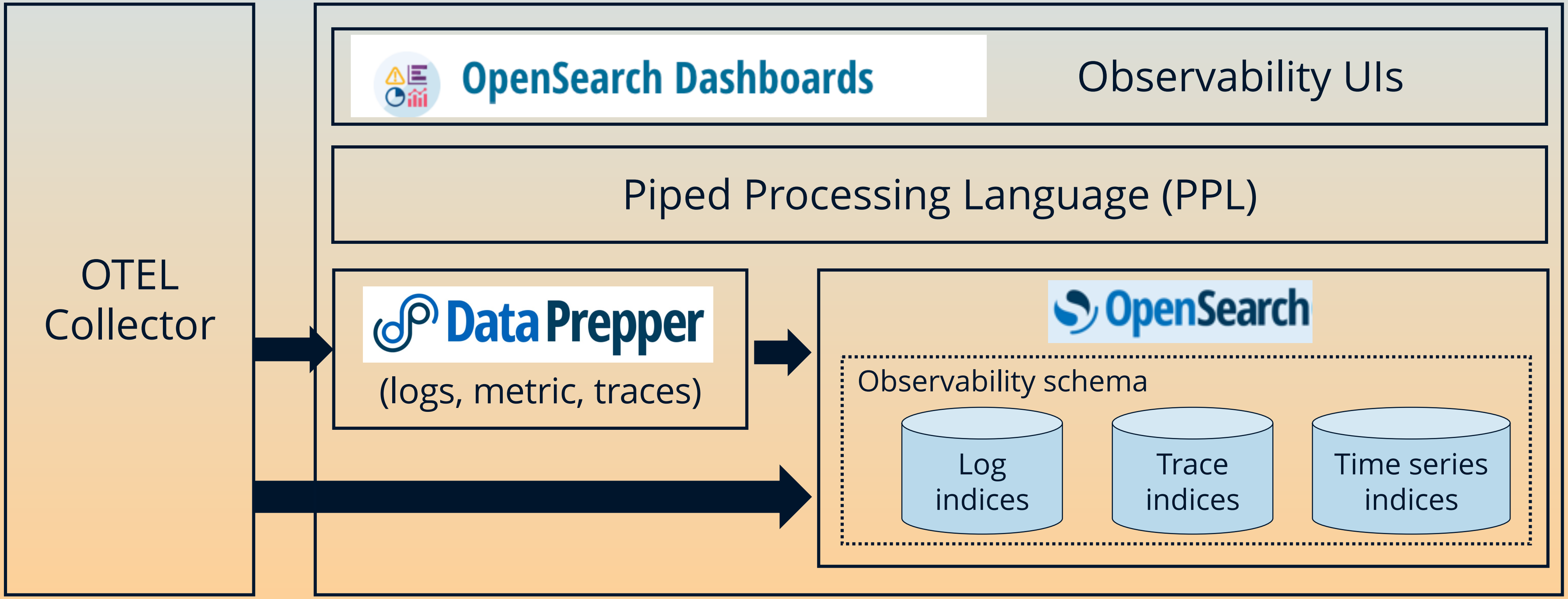
Count ↑	Ratio	Anomalies	Sample Log
63	8.26%	0	88.13.107.162 - - [2018-07-31T14:07:04.880Z] "GET /beats/filebeat/filebeat-6.3.2-linux-x86_64.tar.gz HTTP/1.1" 200 8066 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"
71	9.31%	0	247.0.74.113 - - [2018-07-31T13:14:47.310Z] "GET /beats/filebeat/filebeat-6.3.2-linux-x86.tar.gz HTTP/1.1" 200 4606 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"
79	10.35%	0	198.72.109.109 - - [2018-07-31T12:28:59.760Z] "GET /beats/filebeat HTTP/1.1" 200 1733 "-" "Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24"
87	11.4%	0	212.41.210.59 - - [2018-07-31T11:14:26.268Z] "GET /beats/filebeat/filebeat-6.3.2-linux-x86.tar.gz HTTP/1.1" 200 5689 "-" "Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24"
88	11.53%	0	50.184.59.162 - - [2018-07-31T11:30:42.199Z] "GET /beats/filebeat HTTP/1.1" 200 4892 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"

Events (63)

```
message
88.13.107.162 - - [2018-07-31T14:07:04.880Z] "GET /beats/filebeat/filebeat-6.3.2-linux-x86_64.tar.gz HTT
P/1.1" 200 8066 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"
166.168.152.39 - - [2018-07-31T16:33:44.713Z] "GET /beats/filebeat/filebeat-6.3.2-linux-x86_64.tar.gz HT
TP/1.1" 200 2490 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"
37.68.248.33 - - [2018-07-31T14:30:50.265Z] "GET /beats/filebeat/filebeat-6.3.2-linux-x86_64.tar.gz HTTP
/1.1" 200 4628 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"
195.204.191.118 - - [2018-07-31T08:53:19.636Z] "GET /beats/filebeat/filebeat-6.3.2-linux-x86_64.tar.gz H
TTP/1.1" 200 9053 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"
160.189.170.26 - - [2018-08-01T16:35:49.643Z] "GET /beats/filebeat/filebeat-6.3.2-linux-x86_64.tar.gz HT
```

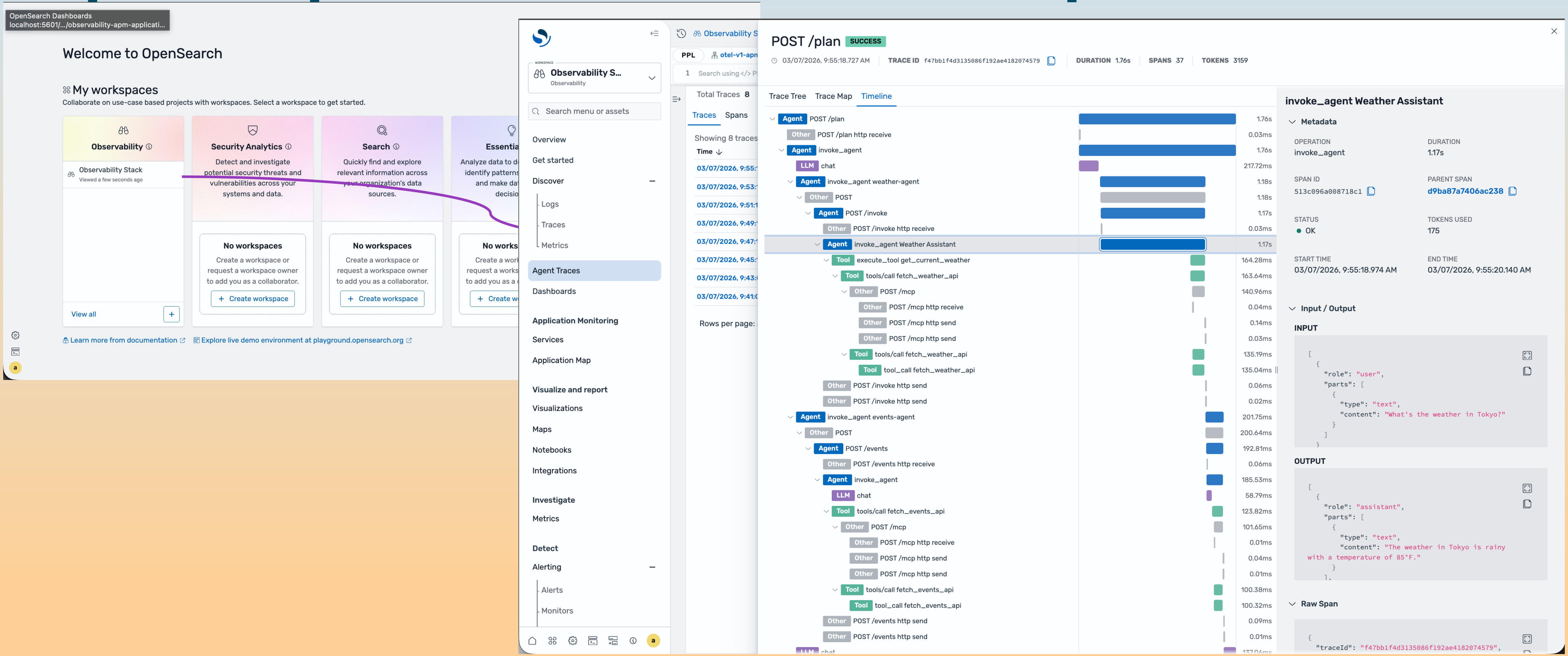
Observability architecture

Open source | Open standards | Customizable



OpenSearch v3

Improved OpenSearch Dashboards User Experience



OpenSearch Dashboards
localhost:5601/.../observability-apm-applicati...

Welcome to OpenSearch

My workspaces

Collaborate on use-case based projects with workspaces. Select a workspace to get started.

Observability

Observability Stack
Viewed a few seconds ago

View all

Security Analytics

Detect and investigate potential security threats and vulnerabilities across your systems and data.

No workspaces

Create a workspace or request a workspace owner to add you as a collaborator.

+ Create workspace

Search

Quickly find and explore relevant information across your organization's data sources.

No workspaces

Create a workspace or request a workspace owner to add you as a collaborator.

+ Create workspace

Essentia

Analyze data to d identify patterns and make dat decisio

No works

Create a work request a works to add you as a

+ Create w

Learn more from documentation

Explore live demo environment at playground.opensearch.org

Observability S

PPL otel-v1-apm

1 Search using </> Pl

Total Traces 8

Traces Spans

Showing 8 traces

Time

03/07/2026, 9:55:
03/07/2026, 9:53:
03/07/2026, 9:51:1
03/07/2026, 9:49:
03/07/2026, 9:47:
03/07/2026, 9:45:
03/07/2026, 9:43:
03/07/2026, 9:41:0

Rows per page:

POST /plan SUCCESS

03/07/2026, 9:55:18.727 AM | TRACE ID f47bb1f4d3135086f192ae4182074579 | DURATION 1.76s | SPANS 37 | TOKENS 3159

Trace Tree | Trace Map | Timeline

```
Agent POST /plan                                  1.76s
  Other POST /plan http receive                   0.03ms
  Agent invoke_agent                              1.76s
    LLM chat                                      217.72ms
    Agent invoke_agent weather-agent              1.18s
      Other POST                                  1.18s
        Agent POST /invoke                        1.17s
          Other POST /invoke http receive         0.03ms
          Agent invoke_agent Weather Assistant    1.17s
            Tool execute_tool get_current_weather 164.28ms
              Tool tools/call fetch_weather_api   163.64ms
                Other POST /mcp                   140.96ms
                  Other POST /mcp http receive    0.04ms
                  Other POST /mcp http send       0.14ms
                  Other POST /mcp http send       0.03ms
                Tool tools/call fetch_weather_api 135.19ms
                  Tool tool_call fetch_weather_api 135.04ms
          Other POST /invoke http send            0.06ms
          Other POST /invoke http send            0.02ms
    Agent invoke_agent events-agent               201.75ms
      Other POST                                  200.64ms
        Agent POST /events                        192.81ms
          Other POST /events http receive         0.06ms
          Agent invoke_agent                      185.53ms
            LLM chat                              58.79ms
            Tool tools/call fetch_events_api      123.82ms
              Other POST /mcp                     101.65ms
                Other POST /mcp http receive      0.01ms
                Other POST /mcp http send         0.04ms
                Other POST /mcp http send         0.01ms
              Tool tools/call fetch_events_api    100.38ms
                Tool tool_call fetch_events_api   100.32ms
          Other POST /events http send            0.09ms
          Other POST /events http send            0.01ms
    LLM chat                                      137.06ms
```

invoke_agent Weather Assistant

Metadata

OPERATION invoke_agent	DURATION 1.17s
SPAN ID 513c096a008718c1	PARENT SPAN d9ba87a7406ac238
STATUS ● OK	TOKENS USED 175
START TIME 03/07/2026, 9:55:18.974 AM	END TIME 03/07/2026, 9:55:20.140 AM

Input / Output

INPUT

```
[
  {
    "role": "user",
    "parts": [
      {
        "type": "text",
        "content": "What's the weather in Tokyo?"
      }
    ]
  }
```

OUTPUT

```
[
  {
    "role": "assistant",
    "parts": [
      {
        "type": "text",
        "content": "The weather in Tokyo is rainy
with a temperature of 85°F."
      }
    ],
```

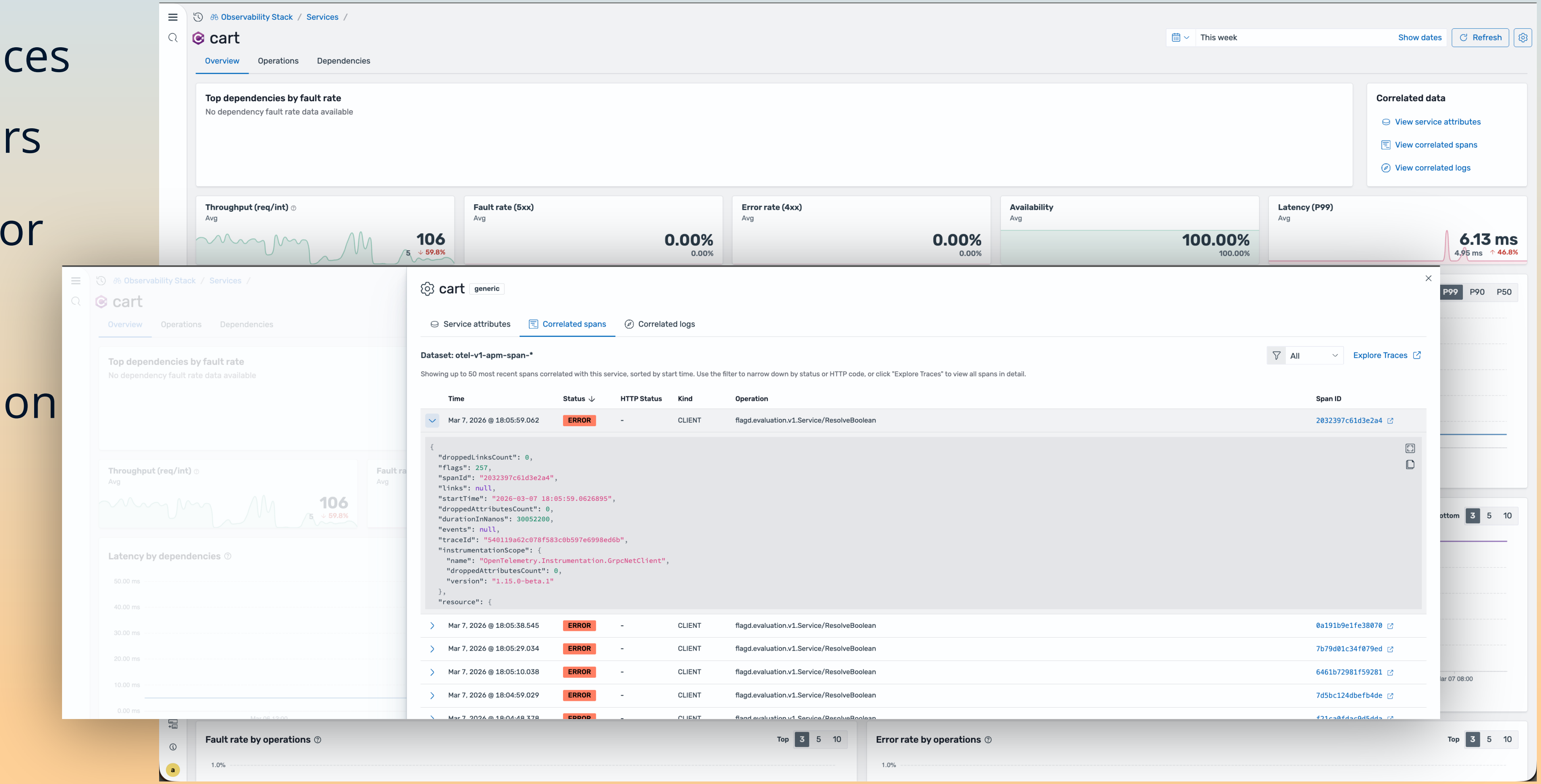
Raw Span

```
{
  "traceId": "f47bb1f4d3135086f192ae4182074579",
```

OpenSearch v3

More innovations for observability use cases

- Correlate logs and traces across multiple clusters
- Cross-cluster search for traces
- Trace-to-logs correlation
- Query trace data with Discover Traces



Dataset: otel-v1-apm-span-*

Showing up to 50 most recent spans correlated with this service, sorted by start time. Use the filter to narrow down by status or HTTP code, or click "Explore Traces" to view all spans in detail.

Time	Status ↓	HTTP Status	Kind	Operation	Span ID
Mar 7, 2026 @ 18:05:59.062	ERROR	-	CLIENT	flagd.evaluation.v1.Service/ResolveBoolean	2032397c61d3e2a4

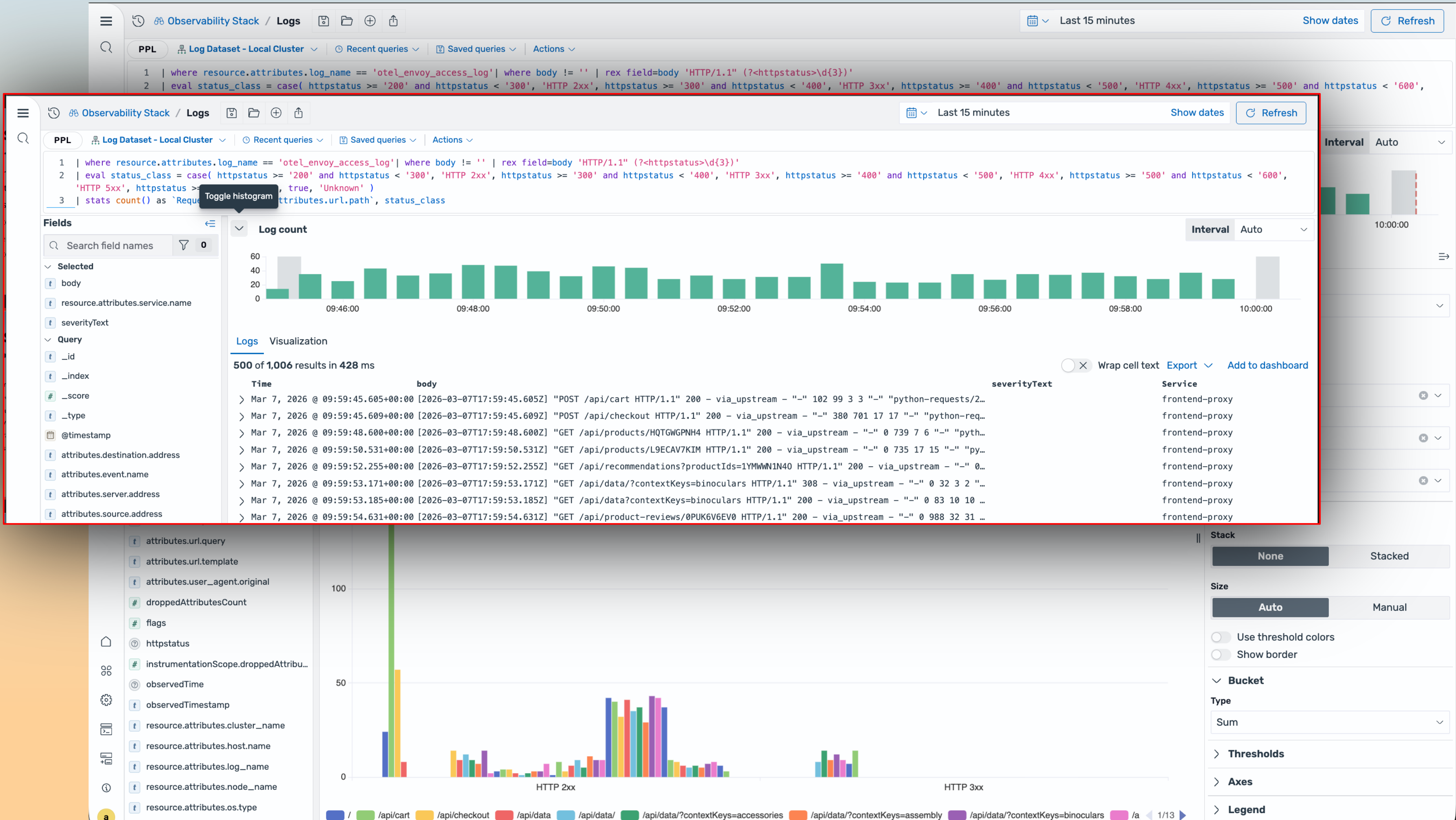
```
{
  "droppedLinksCount": 0,
  "flags": 257,
  "spanId": "2032397c61d3e2a4",
  "links": null,
  "startTime": "2026-03-07 18:05:59.0626895",
  "droppedAttributesCount": 0,
  "durationInNanos": 30052200,
  "events": null,
  "traceId": "540119a62c078f583c0b597e6998ed6b",
  "instrumentationScope": {
    "name": "OpenTelemetry.Instrumentation.GrpcNetClient",
    "droppedAttributesCount": 0,
    "version": "1.15.0-beta.1"
  },
  "resource": {
```

Time	Status ↓	HTTP Status	Kind	Operation	Span ID
Mar 7, 2026 @ 18:05:38.545	ERROR	-	CLIENT	flagd.evaluation.v1.Service/ResolveBoolean	0a191b9e1fe38070
Mar 7, 2026 @ 18:05:29.034	ERROR	-	CLIENT	flagd.evaluation.v1.Service/ResolveBoolean	7b79d01c34f079ed
Mar 7, 2026 @ 18:05:10.038	ERROR	-	CLIENT	flagd.evaluation.v1.Service/ResolveBoolean	6461b72981f59281
Mar 7, 2026 @ 18:04:59.029	ERROR	-	CLIENT	flagd.evaluation.v1.Service/ResolveBoolean	7d5bc124dbefb4de
Mar 7, 2026 @ 18:04:48.378	ERROR	-	CLIENT	flagd.evaluation.v1.Service/ResolveBoolean	f21ca0fdac9d5dda

OpenSearch v3

Piped Processing Language (PPL) query enhancements

- More than 30 added PPL commands and functions
- nested JSON
- unstructured log data at query time
- Analyze time-series data
- new statistical functions



```
| where resource.attributes.log_name == 'otel_envoy_access_log'| where body != '' | rex field=body 'HTTP/1.1" (?<httpstatus>\d{3})'
| eval status_class = case( httpstatus >= '200' and httpstatus < '300', 'HTTP 2xx', httpstatus >= '300' and httpstatus < '400', 'HTTP 3xx', httpstatus >= '400' and httpstatus < '500', 'HTTP 4xx', httpstatus >= '500' and httpstatus < '600',
'HTTP 5xx', httpstatus >                  , true, 'Unknown' )
| stats count() as `Reque                  ttributes.url.path`, status_class
```

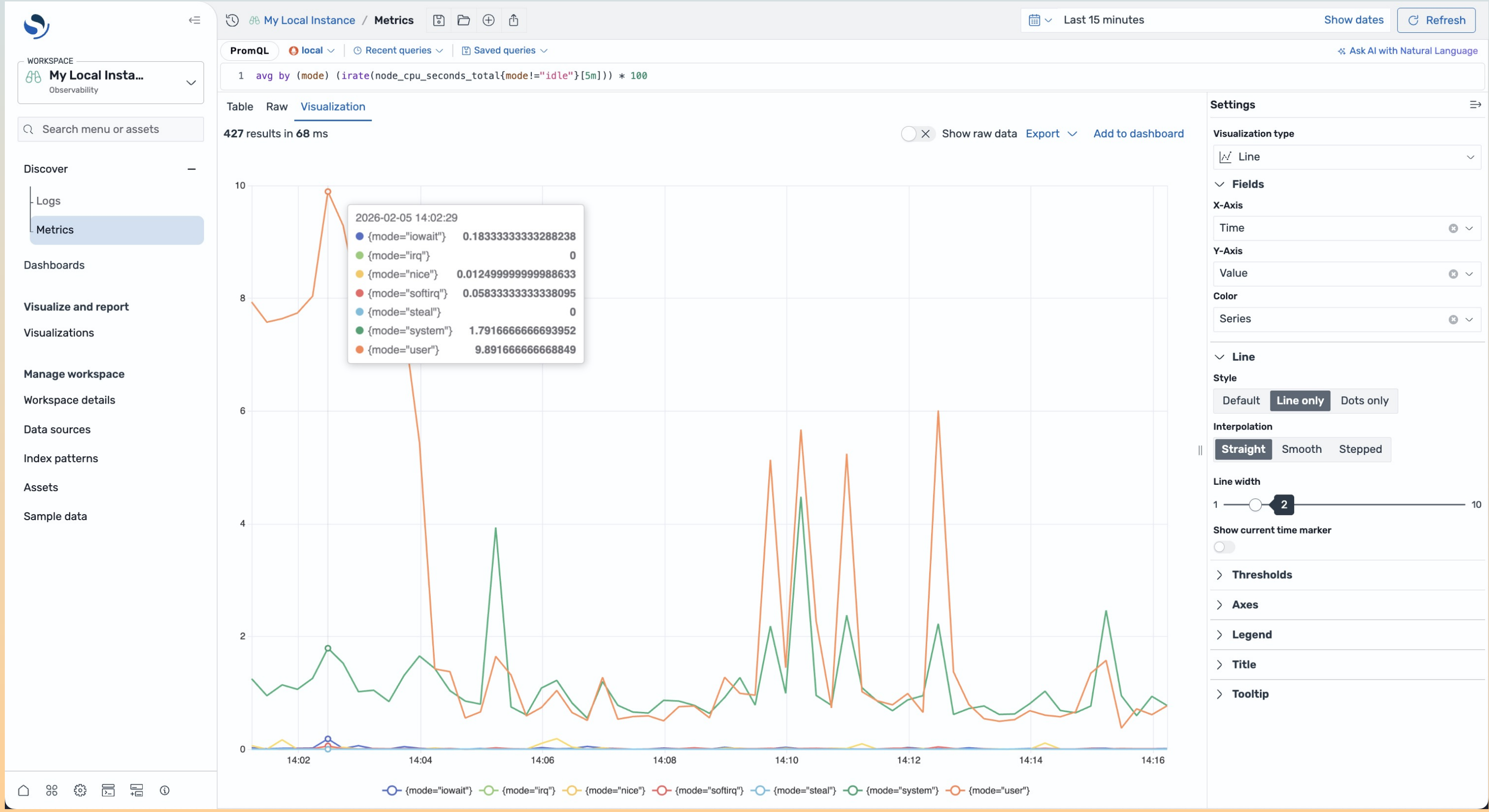
500 of 1,006 results in 428 ms

```
Time                               body                                                                                                  severityText    Service
Mar 7, 2026 @ 09:59:45.605+00:00 [2026-03-07T17:59:45.605Z] "POST /api/cart HTTP/1.1" 200 - via_upstream - "-" 102 99 3 3 "-" "python-requests/2...          frontend-proxy
Mar 7, 2026 @ 09:59:45.609+00:00 [2026-03-07T17:59:45.609Z] "POST /api/checkout HTTP/1.1" 200 - via_upstream - "-" 380 701 17 17 "-" "python-req...        frontend-proxy
Mar 7, 2026 @ 09:59:48.600+00:00 [2026-03-07T17:59:48.600Z] "GET /api/products/HQTGWGPNH4 HTTP/1.1" 200 - via_upstream - "-" 0 739 7 6 "-" "pyth...       frontend-proxy
Mar 7, 2026 @ 09:59:50.531+00:00 [2026-03-07T17:59:50.531Z] "GET /api/products/L9ECAV7KIM HTTP/1.1" 200 - via_upstream - "-" 0 735 17 15 "-" "py...      frontend-proxy
Mar 7, 2026 @ 09:59:52.255+00:00 [2026-03-07T17:59:52.255Z] "GET /api/recommendations?productIds=1YMWWN1N40 HTTP/1.1" 200 - via_upstream - "-" 0...      frontend-proxy
Mar 7, 2026 @ 09:59:53.171+00:00 [2026-03-07T17:59:53.171Z] "GET /api/data/?contextKeys=binoculars HTTP/1.1" 308 - via_upstream - "-" 0 32 3 2 "...      frontend-proxy
Mar 7, 2026 @ 09:59:53.185+00:00 [2026-03-07T17:59:53.185Z] "GET /api/data?contextKeys=binoculars HTTP/1.1" 200 - via_upstream - "-" 0 83 10 10 ...      frontend-proxy
Mar 7, 2026 @ 09:59:54.631+00:00 [2026-03-07T17:59:54.631Z] "GET /api/product-reviews/0PUK6V6EV0 HTTP/1.1" 200 - via_upstream - "-" 0 988 32 31 ...     frontend-proxy
```

OpenSearch v3

Prometheus native support + PromQL

- Introduced in OpenSearch v3.5
- Support for native PromQL queries
- Quick metrics analysis with saved queries
- Easy add to dashboards with existing visualizations
- Support for 8+ visualization types



```
avg by (mode) (irate(node_cpu_seconds_total{mode!="idle"}[5m])) * 100
```

427 results in 68 ms

2026-02-05 14:02:29

● {mode="iowait"}	0.18333333333288238
● {mode="irq"}	0
● {mode="nice"}	0.012499999999988633
● {mode="softirq"}	0.05833333333338095
● {mode="steal"}	0
● {mode="system"}	1.7916666666693952
● {mode="user"}	9.891666666668849

Visualization type: Line

Fields

X-Axis
Time

Y-Axis
Value

Color: Series

Line

Style: Default [✓ Line only] Dots only

Interpolation: [✓ Straight] Smooth Stepped

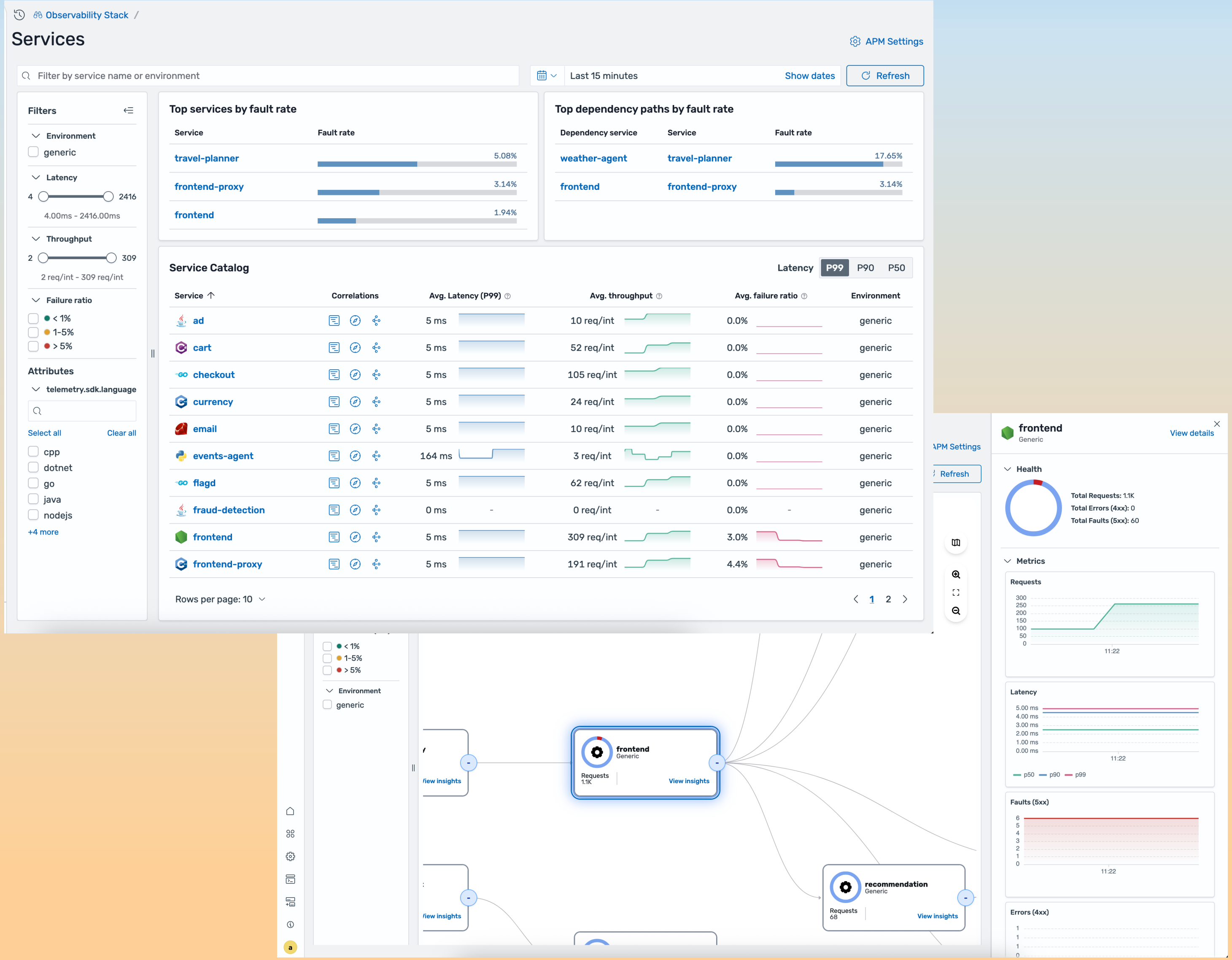
Line width: 1 — 2 — 10

Show current time marker ○

OpenSearch v3

APM and Tracing

- Support for Application Performance Monitoring
- OpenSearch stores Logs and Traces
- Prometheus stores Request, Error & Duration (RED Metrics)
- Comes with out of box topology map based on Application traces
- Supports attribute filtering, in-context correlations and more…



Top services by fault rate

Service	Fault rate
travel-planner	5.08%
frontend-proxy	3.14%
frontend	1.94%

Top dependency paths by fault rate

Dependency service	Service	Fault rate
weather-agent	travel-planner	17.65%
frontend	frontend-proxy	3.14%

Service Catalog

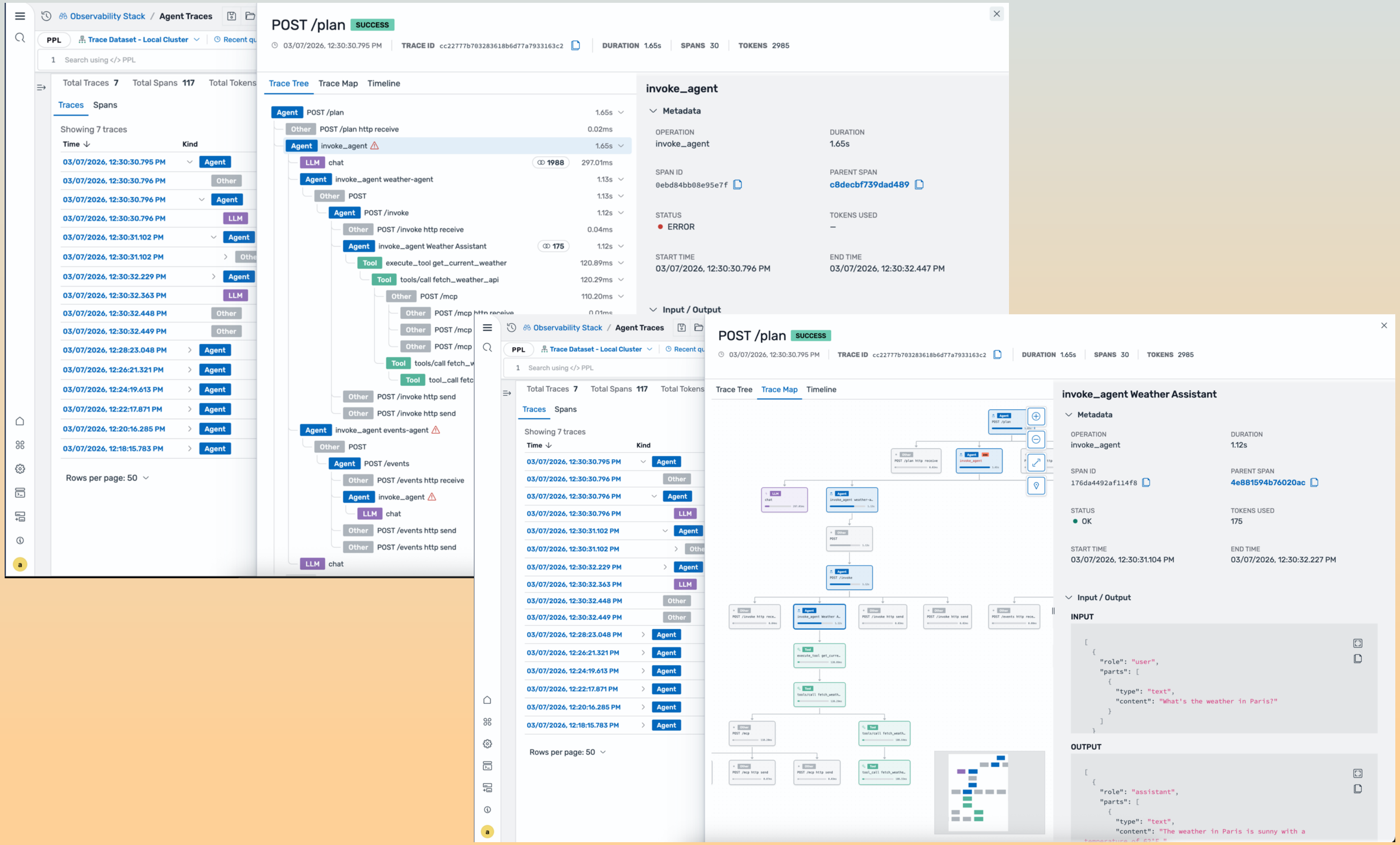
Service ↑	Correlations	Avg. Latency (P99)	Avg. throughput	Avg. failure ratio	Environment
ad		5 ms	10 req/int	0.0%	generic
cart		5 ms	52 req/int	0.0%	generic
checkout		5 ms	105 req/int	0.0%	generic
currency		5 ms	24 req/int	0.0%	generic
email		5 ms	10 req/int	0.0%	generic
events-agent		164 ms	3 req/int	0.0%	generic
flagd		5 ms	62 req/int	0.0%	generic
fraud-detection		0 ms -	0 req/int -	0.0% -	generic
frontend		5 ms	309 req/int	3.0%	generic
frontend-proxy		5 ms	191 req/int	4.4%	generic

Total Requests: 1.1K
Total Errors (4xx): 0
Total Faults (5xx): 60

Agentic AI - Observability, Analytics and Evals Platform

AI observability: from development to production

- Agent Traces, Span
- Query with OpenSearch PPL
- Dive deep into Agent Path for LLM, agent, embeddings and API calls
- Quick search on errors with Facets
- Summary stats on token count usages, P50/P99 Latencies ..



Observability Stack / Agent Traces

PPL | Trace Dataset - Local Cluster | Recent qu

1 Search using </> PPL

Total Traces 7 | Total Spans 117 | Total Tokens

Traces | Spans

Showing 7 traces

Time ↓	Kind
03/07/2026, 12:30:30.795 PM	Agent
03/07/2026, 12:30:30.796 PM	Other
03/07/2026, 12:30:30.796 PM	Agent
03/07/2026, 12:30:30.796 PM	LLM
03/07/2026, 12:30:31.102 PM	Agent
03/07/2026, 12:30:31.102 PM	Othe
03/07/2026, 12:30:32.229 PM	Agent
03/07/2026, 12:30:32.363 PM	LLM
03/07/2026, 12:30:32.448 PM	Other
03/07/2026, 12:30:32.449 PM	Other
03/07/2026, 12:28:23.048 PM	Agent
03/07/2026, 12:26:21.321 PM	Agent
03/07/2026, 12:24:19.613 PM	Agent
03/07/2026, 12:22:17.871 PM	Agent
03/07/2026, 12:20:16.285 PM	Agent
03/07/2026, 12:18:15.783 PM	Agent

Rows per page: 50

POST /plan SUCCESS

03/07/2026, 12:30:30.795 PM | TRACE ID cc22777b703283618b6d77a7933163c2 | DURATION 1.65s | SPANS 30 | TOKENS 2985

Trace Tree | Trace Map | Timeline

```
Agent POST /plan                                        1.65s
  Other POST /plan http receive                         0.02ms
  Agent invoke_agent ⚠                                  1.65s
    LLM chat                                 1988     297.01ms
    Agent invoke_agent weather-agent                    1.13s
      Other POST                                        1.13s
        Agent POST /invoke                              1.12s
          Other POST /invoke http receive               0.04ms
          Agent invoke_agent Weather Assistant  175     1.12s
            Tool execute_tool get_current_weather     120.89ms
              Tool tools/call fetch_weather_api       120.29ms
                Other POST /mcp                       110.20ms
                  Other POST /mcp http receive          0.01ms
                  Other POST /mcp
                  Other POST /mcp
                Tool tools/call fetch_w
                  Tool tool_call fetc
          Other POST /invoke http send
          Other POST /invoke http send
    Agent invoke_agent events-agent ⚠
      Other POST
        Agent POST /events
          Other POST /events http receive
          Agent invoke_agent ⚠
            LLM chat
          Other POST /events http send
          Other POST /events http send
    LLM chat
```

invoke_agent

Metadata

OPERATION invoke_agent	DURATION 1.65s
SPAN ID 0ebd84bb08e95e7f	PARENT SPAN c8decbf739dad489
STATUS ● ERROR	TOKENS USED –
START TIME 03/07/2026, 12:30:30.796 PM	END TIME 03/07/2026, 12:30:32.447 PM

Input / Output

Observability Stack / Agent Traces

PPL | Trace Dataset - Local Cluster | Recent qu

1 Search using </> PPL

Total Traces 7 | Total Spans 117 | Total Tokens

Traces | Spans

Showing 7 traces

Rows per page: 50

POST /plan SUCCESS

03/07/2026, 12:30:30.795 PM | TRACE ID cc22777b703283618b6d77a7933163c2 | DURATION 1.65s | SPANS 30 | TOKENS 2985

Trace Tree | Trace Map | Timeline

invoke_agent Weather Assistant

Metadata

OPERATION invoke_agent	DURATION 1.12s
SPAN ID 176da4492af114f8	PARENT SPAN 4e881594b76020ac
STATUS ● OK	TOKENS USED 175
START TIME 03/07/2026, 12:30:31.104 PM	END TIME 03/07/2026, 12:30:32.227 PM

Input / Output

INPUT

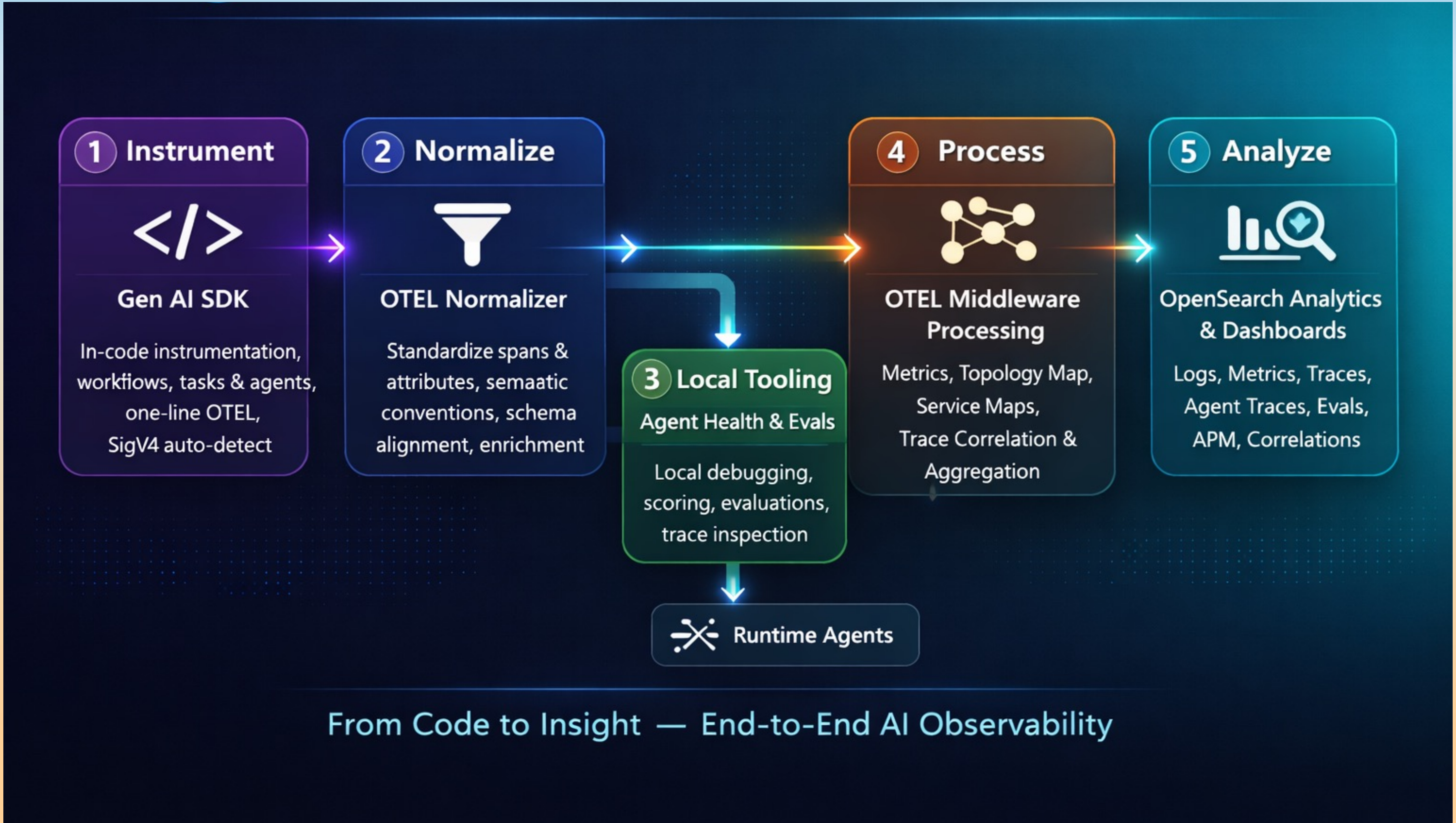
```
[
  {
    "role": "user",
    "parts": [
      {
        "type": "text",
        "content": "What's the weather in Paris?"
      }
    ]
  }
```

OUTPUT

```
[
  {
    "role": "assistant",
    "parts": [
      {
        "type": "text",
        "content": "The weather in Paris is sunny with a
temperature of 62°F."
```

Agentic AI – Observability Platform

OpenSearch



1 Instrument

Gen AI SDK

In-code instrumentation, workflows, tasks & agents, one-line OTEL, SigV4 auto-detect

2 Normalize

OTEL Normalizer

Standardize spans & attributes, semaatic conventions, schema alignment, enrichment

3 Local Tooling

Agent Health & Evals

Local debugging, scoring, evaluations, trace inspection

Runtime Agents

4 Process

OTEL Middleware Processing

Metrics, Topology Map, Service Maps, Trace Correlation & Aggregation

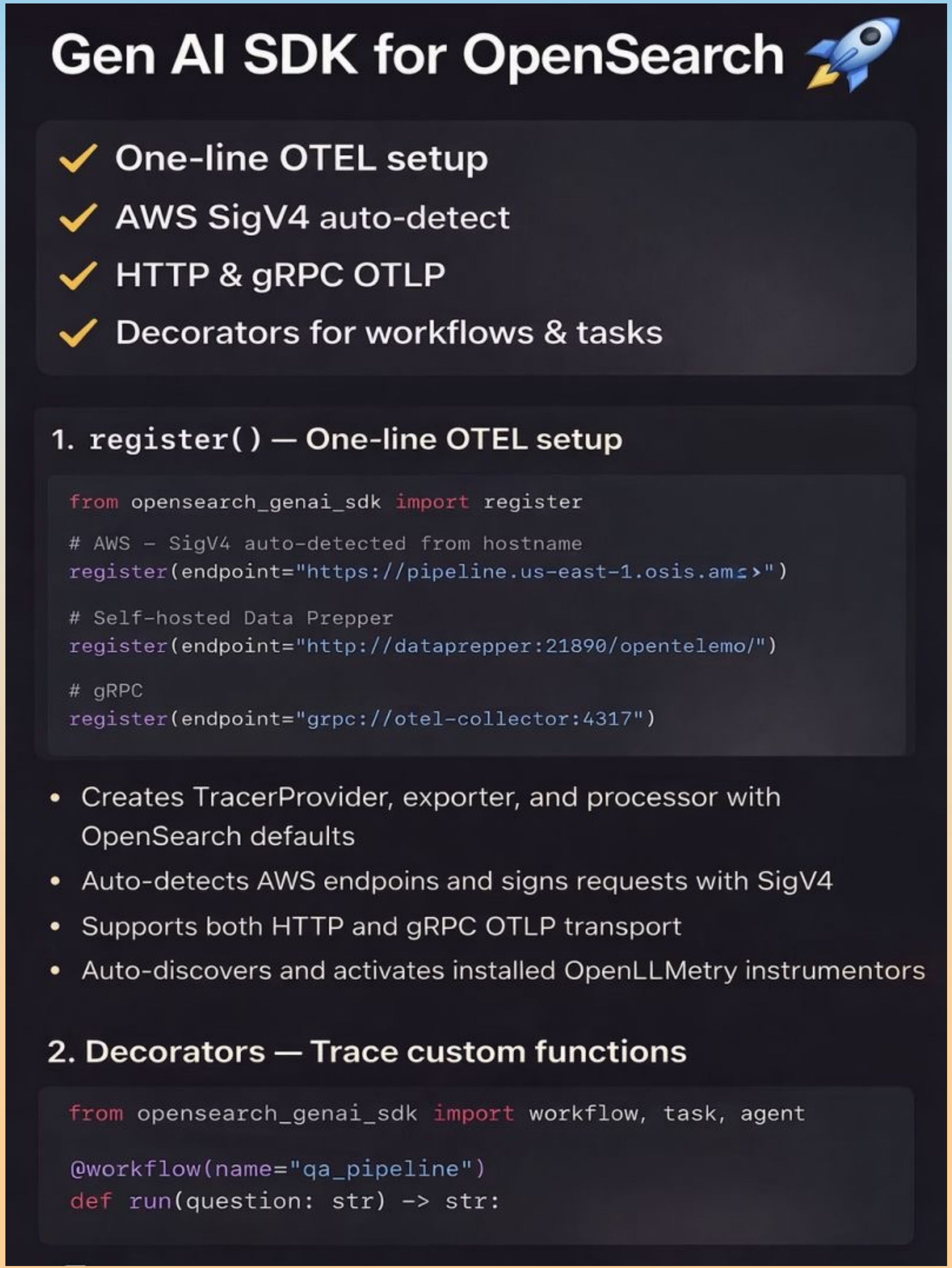
5 Analyze

OpenSearch Analytics & Dashboards

Logs, Metrics, Traces, Agent Traces, Evals, APM, Correlations

From Code to Insight — End-to-End AI Observability

AI observability: from development to production



Gen AI SDK for OpenSearch 🚀

- ✓ One-line OTEL setup
- ✓ AWS SigV4 auto-detect
- ✓ HTTP & gRPC OTLP
- ✓ Decorators for workflows & tasks

1. register() — One-line OTEL setup

```
from opensearch_genai_sdk import register

# AWS - SigV4 auto-detected from hostname
register(endpoint="https://pipeline.us-east-1.osis.am≤›")

# Self-hosted Data Prepper
register(endpoint="http://dataprepper:21890/opentelemo/")

# gRPC
register(endpoint="grpc://otel-collector:4317")
```

- Creates TracerProvider, exporter, and processor with OpenSearch defaults
- Auto-detects AWS endpoins and signs requests with SigV4
- Supports both HTTP and gRPC OTLP transport
- Auto-discovers and activates installed OpenLLMetry instrumentors

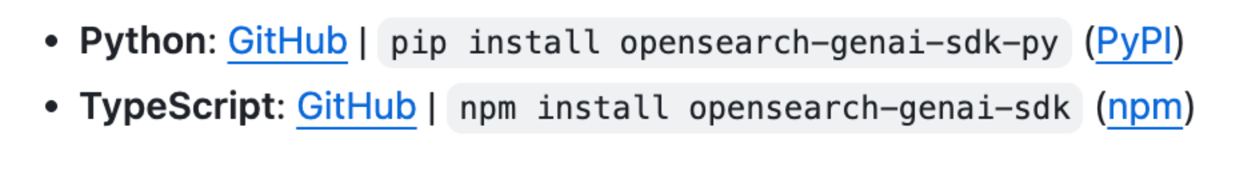
2. Decorators — Trace custom functions

```
from opensearch_genai_sdk import workflow, task, agent

@workflow(name="qa_pipeline")
def run(question: str) -> str:
```

Vamsi Manohar proposed a Gen AI SDK for OpenSearch to standardize evaluation and observability for agentic applications using an OTEL-native approach.

The goal:

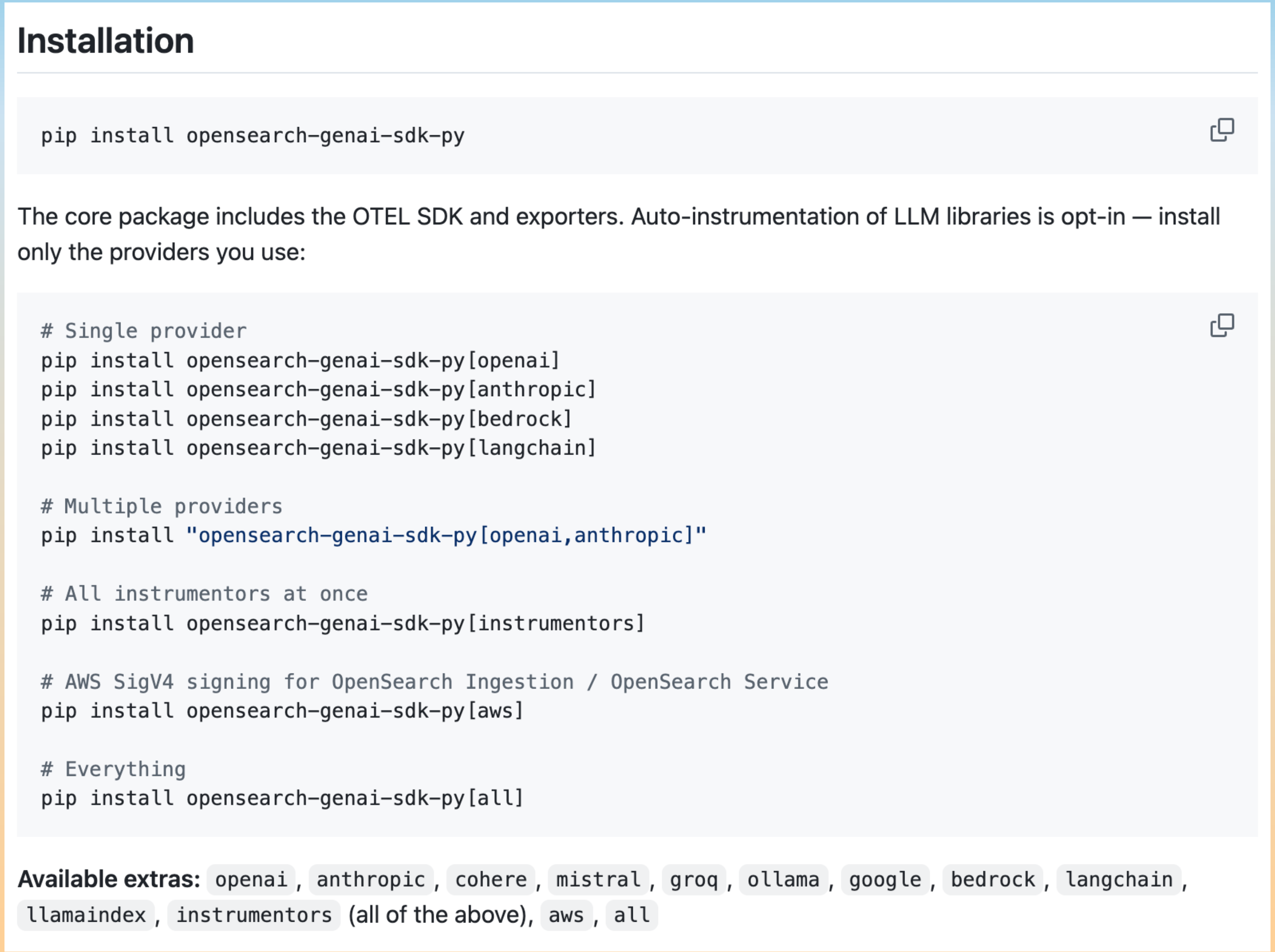
- Simple instrumentation
- Built-in scoring & evaluation export
- Secure OpenSearch ingestion
- Python & TypeScript support



- **Python:** GitHub | `pip install opensearch-genai-sdk-py` (PyPI)
- **TypeScript:** GitHub | `npm install opensearch-genai-sdk` (npm)

https://github.com/opensearch-project/dashboards-observability/issues/2591

AI observability: instrumentation – Gen_AI SDK



Installation

```
pip install opensearch-genai-sdk-py
```

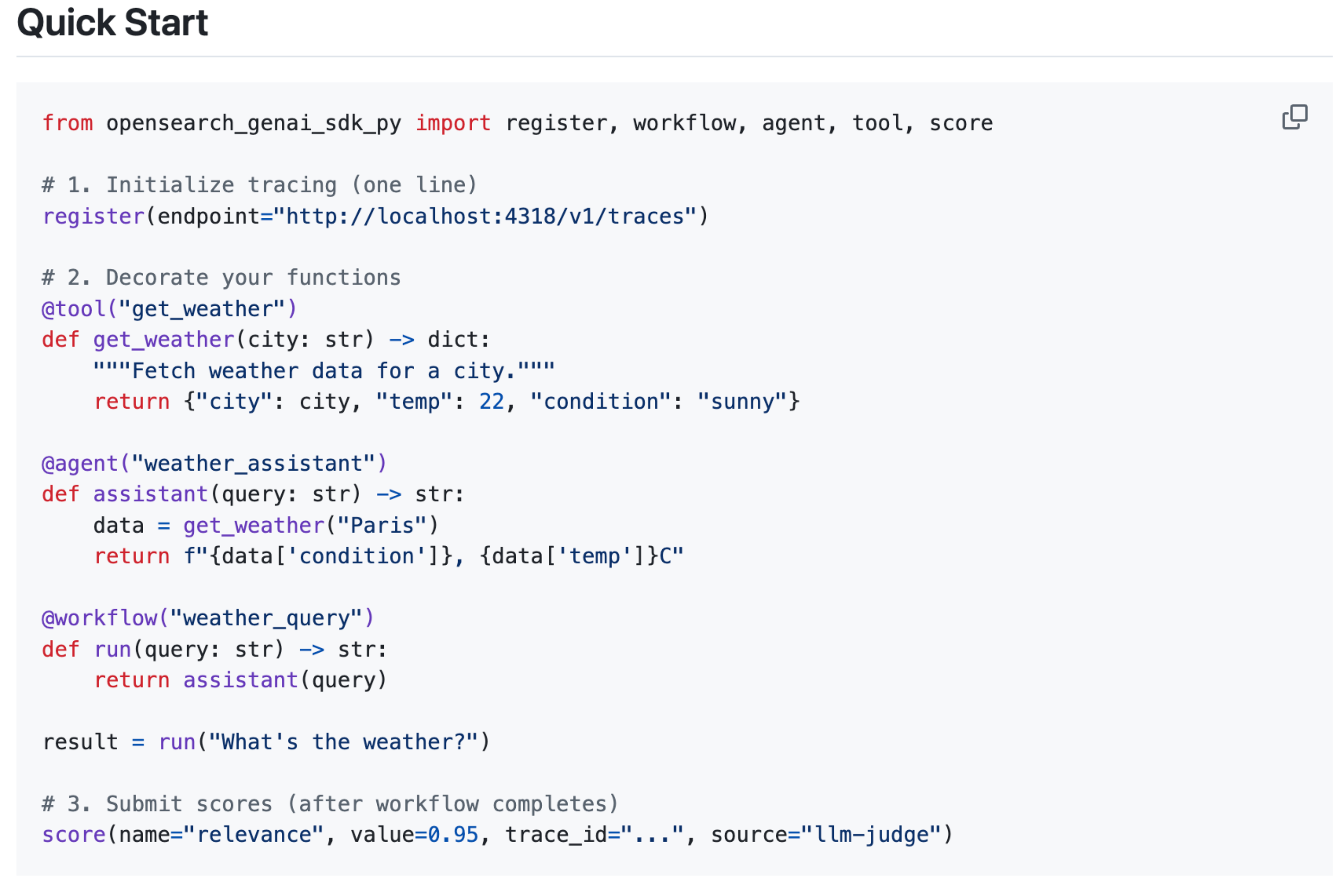
The core package includes the OTEL SDK and exporters. Auto-instrumentation of LLM libraries is opt-in — install only the providers you use:

```
# Single provider
pip install opensearch-genai-sdk-py[openai]
pip install opensearch-genai-sdk-py[anthropic]
pip install opensearch-genai-sdk-py[bedrock]
pip install opensearch-genai-sdk-py[langchain]

# Multiple providers
pip install "opensearch-genai-sdk-py[openai,anthropic]"

# All instrumentors at once
pip install opensearch-genai-sdk-py[instrumentors]

# AWS SigV4 signing for OpenSearch Ingestion / OpenSearch Service
pip install opensearch-genai-sdk-py[aws]

# Everything
pip install opensearch-genai-sdk-py[all]
```

Available extras: `openai`, `anthropic`, `cohere`, `mistral`, `groq`, `ollama`, `google`, `bedrock`, `langchain`, `llamaindex`, `instrumentors` (all of the above), `aws`, `all`

https://github.com/opensearch-project/dashboards-observability/issues/2591

AI observability: instrumentation – Gen_AI SDK



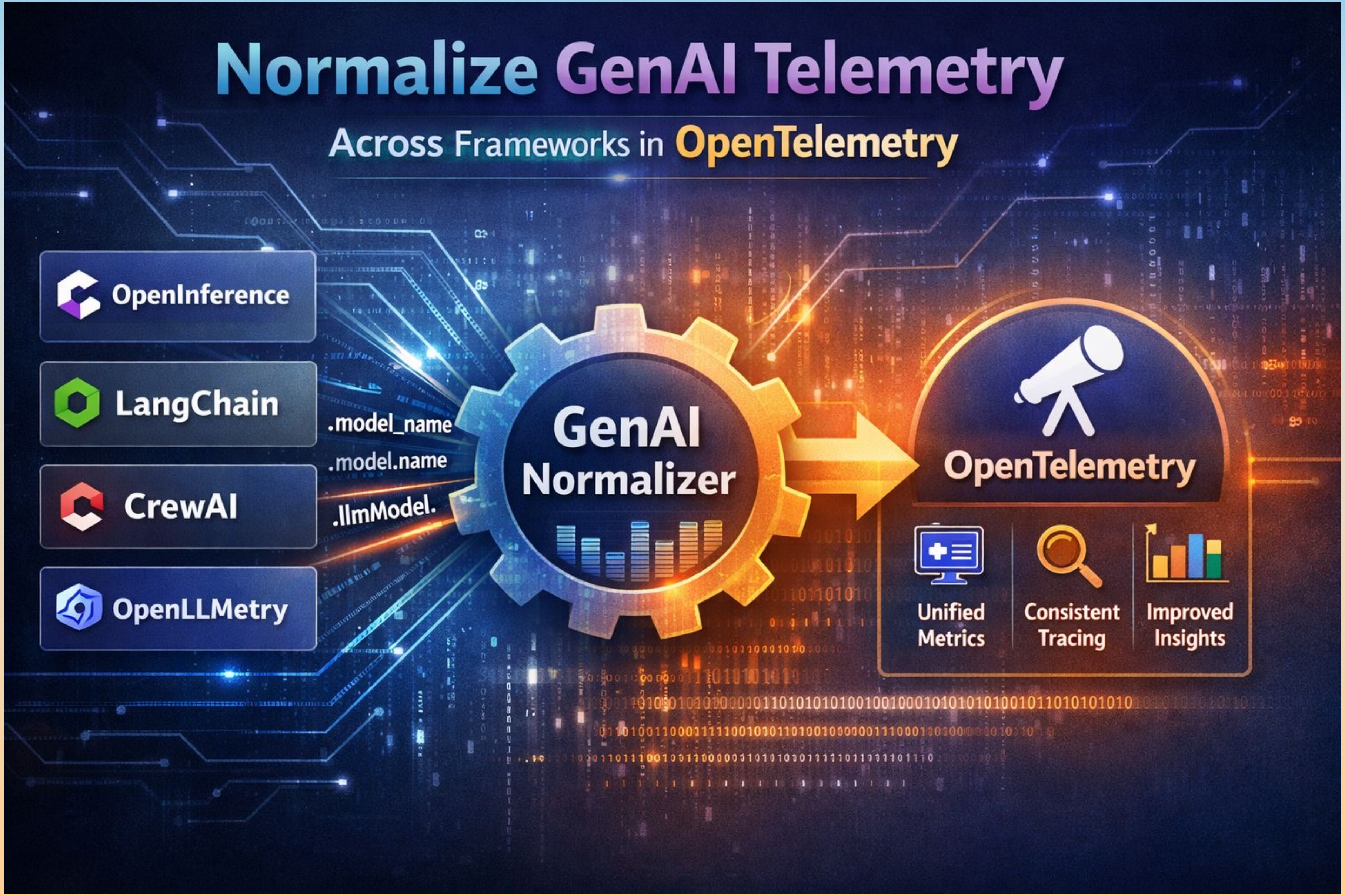
Quick Start

```python
from opensearch_genai_sdk_py import register, workflow, agent, tool, score

# 1. Initialize tracing (one line)
register(endpoint="http://localhost:4318/v1/traces")

# 2. Decorate your functions
@tool("get_weather")
def get_weather(city: str) -> dict:
    """Fetch weather data for a city."""
    return {"city": city, "temp": 22, "condition": "sunny"}

@agent("weather_assistant")
def assistant(query: str) -> str:
    data = get_weather("Paris")
    return f"{data['condition']}, {data['temp']}C"

@workflow("weather_query")
def run(query: str) -> str:
    return assistant(query)

result = run("What's the weather?")

# 3. Submit scores (after workflow completes)
score(name="relevance", value=0.95, trace_id="...", source="llm-judge")
```

https://github.com/opensearch-project/dashboards-observability/issues/2591

AI observability: migration | OTEL Normalization

OpenSearch



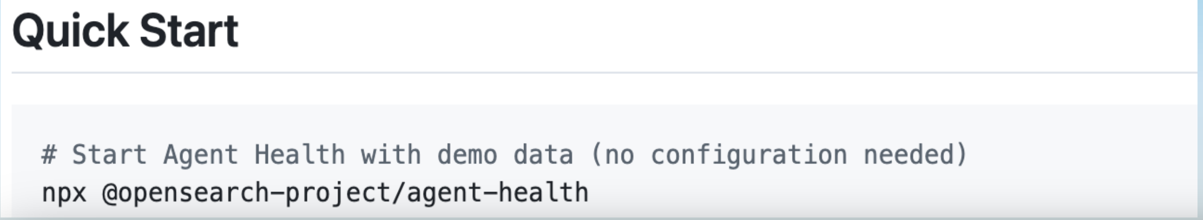
Kyle Hounslow created a new **genainormalizer processor** has been proposed for **OpenTelemetry Collector Contrib**

To standardize GenAI attributes across frameworks like

OpenInference, OpenLLMetry, LangChain, CrewAI, and others

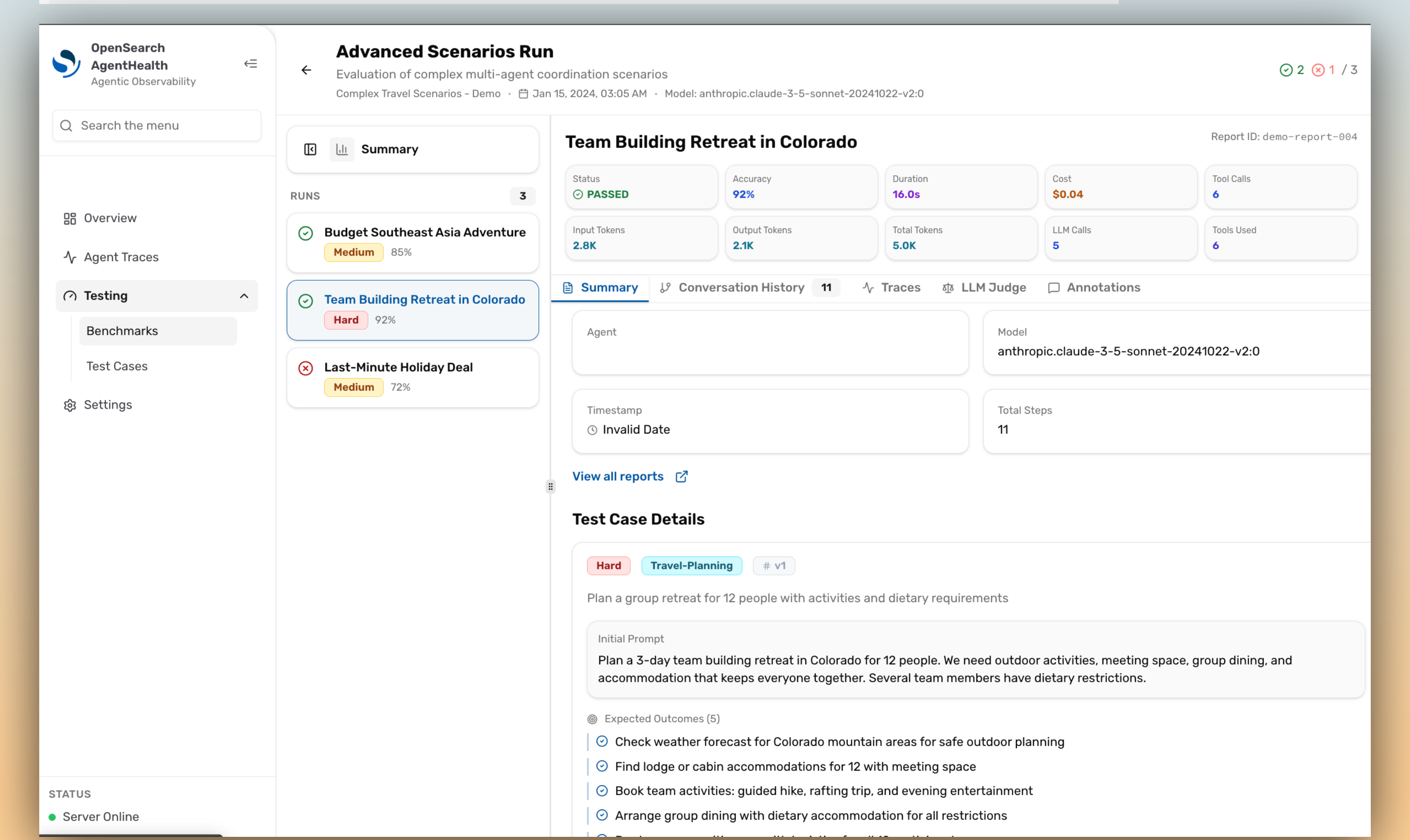
https://github.com/open-telemetry/**opentelemetry-collector-contrib/issues/46069**

AI observability: migration | Agent-Health



Quick Start

```
# Start Agent Health with demo data (no configuration needed)
npx @opensearch-project/agent-health
```



Advanced Scenarios Run

Evaluation of complex multi-agent coordination scenarios

Complex Travel Scenarios - Demo · Jan 15, 2024, 03:05 AM · Model: anthropic.claude-3-5-sonnet-20241022-v2:0

✓ 2 ⊗ 1 / 3

Summary

RUNS 3

- ✓ Budget Southeast Asia Adventure — Medium 85%
- ✓ Team Building Retreat in Colorado — Hard 92%
- ⊗ Last-Minute Holiday Deal — Medium 72%

Team Building Retreat in Colorado

Report ID: demo-report-004

Status	Accuracy	Duration	Cost	Tool Calls
✓ PASSED	92%	16.0s	$0.04	6

Input Tokens	Output Tokens	Total Tokens	LLM Calls	Tools Used
2.8K	2.1K	5.0K	5	6

Summary | Conversation History 11 | Traces | LLM Judge | Annotations

Agent

Model: anthropic.claude-3-5-sonnet-20241022-v2:0

Timestamp: Invalid Date

Total Steps: 11

View all reports

Test Case Details

Hard | Travel-Planning | # v1

Plan a group retreat for 12 people with activities and dietary requirements

Initial Prompt

Plan a 3-day team building retreat in Colorado for 12 people. We need outdoor activities, meeting space, group dining, and accommodation that keeps everyone together. Several team members have dietary restrictions.

Expected Outcomes (5)

- ✓ Check weather forecast for Colorado mountain areas for safe outdoor planning
- ✓ Find lodge or cabin accommodations for 12 with meeting space
- ✓ Book team activities: guided hike, rafting trip, and evening entertainment
- ✓ Arrange group dining with dietary accommodation for all restrictions

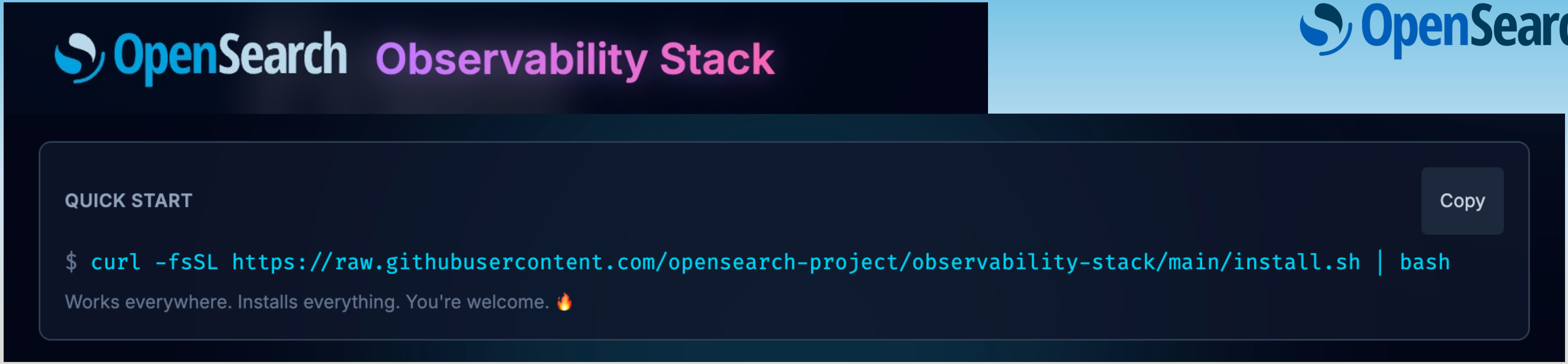
Megha Goyal, created

Agent Health is an evaluation and observability framework for AI agents.

It helps you measure agent performance through "Golden Path" trajectory comparison; where an LLM judge evaluates agent actions against expected outcomes.

https://github.com/open-telemetry/**opentelemetry-collector-contrib**/**issues/46069**

OpenSearch



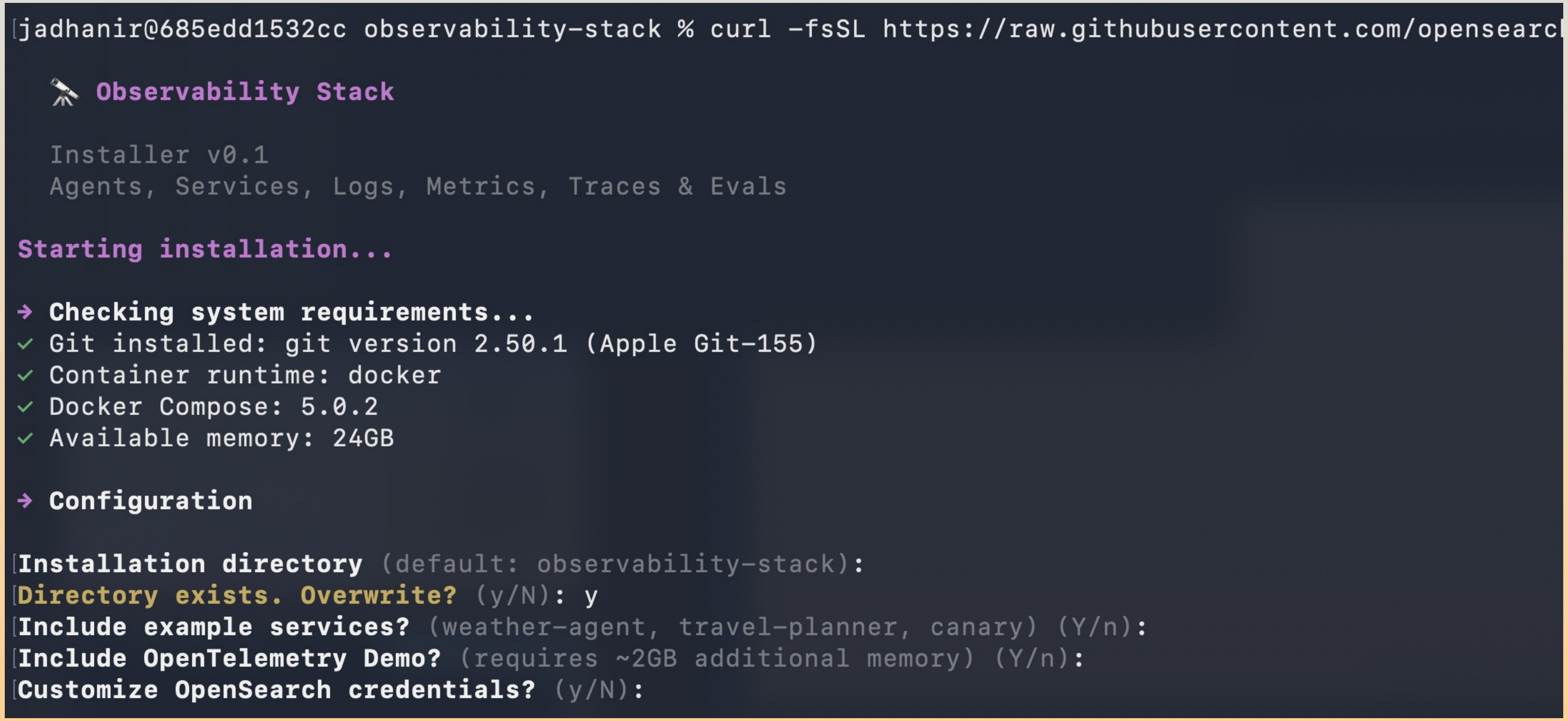
OpenSearch Observability Stack

QUICK START

Copy

```
$ curl -fsSL https://raw.githubusercontent.com/opensearch-project/observability-stack/main/install.sh | bash
```

Works everywhere. Installs everything. You're welcome. 🔥



```
jadhanir@685edd1532cc observability-stack % curl -fsSL https://raw.githubusercontent.com/opensearc

  🔭 Observability Stack

  Installer v0.1
  Agents, Services, Logs, Metrics, Traces & Evals

Starting installation...

→ Checking system requirements...
✓ Git installed: git version 2.50.1 (Apple Git-155)
✓ Container runtime: docker
✓ Docker Compose: 5.0.2
✓ Available memory: 24GB

→ Configuration

Installation directory (default: observability-stack):
Directory exists. Overwrite? (y/N): y
Include example services? (weather-agent, travel-planner, canary) (Y/n):
Include OpenTelemetry Demo? (requires ~2GB additional memory) (Y/n):
Customize OpenSearch credentials? (y/N):
```

https://observability.opensearch.org/

Demo

The road ahead

2026 Roadmap

- OpenTelemetry compatibility
- Native TSDB support
- Unified analytics and developer experience
- Unified alerting
- Natural language querying
- Agent observability
- And much more...

Dotan Horovits · You
CNCF Ambassador | OpenSearch Ambassador | Open Source Ad...
4d ·

OpenSearch Project #roadmap 2026 is out!
The Technical Steering Committee (TSC) has compiled this to give the community greater visibility into what's coming next, across the ...more

Itai Binyamin and 36 others 2 comments · 5 reposts

Blog

Announcing OpenSearch Observability TAG: Shaping Open Source Observability Together

By Dotan Horovits, Shenoy Pratik Gurudatt, Orcun Berkem

| September 16, 2025

Find us at

Forum

Events

Blog

Github

Slack

Thanks!