(Open Source) Observability from Scratch

Josh Lee • Altinity • May 1 2025 • OpenSearchCon Europe

Josh Lee

Open Source Advocate

Altinity

Observability is our ability to <u>understand</u> a system from its outputs alone

“There are only two signals: metrics and (structured) logs”

— paraphrased from Charity Majors, Honeycomb

A Typical Request Log

```
2024-07-01 09:35:34 GET /home 200 ...
```

Adding Duration

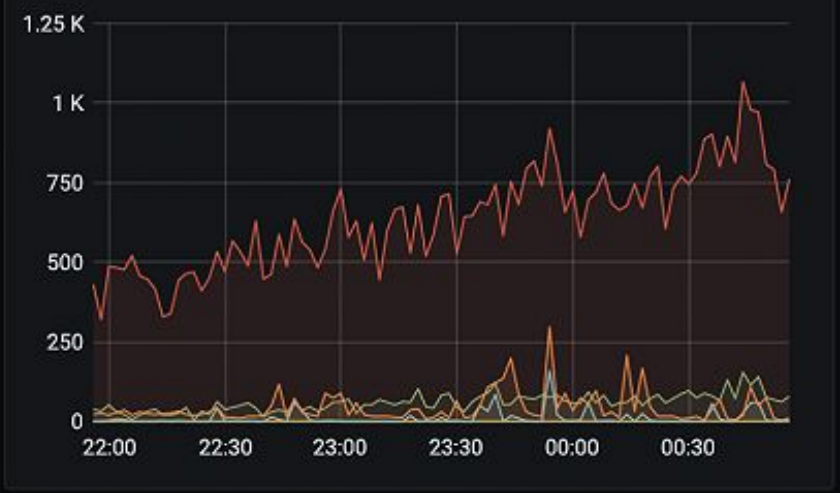
```
2024-07-01 09:35:34 231ms GET /home 200
```

Controller Class | All | Namespace | live | Ingress | All | Ingress Pod | All | Config Reloads

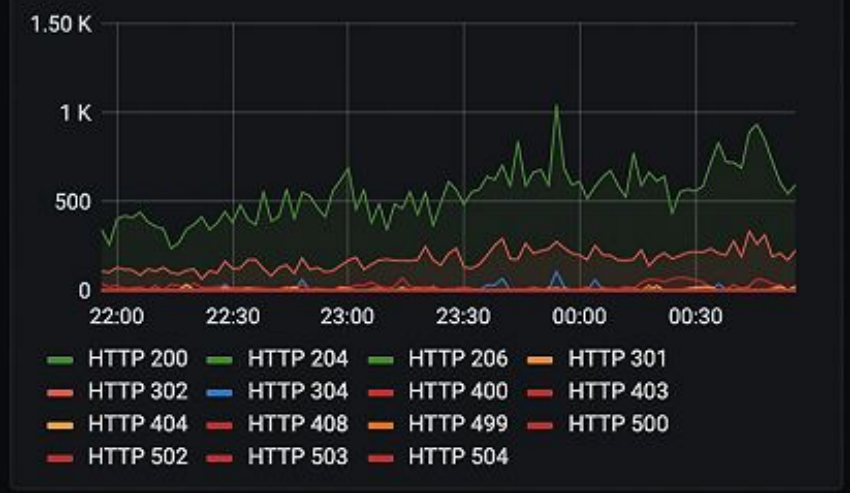
Overview

Requests (period)	% Succe...	Conns (...	Reqs (2...	% Success (2m)	HTTP 1/2xx (2m)	HTTP 3xx (2m)	HTTP 4xx (2m)	HTTP 5xx (2m)
67.3 K	97.0%	311	850	99.3%	592	228	30	0

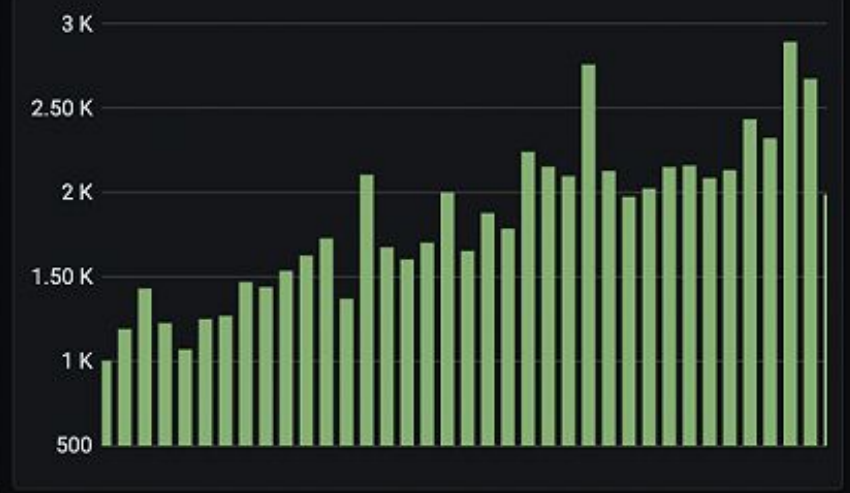
HTTP Requests / Ingress



HTTP Status Codes



Total HTTP Requests

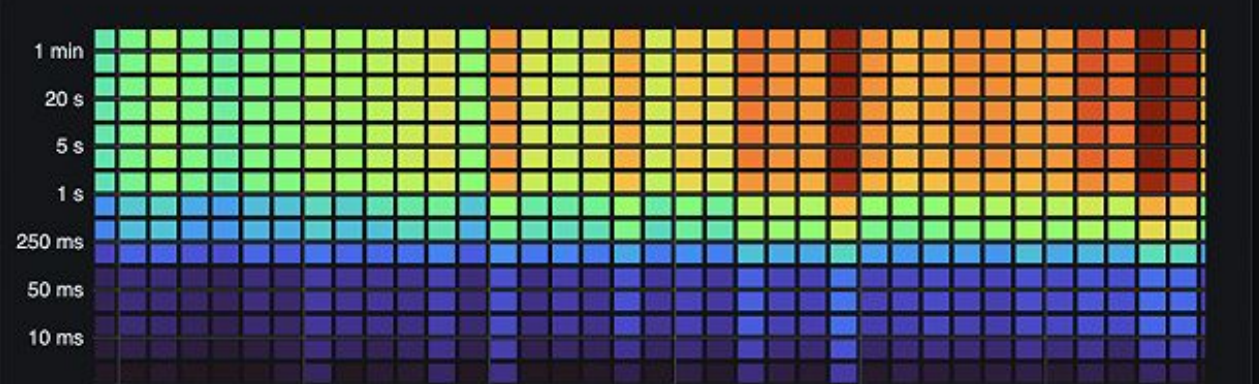


Latency

Latency (Average Percentiles)

	min	max	avg
0.99	4.3 s	42.3 s	6.5 s
0.95	2.4 s	4.5 s	2.8 s
0.9	2.1 s	2.7 s	2.3 s
0.5	357.2 ms	631.6 ms	443.7 ms
0.75	1.0 s	1.8 s	1.5 s
Average	3.2 s	5.1 s	4.1 s

Latency Heatmap



Back to our log…

```
2024-07-01 09:35:34 231ms GET /home 200
```

Back to our log…

```
Request:123 2024-07-01 09:35:34 231ms GET /home
200
```

Connecting the trace:

```
Trace:4ea3 Span:123 2024-07-01 09:35:34 231ms GET
/home 200

Trace:4ea3 Span:456 ParentSpan:123 2024-07-01
09:35:34 201ms GET /api/users 201
```

ER UI | Search | Compare | System Architecture | Monitor | Lookup by Trace ID... | About Jaege

frontend: HTTP GET ca28836

Find... | Trace Timeli

tart **May 31 2023, 11:36:59.537** | Duration **24.95ms** | Services **4** | Depth **7** | Total Spans **17**

e & Operation	0µs	6.24ms	12.48ms	18.71ms
ntend HTTP GET				
frontend grpc.oteldemo.RecommendationService/List...	17.69ms			
recommendationservice /oteldemo.Recomm...	14.12ms			
recommendationservice get_product_list	13.61ms			
recommendationservice /oteldemo...	11.54ms | recommendationservice::/oteldemo.FeatureFlagService/GetFl			
featureflagservice /oteldemo.Fe...	2.23ms			
featureflagservice featurefl...	1.93ms			
recommendationservice /oteldemo...	1.62ms			
productcatalogservice otelde...	29µs			
frontend grpc.oteldemo.ProductCatalogService/GetPr...	5.48ms			
productcatalogservice oteldemo.ProductCat...	27µs			
frontend grpc.oteldemo.ProductCatalogService/GetPr...	4.42ms			
productcatalogservice oteldemo.ProductCat...	9µs			
frontend grpc.oteldemo.ProductCatalogService/GetPr...	5.44ms			
productcatalogservice oteldemo.ProductCat...	21µs			
frontend grpc.oteldemo.ProductCatalogService/GetPr...	3.16ms			
productcatalogservice oteldemo.ProductCat...	20µs			

Observability is not any one signal…

Metrics

Aggregable

Is there a problem?

Traces

Request-Scoped

Where is the problem?

Logs

Verbose, time-stamped records

What is the problem?

Distributed Tracing is the *"Killer App"*

Understand complete request flows

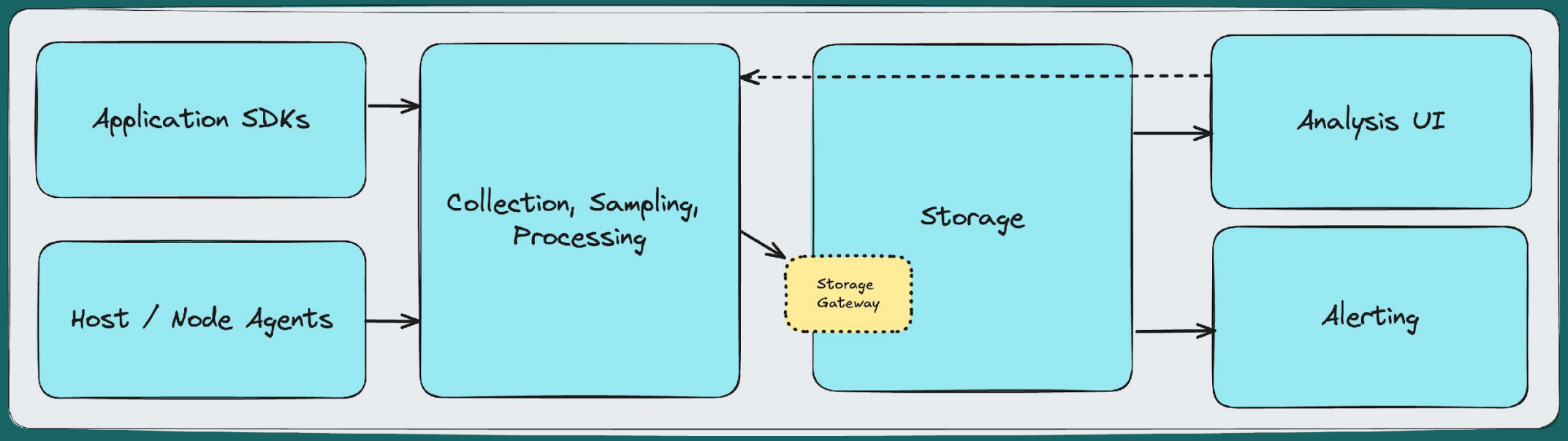
Create a real-time map of system topology and dependencies

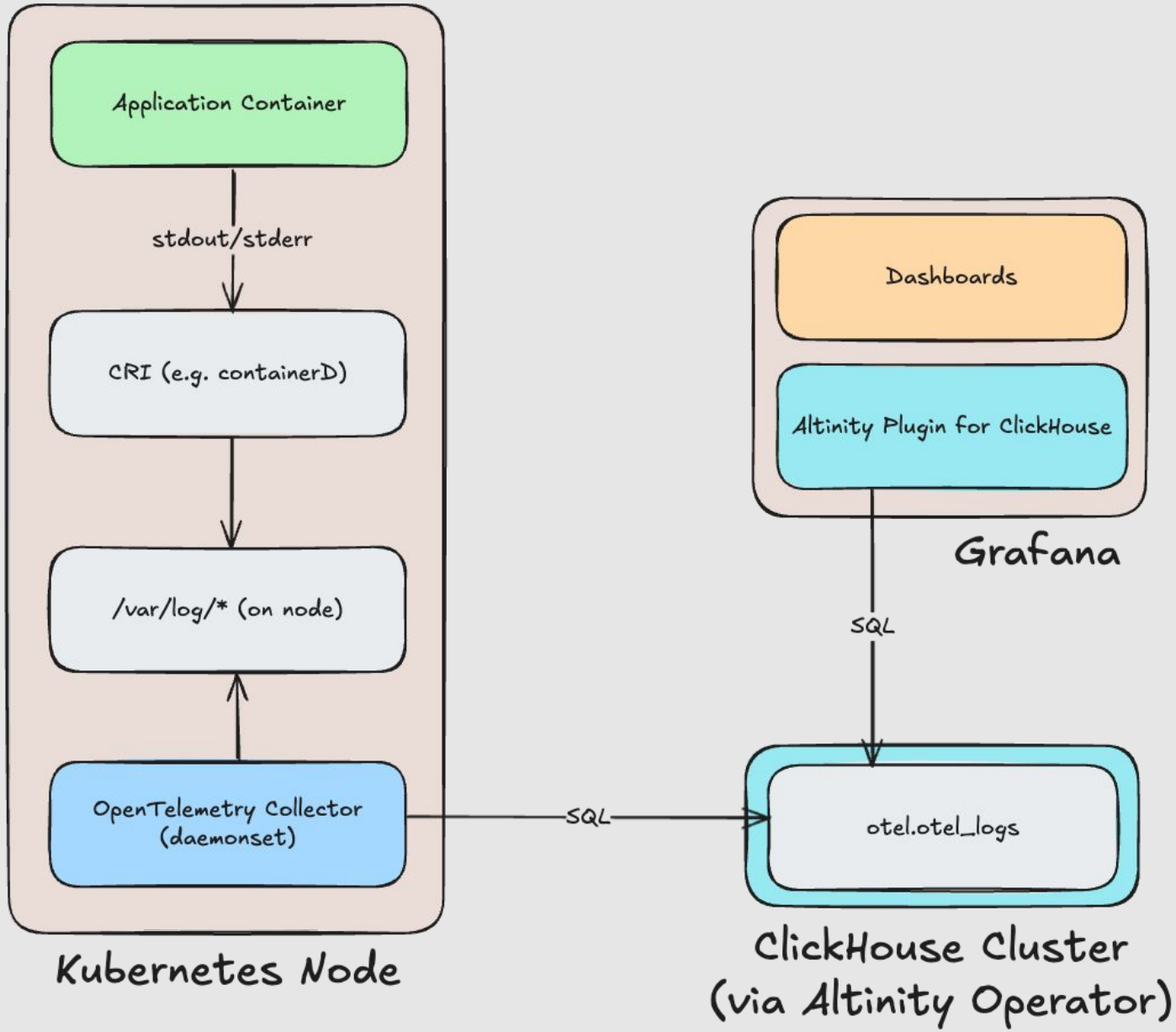
Derive metrics from the richness of trace metadata

Enrich logs and metrics with context

15

A complete observability solution



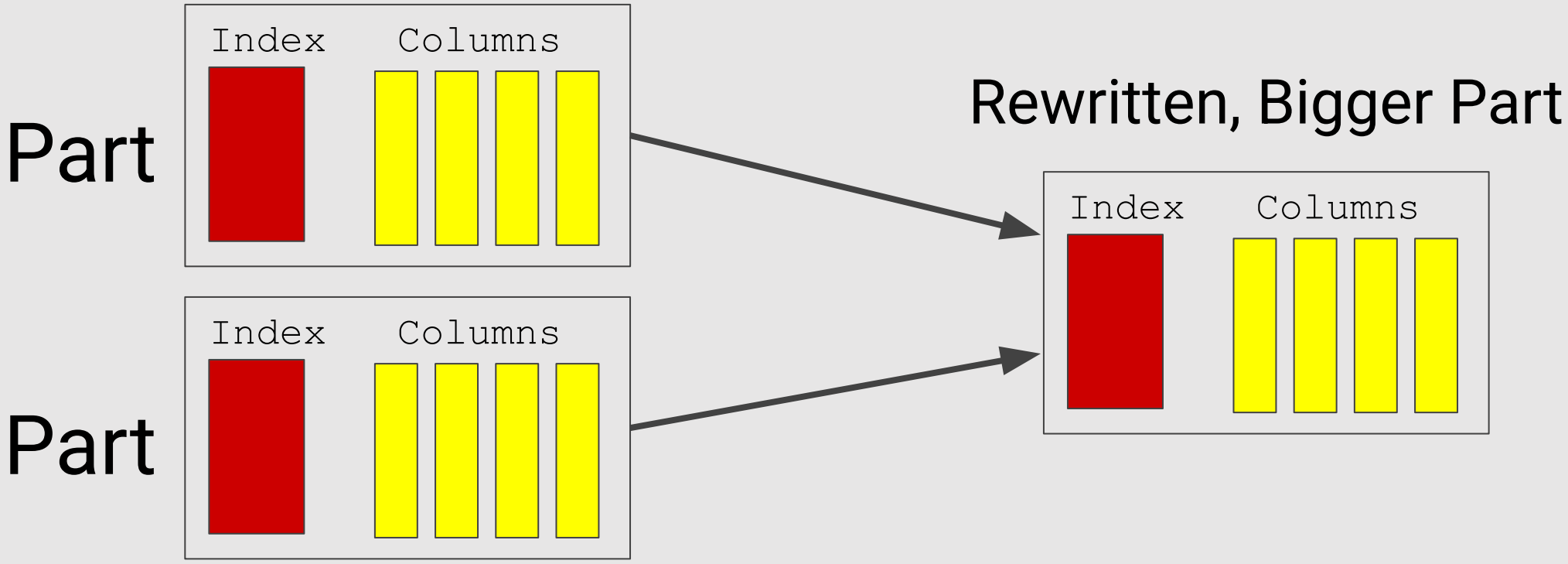

Application Container
stdout/stderr
CRI (e.g. containerD)
/var/log/* (on node)
OpenTelemetry Collector
(daemonset)
Kubernetes Node
SQL
Dashboards
Altinity Plugin for ClickHouse
Grafana
SQL
otel.otel_logs
ClickHouse Cluster
(via Altinity Operator)

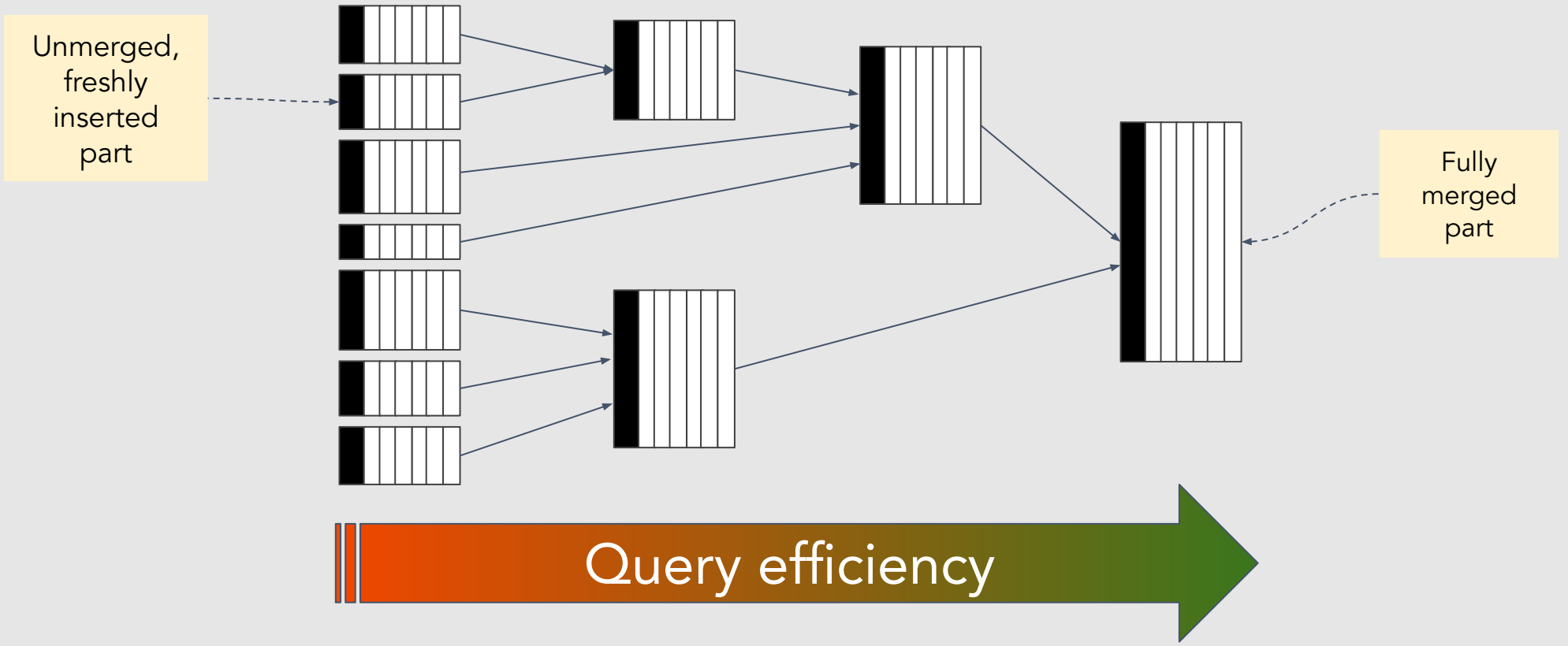
Introducing ClickHouse

SQL-compatible
Massively scalable
Really, really fast
Apache 2.0

Telemetry is WORM

Write-Once, Read-Many


Part
Index
Columns
Part
Index
Columns
Rewritten, Bigger Part
Index
Columns


Unmerged, freshly inserted part
Fully merged part
Query efficiency

ClickHouse for Observability

How does this help?

- Fast writes
- Time-friendly
- Easy cleanup
- Cost-effective

Integrations

- Grafana Datasource Plugin
- Jaeger w/ ClickHouse backend
- qryn (Prometheus, Loki, Tempo, +more)
- Kafka table engine
- OpenTelemetry Exporter

Observability for ClickHouse

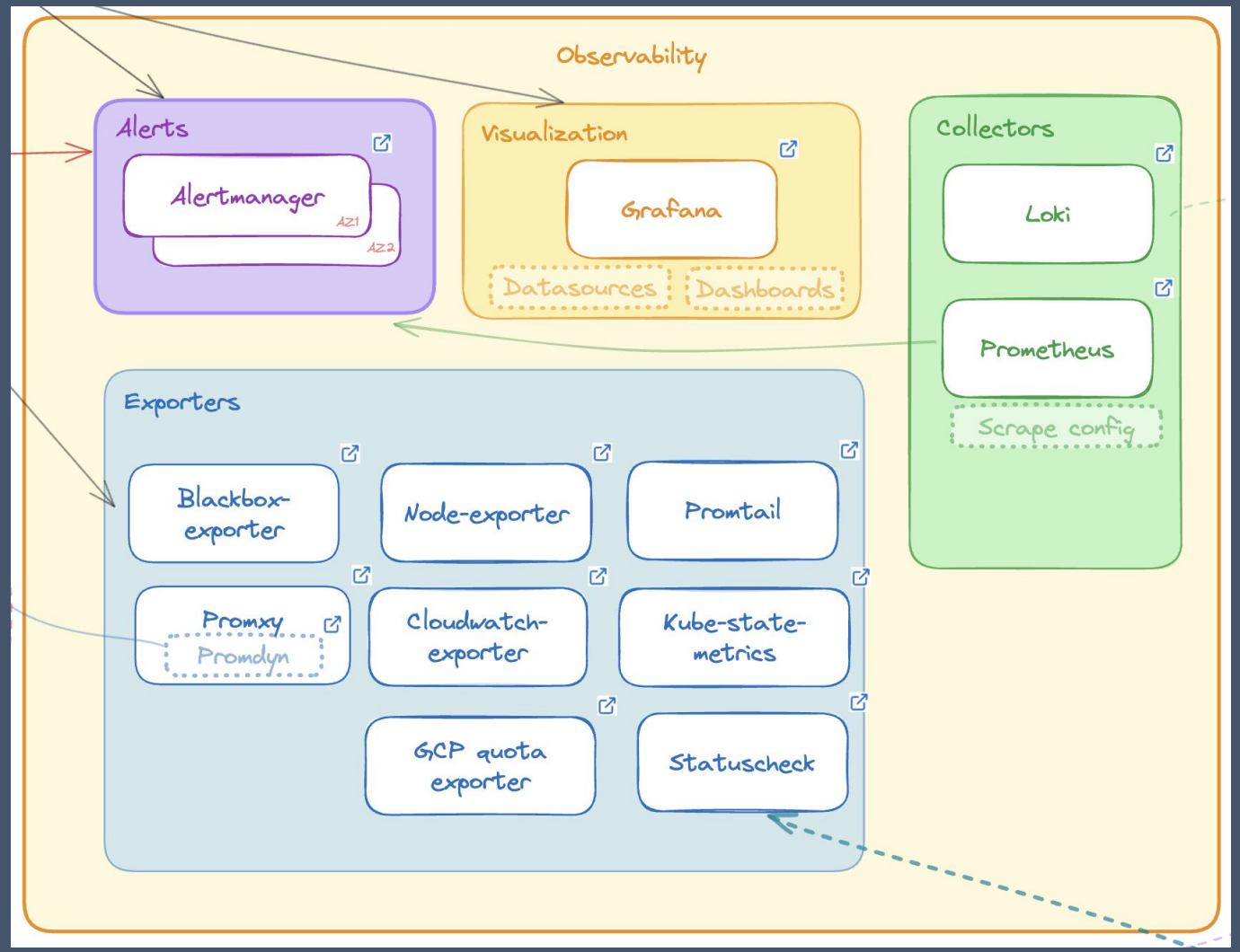
```
SELECT
    query_id,
    query_duration_ms,
    user,
    left(query, 20),
    written_bytes
FROM system.query_log
ORDER BY event_date DESC
LIMIT 1

Query id: e0214d77-e6b8-4e94-b34b-96dd13432364

    ┌─query_id─────────────────────────────┬─query_duration_ms─┬─user────┬─left(query, 20)─────┬─written_bytes─┐
1.  │ 605f93cf-5586-435b-8ffe-988b11cbc539 │                58 │ default │ INSERT INTO otel_tra │        726906 │
    └──────────────────────────────────────┴───────────────────┴─────────┴─────────────────────┴───────────────┘

1 row in set. Elapsed: 0.017 sec. Processed 12.15 thousand rows, 7.99 MB (725.40 thousand rows/s., 477.05 MB/s.)
Peak memory usage: 16.69 MiB.
```

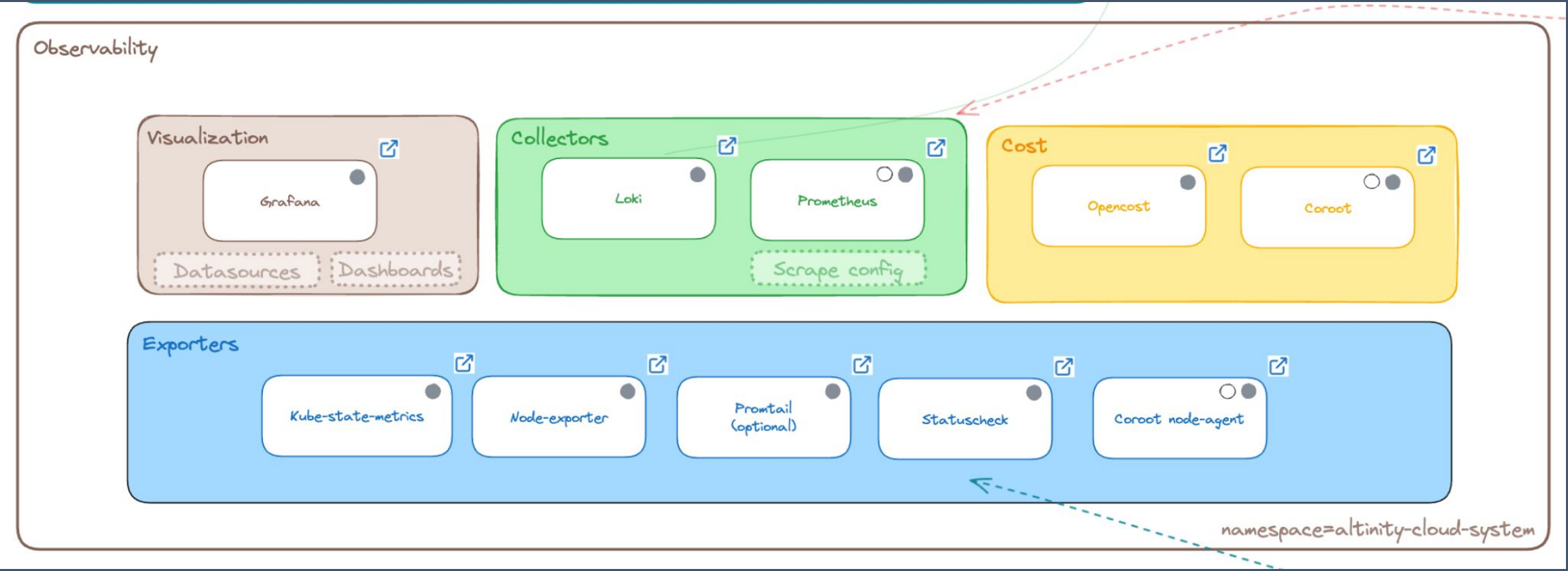
Observability for ClickHouse

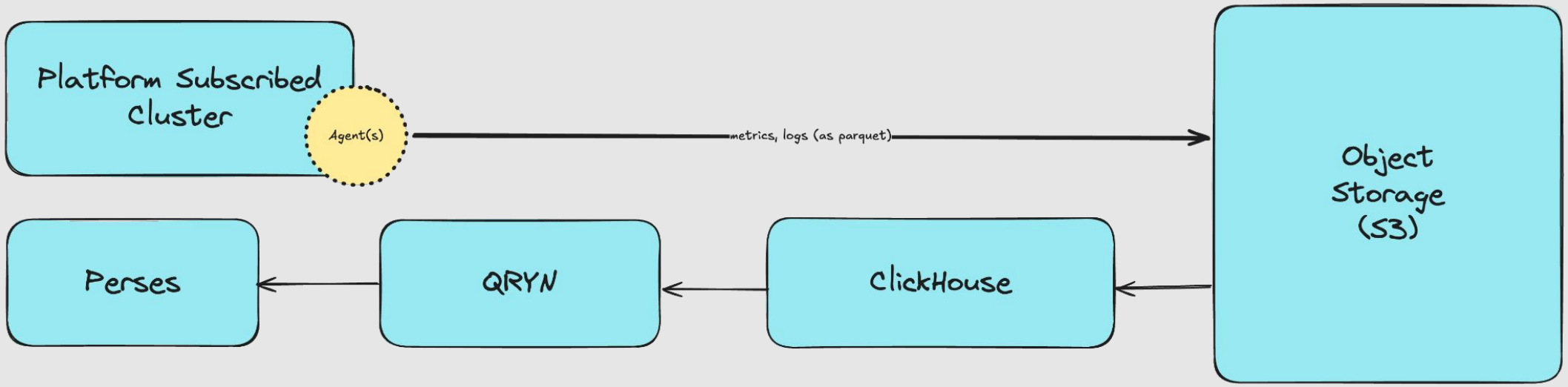
System tables
Built-in metrics
Built-in tracing
Built-in logs
Altinity Operator
Grafana Dashboards

System Tables

name	database	parts	total_rows
asynchronous_metric_log	system	84	31806721
metric_log	system	78	40947
part_log	system	81	146948
query_log	system	133	74777
query_views_log	system	133	8063
trace_log	system	85	1891241


Observability
Alerts
Alertmanager
AZ1
AZ2
Visualization
Grafana
Datasources
Dashboards
Collectors
Loki
Prometheus
Scrape config
Exporters
Blackbox-exporter
Node-exporter
Promtail
Promxy
Promdyn
Cloudwatch-exporter
Kube-state-metrics
GCP quota exporter
Statuscheck


Observability
Visualization
Grafana
Datasources
Dashboards
Collectors
Loki
Prometheus
Scrape config
Cost
Opencost
Coroot
Exporters
Kube-state-metrics
Node-exporter
Promtail
(optional)
Statuscheck
Coroot node-agent
namespace=altinity-cloud-system


Platform Subscribed Cluster
Agent(s)
metrics, logs (as parquet)
Object Storage (S3)
Perses
QRYN
ClickHouse

Dashboards

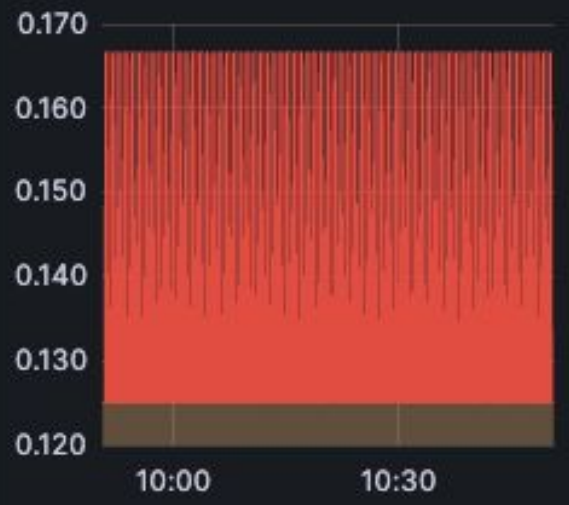
- grafana
  - anywhere-dashboards
    - anywhere-cloud-connect.json
    - anywhere-etcd.json
    - anywhere.json
  - dashboards
    - acm-metrics-renderer.json
    - anywhere-clickhouse-backup.json
    - anywhere-clickhouse-kafka.json
    - anywhere-clickhouse-operator.json

initial user | All | query_kind | All | min_duration_ms | Enter variable value | max_duration_ms | Enter variable value

All

Top charts

Top 5 request's rate by type: All; user: All; query kind: All



	max	avg
SELECT name, bytes_allocated, element_count, load_factor FROM system.dictionaries	0.167	0.135
SELECT database, table, sum(bytes_on_disk) AS bytes, count() AS parts, sum(rows) AS rows FROM system.p...	0.167	0.135
SELECT value, metric FROM system.metrics	0.167	0.135
SELECT value, event FROM system.events	0.167	0.135
SELECT value, metric FROM system.asynchronous_metrics	0.167	0.135

Top slow queries by type: All; user: All; qu...

query	duration_ms
SELECT name, bytes_allocated, element_count, load_factor FROM system.dictionaries	0.08
SELECT value, metric FROM system.metrics	0.21
SELECT value, metric FROM	

Top memory consumers by type: All; user:...

query	usage	count
system.dictionaries		
SELECT version()	8.95 KiB	480
SELECT database, table, sum(bytes_on_disk) AS bytes, count() AS parts,		

Top failed queries by user: All; query kind:...

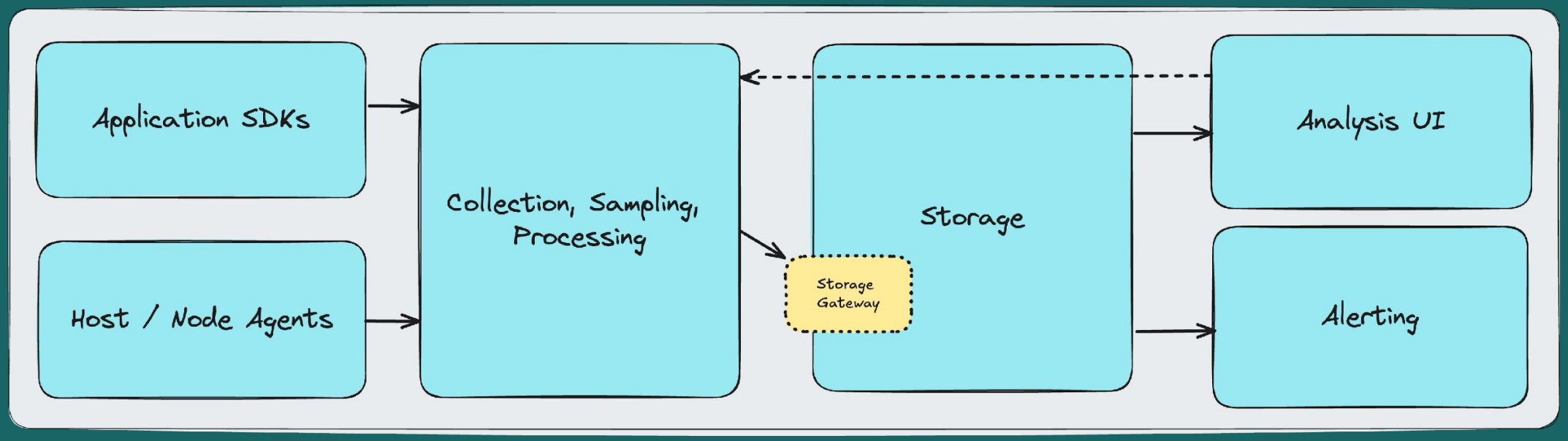
No data to show

Alerting

Alerts for known troublemakers

CrashLoopBackOff
KubeJobFailed
KubeDaemonSetRolloutStuck
KubePersistentVolumeUsage
PrometheusIsDown

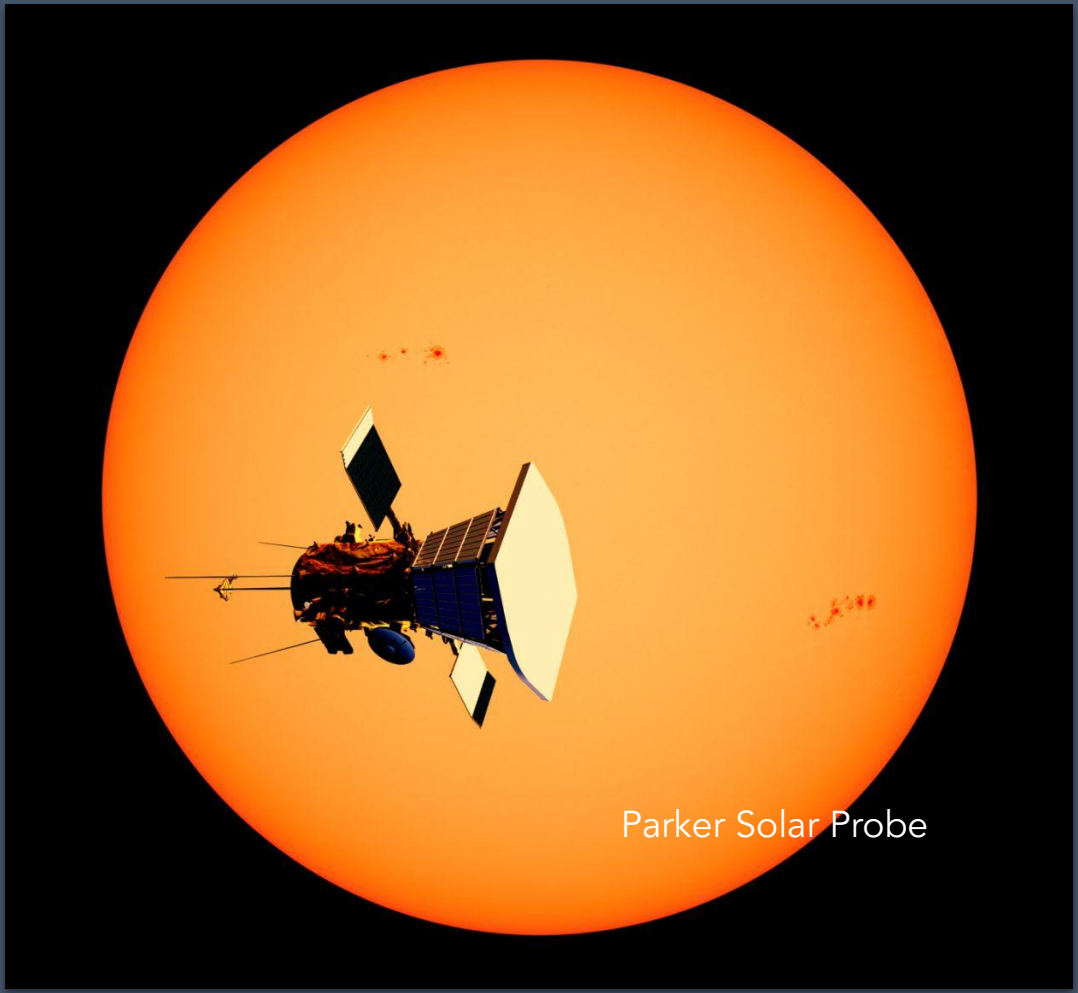
A complete observability solution



eBPF

Provides external observability into any syscalls made by a target process.

Allows network request mapping.



Coroot 1.5.0 is available 🎉 Changelog (how to upgrade)

coroot :~#

default

search for apps and nodes

last hour

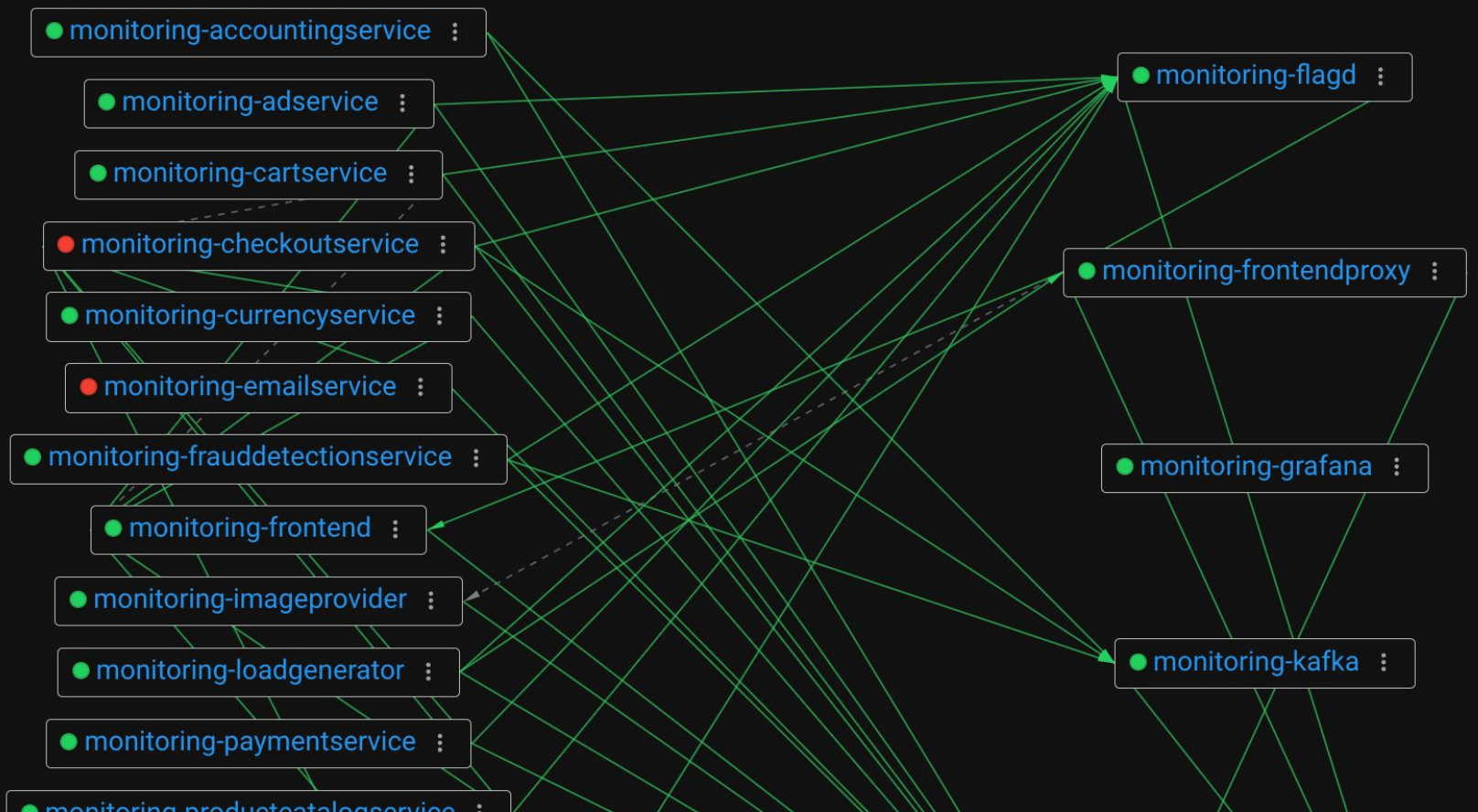
Overview

HEALTH | SERVICE MAP | TRACES | NODES | DEPLOYMENTS | COSTS

search

namespaces: monitoring

application | control-plane | monitoring +



Data Ingestion

Ingesting data with **qryn** is easy and painless thanks to our **polyglot** design.

Use any **Agent** or **Library** compatible with *Opentelemetry, Loki, Datadog, Elastic, Prometheus, Tempo, Graphite, Pyroscope & more*

Logs | Metrics | Telemetry | Profiling | Custom

Get started using the Log Ingestion section

API Support

Food for Thought…

1. Continuity
2. Annotations, Normalization, Filtering (ETL to ELT)
3. Combining App & Infra Views
4. Advanced Analytics
5. Data-mining + Model Training

40

Wrapping Up

Start with what you have

Focus on known troublemakers

Guard against axes of change

Never stop adapting

Thank You OpenSearch Community!

ALTINITY®

Connect with me

We're seeking OSS contributors!
Join our Slack
altinity.com/slack