

Taming Tenancy, Cost and Architecture at Collibra

Through OpenTelemetry and our
Telemetry Backbone

Alex Van Boxel

Almost 30 years in the sector

Mostly as Software Engineer

Web - 3D - Middleware - Mobile - Big Data

More recent as Architect

Data - SRE - Infrastructure

Community

Apache Beam contributor

OpenTelemetry Collector contributor

Collibra

Principal Systems Architect



Collibra

**A data intelligence platform
powered by active metadata**

AI Governance

Data Catalog

Data Governance

Data Lineage

Data Notebook

Data Privacy

Data Quality & Observability

Protect

"How much does X cost?"

Marks

Automatic

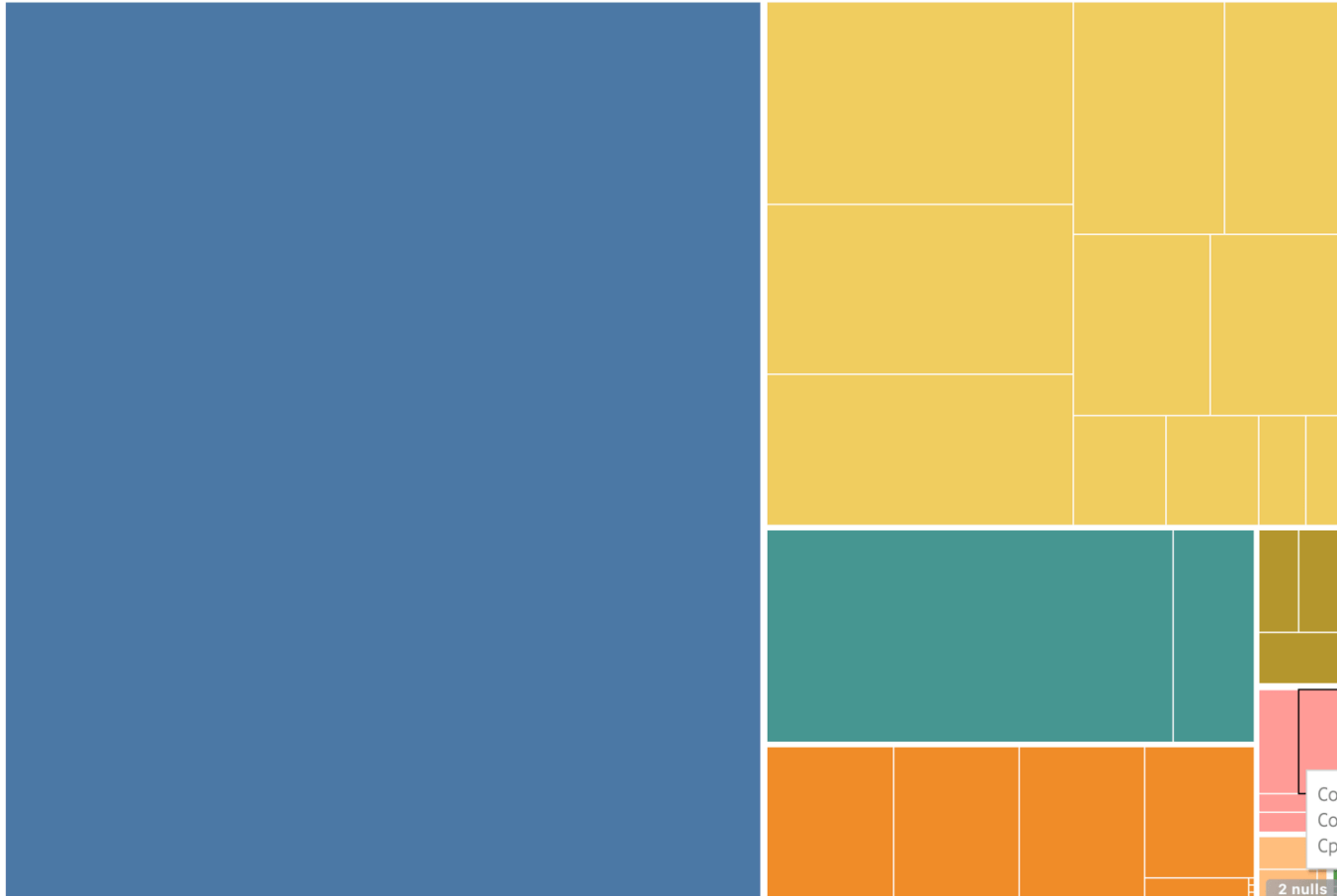
Color Size Label

Detail Tooltip

Collibra C4 System

SUM(Cpu Reservation Vcpu...

Collibra C4 Container



for heat maps use:

Measure Measure Dimension

Choose for me

Collibra C4 Container: colk8snode
Collibra C4 System: telemetry
Cpu Reservation Vcpu Hr: 326

2 nulls

"Compare the Usage and Reservation?"

Pages

Filters

- SUM(Container Cpu Usage Vcpu Hr)
- SUM(Cpu Reservation Vcpu Hr)
- Action (C4 System, K8S Container Na..)
- Action (C4 System, K8S Container Na..)

Marks

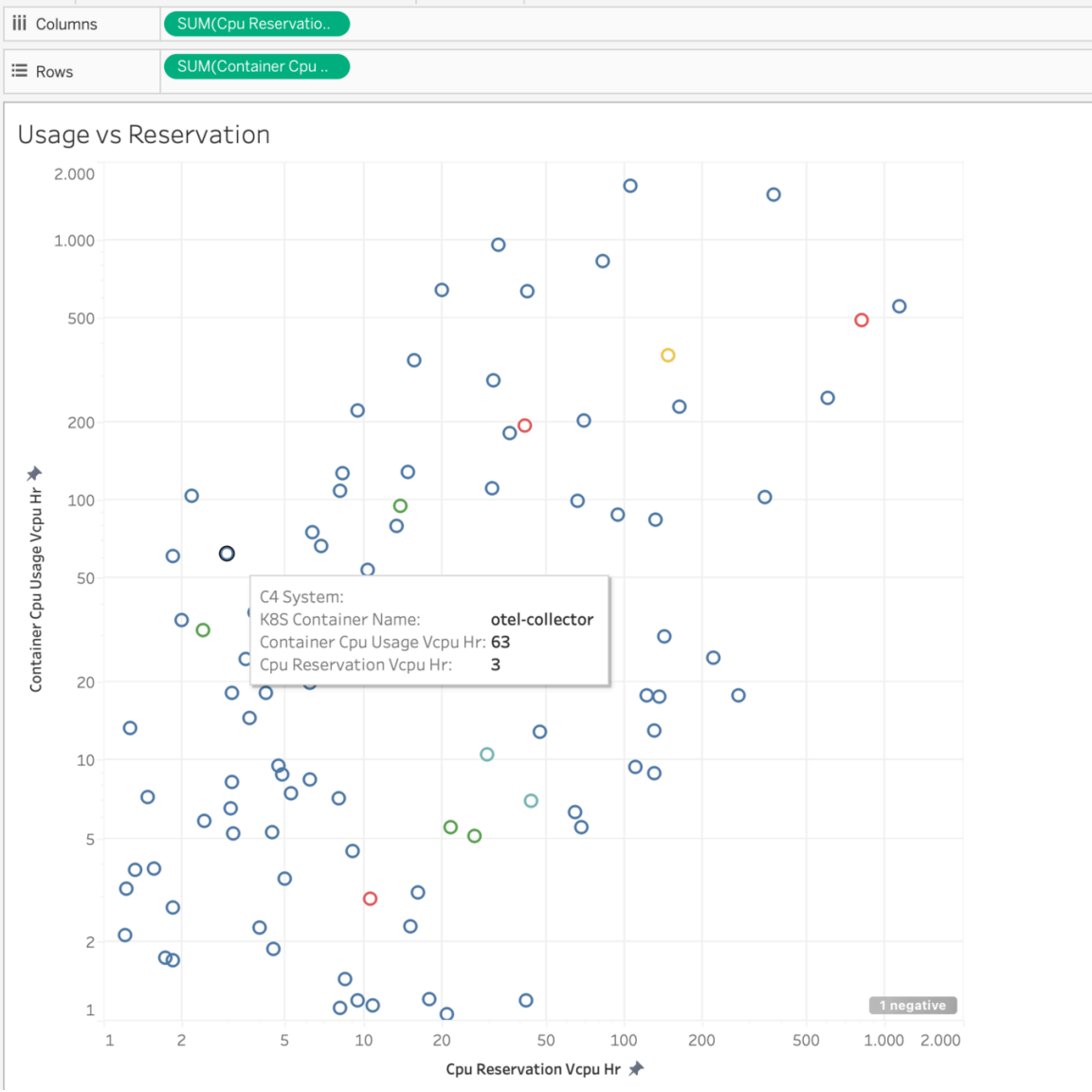
Automatic

Color Size Label

Detail Tooltip Shape

C4 System

K8S Container Name



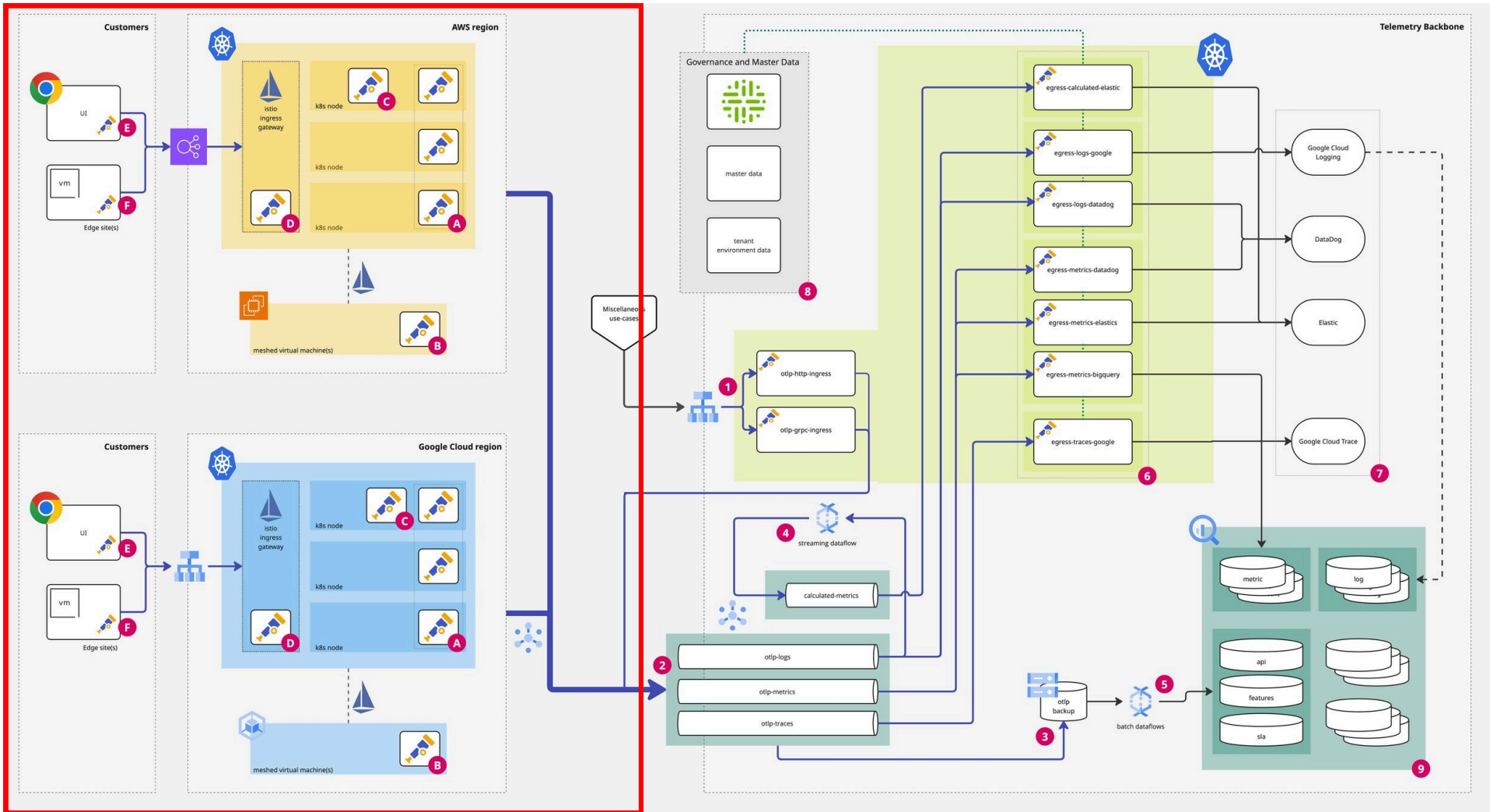
**Trend analysis per
tenant**

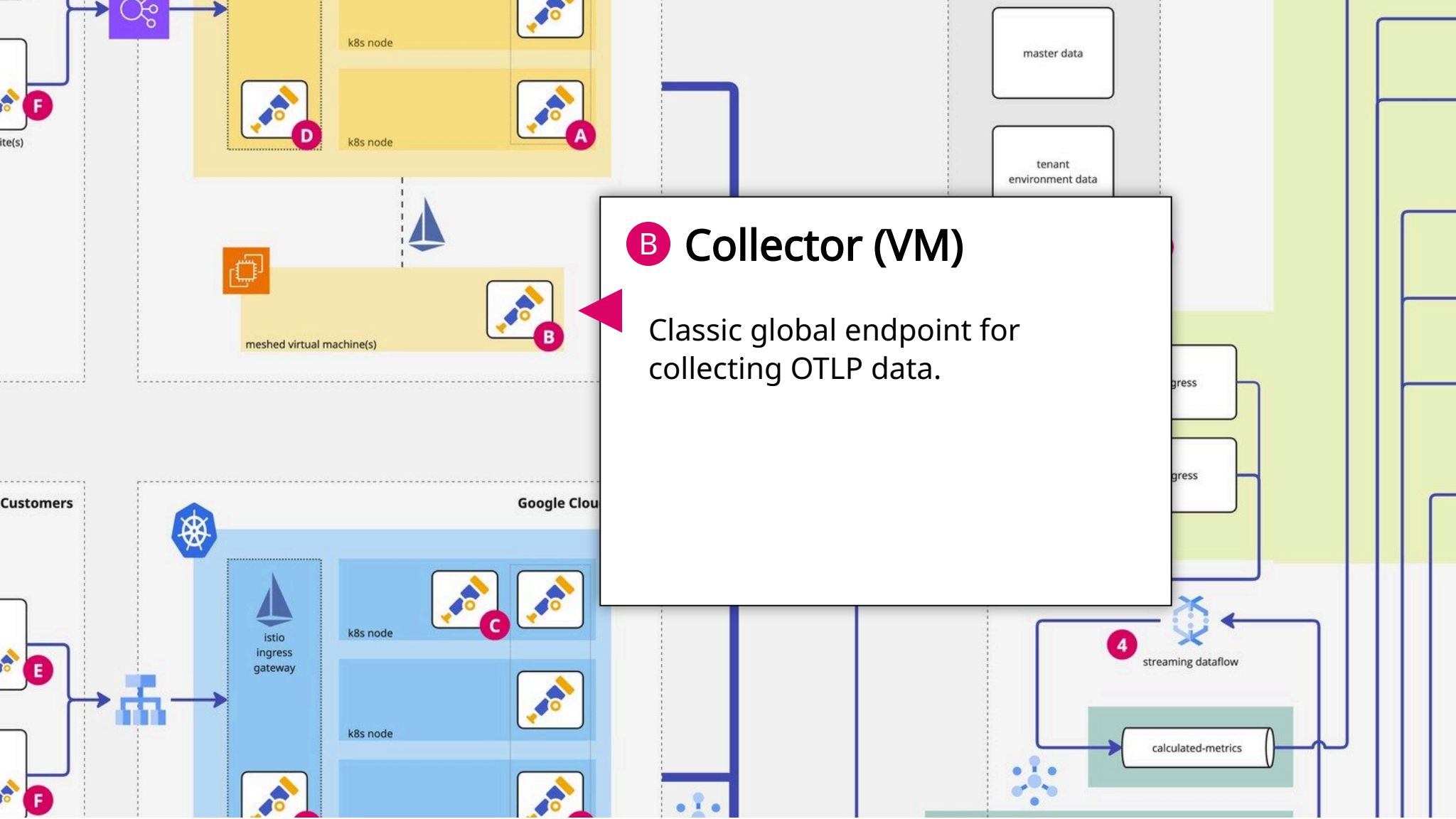
Architecture and Collection at the Edge

Collibra Architecture

- **On-prem heritage:** single deployable monolith → hosted SaaS on VMs (single-tenant isolation for free)
- **Kubernetes shift:** microservices for team velocity & polyglot (Python now dominant in AI)
- **The pivot:** "container per service per tenant" was unsustainable → shared multi-tenant on K8s

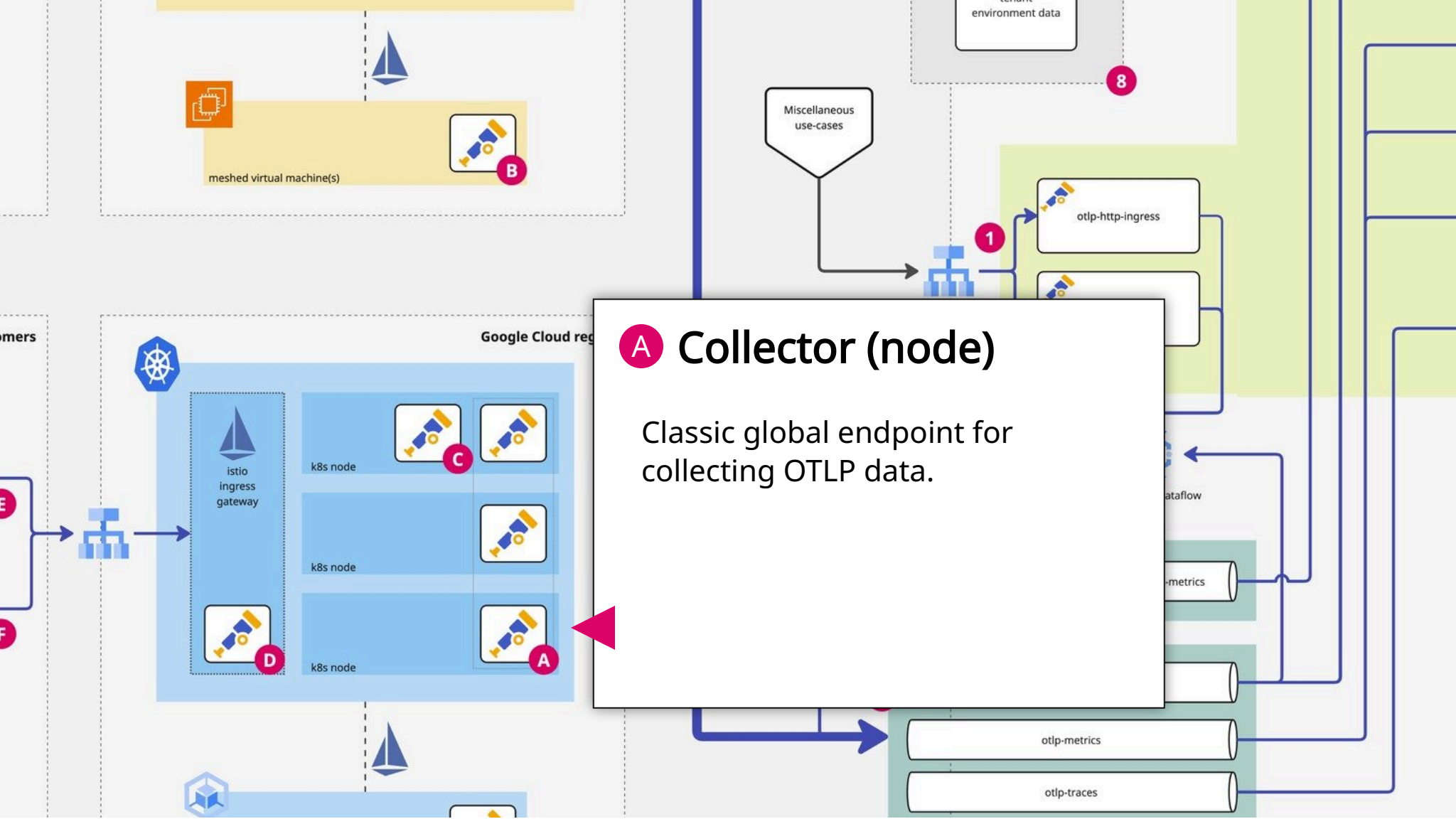
Shared multi-tenancy saves cost — but makes it harder to figure out the cost per tenant... how can we solve this?





B Collector (VM)

Classic global endpoint for collecting OTLP data.

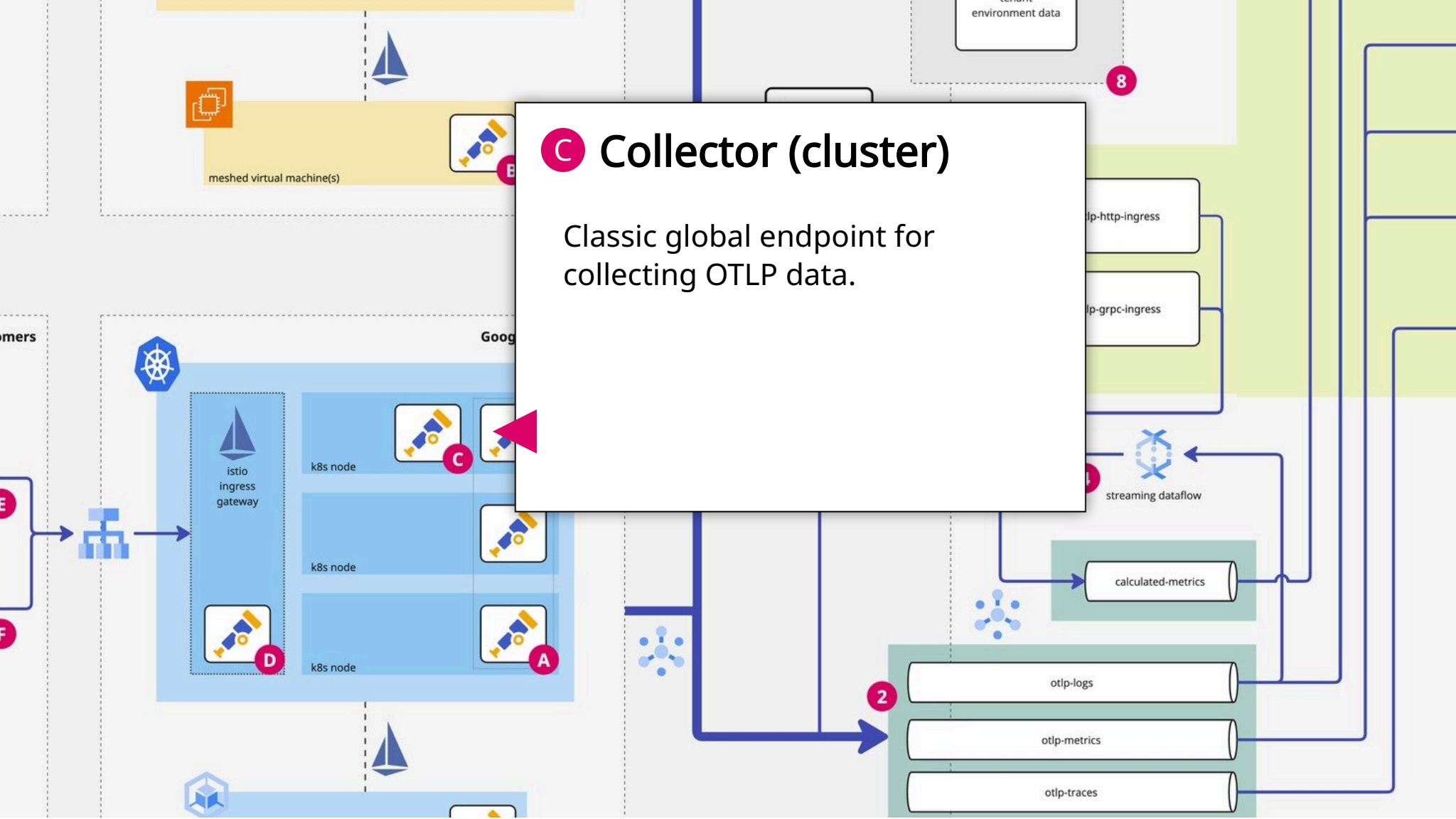


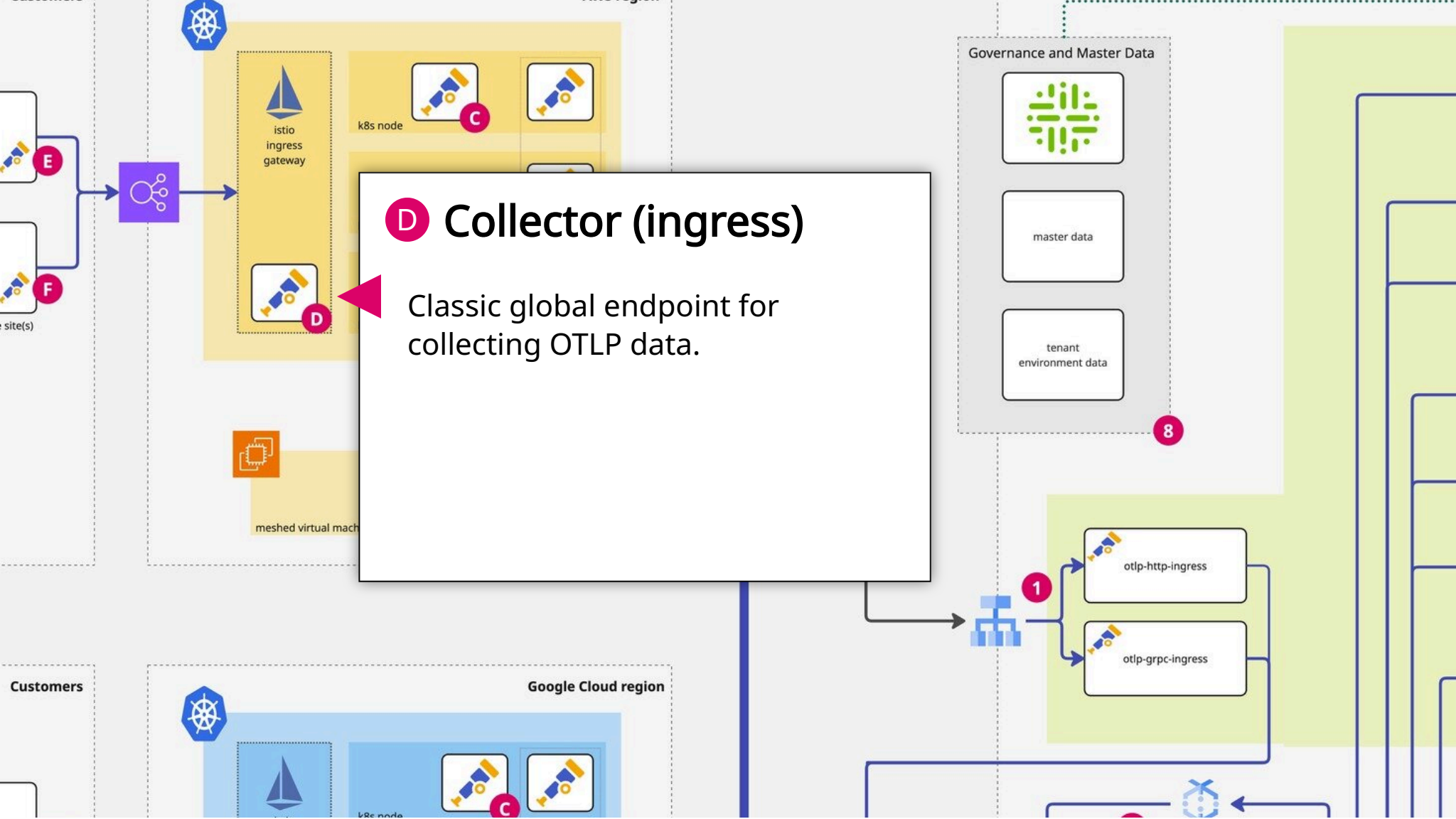
A Collector (node)

Classic global endpoint for collecting OTLP data.

C Collector (cluster)

Classic global endpoint for collecting OTLP data.





D Collector (ingress)

Classic global endpoint for collecting OTLP data.

Governance and Master Data



master data

tenant environment data

8

1

otlp-http-ingress

otlp-grpc-ingress

C

E

F

site(s)

istio ingress gateway

k8s node



D



meshed virtual machine

Google Cloud region

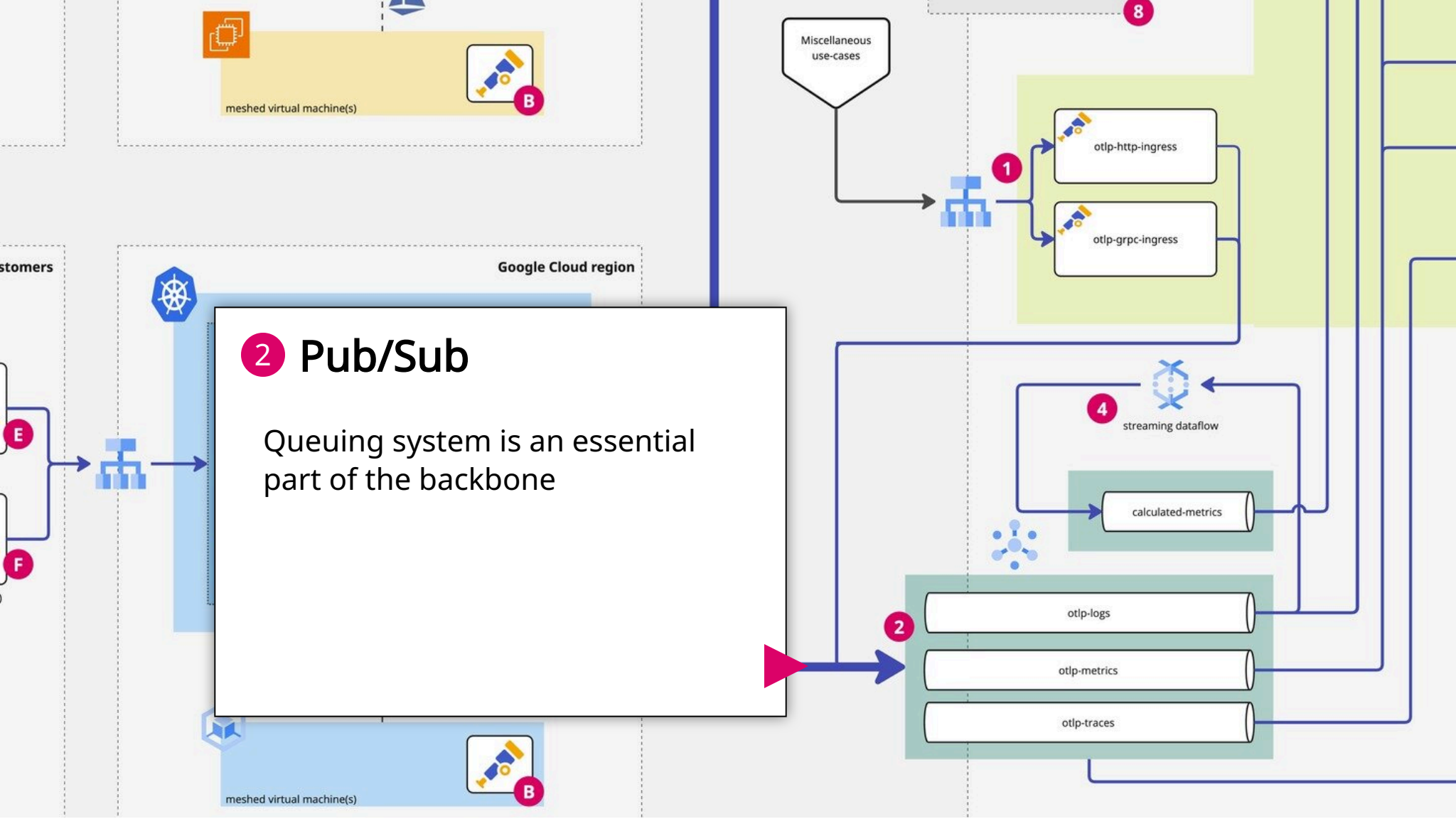


k8s node

Customers

2 Pub/Sub

Queuing system is an essential part of the backbone



Golden Signals

OpenTelemetry Attributes

Gold Attr. #1: Tenant

`collibra.tenant.environment_id`

- VMs Resource Attributes - Configured at the collector
- Pod Resource Attributes - For single tenant pods
- Multi-tenant Pod - Signal Attributes

Gold Attr. #1: Tenant

`collibra.tenant.environment_id`

- VMs Resource Attributes - Configured at the collector
- Pod Resource Attributes - For single tenant pods
- Multi-tenant Pod - Signal Attributes

```
1 {  
2   "event_name": "workflow:started",  
3   "tenant_environment_id": "...",  
4   "asset_id": "..."  
5 }
```

CSTE - Collibra
Structured Telemetry
Event: Events are our
golden signal

Gold Attr. #1: Tenant

`collibra.tenant.environment_id`

- VMs Resource Attributes - Configured at the collector
- Pod Resource Attributes - For single tenant pods
- Multi-tenant Pod - Signal Attributes

Multi-tenant service? Dev's responsibility to add signals in code, eg. Mapped Diagnostic Context

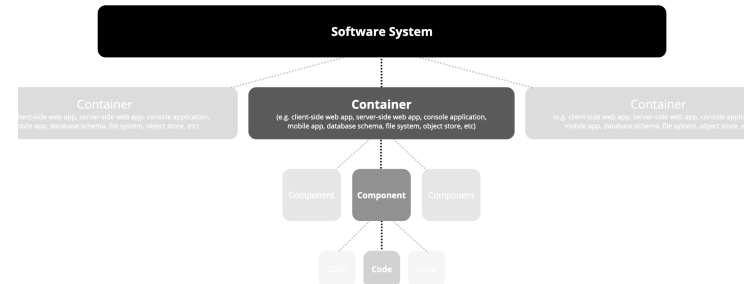
```
1 {
2   "event_name": "workflow:started",
3   "tenant_environment_id": "...",
4   "asset_id": "...",
5 }
```

```
1 MDC.put("tenant_environment_id",
2         ctx.getTenantEnvironmentId());
3 try {
4   // all logs in this thread
5 } finally {
6   MDC.clear();
7 }
```

Gold Attr. #2: Architecture

<https://c4model.com/> - The C4 model is an easy to learn, developer friendly approach to software architecture diagramming (by Simon Brown)

- **System** - logical product capability
- **Container** - service, logical database, topic, module
- **Deployment Node** - where it runs (can nest)



A **software system** is made up of one or more **containers** (applications and data stores), each of which contains one or more **components**, which in turn are implemented by one or more **code** elements (classes, interfaces, objects, functions, etc).

Gold Attr. #2: Architecture

<https://c4model.com/> - The C4 model is an easy to learn, developer friendly approach to software architecture diagramming (by Simon Brown)

- **System** - logical product capability
- **Container** - service, logical database, topic, module
- **Deployment Node** - where it runs (can nest)

```
collibra.c4.system  
collibra.c4.container  
collibra.c4.deployment
```

Gold Attr. #2: Architecture

- Pod Resource Attributes - Easy with 1:1 mapping

```
1 labels:  
2   c4.collibra.com/system: telemetry  
3   c4.collibra.com/container: colkyverno
```

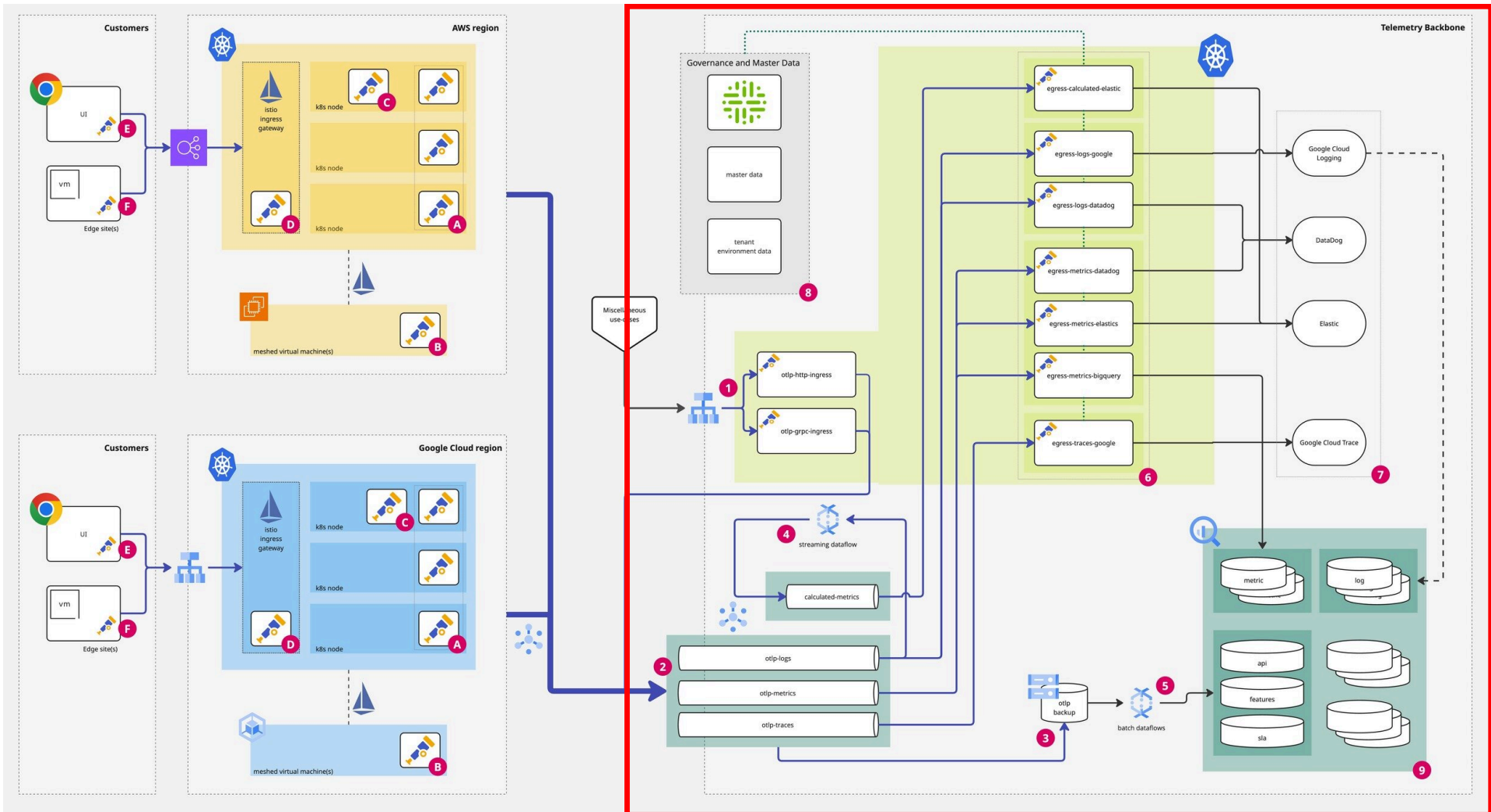


```
1 collibra.c4.system: telemetry  
2 collibra.c4.container: colkyverno
```

- **Modular Monoliths Signal Attributes** - It's not only out single tenant core, but also k8s jobs

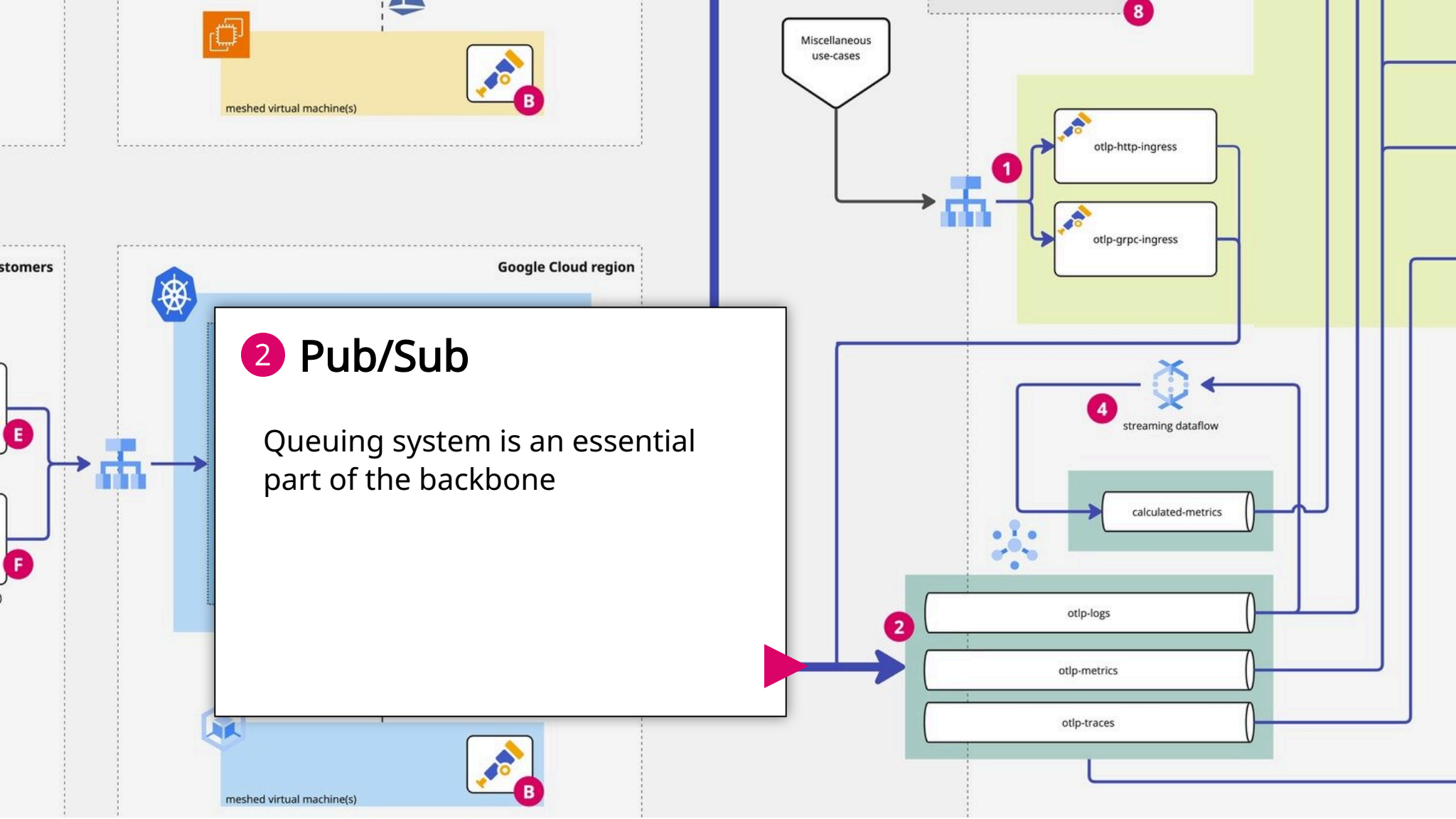
Modular Monoliths (it becomes the responsibility for devs

Telemetry Backbone Enrichment and Routing



2 Pub/Sub

Queuing system is an essential part of the backbone



AWS region



istio
ingress
gateway

8 Master Data

Can be sourced from different systems to merge into the data



D



meshed virtual machine(s)



B

Miscellaneous
use-cases

Governance and Master Data



master data

tenant
environment data

8

1

otlp-http-ingress

otlp-grpc-ingress

Google Cloud region



Governance and Master Data



7 Pipelines and Backends

Parallel pipelines do the processing, enrichment, filtering, calculation and backup to our backends

otlp-grpc-ingress

egress-calculated-elastic

egress-logs-google

egress-logs-datadog

egress-metrics-datadog

egress-metrics-elastics

egress-metrics-bigquery

egress-traces-google

Google Cloud Logging

DataDog

Elastic

Google Cloud Trace

6

7



In-Flight Enrichment from Master Data

- JSON field promotion: body fields → signal OpenTelemetry attributes
- Master data lookup: keyed by `collibra.tenant.environment_id`


Devs don't need to know contract terms or support levels — they just log the tenant environment ID, and the backbone dynamically infers and injects the rest.

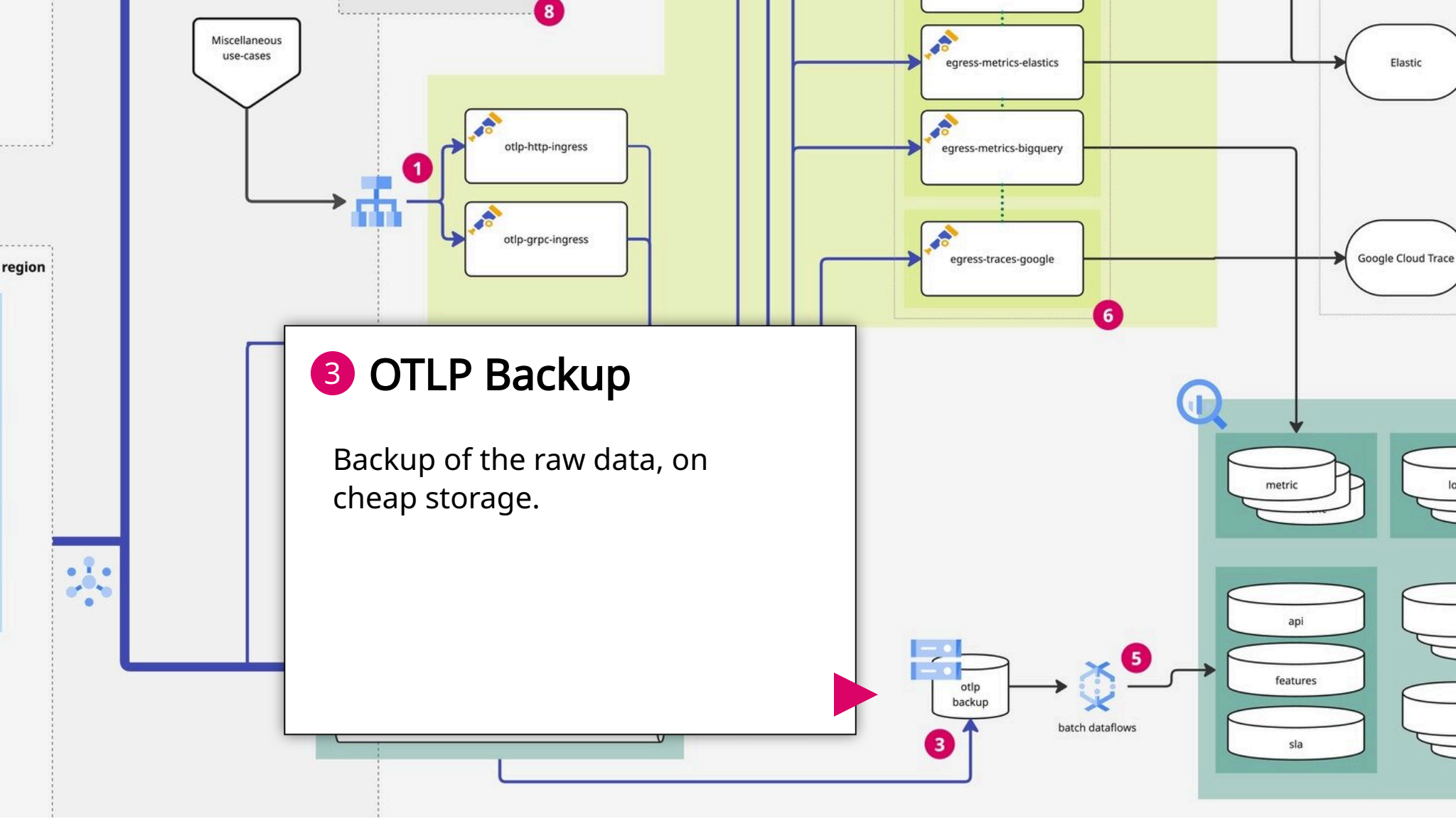


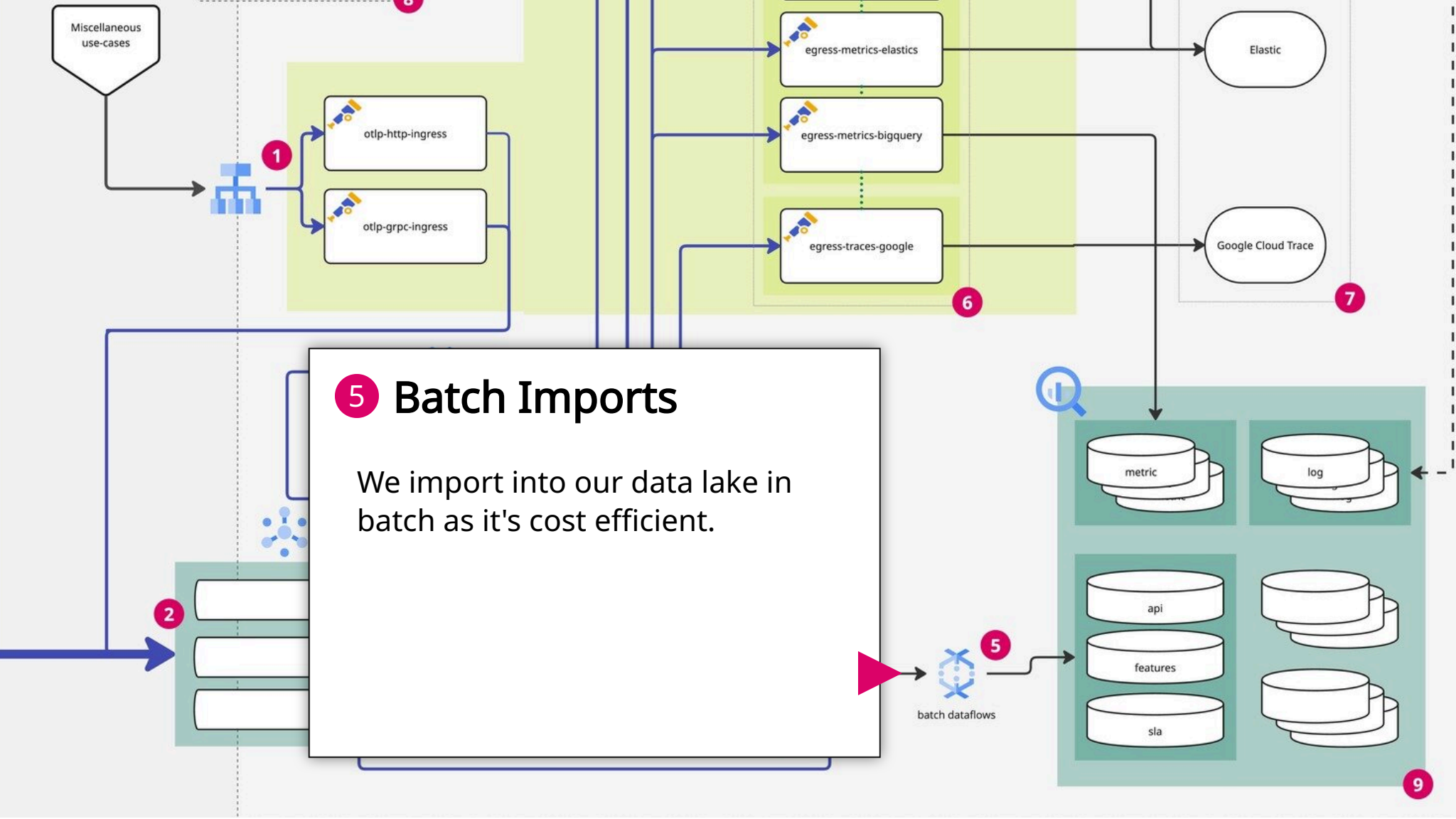
OpenAPI Reverse- Mapping

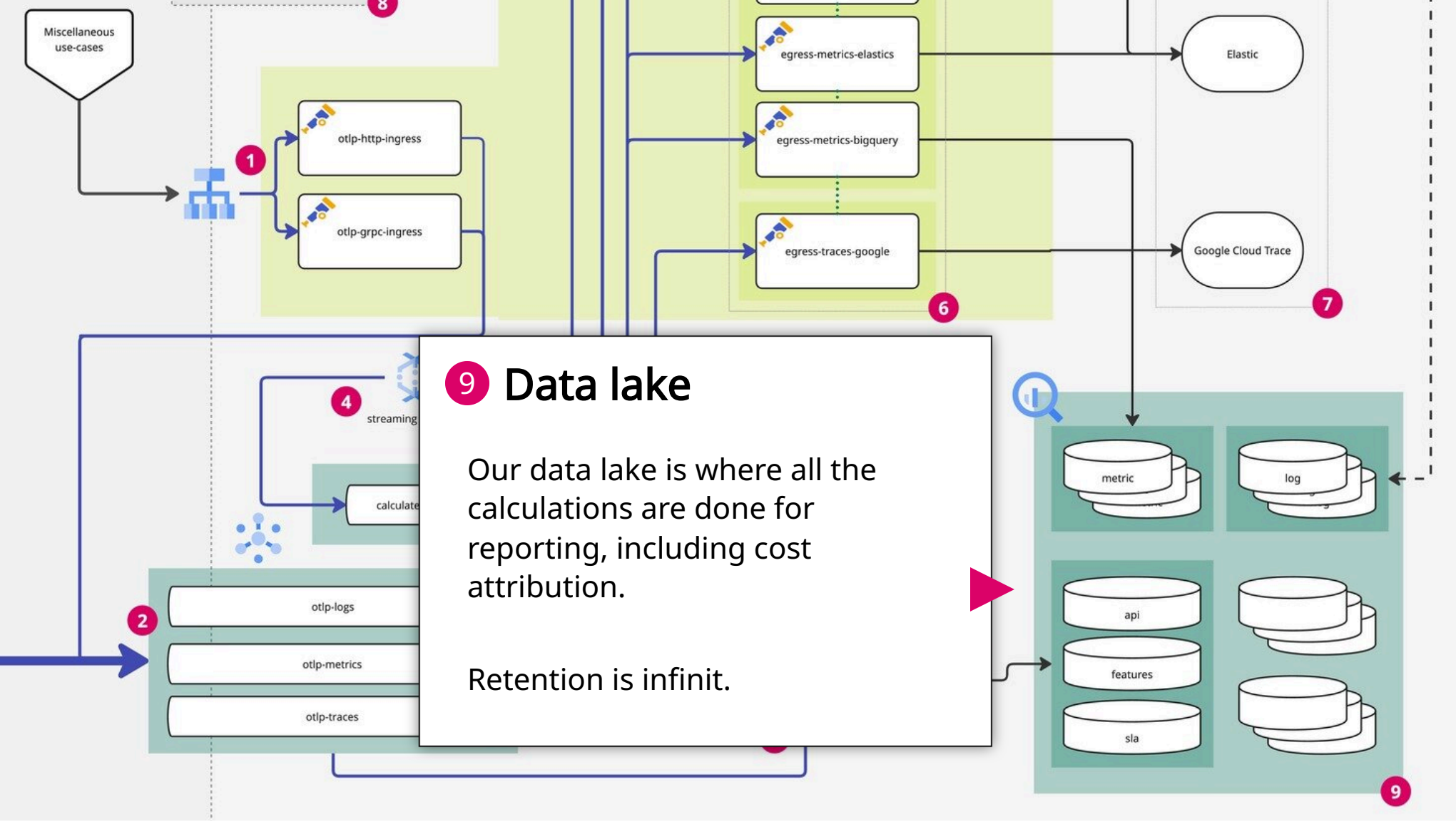


"We don't measure
URLs. We measure
contracts."

- URL cardinality explosion bloats metric DBs and costs
 - Reverse-map Istio URL + method → OpenAPI operationId
 - Low-cardinality, semantic endpoint stream → automated SLOs across all microservices
 - Also: aggressively drop runtime spam & infra-sweep noise before vendors see it
- 








Cost Attribution - Closing the Loop

- **Telemetry volume cost** — aggregate signal volume per C4 system × tenant
- **Compute cost slicing** — CPU / mem / disk / network by tenant and C4
- **C4-aware provisioning** — Collibra Infra CRDs carry C4 metadata; cloud billing maps to logical owner

Open problem: defensible "virtual dollar" formula for cross-team chargebacks.

Semantic Conventions and wiring

Semantic Conventions

Collibra    AI Copilot 

Engineering / Architecture

OpenTelemetry Registry

Business Asset Domain ⓘ | Business Steward:  Alex Van Boxel













   **Actions** ▾

Overview **Assets** Responsibilities History Attachments

OpenTelemetry Registry [Select view](#) ▾



<input type="checkbox"/> Name ↑	Status	Asset Type
<input type="checkbox"/> ▾ Collibra Core Attributes	 Approval Pending	 OpenTelemetry Attribute Group
<input type="checkbox"/> collibra.host.role	Obsolete	 OpenTelemetry Attribute
<input type="checkbox"/> collibra.tenant.environment_id	Accepted	 OpenTelemetry Attribute
<input type="checkbox"/> collibra.tenant.environment_name	Accepted	 OpenTelemetry Attribute
<input type="checkbox"/> collibra.tenant.environment_purpose	Accepted	 OpenTelemetry Attribute
<input type="checkbox"/> collibra.tenant.group_id	Accepted	 OpenTelemetry Attribute
<input type="checkbox"/> > Collibra Deprecated Attributes	Approval Pending	 OpenTelemetry Attribute Group
<input type="checkbox"/> ▾ Collibra Network Attributes	Approval Pending	 OpenTelemetry Attribute Group
<input type="checkbox"/> collibra.network.component	In Progress	 OpenTelemetry Attribute
<input type="checkbox"/> collibra.network.direction	In Progress	 OpenTelemetry Attribute
<input type="checkbox"/> > Collibra Tenant Management Attributes	Approval Pending	 OpenTelemetry Attribute Group

Semantic Conventions

Collibra Engineering / Architecture / Semantic Conventions

collibra.network.envoy_access_log.v1

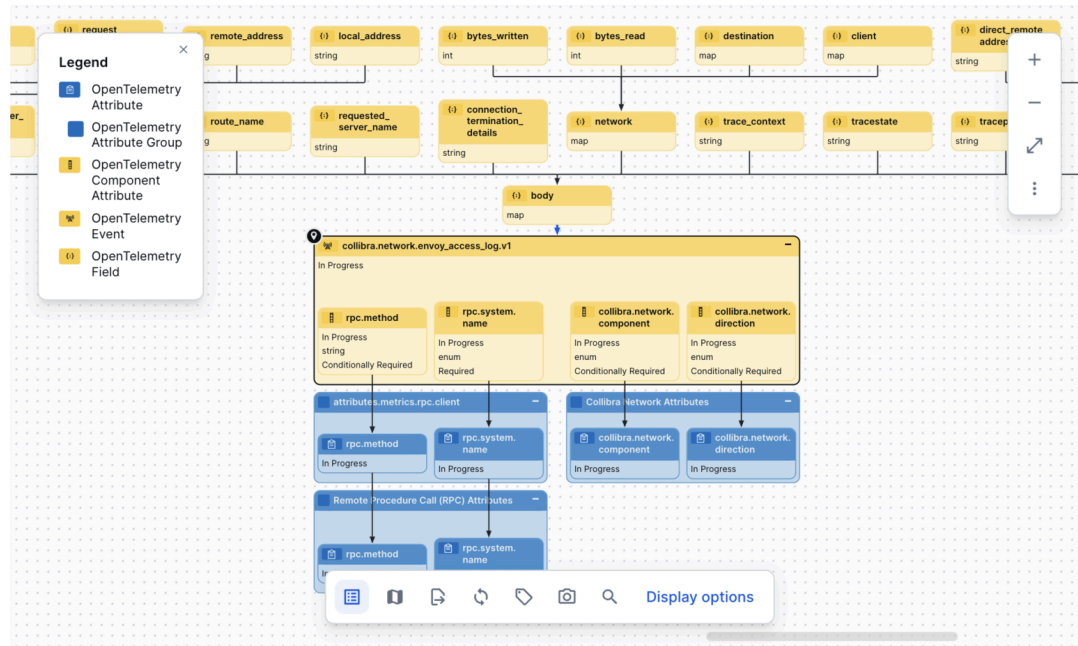
OpenTelemetry Event In Progress

Actions

Summary **Diagram** Pictures Responsibilities History Attachments

Switch view **OpenTelemetry Event**

Create new view



collibra.network.envoy_access_log.v1

Semantic Conventions

Created on Apr 22, 2026

Last modified Apr 22, 2026

Status In Progress

Overview

Description

Envoy proxy access log event, capturing HTTP request and response details emitted by Istio's Envoy's access logging subsystem. This log could be enriched with additional reverse OpenAPI operations in the backbone.

contains Data Element

body

has attribute OpenTelemetry Component Attribute

[collibra.network.component](#)
[collibra.network.direction](#)
[rpc.method](#)
[rpc.system.name](#)

Wiring - SemConv + Weaver

```
1 groups:
2   - id: event.collibra.network
3     stability: development
4     type: event
5     name: collibra.network.envoy_access_log.v1
6     brief: >
7       Envoy proxy access log event, capturing HTTP request and response details
8       emitted by Istio's Envoy's access logging subsystem. This log could be
9       enriched with additional revers OpenAPI operations in the backbone.
10    body:
11      # this body contains the envoy access logs configured in the following values file:
12      # https://github.com/collibra/cloud-platform-workload-helm-chart/blob/main/values.yaml (see accessL
13      id: collibra.network.envoy_access_log.v1
14      requirement_level: required
15      stability: development
16      type: map
17      fields:
18        - id: source
19          type: string
20          stability: development
21          brief: Identifies the source of the access log entry.
22          requirement_level: recommended
23          examples: ["envoy"]
24    - id: timestamp
```

Wiring - More YAML

```
1  apiVersion: wire.collibra.com/v1
2  kind: System
3  metadata:
4    name: telemetry
5  containers:
6    "colk8snode":
7      deploymentNodeRef: [ "chi-regional" ]
8    "colk8sevents":
9      deploymentNodeRef: [ "chi-regional" ]
10   "colhttpingress":
11     deploymentNodeRef: [ "chi-regional" ]
12     namespace: "chi-opentelemetry"
13     paths:
14       - apiRef: "telemetry.otlp.public.v1"
15         ingress:
16           # will require additional service
17           prefix: "/otlp/"
18           gateways: ["public"]
19   "colistiotraces":
20     deploymentNodeRef: [ "chi-regional" ]
21   "colk8smetrics":
22     deploymentNodeRef: [ "chi-regional" ]
23   "colkyverno":
24     deploymentNodeRef: [ "chi-regional" ]
```

**Future
and takeaway**

OpAMP — Pushing Control to the Collection Edge

- Bandwidth problem: backbone filtering saves on vendors, but raw telemetry still costs WAN egress
- OpAMP: dynamic management & configuration of the entire collector fleet
- Adaptive edge sampling: normal tenant → aggressive sampling; incident → dial up fidelity at source

Key Takeaways

① Golden attributes on day one

Define tenancy and architecture dimensions before you split into microservices, not after.

② Decouple with a backbone

Buffer-first ingestion (Pub/Sub) + centralized enrichment unlocks both ops and FinOps / BI.

③ Invest in semantic contracts

They structure your signals today and become the foundation for AI diagnostic agents tomorrow.