

Ensure trust in your data from pipelines to prompts

DATAVERSITY Demo Day: Data Quality / DataOps

October 21, 2026



DATADOG

Agenda

01 Intro to Datadog

02 Why reliable data matters now

03 Data Observability

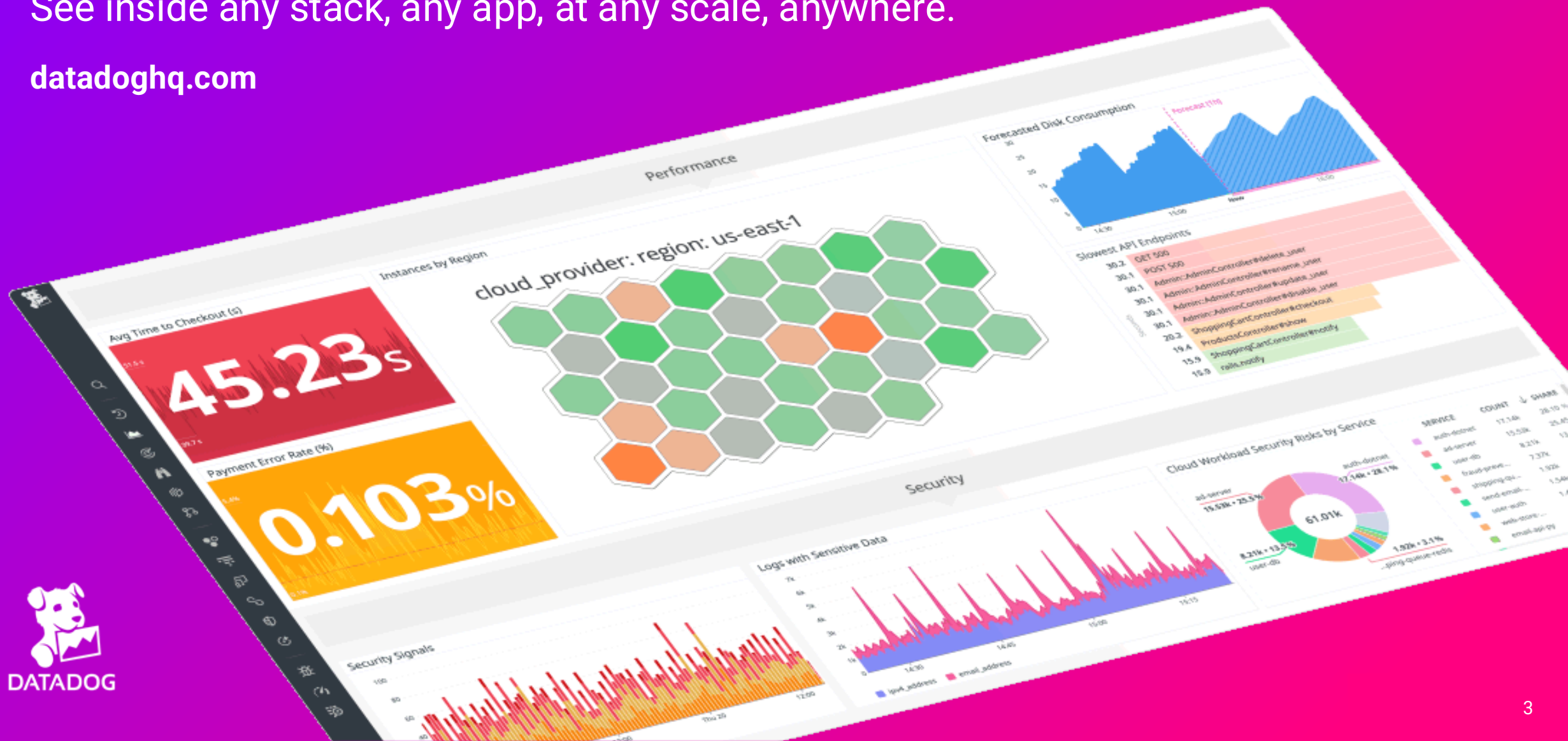
04 Live Demo

05 Q&A

Modern monitoring & security

See inside any stack, any app, at any scale, anywhere.

datadoghq.com



One platform across your stack

Data Observability builds on the same platform



Infrastructure

Hosts, containers, clusters and cloud services in real time.



Applications

Traces that follow every request across your services.



Logs

Search and analyze logs alongside every other signal.



Security

Threats and misconfigurations, in the same context.

Datadog today

33,200

customers

30%

of revenue invested in R&D
on a non-GAAP basis*

*Represents Mar-26 trailing-twelve-month non-GAAP R&D expense as a percent of revenue. Mar-26 trailing-twelve-month GAAP R&D expense was \$1,643M and GAAP R&D expense as a percent of revenue was 45%. Mar-26 trailing-twelve-month non-GAAP R&D expense may be reconciled to GAAP R&D expense, by deducting stock-based compensation expense of \$487M and employer payroll taxes on employee stock transactions of \$42M from GAAP R&D expense.



Uber Freight



SONOS



zendesk



FICO

coupang



CONDÉ NAST



Poll

What's your biggest concern with maintaining reliable data?

- a) Increasing costs
- b) Finding bad data once it's too late
- c) Taking too long to find root causes
- d) Mapping complex pipelines and dependencies
- e) Other

AI is changing the demands on data

SQL, Dashboards & Reporting



Agents & LLMs

AI is running directly on your data

89%

of organizations deploy AI
directly on their data
warehouse

#1 Blocker = Data Quality

How bad data shows up

Trust erodes before you can fix it



A broken dashboard

Numbers go stale or wrong, and decisions get made on them anyway.



A customer complaint

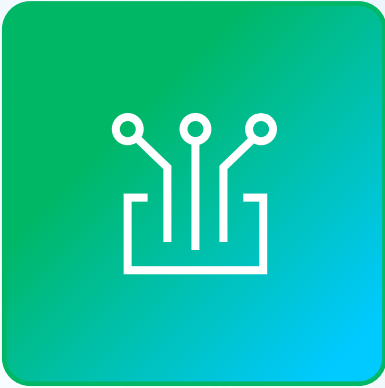
Customers notice the problem before your team does.



An AI model's wrong answers

Models quietly produce wrong answers, with full confidence.

Challenges with maintaining reliable data pipelines



Data teams are the last to know when something breaks



Existing tools make it hard to understand impact or find root causes



Teams forced into trade offs between cost & performance

POLL

How does your team usually find out about a data issue?

- a) A stakeholder reports it
- b) Manual spot checks
- c) Tests in the pipeline
- d) Automated monitoring and alerts

Datadog Data Observability

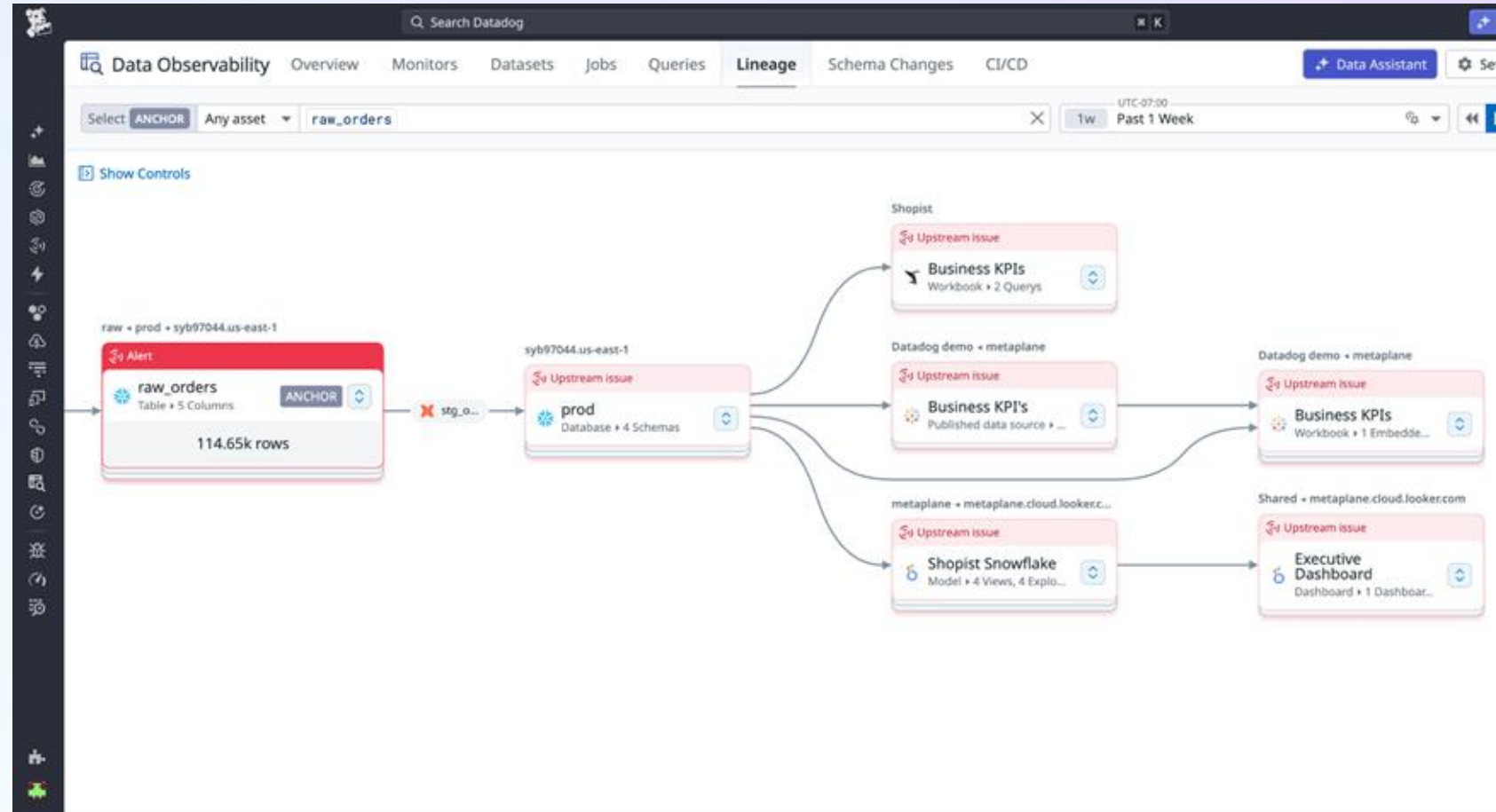
Detect, resolve and optimize, all in one place

Data Observability

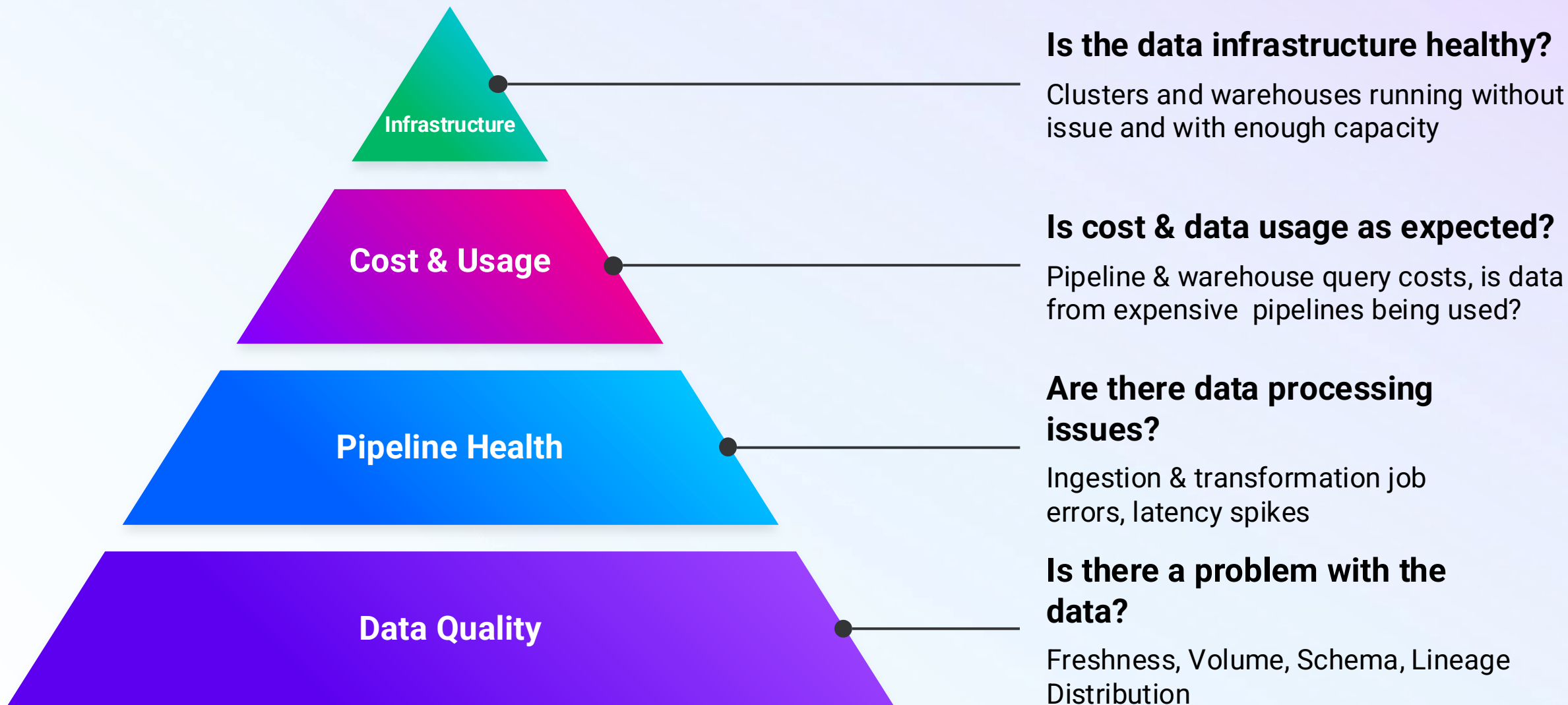
Detect Data Quality and Pipeline Issues Early

Pinpoint Root Causes With End-to-End Lineage

Control Costs With Optimization Recommendations



Our Vision of Complete Data Observability



Our Vision of Complete Data Observability



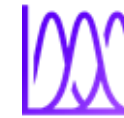
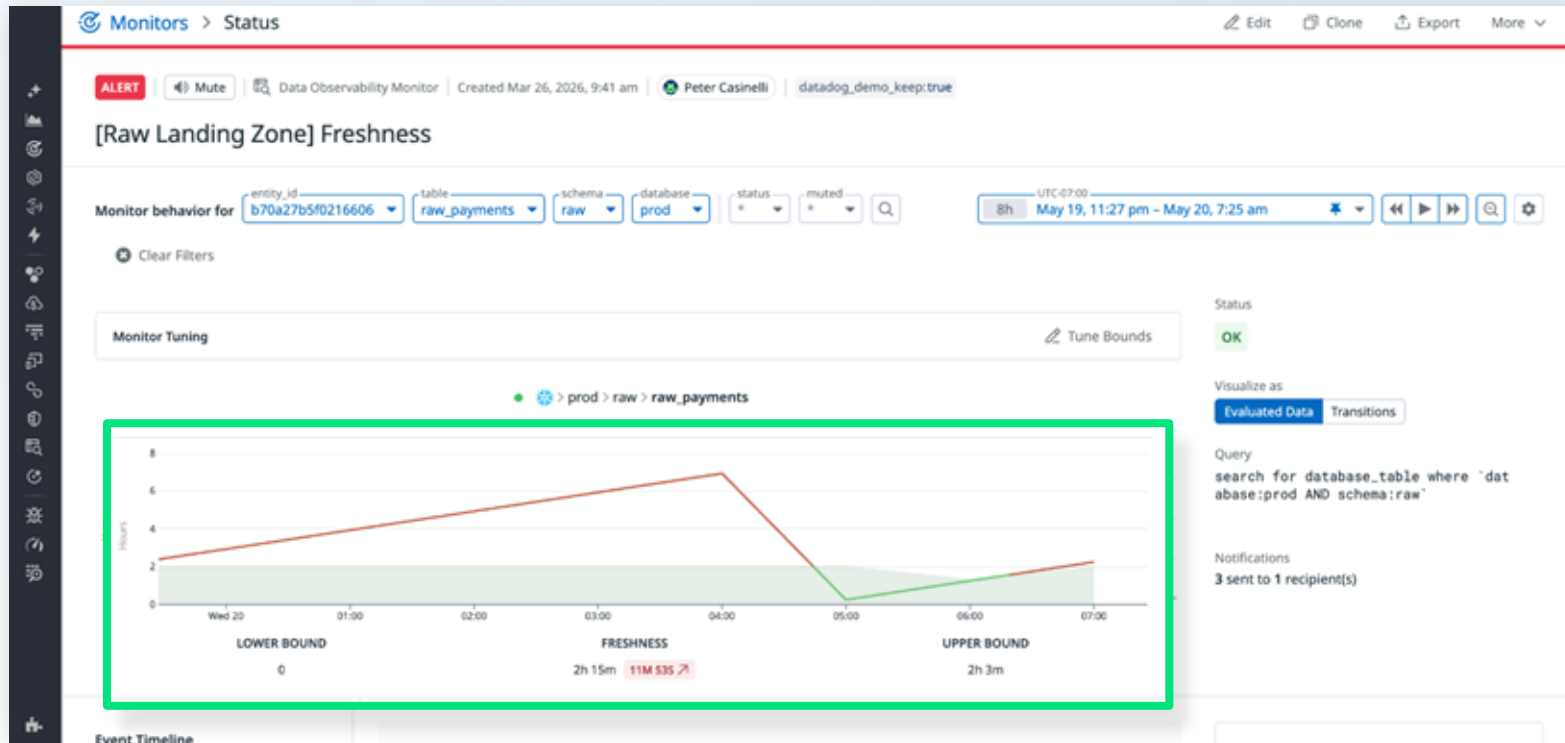
Data Quality

Is there a problem with the data?

Freshness, Volume, Schema, Lineage
Distribution

Is there a problem with the data?

Anomaly detection that learns from your data – flagging issues before stakeholders do



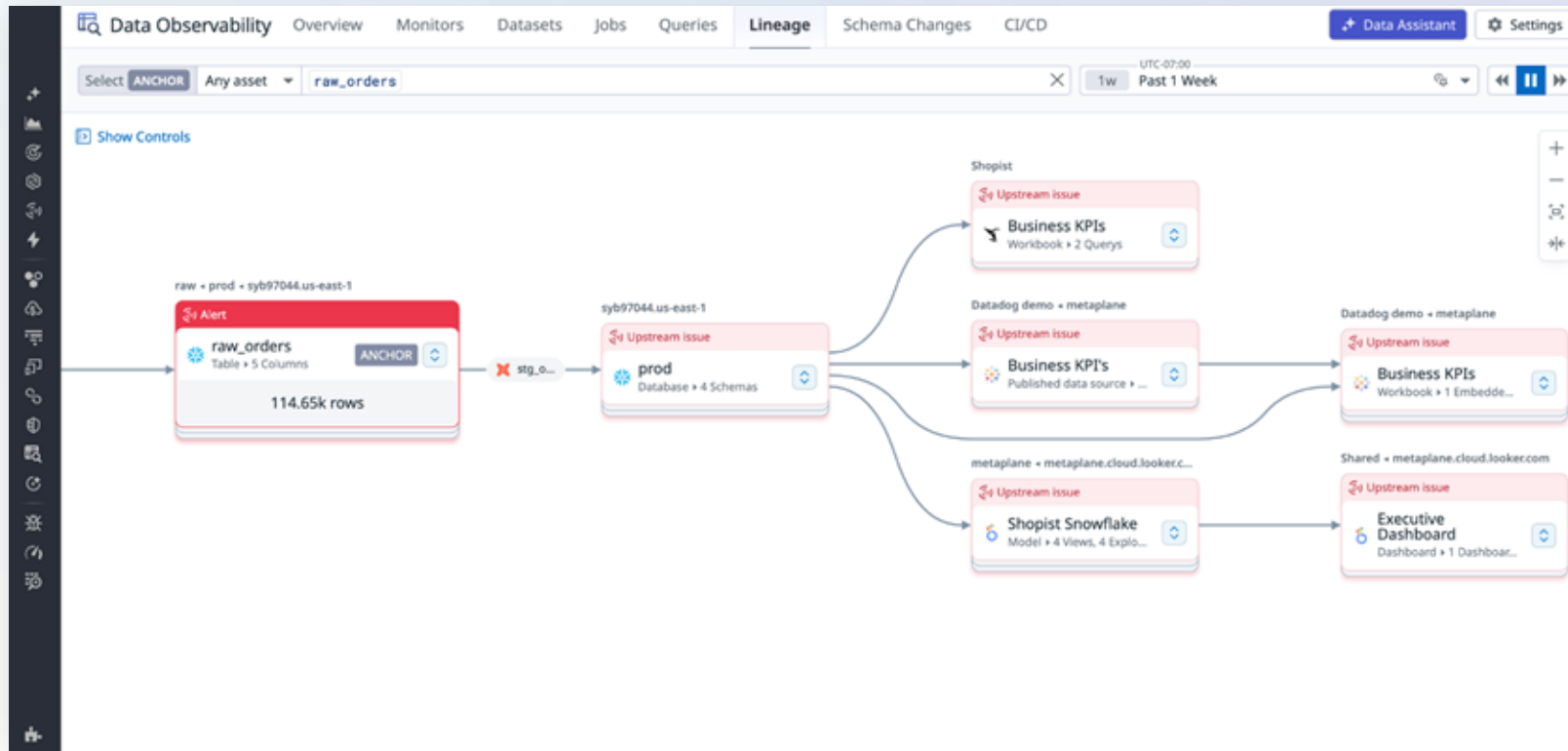
ML-powered anomaly detection

Learns from your data and surfaces silent failures early

- Freshness
- Volume
- Schema
- Distribution

Is there a problem with the data?

End-to-end lineage to assess blast radius and pinpoint root causes

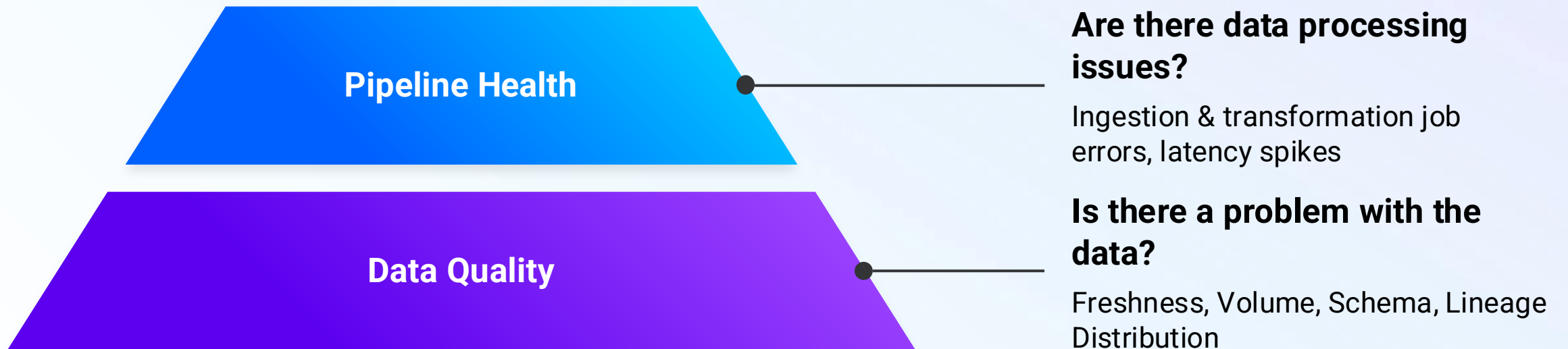


End-to-end lineage

Pinpoint root cause and blast radius in minutes

- Trace an issue upstream to its source
- See every downstream table and dashboard affected

Components of Data Observability



Are there data processing issues?

01

Extract

Sources arrive late or partial. Seeing it here stops gaps before they spread.

02

Transform

Logic breaks or slows. Seeing it here ties errors to the exact step.

03

Load

Jobs fail or stall. Seeing it here flags gaps before tables go stale.

04

Deliver

Dashboards and models use the output. Seeing it here shows who is affected.

Are there data processing issues?

Catch ingestion and transformation failures the moment they happen

The screenshot displays a Data Observability interface with a 'Lineage' tab selected. It shows a task named 'shopist_raw_hourly' (Dag: 1 Task) with environment 'env: iceberg-test-v3-mwaa'. The task is linked to a table 'raw_orders' (Table: 5 Columns) containing 123.77k rows. A red box indicates 'Last run failed'.

On the right, a detailed view of the task 'Shopist Raw Hourly' is shown for trace_id 5376788162787839752. The 'Flame Graph' view shows a duration of 26m 19s. A legend indicates 3 errors. The error details for 'insert_raw_support_tickets' show a 'ZeroDivisionError: division by zero' at line 87 in 'transform_metrics'.

Trace: Tasks **Flame Graph** Waterfall Span List 23

Filter spans by any attribute ☐ Errors 3 Color by Service Hide Legend

shopist_raw_hourly 26m 19s

Shopist Raw Hou... airflow... insert_raw_supp... 977ms < 0.1% total exec time

Span: Overview **Errors 3** Logs 14

Shopist Raw Hourly > insert_raw_support_tickets

Pretty Raw Investigate With Bits AI

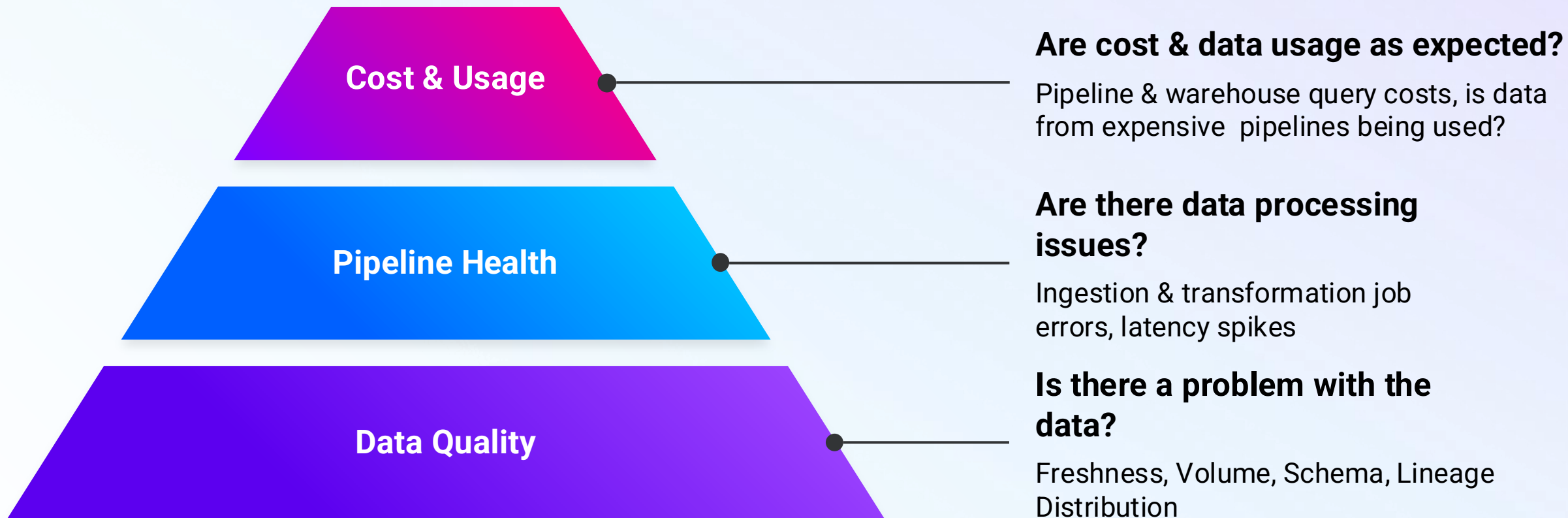
ZeroDivisionError: division by zero

Traceback (most recent call last):

Show 5 third-party frames

File "dags/shopist_raw_hourly.py", line 87, in transform_metrics

Components of Data Observability



Is cost and usage as expected?

Data costs grow quietly, one job and one query at a time



Know what drives spend

Which jobs, queries and warehouses account for the bill, and how that changes over time.



Spot inefficiency

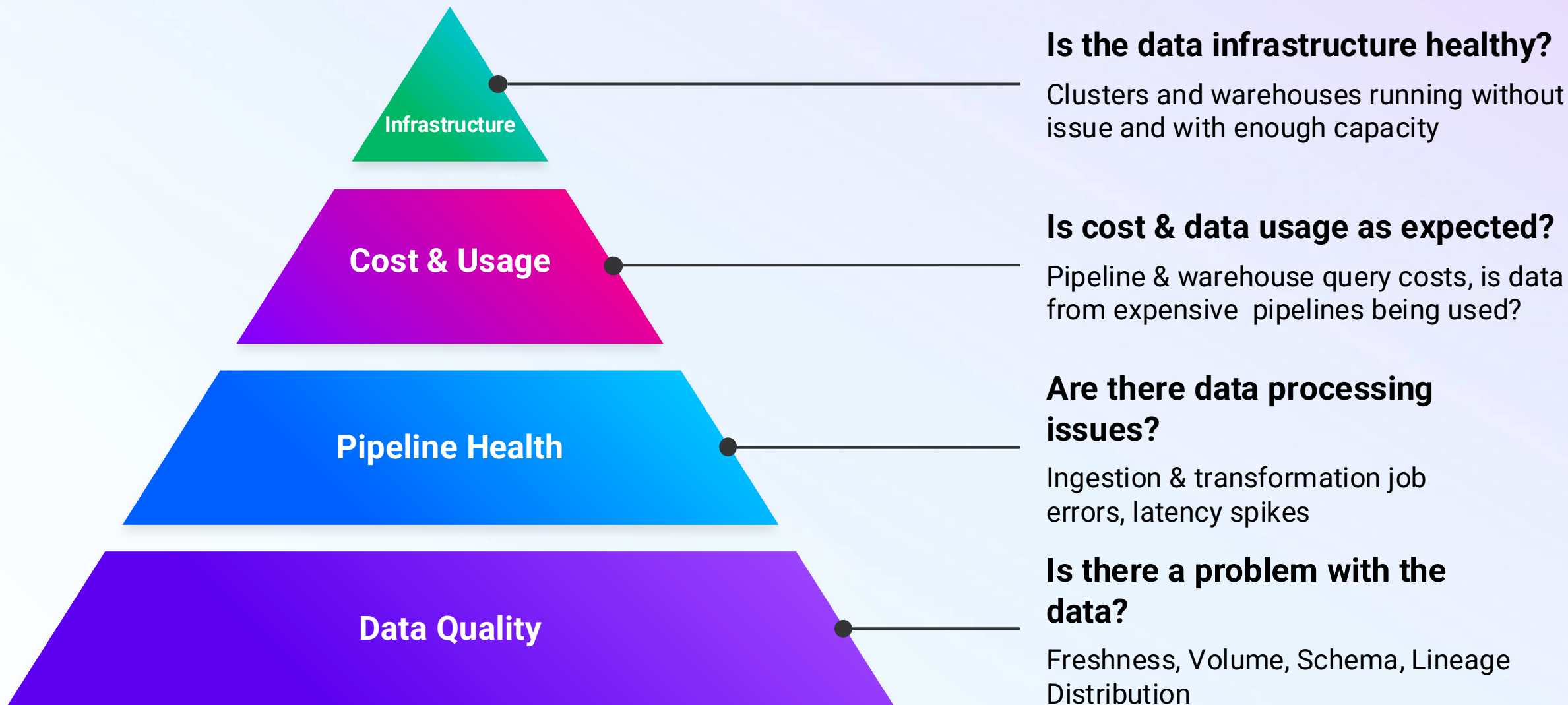
Runaway queries, oversized compute and slow jobs that cost more than they should.



Match cost to value

Whether data from expensive pipelines is actually used by anyone downstream.

Components of Data Observability



Is the data infrastructure healthy?

Every pipeline depends on the compute and storage underneath it



Availability

Whether clusters and warehouses are up and responding when jobs need them.



Capacity

Whether there is enough headroom for peak workloads, without paying for idle resources.

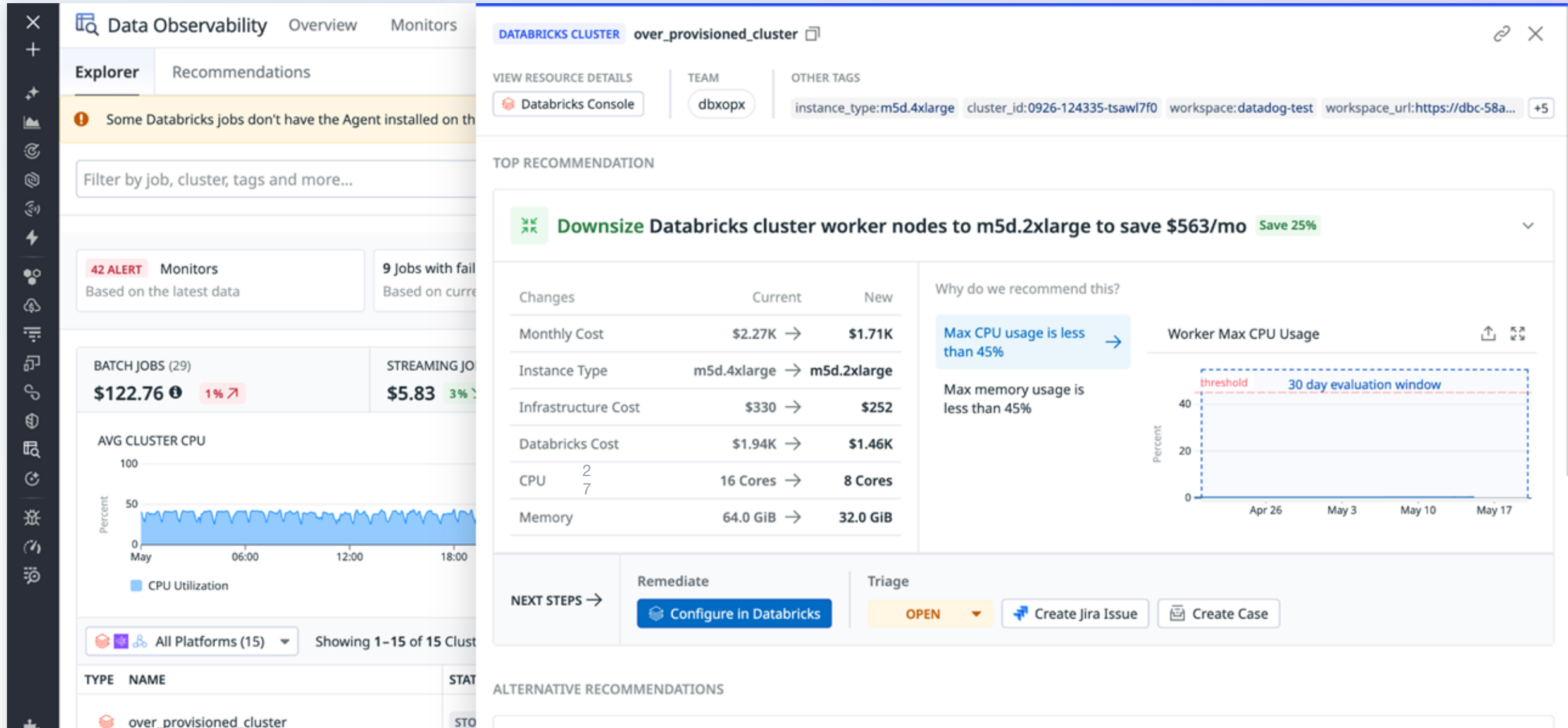


Contention

Whether workloads are competing for the same resources and slowing each other down.

Optimize infrastructure with actionable recommendations

Optimize infrastructure with actionable recommendations



POLL

Which area would help your team most, first?

- a) Data quality
- b) Pipeline health
- c) Cost and usage
- d) Infrastructure

Live demo

Data Observability in action

Key takeaways



AI raises the stakes

AI is changing what's asked of your data.



Tools fall short

Existing tool sets make reliable pipelines hard to maintain.



One place to fix it

Detect, resolve and optimize data issues before they cascade downstream.

**Catch bad data before it
reaches your AI models,
dashboards and customers.**

Get started today



Book a demo for Data Observability

