

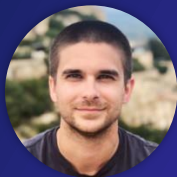


HOW TO ACHIEVE

Proactive Data Observability for your Lakehouse



Josh Benamram
Co-Founder and CEO



Ryan Yackel
Chief Product Evangelist



#standwithukraine

What we'll cover today



5 Steps for
Proactive Observability

10 min



Platform Demo

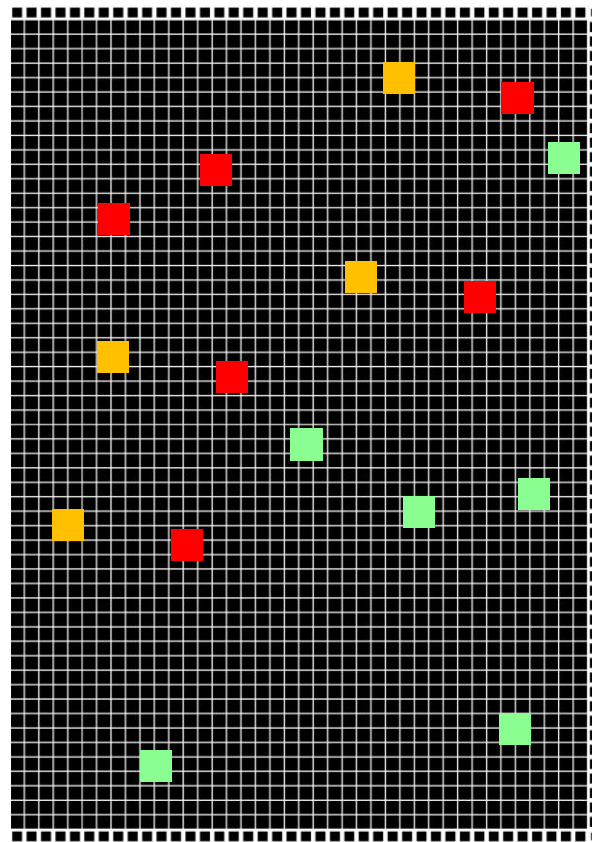
15 min



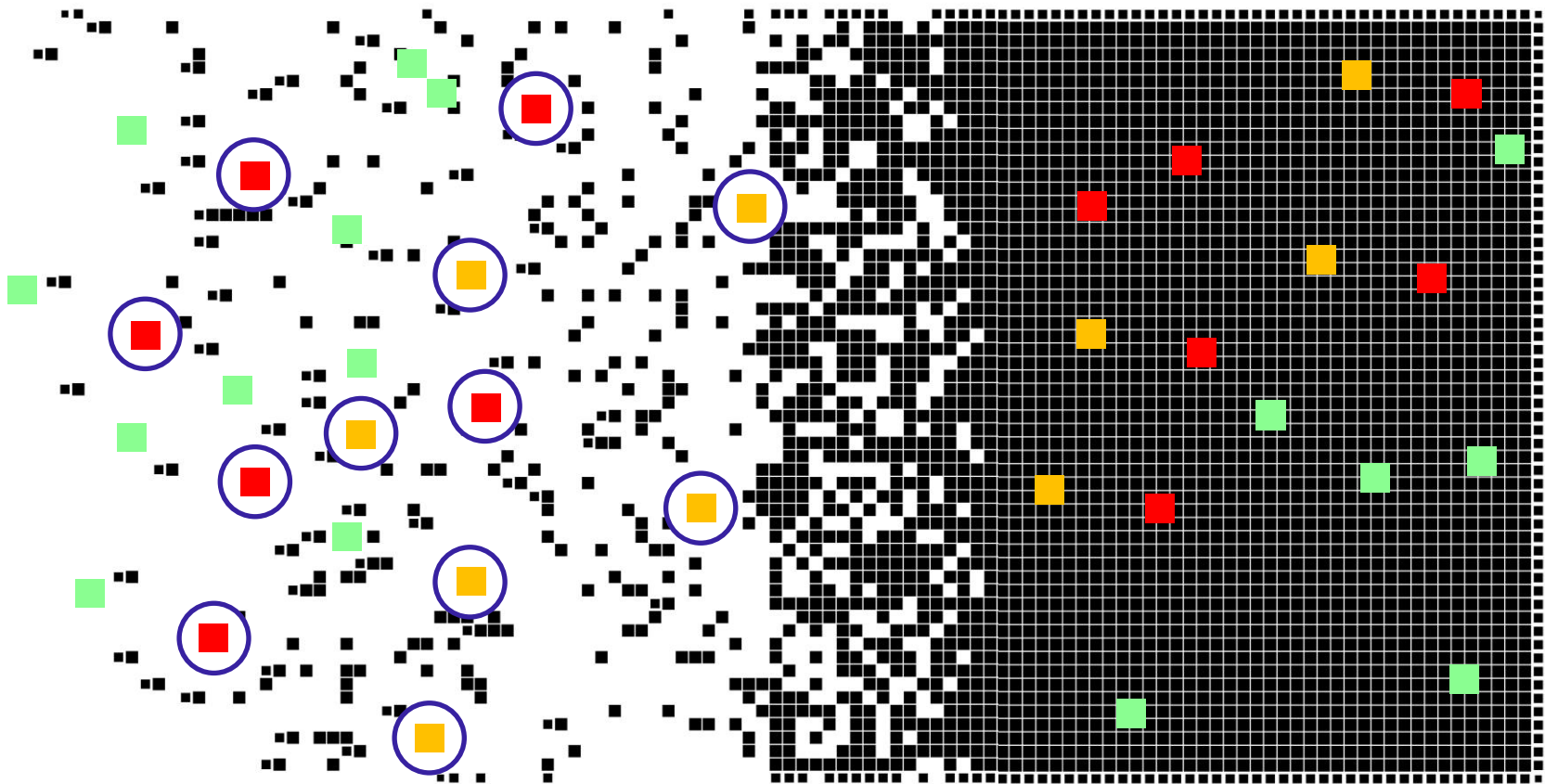
Q&A Discussion

5 min

Lakehouse



Lakehouse



Data

Lakehouse



Data Engineer

“

90% of our problems come from INPUT data problems.
This propagates to the system and blows up further down the line.

CTO, Investment Advisory

“

We not only need to monitor pipelines, but we also need to get
ahead of data issues, so our customers won't experience bad data.

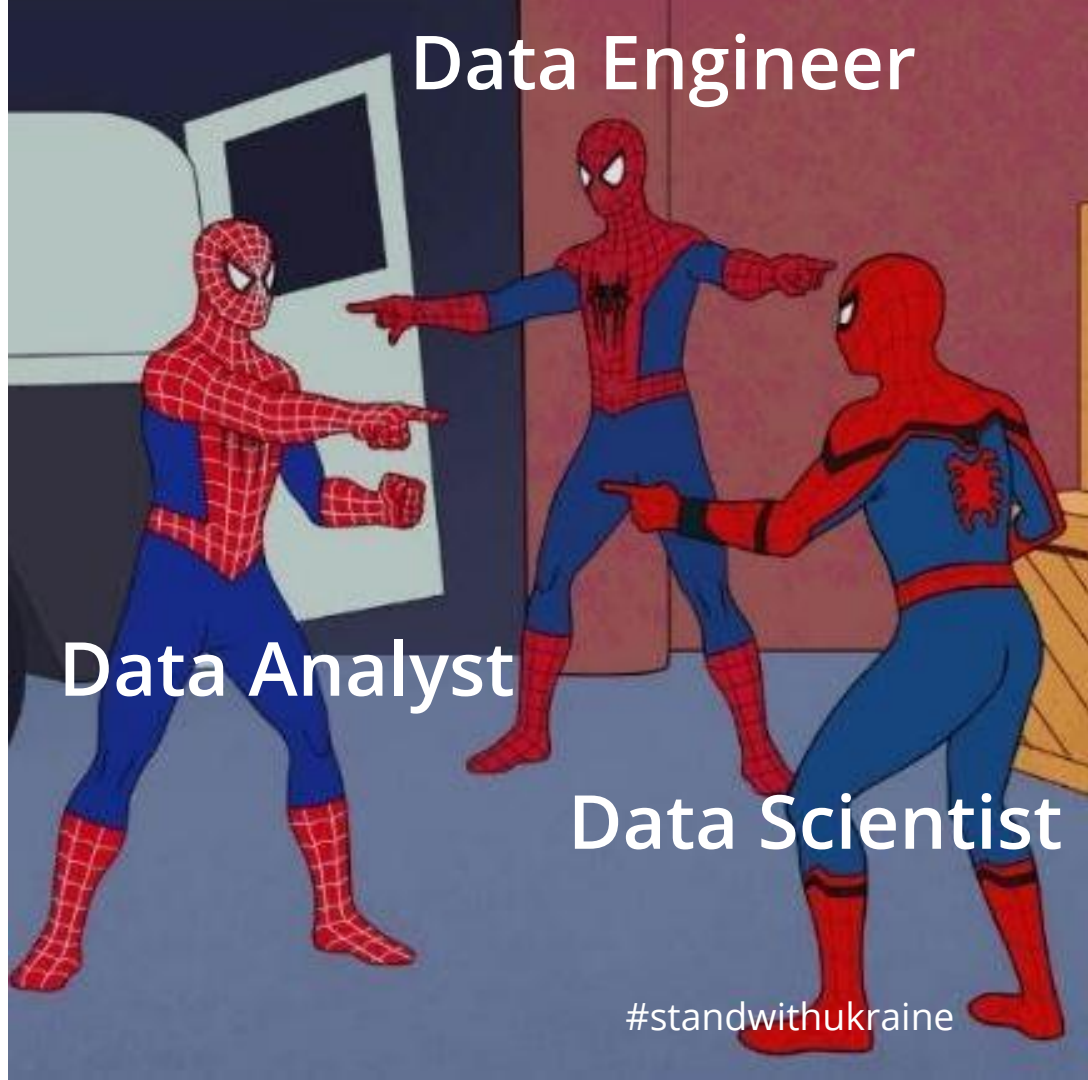
VP of Data Engineering, Online Sports Betting

“

We spend a lot of time fixing operational issues due to fragmentation of our data stack.
Tracking data quality, lineage, and schema changes becomes a nightmare.

Software Data Engineer, Cryptocurrency Exchange

“OK, why's the
data wrong...”

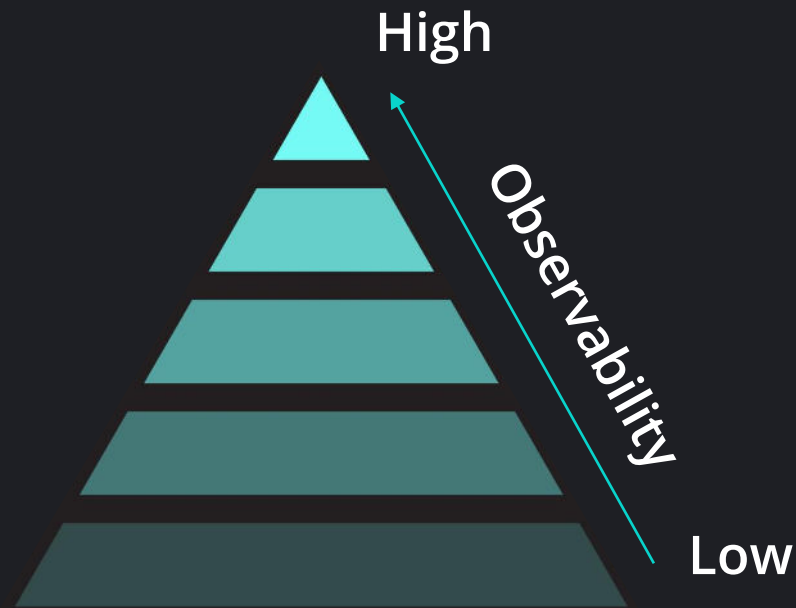


Data Engineer

Data Analyst

Data Scientist

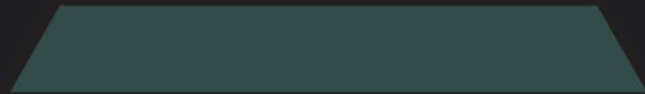
5 Steps to Proactive Data Observability





#standwithukraine

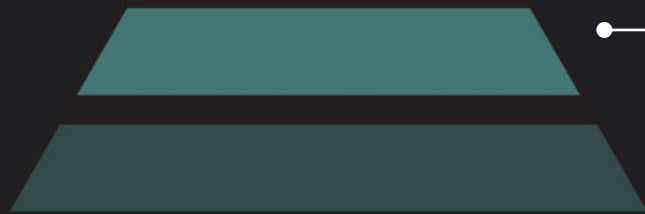
Step 1: Understand Pipeline Health



Pipeline Execution

Is data flowing through my pipelines?

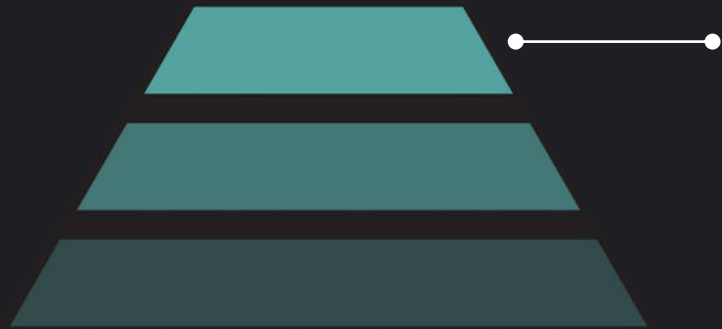
Step 2: Alert on Pipeline Latency



Pipeline Latency

Is data arriving in a usable window of time?

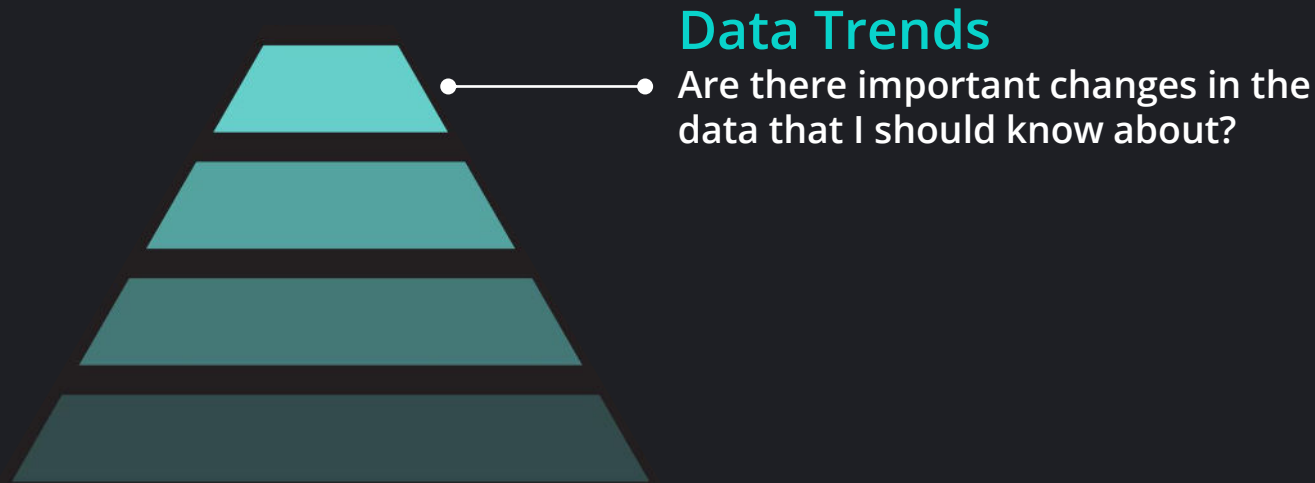
Step 3: Check Data Sanity



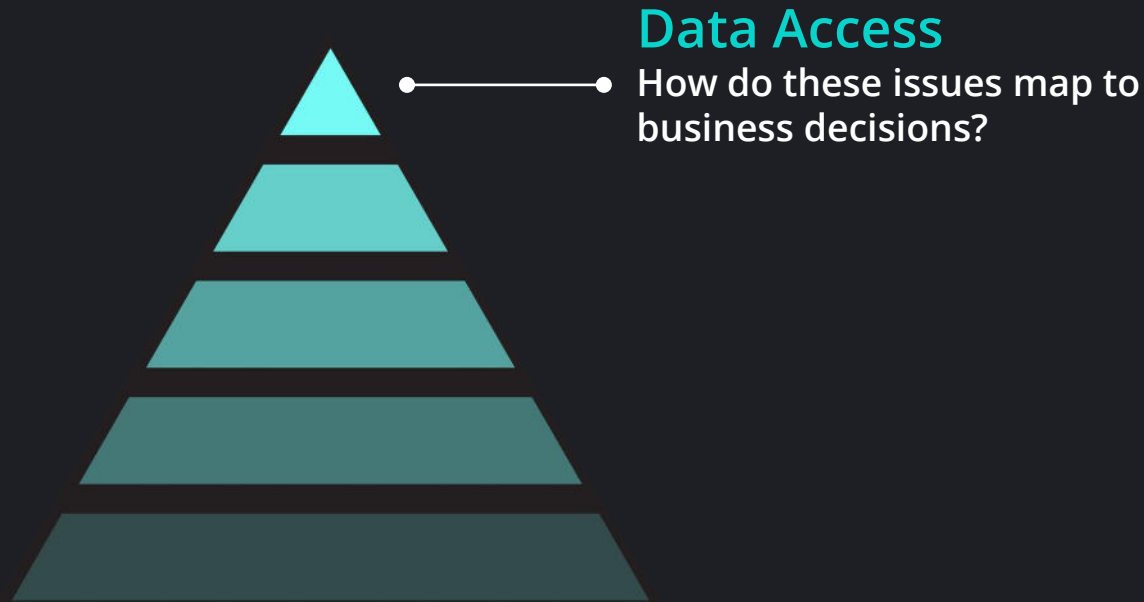
Data Sanity

Is the data coming through valid and complete? Are there errors in the data itself?

Step 4: Analyze Data Trends



Step 5: Identify Business Impacts



5 Steps to Proactive Data Observability

Data Access

How do these issues map to business decisions?

Data Sanity

Is the data coming through valid and complete? Are there errors in the data itself?

Pipeline Execution

Is data flowing through my pipelines?



Data Trends

Are there important changes in the data that I should know about?

Pipeline Latency

Is data arriving in a usable window of time?

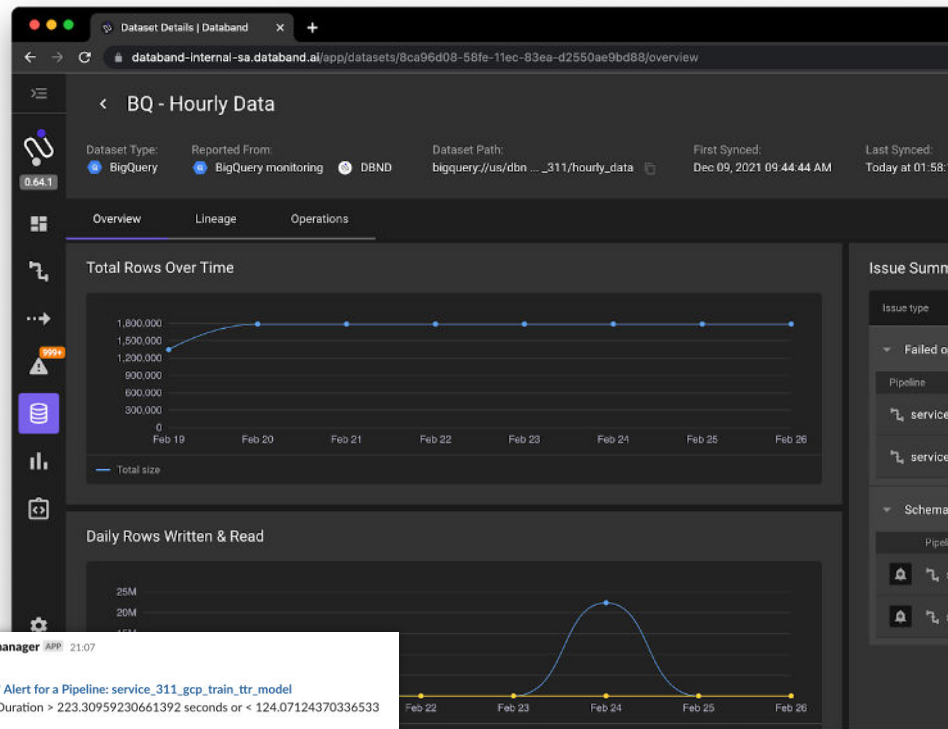
Live Demo

Proactive Data Observability



Josh Benamram

Co-Founder and CEO



DBND Alertmanager APP 21:07

ALERT!!!

***HIGH* Alert for a Pipeline: service_311_gcp_train_ttr_model**

Summary: Duration > 223.30959230661392 seconds or < 124.07124370336533 seconds

- Alert Trigger Time: UTC 2022-02-27 02:06:57
- Run Start Time: UTC 2022-02-27 02:02:22
- Run Name: trig_2022-02-26 02:00:00
- Triggered by: databand
- Airflow Tags: python, bigquery, gcp, demo, project: Service 311
- Affected Datasets:

All Alerts

Show less



Josh Benamram • 1st

Co-Founder & CEO at Databand | Forbes 30 Under 30

5d • 🌐

Databand.ai stands with our fellow Ukrainian teammates and all the people of Ukraine.

As our Ukrainian team members navigate this challenging time, Databand is offering help in every way possible and aiming to rally support.

Our team members have contributed to various organizations, and you can help too! Links below to the Ukraine Humanitarian Fund on GoFundMe and The National Bank of Ukraine's humanitarian fund. We encourage all our connections to give what they can to help those impacted by this conflict.

[#standwithukraine](#)

<https://lnkd.in/daXun-iv> ✓

<https://lnkd.in/dBdEXH4y> ✓



<https://lnkd.in/daXun-iv>

<https://lnkd.in/dBdEXH4y>

[#standwithukraine](#)