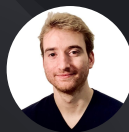
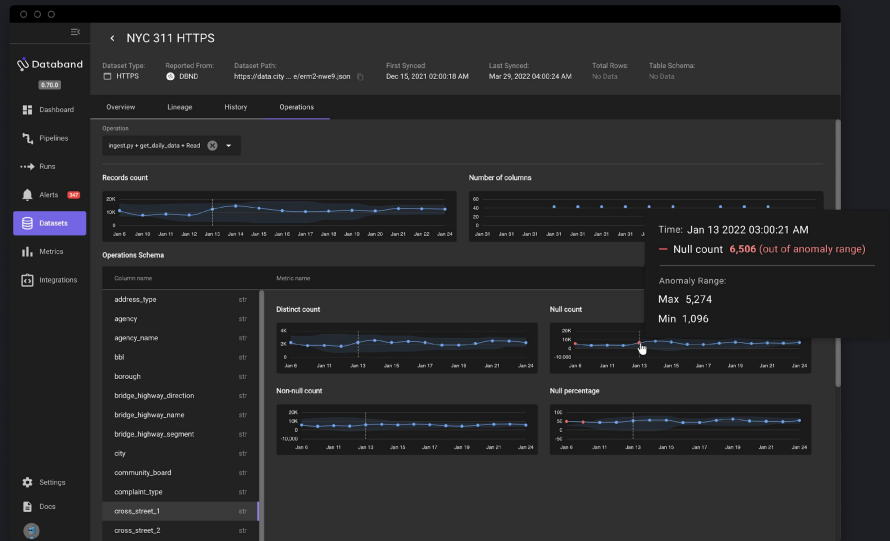




Data Pipeline Reliability



—
Josh Benamram
CEO

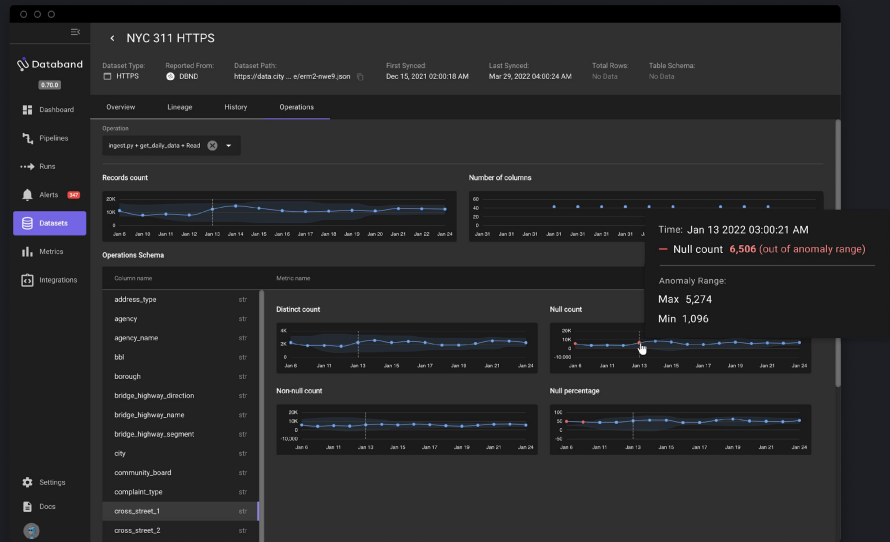


//

Data pipeline reliability means that data is delivered **complete, accurate, and timely**. It is the foundation for building data quality and trust across the analytics and data science organization.



Data Pipeline Reliability



Process Quality

- **Job Performance**
Are queries and jobs running efficiently?
- **Pipeline Latency**
Are pipelines running on time?
- **Pipeline Execution**
Is data flowing?

Data Quality

- **Data Content**
Are there significant changes or issues within the data?
- **Data Structure**
Is the data shape valid and complete?
- **Data Freshness**
Is data arriving on time?

Our solution

Data Monitoring & Incident Response

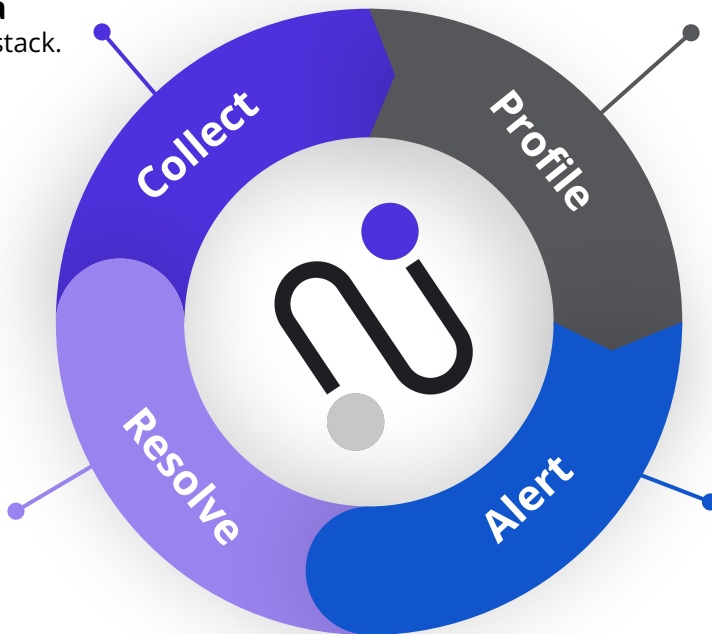
1. Automatically collect metadata

From all key solutions in the modern data stack.



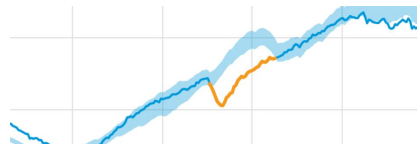
4. Resolve through automation

Create smart workflows to remediate data quality issues and keep SLAs on track




2. Build historical baseline

Based on common data pipeline behavior.

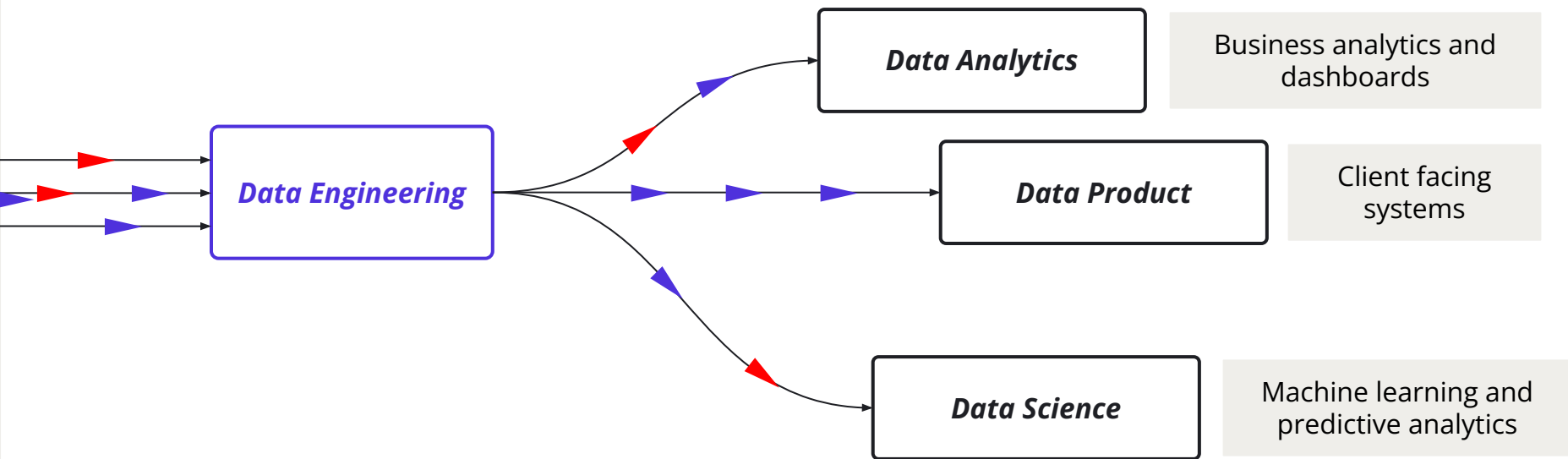


3. Alert on anomalies and rules

Based on deviations or breaches.

-  ***HIGH* Alert for a Pipeline: service_311_get_data**
- **Summary:** Duration > 76.39938569068909 seconds or < 6:
 - **Alert Trigger Time:** Feb 02, 2022 8:11:41AM
 - **Run Start Time:** Feb 02, 2022 8:10:06AM
 - **Triggered by:** Databand
 - **Affected Datasets:**
 - Writes into dataset BQ - Hourly Data

Data reliability starts at **data platform & engineering**



**Process Quality
Control**



**Data Quality
Control**



Data Platform SLA



Demo