Databand

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.
Databand

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.

2. **Build historical baseline**
   Based on common data pipeline behavior.

3. **Alert on anomalies and rules**
   Based on deviations or breaches.

4. **Resolve through automation**
   Create smart workflows to remediate data quality issues and keep SLAs on track
Data reliability starts at **data platform & engineering**

- **Data Engineering**
- **Data Analytics**
  - Business analytics and dashboards
- **Data Product**
  - Client facing systems
- **Data Science**
  - Machine learning and predictive analytics
Process Quality Control + Data Quality Control = Data Platform SLA