Databand
Streamline your DataOps
# About Us

## Company

Founded 2018, 20 Engineers

## Team

Data and ML Enthusiasts
Data Engineering, ML Engineering, Data Science, Data Product

## Funding

Well Funded, Series A Stage, Leading VCs

## Stage

Daily production use in some of the world's most exciting data engineering teams

Fortune50 enterprises to 50 person startups
Contributions

- Functional DAGs (AIP31)
- Scheduler Optimizations
- In-Process Executor
Problem Statement:
In data engineering teams today, it’s hard to know that pipelines are reliable.
Data Engineering Stages

**Identify Need**
- Deciding that data is worth it, realizing that analysts and scientists need help, assigning the appropriate resources.

**Build**
- Investing in tools that get data from point A to point B, in the structure that's needed for analysis or ML.

**Make Sure it Works**
- Investing in the Ops and CI/CD practices required to ensure reliability and safe iteration.

(Most Airflow companies here, part of why Airflow was selected)
DataOps Observability Pyramid of Needs

1. Data Access
   - How do issues below, combined with org policies, map to problems in how people are actually working with the data?

2. Data Trends
   - Are there important changes in the data that I should know about?

3. Data Sanity
   - Is the data coming through valid and complete? Are there errors in the data itself?

4. Pipeline Latency
   - Is data arriving in a usable window of time?

5. Pipeline Execution
   - Is data flowing through my pipelines?
What makes it tough?
Pipelines Are More Complex

Before

- SQL
- Informatica
- Tableau

Today

- Airflow
- Google Cloud Composer
- Kubernetes
- Jupyter
- Spark
- mflow
- Databricks
- Amazon EMR
- Python
- Java
- Scala
- SQL
- Snowflake
- Google BigQuery
- Amazon Redshift
- MySQL
- Presto
- Parquet
- TensorFlow
- Pandas
- Presto
Standards Are Not Yet Defined

DevOps

DataOps
Data Engineering Time is Precious

2x more job openings for data engineers than data scientists

50% of companies don’t have enough backend data engineers

180 hours a week spent on pipeline troubleshooting at avg company

Stats from PulseQ&A research firm, 2018 dataops survey
Salary figures from Indeed queries
Databand helps data engineering teams ensure **reliable delivery** of **quality data**.

By providing deep monitoring on the health of data pipelines (observability).

Collect, organize, and alert on pipeline metadata
How it Works

Orchestrator

Task Engine

Data

DAG Status API

Task Internals API

Data Quality API

Databand Server
Demo
Instant View on Data Pipeline Health

Product Screenshots

- Run Tracking
- Error/Log Collection
- Data Profiling
- Alerting
- Metric Reports
- Data Profile Comparisons
Contact us to learn more!
contact@databand.ai

Always open to discussion, ideas, feedback