

A DataOps and automation platform

Data team challenges

A data team typically faces two main challenges in their effort to generate data for business needs:

- ◆ Orchestrating and automating the data loading and data transformation pipeline.
- ◆ Reliability challenges.

Root cause

These challenges arise partly because the automation platform and the underlying abstraction are not offering adequate support.

E.g., the DAG-of-jobs abstraction seen with Airflow, etc. is a workflow automation abstraction and is not directly aware of the data processing nature of data pipelines. Consequently, more work is left to developers when building data pipelines.

Due to limited platform support, reliability is an engineering effort rather than a by-design property of the pipeline. Every corner case you do not cover can cause reliability problems.

Our approach

We start by replacing the traditional DAG-of-tasks abstraction with a new one: "Catalog-based dependency". We have a formal and simple process to convert requirements into a design that fits this approach. Our platform Trel, fully supports implementing such designs.

Benefits

Under our guidelines, your pipelines are ELT, and they gain reliability by following the design guidelines. Furthermore, the implementation is highly streamlined, with lightweight jobs that rely on the platform for orchestration, input selection, resource allocation, dependencies, etc.

At our platform's center is a powerful data catalog with time-travel capabilities. We also have safe workflows for CI/CD, schema migration, failed job recovery, collaboration, etc.

Trel pipelines are easy to design, build and explain. It is suitable for multi-cloud data warehousing, big data as well as machine learning pipelines.

Contact Us

- 📍 Fremont, CA
- ✉ sales@cumulativedata.com
- ☎ +1 407 374 3875