TETRASCIENCE

DATA INTEGRATION PLATFORM

Pipelines for Data Processing

Raw data is generally not usable as-is, especially when combining data sets from multiple sources or instruments. Different formats. Different taxonomies and naming conventions. Nothing matches. And the raw data is often stored in multiple locations, making it inaccessible. Whether the goal is transforming the data or moving the data, we have the technology to automatically process raw data, saving time and creating value.

Automate experimental data processing

Flexible
tailor data flows based on your business logic

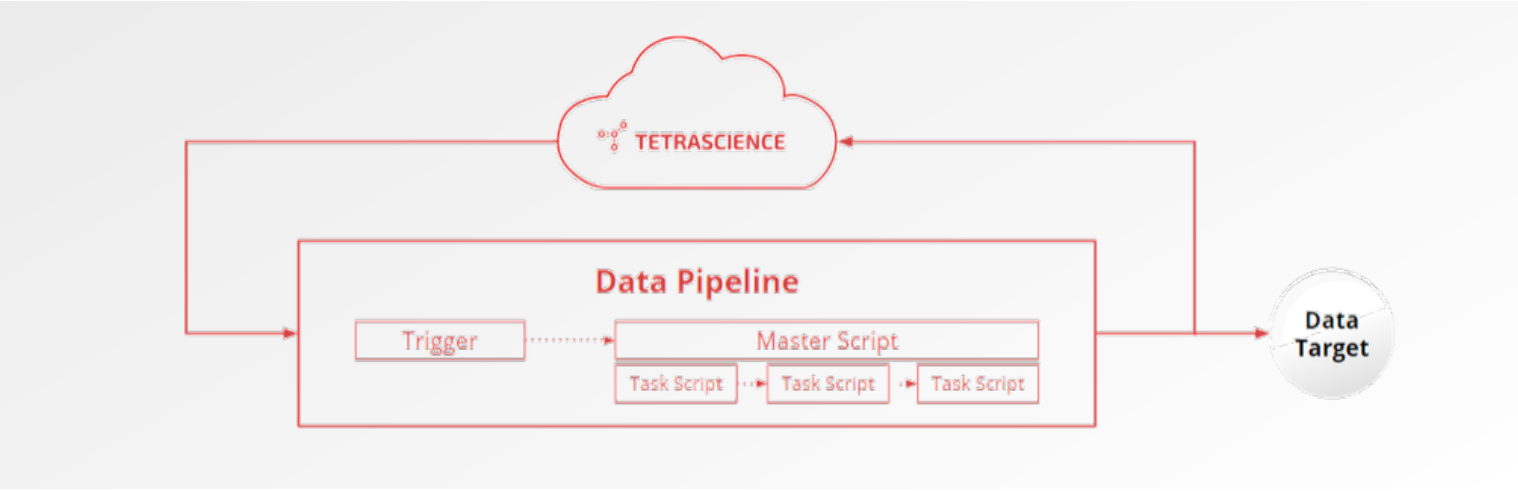
Productized
developer-friendly, using open source language and SDKs

Scalable
add any number of pipelines; intuitive to set up and monitor

Secure
uses industry and AWS/Docker best practices

DATA PIPELINES

A data pipeline is a powerful automation technology to process data. Processing generally refers to either transforming data in various ways (see next page for examples) or moving data from one location to another.



Triggers

Triggers control when and under what circumstance a data pipeline is initiated. Triggers are configured directly within the TetraScience User Interface. Sophisticated trigger logic enables custom business rules tailored to your needs. Continue to use your existing data pipelines with our SDK

Master/Task script

Written in Javascript / NodeJS, a general-purpose language, master script natively supports branching, loop, if-else, and other complex data flow logic. Run tasks in parallel and control the concurrency. Task Script natively supports Python and Javascript / NodeJS; configure your own Docker image to expand options

Data pipelines

Pipelines consist of multiple steps, each implemented using your preferred programming language. Pipeline status can be viewed directly within the TetraScience dashboard. If a pipeline fails, the dashboard will show a failed status and alert users

Data targets

Use pipelines to move selected data to a selected target or targets. Whether the destination is an ELN, LIMS, CRO/CDMO, back to the Data Lake after a calculation, or over to an automation for next instructions, pipelines orchestrate the flow of your data

TAILORING DATA PIPELINES TO YOUR BUSINESS

Here are examples of the most popular ways customers use data pipelines to transform and move their experimental data.

Data integrity

Perform quality checks and validation, improving data integrity. Notify the CRO and/or the scientists of issues so they can take corrective action

Data enrichment

Merge metadata from different sources, enriching the original data set. Keep your sensitive information within your control while sharing pertinent information, like with CROs

Data science + ML

Run calculations and execute data science scripts, generating analytical results on the fly. Update machine learning model weights using real-time data

Closed-loop control

Connect automation across the physical and digital worlds. For example, modify lab automation robot behavior based on instrument readouts, with human oversight as needed

Report generation

Periodically generate reports and deliver to designated users. For example, on Friday afternoons, compile the experiments completed during the week and send to the team lead

File conversion

Convert data files from one format to another. For example, Waters Empower raw data to an intermediate format. Or various formats into Allotrope Data Format (ADF) files

ADDITIONAL INFORMATION

Developer friendly

Develop and test your pipelines in your local development environment. Collaborate with your team using tools like Github or Gitlab, and iterate. Use your favorite CI/CD tool to test, build the artifact, and then upload to TetraScience. We will then verify, deploy your code / container, and orchestrate them in the cloud.

Continue to use your existing data pipelines by taking advantage of our SDK. Write the files and artifacts back into the TetraScience Data Lake and apply metadata and tags. This gives you a centralized place for your experimental data, whether raw data, quality control data, artifacts, or analysis results.

Best-in-breed tech stack

Our pipelines are built on AWS technologies (for horizontal scaling) and Docker (for vertical scaling). We believe this combination allows for maximum scaling and flexibility.

- AWS Lambda lets you run code without provisioning or managing servers and start a large number of executions in parallel
- Docker allows you to containerize task scripts to bypass memory and duration limits of AWS Lambda functions, use your own binary packages, and use the different languages you need to get the job done

TETRASCIENCE

TetraScience is the leader in transforming the Digital Lab. We provide an advanced data engineering solution that makes life sciences R&D data truly accessible and actionable. More than 80 leading pharmaceutical and biotech companies rely on our cloud-native Data Integration Platform and applications to automatically centralize and harmonize their experimental data, preparing it so you can focus on value-add activities like data science and AI, trend identification, and anomaly detection and response. Activate the flow of your data.

www.tetrascience.com
hello@tetrascience.com
@TetraScience

294 Washington St.
11th Floor
Boston, MA 02108