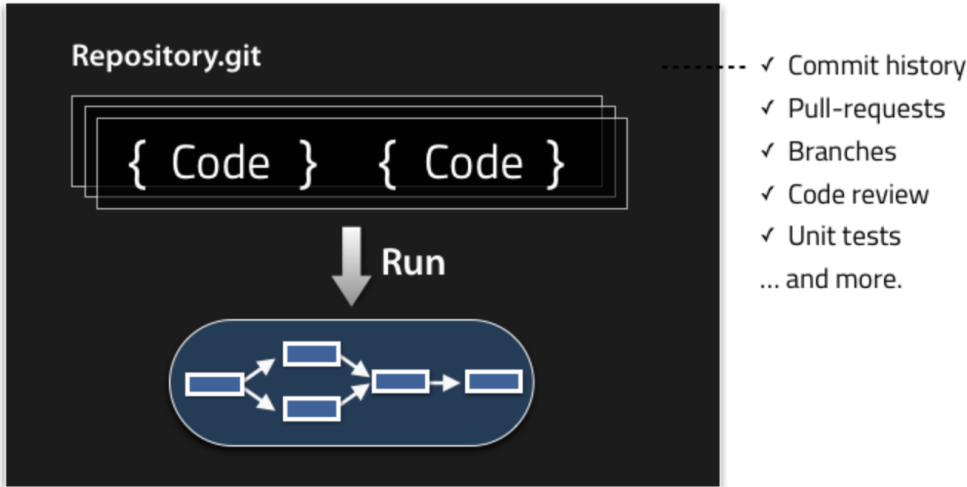


About Treasure Workflow

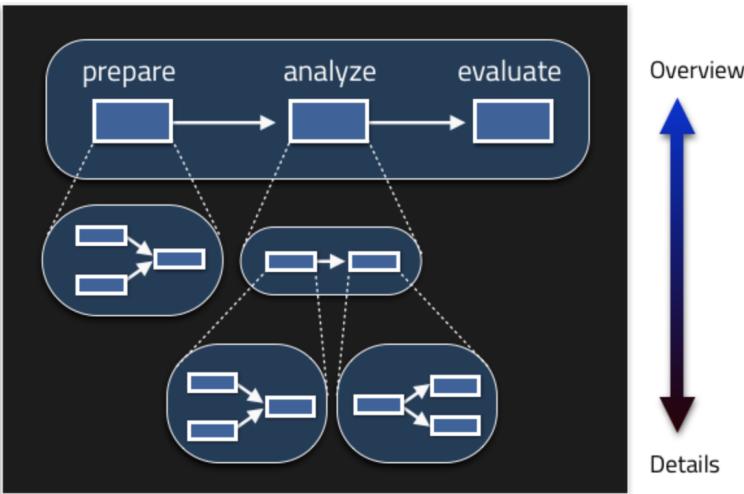
Treasure Workflow is a multi-cloud workflow engine that can orchestrate tasks not only on Treasure Data but across a variety of cloud infrastructures such as AWS and Google Cloud Platform. Treasure Workflow extends and enhances the capabilities of the highly reputable open-source workflow program, Digidag.

The development of data applications such as smart retargeting, A/B testing with customer goals, and omnichannel marketing can involve hundreds of steps, each of which might require complex transformations or dependencies. These challenges to data management and collaboration can quickly become cumbersome.

You can create workflows to run efficient queries against—for example—your customer data and schedule tasks that feed into audience identification, profiling, and tracking.



- ✓ Commit history
- ✓ Pull-requests
- ✓ Branches
- ✓ Code review
- ✓ Unit tests
- ... and more.



Integrate with and organize your organization's data, run SQL analysis across that data regardless of the scale, and then create repeatable insight by saving queries that disseminate data.

Features and Benefits

Treasure Workflow allows you to:

- Create a workflow, which defines the order in which processing tasks run
- Design with scheduled processing flows in mind
- Parameterize for easy cloning, sharing, and re-use
- Develop locally, push to Treasure Data to run on a scheduled basis
- Manage error handling more easily
- Configure tasks that can operate nearly every part of the TD system, including:
 - Importing data batch jobs using Integrations
 - Running Presto and Hive queries
 - Create or append to tables
 - Result export to other systems

With Treasure Data, you can improve your ability to create internal Data Applications and gain the following benefits:

Organizing your team's work

As your number of scheduled queries or CRON jobs increases, it becomes harder for organizations to keep track of "what is this one doing?". Being able to define tasks into organized workflows and projects allows you to immediately know the context that a given query is operating in.

Managing error handling and establishing automated notifications

Many times we see significantly large queries and scripts operating in our customer's systems. These can be 100s or even 1000s of lines long. When errors occur in these SQL queries, it can be incredibly difficult to debug. By breaking your large queries into workflows of smaller dependent queries, it becomes much easier to figure out which part of your logic has broken.

You can receive notifications when any part of your workflow fails, and thus fix it quickly. You can also specify to receive notification of successful workflow runs, or workflow runs that do not complete within specified time boundaries.

Reducing end-to-end latency

By ensuring that dependencies between steps are properly kept, you can create processing pipelines for live data use cases such as reducing KPI updates from daily to hourly to more frequently.

Improving Collaboration and Re-Usability

Parameterization is deeply embedded within Treasure Workflow, so that, as an analyst, you can create a reusable workflow template. You can use your template workflow for future, additional analysis. Stop re-creating SQL statements for similar requests, and start templating your work for easier re-use.

Also, by organizing your queries, it becomes much easier to onboard new employees into your organization or into an ongoing project. You can use Treasure Workflow to group tasks together. New collaborators can more quickly understand the general "why" of a query before digging into its specific logic.

Digdag vs Treasure Workflow

Treasure Workflow currently allows for most of the functionality that Digdag, the underlying open source project, allows. But, there are a few exceptions. Some Digdag operators and functionality are not yet enabled when you submit workflows to the Treasure Workflow cloud environment. The following options are not allowed because shared processing and local disk are used:

- **embulk>**: Use embulk> for running arbitrary embulk jobs. However, you can use `td_load>` to import bulk data into Treasure Data.
- **download_file**: Typically, you might use the `download_file` parameter with the `td>` and other operators for downloading files locally. Instead, you can use the normal Treasure Data result export functionality

Treasure Workflow and Profile Sets in Audience Studio

Workflow is used in [Audience Studio](#). You can [create workflows for your source data](#) in preparation for creating master segments. You can also use Treasure Workflow in [predictive scoring](#) to refine audiences and segments and when you send segmented data to other systems. Treasure Data generates workflows that you can view.