Gladier

Release 0.8.0

The Gladier Team

Jan 27, 2023

CONTENTS:

CHAPTER

ONE

GLADIER

Gladier (the "GLobus Architecture for Data-Intensive Experimental Research") is a Python toolkit for developing data collection, analysis, and publication pipelines ("flows") for experimental facilities. A flow might, for example:

- Retrieve data from an instrument, verify its quality, extract metadata, and publish data+metadata to a catalog, or:
- Collect data from a series of experiments, train a machine learning model, and deploy the model to an instrument.

In these and many other applications, Gladier makes it easy to specify what actions to perform, and where, and then to execute those actions reliably and securely.

Gladier builds on the powerful cloud-hosted Globus platform, including Globus automation services for reliable and scalable flow execution; Globus Auth for secure distributed operation; and services like Globus Transfer, funcX, and Globus Search to implement data transfer, compute, cataloging, and other actions.

You can read more about Gladier, including example applications, in a technical report; here we focus on how to install and use it, and provide pointers to sample code.

1.1 Installation

Gladier requires Python 3.6 and higher. For a modern version of python, see the official Python Installation Guide.

Gladier includes two main packages available through PyPi:

- Gladier – The core development tooling for building and running flows
- Gladier Tools – An optional set of plugable tools for common tasks

With pip installed, you can do the following:

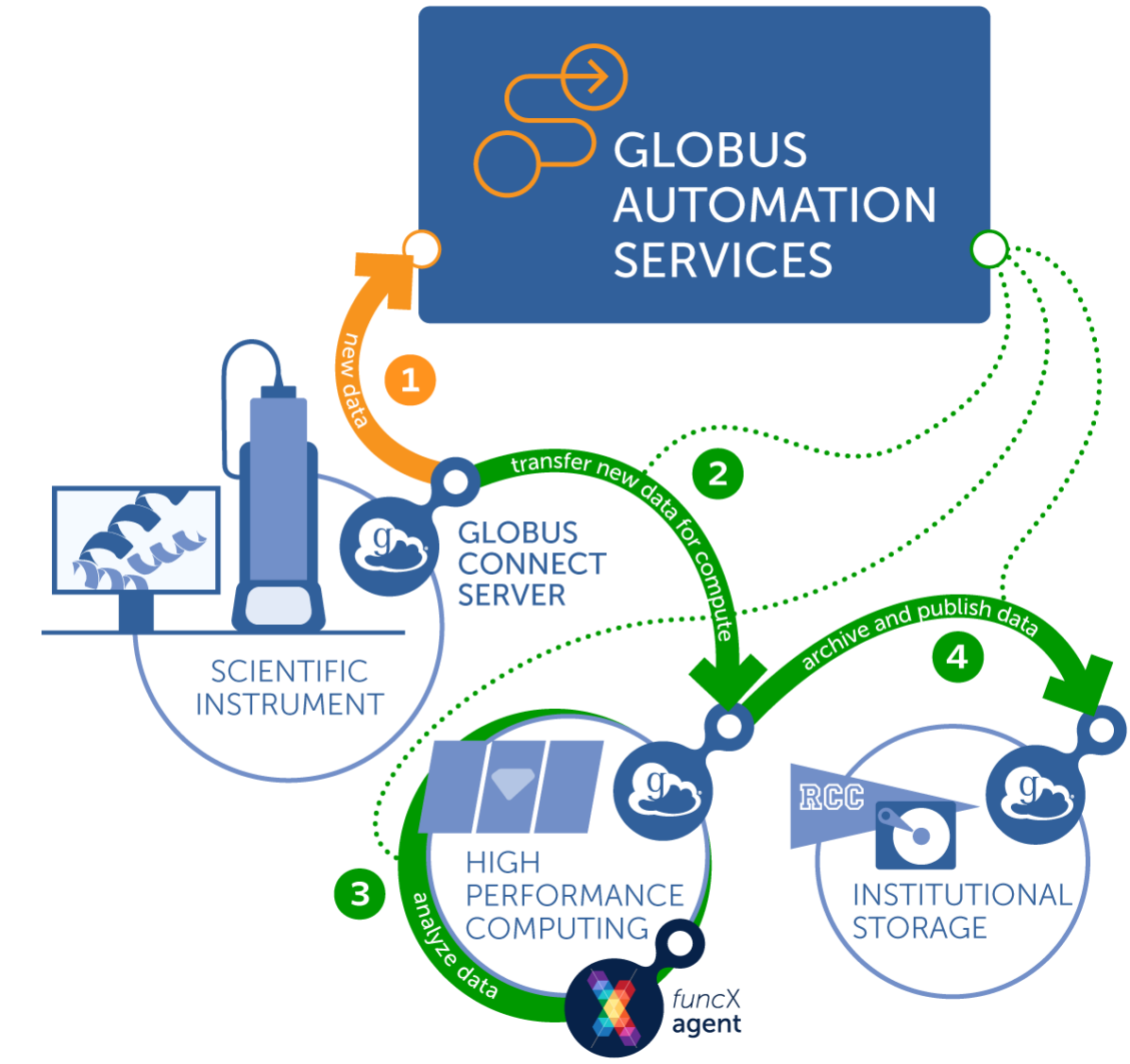
```
pip install gladier gladier-tools
```

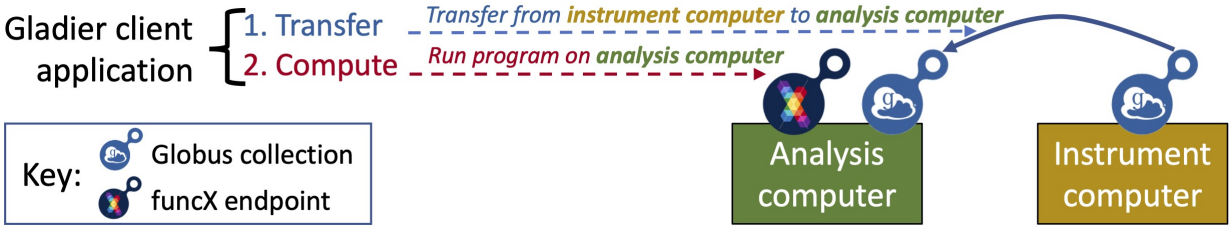
1.2 Setup

We first consider a simple Gladier Flow that transfers data from an instrument computer to an analysis computer, and then runs an analysis function on the analyasis computer. The flow thus comprises two steps:

- Transfer: Copy data to the analysis computer.
- Compute: Run the analysis function on the data copied in the first step.

The first step involves a Globus Transfer action, and the second a FuncX Compute action. In order for your computer to be accessible by Globus, it needs to be running a Globus collection. See the Globus Connect Personal section below. For the second part, you need to set up a FuncX endpoint to tell FuncX where it should execute functions. See the funcX Endpoint section below for details.


GLOBUS
AUTOMATION
SERVICES
new data
1
transfer new data for compute
2
GLOBUS
CONNECT
SERVER
SCIENTIFIC
INSTRUMENT
archive and publish data
4
HIGH
PERFORMANCE
COMPUTING
3
analyze data
funcX
agent
RCC
INSTITUTIONAL
STORAGE


Gladier client application
1. Transfer
2. Compute
Transfer from instrument computer to analysis computer
Run program on analysis computer
Key:
Globus collection
funcX endpoint
Analysis computer
Instrument computer

1.2.1 Globus Connect Personal

To run tools like `gladier_tools.globus.Transfer`, we need access to a Globus collection. Follow the Globus Connect Personal instructions to set up your own Globus Connect Personal endpoint and configure a collection, then navigate to Globus Web App collections to see details about the collections to which you have access.

Note: Your Globus endpoint should be accessible on the same machine you install your FuncX endpoint.

1.2.2 FuncX Endpoint

FuncX Endpoint We also need a funcX endpoint on which to run funcx functions. Follow the FuncX Endpoint instructions to set up your own endpoint. Once an endpoint has been configured, you can:

- run `funcx-endpoint` to list all endpoints to which you have access
- run `funcx-endpoint start <endpoint-name>` to start an endpoint

1.3 Running Flows

We provide a step-by-step guide to creating and running a simple two-step flow that:

1. runs a program on a directory (specifically, Tar, to create an archive file) and
2. transfers the transformed file to another computer.

Provided you have already performed the *Setup*, you can run this full example in the Tar and Transfer section in *Examples*. The following code, which implements the Gladier Client for this example, is a typical Gladier Flow definition.

```
@generate_flow_definition
class TarAndTransfer(GladierBaseClient):
    gladier_tools = [
        'gladier_tools.posix.Tar',
        'gladier_tools.globus.Transfer',
    ]
```

The new class has two main attributes:

- `gladier_tools` – Defines the tools that are to be used this client – in this case, the `gladier_tools.posix.Tar` and `gladier_tools.globus.Transfer` tools defined in Examples
- `@generate_flow_definition` – Splices together flows on each tool into one single runnable flow

Each Gladier tool sets its own FuncX Functions, required input, and separate Globus Flow. The Gladier Client gathers everything into one place. It handles registering functions, ensuring each tools input requirements are met, and deploying/updating the final combined Globus flow. After that, running the flow is as simple as these two lines:

```
tar_transfer_cli = TarAndTransfer()
flow = tar_transfer_cli.run_flow()
```

The first line instantiates the client with default values, which means that it will automatically trigger a Globus Auth login and register Flows and FuncX functions as needed.

The second line, `flow = hello_cli.run_flow()`, is responsible for the following things

- Gather default input from each tool

- Validate input for each tool
- Trigger an initial login for access to FuncX and the Flows Service
- Register funcX functions (and re-register them, if they were previously registered and have changed)
- Register the flow_definition (and update it, if it was previously registered, and has changed)
- Trigger a second login to authorize the deployed flow
- Start the flow

Once these steps have been performed, the flow is started and will perform the two steps one after the other: 1) tar a directory and 2) transfer the resulting archive. Gladier provides tools for tracking a flow's progress and output:

Once this checklist is complete, the flow is stared and will run through each flow-state in sequence. First, taring a directory then transferring the resulting archive. Nothing more is needed from Gladier, however there are extra tools for tracking a flows progress and output.

```
tar_transfer_cli.progress(flow['action_id'])
details = tar_transfer_cli.get_status(flow['action_id'])
pprint(details)
```

`hello_cli.progress()` will periodically query the status of the flow until it finishes. It's a nice way to watch progression as the flow executes. Once the flow has finished, `hello_cli.get_status()` will fetch output from the Globus Flow, so it can be displayed in a readable format.

1.3.1 Re-Running Flows

Gladier makes it easy to change a flow or function. Both Flows and Functions are checksummed on each run so that changes to either can be detected automatically; Gladier will then re-register any function that has changed and update the flow if it has changed. Thus, users can focus on writing functions without worrying if a change has been registered. New tools can also be added to a flow, and tools can be removed from a flow simply by commenting them out.

Each flow or function update is logged in the `INFO` log in Python. Additionally, if you wish to see the full flow definition at any time, it can be queried in an instantiated client like so:

```
tar_transfer_cli = TarAndTransfer()
pprint(tar_transfer_cli.flow_definition)
```

1.3.2 Next Steps

- See the *Examples* section for full code snippets and interactive demos
- See *Gladier Tools* for a list of pre-made tools
- *Tools* for creating your own tools
- *SDK Reference* for detailed information about Gladier SDK components

1.4 Tools

Note: For the Python package containing pre-built and re-usable tools, see *Gladier Tools*

As noted in A simple Gladier application, a Gladier application defines a set of tools that are to be executed by the associated flow.

Gladier Tools are the glue that holds together Globus Flows and FuncX functions. A tool bundles everything that a FuncX function needs to run, so that the Gladier Client can register the function, check its requirements, and run it inside the Globus Flow.

The following code implements a simple tool, MakeDirs, that runs a single funcX function, makedirs.

```
from gladier import GladierBaseTool, generate_flow_definition


def makedirs(**data):
    """Make a directory on the filesystem"""
    import os
    os.makedirs(data['name'], mode=data['mode'], exist_ok=data['exist_ok'])
    return data['name']


@generate_flow_definition
class MakeDirs(GladierBaseTool):
    """List files on the filesystem"""
    funcx_functions = [makedirs]
    required = ['name']
    flow_input = {
        'mode': 0o777,
        'exist_ok': False
    }
```

In the same module, this can be added to a client by simply adding the `MakeDirs` class to `gladier_tools`:

```
# from gladier import GladierBaseClient
@generate_flow_definition
class ExampleClient(GladierBaseClient):
   gladier_tools = [MakeDirs]
```

A tool is composed of zero or more funcx_functions and a flow_definition. A flow can be specified explicitly, or if it involves just a single FuncX function, can be generated automatically by using the @generate_flow_definition decorator. When applied to tools, this decorator will create a simple one-state flow for the function to run.

The main attributes of a Gladier Tool are here:

- *funcx_functions* (list of callables) – A list of functions this tool uses.
- *flow_input* (dict): Default input that should be used in the flow. This is automatically overridden if the user supplies flow input to the Gladier Client
- *required* (list of strings): A list of critical Funcx flow_input keys that must be present for the function to be run. Gladier will raise an exception if these are not present when the user attempts to run the flow.
- *flow_definition* (dict): (Optional) A complete Globus Flow for running this tool. Provides a built-in flow new users can use to instantly run your tool.

The *Gladier Tools* package provides a set of predefined, generally useful tools for common tasks.

1.5 Flow Generation

For simple flows, Gladier can handle the process of stitching together the flows of several tools into one big flow, or even generating a flow from scratch for a simple Gladier tool. Flow generation does not replace manually authored flows, but provides a path for automating simple flows that don't require a lot of branching.

Flow generation is available for both clients and tools with the `@generate_flow_definition` decorator. Although it's applied the same way, the decorator behaves a little differently for a client and a tool.

- Gladier Tool – Generates a flow for each function defined in `funcx_functions`
- Gladier Client – Combines flows for each Gladier Tool defined in `gladier_tools`

1.5.1 Flow Generation on Gladier Tools

For Gladier Tools that only need one flow step per FuncX Function, flow generation can be a good option. Gladier will automatically determine information about the functions on the tool, and incorporate them into the flow. The usage looks like this:

```
from gladier import GladierBaseTool, generate_flow_definition


def ls(data):
    """Do an 'ls' on the filesystem, given a ``dir``"""
    import os
    the_dir = data.get('dir', '~')
    return os.listdir(the_dir)


@generate_flow_definition
class FileSystemListCommand(GladierBaseTool):
    """List files on the filesystem"""
    funcx_functions = [ls]
```

The decorator `@generate_flow_definition` will automatically set the `flow_definition` attribute when a GladierClient includes it in `tools`. The flow it generates will be identical to the following:

```
{
  "Comment": "List files on the filesystem",
  "StartAt": "Ls",
  "States": {
    "Ls": {
      "ActionScope": "https://auth.globus.org/scopes/facd7ccc-c5f4-42aa-916b-
↪a0e270e2c2a9/automate2",
      "ActionUrl": "https://api.funcx.org/automate",
      "Comment": "Do an 'ls' on the filesystem, given a ``dir``",
      "Type": "Action",
      "ExceptionOnActionFailure": false,
      "Parameters": {
        "tasks": [
```

(continues on next page)

(continued from previous page)

```
          {
            "endpoint.$": "$.input.funcx_endpoint_compute",
            "func.$": "$.input.ls_funcx_id",
            "payload.$": "$.input"
          }
        ]
      },
      "End": true,
      "ResultPath": "$.Ls",
      "WaitTime": 300
    }
  }
}
```

Flow Generation Modifiers

You'll notice that flow generation makes some assumptions you might want to change. For example, `ls` is not a compute heavy task, and might be better to run on the head node. Overriding some attributes of the flow can be done with `modifiers`.

```
@generate_flow_definition(modifiers={
    ls: {'endpoint': 'funcx_endpoint_non_compute'}
})
class FileSystemListCommand(GladierBaseTool):
    """List files on the filesystem"""
    funcx_functions = [ls]
```

1.5.2 Flow Generation on Gladier Clients

When Flow Generation is applied to Gladier Clients it does not generate the flow from scratch, but instead combines the flow on each Gladier Tool into one big flow.

Note that modifiers can also be used with Gladier Clients to customize some attributes.

```
@generate_flow_definition(modifiers={
   'generate_metadata': {'endpoint': 'funcx_endpoint_non_compute'},
   'publish_to_search': {'endpoint': 'funcx_endpoint_non_compute',
                         'payload': 'generate_metadata'}
})
class ProcessData(GladierBaseClient):
    gladier_tools = [
        'mytools.TransferData',
        'mytools.processing.ProcessData',
        'mytools.processing.GenerateGraphs',
        'mytools.processing.GenerateMetadata',
        'mytools.processing.PublishResults',
    ]
```

Modifiers

A complete reference is coming soon!

1.6 Passing Payloads

This section shows how to pass outputs of one FuncX Function to another using Globus Flows.

1.6.1 Simple Payloads

Sometimes you may need a state that depends on the output from another state. This can either be to split up complex functions, or make FuncX Parallelize payloads into multiple simultanious tasks.

By default, all Gladier Tools take input from the main input source, defined as `$.input`. However, by using modifiers (see more in *Flow Generation*), this default can be changed to use the output of another funcx function.

```
@generate_flow_definition(modifiers={
  my_second_function: {'payload': '$.MyFirstFunction.details.result[0]'},
})
class MyTool(GladierBaseTool):
    funcx_functions = [
        my_first_function,
        my_second_function,
    ]
```

In the example above, the first function is given the full input to work with on `$.input`. The output of `my_first_function` will be produced with the name `MyFirstFunction.details.result` as a list of FuncX task results. By default, only one FuncX task is run per-function, so typically this will be a list with only one entry. The path `$.MyFirstFunction.details.result[0]` references the exact output returned by a single invocation of `my_first_function`.

Note: Gladier Automatically creates flow state names by translating them from snake case to upper camel case. For example, `my_first_function` results in the state name `MyFirstFunction`

When `my_first_function` finishes and `my_second_function` begins, it will be given the input stored in `$.MyFirstFunction.details.result[0]`. This value **MUST** be a dictionary containing expected parameters in `my_second_function`, otherwise a flow exception will be raised and the flow will be marked as a failure.

Warning: When using function outputs as payloads with `ExceptionOnActionFailure: false`, this can result in cascading failures where the stringified exception results are used as input to the next function. It's recommended you either set `ExceptionOnActionFailure: true` or pass payloads as `$.MyFirstFunction.details`.

1.6.2 Multiple FuncX Tasks

FuncX is built to run many tasks in parallel. You can instruct Gladier to pass multiple task payloads with the `tasks` modifier. However, at this level FuncX also needs an expcilit FuncX endpoint and Function ID for each task it will process. It's common to use one FuncX function to build the list of payloads to be run in parallel.

```
def parallel_workload_input_builder(funcx_endpoint_compute, parallel_workload_funcx_id,␣
↪parallel_workloads, **data):
    return [{
        'endpoint': funcx_endpoint_compute,
        'function': parallel_workload_funcx_id,
        'payload': payload,
    } for payload in parallel_workloads]


def parallel_workload(name, **data):
    import time
    return f'{name} finished at {time.time()}!'


@generate_flow_definition(modifiers={
    parallel_workload: {'tasks': '$.ParallelWorkloadInputBuilder.details.result[0]'},
})
class ParallelWorkloadsTool(GladierBaseTool):
    funcx_functions = [
        parallel_workload_input_builder,
        parallel_workload,
    ]
    required_input = [
        'funcx_endpoint_compute',
        'parallel_workloads',
        'parallel_workload_funcx_id'
    ]
```

Above, the `parallel_workload_input_builder` function is run first and generates the list of FuncX tasks. This can be an arbitrarily long list determined at runtime. Each task in the list must contain three elements: `endpoint`, `function` and `payload`.

`endpoint` above is typically specified by the user at input time, and is by default `funcx_endpoint_compute`. But the FuncX function is updated by Gladier every change, and the name is determined automatically. By default, Gladier appends `_funcx_id` to the end of each of the `funcx_function` definitions and automaticaly adds them to `$.input`. `parallel_workload_funcx_id` can be determined above using this method, or one can verify via the flow output.

`payload` must be a dictonary containing keyword parameters for the function which match the function signature. This is similar to all other FuncX functions used in Gladier, which are called with all input data specified on `$.input`.

When `parallel_workload` runs, it will execute all tasks in parallel, or by any rules defined by your particular FuncX endpoint. Each of the outputs will be listed in `$.ParallelWorkload.details.result` once all tasks finish. If any task fails, a stack trace will be returned as a string. If all tasks fail, the flow will be marked as "FAILED".

A full example of the Flow Definition as JSON output is below:

```
{
    "Comment": "Flow with states: ParallelWorkloadInputBuilder, ParallelWorkload",
    "StartAt": "ParallelWorkloadInputBuilder",
```

(continues on next page)

(continued from previous page)

```
    "States": {
        "ParallelWorkloadInputBuilder": {
        "Comment": null,
        "Type": "Action",
        "ActionUrl": "https://automate.funcx.org",
        "ActionScope": "https://auth.globus.org/scopes/b3db7e59-a6f1-4947-95c2-
↪59d6b7a70f8c/action_all",
        "ExceptionOnActionFailure": false,
        "Parameters": {
            "tasks": [
            {
                "endpoint.$": "$.input.funcx_endpoint_compute",
                "function.$": "$.input.parallel_workload_input_builder_funcx_id",
                "payload.$": "$.input"
            }
            ]
        },
        "ResultPath": "$.ParallelWorkloadInputBuilder",
        "WaitTime": 300,
        "Next": "ParallelWorkload"
        },
        "ParallelWorkload": {
        "Comment": null,
        "Type": "Action",
        "ActionUrl": "https://automate.funcx.org",
        "ActionScope": "https://auth.globus.org/scopes/b3db7e59-a6f1-4947-95c2-
↪59d6b7a70f8c/action_all",
        "Parameters": {
            "tasks.$": "$.ParallelWorkloadInputBuilder.details.result[0]"
        },
        "ResultPath": "$.ParallelWorkload",
        "WaitTime": 300,
        "ExceptionOnActionFailure": true,
        "End": true
        }
    }
}
```

1.6.3 Parallel Processing Example

Below is a full runnable example, using the FuncX tutorial endpoint.

```
from gladier import GladierBaseClient, GladierBaseTool, generate_flow_definition
from pprint import pprint


def parallel_workload_input_builder(funcx_endpoint_compute, parallel_workload_funcx_id,␣
↪parallel_workloads, **data):
    return [{
        'endpoint': funcx_endpoint_compute,
        'function': parallel_workload_funcx_id,
```

(continues on next page)

(continued from previous page)

```
        'payload': payload,
    } for payload in parallel_workloads]


def parallel_workload(name, **data):
    import time
    return f'{name} finished at {time.time()}!'


@generate_flow_definition(modifiers={
    parallel_workload: {'tasks': '$.ParallelWorkloadInputBuilder.details.result[0]'},
})
class ParallelWorkloadsTool(GladierBaseTool):
    funcx_functions = [
        parallel_workload_input_builder,
        parallel_workload,
    ]
    required_input = [
        'funcx_endpoint_compute',
        'parallel_workloads',
        'parallel_workload_funcx_id'
    ]


@generate_flow_definition
class ParallelWorkloadsClient(GladierBaseClient):
    gladier_tools = [
        ParallelWorkloadsTool,
    ]


if __name__ == '__main__':
    flow_input = {
        'input': {
            'parallel_workloads': [
                {'name': 'foo'},
                {'name': 'bar'},
                {'name': 'baz'},
            ],
            'funcx_endpoint_compute': '553e7b64-0480-473c-beef-be762ba979a9',
        }
    }
    work_flow = ParallelWorkloadsClient()
    pprint(work_flow.flow_definition)

    flow = work_flow.run_flow(flow_input=flow_input)
    run_id = flow['run_id']
    work_flow.progress(run_id)
    pprint(work_flow.get_status(run_id))
```

1.7 Auth in Gladier

By Default, Gladier will automatically initiate a login as needed. This behavior can be customized if Glaider needs to be used as part of a larger app.

1.7.1 Scopes

Gladier requires scopes for the following Services:

- **Globus Flows Service**
  - *https://auth.globus.org/scopes/eec9b274-0c81-4334-bdc2-54e90e689b9a/run*
  - *https://auth.globus.org/scopes/eec9b274-0c81-4334-bdc2-54e90e689b9a/run_manage*
  - *https://auth.globus.org/scopes/eec9b274-0c81-4334-bdc2-54e90e689b9a/view_flows*
  - *https://auth.globus.org/scopes/eec9b274-0c81-4334-bdc2-54e90e689b9a/run_status*
  - *https://auth.globus.org/scopes/eec9b274-0c81-4334-bdc2-54e90e689b9a/manage_flows*
- **FuncX**
  - *openid*
  - *urn:globus:auth:scope:search.api.globus.org:all*
  - *https://auth.globus.org/scopes/facd7ccc-c5f4-42aa-916b-a0e270e2c2a9/all*
- **Deployed Flow**
  - Scope varies per-flow

Note that the Deployed Flow will be unique for each Glaider Client, and the scope will not exist until the flow is deployed. This sometimes requires multiple logins with Globus, first to fetch the base flows service scopes to deploy the flow, then a second login to get tokens to run the newly deployed flow.

Note also, that dependent scopes underlying the deployed flow may also change if the flow is modified to add additional services. For example, a flow could be initially deployed to do a simple transfer task, then modified and run again but with an additional search ingest task. If this happens, a new login must take place in order for the modified flow to be run again.

1.7.2 Storage

By default, tokens in Gladier are stored in `~/.gladier-secrets.cfg`

1.7.3 Customizing Auth

The default behavior of Auth in Gladier can be changed by passing a custom Login Manager into any Gladier Client:

```
from gladier import CallbackLoginManager

def callback(scopes: List[str]) -> Mapping[str, Union[AccessTokenAuthorizer,␣
↪RefreshTokenAuthorizer]]:
    authorizers_by_scope = do_my_login(scopes)
    return authorizers_by_scope
```

(continues on next page)

(continued from previous page)

```
callback_login_manager = CallbackLoginManager(
    # A dict of authorizers mapped by scope can be provided if available
    initial_authorizers,
    # If additional logins are needed, the callback is called.
    callback=callback
)

MyGladierClient(login_manager=callback_login_manager)
```

`my_custom_login_function` should be capable of both completeing a Globus Auth flow and storing tokens for future invacations. Ideally, `initial_authorizers` will contain all of the scopes needed so no login is needed.

1.7.4 Complete example

The complete example below uses Fair Research Login to demontrate a customized login flow. Please note, this example assumes customization using a Native App. Those using the authorization code grant, such as in a webservice, must modify their app accordingly.

```
from typing import List, Mapping, Union
from globus_sdk import AccessTokenAuthorizer, RefreshTokenAuthorizer
from fair_research_login import NativeClient, LoadError
from gladier import generate_flow_definition, GladierBaseClient, CallbackLoginManager


# A simple shell tool will be used for demonstration
@generate_flow_definition
class GladierTestClient(GladierBaseClient):
    gladier_tools = [
        "gladier_tools.posix.shell_cmd.ShellCmdTool",
    ]


# Fair Research Login is used for simplicity
frl = NativeClient(client_id="7414f0b4-7d05-4bb6-bb00-076fa3f17cf5")

try:
    # Try to use a previous login to avoid a new login flow
    initial_authorizers = frl.get_authorizers_by_scope()
except LoadError:
    # Passing in an empty dict will trigger the callback below
    initial_authorizers = {}


def callback(scopes: List[str]
             ) -> Mapping[str, Union[AccessTokenAuthorizer, RefreshTokenAuthorizer]]:
    # 'force' is used for any underlying scope changes. For example, if a flow adds␣
↪transfer
    # functionality since it was last run, running it again would require a re-login.
    frl.login(requested_scopes=scopes, force=True, refresh_tokens=True)
    return frl.get_authorizers_by_scope()
```

(continues on next page)

(continued from previous page)

```
custom_login_manager = CallbackLoginManager(
    initial_authorizers,
    # If additional logins are needed, the callback is used.
    callback=callback
)

# Pass in any custom login manager to modify the behavior. Everything else stays the␣
↪same.
client = GladierTestClient(login_manager=custom_login_manager)
run = client.run_flow(flow_input={
    "input": {
        "args": "echo 'Hello Custom Login!'",
        "capture_output": True,
        "funcx_endpoint_compute": "4b116d3c-1703-4f8f-9f6f-39921e5864df",
    }
})

run_id = run['run_id']
client.progress(run_id)
print(f"Flow result: {client.get_status(run_id)['status']}")
```

CHAPTER

TWO

GLADIER TOOLS

Gladier tools is an optional package containing commonly used operations. You can install Gladier Tools with the following:

```
pip install gladier-tools
```

Once the Gladier Tools packages is installed, individual tools can be added to your client like so:

```
@generate_flow_definition
class TarAndTransfer(GladierBaseClient):
    gladier_tools = [
        'gladier_tools.posix.Tar',
        'gladier_tools.globus.Transfer',
    ]
```

The following Gladier Tools are available:

```
# Posix
'gladier_tools.posix.Tar',
'gladier_tools.posix.UnTar',
'gladier_tools.posix.Encrypt',
'gladier_tools.posix.Decrypt',
'gladier_tools.posix.AsymmetricEncrypt',
'gladier_tools.posix.AsymmetricDecrypt',
'gladier_tools.posix.HttpsDownloadFile',

# Globus
'gladier_tools.globus.Transfer',

# Publish
'gladier_tools.publish.Publish',
```

See the sections below for detailed information about each tool.

2.1 Globus Tools

Globus Tools map to the list of Globus Maintained Action Providers

They may not even contain FuncX functions, but are rather a convenient way to add common flow operations to your Gladier Client without copying the JSON yourself.

2.1.1 Globus Transfer

class gladier_tools.globus.transfer.**Transfer**(*alias: Optional[str] = None*, *alias_class: Optional[ToolAlias] = None*)

Bases: GladierBaseTool

Transfer is a simple single-state flow with no FuncX Functions, which talks directly to the Transfer Action Provider. It transfers only a single file or directory.

Parameters

- **transfer_source_endpoint_id** – Globus Source Endpoint UUID
- **transfer_destination_endpoint_id** – Globus Destination Endpoint UUID
- **transfer_source_path** – Globus Source Path
- **transfer_destination_path** – Globus Destination Path
- **transfer_recursive** – True if this is a directory, false otherwise.

2.2 Publish Tools

Publish Tools talk to both Globus Search and a Globus Transfer as one single operation. Typically, this is the final state in a processing flow as a way to submit data for viewing in a portal.

2.2.1 Publish

Publish takes files or directories and transfers them to another publication endpoint and ingests metadata about the files into Globus Search. Both the index in Globus Search and the Globus Endpoint in Globus Pilot must be setup first.

Setup only needs to be done once, then publish can be used freely afterwards. Both Globus Pilot and Globus Search CLI can be installed with the following:

```
pip install globus-search-cli globus-pilot
```

Globus Search CLI is responsible for setting up indices in Globus Search. Globus Pilot is a tool which generates metadata about files and directories and handles both the transfer to an endpoint and the ingest to search.

Create the index with the following:

```
globus-search index create my-index
```

Setup your publication endpoint with the index you created above with:

```
globus-pilot index setup <UUID from the step above>
```

After that, you should be ready to publish to your data. See documentation for both of the tools above here:

- Globus Search CLI
- Globus Pilot

class gladier_tools.publish.**Publish**(*alias: Optional[str] = None*, *alias_class: Optional[ToolAlias] = None*)

Bases: `GladierBaseTool`

This function uses the globus-pilot tool to generate metadata compatible with portals on https://acdc.alcf.anl.gov/. Requires globus_pilot>=0.6.0.

FuncX Functions:

- publish_gather_metadata (funcx_endpoint_non_compute)

Publication happens in three steps:

- PublishGatherMetadata – A funcx function which uses globus-pilot to gather metadata on files or folders
- PublishTransfer – Transfers data to the Globus Endpoint selected in Globus Pilot
- PublishIngest – Ingest metadata gathered in fist step to Globus Search

Note: This tool needs internet access to fetch Pilot configuration records, which contain the destination endpoint and other project info. The default FuncX endpoint name is *funcx_endpoint_non_compute*. You can change this with the following modifier:

```
@generate_flow_definition(modifiers={
    'publish_gather_metadata': {'endpoint': 'funcx_endpoint_non_compute'},
})
```

More details on modifiers can be found at https://gladier.readthedocs.io/en/latest/gladier/flow_generation.html

NOTE: This tool nests input under the 'pilot' keyword. Submit your input as the following:

```
{
    'input': {
        'pilot': {
            'dataset': 'foo',
            'index': 'my-search-index-uuid',
            'project': 'my-pilot-project',
            'source_globus_endpoint': 'ddb59aef-6d04-11e5-ba46-22000b92c6ec',
        }
}
```

Parameters

- **dataset** – Path to file or directory. Used by Pilot to gather metadata, and set as the source for transfer to the publication endpoint configured in Pilot.
- **destination** – relative location under project directory to place dataset (Default /)
- **source_globus_endpoint** – The Globus Endpoint of the machine where you are executing
- **source_collection_basepath** – If using a guest collection, the posix path of the guest collection. Used to translate source paths for the transfer step. (Default: /)
- **index** – The index to ingest this dataset in Globus Search
- **project** – The Pilot project to use for this dataset
- **groups** – A list of additional groups to make these records visible_to.

- **funcx_endpoint_non_compute** – A funcX endpoint uuid for gathering metadata. Requires internet access.

Requires: the 'globus-pilot' package to be installed.

2.3 Posix Tools

Below are Gladier Tools for common operations on posix filesystems.

2.3.1 Tar

class gladier_tools.posix.tar.**Tar**(*alias: str = None*, *alias_class: ToolAlias = None*)

Bases:

The Tar tool makes it possible to create Tar archives from folders. FuncX Functions:

- tar (funcx_endpoint_compute)

Parameters

- **tar_input** – Input directory to archive.
- **tar_output** – (optional) output file to save the new archive. Defaults to the original input file with an extension (myfile.tgz) if not given.
- **funcx_endpoint_compute** – By default, uses the `compute` funcx endpoint.

Returns path
The name of the newly created archive.

2.3.2 Untar

class gladier_tools.posix.untar.**UnTar**(*alias: str = None*, *alias_class: ToolAlias = None*)

Bases:

The UnTar tool makes it possible to extract data from Tar archives. FuncX Functions:

- untar (funcx_endpoint_compute)

Parameters

- **untar_input** – Input directory to archive.
- **untar_output** – (optional) output file to save the new archive. Defaults to the original # noqa input file with an extension '.tgz' removed.
- **funcx_endpoint_compute** – By default, uses the `compute` funcx endpoint. # noqa

Returns path
The output location of the extracted archive

Raises

- `ValueError` – If any files within the tar would extract to a non-relative location
- `FileNotFoundError` – If the file does not exist.

2.3.3 Shell CMD

class gladier_tools.posix.shell_cmd.**ShellCmdTool**(*alias: str = None*, *alias_class: ToolAlias = None*)

Bases:

Run a command in a shell with various options. Suitable for use as a funcx function. On anything other than exit success (returning 0), an exception will be raised, the stack trace will be propogated, and the flow will be halted. On exit success, the return code, stdout, and sterr are returned in a list in the format:

[0, "Standard Output Text", "Standard Error Text"]

Note: stdout and sterr will always be None unless `capture_output` is set.

Parameters

- **args** – Arguments to the command. Can be either a list of strings containing command and parameters, or a single string with command and parameters together.
- **arg_sep_char** – If the args val is a string, the character that separates the various arguments. By default, it is space.
- **capture_output** – Whether output should be captured. If so, the return value will contain captured text. Beware of capturing too much text, especially in funcx and Globus Flows use cases where output size is limited.
- **cwd** – [Optional] The directory to run the command from. When used in funcx,the run directory may be unpredictable based on how the funcx endpoint is started and configured.
- **env** – [Optional] A dictionary of environment variables to set
- **timeout** – How long the command should be allowed to run. Default is 60 seconds.
- **exception_on_error** – If the command fails, should an exception be raised or should just the (presumably non-zero) return code be returned. Defaults to True.
- **input_path** – A path to a file which should be used as the standard input to the command. When not provided, stdin does not exist so attempts to read from it will result in an error.
- **output_path** – A path to a file which should be used to capture the output (stdout) of the command. This cannot be set if *capture_output* is set.
- **error_path** – A path to a file which should be used to capture the error output (stderr) of the command. This cannot be set if *capture_output* is set.
- ****kwargs** –
- **nil** –

Returns

A tuple containing the return code of the command execution, the string of the standard out and standard error (if capture_output is True)

2.3.4 HTTPS Download File

class gladier_tools.posix.https_download_file.**HttpsDownloadFile**(*alias: str = None*, *alias_class: ToolAlias = None*)

Bases:

2.3.5 Encrypt

class gladier_tools.posix.encrypt.**Encrypt**(*alias: str = None*, *alias_class: ToolAlias = None*)

Bases:

The Encrypt tool takes in a file and a password to perform 128-bit AES symmetric key encryption on the file. The original contents of the file are overwritten with the encrypted text. Adds an extension (.aes) to the name of the file. It has not been found to be compatible with 3rd party encryption/decryption tools.

FuncX Functions:

- encrypt (funcx_endpoint_compute)

Parameters

- **encrypt_input** – Path to the file which needs to be encrypted.
- **encrypt_output** – Custom path to outputfile. Default is the same file with the '.aes' suffix added
- **encrypt_key** – Symmetric key or "password" which can be used to decrypt the encrypted file.
- **funcx_endpoint_compute** – By default, uses the `compute` funcx endpoint.

Returns output_path
Location of the encrypted file.

2.3.6 Decrypt

class gladier_tools.posix.decrypt.**Decrypt**(*alias: str = None*, *alias_class: ToolAlias = None*)

Bases:

Decrypt tool takes in an encrypted file and a password to perform decryption on the file. The decryption only works on files that have been encrypted by the Gladier Encrypt tool. It has not been found to be compatible with 3rd party encryption/decryption tools.

FuncX Functions:

- decrypt (funcx_endpoint_compute)

Parameters

- **decrypt_input** – Path to the file which needs to be decrypted.
- **decrypt_key** – Symmetric key or "password" which will be used to decrypt the encrypted file. Must be the same key that was used during encryption.
- **decrypt_output** – (optional) The full path to the decrypted file. If not provided, the decrypted file will have the same name as the input file, with the last 4 characters truncated(assuming it was a .aes file).

- **funcx_endpoint_compute** – By default, uses the compute funcx endpoint.

Returns output_path
Location of the decrypted file.

2.3.7 Asymmetric Encrypt

class gladier_tools.posix.asymmetric_encrypt.**AsymmetricEncrypt**(*alias: str = None*, *alias_class: ToolAlias = None*)

Bases:

The Asymmetric Encrypt tool takes in a file and the path to the RSA public key to perform RSA encryption on the file. Adds an extension (.rsa) to the name of the file. It has not been found to be compatible with 3rd party encryption/decryption tools.

FuncX Functions:

- asymmetric_encrypt (funcx_endpoint_compute)

Parameters

- **public_key_path** – Path to the .pub file which contains the RSA public key. Defaults to ~/.ssh/id_rsa.pub
- **asym_encrypt_file** – File which needs to be encrypted.
- **funcx_endpoint_compute** – By default, uses the compute funcx endpoint.

Returns output_path
Location of the encrypted file.

2.3.8 Asymmetric Decrypt

class gladier_tools.posix.asymmetric_decrypt.**AsymmetricDecrypt**(*alias: str = None*, *alias_class: ToolAlias = None*)

Bases:

The Asymmetric Decrypt tool takes in a file encrypted by the asymmetric encryption tool and the path to the RSA public key to perform decryption on the file. The output file can be passed in as a flow_input. If no output file passed, the last 4 characters (.rsa) of the input file are removed. It has not been found to be compatible with 3rd party encryption/decryption tools.

FuncX Functions:

- asymmetric_decrypt (funcx_endpoint_compute)

Parameters

- **private_key_path** – Path to the id_rsa file which contains the RSA private key. Defaults to ~/.ssh/id_rsa if not passed in.
- **asym_decrypt_file** – File which needs to be decrypted.
- **funcx_endpoint_compute** – By default, uses the compute funcx endpoint.
- **asym_decrypt_password** – (Optional) If the private file is password protected, pass it in through this argument.
- **output_file** – (Optional) Path to the output file which holds the decrypted contents.

Returns output_path
Location of the decrypted file.

2.4 See Also

- *Tools*

CHAPTER

THREE

EXAMPLES

3.1 Pre Requisites

If you want to run the examples below on your own computer, you need to follow the *Setup* guide. The *Setup* guide covers setting up FuncX and GCP on the machine you are using.

Note: For a quick 5-minute start, you can run the Binder Examples, which are all run in a pre-built environment and require no setup.

3.2 Example Flows

3.2.1 Tar and Transfer

This example highlights the usage of the Tar and Transfer Gladier tools in a single flow. The tar step creates an archive of the input file, and the transfer step transfers the archived file between Globus Endpoints.

```
from gladier import GladierBaseClient, generate_flow_definition
from pprint import pprint


@generate_flow_definition
class TarAndTransfer(GladierBaseClient):
    gladier_tools = [
        'gladier_tools.posix.Tar',
        'gladier_tools.globus.Transfer',
    ]


if __name__ == '__main__':
    flow_input = {
        'input': {
            # The directory below should exist with files. It will be archived by the␣
↪Tar Tool.
            'tar_input': '~/myfiles',
            # Set this to your own funcx endpoint where you want to tar files
            # 'funcx_endpoint_compute': '',
            # Set this to the globus endpoint where your tarred archive has been created
```

(continues on next page)

(continued from previous page)

```
            # 'transfer_source_endpoint_id': '',
            # By default, this will transfer the tar file to Globus Tutorial Endpoint 1
            'transfer_destination_endpoint_id': 'ddb59aef-6d04-11e5-ba46-22000b92c6ec',
            # By default, the Tar Tool will append '.tgz' to the archive it creates
            'transfer_source_path': '~/myfiles.tgz',
            'transfer_destination_path': '~/my_archives/myfiles.tgz',
            'transfer_recursive': False,
        }
    }
    # Instantiate the client
    tar_and_transfer = TarAndTransfer()

    # Optionally, print the flow definition
    # pprint(tar_and_transfer.flow_definition)

    # Run the flow
    flow = tar_and_transfer.run_flow(flow_input=flow_input)

    # Track the progress
    action_id = flow['action_id']
    tar_and_transfer.progress(action_id)
    pprint(tar_and_transfer.get_status(action_id))
```

Steps

Start the workflow by overriding the Gladier Client with a list of the tools you want to use. In this case, we want to use the Tar, and Transfer tools.

```
@generate_flow_definition
class TarAndTransfer(GladierBaseClient):
   gladier_tools = [
      'gladier_tools.posix.Tar',
      'gladier_tools.globus.Transfer',
   ]
```

The `@generate_flow_definition` decorator takes the flow definitions of the individual tools and constructs a flow definition for the new flow, so we do not have to define a custom flow definition. To view the constructed flow definition, use `pprint(tat.flow_definition)`. Refer to the Flow Generation doc for more details on the same.

The next step is to define the input for the flow. It might be helpful to refer to the docs for each of the tools to find out what needs to be passed in as input. For example, here are the documentations for the Tar and Transfer tools. Feel free to use the below blueprint:

```
flow_input = {
   'input': {
      # Set this to the file/folder that has to be tarred
      'tar_input': '',
      # Set this to your own funcx endpoint where you want to tar files
      'funcx_endpoint_compute': '',
      # Set this to the globus endpoint where your tarred archive has been created
      'transfer_source_endpoint_id': '',
      # By default, this will transfer the tar file to Globus Tutorial Endpoint 1
```

(continues on next page)

(continued from previous page)

```
        'transfer_destination_endpoint_id': 'ddb59aef-6d04-11e5-ba46-22000b92c6ec',
        # Provide the paths to the tarred file as input for the transfer step
        'transfer_source_path': '',
        'transfer_destination_path': '',
        'transfer_recursive': False,
    }
}
```

All that is left is to create an instance of the `GladierBaseClient` class and run the flow. Use the below code to view the progress of the flow:

```
tat = TarAndTransfer()
pprint(tat.flow_definition)
flow = tat.run_flow(flow_input=flow_input)
action_id = flow['action_id']
tat.progress(action_id)
pprint(tat.get_status(action_id))
```

3.2.2 Encrypt and Transfer

This example highlights the usage of built in tools (Tar, Encrypt and Transfer) along with a custom defined tool (MakeFiles) in the same flow.

```
from gladier import GladierBaseClient, generate_flow_definition, GladierBaseTool
from pprint import pprint


def make_files(**data):
    from pathlib import Path

    input_path = Path(data['make_input']).expanduser()
    input_path.mkdir(exist_ok=True, parents=True)

    for number in range(3):
        with open(input_path / f'file{number}.txt', 'w') as f:
            f.write(f'This is file no. {number}')

    return str(input_path)


@generate_flow_definition
class MakeFiles(GladierBaseTool):
    funcx_functions = [make_files]
    required_input = [
        'make_input',
        'funcx_endpoint_compute'
    ]


@generate_flow_definition
class EncryptAndTransfer(GladierBaseClient):
```

(continues on next page)

(continued from previous page)

```
    gladier_tools = [
        MakeFiles,
        'gladier_tools.posix.Tar',
        'gladier_tools.posix.Encrypt',
        'gladier_tools.globus.Transfer',
    ]


if __name__ == '__main__':
    flow_input = {
        'input': {
            # Set this to the folder in which you want to run the makeFiles function in
            'make_input': '/tmp/myfiles',
            # Set this to the same folder as above
            'tar_input': '/tmp/myfiles',
            # Set this to the resultant archive of the above folder
            'encrypt_input': '/tmp/myfiles.tgz',
            # Set this to the symmetric key you want to use to encrypt/decrypt the file
            'encrypt_key': 'my_secret',
            # Set this to your own funcx endpoint where you want to encrypt files
            # 'funcx_endpoint_compute': '',
            # Set this to the globus endpoint where your encrypted archive has been␣
↪created
            # 'transfer_source_endpoint_id': '',
            # By default, this will transfer the encrypt file to Globus Tutorial␣
↪Endpoint 1
            'transfer_destination_endpoint_id': 'ddb59aef-6d04-11e5-ba46-22000b92c6ec',
            'transfer_source_path': '/tmp/myfiles.tgz.aes',
            'transfer_destination_path': 'my_encrypted_files/myfiles.tgz.aes',
            'transfer_recursive': False,
        }
    }

    # Create the Client
    encrypt_and_transfer = EncryptAndTransfer()

    # Optionally print the flow definition
    # pprint(encrypt_and_transfer.flow_definition)

    # Run the flow
    flow = encrypt_and_transfer.run_flow(flow_input=flow_input)

    # Track progress
    action_id = flow['action_id']
    encrypt_and_transfer.progress(action_id)
    pprint(encrypt_and_transfer.get_status(action_id))
```

Steps

We start by writing our own custom tool. For the example, MakeFiles is a trivial tool that creates three files in a given folder. The funcx function that achieves this functionality is `make_files`. We then define the tool using `GladierBaseTool`.

```
@generate_flow_definition
class MakeFiles(GladierBaseTool):
    funcx_functions = [make_files]
    required_input = [
        'make_input',
        'funcx_endpoint_compute'
    ]
```

Defining a workflow is similar to the previous case, where we override the `GladierBaseClient` class. In this case, we want to use the MakeFiles, Tar, Encrypt, and Transfer tools.

```
@generate_flow_definition
class CustomTransfer(GladierBaseClient):
    gladier_tools = [
        MakeFiles,
        'gladier_tools.posix.Tar',
        'gladier_tools.posix.Encrypt',
        'gladier_tools.globus.Transfer',
    ]
```

The `@generate_flow_definition` decorator takes the flow definitions of the individual tools and constructs a flow definition for the new flow, so we do not have to define a custom flow definition. To view the constructed flow definition, use `pprint(ct.flow_definition)`. Refer to the Flow Generation doc for more details on the same.

The next step is to define the input for the flow. It might be helpful to refer to the docs for each of the tools to find out what needs to be passed in as input. For example, here are the documentations for the Tar, Encrypt and Transfer tools. Feel free to use the below blueprint:

```
flow_input = {
    'input': {
        # Set this to the folder in which you want to run the makeFiles function in
        'make_input': '',
        # Set this to the same folder as above
        'tar_input': '',
        # Set this to the resultant archive of the above folder
        'encrypt_input': '',
        # Set this to the symmetric key you want to use to encrypt/decrypt the file
        'encrypt_key': '',
        # Set this to your own funcx endpoint where you want to encrypt files
        'funcx_endpoint_compute': '',
        # Set this to the globus endpoint where your encrypted archive has been created
        'transfer_source_endpoint_id': '',
        # By default, this will transfer the encrypt file to Globus Tutorial Endpoint 1
        'transfer_destination_endpoint_id': 'ddb59aef-6d04-11e5-ba46-22000b92c6ec',
        'transfer_source_path': '',
        'transfer_destination_path': '',
        'transfer_recursive': False,
    }
}
```

All that is left is to create an instance of the `GladierBaseClient` class and run the flow. Use the below code to view the progress of the flow:

```
ct = CustomTransfer()
pprint(ct.flow_definition)
flow = ct.run_flow(flow_input=flow_input)
action_id = flow['action_id']
ct.progress(action_id)
pprint(ct.get_status(action_id))
```

CHAPTER

FOUR

SDK REFERENCE

4.1 Gladier Base Client

class gladier.client.**GladierBaseClient**(*authorizers: Optional[Mapping[str, Union[AccessTokenAuthorizer, RefreshTokenAuthorizer]]] = None*, *auto_login: bool = True*, *auto_registration: bool = True*, *login_manager: Optional[BaseLoginManager] = None*, *flows_manager: Optional[FlowsManager] = None*)

Bases: `object`

The Gladier Client ties together commonly used funcx functions and basic flows with auto-registration tools to make complex tasks easy to automate.

This class is intended to be subclassed as follows:

```
@generate_flow_definition
class MyGladierClient(GladierBaseClient):
    gladier_tools = [MyTool]
```

And used like the following:

```
my_gc = MyGladierClient()
flow = my_gc.run_flow(flow_input={"flow_input": {"my_field": "foo"}})
run_id = flow["run_id"]
my_gc.progress(run_id)
pprint(tar_and_transfer.get_status(run_id))
```

The following class variables can be set on clients to change their behavior when deploying and running flows.

- **glaider_tools (default: [])**
  - A list of Gladier Tools to build a working flow_defitinion. Each tool's minimum input must be satisfied prior to running the flow. Can be used with the @generate_flow_definition decorator to automatically chain together flow definitions present on each tool in linear order.
- **flow_definition (default: {})**
  - An explicit flow definition to use for this client. Cannot be used with @generate_flow_definition
- **secret_config_filename (default: `~/.gladier-secrets.cfg`)**
  - Storage are for Globus Tokens and general storage
- **app_name (default: 'Gladier Client')**
  - The app name used during a login flow

- **client_id**
  - The Globus Client ID used for native logins
- **globus_group (default: None)**
  - A Globus Group to be applied to all flow/run permissions. Group will automatically be added to flow_viewers, flow_starters, flow_administrators, run_managers, run_monitors
- **subscription_id (default: None)**
  - The subscription id associated with this flow
- **alias_class (default: gladier.utils.tool_alias.StateSuffixVariablePrefix)**
  - The default class used to for applying aliases to Tools

Default options are intended for CLI usage and maximum user convenience.

Parameters

- **auto_registration** – Automatically register functions or flows if they are not previously registered or obsolete.
- **login_manager** – Class defining login behavior. Defaults to AutoLoginManager, and will auto-login when additional scopes are needed.
- **flows_manager** – A flows manager class with customized behavior. Attrs like group and login_manager will automatically be set if None

Raises

gladier.exc.AuthException – if authorizers given are insufficient

`login()`

Call `login()` on the configured login manager. Attepmts to prepare the user to run flows, but may require being called twice if a flow is not yet deployed. Automatically called internally by `run_flow()` if required.

`logout()`

Call `logout()` on the login manager, to revoke saved tokens and deactivate the current flow. The flow_id and function ids/checksums are unaffected, and can be re-used after another invocation of `login()`.

`get_input()` → dict

Get funcx function ids, funcx endpoints, and each tool's default input. Default input may not be enough to run the flow. For example if a tool does processing on a local filesystem, the file will always need to be provided by the user when calling run_flow().

Defaults rely on GladierBaseTool.flow_input defined separately for each tool.

Returns

input for a flow wrapped in an 'input' dict. For example: {'input': {'foo': 'bar'}}

`get_status(action_id: str)`

Get the current status of the automate flow. Attempts to do additional work on funcx functions to deserialize any exception output.

Parameters

action_id – The globus action UUID used for this flow. The Automate flow id is always the flow_id configured for this tool.

Raises

Globus Automate exceptions from self.flows_client.flow_action_status

Returns

a Globus Automate status object (with varying state structures)

progress(*action_id*, *callback=None*)

Continuously call self.get_status() until the flow completes. Each status response is used as a parameter to the provided callback, by default will use the builtin callback to print the current state to stdout.

Parameters

- **action_id** – The action id for a running flow. The flow is automatically pulled based on the current tool's flow_definition.
- **callback** – The function to call with the result from self.get_status. Must take a single parameter: mycallback(self.get_status())

4.2 Flows Manager

Below is the class Gladier uses to track local changes to flows.

class gladier.**FlowsManager**(*flow_id: ~typing.Optional[str] = None*, *flow_definition: ~typing.Optional[dict] = None*, *flow_schema: ~typing.Optional[dict] = None*, *flow_title: ~typing.Optional[str] = None*, *globus_group: ~typing.Optional[str] = None*, *subscription_id: ~typing.Optional[str] = None*, *on_change: ~typing.Callable = <function ensure_flow_registered>*, *redeploy_on_404: bool = True*, ***kwargs*)

Bases: `ServiceManager`

The flows manager tracks an externally defined flow_definition and ensures it stays up-to-date with a deployment in the flows service. It can be run without a flow_id in which case it will deploy its own flow if a stored flow_id does not exist.

The flow checksum is evaluated inside `run_flow()`, and any updates to the flow will take place before the flow is started. The registration behavior can be customized by passing in a function for `on_change`. Default behavior will call `register_flow()`, None will disable changing the flow before starting it. The `on_change` signature is below:

```
def on_change_callback(flows_manager_instance: FlowsManager,
                       exc: gladier.exc.RegistrationException) -> None:
```

Parameters

- **flow_id** – Explicit flow id to use. None will result in deploying a new flow
- **flow_definition** – Flow definiton that should be used. Usually set dynamically at runtime when used with a Gladier Client
- **flow_schema** – The schema to be used alongside the flow definition
- **flow_title** – The title for the Globus Flow
- **globus_group** – A Globus Group UUID. Used to grant all flow and run permissions
- **subscription_id** – (deprecated) Subscription ID has no effect and will be removed in a future version.
- **on_change** – callback on checksum mismatch or missing flow id. Default registers/deploys flow, `None` takes no action and attempts to run "obselete" flows.
- **redeploy_on_404** – Deploy a new flow if attempting to run the current flow ID results in 404. Behavior is disabled if an explicit flow_id is specified.

When used with a Gladier Client, following items will be auto-configured and should not be set explicitly in the constructor:

- flow_definition
- flow_schema

Note: The FlowsManager class cannot be used to run flows outside of Gladier Clients due to internal class storage requirements. Doing so will result in an exception.

CHAPTER

FIVE

UPGRADE MIGRATIONS

5.1 Migrating to v0.5.0

The only major change in v0.5.0 was the removal of the HelloWorld Tools from the main Gladier package. The following are no longer present:

- `gladier.tools.hello_world.HelloWorld`
- `gladier.tools.manifest.ManifestTransfer`
- `gladier.tools.manifest.ManifestToFuncXTasks`

There currently aren't plans to rewrite them in the Gladier Tools package, but open an issue if you would like us to consider changing that!

5.2 Migrating to v0.4.0

Gladier v0.3.x depended on FuncX v0.0.5 and FuncX Endpoint v0.0.3. Gladier v0.4.x now uses Funcx v0.2.3-v0.3.0+ (funcx-endpoint v0.2.3-v0.3.0+). There are a number of breaking changes between these two versions of FuncX, including funcx endpoints, flow definitions, and backend services.

5.2.1 FuncX Endpoints

All FuncX endpoints will need to be recreated with the never version of FuncX. Gladier typically names these endpoints as the following:

- `funcx_endpoint_non_compute`
- `funcx_endpoint_compute`

Since these use different backend services, using endpoints that don't match the FuncX version will result in errors. Using 0.0.3 endponits on 0.2.3+ will result in permission denied, using 0.2.3+ on 0.0.3 will result in Server 500s.

5.2.2 Argument Passing and Function Definitions

Previously, all arguments in a Flow were passed to FuncX functions as a dict. It looked like the following:

```
'Parameters': {'tasks': [{'endpoint.$': '$.input.funcx_endpoint_non_compute',
                          'function': '8227609b-4869-4c6f-9a1b-87dc49fcc687',
                          'payload.$': '$.input'}]},

def my_function(data):
    ...
```

In the above, `data` would get the entire dict from $.input, which was typically whatever input was passed to start the flow. In the new version of FuncX, this has changed. All arguments are either positional or keyword arguments and should be named. This is difficult in automate, since naming arguments requires specifying them explicitly in the flow definition. An easy migration path is the following:

```
'Parameters': {'tasks': [{'endpoint.$': '$.input.funcx_endpoint_non_compute',
                          'function': '8227609b-4869-4c6f-9a1b-87dc49fcc687',
                          'payload.$': '$.input'}]},

def my_function(**data):
    ...
```

Changing data to a keyword argument will allow re-creating the same behavior as before.

5.2.3 FuncX Functions

Like FuncX Endpoints, FuncX Functions also need to be changed between versions. This is an automatic process in most cases if you are running the latest version of Gladier and saw a big giant warning when upgrading. Gladier will automatically delete funcx functions that don't match the newly supported version of FuncX Gladier uses.

However, it's necessary to do a manual upgrade to remove these functions in some cases. To upgrade manually, edit the file `~/.gladier-secrets.cfg`, and remove all config items that end in `funcx_id` and `funcx_id_checksum`:

```
hello_world_funcx_id = 3bccfcdb-bc0e-4549-9297-8e08c6f50bd5
hello_world_funcx_id_checksum =␣
↪c590423de52051e7b7bb044dc173673d2c9ad965f7f71bee665494815b3a2046
```

5.2.4 Flow Definitions

Some items in Automate flow definitions also changed. See below for a list of the attributes.

FuncX Version 0.0.5 flow definitions:

- `ActionUrl` – 'https://api.funcx.org/automate'
- `ActionScope` – 'https://auth.globus.org/scopes/facd7ccc-c5f4-42aa-916b-a0e270e2c2a9/automate2'

FuncX Version 0.2.3+ flow definitions:

- `ActionUrl` – 'https://automate.funcx.org'
- `ActionScope` – 'https://auth.globus.org/scopes/b3db7e59-a6f1-4947-95c2-59d6b7a70f8c/action_all'

Additionally for FuncX Payloads, Function UUIDs are passed with a different name.

'func.$': '$.input.'

Needs to be changed to:

'function.$': '$.input.'

5.2.5 FuncX Flow Result Format

The format of the return value from FuncX functions has changed format. This only affects Flow states that depend on the output of a FuncX function/flow state.

Previous flow states were not returned in a list, and were referenced with the following:

```
'InputPath': '$.MyFuncXFunctionOutput.details.result',
```

FuncX now returns these in a list, and they need to be index. The above needs to be changed to the following:

```
'InputPath': '$.MyFuncXFunctionOutput.details.result[0]',
```

CHAPTER

SIX

GLOSSARY

Gladier builds on a science services framework, in the form of Globus Auth, Transfer, Search, Groups, and Flows, plus funcX function-as-a-service. These services provide a reliable, secure, and high-performance substrate to access and manage data and computing resources. Here we highlight several of these services and describe how they are used to create Gladier deployments.

6.1 Globus

Globus provides a collection of data services built for science including: Globus Auth, Transfer, Search, Groups, and Flows, and funcX to enable distributed function-as-a-service execution. Globus Services are highly reliable, professionally operated cloud-hosted services that support the work of over 150,000 researchers worldwide as foundational capabilities for scientific applications and workflows; using them greatly reduces the burden on local systems, administrators, and programmers.

6.2 Globus Flows

Flows addresses the problem of securely and reliably automating sequences of data management tasks that may span locations, storage systems, administrative domains, and timescales, and integrate both mechanical and human inputs. Client libraries deployed on Globus endpoints and other sources enable the detection of events and invocation of a flow. The Flows service manages execution of user-supplied automation flows either manually or as a result of data events, and the invocation of actions from those automation flows, including actions provided by Globus endpoints and services. The service is extensible via the definition of new events and actions to meet the needs of specific communities.

6.3 FuncX

funcX is a function-as-a-service platform that implements a federated compute substrate, enabling computation to be registered as Python functions and invocations to be dispatched to remote computers for execution. The service provides a single point-of-contact, supporting function registration, sharing, and discovery as well as reliable and secure execution on connected endpoints. The funcX endpoint software, built on Parsl, allows functions to be executed in containers and for resources to be dynamically provisioned on cloud and cluster systems. These funcX endpoints provide serverless capabilities whereby researchers fire-and-forget tasks that are dynamically allocated across the supercomputer using an opportunistic backfill queue to utilize spare capacity.

6.4 Globus Queues and Triggers

Queues provides a reliable, cloud-based mechanism to manage and store events. The Queues service allows users to provision a dedicated queue for their instrument. Clients can then raise events to the queue using HTTP POST requests where they will be maintained until a subscriber consumes them. This enables experimental facilities and instruments to raise events as data are created without requiring heavy-weight installations on the edge device.

The **Triggers** service provides a cloud-based consumer of Queues events. Users can configure a Trigger to monitor a queue and initiate Flows as events are received. To create a trigger one defines:

- An event queue for the trigger to monitor
- A condition for when the trigger will fire
- An action to perform when the condition is met (e.g., a flow uuid)
- A template to create an input JSON document for the action. This often includes default values.

The combination of the **Queues** and **Triggers** services simplifies creating new Gladier deployments.

6.5 Globus Transfer

Transfer implements a location-agnostic data substrate that enables data to be accessed, shared, and moved among disparate storage systems, including at instruments, supercomputers, and on data services. Globus Transfer allows users and applications to modify data access permissions on remote storage systems and to move data reliably and securely between systems via a single API.

6.6 Globus Auth

Auth Auth allows for users to delegate permissions for clients to access services in the Gladier architecture, and for services to access other services on their behalf as well. For example, it allows Globus Flows to manage Globus Transfers and to execute analyses via funcX on systems accessible only to the user.

CHAPTER

SEVEN

CHANGELOG

All notable changes to this project will be documented in this file. See standard-version for commit guidelines.

7.1 0.8.0 (2023-01-05)

7.1.1 BREAKING CHANGES

- Requires a new login after upgrading
- Passing Authorizers to Gladier Clients has been deprecated in favor of using the new Login Manager system
- Passing `auto_login` is deprecated and will be removed
  - Disabling automatic login can be replicated by using a login manager with `CallbackLoginManager(..., callback=None)`. See Cusomizing Auth for more details.
- Gladier "public" configs have been removed
  - public configs were undocumented and shouldn't affect any normal Gladier users

7.1.2 Features

- Add "flow_transition_states" to BaseTools for determining Choice state (7053e75)
- Add login customization for use within larger apps (39c0a0c)
- Add support for python 3.10 (eaf3fec)
- Allow passing custom flow managers to Gladier Clients (a2fdead)
- Make the flows manager available for public usage (b905e7c)
- Login Manager and config overhauls (6016abc)
- Update Client ID from an older version (8b4393c)

7.2 0.8.0b2 (2022-11-08)

7.2.1 Features

- Add login customization for use within larger apps (39c0a0c)
- Allow passing custom flow managers to Gladier Clients (a2fdead)
- Make the flows manager available for public usage (b905e7c)

7.3 0.8.0b1 (2022-10-26)

7.3.1 BREAKING CHANGES

- Requires a new login after upgrading.
- Gladier "public" configs have been removed.

7.3.2 Features

- Support for writing tools using Flow Choice states.
- Add "flow_transition_states" to BaseTools for determining Choice state (7053e75)
- Add support for python 3.10 (eaf3fec)
- Login Manager and config overhauls (6016abc)
- Update Client ID from an older version (8b4393c)

7.3.3 0.7.1 (2022-08-25)

7.3.4 Bug Fixes

- Error on first time flow deployment if using group perms (b444f2b)

7.4 0.7.0 (2022-08-22)

7.4.1 BREAKING CHANGES

- Older funcx versions before v1.0 are no longer supported.
  - No code changes are required to migrate to Gladier v0.7.0 or FuncX v1.0

7.4.2 Features

- Add support for globus-automate-client 0.16 (dc9e82c)
- Upgrade to funcx v1 (c947077)

7.4.3 Bug Fixes

- Aliases not working with tools that use `@generate_flow_definition` (da30756)

7.4.4 0.6.3 (2022-07-20)

7.4.5 Bug Fixes

- Fix error when docstring is too long (a19a4bf)

7.4.6 0.6.2 (2022-05-06)

7.4.7 Bug Fixes

- Possible flows client 401 due to client caching (3b5a307)

7.4.8 0.6.1 (2022-05-05)

7.4.9 Features

- Add support for globus-automate-client 0.15.x (9ecd2e1)

7.5 0.6.0 (2022-05-05)

7.5.1 BREAKING CHANGES

- The following versions of FuncX and the Globus Automate Client will no longer be supported:
- Globus Automate Client: Requires 0.13.0 and above
- FuncX: Requires 0.3.6 and above

Older versions of these packages are only compatible with Globus SDK v2, and require updating any code that relies on the older Globus SDK version. See the SDK upgrade guide here:

7.5.2 Features

- Added 'aliased' tool chaining feature (48cbaaf)

7.5.3 Bug Fixes

- Add packaging dependency (95f6ec1)
- Added check when adding tool without flow states (c612383)
- Redeploy flow on 404s (6c0d3a7)
- tools/flows client not properly being cached in gladier clients (a6e5cec)
- Drop support for old versions of globus-automate-client/funcx (5d17d96), closes /globus-sdk-python.readthedocs.io/en/stable/upgrading.html#from-1-x-or-2-x-to-3-0

7.6 0.6.0b2 (2022-02-17)

7.6.1 Bug Fixes

- Add packaging dependency (95f6ec1)

7.7 0.6.0b1 (2022-02-17)

7.7.1 BREAKING CHANGES

- The following versions of FuncX and the Globus Automate Client will no longer be supported:
- Globus Automate Client: Requires 0.13.0 and above
- FuncX: Requires 0.3.6 and above

Older versions of these packages are only compatible with Globus SDK v2, and require updating any code that relies on the older Globus SDK version. See the SDK upgrade guide here:

7.7.2 Features

- Added 'aliased' tool chaining feature (48cbaaf)

7.7.3 Bug Fixes

- Added check when adding tool without flow states (c612383)
- Redeploy flow on 404s (6c0d3a7)
- tools/flows client not properly being cached in gladier clients (a6e5cec)
- Drop support for old versions of globus-automate-client/funcx (5d17d96), closes /globus-sdk-python.readthedocs.io/en/stable/upgrading.html#from-1-x-or-2-x-to-3-0

7.7.4 0.5.4 (2021-11-15)

7.7.5 Bug Fixes

- Only apply migrations when needed (670daea)

7.7.6 0.5.3 (2021-09-14)

7.7.7 Bug Fixes

- Limits run label length to 64 chars (53cd20f), closes #146

7.7.8 0.5.2 (2021-08-23)

7.7.9 Features

- Expanded flow modifiers to accept all top level state fields (3b3135f)

7.7.10 Bug Fixes

- Flow Modifier errors not propagating Client or Tool names (1af9724)
- Remove funcx-endpoint version check (8fe88a3)

7.7.11 0.5.1 (2021-08-19)

7.7.12 Bug Fixes

- Deploying new flows with the latest version of the flows service (afa5adf)

7.8 0.5.0 (2021-08-05)

7.8.1 BREAKING CHANGES

- Removal of older introductory testing tools

7.8.2 Bug Fixes

- logout not properly clearing authorizers cache (05b0d6c)
- Pass an 'empty' schema by default to fulfill automate requirement (bf9eb80)
- pin automate version to avoid future incompatible releases (b736fa2)
- Raise better exception when no flow definition set on tool (eb8ac03)
- Remove old "Hello World" tools. We have better ones now. (3a889f9)

7.8.3 0.4.1 (2021-07-20)

7.8.4 Features

- Added get_run_url for fetching the link to a running flow in the Globus webapp
- Arguments to run_flow support pass through args to the flows service

7.9 0.4.0 (2021-07-19)

7.9.1 BREAKING CHANGES

- – This will break all current funcx functions without modification. Everyone will need to upgrade to the new funcX endpoint package wherever they are executing functions. See the full migration doc in Migrating to V0.4.0

7.9.2 Features

- Upgrade to FuncX 0.2.3 (from 0.0.5) (83507f7)

7.9.3 0.3.5 (2021-07-14)

7.9.4 Features

- Added config migration system to Gladier (cdc7875)

7.9.5 0.3.4 (2021-07-09)

7.9.6 Bug Fixes

- gladier improperly falling back onto FuncX authorizers (3976a3a)

7.9.7 0.3.3 (2021-06-18)

7.9.8 Bug Fixes

- Tools with more than two states would raise error with flow gen (dd6586e)
- when user adds new AP to flow, Gladier now handles re-auth (e24a372)

7.9.9 0.3.2 (2021-06-17)

7.9.10 Bug Fixes

- add funcx-endpoint==0.0.3 to Gladier requirements (8ac47b8)

7.9.11 0.3.1 (2021-06-04)

7.9.12 Bug Fixes

- Fixed bug when instantiating two Gladier Clients (3aca2fe)

7.9.13 0.3.0 (2021-05-28)

7.9.14 Features

- Support flow modifiers, payload dict modifiers (a05b77a)
- FuncX ids and Flow ids are now saved in ~/.gladier-secrets.cfg instead of ./gladier.cfg
- Added support for setting Groups
- Added support for setting subscription ids

7.9.15 Bug Fixes

- Added changes lost from previous merges to fix client (c707d32)

7.9.16 0.2.0 - May 17, 2021

- Changed name `gladier.defaults.GladierDefaults` to `gladier.base.GladierBaseTool`
- Changed name `gladier.client.GladierClient` to `gladier.client.GladierBaseClient`
- Added a lot more documentation to the read-the-docs page!

7.9.17 0.0.1 - Apr 5, 2021

- Initial Release!

CHAPTER

EIGHT

INDICES AND TABLES

- genindex
- modindex
- search

INDEX