Research Automation with Globus Flows

Ada Nikolaidis
ada@globus.org
Automation with Globus Flows

• **Intro**
  – What is it?
  – Key concepts
  – How to use flows

• **Dive Deeper**
  – Automating Instrument Data at Scale (tomorrow)
  – Q&A
  – Office Hours
What is Globus Flows?

Flows Service

• A platform for managed, secure, reliable task orchestration
• Create, share, and run “flows”
What’s a Flow?

A sequence of steps...

- Hosted
- Reusable
- Flexible
- Shareable
How Do You Use a Flow?

Each time a flow is started, it’s called a *run*

- Start from Web App, CLI, Python SDK, API
- Provide input
- Performs a series of actions with that input
- Manage the run (evaluate, share)
What’s an Action?

An *action* is a special type of state

- AWS Step Functions
- “States Language”
- Basic states like “Wait”, “Pass”, “Choice”
What’s an Action?

...So what’s an *action*?

- Operations you can perform on other resources and services
- *Managed* interactions
  - Authenticate, authorize, validate, store, auto-retry
Globus-Provided Actions

- transfer
- delete
- mkdir
- ls
- stat
- identifier
- ingest
- delete
- notify
- input
- ACLs
- compute
Examples of Flows

• **Two Stage Globus Transfer**
  – Transfer to temp location, then transfer to final destination

• **Move (Copy and Delete)**
  – Transfer to a destination and delete files on the source

• **Transfer and Set Permissions (coming soon)**
  – Transfer to a destination and share with users and groups
Demo: Running flows
What happens when you run a flow?
Is this user allowed to run this flow?

Check input using flow's input schema

Register the run with the specified settings

Start the first (or next) action

...wait for the action to finish...

Is there another action after this?

The run has finished
Most often: A mapped collection
Notification: Runner receives an email
To grant permission: “resume”

Tip #1: Resume with the same application that started the run
Tip #2: Use guest collections when possible
Anatomy of a Flow

Definition
- Operations
- Transformations
- Conditions

Input Schema
- Data requirements

Three key benefits:
1. Prevent errors
2. Guided input
3. Flow validation
Anatomy of an Flow Definition

```
{
    "StartAt": "MyTransfer",
    "Comment": "A simple file transfer flow using the Globus Transfer Action",
    "States": {
        "MyTransfer": {
            "Type": "Action",
            "ActionUrl": "https://transfer.actions.globus.org/transfer",
            "Parameters": {
                "source_endpoint": "6c54cade-bde5-45c1-bdea-f4bd71dba2cc",
                "destination_endpoint": "31ce9ba0-176d-45a5-add3-f37d233ba47d",
                "DATA": [
                    {
                        "source_path.$": "$source_path",
                        "destination_path.$": "$destination_path"
                    }
                ],
                "ResultPath": "$MyTransferResult",
                "WaitTime": 300,
                "End": true
            }
        }
    }
}
```
Anatomy of an Action

Here is a code snippet that demonstrates a simple file transfer flow using the Globus Transfer Action:

```json
"StartAt": "MyTransfer",
"Comment": "A simple file transfer flow using the Globus Transfer Action",
"States": {
  "MyTransfer": {
    "Type": "Action",
    "ActionUrl": "https://transfer.actions.globus.org/transfer",
    "Parameters": {
      "source_endpoint": "6c54cade-bde5-45c1-bdea-f4bd71db2cc",
      "destination_endpoint": "31ce9ba0-176d-45a5-add3-f37d233b47d",
      "DATA": [
        {
          "source_path.$": ".source_path",
          "destination_path.$": ".destination_path"
        }
      ],
      "ResultPath": "$.MyTransferResult",
      "WaitTime": 300,
      "End": true
    }
  }
}
```
Flows at Scale

Command Line Interface
Ad hoc scripting and integration

Python SDK
Applications, Services, Portals
New: Enhanced Validation

Enhanced Validation (Beta)

• Available in: Globus CLI, Python SDK, API

• Catches errors that were previously impossible:
  – Invalid references to input
  – Invalid parameters to actions
  – Conflicts in state type
Concepts

<table>
<thead>
<tr>
<th>Flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run</td>
</tr>
<tr>
<td>Action</td>
</tr>
<tr>
<td>Action Provider</td>
</tr>
</tbody>
</table>
Q&A
Thank You!

Don’t Forget…

• Advanced session: *tomorrow*
  – Automating Instrument Data at Scale

• Office Hours: *today and tomorrow*