

Data Staging with Globus

Students:

Blake Ehrenbeck

Zhen Huang

Education: BS 2019, Computer Science, Illinois Institute of Technology

Supervisors:

Bill Allcock

Lisa Childers

Paul Rich

Goal:

- Enable users of Cobalt to submit jobs with *qsub* (the command that submits jobs to the scheduler) via a REST API
- For jobs requiring data staging, enable data movement to be orchestrated at runtime without user intervention
 - Why?
 - Current Data Transfers need explicit user intervention
 - Jobs may sit in the queue longer than a user is willing to wait to facilitate transfers

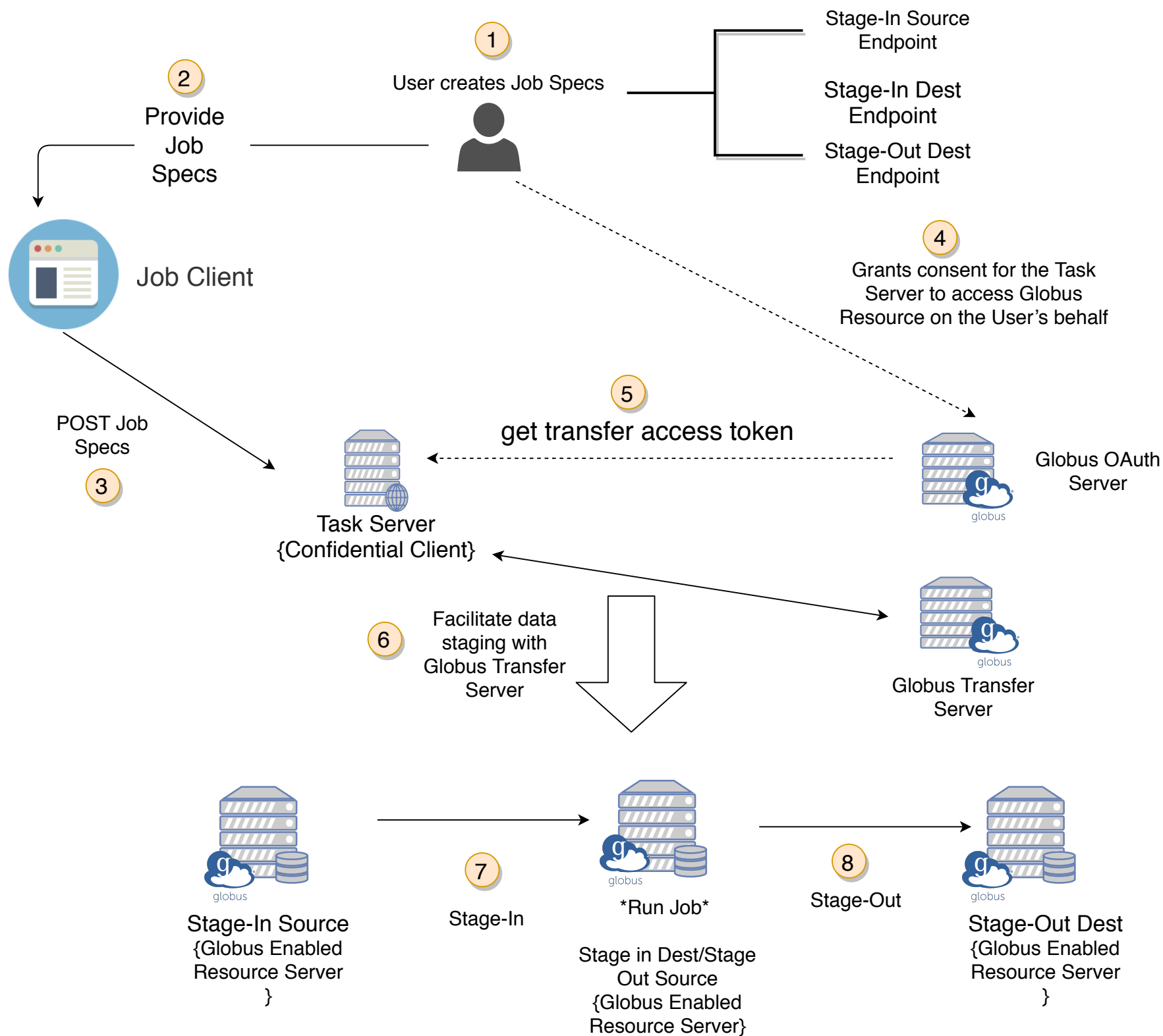
Approach:

- Enable users provide job specs via REST API
- Using Globus resource to staging data on user behalf
- Refresh the access token if it expired during prolonged queue wait time

Three Components

- **Globus OAuth and Transfer Resource Servers** (for Data Staging)
 - Third party login (Google, ALCF)
 - Authorized endpoints file transferring
- **Job Client** (Get user's job specs)
 - Mocks a barebones job submission interface
 - Accepts three data staging endpoints (all Globus-enabled)
- **Task Server** (*Task Server* REST API)
 - Handles data staging with inputs from a client (we used *Job Client*)
 - After getting consent via OAuth it's able to move data on behalf of the user
 - Designed to be used by any client able to make REST calls
 - To be integrated into the Cobalt simulator

[Diagrammed on the next page.]



Demo

Future Work

There are still three more weeks before we leave. So our next steps are to integrate the Task Server API into the Cobalt Scheduler simulator. The new job client will be *qsub*.

Questions?

Job 1 submission from job client:

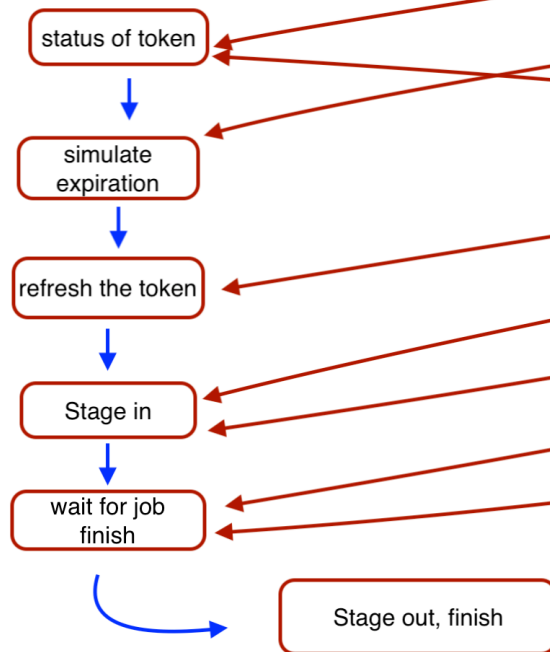
The screenshot shows the 'Transfer Files | Globus' window. The 'Job Client' section contains three rows of input fields. The first row is for 'Stage In Source' with the value 'ddb59aef-6d04-11e5-ba46-22000b92c6ec||~/music/'. The second row is for 'Stage In Destination' with the value '097b8756-83fd-11e8-9536-0a6d4e044368||~/Desktop/Testing/music'. The third row is for 'Stage Out Destination' with the value 'ddb59af0-6d04-11e5-ba46-22000b92c6ec||~/music'. Yellow circles highlight the endpoint IDs and data paths in each row. Yellow arrows point from the labels 'Endpoint ID' and 'Data path' to the corresponding parts of the input strings. A blue 'Submit' button is located at the bottom right.

Label	Value
Stage In Source	ddb59aef-6d04-11e5-ba46-22000b92c6ec ~/music/
Stage In Destination	097b8756-83fd-11e8-9536-0a6d4e044368 ~/Desktop/Testing/music
Stage Out Destination	ddb59af0-6d04-11e5-ba46-22000b92c6ec ~/music

Submit

Task Server (single job):

UUID	Runtime	Status
30b4e148-8e91-11e8-858d-a45e60d8927d	25.63124442299886	SUCCESS



Toggle Console

10:58:07 AM» <30b4e148-8e91-11e8-858d-a45e60d8927d> is submitted with wait time : 9 seconds **Step 0**

10:58:14 AM» <30b4e148-8e91-11e8-858d-a45e60d8927d> The transfer access token now is active after user login. **Step 1**

10:58:14 AM» <30b4e148-8e91-11e8-858d-a45e60d8927d> Now revoking the transfer access token to simulate token expiration due to long queue wait **Step 1**

10:58:14 AM» <30b4e148-8e91-11e8-858d-a45e60d8927d> The transfer access token is now inactive. **Step 1**

10:58:17 AM» TaskID:<30b4e148-8e91-11e8-858d-a45e60d8927d> The transfer access token now is active after using refresh_token **Step 1**

10:58:18 AM» <30b4e148-8e91-11e8-858d-a45e60d8927d>Queue wait is done, now initiating Stage in.... **Step 2**

10:58:28 AM» <30b4e148-8e91-11e8-858d-a45e60d8927d> Stage In succeeded **Step 2**

10:58:28 AM» <30b4e148-8e91-11e8-858d-a45e60d8927d>Running the job **Step 3**

10:58:31 AM» <30b4e148-8e91-11e8-858d-a45e60d8927d>Job is done **Step 3**

10:58:31 AM» <30b4e148-8e91-11e8-858d-a45e60d8927d> Initiating Stage out.... **Step 4**

10:58:42 AM» <30b4e148-8e91-11e8-858d-a45e60d8927d> Stage Out succeeded **Step 4**

- Step 0** Job Submission
- Step 1** Credentials
- Step 2** Data stage in
- Step 3** Running job
- Step 4** Data stage out

Job 2:

<

>

localhost:8080

Transfer Files | Globus

localhost:8080

Job Client

Stage In Source

ddb59aef-6d04-11e5-ba46-22000b92c6ec|~/movie/

Stage In Destination

097b8756-83fd-11e8-9536-0a6d4e044368|~/Desktop/Testing/movie

Stage Out Destination

ddb59af0-6d04-11e5-ba46-22000b92c6ec|~/movie

Submit Tra

log from

(log for multi-jobs)

localhost

localhost:8080/status?taskId=17bda0cc-8e8e-11e8-96f3-a45e60d89...

localhost:8080/status?taskId=1b3772fa-8e8e-11e8-b467-a45e60d...

Transfer Files | Globus

Status

Job Submitted! Job ID is : 17bda0cc-8e8e-11e8-96f3-a45e60d8927d with a delay of 8 seconds.

UUID	Runtime	Status
17bda0cc-8e8e-11e8-96f3-a45e60d8927d	null	
1b3772fa-8e8e-11e8-b467-a45e60d8927d	null	

Step 0

Job Submission

Step 1

Credentials

Step 2

Data stage in

Step 3

Running job

Step 4

Data stage out

Toggle Console

10:35:57 AM» User logged in... Step 0

10:35:57 AM» <17bda0cc-8e8e-11e8-96f3-a45e60d8927d> is submitted with wait time : 8 seconds Step 0

10:36:03 AM» <17bda0cc-8e8e-11e8-96f3-a45e60d8927d> The transfer access token now is active after user login. Step 1

10:36:03 AM» <17bda0cc-8e8e-11e8-96f3-a45e60d8927d> Now revoking the transfer access token to simulate token expiration due to long queue wait Step 1

10:36:03 AM» <17bda0cc-8e8e-11e8-96f3-a45e60d8927d> The transfer access token is now inactive. Step 1

10:36:07 AM» TaskID:<17bda0cc-8e8e-11e8-96f3-a45e60d8927d> The transfer access token now is active after using refresh_token Step 1

10:36:08 AM» <17bda0cc-8e8e-11e8-96f3-a45e60d8927d> Queue wait is done, now initiating Stage in.... Step 2

10:36:09 AM» <1b3772fa-8e8e-11e8-b467-a45e60d8927d> The transfer access token now is active after user login. Step 1

10:36:09 AM» <1b3772fa-8e8e-11e8-b467-a45e60d8927d> Now revoking the transfer access token to simulate token expiration due to long queue wait Step 1

10:36:09 AM» <1b3772fa-8e8e-11e8-b467-a45e60d8927d> The transfer access token is now inactive. Step 1

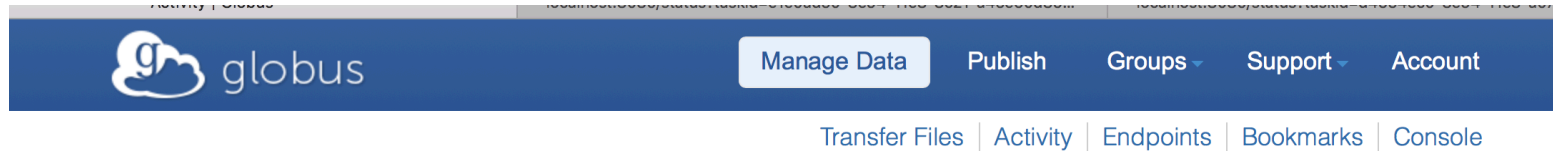
10:36:13 AM» TaskID:<1b3772fa-8e8e-11e8-b467-a45e60d8927d> The transfer access token now is active after using refresh_token Step 1

10:36:14 AM» <1b3772fa-8e8e-11e8-b467-a45e60d8927d> Queue wait is

Job 1

Job 2

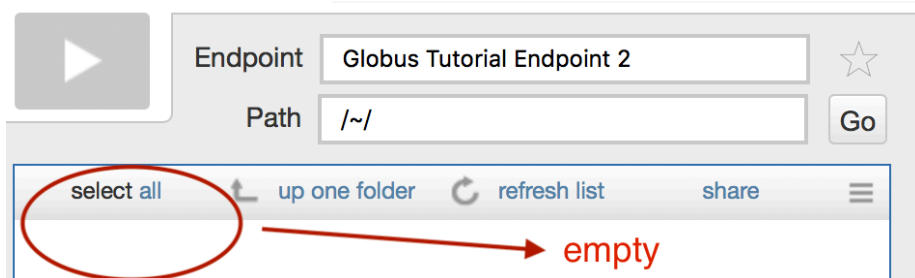
Success!



Activity



before:



finished

