SDSC Summer Institute 2015: Workflow Management

Hands-on Exercises 2: Searching, Moving and Pre-processing Data

Goals: Search, move and pre-process data via Kepler

Goal 1: Find data using a REST service.

1	Start Kepler
2	Search and drag *RESTService* actor to the canvas.
3	Configure the actor's parameters. Set *serviceSiteURL* to be http://www.biomart.org/biomart/martservice and set *methodType* to be Get
4	Search and drag *SDF* director to the canvas.
5	Search and drag *String Constant* actor to the canvas. Configure its *value* parameter to be : type=datasets,mart=ensembl
6	Connect the output port of *String Constant* actor to the *paramInputPort* input port of the *RESTService* actor.
7	Search and drag *Display* actor to the canvas.
8	Connect the output port of the *RESTService* actor to the input port of the *Display* actor.
9	Run the workflow to see the search results.

Goal 2: Move data between computers using Secure Copy command.

1	Open a new blank workflow canvas.
2	Add *SSH File Copier* actor to canvas. Configure its *recursive* parameter to be true by checking the box.
3	Search, drag and drop three "StringParameter" actors to canvas.
4	Right click on one parameter, select "Customize Name", and customize its name to be TargetHost.
5	Double click on the parameter and set its value to be LOGIN@gordon.sdsc.edu, where LOGIN should be replaced with the login id assigned to you.
6	Repeat the above two steps to customize name of the second parameter to be "inputFolder", and set its value to be $HOME/GPU_test
7	Similarly, customize name of the third parameter to be "remoteDir", and set its value to be /home/LOGIN, where LOGIN is the login id assigned to you.
8	Search and drag *SDF* director to the canvas.
9	Search and drag two *String Constant* actors to the canvas.
10	Change the name of one of the *String Constant* actors to "inputFolder": Right-click on the actor, select Customize Name, then type "inputFolder". Then double click on the actor and configure its *value* parameter to be $inputFolder
11	Change the name of the other *String Constant* actor to "remoteDir". Then double click and configure its *value* parameter to be $TargetHost:$remoteDir
12	Search and drag *Display* actor to the canvas.
13	Connect the *succ* port of the *SSH File Copier* actor to the input port of *Display* actor. You may also connect a second Display actor to the *error* port of SSH File Copier, if you wish.
14	Connect the output port of the *inputFolder* actor to the *source* port of the *SSH File Copier* actor.
15	Connect the output of the *remoteDir* actor to the *target* port of the *SSH File Copier* actor.
16	Run the workflow. Enter the password when prompted.