Hands-on 2: Searching and Moving Data

Shweta Purawat

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

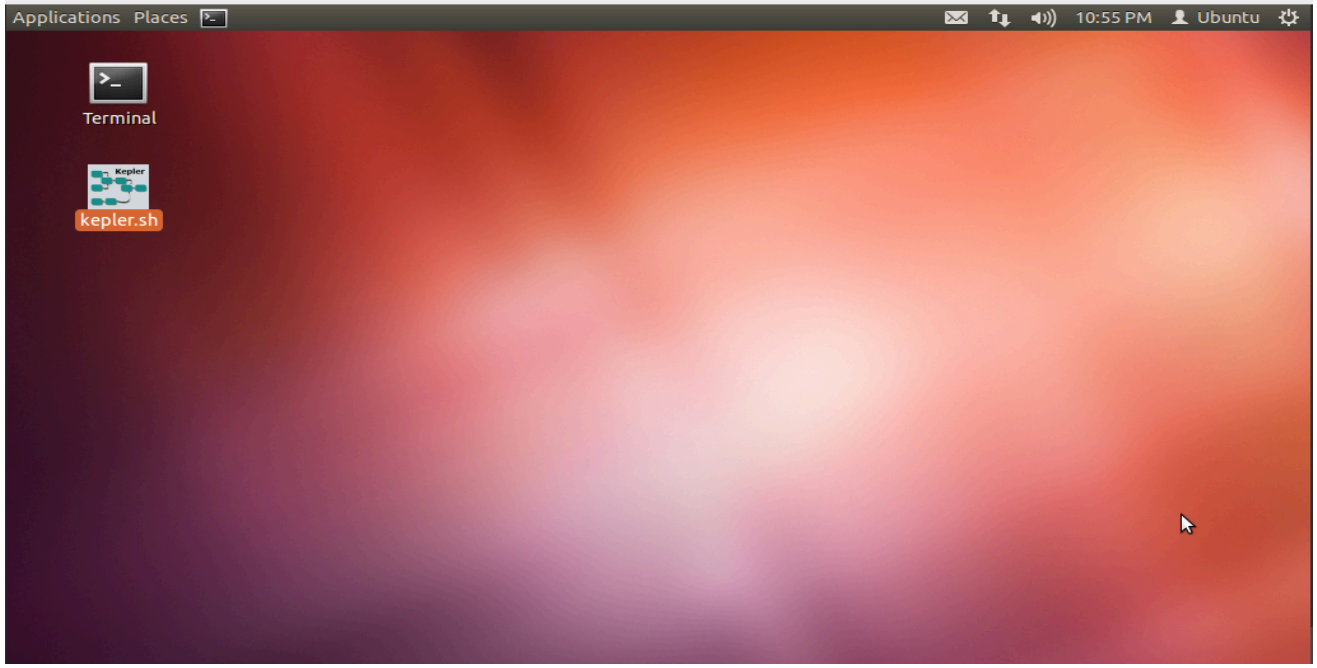
WorDS Center

Hands-On Examples

1. Finding data
2. Moving data

bioKepler Virtual Box Instance

- You should see the following screen



- Double click on the Kepler icon to start Kepler

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

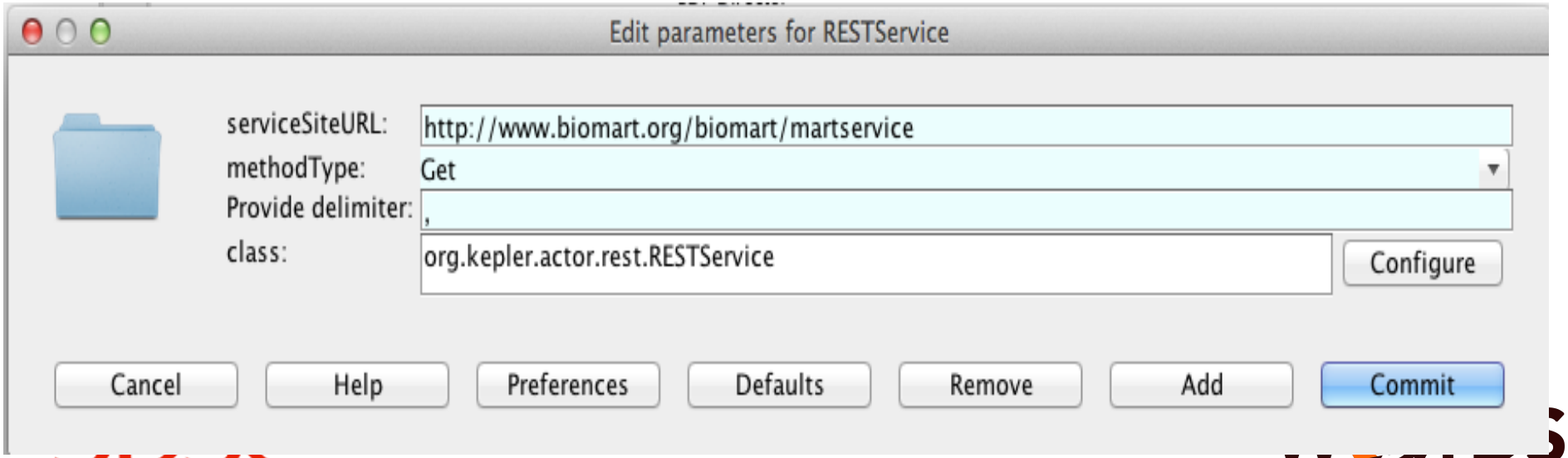
WorDS Center

1. Finding Data using a REST Service

Goal: Create a workflow to fetch the datasets information for a particular type from biomart using their REST Service

Step 1: Add RESTService Actor

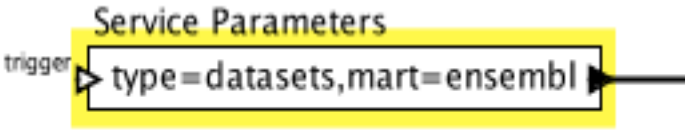
- Search and drag RESTService actor on the canvas
- Configure the actor:
  - **serviceSiteURL:** http://www.biomart.org/biomart/martservice
  - **methodType:** Get

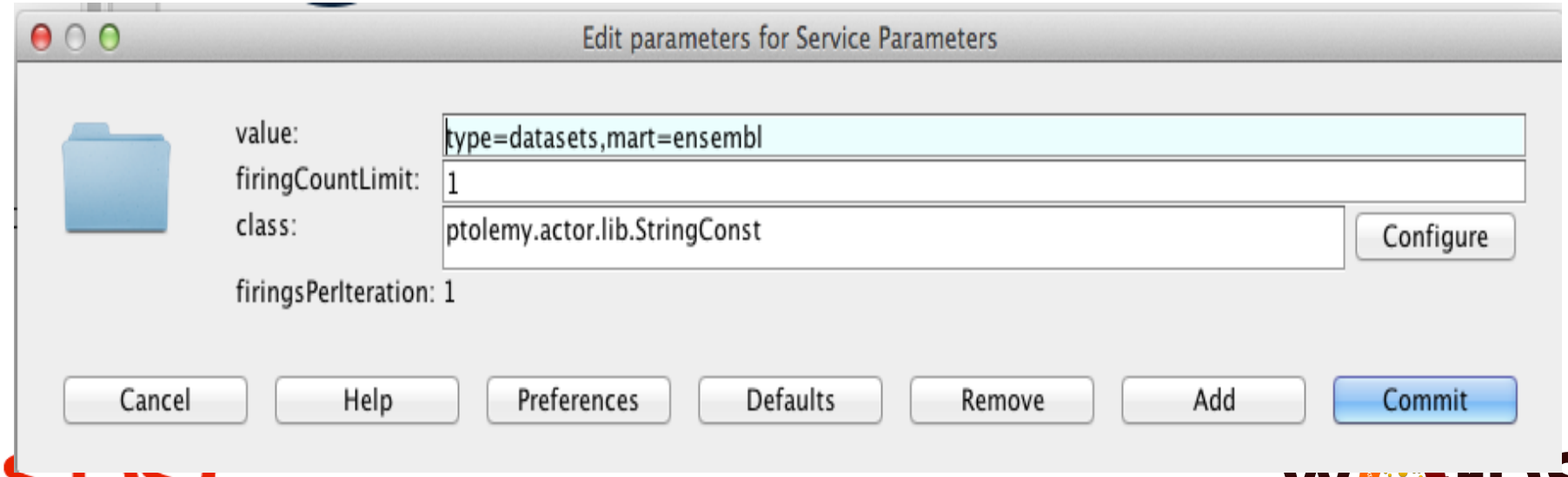


Step 2: Add Director and String Constants

- Add SDF Director
- Add Input actor
  - Add String Constant actor
  - configure

  Value:type=datasets,mart=ensembl

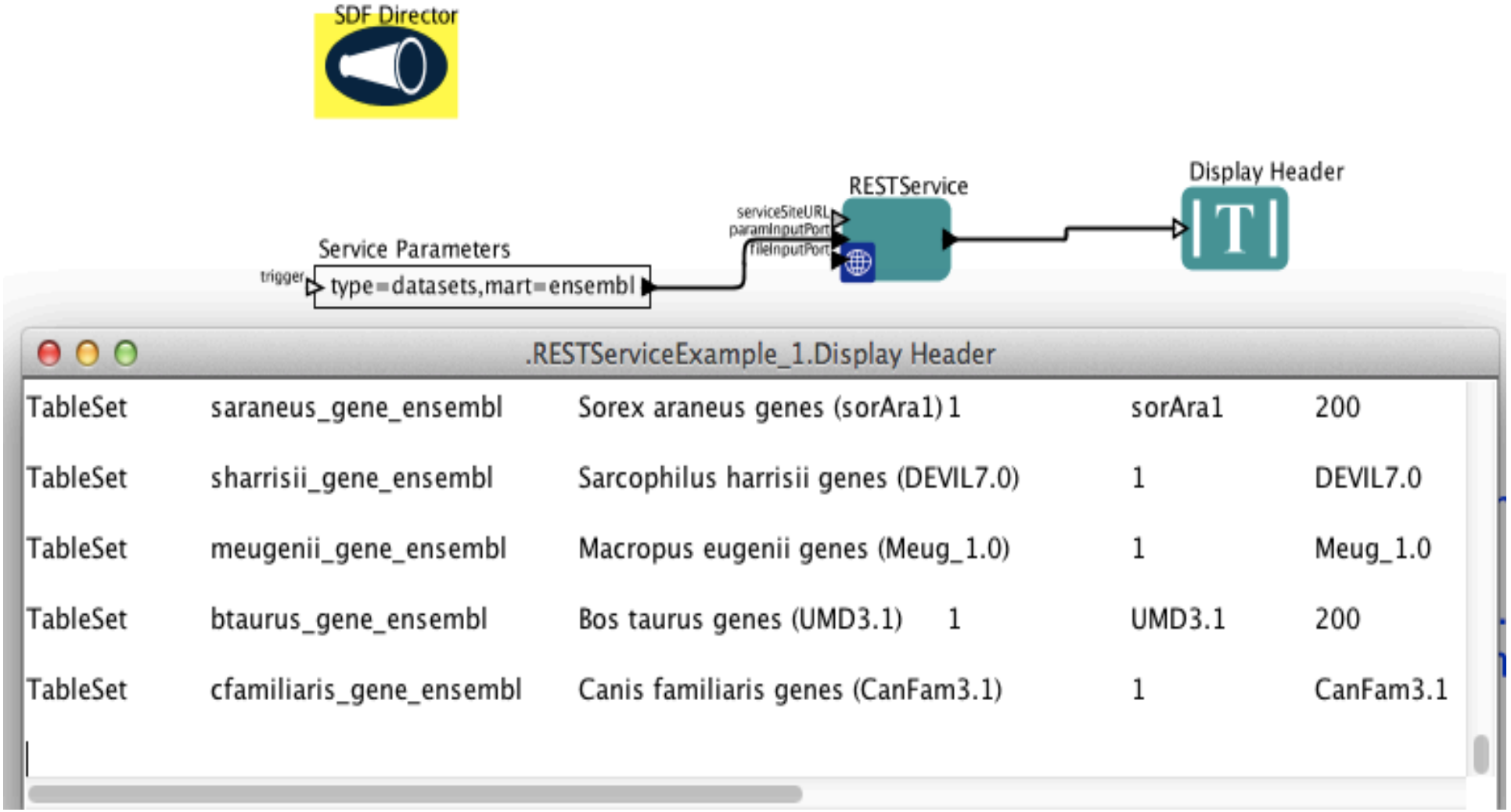




Step 3: Add Display actor

- Add Display
- Connect input and output actors to the RESTService actor

Step 4: Run Finished Workflow



2. Moving Data using Secure Copy Command

Goal: Create a workflow to securely copy data files from local machine to remote GPU cluster

scp –r $inputFolder $TargetHost: remoteDir

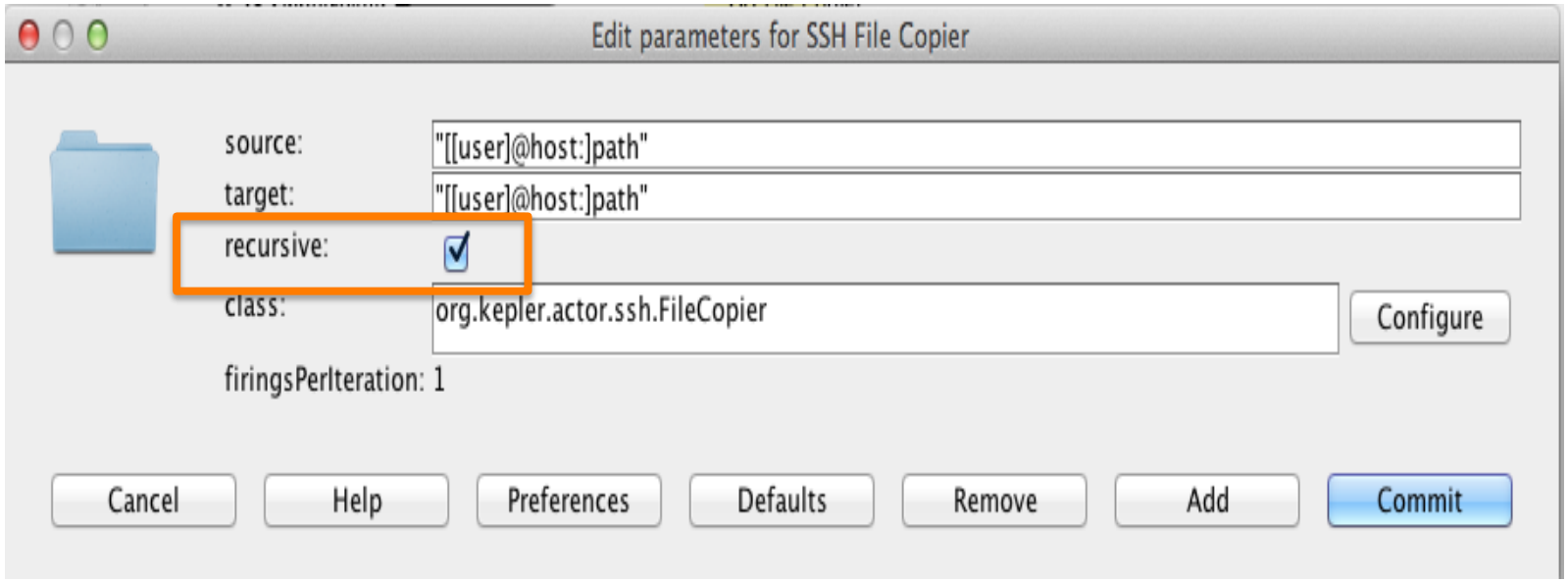
Source TargetHost Destination

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

WorDS Center

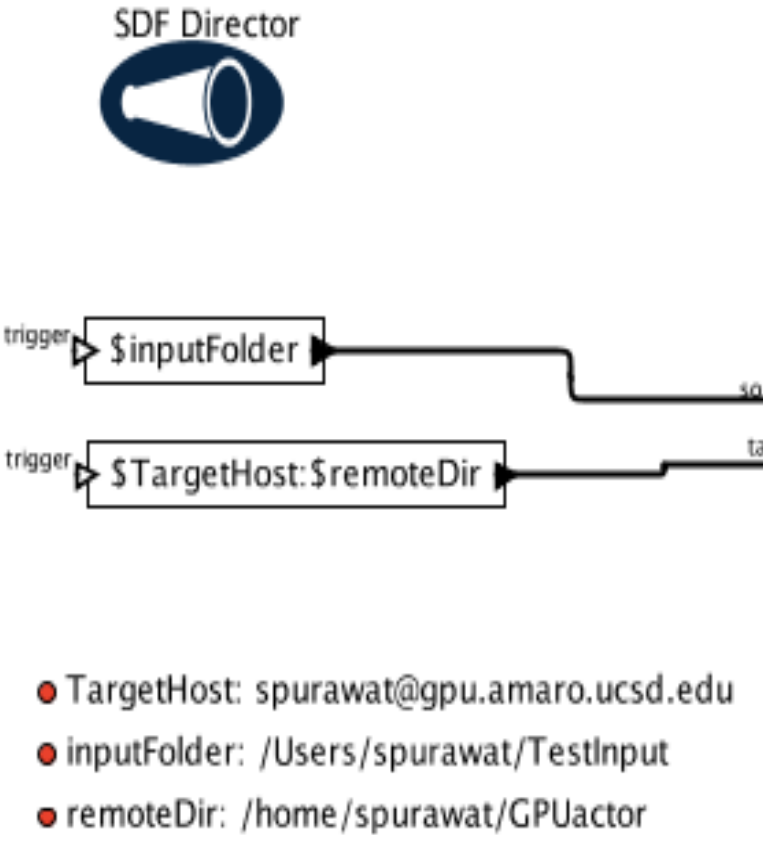
Step 1: Add SSH File Copier

- Add SSH File Copier actor
- Configuration: Set recursive true



Step 2: Add Parameters

- Add Parameters:
  - Search, drag and drop three "StringParameter" onto workflow canvas.
  - Right Click on a parameter and Customize Name: **TargetHost**
  - Double click on the “TargetHost” and set
    Value: **LOGIN@gordon.sdsc.edu**
  - Similarly customize name of other parameter
    inputFolder, value: **$HOME/GPU_test**
    remoteDir, Value: **/home/LOGIN**



Step 3: Add Director and String Constant Input actors

- Add SDF Director
- Add Input actors
  - Search and drag two String Constant actors
  - Configure: Double click and set input actor1

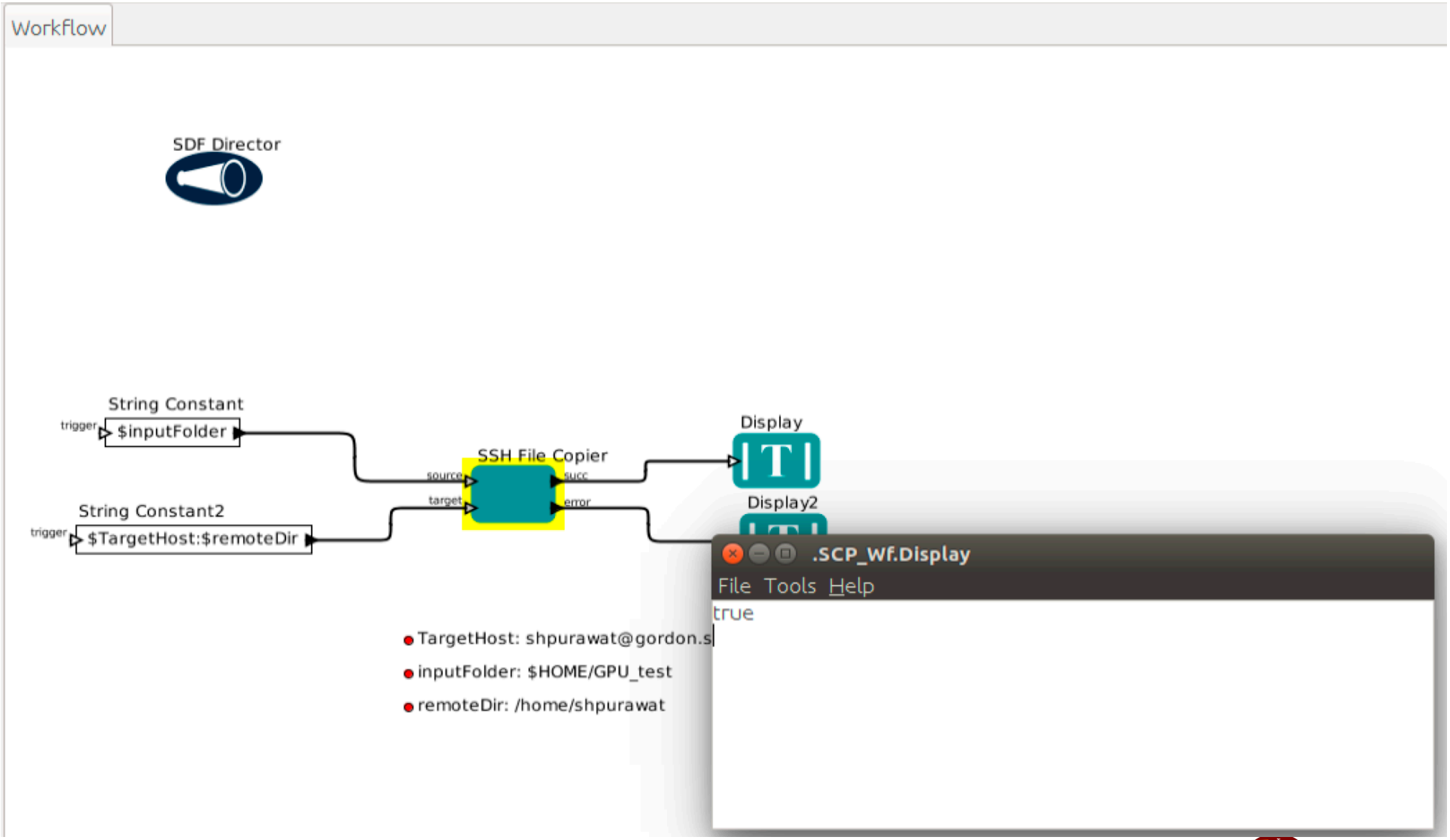
    value: **$inputFolder**
  - Configure: Double click and set input actor2

    value: **$TargetHost:$remoteDir**

Step 4: Display DataFile transfer success

- Add Display
- Connections:
  - Connect "**succ**" port of the SSH copier actor to input port of Display actor
  - Connect output of "inputFolder" actor to "**source**" port of the SSH copier actor
  - Connect output of "TargetHost:remoteDir" actor to "**target**" port of the SSH copier actor

Step 5: Run Finished Workflow



Next Up:

Analyzing Data Locally and Remotely

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

WorDS Center