

# Hands-on 2: Searching and Moving Data

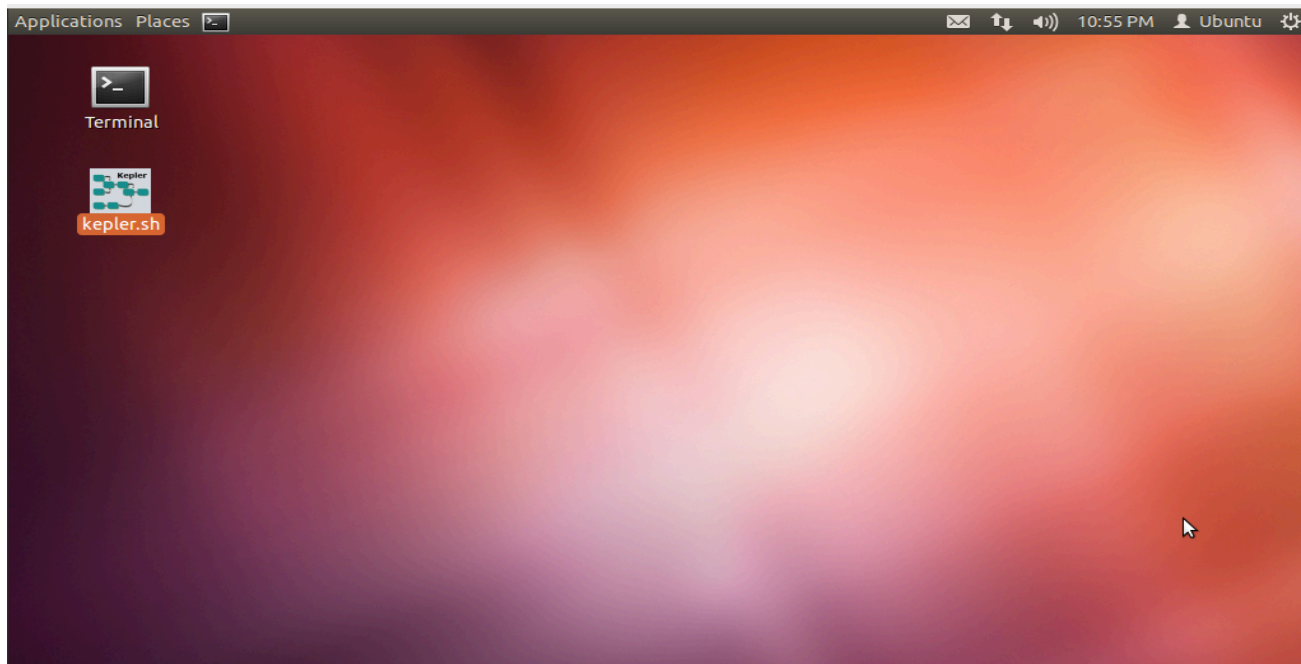
Shweta Purawat

# 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

# 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

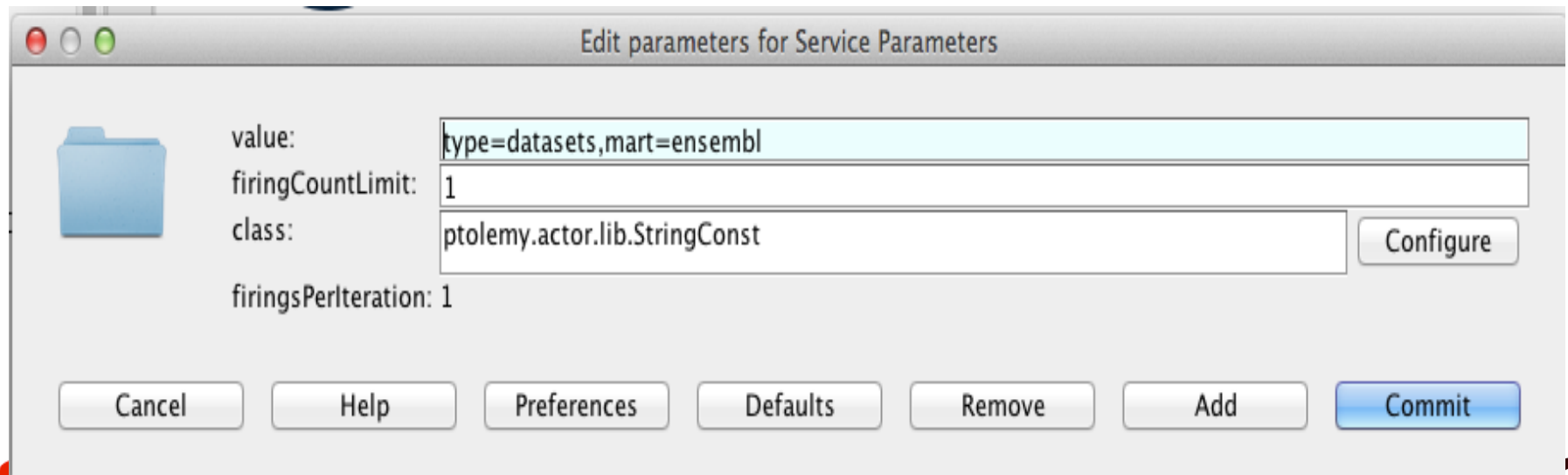
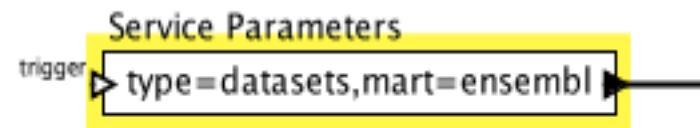
# Step 1: Add RESTService Actor

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

A screenshot of a software window titled "Edit parameters for RESTService". The window contains a folder icon on the left and four input fields on the right. The first field is labeled "serviceSiteURL:" and contains the text "http://www.biomart.org/biomart/martservice". The second field is labeled "methodType:" and contains the text "Get". The third field is labeled "Provide delimiter:" and contains a comma ",". The fourth field is labeled "class:" and contains the text "org.kepler.actor.rest.RESTService". To the right of the "class:" field is a "Configure" button. At the bottom of the window are several buttons: "Cancel", "Help", "Preferences", "Defaults", "Remove", "Add", and "Commit".

# 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



.RESTServiceExample_1.Display Header				
TableSet	saraneus_gene_ensembl	Sorex araneus genes (sorAra1)	1	sorAra1 200
TableSet	sharrisii_gene_ensembl	Sarcophilus harrisii genes (DEVIL7.0)	1	DEVIL7.0
TableSet	meugenii_gene_ensembl	Macropus eugenii genes (Meug_1.0)	1	Meug_1.0
TableSet	btaurus_gene_ensembl	Bos taurus genes (UMD3.1)	1	UMD3.1 200
TableSet	cfamiliaris_gene_ensembl	Canis familiaris genes (CanFam3.1)	1	CanFam3.1



## 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

# Step 1: Add SSH File Copier

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



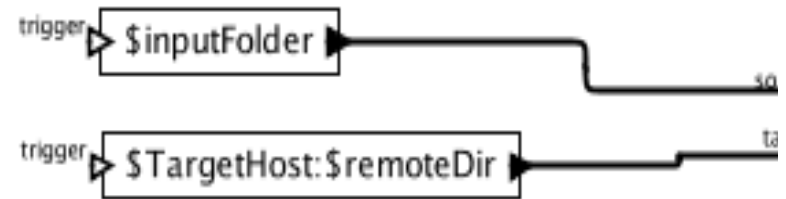
A screenshot of the 'Edit parameters for SSH File Copier' dialog box. The dialog has a folder icon on the left. The parameters are as follows:

- source: "[[user]@host:]path"
- target: "[[user]@host:]path"
- recursive: ☒ (This row is highlighted with an orange box)
- class: org.kepler.actor.ssh.FileCopier
- firingsPerIteration: 1

Buttons at the bottom include Cancel, Help, Preferences, Defaults, Remove, Add, and Commit. A 'Configure' button is next to the class field.

## 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**



- TargetHost: spurawat@gpu.amaro.ucsd.edu
- inputFolder: /Users/spurawat/TestInput
- remoteDir: /home/spurawat/GPUactor

## Step 3: Add Director and String Constant Input actors

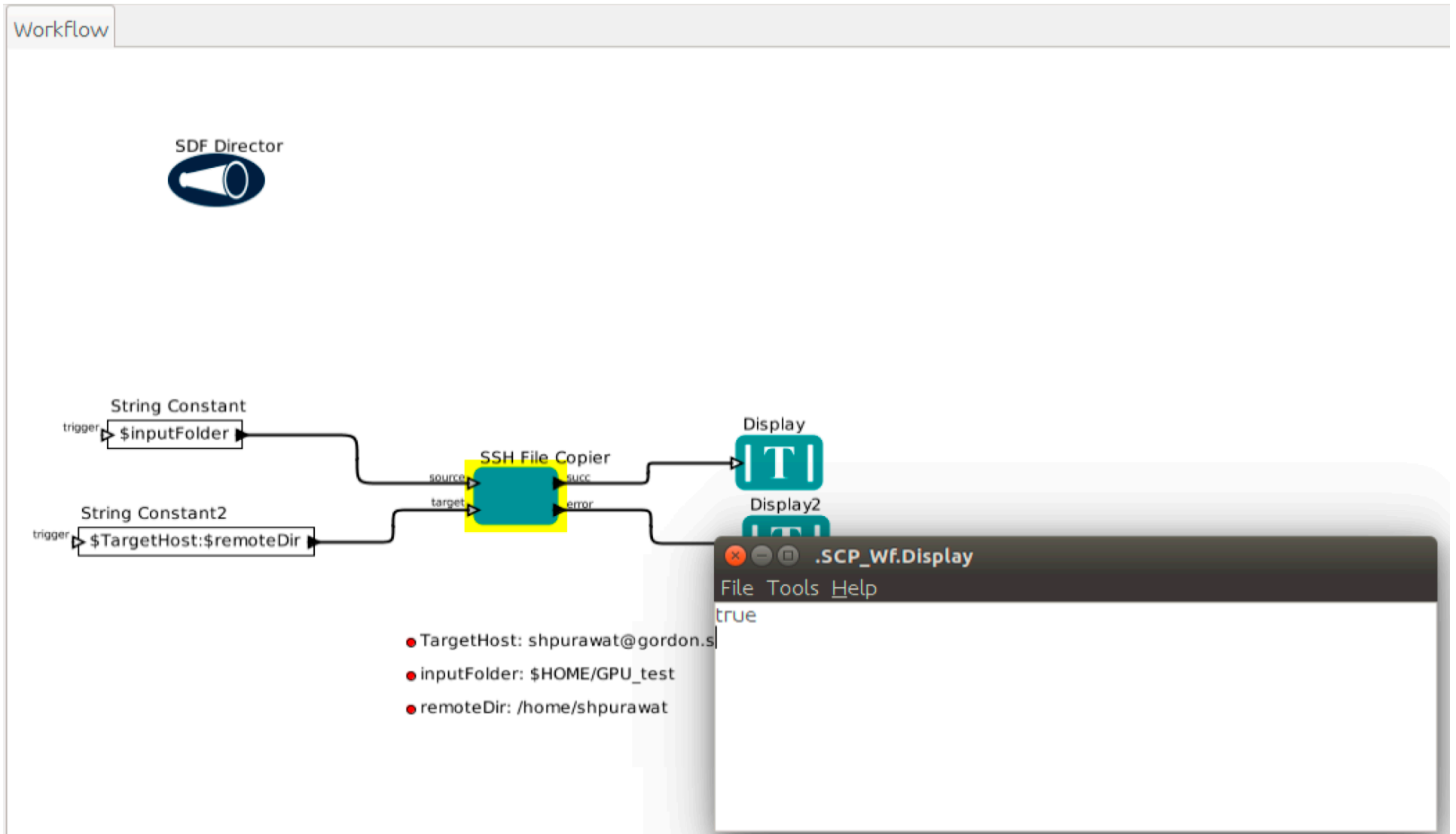
- Add SDF Director
- Add Input actors
  - Search and drag two String Constant actors
  - Configure: Double click and set input actor1 value: **\$inputFolder**
  - Configure: Double click and set input actor2 value: **\$TargetHost:\$remoteDir**

# 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