MERLIN ETL Process

Ingest dataset

* Log into 192.168.104.11 as **jreeme**/**password**
* Note: some terminals exit out after a specified period of inactivity which kills the ETL process. So, setup a screen session to start the ETL process.
  + [jreeme@build]$ **screen -ls** <lists existing screens>
  + **screen -r *<id>*** <if screen exists use this command>
  + **screen** <if none exist, use this command>
* **staging-smahoney**
* Note: before you can add a dataset, the data has to be pre-staged into a specific area (/mnt/merlin.merlin-etl/data/); see section "Stage data for Merlin ETL process".
* **add-dataset *<Agency> <Case> <Location of data file>***
  + Ex. add-dataset KeyW JEB /mnt/merlin/merlin-etl/data/jeb.tgz
* **run-etl *<id from console>***
* Note: the console spits out status messages on the progress of the ETL process, you can also see this at 192.168.104.7:8088/cluster.

Set permissions

* Someone with admin privileges has to go into MOSAIC and set which groups can see this dataset.
* Another method is to use Amino 192.168.104.7:3000 root/password
* Create anew Role if needed
* Create a new User if needed
* Asscoiate role with User
* Find the ingest id
  + Log into the Loopback API Explorer 192.168.104.7:3000/explorer/
  + Dataset
  + Select Get
  + Select Try it Out
  + Record datasetUID
* Add dataset id to the desired AminoRole
  + Select AminaRole
  + Select GET
  + Try it Out
  + Copy the role json
  + Select PUT
  + Paste copied
  + Add your new UID into the datsets array
  + Try it out!

Stage data for Merlin ETL process

* Log onto 192.168.104.44 with **jreeme**/**password** (jreeme@nfs-server)
* Use WinSCP to copy data to the /tmp directory or any other valid method to get data to desired location.
* **sudo cp /tmp/*<datafile.tar.gz>* /nfs-public/parrot-scif-sm/hadoop-datavolume/merlin-etl/data/**
* **chown root:root *<datafile.tar.gz>***

Update stack

* Log into build server 192.168.104.11 as **jreeme**/**password**
* **cd src/merlin-stack/firmament/deploy/vmware/vmware.parrot-scif.keyw**
* **f p b -i full-stack-thin-sm.json**