Access to MASS for Collaboration

Roger Milton & Martin Ouldridge

External Access to MASS

Contents

- Overview of MASS and MOOSE
- MOOSE data organisation
- Some example commands
- Getting access

What is MASS?

- Main Met Office resilient tape-based archiving system
- HPSS library-management system
- 2 automated tape-libraries, ~50 PB each
- Duplexed copies
- 1 PB cache
- Bespoke user-interface ‘MOOSE’
- Recent stats:
  - > 25 PB archived
  - ~40 TB written per day, ~60 TB accessed per day
  - ~50% of retrieves served from cache

MASS upgrade for new HPC

- Will store up to 600 PB of primary data by 2020
- Multiple automated tape libraries in 2 IT Halls
  - At least 240 tape drives
  - At least 55,000 tapes
- At least 6.5 PB cache by 2017

- 345 TB – 445 TB per day expected to be archived by 2017
- 150 TB per day expected to be restored by 2017
- First phase of the upgrade in production by October 2015

MOOSE – the MASS client

- Common interface for all users
- Logical separation from vendor-specific interface
- Physical buffering, enhancing availability
- User account management
- Access-control
- (Some) cataloguing
- Filtered retrieval for PP (UM Post Processing) and netCDF file-types
- Manages collocation of related data on physical media

Access for external collaborators

- External collaborators can:
  - Have access based on a MASS “project” basis
  - List contents of data sets
  - Retrieve files
  - Filter files
  - Get information about data sets:
    - Ownership, quality-assessments, comments etc.
- They cannot:
  - Store or overwrite files
  - Move, rename or delete files
  - Change metadata about files or data-sets

External Access to MASS

- MOOSE data organisation

MOOSE data organisation

- MOOSE is the client used to search and retrieve data from MASS
- Uses a directory hierarchy
- No fully searchable index exists
- Data is stored in **Classes**, **Sets** and **Collections**
- Data is accessed via MOOSE URIs (Uniform Resource Identifier) – like *file paths*
- MOOSE provides ownership and access permissions to the data
- MOOSE on JASMIN provides read-only access

An example MOOSE URI

- `moose:/crum/antie/apa.pp/<files>`

- `moose:` - indicates a MOOSE URI
  - can be abbreviated to "`moo:`" or just "`:`"

- `crum` – is the data *Class*
- `antie` - is the data *Set*
- `apa.pp` - is the data *Collection*
  - files are stored in the Collection

MOOSE data Classes

- `moose:/crum/antie/apa.pp`
- Each data Class controls the way in which data can be added to it, e.g. what are valid names for its sub-directories, and how many levels of sub-directory are allowed.
- Available Classes are:
  - `opfc` – operational forecasts
  - `devfc` – development forecasts
  - `crum` – Climate Research using the UM
  - `misc` - miscellaneous
  - `adhoc` – unstructured data (limited JASMIN access)

MOOSE data Sets

- `moose:/crum/antie/apa.pp`

- *Usually* corresponds to a UM run
- Each Set has an owner
- Sets can be associated with Projects
- MOOSE users on JASMIN are given access to Projects
- *Some* naming conventions for Sets exist:
  - E.g. `moose:/crum/x*` are climate model runs on MONSooN

MOOSE data Collections

- `moose:/crum/antie/apa.pp`

- Is a sub-directory which contains files
- The **suffix** indicates the type of files contained in the Collection
- Permitted suffixes are:
  - `.pp` – contains PP (UM Post Processing) files
  - `.nc.file` – contains netCDF files
  - `.file` – contains files

External Access to MASS

- Some example commands
  - Access to the MOOSE client happens through the command-line tool called “moo”.

First: login to JASMIN

```
roger@metoffice:> exec ssh-agent $SHELL
roger@metoffice:> ssh-add ~/.ssh/jasmin_id_rsa
Enter passphrase for
   /home/metoffice/roger/.ssh/jasmin_id_rsa:
Identity added: ...

roger@metoffice:> ssh -A rogerm@jasmin-login1.ceda.ac.uk
...
[rogerm@jasmin-login1 ~]$ ssh mass-cli1
```

Check system status

```
[rogerm@mass-cli1 ~]$ moo sysinfo
Controller: exxmobsslc01.metoffice.gov.uk
```

- you can use this to check the MASS system is available.

Help command (“moo help”)

```
$ moo help
Usage: moo <sub-command> [options] [arguments]

Met Office Operational Storage Environment (MOOSE) command-line client.
Type "moo help <SUB-COMMAND>" for help on a specific sub-command.

Available sub-commands (note that some sub-commands are not available
to some types of user and on some client-platforms):
  comment
        for creating, modifying, or retrieving comments applied to data sets and collections
  dls, displs, disposals
        List data which has been marked for deletion using the dispose command
  filter
        atomic retrieval (.nc files only)
  get
        retrieves files from the archive
  help, h, ?, --help, --usage
        prints help for this command or its applications
  help, h, ?, --help, --usage
        prints help for this command or its applications
  ls    ...
```

More detailed help

```
$ moo help ls
ls, list: prints a listing of directories/files in the archive
Usage: moo ls [URI ...]

Prints a listing of the files and directories in one or more
   locations
(URI) in the Moose archive.

Valid options:
  --access-time, --atime, -u
        displays time of last-access rather than archival
  --all, -a, -e
        prints all items
  --directory, -d
        displays information for the selected directory
  --help, -h
        prints help and exits
...
```

List my project memberships

```
$ moo projlist –long
```

```
project-jasmin-GA3          katie.kxxx          GA3 data transfers
project-crum-an-access      roger.mxxxxx        early adopter access to /crum/an...
project-ceda-access         roger.mxxxxx        CEDA staff access
project-badc-test           martin.oxxxxxxxx    Project for testing BADC access
```

List project information

```
$ moo projinfo project-badc-test --long
project-badc-test   martin.oxxxxxxx  Collaboration
                         Project for testing BADC access
Associated sets:

Access rules:
   Read /crum/x*                 moose:/crum/x     read
   Read /crum/qtes*              moose:/crum/qtes  read
   Read /crum/t*                 moose:/crum/t     read
```

Listing with wildcards

```
$ moo ls -l moose:/crum/qtes*

S martin.oxxxxxxx  0.14 GBP 1282250062 2014-11-25
                           16:43:36 GMT moose:/crum/qtest
```

Listing data Classes

```
$ moo ls :  # ":" is short for "moose:"
moose:/adhoc
moose:/crum
moose:/devfc
moose:/misc
moose:/opfc
```

Obtaining Set information (1)

```
$ moo setinfo  moose:/crum/qtest

Information for : moose:/crum/qtest
---------------------------------------------
Owner: martin.oxxxxxxxx
Size: 1282250062
Cost: 0.14GBP
Category: UNCATEGORISED
Tags:
Moose Comments: "This is a comment"
Protection Level: Managed Protection Level
```

Obtaining Set information (2)

```
$ moo ls -l moose:/crum/qtest

C martin.oxxxxxxxx  0.14 GBP  1272196880 2014-11-25 16:43:36
   GMT moose:/crum/qtest/ada.file
C martin.oxxxxxxxx  0.00 GBP  1610256 2014-03-21 15:37:10 GMT
   moose:/crum/qtest/ama.pp
C martin.oxxxxxxxx  0.00 GBP  8442926 2014-03-21 15:51:42 GMT
   moose:/crum/qtest/ana.nc.file
```

Getting a single File (should you want to)

```
$ moo get moose:/crum/qtest/ama.pp/small.pp small.pp
### get, command-id=178814452, estimated-
  cost=1468736byte(s), files=1, media=1
See /home/users/rogerm for conditions of use.


$ ls -l small.pp
-rw-r--r-- 1 rogerm users 1468736 Jan 20 14:34 small.pp
```

Getting a whole Collection (should you want to)

```
$ mkdir local_collection
$ moo get moose:/crum/qtest/ama.pp local_collection
### get, command-id=178988366, estimated-
  cost=1610256byte(s), files=2, media=0
See /home/users/rogerm/local_collection for conditions
  of use.

$ ls -l local_collection
total 2160
-rw-r--r-- 1 rogerm users 1468736 Jan 21 10:28 small.pp
-rw-r--r-- 1 rogerm users  141520 Jan 21 10:28 test.pp
-rw-r--r-- 1 rogerm users 551 Jan 21 10:28 data_licence_178988366.lic
```

Selecting from a Collection (more efficient)

- Create a query file to specify which fields you require:

```
$ cat query_file.txt
begin
  # 5216   90 TOTAL PRECIPITATION RATE     KG/M2/S
  stash=5216
  # only want data from 1990-91
  year=[1990..1991]
end
```

Selecting from a Collection (1)

- Run the query using “moo select”

```
$ mkdir precip_results_90-91
$ moo select query_file.txt moose:/crum/antie/apa.pp
  precip_results_90-91
### select, command-id=179201192, estimated-
  cost=79822080byte(s), files=24, media=1
See /home/users/rogerm/precip_results_90-91 for
  conditions of use.
```

Selecting from a Collection (2)

- Look at the results

```
$ moo ls –l moose:/crum/antie/apa.pp/antiea.pa1990apr.pp
F paul.exxxxxxx  0.05 GBP  430631528 2013-12-07 01:35:45
   GMT moose:/crum/antie/apa.pp/antiea.pa1990apr.pp

$ ls -l precip_results_90-91/antiea.pa1990apr.pp
-rw-r--r-- 1 rogerm users 3325920 Jan 22 11:47
   precip_results_90-91/antiea.pa1990apr.pp
```

The licence file

```
$ ls -l local_collection/*.lic
-rw-r--r-- 1 rogerm users 551 Jan 21 10:28
                          local_collection/data_licence_178988366.lic
$ cat local_collection/data_licence_178988366.lic
Retrieval of Data from UK Met Office Archive
============================================
Details of Licence, Terms & Conditions
Data retrieved by: badc.rxxxx.mxxxxxx
Date: Wed Jan 21 10:28:41 GMT 2015
Command-ID: 178988366
Data retrieved from Data-Set: '/crum/qtest' is made available under the
                          following licence(s),
details of which can be viewed by following the Terms & Conditions link(s)
   below.
  Licence: 'Collaboration Licence' granted under project 'project-badc-test';
          T&C:
   http://collab.metoffice.gov.uk/twiki/bin/view/Support/ExMASSTandCs
```

Further information on usage

- Main help page (on collaborative wiki)

http://collab.metoffice.gov.uk/twiki/bin/view/Support/ExternalAccessToMASS

- Links to many documents, including:
  - Getting started with MOOSE on MONSooN and JASMIN (a gentle MOOSE introduction)

http://collab.metoffice.gov.uk/twiki/bin/view/Support/MooseOnMonsoon

  - Full MOOSE User Guide

http://collab.metoffice.gov.uk/twiki/bin/viewfile/Static/MASS/user_guide.html

External Access to MASS

- Getting access

Getting Access

1. Get a CEDA/BADC account
2. Get a JASMIN login account
3. Apply for access to the MOOSE resources on JASMIN
   - (existing JASMIN users will also need to do this)
4. [non-Met Office users:] Arrange for a Met Office Sponsor to initiate a request for a MASS account.
5. [Met Office employees:] Request an external access to MASS account.

See

http://collab.metoffice.gov.uk/twiki/bin/view/Support/ExternalAccessToMASS

or e-mail monsoon@metoffice.gov.uk to get started.

Questions and answers