

***/p/lustre** File System Rollout**

including information on /p/Lustre and /p/gpfs quotas

March 2019 LC User Meeting

Ines Heinz
Tim Fahey

March 28, 2019



Lustre Quotas – aka Tiers

- What /p/lustre file systems are available
- What is a Tier
- What happens if a user hits their quota
- Why is the info needed
- Where are my files

Lustre filesystems are provided by LC.
The programs manage who gets Tier3.

What /p/lustres are available

- Each lustre is unique and has it's own quota system
- The Tier1 and Tier2 quotas are the same on each lustre to make it easier to understand and manage.

CZ	RZ	SCF
/p/lustre1	/p/lustre1	/p/lustre1
/p/lustre2 (formerly known as lscratchh*)		/p/lustre2
/p/lustre3 (mounted on catalyst, pascal)		

*Users will be able to use the /p/lscratchh symlink to get to their files indefinitely so they may be confused, but there won't be purging there.

What is a Tier

■ Tier 1

- Default Tier for every user.
- Limits set to 20TB space and 1M inodes
- Defaults chosen based on actual LC data from an Iscratch system

■ Tier 2

- Hotline is authorized to grant this to users who submit Tier 2 email request
- Limits set to 50TB space and 10M inodes
- Each lustre is separate so Tier 2 on one lustre doesn't mean Tier 2 on another lustre

■ Tier 3

- Hotline needs to gather details
- Limits based upon email justification and programmatic need
- Users need to be helped with requesting appropriate timeframe
- Like the airlines with their seats - we are over-subscribed

What happens if a user hits their limits

- **Soft quota**
 - Based on either space or inodes
 - Reminds user about upcoming issue
 - User can still write
 - Starts a 10 day timer
 - If user gets below 90% within those 10 days, they go back to normal
 - After 10 days, they stop being able to write until they get below 90%
 - Files not deleted except
 - User can delete their files
 - Files placed in 0_LC_AutoDelete will be automatically deleted
- **Hard quota**
 - Based on either space or inodes
 - Stop being able to write
 - Get “helpful” messages like “EOQUOT” and “slurmstepd: error”
 - Files not deleted except
 - User can delete their files
 - Files placed in 0_LC_AutoDelete will be automatically deleted
- **Quotas are per user, not per directory**

Tier 3 Request: Why is the info needed?

- <https://myconfluence.llnl.gov/display/RAM/LustreX> (login with OUN+AD)
- <https://lc.llnl.gov/confluence/display/LC/Lustre+Quota+Increase+Requests> (login with LC username and RSA PIN+token code)

LC, I am doing <3D simulations> in support of the <L1 xxx> programmatic effort that is funded through <WCI> and which is due <end of FY18>. My programmatic supervisor is <Andy Anderson - anderson1>. This work requires multiple runs and additional space. Can you please request a Tier 3 increase for me? I understand that this can only be granted for a maximum of a year with the appropriate justification and I understand that when it expires, I will automatically go back to the 50TB/10M Tier 2 allocation.

- Cognizant APD*/DL**: <Derek Wapman - wapman1>
- Zone: <CZ> ## choose one - CZ, RZ, or SCF,
- Directory: <p/lustre1>
- LC Username: <ines>
- Space: <#>TB
- Inode: <#>M
- Expiration: <28-Feb-19>

Thanks! <ines>


*APD = Associate Program Director

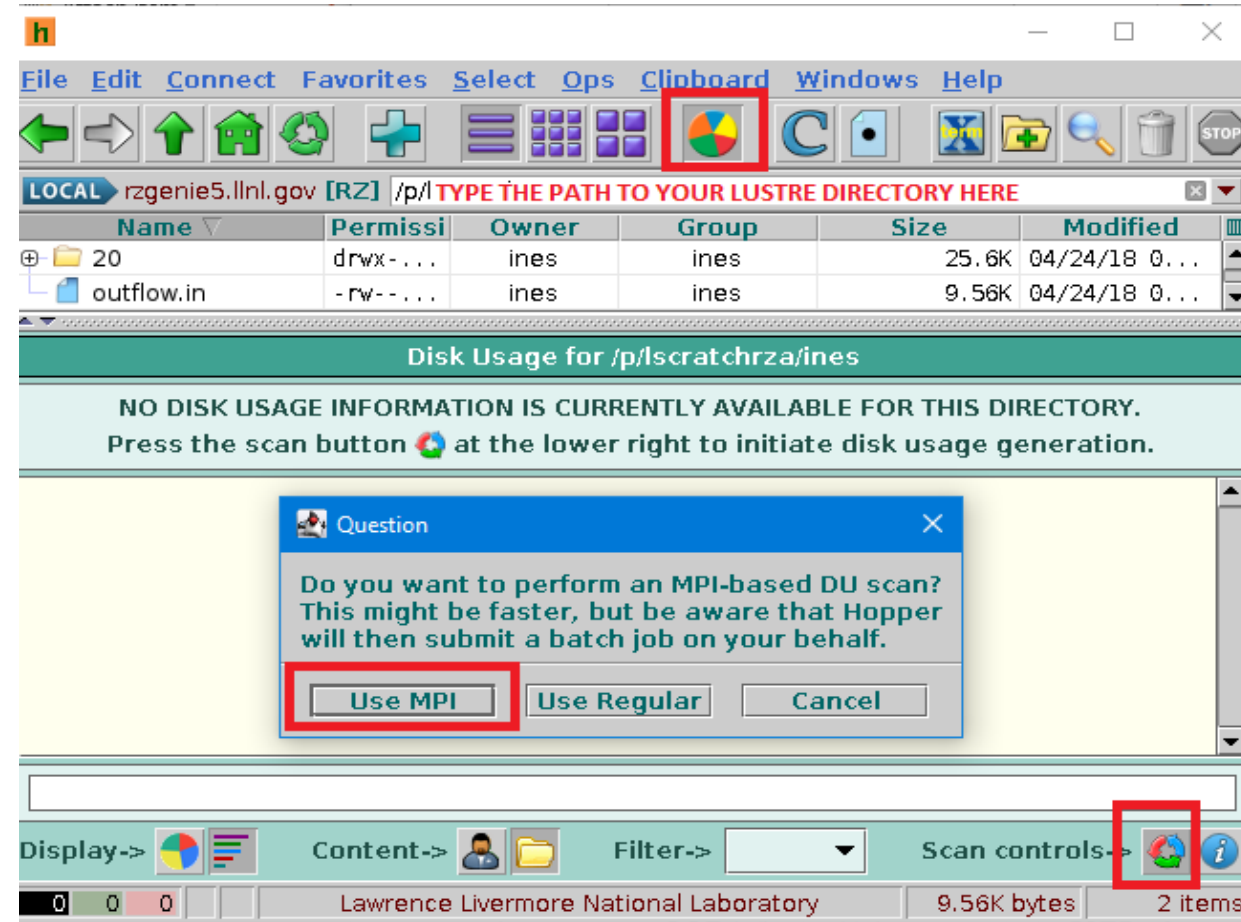
**DL = Division Leader

Why is the info needed

- 3D simulations/Effort/WCI
 - Helps define the programmatic need
- End of FY19
 - Helps define timeframe and urgency
- Cognizant APD/DL
 - Tier 3 is split out amongst the programs
- Zone/Directory
 - Defines where user needs that Tier
- LC Username
 - User can request this for other users
- Space/Inode
 - Need to know how much user needs
 - User asking for “as much as possible” will take much longer to get need filled
- Expiration
 - Can be granted for up to a year, but has to match the need.
 - User asking for a year need to understand that it will slow down their request

Where are my files

1. Log onto a slic machine (oslic, rzslic, or cslic).
2. Launch Hopper 
3. In the address bar of Hopper /p/lustre*/<LCusername>
4. Click the pie chart
5. Click the icon for the Scan controls (at bottom)
6. Click "Use MPI"
7. Will submit a batch job on one of the two batch nodes on the slic cluster in question.
8. Type showq command to monitor status of the job.



The screenshot shows the Hopper web interface. The address bar displays the path `/p/` with a red box around the pie chart icon in the toolbar. Below the address bar is a table listing files and folders:

Name	Permissi	Owner	Group	Size	Modified
20	drwx-...	ines	ines	25.6K	04/24/18 0...
outflow.in	-rw--...	ines	ines	9.56K	04/24/18 0...

Below the table is a section titled "Disk Usage for /p/lscratchrza/ines" with a message: "NO DISK USAGE INFORMATION IS CURRENTLY AVAILABLE FOR THIS DIRECTORY. Press the scan button at the lower right to initiate disk usage generation." A dialog box titled "Question" is open, asking: "Do you want to perform an MPI-based DU scan? This might be faster, but be aware that Hopper will then submit a batch job on your behalf." The "Use MPI" button is highlighted with a red box. At the bottom of the interface, the "Scan controls" section has a red box around the scan button icon.

GPFS Quotas: Lassen and RZAnsel

- /p/gpfs1
- **Tier 1**
 - Default Tier for every user.
 - Limits set to 20TB space and 1M inodes
- **Tier 2**
 - Hotline is authorized to grant this to users who submit Tier 2 email request
 - Limits set to 50TB space and 10M inodes
 - Each gpfs is separate (CZ/RZ/SCF) so Tier 2 on one gpfs doesn't mean Tier 2 on another gpfs
- **Tier 3**
 - Hotline needs to gather details
 - Limits based upon email justification and programmatic need
 - Users need to be helped with requesting appropriate timeframe
 - Like the airlines with their seats - we will over-subscribe

GPFS Quotas: Sierra

- Sierra is a TriLab machine
 - LANL, SNL, and LLNL
 - Managed through ATCC campaign proposals
- /p/gpfs1
- Quota
 - As approved through the ATCC process
 - Programmatic negotiation and approval



Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.