



Oregon State  
University



# USER COMPUTING FOR DUNE

3/31/19

Heidi Schellman, Oregon State University



# Our disks are full and we are sad

## dune persistent storage usage by user

User	Group	Name	Space used (GiB)	# Files
jmalbos	dune	Justo Martin-Albo Simon	17,865	187,829
mrobinso	dune	Matthew Robinson	13,352	205,623
dbrailsf	dune	Dominic Brailsford	11,282	735,899
marshalc	dune	Christopher Marshall	8,388	429,674
iseong	dune	Ilsoo Seong	7,900	1,011
tlord	dune	Tom Lord	7,626	258,923
gyang	dune	Guang Yang	7,409	290,160
tejinc	dune	Tejin Cai	5,484	171,998
yj2429	dune	Yeon-jae Jwa	5,387	217,917
econley	dune	Erin Conley	5,383	186,270
yzhou	dune	Yuyang Zhou	5,279	128,138
dlast	dune	David Last	5,164	17,856



# Strategy?

- Impose user quotas (1 TB?)
- Create group areas to preserve and prioritize important projects
- Any sample  $> 1$  TB needs to be documented and preserved using sam4users
  - ▣ Needs effort to document and train
- Once in sam datasets data can migrate to other sites



# Big S+C is watching you

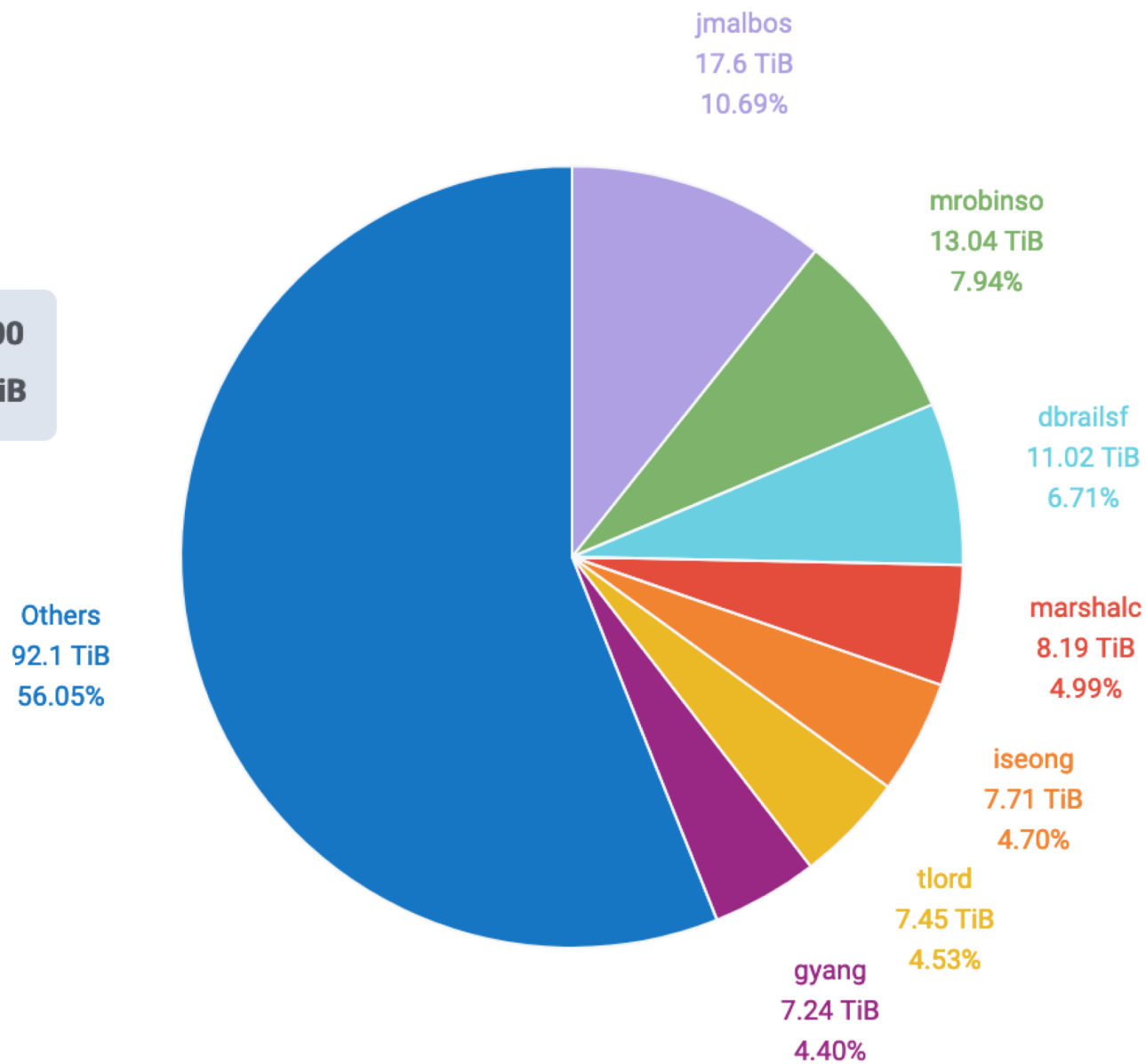
- [http://fndca3a.fnal.gov/cgi-bin/space\\_usage\\_by\\_user.cgi.py?key=dune](http://fndca3a.fnal.gov/cgi-bin/space_usage_by_user.cgi.py?key=dune)
- <https://fifemon.fnal.gov/monitor/d/000000175/dcachel-persistent-usage-by-vo?orgld=1&var-VO=dune&from=1551483205769&to=1554071605769&panellid=5&fullscreen>



# Usage per user (GiB)

2019-03-14 12:50:00

dune: 167.89 TiB



# Small files

6

- Dcache and small files do not get along
- MINERvA sped up analysis by factor of  $\sim 10$  by moving to larger files
- How large are user files?
- Can they merge them?



# Large files

- Some of the largest users are producing large files (good)
- But they are not art so no metadata.
- Need to generate metadata and back these puppies up.



# Metadata

- I teach responsible conduct of research
- You do not create huge samples, put them on unbacked up disk, not catalog them and then do science with them. Your boss swore to the NSF and DOE you would not do this... (at the same time he/she promised to mentor postdocs and students)
- We need a documented, easy, but enforced way to describe and archive large samples.





# Disk resources

9

- We have disk in the UK, at CERN and other US sites.
- We probably need more analysis disk at FNAI
- How can users use these resources transparently?
  - ▣ Make datasets with them
  - ▣ Use rucio to move and catalog
  - ▣ Use sam to find them

