

Getting Started with the National Systems
Video Series



computecanada

File Systems

by Pawel Pomorski, SHARCNET, University of Waterloo

Overview

Compute Canada provides a wide range of storage options to cover the needs of our very diverse users

Solutions range from high-speed temporary local storage to different kinds of long-term storage

Filesystems on Compute Canada are a shared resource and for this reason must be used responsibly

These filesystems are designed to store a limited number of very large files, typically binary rather than text files

Users must avoid storing thousands of small files (<1MB in size)



Best Practices

Only use text format for files that are smaller than a few megabytes

As far as possible, use local storage for temporary files

If your program must search within a file, it is fastest to do it by first reading it completely before searching, or to use a RAM disk

Regularly clean up your data in the scratch and project spaces, to free up space for new data

If you no longer use certain files but they must be retained, archive and compress them, and if possible copy them elsewhere



Storage Types

Storage Type	Accessibility	Throughput (large operations, >1 MB per operation)	Latency (small operations)	Longevity
Network Filesystem (NFS) /home	All nodes	Poor	High	Long term
Long-Term Parallel Filesystem /project	All nodes	Fair	High	Long term
Short-Term Parallel Filesystem /scratch	All nodes	Fair	High	Short term (periodically cleaned)
Local Filesystem /tmp	Local to the node	Fair	Medium	Very short term
Memory (RAM) Filesystem	Local to the node	Good	Very low	Very short term, cleaned after every job

Filesystem Quotas and Policies

Filesystem	Quotas	Backed up?	Purged?	Available by Default?	Mounted on Compute Nodes?
Home Space /home	50 GB, 500 K files	Yes	No	Yes	Yes
Scratch Space /scratch	20 TB and 1000k files per user, 100 TB and 10M files per group	No	Yes, all files older than a certain number of days	Yes	Yes
Project Space /project	1TB for each cluster per user, can request increase to 10TB 500k files per user	Yes	No	Yes	Yes
Nearline Space	1 to 10 TB per group	No	No	No	No

Checking your usage

check how much space is used in current directory and all subdirectories with disk usage (du) command, using convenient units (-h):

du -h

find all files bigger than 100 MB in current directory and all subdirectories

find -type f -size +100M

Other ways to monitor your file usage will be made available later



Getting more storage

If your needs are not well served by the available storage options please contact us by sending an email to support@computecanada.ca

If you require more storage space than available via default quota, it is possible to request it in our annual RAC competition.

More information at Compute Canada Wiki (docs.computecanada.ca). Search for “Storage and File Management”

