

HPC Data Management: Or how (not) to handle your data in an HPC environment

Marty Kandes, Ph.D.

Computational & Data Science Research Specialist
High-Performance Computing User Services Group
San Diego Supercomputer Center
University of California, San Diego

CIML Summer Institute
Monday, June 27th, 2022
10:05 AM - 11:35 AM PT

tl;dr: Data has a lifecycle, data management is a lifestyle



Image Credit: Harvard Biomedical Data Management

HPC Data Management: Or how (not) to handle your data in an HPC environment

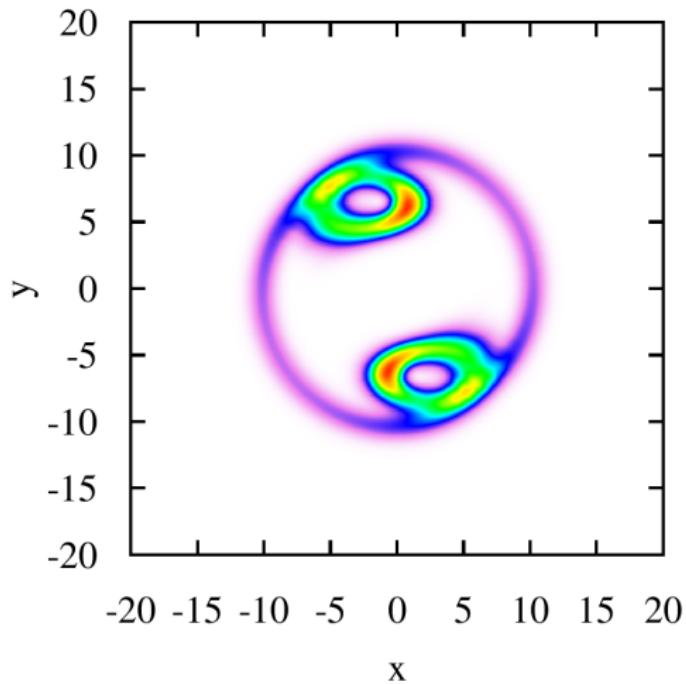
- ▶ Case Study: A simulation-visualization workflow
- ▶ Getting data to (and from) an HPC system
- ▶ Managing data on an HPC system
- ▶ Reading, writing, and preparing data for HPC
- ▶ Preventing data loss

Case Study: A simulation-visualization workflow

Exploring the Dynamics of a Quantum-Mechanical Compton Generator

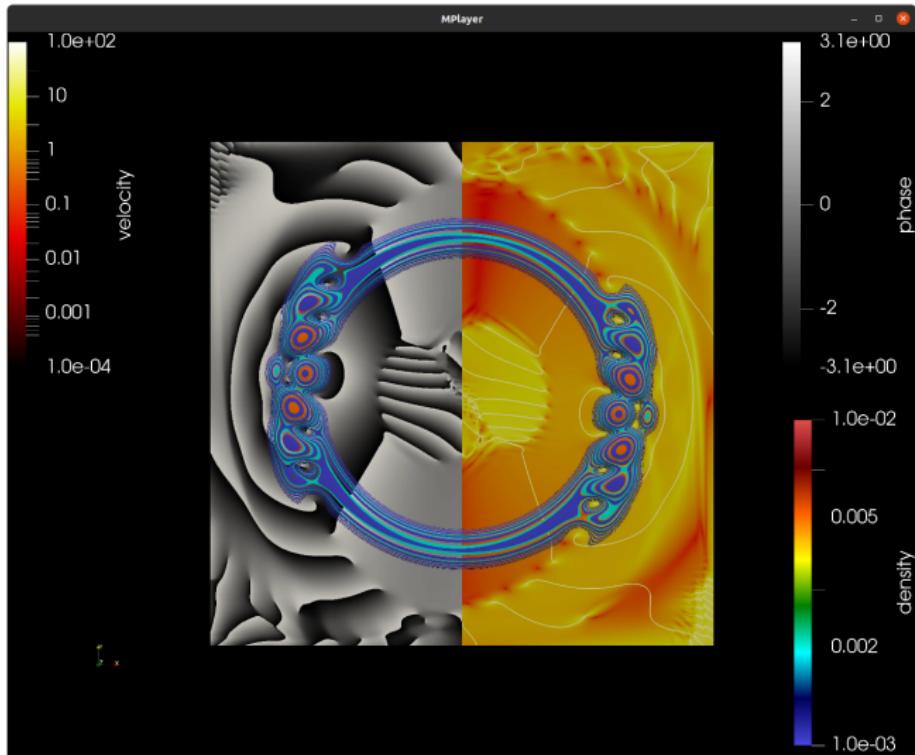
The objective of this XSEDE project is to explore the dynamics of a quantum-mechanical Compton generator through the use of numerical simulations. In this initial research allocation request, we aim to determine the fundamental mechanism by which the Compton-induced drift velocity manifests itself. We hypothesize that the drift velocity likely arises due to the nucleation of quantized vortices within the system. To test this hypothesis, we request **1.8M SUs on Comet for production simulation runs, 32 TB of project storage on Data Oasis to store simulation data, 375K SUs on Maverick for visualization and analysis of this simulated data**, and 100K SUs on both Comet and Stampede, respectively, for code development and testing.

Preliminary (Projected) 2D Visualization



[movies/psi2xy.mp4](#)

Full 3D Visualization



[movies/gpse-v059-comet-compton-20170828-004-density-phase-velocity-xyz.mp4](#)

Simulation and Visualization Resources



Simulation on **Comet** @ SDSC



Visualization on **Maverick** @ TACC

```
mkanade@bar@task: ~/Dropbox/software/projects/gpse/source
```

```
995 ! .... BEGIN MAIN TIME PROPAGATION LOOP
996
997 ! Initialize simulation time
998 tN = t0
999
1000 IF (.reanOn == .EOF, .FALSE.,) THEN ! start main simulation time
1001 ! propagation loop
1002
1003 DO n = 0, nSteps
1004
1005 CALL MPI_BARRIER(MPI_COMM_WORLD, mpiError)
1006
1007 Compute partial relations to file from MPI_MASTER; write
1008 wave function and external potential to file from MPI_MASTER
1009 IF (MODULO(n, nWrite) == 0) THEN
1010
1011 Compute partial base expectation values locally on each
1012 MPI process; reduce partial base expectation values from
1013 all MPI processes to MPI_MASTER to get full base
1014 expectation values
1015 CALL evau_compute_base(MPI_MASTER, mpiReal, mpiError, 1,&
1016 & f0Order, nXa, nXb, nXc, nra, nrb, nrc, nza, nzb, &
1017 & nZc, nDx, nDy, nDz, g5, Xa, Ya, Za, &
1018 & Vxa3d, Ps3d)
1019
1020 Compute derived expectation values, uncertainties from base
```

Welcome to the TACC Visualization Portal
Simple access to TACC's Vis Resources

Features:

- Run an interactive, web-based visualization
- Python / Jupyter Notebook Integration
- R Studio Integration
- Run on Maverick, Stampede and Stampede-KNL, and Wrangler
- Visualization jobs submission and monitoring
- Current resource usage and allocation view

Authentications:

- TACC User Portal User
- XSEDE User Portal User

Username: Password: Login

Job Submission VNC Visualization Session

Simulation data, I/O, and storage requirements

- ▶ What is the input data for a simulation?
- ▶ How much output data is generated during a simulation?
- ▶ What is the file format of the data?
- ▶ Does the simulation code use serial or parallel I/O methods?
- ▶ Where should the input and output data be stored?
- ▶ What are there storage limits on these filesystems?
- ▶ How should the data for a simulation be organized?
- ▶ Should the data be compressed?

Visualization data, I/O, and storage requirements

- ▶ How much simulation data needs to be visualized?
- ▶ Where should this simulation data be stored?
- ▶ What are there storage limits on these filesystems?
- ▶ How should the simulation data be transferred?
- ▶ Where should the visualization data be stored?
- ▶ Will the visualization data undergo further post-processing?

Simulation and data transfer workflow

- ▶ `code/run-gpse-comet.sh`
- ▶ `code/gpse.input`
- ▶ `code/compress-wave-functions.sh`
- ▶ `code/gridftp-transfer.sh`

Simulation batch job script

```
mkandes@hardtack: ~/Dropbox/work/ucsd/sdsc/presentations/2022/sdsc/clml/data-...
#!/usr/bin/env bash

#SBATCH --job-name=gpse-v0.5.9-comet-compton-20170828-017
#SBATCH --account=sdu118
#SBATCH --partition=compute
#SBATCH --nodes=6
#SBATCH --ntasks-per-node=24
#SBATCH --cpus-per-task=1
#SBATCH --mem=120G
#SBATCH --time=48:00:00
#SBATCH --output=%x.o%j.%N

declare -xr LOCAL_SCRATCH_DIR="/scratch/${USER}/$(SLURM_JOB_ID)"
declare -xr LUSTRE_SCRATCH_DIR="/oasis/scratch/comet/${USER}/temp_project/sdull18"
declare -xr LUSTRE_PROJECTS_DIR="/oasis/projects/nsf/sdull18/${USER}"

declare -xr COMPILER_MODULE='intel/2016.3.210'
declare -xr MPI_MODULE='intelmpi/2016.3.210'
declare -xr GNUTOOLS_MODULE='gnutools/2.69'

module purge
module load "${COMPILER_MODULE}"
module load "${MPI_MODULE}"
module load "${GNUTOOLS_MODULE}"
module list
export PATH="${HOME}/software/gpse/0.5.9/bin:${PATH}"
printenv

mkdir -p "${LUSTRE_SCRATCH_DIR}/${SLURM_JOB_NAME}"
cd "${LUSTRE_SCRATCH_DIR}/${SLURM_JOB_NAME}"
cp "${SLURM_SUBMIT_DIR}/gpse.input" ./"

time -p mpirun -n "${SLURM_NTASKS}" gpse.x
```

1,1

All

Simulation input file

```
mkaned@hardtack: ~/Dropbox/work/ucsd/sdsc/presentations/2022/sdsc/cimi/data-management/code
gpseIn
itpOn      = .FALSE.          ! Perform imaginary time propagation? .TRUE. = Yes ; .FALSE. = No
rk4Lambda  = 2                ! 1 = Tan-Chen-1 ; 2 = Classical 4th-Order Runge-Kutta ; 3 = Tan-Chen Lambda-3 ; 4 = England ; 5 = Tan-Chen-5
fdOrder    = 2                ! 2 = 2nd-Order Central Differences ( CD ) ; 4 = 4th-Order CD ; 6 = 6th-Order CD ; 8 = 8th-Order CD
nsteps     = 155520           ! Total number of time steps in simulation
nwrite     = 648               ! Number of IO writes to disk; i.e., number of time steps between writes to disk
nX         = 1152              ! Number of grid points along the x-axis
nY         = 1152              ! Number of grid points along the y-axis
nZ         = 576               ! Number of grid points along the z-axis
dNx        = 1                ! Write out wave function only every dNx grid points along the x-axis
dNy        = 1                ! Write out wave function only every dNy grid points along the y-axis
dNz        = 1                ! Write out wave function only every dNz grid points along the z-axis
t0         = 0.0               ! Time at the beginning of the simulation
tF         = 30.0              ! Time at the end of the simulation; only use for compton generator
x0         = 0.0               ! X-coordinate of the centre of the grid
y0         = 0.0               ! Y-coordinate of the centre of the grid
z0         = 0.0               ! Z-coordinate of the centre of the grid
dT         = 1.929812345679e-4 ! Interval of a time step
dX         = 0.027777777777 ! Distance between grid points along the x-axis
dY         = 0.027777777777 ! Distance between grid points along the y-axis
dZ         = 0.027777777777 ! Distance between grid points along the z-axis
x0rrf      = 0.0              ! X-coordinate of the rotating reference frame's origin
y0rrf      = 0.0              ! Y-coordinate of the rotating reference frame's origin
z0rrf      = 0.0              ! Z-coordinate of the rotating reference frame's origin
wX         = 0.0               ! X-component of the rotating reference frame's angular velocity vector
wY         = 0.0               ! Y-component of the rotating reference frame's angular velocity vector
wZ         = 0.5               ! Z-component of the rotating reference frame's angular velocity vector
g5         = 0.0               ! NonLinear atom-atom interaction coupling constant
psiInput   = 1                ! 0 = No input wave function; 1 = Read wave function from .bin file; 2 = Read wave function from .vtk file ( not available yet )
psiOutput  = 3                ! 0 = No output wave function; 1 = Write wave function to .bin file; 2 = Write wave function to .vtk file; 5 = Write int den to .splot
psiFileNo  = 500              ! Set psiFileNo to input wave function file number
psiInit    = 3                ! 0 = Isotropic 3D SHO ; 1 = Anisotropic 3D SHO ; 2 = Axisymmetric 3D SHO ; 3 = Approx 3D SHOR
nxpsi     = 0                 ! Degree of Hermite polynomial used to define anisotropic SHO wave function along x-axis
nypsi     = 0                 ! Degree of Hermite polynomial used to define anisotropic SHO wave function along y-axis
nzpsi     = 0                 ! Degree of Hermite polynomial used to define both anisotropic and axially-symmetric SHO wave functions along z-axis
nrpsi     = 0                 ! Degree of (associated) Laguerre polynomials used to define radial components of isotropic and axially-symmetric SHO wave functions
mlpsi     = 0                 ! Projection of orbital angular momentum along z-axis for axially-symmetric SHO wave function
x0psi     = 0.0               ! X-coordinate of origin used to define initial wave function
y0psi     = 0.0               ! Y-coordinate of origin used to define initial wave function
z0psi     = 0.0               ! Z-coordinate of origin used to define initial wave function
r0psi     = 10.0              ! Radius of psiInit = 3 SHOR approx wavefunction
wXpsi    = 0.0               ! Angular frequency of SHO potential along x-axis used to define anisotropic SHO wave function
wYpsi    = 0.0               ! Angular frequency of SHO potential along y-axis used to define anisotropic SHO wave function
wZpsi    = 1.0                ! Angular frequency of SHO potential along z-axis used to define both anisotropic and axially-symmetric SHO wave functions
wRpsi    = 1.0                ! Radial angular frequency of axially-symmetric SHO or SHOR potential used to define axially-symmetric SHO or approx SHOR wave functions
pxpsi    = 0.0               ! Initial linear momentum boost along x-axis
pypsi    = 0.0               ! Initial linear momentum boost along y-axis
gpse.input" 64L, 5703C
1,2          Top
```

Post-simulation data compression script

```
mkandes@hardtack: ~/Dropbox/work/ucsd/sdsc/presentations/2022/sdsc/ciml/data...  
#!/usr/bin/env bash  
  
declare -xr LUSTRE_SCRATCH_DIR="/oasis/scratch/comet/${USER}/temp_project"  
declare -xr LUSTRE_PROJECTS_DIR="/oasis/projects/nsf/sdu118/${USER}"  
  
declare -xr JOB_NAME='gpse-v0.5.9-comet-compton-20170828-013'  
  
declare -xir START_TIMESTEP=1000  
declare -xir END_TIMESTEP=1240  
  
cd "${LUSTRE_SCRATCH_DIR}/sdu118/${JOB_NAME}"  
  
for ((x="${START_TIMESTEP}"; x<="${END_TIMESTEP}"; x++)); do  
    mkdir -p psi-$x  
    mv psi-$x-*vtk psi-$x  
    zip -r psi-$x.zip psi-$x  
    rm -rf psi-$x  
done  
1,1 All
```

Data transfer script

```
mckandes@hardtack: ~/Dropbox/work/ucsd/sdsc/presentations/2022/sdsc/ciml/data...     
```

```
#!/usr/bin/env bash

declare -xr SOURCE_ENDPOINT='oasis-dm.sdsc.xsede.org:2811'
declare -xr SOURCE_DIR='/oasis/projects/nsf/sdull18/mkandes/gpse-v0.5.9-comet-compton-20170828-001'

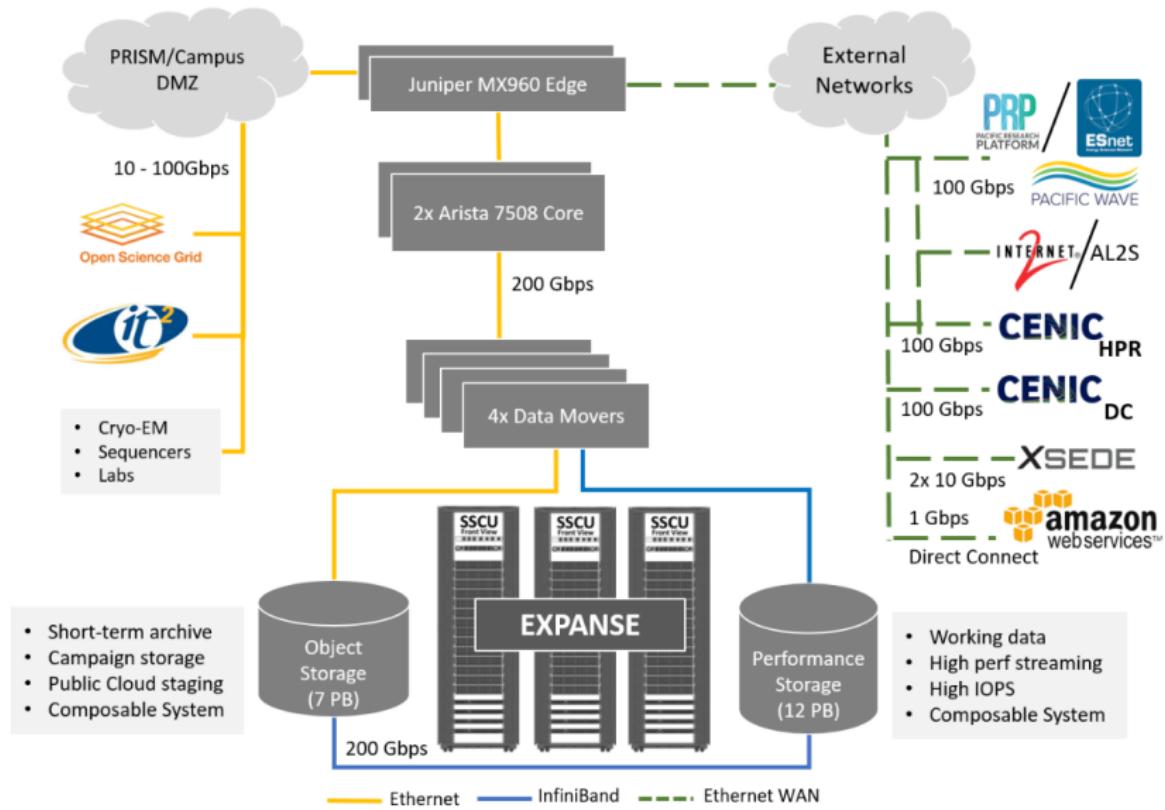
declare -xr DESTINATION_ENDPOINT='gridftp.stampede2.tacc.xsede.org:2811'
declare -xr DESTINATION_DIR='/work/03216/mckandes/maverick/gpse-v0.5.9-comet-compton-20170828-001'

module load globus

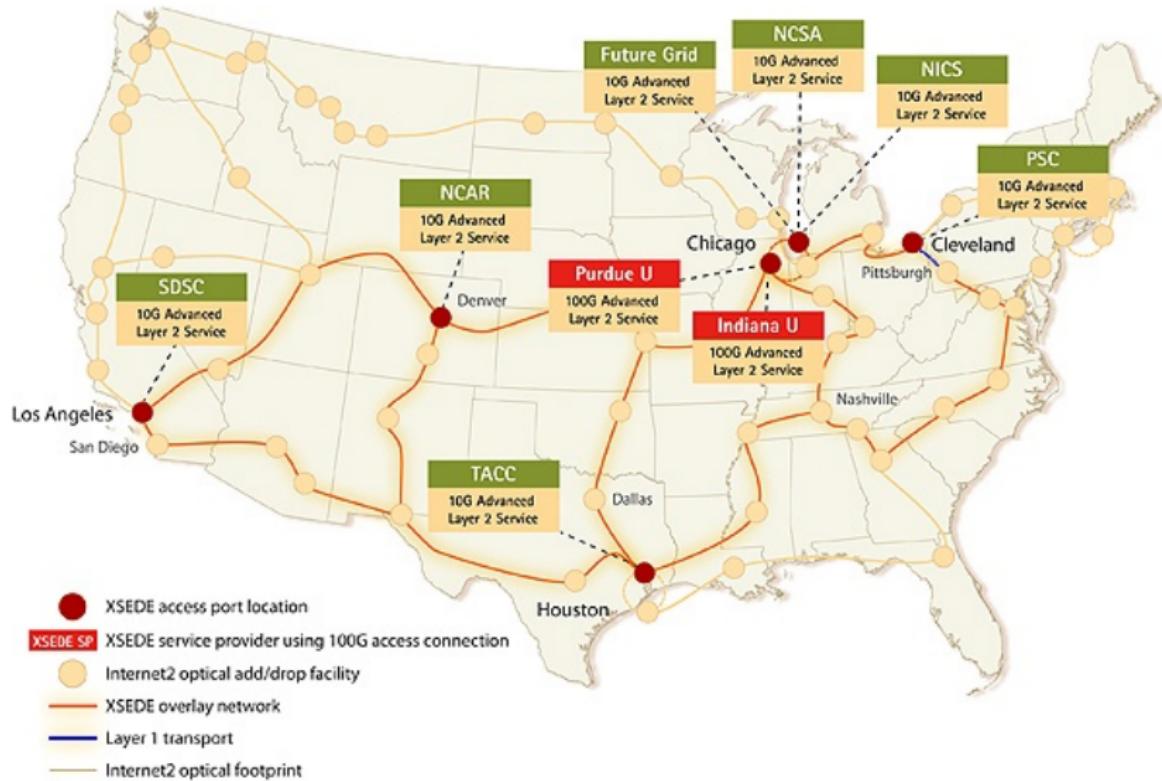
for timestep in {1000..1024}; do
    globus-url-copy -vb -stripe -tcp-bs 8m -p 4 "gsiftp://${SOURCE_ENDPOINT}/${SOURCE_DIR}/psi-${timestep}.zip" "gsiftp://${DESTINATION_ENDPOINT}/${DESTINATION_DIR}/psi-${timestep}.zip"
done
```

1,1 Top

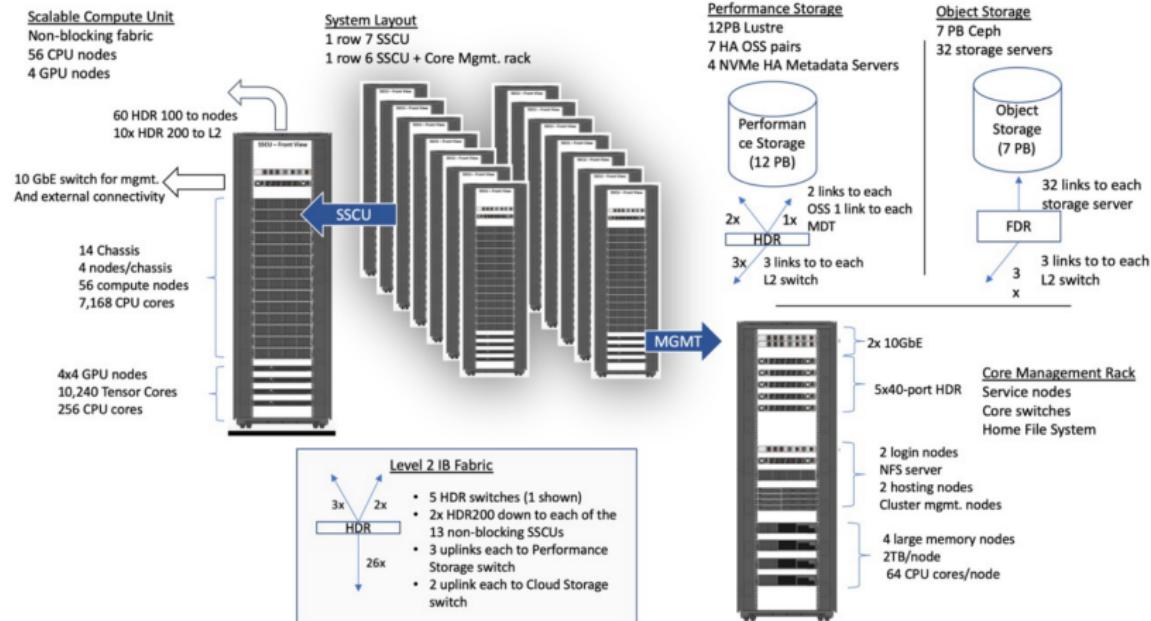
Expanse storage systems and external networks



Research and Education Networks



Expanse System Architecture



Expanse Login Nodes

```
mkandes@login01:~$ ssh expanse
Welcome to Bright release      9.0
                                         Based on Rocky Linux 8
                                         ID: #000002

-----
WELCOME TO
  

-----
Use the following commands to adjust your environment:  

'module avail'          - show available modules  

'module add <module>'    - adds a module to your environment for this session  

'module initadd <module>' - configure module to be loaded at every login  

-----
Last login: Sat Jun 25 11:06:48 2022 from 208.58.214.56
[mkandes@login01 ~]$
```

Expanse \$HOME Network File System (NFS)

```
mkandes@login01 ~]$ pwd
/home/mkandes
[mkandes@login01 ~]$ cat /etc/auto.home | grep mkandes
mkandes -fstype=bind :/expanse/nfs/home2/mkandes
mkandes test -fstype=bind :/expanse/nfs/home4/mkandes_test
[mkandes@login01 ~]$ df -Th | grep mkandes
10.22.100.112:/pool2/home/mkandes          nfs      205T  9.7T  195T   5% /home/mkandes
[mkandes@login01 ~]$ ls
benchmarks           scripts
cm                  software
data                spack-bootstrap-new-instance.sh
install-expanse-core-packages.R             tmp
NAMD_2.14_Source.tar.gz                    vmd-1.9.3.bin.LINUXAMD64-CUDA8-OptiX4-OSPRay111p1.opengl.tar.gz
projects
[mkandes@login01 ~]$
```

- ▶ Usage: Source code, compiled software binaries and libraries, small input/output files, batch job scripts, other important files such as containers
- ▶ Quota: approximately 100 GB
- ▶ Backup: **ZFS** snapshots performed once or twice per month
- ▶ Purge: 90 days after allocation expiration

df

mkanedes@login02:~

DF(1) User Commands DF(1)

NAME
df - report file system disk space usage

SYNOPSIS
`df [OPTION]... [FILE]...`

DESCRIPTION
This manual page documents the GNU version of `df`. `df` displays the amount of disk space available on the file system containing each file name argument. If no file name is given, the space available on all currently mounted file systems is shown. Disk space is shown in 1K blocks by default, unless the environment variable `POSIXLY_CORRECT` is set, in which case 512-byte blocks are used.

If an argument is the absolute file name of a disk device node containing a mounted file system, `df` shows the space available on that file system rather than on the file system containing the device node. This version of `df` cannot show the space available on unmounted file systems, because on most kinds of systems doing so requires very non-portable intimate knowledge of file system structures.

Manual page df(1) line 1 (press h for help or q to quit)

Type(s) of I/O

```
[mkandes@login01 ~]$ cd ~/software/gpse/0.6.5/mpi-omp/gcc/10.2.0/openmpi/4.0.4/gpse...
[mkandes@login01 gpse]$ ls
build  CHANGELOG  gpse.input  gpse.x  LICENSE  Makefile  README  source
[mkandes@login01 gpse]$ cd source/
[mkandes@login01 source]$ ls
evua.f90  grid.f90  io.f90  pmca.f90  rot.f90
gpse.f90  grk4.f90  math.f90  psi.f90  vex.f90
[mkandes@login01 source]$ grep -H psiOutput *.f90
gpse.f90:          & psiInput, psiOutput, psiFileNo, psiFileNoChkpt, psiInit, &
gpse.f90:          IF (psiOutput == 1) THEN
gpse.f90:          ELSE IF (psiOutput == 2) THEN
gpse.f90:          ELSE IF (psiOutput == 3) THEN
gpse.f90:          ELSE IF (psiOutput == 4) THEN
gpse.f90:          ELSE IF (psiOutput == 5) THEN
gpse.f90:              WRITE(UNIT=OUTPUT_UNIT, FMT=*) '#           psiOutput = ', psiOut
put
gpse.f90:      CALL MPI_BCAST(psiOutput, 1, mpiInt, mpiMaster, MPI_COMM_WORLD, m
piError)
psi.f90:      INTEGER, PUBLIC :: psiOutput = -1
psi.f90:      psiOutput is a PUBLIC, INTEGER-valued variable that sets the file
[mkandes@login01 source]$
```

Simple parallel I/O - one file per process

```
mkandes@login01:~/software/gpse/0.6.5/mpi-omp/gcc/10.2.0/openmpi/4.0.4/gpse/s... END DO

psiFileNo = psiFileNo + 1

! Use simple parallel I/O to write out the complete wave
! function simultaneously to many legacy VTK files that
! only contain a single, one-dimensional slab of the wave
! function overseen by an MPI process
ELSE IF (psiOutput == 3) THEN

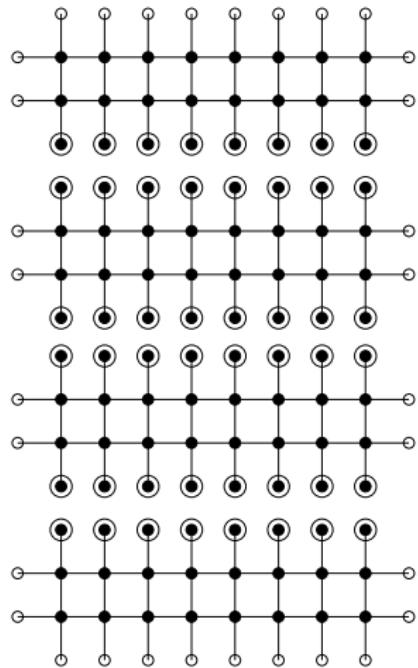
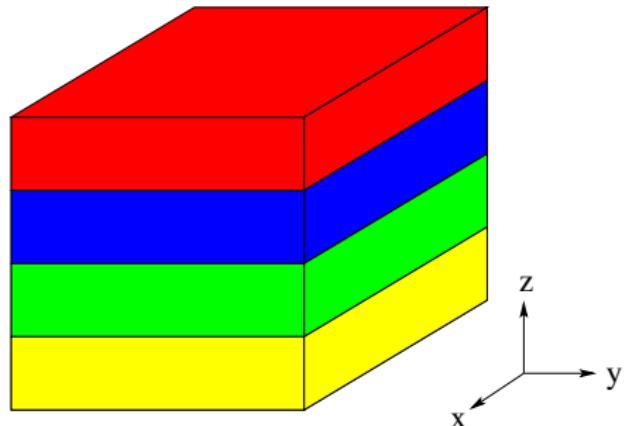
    CALL io_write_vtk('psi-', psiFileNo, mpiRank, nX, &
                      & nXa, nXb, nXbc, dNx, nY, nYa, nYb, nYbc, dNy, &
                      & nZ, nZa, nZb, nZbc, dNz, Xa, Ya, Za, Psi3a)
    psiFileNo = psiFileNo + 1

! Use MPI-I/O to write the wave function out to a single
! binary file in parallel
ELSE IF (psiOutput == 4) THEN

    WRITE(UNIT=fileUnitChar, FMT='(I4.4)') psiFileNo
    CALL MPI_FILE_OPEN ( MPI_COMM_WORLD, &
                        & TRIM ( 'psi-'//fileUnitChar//'.bin' ), &
                        & MPI_MODE_CREATE + MPI_MODE_WRONLY, &

1155,25      53%
```

1D (Slab) Domain Decomposition



Simulation data - one zip archive per timestep

```
[mkandes@login02 ~]$ cd /expanse/lustre/projects/use300/mkandes/data/compton/gpse-...
[mkandes@login02 compton]$ ls
gpse-v0.5.9-comet-compton-20170828-001  gpse-v0.5.9-comet-compton-20170828-012
gpse-v0.5.9-comet-compton-20170828-009  gpse-v0.5.9-comet-compton-20170828-013
[mkandes@login02 compton]$ cd gpse-v0.5.9-comet-compton-20170828-001
[mkandes@login02 gpse-v0.5.9-comet-compton-20170828-001]$ ls
0143      psi-1048.zip  psi-1097.zip  psi-1145.zip  psi-1194.zip
psi-1000.zip  psi-1049.zip  psi-1098.zip  psi-1146.zip  psi-1195.zip
psi-1001.zip  psi-1050.zip  psi-1099.zip  psi-1147.zip  psi-1196.zip
psi-1002.zip  psi-1051.zip  psi-1100.zip  psi-1148.zip  psi-1197.zip
psi-1003.zip  psi-1052.zip  psi-1101.zip  psi-1149.zip  psi-1198.zip
psi-1004.zip  psi-1053.zip  psi-1102.zip  psi-1150.zip  psi-1199.zip
psi-1005.zip  psi-1054.zip  psi-1103.zip  psi-1151.zip  psi-1200.zip
psi-1006.zip  psi-1055.zip  psi-1104.zip  psi-1152.zip  psi-1201.zip
psi-1007.zip  psi-1056.zip  psi-1105.zip  psi-1153.zip  psi-1202.zip
psi-1008.zip  psi-1057.zip  psi-1106.zip  psi-1154.zip  psi-1203.zip
psi-1009.zip  psi-1058.zip  psi-1107.zip  psi-1155.zip  psi-1204.zip
psi-1010.zip  psi-1059.zip  psi-1108.zip  psi-1156.zip  psi-1205.zip
psi-1011.zip  psi-1060.zip  psi-1109.zip  psi-1157.zip  psi-1206.zip
psi-1012.zip  psi-1061.zip  psi-1110.zip  psi-1158.zip  psi-1207.zip
psi-1013.zip  psi-1062.zip  psi-1111.zip  psi-1159.zip  psi-1208.zip
psi-1014.zip  psi-1063.zip  psi-1112.zip  psi-1160.zip  psi-1209.zip
psi-1015.zip  psi-1064.zip  psi-1113.zip  psi-1161.zip  psi-1210.zip
psi-1016.zip  psi-1065.zip  psi-1114.zip  psi-1162.zip  psi-1211.zip
```

zip

```
mkaned@login02:~ ZIP(1L) ZIP(1L)

NAME
    zip - package and compress (archive) files

SYNOPSIS
    zip [-aABcdDeEfFghjklLmoqrRSTuvVwXyz!@$] [--longoption ...] [-b path]
        [-n suffixes] [-t date] [-tt date] [zipfile [file ...]] [-xi list]

    zipcloak (see separate man page)

    zipnote (see separate man page)

    zipsplit (see separate man page)

    Note: Command line processing in zip has been changed to support long
          options and handle all options and arguments more consistently. Some
          old command lines that depend on command line inconsistencies may no
          longer work.

DESCRIPTION
    zip is a compression and file packaging utility for Unix, VMS, MSDOS,
    OS/2, Windows 9x/NT/XP, Minix, Atari, Macintosh, Amiga, and Acorn RISC
    Manual page zip(1) line 1 (press h for help or q to quit)
```

unzip

```
mkaned@login02:~ UNZIP(1L) UNZIP(1L)

UNZIP(1L)
NAME
    unzip - list, test and extract compressed files in a ZIP archive

SYNOPSIS
    unzip [ -Z ] [ -cflptTuvz[abjnoqsCDKLMUVWX$/:^] ] file[.zip] [file(s) ...]
    [ -x xfile(s) ... ] [ -d exdir]

DESCRIPTION
    unzip will list, test, or extract files from a ZIP archive, commonly
    found on MS-DOS systems. The default behavior (with no options) is to
    extract into the current directory (and subdirectories below it) all
    files from the specified ZIP archive. A companion program, zip(1L),
    creates ZIP archives; both programs are compatible with archives cre-
    ated by PKWARE's PKZIP and PKUNZIP for MS-DOS, but in many cases the
    program options or default behaviors differ.

ARGUMENTS
    file[.zip]
        Path of the ZIP archive(s). If the file specification is a
        wildcard, each matching file is processed in an order determined
        by the operating system (or file system). Only the filename can
```

Manual page unzip(1) line 1 (press h for help or q to quit)

Simulation data - one vtk file per slab (process)

```
mkandes@login02:/expanse/lustre/projects/use300/mkandes/data/compton/gpse-...
psi-1033.zip psi-1082.zip psi-1131.zip psi-1179.zip psi-1228.zip
psi-1034.zip psi-1083.zip psi-1132.zip psi-1180.zip psi-1229.zip
psi-1035.zip psi-1084.zip psi-1133.zip psi-1181.zip psi-1230.zip
psi-1036.zip psi-1085.zip psi-1134.zip psi-1182.zip psi-1231.zip
psi-1037.zip psi-1086.zip psi-1134.zip psi-1183.zip psi-1232.zip
psi-1038.zip psi-1087.zip psi-1135.zip psi-1184.zip psi-1233.zip
psi-1039.zip psi-1088.zip psi-1136.zip psi-1185.zip psi-1234.zip
psi-1040.zip psi-1089.zip psi-1137.zip psi-1186.zip psi-1235.zip
psi-1041.zip psi-1090.zip psi-1138.zip psi-1187.zip psi-1236.zip
psi-1042.zip psi-1091.zip psi-1139.zip psi-1188.zip psi-1237.zip
psi-1043.zip psi-1092.zip psi-1140.zip psi-1189.zip psi-1238.zip
psi-1044.zip psi-1093.zip psi-1141.zip psi-1190.zip psi-1239.zip
psi-1045.zip psi-1094.zip psi-1142.zip psi-1191.zip psi-1240.zip
psi-1046.zip psi-1095.zip psi-1143.zip psi-1192.zip
psi-1047.zip psi-1096.zip psi-1144.zip psi-1193.zip
[mkandes@login02 gpse-v0.5.9-comet-compton-20170828-001]$ cd psi-1134/
[mkandes@login02 psi-1134]$ ls
psi-1134-0000.vtk psi-1134-0105.vtk psi-1134-0210.vtk psi-1134_124.vtu
psi-1134-0001.vtk psi-1134-0106.vtk psi-1134-0211.vtk psi-1134_125.vtu
psi-1134-0002.vtk psi-1134-0107.vtk psi-1134-0212.vtk psi-1134_126.vtu
psi-1134-0003.vtk psi-1134-0108.vtk psi-1134-0213.vtk psi-1134_127.vtu
psi-1134-0004.vtk psi-1134-0109.vtk psi-1134-0214.vtk psi-1134_128.vtu
psi-1134-0005.vtk psi-1134-0110.vtk psi-1134-0215.vtk psi-1134_129.vtu
psi-1134-0006.vtk psi-1134-0111.vtk psi-1134-0216.vtk psi-1134_130.vtu
```

Simulation data - vtk file format

```
mkandes@login02:/expanse/lustre/projects/use300/mkandes/data/compton/gpse-...  
psi-1134-0103.vtk  psi-1134-0208.vtk  psi-1134_122.vtu  
psi-1134-0104.vtk  psi-1134-0209.vtk  psi-1134_123.vtu  
[mkandes@login02 psi-1134]$ head -20 psi-1134-0197.vtk  
# vtk DataFile Version 3.0  
STANDARD LEGACY VTK FORMAT  
ASCII  
DATASET RECTILINEAR_GRID  
DIMENSIONS 1152 1152 2  
X_COORDINATES 1152 double  
-15.9861111066350  
-15.9583333288650  
-15.9305555510950  
-15.9027777733250  
-15.8749999955550  
-15.8472222177850  
-15.8194444400150  
-15.7916666622450  
-15.7638888844750  
-15.7361111067050  
-15.7083333289350  
-15.6805555511650  
-15.6527777733950  
-15.6249999956250  
[mkandes@login02 psi-1134]$
```

vtk vs. vtu

```
mkanedes@login02:/expanse/lustre/projects/use300/mkanedes/data/compton/gpse-...
```

| File Type | User | File Name | Size | Date | File ID |
|------------|------------|-----------|------|-------------|-------------------|
| -rw-r--r-- | 1 mkanedes | use300 | 127M | Aug 29 2017 | psi-1134-0039.vtk |
| -rw-r--r-- | 1 mkanedes | use300 | 127M | Aug 29 2017 | psi-1134-0132.vtk |
| -rw-r--r-- | 1 mkanedes | use300 | 127M | Aug 29 2017 | psi-1134-0069.vtk |
| -rw-r--r-- | 1 mkanedes | use300 | 127M | Aug 29 2017 | psi-1134-0044.vtk |
| -rw-r--r-- | 1 mkanedes | use300 | 127M | Aug 29 2017 | psi-1134-0041.vtk |
| -rw-r--r-- | 1 mkanedes | use300 | 127M | Aug 29 2017 | psi-1134-0036.vtk |
| -rw-r--r-- | 1 mkanedes | use300 | 127M | Aug 29 2017 | psi-1134-0061.vtk |
| -rw-r--r-- | 1 mkanedes | use300 | 127M | Aug 29 2017 | psi-1134-0042.vtk |
| -rw-r--r-- | 1 mkanedes | use300 | 127M | Aug 29 2017 | psi-1134-0035.vtk |
| -rw-r--r-- | 1 mkanedes | use300 | 127M | Aug 29 2017 | psi-1134-0040.vtk |
| -rw-r--r-- | 1 mkanedes | use300 | 127M | Aug 29 2017 | psi-1134-0037.vtk |
| -rw-r--r-- | 1 mkanedes | use300 | 127M | Aug 29 2017 | psi-1134-0045.vtk |
| drwxr-sr-x | 4 mkanedes | use300 | 111K | Jul 9 2021 | .. |
| -rw-r--r-- | 1 mkanedes | use300 | 531M | Jul 9 2021 | psi-1134_98.vtu |
| -rw-r--r-- | 1 mkanedes | use300 | 531M | Jul 9 2021 | psi-1134_77.vtu |
| -rw-r--r-- | 1 mkanedes | use300 | 520M | Jul 9 2021 | psi-1134_59.vtu |
| -rw-r--r-- | 1 mkanedes | use300 | 528M | Jul 9 2021 | psi-1134_55.vtu |
| -rw-r--r-- | 1 mkanedes | use300 | 523M | Jul 9 2021 | psi-1134_4.vtu |
| -rw-r--r-- | 1 mkanedes | use300 | 523M | Jul 9 2021 | psi-1134_35.vtu |
| -rw-r--r-- | 1 mkanedes | use300 | 523M | Jul 9 2021 | psi-1134_23.vtu |
| -rw-r--r-- | 1 mkanedes | use300 | 516M | Jul 9 2021 | psi-1134_18.vtu |
| -rw-r--r-- | 1 mkanedes | use300 | 531M | Jul 9 2021 | psi-1134_102.vtu |
| -rw-r--r-- | 1 mkanedes | use300 | 531M | Jul 9 2021 | psi-1134_96.vtu |
| -rw-r--r-- | 1 mkanedes | use300 | 528M | Jul 9 2021 | psi-1134_95.vtu |

Visualization data - vtu file format

```
[mkandes@login02 psi-1134]$ head -11 psi-1134_98.vtu
<VTKFile type="UnstructuredGrid" version="2.2" byte_order="LittleEndian" header_
type="UInt64">
  <UnstructuredGrid>
    <Piece NumberOfPoints="6223392" NumberOfCells="3069504"
      >
        <PointData Scalars="RePsi" GlobalIds="__D3__GlobalNodeIds">
          <DataArray type="Float64" Name="ImPsi" format="appended" RangeMin="-0.00
12450472664" RangeMax="0.00065853427111" offset="0" />
          <DataArray type="Float64" Name="RePsi" format="appended" RangeMin="-0.00
093535214815" RangeMax="0.001508906848" offset="49787144" />
          <DataArray type="Int64" IdType="1" Name="__D3__GlobalNodeIds" format=
appended" RangeMin="412086382" RangeMax="686755584" offset
="99574288" />
          <DataArray type="UInt8" Name="vtkGhostType" format="appended" RangeMin=
0" RangeMax="1" offset="149361432" />
        </PointData>
        <CellData>
          <DataArray type="UInt8" Name="vtkGhostType" format="appended" RangeMin=
0" RangeMax="1" offset="155584832" />
        </CellData>
      </Piece>
    </UnstructuredGrid>
  </VTKFile>
[mkandes@login02 psi-1134]$
```

Simulation data storage requirements

```
mkandes@login02:expanse/lustre/projects/use300/mkandes/data/compton
psi-1134-0103.vtk  psi-1134-0208.vtk  psi-1134_122.vtu
psi-1134-0104.vtk  psi-1134-0209.vtk  psi-1134_123.vtu
[mkandes@login02 psi-1134]$ pwd
/expanse/lustre/projects/use300/mkandes/data/compton/gpse-v0.5.9-comet-compton-2
0170828-001/psi-1134
[mkandes@login02 psi-1134]$ du -h
113G .
[mkandes@login02 psi-1134]$ cd ../
[mkandes@login02 gpse-v0.5.9-comet-compton-20170828-001]$ du -h
113G ./psi-1134
20G ./0143
3.3T .
[mkandes@login02 gpse-v0.5.9-comet-compton-20170828-001]$ cd ../
[mkandes@login02 compton]$ du -h
3.2T ./gpse-v0.5.9-comet-compton-20170828-012
113G ./gpse-v0.5.9-comet-compton-20170828-001/psi-1134
20G ./gpse-v0.5.9-comet-compton-20170828-001/0143
3.3T ./gpse-v0.5.9-comet-compton-20170828-001
4.4G ./gpse-v0.5.9-comet-compton-20170828-013/psi-1000
1.6T ./gpse-v0.5.9-comet-compton-20170828-013
20G ./gpse-v0.5.9-comet-compton-20170828-009/0143
3.2T ./gpse-v0.5.9-comet-compton-20170828-009
12T .
[mkandes@login02 compton]$
```

Expanse Lustre Filesystem - /expanse/lustre/projects

```
mkandes@login02:~/expanse/lustre/projects/csd403/mkandes
[mkandes@login02 -]$ df -Th | grep lustre
10.22.101.123:0z1b:10.22.101.124@021b:/expanse/scratch  lustre      9.2P  2.0P  7.3P  21% /expanse/lustre/scratch
10.22.101.123:0z1b:10.22.101.124@021b:/expanse/projects  lustre      9.2P  2.0P  7.3P  21% /expanse/lustre/projects
[mkandes@login02 -]$ expanse-client user -p

Resource  expanse

NAME      STATE   PROJECT   TG PROJECT      USED   AVAILABLE   USED BY PROJECT
-----
mkandes  allow    csd403   TG-IBN140002     117    13712000      9136788
mkandes  allow    sds166   TG-STA160003      8     100000       56435
mkandes  allow    sds184   TG-TRA210003      1      5000        1308
mkandes  allow    use300          1907552    5050000      3432564
[mkandes@login02 -]$ cd /expanse/lustre/projects/use300/mkandes
[mkandes@login02 mkandes]$ ls
backups  containers  data  tickets
[mkandes@login02 mkandes]$ cd /expanse/lustre/projects/csd403/mkandes
[mkandes@login02 mkandes]$ ls
nemar
[mkandes@login02 mkandes]$ du -h nemar/
96K      nemar/.ipynb_checkpoints
^C
[mkandes@login02 mkandes]$
```

- ▶ Usage: Parallel (MPI) I/O, store large input/output files
- ▶ Quota: 500 GB per project/group (default)
- ▶ Backup: None
- ▶ Purge: 90 days after allocation expiration

Murphy's Law - Lustre Edition

User News - XSEDE

https://www.xsede.org/news/user-news/-/news/item/14031

Github - mkandes Github - sdsc Email - UCSD User Tickets - XSEDE XDCDB Admin - XSEDE ZenDesk - SDSC Splunk - SDSC SDSC - FORGE

SDSC Expanse: Lustre Filesystem Status Update 8 - May 16, 2022

Outage start 04/25/2022 08:00 PDT
Anticipated end 04/28/2022 20:00 PDT
Outage type Partial outage

Update 8
Posted by Mahidhar Tatineni on 05/17/2022 07:02 UTC
During today's Expanse maintenance we were able to mount 4 out of the 7 Object Storage Targets (OSTs) that have been inaccessible since our Lustre maintenance. This will make files on the recovered OSTs accessible to users. With this recovery, 69 out of the 72 OSTs on the system are now available. We continue to work to restore the remaining 3 OSTs. Please email help@xsede.org if you have any questions.

Update 7
Posted by Mahidhar Tatineni on 05/06/2022 00:58 UTC
We are continuing work to restore Lustre OST targets on Expanse. Since our last update we took one more OST offline (total of 7 offline right now). At present, working with our filesystem vendor, the OSTs with issues have been isolated from the rest of the system and we are going through recovery attempts. Currently, we do not have an estimate time for completion or know if there is any data loss given the complexity of the recovery, but will post updates as we have information.

The Lustre filesystem is available for use with the 65 OSTs that are online. Users are requested to check their input files and if they are inaccessible, setup a new directory with fresh inputs to continue their runs. Any new simulations and new data created will not use the problem OSTs and are expected to complete successfully. Similarly Globus transfers from directories with inaccessible files will fail and will require a fresh directory.

For ALL future jobs that need the Lustre filesystem please add:

```
#SBATCH --constraint="lustre"
```

to your scripts. This constraint will help us better manage any future Lustre

[Login to manage subscriptions](#)

Search user news

Select categories

- Anvil (Purdue)
- Bridges-2 (PSC)
- Bridges (PSC)
- Comet (SDSC)
- Delta (NCSA)

Keywords

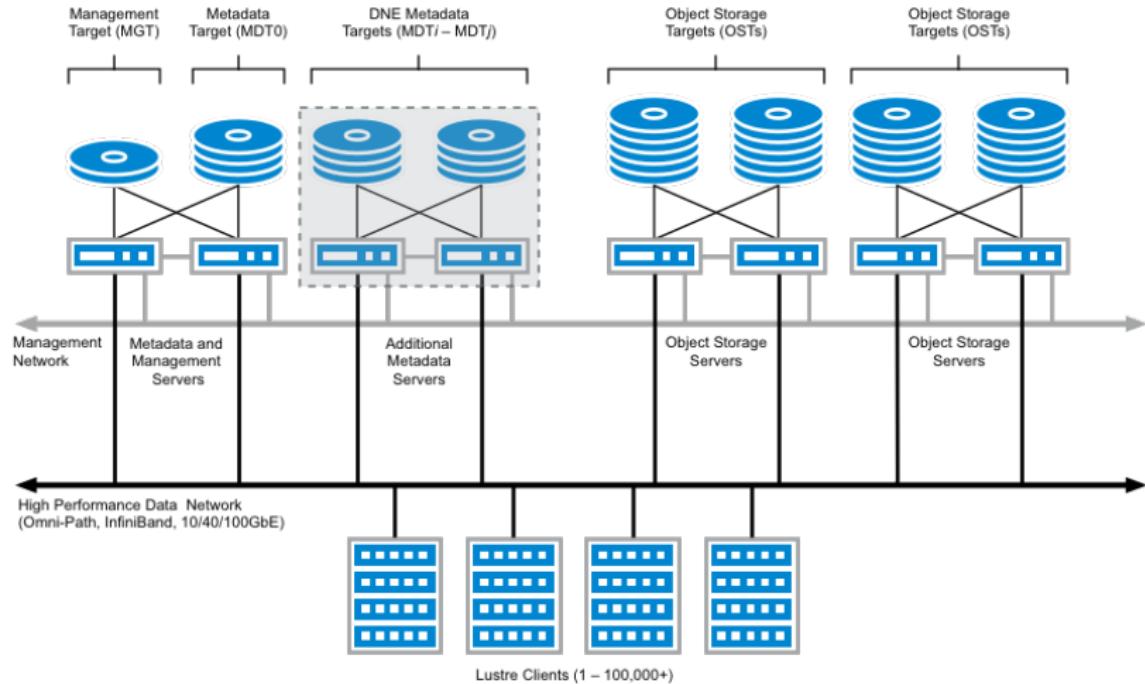
News item types

- Announcement
- Event
- Outage

Contact Information

XSEDE Helpdesk

Lustre Filesystem Architecture



lfs

```
mkandes@login01:~ mkandes@login01:~  
lfs(1) user utilities lfs(1)  
  
NAME  
lfs - client utility for Lustre-specific file layout and other  
attributes  
  
SYNOPSIS  
lfs changelog [--follow] <mdtname> [startrec [endrec]]  
lfs changelog_clear <mdtname> <id> <endrec>  
lfs check <mds|osts|servers>  
lfs data_version [-nrw] <filename>  
lfs df [-ihlv] [--pool|-p <fsname>[.<pool>]] [path]  
lfs fid2path [--link <linkno>] <fsname|rootpath> <fid> ...  
lfs find <directory>|<filename> ...  
    [[!] --atime|-A [-+]n] [[!] --ctime|-C [-+]n]  
    [[!] --mtime|-M [-+]n]  
    [[!] --component-count|--comp-count [+]-n]  
    [[!] --component-flags|--comp-flags <[^]flag,...>]  
    [[!] --component-end|--comp-end|-E [+]-n[KMGTPe]]  
    [[!] --component-start|--comp-start [+]-n[KMGTPe]]  
    [[!] --mirror-count|-N [+]-n]  
    [[!] --mirror-state <[^]state>]  
    [[!] --gid|-g|--group|-G <gname>|<gid>]  
Manual page lfs(1) line 1 (press h for help or q to quit)
```

lfs quota (by group)

```
[mkandes@login01 ~]$ lfs quota -g use300 -h /expanse/lustre/projects/
Disk quotas for grp use300 (gid 300):
  Filesystem    used   quota   limit   grace   files   quota   limit   grace
/expanse/lustre/projects/
        24.88T      0k     50T      - 16430928      0      0      -
gid 300 is using default file quota setting
[mkandes@login01 ~]$ lfs quota -g csd403 -h /expanse/lustre/projects/
Disk quotas for grp csd403 (gid 7511):
  Filesystem    used   quota   limit   grace   files   quota   limit   grace
/expanse/lustre/projects/
        134.7G      0k     50T      -  573504      0      0      -
gid 7511 is using default file quota setting
[mkandes@login01 ~]$ lfs quota -g sds184 -h /expanse/lustre/projects/
Disk quotas for grp sds184 (gid 11905):
  Filesystem    used   quota   limit   grace   files   quota   limit   grace
/expanse/lustre/projects/
        5.96G      0k     50T      -   60751      0      0      -
gid 11905 is using default file quota setting
[mkandes@login01 ~]$
```

lfs quota (by user)

```
[mkandes@login01 ~]$ lfs quota -u mkandes -h /expanse/lustre/projects/
Disk quotas for usr mkandes (uid 501506):
  Filesystem    used   quota   limit   grace   files   quota   limit   grace
/expanse/lustre/projects/
          16.61T      0k      0k      - 12575615      0      0      -
uid 501506 is using default block quota setting
uid 501506 is using default file quota setting
[mkandes@login01 ~]$ lfs quota -u mkandes -h /expanse/lustre/scratch/mkandes/temp_project/
Disk quotas for usr mkandes (uid 501506):
  Filesystem    used   quota   limit   grace   files   quota   limit   grace
/expanse/lustre/scratch/mkandes/temp_project/
          16.61T      0k      0k      - 12575615      0      0      -
uid 501506 is using default block quota setting
uid 501506 is using default file quota setting
[mkandes@login01 ~]$ █
```

Expanse Lustre Filesystem - /expanse/lustre/scratch

- ▶ Usage: Parallel (MPI) I/O, storing large (temporary) files
- ▶ Quota: Up to 10 TB per user (default)
- ▶ Backup: None
- ▶ Purge: 90 days after file creation

Data transfer script - GridFTP

```
mkandes@hardtack: ~/Dropbox/work/ucsd/sdsc/presentations/2022/sdsc/ciml/data...     
```

```
#!/usr/bin/env bash

declare -xr SOURCE_ENDPOINT='oasis-dm.sdsc.xsede.org:2811'
declare -xr SOURCE_DIR='/oasis/projects/nsf/sdull18/mkandes/gpse-v0.5.9-comet-compton-20170828-001'

declare -xr DESTINATION_ENDPOINT='gridftp.stampede2.tacc.xsede.org:2811'
declare -xr DESTINATION_DIR='/work/03216/mckandes/maverick/gpse-v0.5.9-comet-compton-20170828-001'

module load globus

for timestep in {1000..1024}; do
    globus-url-copy -vb -stripe -tcp-bs 8m -p 4 "gsiftp://${SOURCE_ENDPOINT}/${SOURCE_DIR}/psi-${timestep}.zip" "gsiftp://${DESTINATION_ENDPOINT}/${DESTINATION_DIR}/psi-${timestep}.zip"
done
```

1,1 Top

globus-url-copy

```
mkaned@login01:~
```

GLOBUS-URL-COPY(1) Grid Community Toolkit Manual GLOBUS-URL-COPY(1)

NAME
globus-url-copy - globus-url-copy

SYNOPSIS
globus-url-copy [options] SOURCE-URL DESTINATION-URL

DESCRIPTION
The **globus-url-copy** program is a command line tool for multi-protocol data movement. It supports gsiftp:// (GridFTP), ftp://, http://, https://, sshftp:// and file:/// protocol specifiers in the URL.

OPTIONS

- help, -usage
Print help.
- versions
Print the versions of all modules that this program uses
- c, -continue-on-error
Do not die after any errors. By default, program will exit after most errors.

Manual page globus-url-copy(1) line 1 (press h for help or q to quit)

Globus Toolkit

The screenshot shows a web browser window for the Globus Toolkit. The title bar reads "Toolkit | Globus". The address bar shows the URL "https://toolkit.globus.org". Below the address bar, there is a horizontal bar with several links: GitHub - mkandes, GitHub - sdsc, Email - UCSD, User Tickets - XSEDE, XDCDB Admin - XSEDE, ZenDesk - SDSC, Splunk - SDSC, SDSC - FORGE, and a "More" link. The main content area features the Globus logo (a blue 'g' inside a white cloud-like shape) and the text "globus". Below this is a large image of a small globe with a blue 'g' and a red 'T' on it, set against a dark background. The globe has a brown base with the text "GLOBUS TOOLKIT 1998-2018". To the right of the image, the text "Globus Toolkit is Retired" is displayed. Below this, a paragraph states: "After a two decade run, the **Globus Toolkit** is no longer available as a do-it-yourself distributed computing toolkit, but its spirit lives on in a mature, full-featured and easy to use service for research data management – [Globus.org!](#)". Three call-to-action buttons are shown below: "Start using Globus to move data (it's easy and free!)", "Get comprehensive services and support with a Globus subscription", and "Get guidance on how to transition from Globus Toolkit to Globus.org". The bottom of the page contains standard browser navigation icons.

Grid Community Toolkit

The screenshot shows a web browser window with the following details:

- Title Bar:** Overview # | Grid Community Forum
- Address Bar:** https://gridcf.org
- Toolbar:** Back, Forward, Stop, Refresh, Home, etc.
- Tab Bar:** GitHub - mkandes, GitHub - sdsc, Email - UCSD, User Tickets - XSEDE, XDCDB Admin - XSEDE, ZenDesk - SDSC, Splunk - SDSC, SDSC - FORGE, etc.
- Content Area:**
 - ## Grid Community Forum

Community-based support for core software packages in grid computing
 - ## Overview

The Grid Community Forum (GridCF) is a global community that provides support for core grid software.

Specifically, the GridCF is attempting to support a software stack christened the [Grid Community Toolkit \(GCT\)](#). The GCT is an open-source fork of the venerable [Globus Toolkit](#) created by the [Globus Alliance](#). The GCT is derived from the Globus Toolkit, but is not the Globus Toolkit. Further, the GridCF is not a part of the Globus Alliance.

The GridCF is a nascent organization: we are looking for energetic contributors across a broad range of technical skills. Check out our [governance doc](#) and [join us on GitHub!](#)
 - ## Objectives

The creation of the GridCF and the GCT fork of the Globus Toolkit was motivated by the announcement of the upcoming [end-of-support](#) of the Globus Toolkit in January 2018.

With this effort, we aim to:
- Bottom Navigation:** Back, Forward, Home, etc.

Globus

The screenshot shows the official website for Globus, a non-profit service. The header includes a search bar, a navigation bar with links like 'I Want To...', 'Pricing', 'Resources', 'Support', 'About', and 'Log In', and a toolbar with various browser extensions.

The main content features a large blue background with a network graph pattern. On the left, there's a stylized orange 'g' logo icon. The central message is "Introducing Flows" with the subtitle "Our new service that accelerates discovery". Below this, a horizontal row of circular icons represents different service components: Auth (User icon), Transfer (Briefcase icon), Web Form (Form icon), Describe (Document icon), Identifier (Folder icon), Share (User icon with a plus sign), and Search (Search icon). Below each icon is its corresponding service name: GET CREDENTIALS, TRANSFER DATA, USER INPUT, GET METADATA, MINT DOI, SET POLICY, and CREATE INDEX. A "LEARN MORE ABOUT FLOWS" button is located at the bottom of this section.

At the bottom, the tagline "Research data management simplified." is displayed, along with three prominent buttons: "TRANSFER" (with a briefcase icon), "SHARE" (with a user icon), and "BUILD" (with a hexagonal icon).

Globus Log In

g Log in using Globus +

https://auth.globus.org/p/login?client_id=b9ba3e72-768f-4ddb-952d-e0bb7305e2c7&scope=...

GitHub - mkandes GitHub - sdsc Email - UCSD User Tickets - XSEDE XDCDB Admin - XSEDE ZenDesk - SDSC Splunk - SDSC SDSC - FORGE >

 globus

Log in to use Globus Web App

Use your existing organizational login
e.g., university, national lab, facility, project

XSEDE

By selecting Continue, you agree to Globus [terms of service](#) and [privacy policy](#).

[Continue](#)



Globus uses CILogon to enable you to Log In from this organization. By clicking Continue, you agree to the [CILogon privacy policy](#) and you agree to share your username, email address, and affiliation with CILogon and Globus. You also agree for CILogon to issue a certificate that allows Globus to act on your behalf.

OR

 Sign in with Google  Sign in with ORCID iD

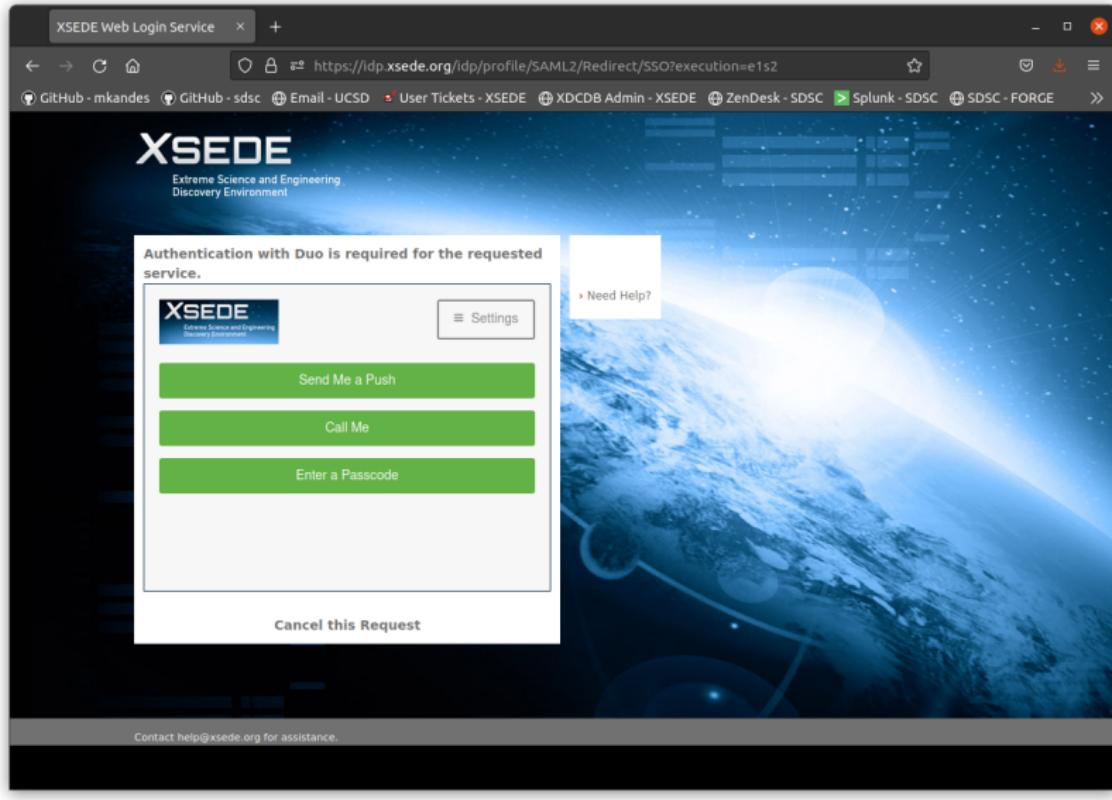
Didn't find your organization? Then use [Globus ID](#) to sign in. ([What's this?](#))

XSEDE CILogon

The screenshot shows a web browser window with the following details:

- Title Bar:** XSEDE Web Login Service
- Address Bar:** https://idp.xsede.org/idp/profile/SAML2/Redirect/SSO?execution=e1s1
- Toolbar:** GitHub - mkandes, GitHub - sdsc, Email - UCSD, User Tickets - XSEDE, XDCDB Admin - XSEDE, ZenDesk - SDSC, Splunk - SDSC, SDSC - FORGE
- Content Area:**
 - XSEDE Logo:** Extreme Science and Engineering Discovery Environment
 - CILogon Form:** Login to CILogon. It includes fields for XSEDE Username (mckandes) and XSEDE Password (redacted), a "Don't Remember Login" checkbox, and a red "Login" button.
 - CILogon Sidebar:** Features the CILogon logo and the text: "CILogon facilitates secure access to Cyberinfrastructure (CI)." It also includes links for "Sign Up For An Account," "Forgot your password?", and "Need Help?"
 - Contact Information:** At the bottom left, it says "Contact help@xsede.org for assistance."

XSEDE Duo Two-Factor Authentication



Globus File Manager

The screenshot shows the Globus File Manager interface running in a web browser window. The URL is <https://app.globus.org/file-manager>. The browser's address bar also lists other tabs: GitHub - mkandes, GitHub - sdsc, Email - UCSD, User Tickets - XSEDE, XDCDB Admin - XSEDE, ZenDesk - SDSC, Splunk - SDSC, SDSC - FORGE.

The main interface includes:

- Left Sidebar:** A vertical sidebar with icons and labels for: FILE MANAGER (selected), BOOKMARKS, ACTIVITY, COLLECTIONS, GROUPS, CONSOLE, FLOWS, ACCOUNT, LOGOUT, and HELP.
- Header:** A top navigation bar with a back/forward button, refresh, search, and a star icon for bookmarks.
- Toolbar:** A horizontal toolbar with "select all", "up one folder", "refresh list", and a "view" button.
- Search:** A search bar labeled "Collection" and "Path".
- Content Area:** A large central area with a placeholder message: "Search for a collection to begin" and a "Get started by taking a short tour." button.
- Right Panel:** A context menu or sidebar with the following options:
 - Share
 - Transfer or Sync to...
 - New Folder
 - Rename
 - Delete Selected
 - Download
 - Open
 - Upload
 - Get Link
 - Show Hidden Items
 - Manage Activation

Globus Collection Search

The screenshot shows the Globus Collection Search interface. On the left is a vertical sidebar with icons for Collection Search, File Manager, Bookmarks, Activity, Collections, Groups, Console, Flows, Account, Logout, and Help. The main area has a header with tabs for Collection Search and a search bar containing "Collection". Below the search bar is a placeholder text "Start typing the name of a data collection or select one below". Underneath are tabs for Recent, Bookmarks, Your Collections, and Shared With You. Two collections are listed: "XSEDE TACC stampede2" and "XSEDE Expanse". To the right of these lists is a section titled "Move Data To and From Your Computer" featuring an illustration of a laptop with a cloud icon and the text "Globus Connect Personal connects your laptop or other personal computer to Globus with just a few clicks. With Globus Connect Personal you can share and transfer files to/from a local machine – campus server, desktop computer or laptop – even if it's behind a firewall and you don't have administrator privileges." A "Get Globus Connect Personal" button with a help icon is also present.

Collection Search | Globus

https://app.globus.org/file-manager/collections

GitHub - mkandes GitHub - sdsc Email - UCSD User Tickets - XSEDE XDCDB Admin - XSEDE ZenDesk - SDSC Splunk - SDSC SDSC - FORGE

Collection

Start typing the name of a data collection or select one below

Recent Bookmarks Your Collections Shared With You

XSEDE TACC stampede2

XSEDE Expanse

Move Data To and From Your Computer

Globus Connect Personal connects your laptop or other personal computer to Globus with just a few clicks. With Globus Connect Personal you can share and transfer files to/from a local machine – campus server, desktop computer or laptop – even if it's behind a firewall and you don't have administrator privileges.

Get Globus Connect Personal ⓘ

XSEDE Expanse Globus Collection / Endpoint

The screenshot shows a web browser window titled "File Manager | Globus" with the URL https://app.globus.org/file-manager?origin_id=b256c034-1578-11eb-893e-0a5521ff3f4b. The browser has several tabs open, including GitHub, Email - UCSD, User Tickets - XSEDE, XDCDB Admin - XSEDE, ZenDesk - SDSC, Splunk - SDSC, and SDSC - FORGE.

The main interface is a "File Manager" for the "XSEDE Expanse" collection. On the left is a sidebar with icons for FILE MANAGER, BOOKMARKS, ACTIVITY, COLLECTIONS, GROUPS, CONSOLE, FLOWS, ACCOUNT, LOGOUT, and HELP. The main area shows a search bar with "XSEDE Expanse" and a path input field. Below these are buttons for "select all", "up one folder", "refresh list", and a "view" dropdown menu. A central message says "Please authenticate to access XSEDE Expanse" and provides instructions: "When you press the CONTINUE button below you will be redirected to the collection's login page. After logging in, you will be returned here." A blue "Continue" button is at the bottom of this message. To the right is a context menu with options: Share, Transfer or Sync to..., New Folder, Rename, Delete Selected, Download, Open, Upload, Get Link, Show Hidden Items, and Manage Activation.

XSEDE Client Authorization

XSEDE OAuth 1.0 Client +

https://oa4mp.xsede.org/oauth/authorize?oauth_token=myproxy%3Aoa4mp%2C201

GitHub - mckandes GitHub - sdsc Email - UCSD User Tickets - XSEDE XDCDB Admin - XSEDE ZenDesk - SDSC Splunk - SDSC SDSC - FORGE

XSEDE

Extreme Science and Engineering Discovery Environment

Welcome to the XSEDE Client Authorization Page

Client Access

The XSEDE Client below is requesting access to your XSEDE account. If you approve, please sign in with your XSEDE username and password.

Note: Only members of active XSEDE project allocations will be able to sign in on this page.

CLIENT INFORMATION

The XSEDE Client listed below is requesting access to your XSEDE account. If you approve, please sign in with your XSEDE username and password. For help or clarification on why you are presented with this screen please contact the XSEDE Help Desk.

Name: Globus
URL: http://www.globus.org/

Username: mckandes

>Password:

SIGN IN CANCEL

Please send any questions or comments about this site to help@xsede.org.



XSede Home Page

XSede Home Page

XSEDE Expanse - \$HOME

File Manager | Globus + https://app.globus.org/file-manager?origin_id=b256c034-1578-11eb-893e-0a5521ff3f4b&orig

Github - mkandes Github - sdsc Email - UCSD User Tickets - XSEDE XDCDB Admin - XSEDE ZenDesk - SDSC Splunk - SDSC SDSC - FORGE

File Manager

Collection XSEDE Expanse Path /~/

select all up one folder refresh list

| NAME | LAST MODIFIED | SIZE | |
|---------------------------------|---------------------|----------|---|
| benchmarks | 6/6/2022, 08:06 AM | — | > |
| cm | 6/8/2022, 02:24 PM | — | > |
| data | 6/26/2022, 01:04 PM | — | > |
| install-expanse-core-packages.R | 7/21/2021, 09:15 AM | 2.15 KB | |
| NAMD_2.14_Source.tar.gz | 5/13/2022, 11:29 AM | 57.75 MB | |
| projects | 5/24/2022, 01:45 PM | — | > |
| scripts | 5/24/2022, 01:19 PM | — | > |

Share Transfer or Sync to... New Folder Rename Delete Selected Download Open Upload Get Link Show Hidden Items Manage Activation

Panels

BOOKMARKS ACTIVITY COLLECTIONS GROUPS CONSOLE FLOWS ACCOUNT LOGOUT HELP

XSEDE Expanse - /expanse/lustre/projects

The screenshot shows a web-based file manager interface for XSEDE Expanse. The URL in the address bar is https://app.globus.org/file-manager?origin_id=b256c034-1578-11eb-893e-0a5521ff3f4b&orig. The main content area displays a list of four sub-directories under the path `/expanse/lustre/projects/use300/mkandes/data/compton/`. Each entry includes a folder icon, the directory name, the last modified date, and a 'view' button. To the right of the list is a vertical menu with options: Share, Transfer or Sync to..., New Folder, Rename, Delete Selected, Download, Open, Upload, Get Link, Show Hidden Items, and Manage Activation. On the left, a sidebar provides links to various XSEDE services: GitHub (mkandes, sdsc), Email (UCSD), User Tickets (XSEDE), XDCDB Admin (XSEDE), ZenDesk (SDSC), Splunk (SDSC), SDSC (FORGE), and a Logout link.

| NAME | LAST MODIFIED | SIZE | VIEW |
|----------------------------------------|--------------------|------|------|
| gpse-v0.5.9-comet-compton-20170828-001 | 7/9/2021, 01:26 PM | — | > |
| gpse-v0.5.9-comet-compton-20170828-009 | 7/8/2021, 08:48 PM | — | > |
| gpse-v0.5.9-comet-compton-20170828-012 | 7/8/2021, 04:03 PM | — | > |
| gpse-v0.5.9-comet-compton-20170828-013 | 7/8/2021, 05:26 PM | — | > |

XSEDE Expanse — XSEDE Stampede2

The screenshot shows the XSEDE Expanse File Manager interface running on the Stampede2 system. The main window title is "File Manager | Globus". The URL in the address bar is https://app.globus.org/file-manager?destination_id=ceea5ca0-89a9-11e7-a97f-22000a92523b. The browser tab bar includes links for GitHub (mkandas, sdsc), Email - UCSD, User Tickets - XSEDE, XDCDB Admin - XSEDE, ZenDesk - SDSC, Splunk - SDSC, and SDSC - FORGE.

The File Manager interface has the following components:

- Left Sidebar:** Includes icons for File Manager, Bookmarks, Activity, Collections, Groups, Console, Flows, Account, Logout, and Help.
- Top Bar:** Shows the collection name "XSEDE Expanse", path "/expanse/lustre/projects/use300/mkandas/data/comptc", and current directory "~/". It also features a "Transfer & Timer Options" dropdown and a "Start" button.
- Content Area:** Displays a list of folder entries under "gpse-v0.5.9-comet-compton-20" and a list of files on the right. A context menu is open over the first folder entry.
- Context Menu:** The menu items are: Share, Transfer or Sync to..., New Folder, Rename, Delete Selected, Download, Open, Upload, Get Link, and Show Hidden Items.
- Right Panel:** Shows a list of files with their details: build-gromacs-mpi.slurm, build-gromacs-mpiprime.slurm, build-namd-mpi-linux-x86_64.slurm, build-namd-multicore-linux-x86_64.slurm, build-namd-ofi-linux-x86_64.slurm, build-namd-verbs-linux-x86_64.slurm, and cleanup.sh.

Stampede2 Login Nodes - 2FA

```
mckandes@login3.stampede2:~ mckandes@hardtack:~$ ssh stampede2
The authenticity of host 'stampede2.tacc.utexas.edu (129.114.63.43)' can't be es-
tablished.
ECDSA key fingerprint is SHA256:SegC2YyyftiRpdwhXqNZE+15RyGeFSal4Vuz0HYJ5E8.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'stampede2.tacc.utexas.edu,129.114.63.43' (ECDSA) to
the list of known hosts.

To access the system:

1) If not using ssh-keys, please enter your TACC password at the password prompt
2) At the TACC Token prompt, enter your 6-digit code followed by <return>.

Password:
TACC Token Code:
Last login: Thu Aug 26 09:44:57 2021 from 208.58.214.56
-----
          Welcome to the Stampede2 Supercomputer
          Texas Advanced Computing Center, The University of Texas at Austin
-----
** Unauthorized use/access is prohibited. **

If you log on to this computer system, you acknowledge your awareness
of and concurrence with the UT Austin Acceptable Use Policy. The
```

Stampede2 Login Nodes - Allocations

```
mckandes@login3.stampede2:~
```

| with 80 cores per node for a single job. | | | | | | |
|------------------------------------------------|------------|------------|--------------|------------|------------|-------|
| ----- Project balances for user mckandes ----- | | | | | | |
| Name | Avail SUs | Expires | Name | Avail SUs | Expires | |
| TG-STA160003 | 88357 | 2022-08-03 | TG-IBN140002 | 63407 | 2022-12-31 | |
| TG-DDM160003 | 1630 | 2022-07-21 | | | | |
| ----- Disk quotas for user mckandes ----- | | | | | | |
| Disk | Usage (GB) | Limit | %Used | File Usage | Limit | %Used |
| /home1 | 2.8 | 10.0 | 27.64 | 3204 | 200000 | 1.60 |
| /work2 | 0.0 | 1024.0 | 0.00 | 4 | 3000000 | 0.00 |
| /scratch | 0.0 | 0.0 | 0.00 | 15 | 0 | 0.00 |

* Important: The original Stockyard filesystem (/work and then changed to /work_old) is no longer compatible with S2 lustre. Please submit a ticket for instructions on how to access any data remaining on the original Stockyard filesystem

Tip 105 (See "module help tacc_tips" for features or how to disable)

To bypass any alias (but not shell function) for "foo" do "\foo".

```
login3.stampede2(1001)$ echo $WORK
/work2/03216/mckandes/stampede2
login3.stampede2(1002)$
```

Stampede2 \$WORK Lustre Filesystem

```
mckandes@login3.stampede2:~
```

```
login3.stampede2(1007)$ echo $WORK
/work2/03216/mckandes/stampede2
login3.stampede2(1008)$ df -Th | grep work2
172.29.200.10@o2ib1172:172.29.200.11@o2ib1172:/work    lustre    6.8P  1.6P  5.2P  24% /work2
login3.stampede2(1009)$ lfs quota -u mckandes -h /work2
Disk quotas for usr mckandes (uid 825148):
  Filesystem    used   quota   limit   grace   files   quota   limit   grace
    /work2      16k     0k     1T      -       4      0 3000000      -
login3.stampede2(1010)$ lfs quota -g 03216 -h /work2
Permission denied.
login3.stampede2(1011)$ groups
G-816050 G-81782 G-803761 G-815975 G-817736 G-818260 G-818343 G-818468 G-821172 G-823567
login3.stampede2(1012)$ lfs quota -g G-816050 -h /work2
Disk quotas for grp G-816050 (gid 816050):
  Filesystem    used   quota   limit   grace   files   quota   limit   grace
    /work2      24k     0k     0k      -       6      0      0      -
gid 816050 is using default block quota setting
gid 816050 is using default file quota setting
login3.stampede2(1013)$
```

Globus File Transfer - Setup

The screenshot shows the Globus File Manager interface. The left sidebar contains navigation links: GitHub, File Manager, Bookmarks, Activity, Collections, Groups, Console, Flows, Account, Logout, and Help. The main area displays a file list under the collection "XSEDE Expanse". The file list includes:

- psi-1132.zip (7/8/2021, 09:00 PM, 15.06 GB)
- psi-1133.zip (7/8/2021, 09:40 PM, 15.05 GB)
- psi-1134 (7/9/2021, 07:23 PM, -)
- psi-1134.zip (7/8/2021, 09:38 PM, 15.05 GB) - selected
- psi-1135.zip (7/8/2021, 09:01 PM, 15.05 GB)
- psi-1136.zip (7/8/2021, 10:03 PM, 15.05 GB)
- psi-1137.zip

A context menu is open over the selected file "psi-1134.zip", listing options: Share, Transfer or Sync to..., New Folder, Rename, Delete Selected, Download, Open, Upload, Get Link, and Show Hidden Items. The message "This folder is empty." is displayed in the bottom right corner of the list area.

Globus File Transfer - Initiated

The screenshot shows the Globus File Manager interface. On the left is a sidebar with various icons: Cloud (File Manager), GitHub, Email, User Tickets, XDCDB Admin, ZenDesk, Splunk, SDSC, and FORGE. Below these are Bookmarks, Activity (with 1 notification), Collections, Groups, Console, Flows, Account, Logout, and Help.

The main area is titled "File Manager" and shows a "Collection" named "XSEDE Expanse". The path is listed as "/panse/lustre/projects/use300/mkandas/data/compton/". A green toast notification in the top right corner says "Transfer request submitted successfully" with a "View details" link.

The central workspace displays a list of files and folders:

| File/Folder | Last Modified | Size |
|--------------|--------------------|----------|
| psi-1132.zip | 7/8/2021, 09:00 PM | 15.06 GB |
| psi-1133.zip | 7/8/2021, 09:40 PM | 15.05 GB |
| psi-1134 | 7/9/2021, 07:23 PM | - |
| psi-1134.zip | 7/8/2021, 09:38 PM | 15.05 GB |
| psi-1135.zip | 7/8/2021, 09:01 PM | 15.05 GB |
| psi-1136.zip | 7/8/2021, 10:03 PM | 15.05 GB |
| psi-1137.zip | | |

A context menu is open over the "psi-1134.zip" file, listing options: Share, Transfer or Sync to..., New Folder, Rename, Delete Selected, Download, Open, Upload, Get Link, and Show Hidden Items. A message in the background of the menu says "This folder is empty."

The bottom of the screen features a navigation bar with icons for back, forward, search, and other file management functions.

Globus File Transfer - Details

The screenshot shows the Globus File Transfer interface with the following details:

Task Information:

- Task Label:** XSEDE Expanse to XSEDE TACC stampede2
- Source:** XSEDE Expanse
- Destination:** XSEDE TACC stampede2
- Task ID:** 3273f7d4-f617-11ec-835d-cd84b862b754
- Owner:** Martin Kandes (mckandes@xsede.org)
- Condition:** ACTIVE
- Requested:** 6/27/2022, 05:46 AM
- Deadline:** 6/28/2022, 05:46 AM

Transfer Settings:

- verify file integrity after transfer
- transfer is not encrypted
- overwriting all files on destination

File Transfer Summary:

| | |
|-------|------------------------|
| 1 | Files |
| 0 | Directories |
| 0 B | Bytes Transferred |
| 0 B/s | Effective Speed |
| 0 | Skipped files on sync |
| 0 | Skipped files on error |

[View debug data](#)

Left Sidebar:

- FILE MANAGER
- BOOKMARKS (1)
- ACTIVITY (1)
- COLLECTIONS
- GROUPS
- CONSOLE
- FLOWS
- ACCOUNT
- LOGOUT
- HELP

Top Bar:

- Overview | Activity Detail
- https://app.globus.org/activity/3273f7d4-f617-11ec-835d-cd84b862b754/overview?back=file-transfer
- Github links: GitHub - mckandes, GitHub - sdsc, Email - UCSD, User Tickets - XSEDE, XDCDB Admin - XSEDE, ZenDesk - SDSC, Splunk - SDSC, SDSC - FORGE

Globus File Transfer - Completed

The screenshot shows a browser window for the Globus File Manager at the URL <https://app.globus.org/activity/3273f7d4-f617-11ec-835d-cd84b862b754/overview?back=file-transfer>. The main content area displays a successful transfer task:

XSEDE Expanse to XSEDE TACC stampede2
transfer completed

Task Label: XSEDE Expanse to XSEDE TACC stampede2

Source: XSEDE Expanse

Destination: XSEDE TACC stampede2

Task ID: 3273f7d4-f617-11ec-835d-cd84b862b754

Owner: Martin Kandes (mckandes@xsede.org)

Condition: SUCCEEDED

Requested: 6/27/2022, 05:46 AM

Completed: 6/27/2022, 05:47 AM

Duration: 1 minute 9 seconds

Transfer Settings:

- verify file integrity after transfer
- transfer is not encrypted
- overwriting all files on destination

On the right, a summary box provides the following statistics:

| | |
|-------------|------------------------|
| 1 | Files |
| 0 | Directories |
| 15.05 GB | Bytes Transferred |
| 219.28 MB/s | Effective Speed |
| 0 | Skipped files on sync |
| 0 | Skipped files on error |

[View debug data](#)

The left sidebar contains a navigation menu with the following items:

- FILE MANAGER (selected)
- BOOKMARKS
- ACTIVITY
- COLLECTIONS
- GROUPS
- CONSOLE
- FLOWS
- ACCOUNT
- LOGOUT
- HELP

Globus File Transfer - Event Log

The screenshot shows the Globus File Transfer interface with the title "Activity Detail | Activity" and a URL bar pointing to <https://app.globus.org/activity/3273f7d4-f617-11ec-835d-cd84b862b754/events?back=file-m...>. The top navigation bar includes links for GitHub, Email, User Tickets, XDCDB Admin, ZenDesk, Splunk, and SDSC.

The main area displays the "FILE MANAGER" section for a transfer from "XSEDE Expanse" to "XSEDE TACC stampede2". A green checkmark indicates "transfer completed".

The "Event Log" tab is selected, showing three entries:

- succeeded**: Occurred on 6/27/2022, 05:47 AM. The status was "succeeded". The details show a JSON object with "files_succeeded": 1.
- progress**: Occurred on 6/27/2022, 05:47 AM. The status was "progress". The details show a "View details" button.
- started**: Occurred on 6/27/2022, 05:46 AM. The status was "started". The details show a "View details" button.

The left sidebar contains a vertical menu with icons for FILE MANAGER, BOOKMARKS, ACTIVITY, COLLECTIONS, GROUPS, CONSOLE, FLOWS, ACCOUNT, LOGOUT, and HELP.

Globus Directory Transfer - Setup

The screenshot shows the Globus File Manager interface. On the left is a sidebar with various icons: Cloud, File Manager (selected), Bookmarks, Activity, Collections, Groups, Console, Flows, Account, Logout, and Help.

The main area displays two collections:

- XSEDE Expanse**: Path: /expanse/lustre/projects/use300/mkandas/data/comptc. Contains files: psi-1132.zip (7/8/2021, 09:00 PM, 15.06 GB), psi-1133.zip (7/8/2021, 09:40 PM, 15.05 GB), and a folder **psi-1134** (7/9/2021, 07:23 PM, -).
- XSEDE TACC stampede2**: Path: /work2/03216/mckandes/stampede2/. Contains a file: psi (7/8/2021, 05:47 AM, 15.05 GB).

A context menu is open over the **psi-1134** folder, listing options:

- Share
- Transfer or Sync to...
- New Folder
- Rename
- Delete Selected
- Download
- Open
- Upload
- Get Link
- Show Hidden Items

At the bottom of the interface, there are navigation icons for back, forward, search, and other file management functions.

Globus Directory Transfer - Queued

The screenshot shows a web browser window for the Globus platform at the URL <https://app.globus.org/activity/4247593e-f618-11ec-aede-6f7c2b57b05c/overview?back=file-transfer>. The title bar indicates the task is "task queued".

FILE MANAGER XSEDE Expanse to XSEDE TACC stampede2

Overview Event Log

Task Label XSEDE Expanse to XSEDE TACC stampede2

Source ▶ XSEDE Expanse

Destination ▶ XSEDE TACC stampede2

Task ID 4247593e-f618-11ec-aede-6f7c2b57b05c

Owner Martin Kandes (mckandes@xsede.org)

Condition ACTIVE

Requested 6/27/2022, 05:54 AM

Deadline 6/28/2022, 05:54 AM

Transfer Settings

- verify file integrity after transfer
- transfer is not encrypted
- overwriting all files on destination

Edit Label

Cancel Task

| | |
|-------|------------------------|
| 417 | Files |
| 1 | Directories |
| 0 B | Bytes Transferred |
| 0 B/s | Effective Speed |
| 0 | Skipped files on sync |
| 0 | Skipped files on error |

[View debug data](#)

Globus Directory Transfer - Completed

Overview | Activity Det... [+](#)

<https://app.globus.org/activity/4247593e-f618-11ec-aede-6f7c2b57b05c/overview?back=file-transfer>

[GitHub - mckandes](#) [GitHub - sdsc](#) [Email - UCSD](#) [User Tickets - XSEDE](#) [XDCDB Admin - XSEDE](#) [ZenDesk - SDSC](#) [Splunk - SDSC](#) [SDSC - FORGE](#) [»](#)

 FILE MANAGER  XSEDE Expanse to XSEDE TACC stampede2
transfer completed

 FILE MANAGER

 BOOKMARKS

 ACTIVITY

 COLLECTIONS

 GROUPS

 CONSOLE

 FLOWS

 ACCOUNT

 LOGOUT

 HELP

[Overview](#) [Event Log](#)

Task Label: XSEDE Expanse to XSEDE TACC stampede2

Source: XSEDE Expanse

Destination: XSEDE TACC stampede2

Task ID: 4247593e-f618-11ec-aede-6f7c2b57b05c

Owner: Martin Kandes (mckandes@xsede.org)

Condition: SUCCEEDED

Requested: 6/27/2022, 05:54 AM

Completed: 6/27/2022, 05:57 AM

Duration: 3 minutes 24 seconds

Transfer Settings:

- verify file integrity after transfer
- transfer is not encrypted
- overwriting all files on destination

| | |
|-------------|------------------------|
| 417 | Files |
| 1 | Directories |
| 108.85 GB | Bytes Transferred |
| 534.56 MB/s | Effective Speed |
| 0 | Skipped files on sync |
| 0 | Skipped files on error |

[View debug data](#)

Getting data to (and from) an HPC system

- ▶ Transfer: wget, curl, scp, sftp, rclone
- ▶ Archiving and Compression: gzip, gunzip, tar
- ▶ Integrity: md5sum, sha256sum

Example Dataset

https://ftp.ncbi.nlm.nih.gov/gene/DATA/gene_info.gz

wget

```
mkandes@login02:~ mkandes@login02:~ mkandes@login02:~  
WGET(1)          GNU Wget          WGET(1)  
  
NAME  
Wget - The non-interactive network downloader.  
  
SYNOPSIS  
wget [option]... [URL]...  
  
DESCRIPTION  
GNU Wget is a free utility for non-interactive download of files from  
the Web. It supports HTTP, HTTPS, and FTP protocols, as well as  
retrieval through HTTP proxies.  
  
Wget is non-interactive, meaning that it can work in the background,  
while the user is not logged on. This allows you to start a retrieval  
and disconnect from the system, letting Wget finish the work. By  
contrast, most of the Web browsers require constant user's presence,  
which can be a great hindrance when transferring a lot of data.  
  
Wget can follow links in HTML, XHTML, and CSS pages, to create local  
versions of remote web sites, fully recreating the directory structure  
of the original site. This is sometimes referred to as "recursive  
downloading." While doing that, Wget respects the Robot Exclusion  
Manual page wget(1) line 1 (press h for help or q to quit)
```

curl

```
mkandes@login02:~ curl(1)          Curl Manual          curl(1)

NAME
    curl - transfer a URL

SYNOPSIS
    curl [options / URLs]

DESCRIPTION
    curl is a tool to transfer data from or to a server, using one of the
    supported protocols (DICT, FILE, FTP, FTPS, GOPHER, HTTP, HTTPS, IMAP,
    IMAPS, LDAP, LDAPS, POP3, POP3S, RTMP, RTSP, SCP, SFTP, SMB, SMBS,
    SMTP, SMTPS, TELNET and TFTP). The command is designed to work without
    user interaction.

    curl offers a busload of useful tricks like proxy support, user authentication,
    FTP upload, HTTP post, SSL connections, cookies, file transfer resume,
    Metalink, and more. As you will see below, the number of features will make your head spin!

    curl is powered by libcurl for all transfer-related features. See
    libcurl(3) for details.

Manual page curl(1) line 1 (press h for help or q to quit)
```

scp

```
mkandes@login02:~
```

SCP(1) BSD General Commands Manual SCP(1)

NAME
scp – secure copy (remote file copy program)

SYNOPSIS
scp [~~-346BCpqqrTv~~] [-c cipher] [-F ssh_config] [-i identity_file]
[-J destination] [-l limit] [-o ssh_option] [-P port] [-S program]
source ... target

DESCRIPTION
scp copies files between hosts on a network. It uses ssh(1) for data transfer, and uses the same authentication and provides the same security as ssh(1). scp will ask for passwords or passphrases if they are needed for authentication.

The source and target may be specified as a local pathname, a remote host with optional path in the form [user@]host:[path], or a URI in the form scp://[user@]host[:port][/path]. Local file names can be made explicit using absolute or relative pathnames to avoid scp treating file names containing ':' as host specifiers.

When copying between two remote hosts, if the URI format is used, a port
Manual page scp(1) line 1 (press h for help or q to quit)

sftp

```
mkandes@login02:~
```

SFTP(1) BSD General Commands Manual SFTP(1)

NAME
sftp – secure file transfer program

SYNOPSIS
sftp [-46aCfpqrsv] [-B buffer_size] [-b batchfile] [-c cipher]
[-D sftp_server_path] [-F ssh_config] [-i identity_file]
[-J destination] [-l limit] [-o ssh_option] [-P port]
[-R num_requests] [-S program] [-s subsystem | sftp_server]
destination

DESCRIPTION
sftp is a file transfer program, similar to ftp(1), which performs all operations over an encrypted ssh(1) transport. It may also use many features of ssh, such as public key authentication and compression.

The destination may be specified either as [user@]host[:path] or as a URI in the form sftp://[user@]host[:port][/path].

If the destination includes a path and it is not a directory, sftp will retrieve files automatically if a non-interactive authentication method is used; otherwise it will do so after successful interactive authentication.

Manual page sftp(1) line 1 (press h for help or q to quit)

gzip

```
mkanedes@login02:~
```

GZIP(1) General Commands Manual GZIP(1)

NAME
gzip, gunzip, zcat - compress or expand files

SYNOPSIS
`gzip [-acdfhklLnNrtvV19] [-S suffix] [name ...]`
`gunzip [-acfhlLnNrtvV] [-S suffix] [name ...]`
`zcat [-fhLV] [name ...]`

DESCRIPTION
`Gzip` reduces the size of the named files using Lempel-Ziv coding (LZ77). Whenever possible, each file is replaced by one with the extension `.gz`, while keeping the same ownership modes, access and modification times. (The default extension is `z` for MSDOS, OS/2 FAT, Windows NT FAT and Atari.) If no files are specified, or if a file name is `"-"`, the standard input is compressed to the standard output. `Gzip` will only attempt to compress regular files. In particular, it will ignore symbolic links.

If the compressed file name is too long for its file system, `gzip` truncates it. `Gzip` attempts to truncate only the parts of the file name longer than 3 characters. (A part is delimited by dots.) If the name

Manual page gzip(1) line 1 (press h for help or q to quit)

tar

```
mkanedes@login02:~ mkandes@login02:~  
TAR(1)          GNU TAR Manual          TAR(1)  
  
NAME  
      tar - an archiving utility  
  
SYNOPSIS  
  Traditional usage  
      tar {A|c|d|r|t|u|x}[GnSkUW0mpsMBiajJzZhPlRvwo] [ARG...]  
  
  UNIX-style usage  
      tar -A [OPTIONS] ARCHIVE ARCHIVE  
  
      tar -c [-f ARCHIVE] [OPTIONS] [FILE...]  
  
      tar -d [-f ARCHIVE] [OPTIONS] [FILE...]  
  
      tar -t [-f ARCHIVE] [OPTIONS] [MEMBER...]  
  
      tar -r [-f ARCHIVE] [OPTIONS] [FILE...]  
  
      tar -u [-f ARCHIVE] [OPTIONS] [FILE...]  
  
      tar -x [-f ARCHIVE] [OPTIONS] [MEMBER...]  
Manual page tar(1) line 1 (press h for help or q to quit)
```

md5sum

```
mkandes@login02:~
```

MD5SUM(1) User Commands MD5SUM(1)

NAME
md5sum - compute and check MD5 message digest

SYNOPSIS
`md5sum [OPTION]... [FILE]...`

DESCRIPTION
Print or check MD5 (128-bit) checksums.
With no FILE, or when FILE is -, read standard input.

-b, --binary
read in binary mode

-c, --check
read MD5 sums from the FILES and check them

--tag create a BSD-style checksum

-t, --text
read in text mode (default)

Manual page md5sum(1) line 1 (press h for help or q to quit)

sha256sum

```
mkandes@login02:~
```

SHA256SUM(1) User Commands SHA256SUM(1)

NAME
sha256sum - compute and check SHA256 message digest

SYNOPSIS
`sha256sum [OPTION]... [FILE]...`

DESCRIPTION
Print or check SHA256 (256-bit) checksums.
With no FILE, or when FILE is -, read standard input.

-b, --binary
read in binary mode

-c, --check
read SHA256 sums from the FILES and check them

--tag create a BSD-style checksum

-t, --text
read in text mode (default)

Manual page sha256sum(1) line 1 (press h for help or q to quit)

Managing data on an HPC system

Use local /scratch for data intensive I/O

Reading, writing, and preparing data for HPC

Pre-process raw datasets into ML framework native formats

Preventing data loss

Back up your data regularly! Try `rsync`

Questions?

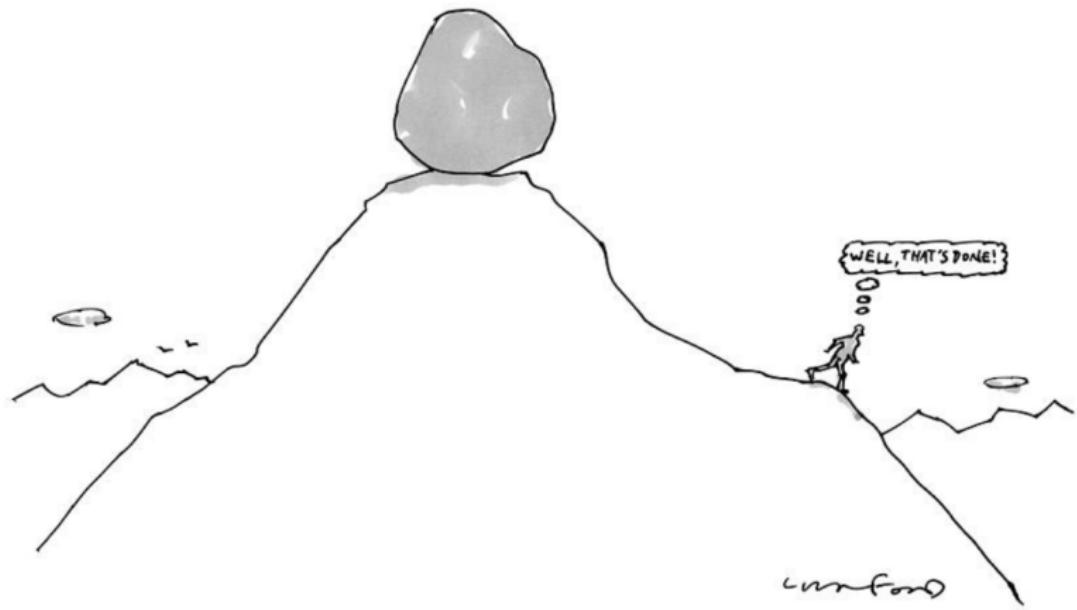


Image Credit: New Yorker - M. Crawford