



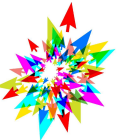
computecanada

New systems: common mistakes to avoid

by Alex Razoumov, WestGrid

Common mistakes: resources

- Production runs - only on compute nodes via the scheduler
 - do not run anything intensive on login nodes or directly on compute nodes
- Only request resources (memory, running time) needed
 - with a bit of a cushion, maybe 120% of the measured values
 - use Slurm command to estimate your completed code's memory usage
\$ sacct -j jobID --format=JobID,JobName,ReqMem,MaxRSS,Elapsed



Common mistakes: code efficiency

- Do not run unoptimized codes
 - use compilation flags -O2, -O3
- Programming language choice
 - compiled languages C/C++, F90, Chapel
 - interpreted or special runtime environment: Python, R, Java
 - use precompiled libraries



Common mistakes: files

- Do not create millions of tiny files
 - organize your code's output
 - use **tar** or even better **dar** (<http://dar.linux.free.fr>, supports indexing and differential archives)
- Do not store large data (more than a few MB) as ASCII files - waste of disk space and bandwidth
 - use a binary format
 - use scientific file formats (NetCDF, HDF5, etc): portability, headers, binary, compression, parallel
 - compress your files
- Use the right filesystem
- Do not leave files you no longer require, archive them elsewhere
- Learn and use parallel I/O

