

COSMA 101:

Demystifying COSMA

ICC Theory Lunch

13th May 2019

Alastair Basden

Purpose of these talks

- To improve capabilities at ICC
- Educate and establish best practices
- To answer questions that you're afraid to ask!
- Make efficient use of COSMA

Contents

- COSMA summary
- Login nodes
- Projects, groups, accounts
- Storage allocation
- Applying for time

What is COSMA?

- The COSmology MAchine
 - Now in its 7th iteration
 - The Memory Intensive service for DiRAC
 - STFC Distributed Research utilizing Advanced Computing

Cosma components

- Compute nodes
- Login nodes
- Data transfer nodes
- Web servers
- Archive servers
- Storage servers
 - Homepage, data, system
 - /snap7 is served by 22 servers (17 shared with /cosma7)
- Consoles/control machines
- Test servers
- VMs
- Analysis nodes

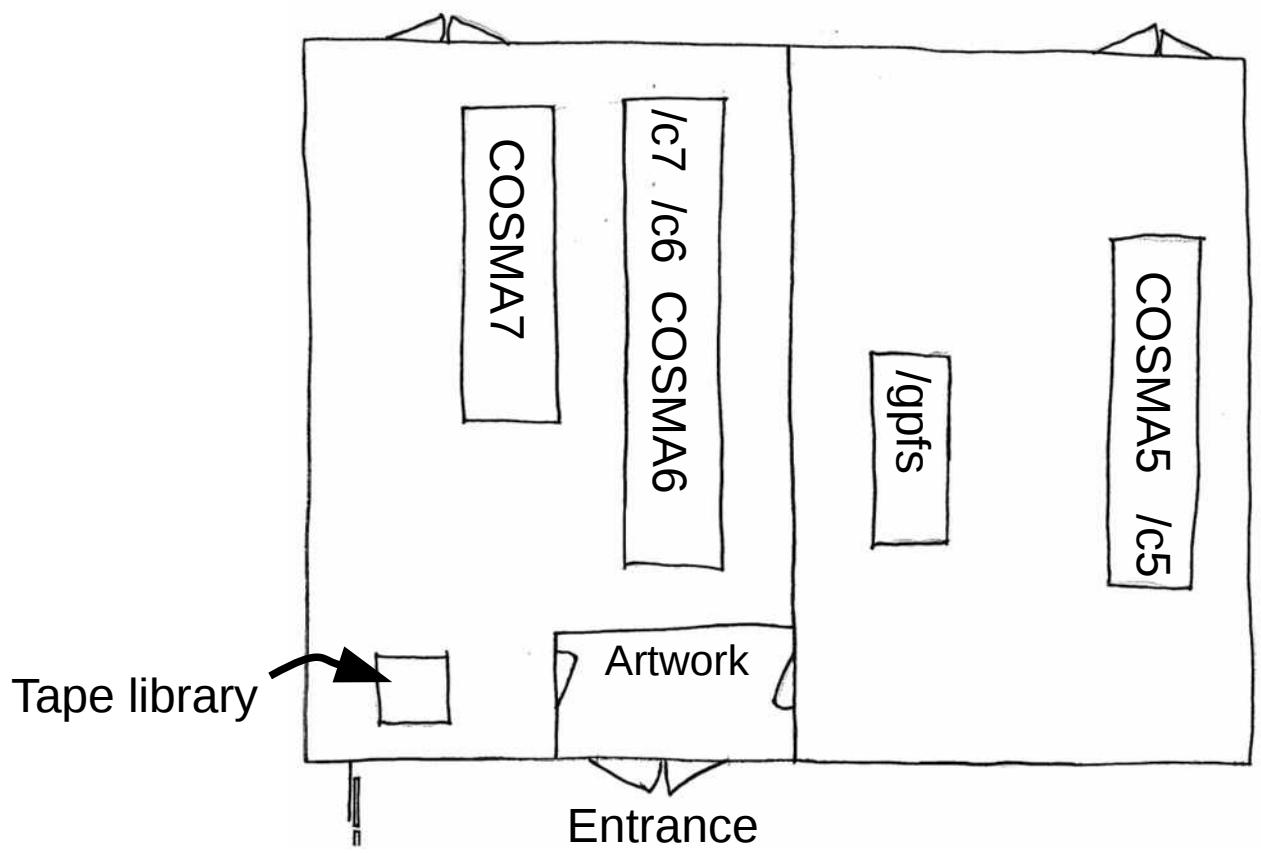
COSMA summary

- COSMA5 – 302 nodes, 16 cores, 128GB RAM
 - 2012
 - Was DiRAC, now ICC
- COSMA6 – 575 nodes, 16 cores, 128GB RAM
 - Identical hardware, gifted in 2016
- COSMA7 – 452 nodes, 28 cores, 512GB RAM
 - 2018/2019

COSMA location

- Arthur Holmes Data Centre





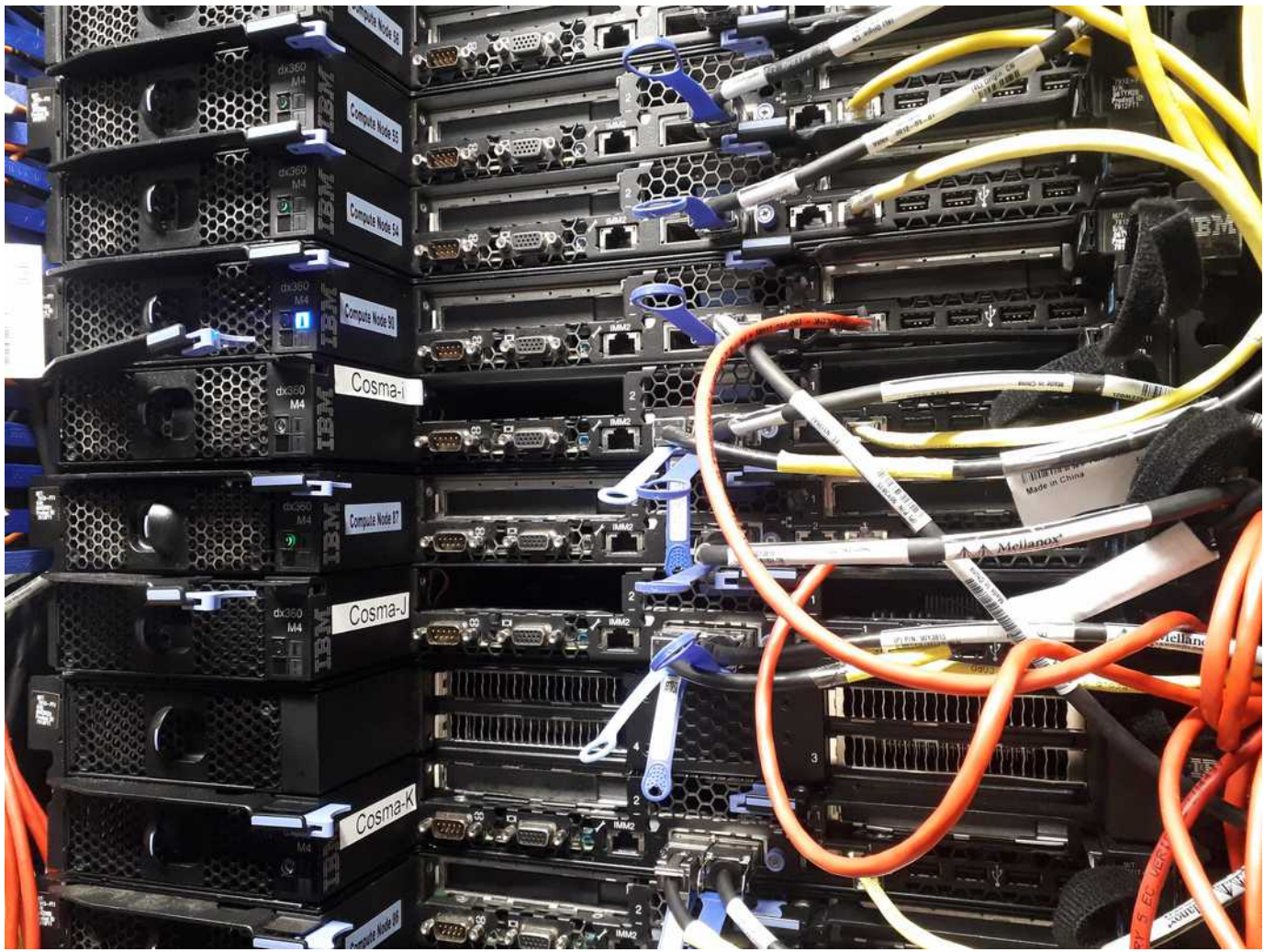
Login nodes

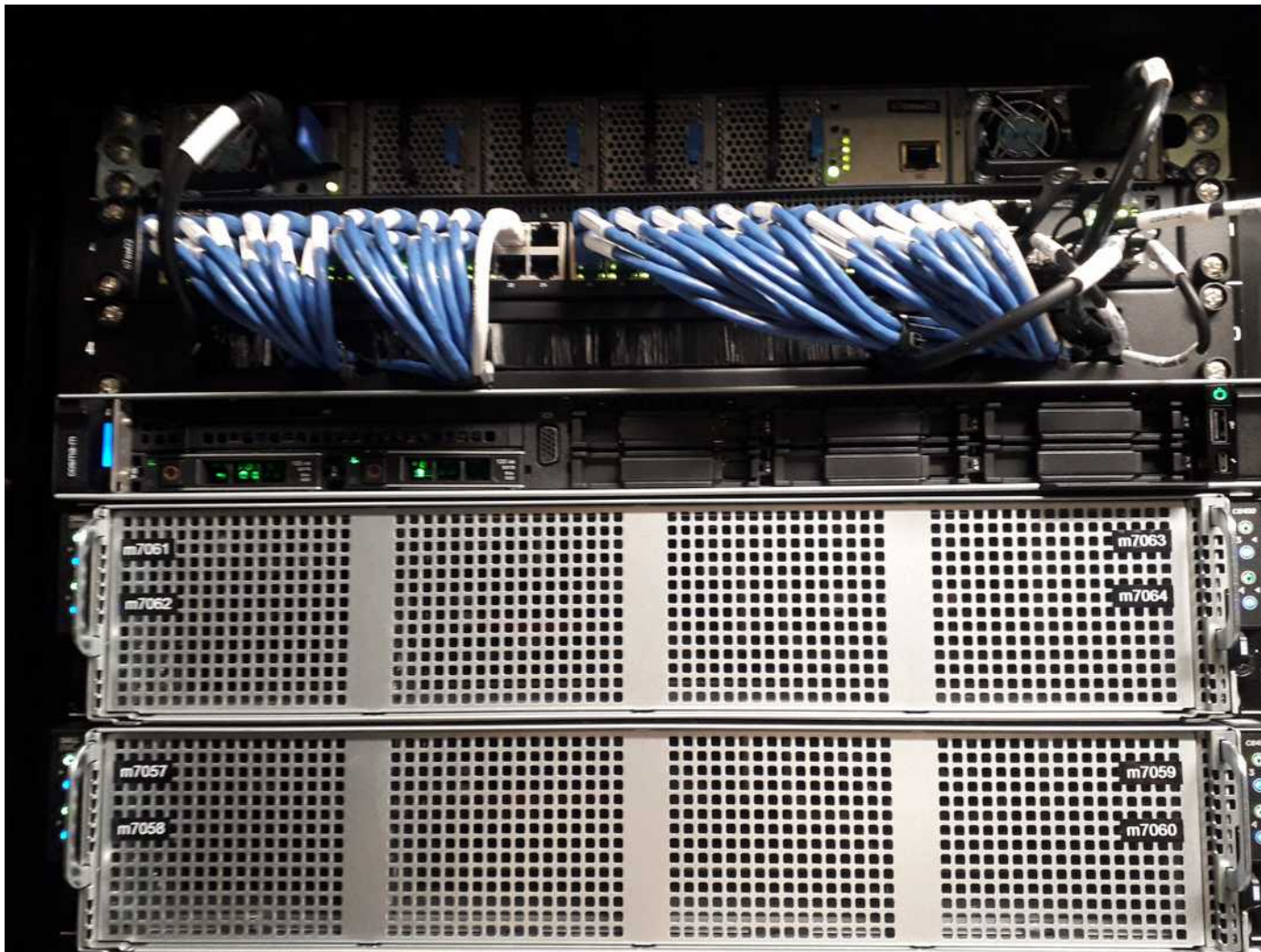
- login.cosma.dur.ac.uk
- login5.cosma.dur.ac.uk
 - Both go randomly to cosma-a or cosma-b, which are cosma5 login nodes.
 - These will shortly be renamed login5a, login5b
- login6.cosma.dur.ac.uk
 - Goes to cosma-i, which is the cosma6 login node.
 - This will shortly be renamed login6a
- login7.cosma.dur.ac.uk
 - Goes to cosma-m and cosma-n, which are the cosma7 login nodes.
 - Will shortly be renamed login7a, login7b.
 - login7c already exists (internally)

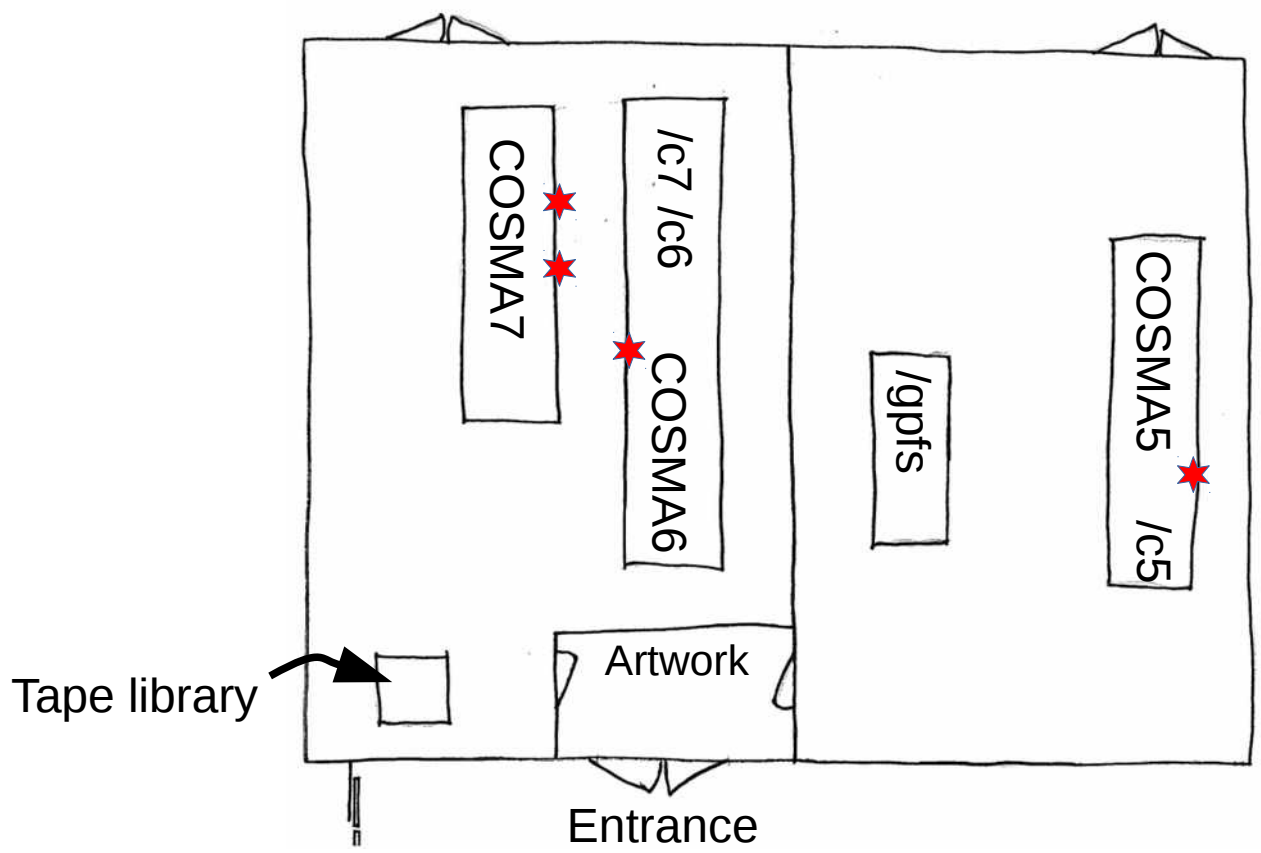
What are login nodes

- Basically, very similar to compute nodes
- More memory
 - 0.5/1.5TB, 3-4x more than compute nodes
- Might be slightly larger physically to allow better Ethernet connectivity (more PCIe cards)
- In the racks with the other nodes









Job submission

- Job submission (by SLURM) can be done to any COSMA, from any login node
- Best practice is to use the login node for the COSMA to which you will submit
 - The login nodes have the same CPUs as their associated compute nodes
 - Reduces issues with compiler optimisation

Login accounts

- A login account for COSMA is requested from SAFE
 - (safe.epcc.ed.ac.uk/dirac)
- To use cosma6 or 7, need to be part of a DiRAC project
 - e.g. dp004/VIRGO
- To use cosma5, need to be in ICC, or collaborating with ICC
 - You will be joined to project “durham”/hpcicc
 - A SAFE account is still required (to keep things simple for us!)

Groups

- You might be part of several projects
 - Your Unix group(s) is (are) named after your project
- The “id” command will tell you which groups you are part of, and your current, effective, group:
 - `uid=20957(dc-basd1) gid=64528(durham) groups=64528(durham),1210(dphsprog),20140(dr004),64526(dp004),64532(lg),64603(cosma7),64605(mad),64607(cosma5),64610(madtesters),1295600001(clusterusers)`
- File permissions/ownership is based on your UID and GID:
 - ls -al:
 - `-rw-r--r-- 1 dc-basd1 durham 4110 Sep 21 2018 users.txt`
 - User and project quotas are derived from file ownership
 - Scattered files are not always easy to find
- newgrp command can be used to change your effective group
 - You will then write files as part of that group
- COSMA5 users will be members of group “durham” (DiRAC calls this hpcicc)
- To use COSMA6/7, you need to be part of a dp group (e.g. dp004)

How best to use COSMA

- If you are doing work for a DiRAC project:
 - Do not use COSMA5
 - We would not get any credit for this
 - Continuation of a DiRAC service at Durham is at stake!
- If your job can use lots of memory:
 - Use COSMA7
 - We report memory used, and have to show that there is a valid requirement for a memory intensive system
- If you can reduce your node count to use more memory per node:
 - Please do so, and use COSMA7
 - You might find it runs faster than expected due to parallel scaling inefficiencies
- If you use less than 128GB/node:
 - Use COSMA6
- If your project is not DiRAC related:
 - Use COSMA5
 - And put in a seedcorn application!

Storage allocation

- You have several storage allocations on Cosma
 - /cosma/home
 - ~10GB quota
 - Backed up
 - Source files, information, etc
 - 37TB XFS / NFS file system
 - /cosma567/data/ [PROJECT] / USERNAME
 - ~10TB quota – can be increased
 - Parallel file systems
 - Lustre or GPFS
 - Multiple redundancy, no backup
 - /snap7/scratch/PROJECT/USERNAME
 - Fast IO for COSMA7. No redundancy, no backup
 - NOT FOR ANY LONG-TERM STORAGE
 - Various other locations depending on project
 - e.g. /madfs, /data/deg1

COSMA power consumption

- COSMA7 consumes ~160kW
 - Single node 80/300W
 - 11W/core
- COSMA6 consumes ~170kW
 - Single node 250W
 - 16W/core
- COSMA5 consumes ~100kW
- ~£500,000/year
 - Assuming cooling efficiency (PUE) 1.1-1.2
- ~2000 Tonnes CO₂ per year (2Gg)
 - Please make good use of COSMA
 - Efficient codes
 - Avoid unnecessary runs



Credit: renewablesfirst.co.uk

COSMA8

- DiRAC3
 - Current expectation within 1 year
 - Durham will again host the Memory Intensive system

Applying for time

- DiRAC has seedcorn time
 - 50,000 core hours
 - To test resources/codes, benchmark
 - Enable code development/testing outside of an established project
- Call for full proposals will open shortly

Next time...

- Parallel file systems