ARCHER Service

Service, Applications, Migration Issues

Andy Turner, EPCC
a.turner@epcc.ed.ac.uk

www.epcc.ed.ac.uk
www.archer.ac.uk

Preview

- Brief overview of ARCHER
- Common migration issues
- Discussion

Discussion

Possible Topics

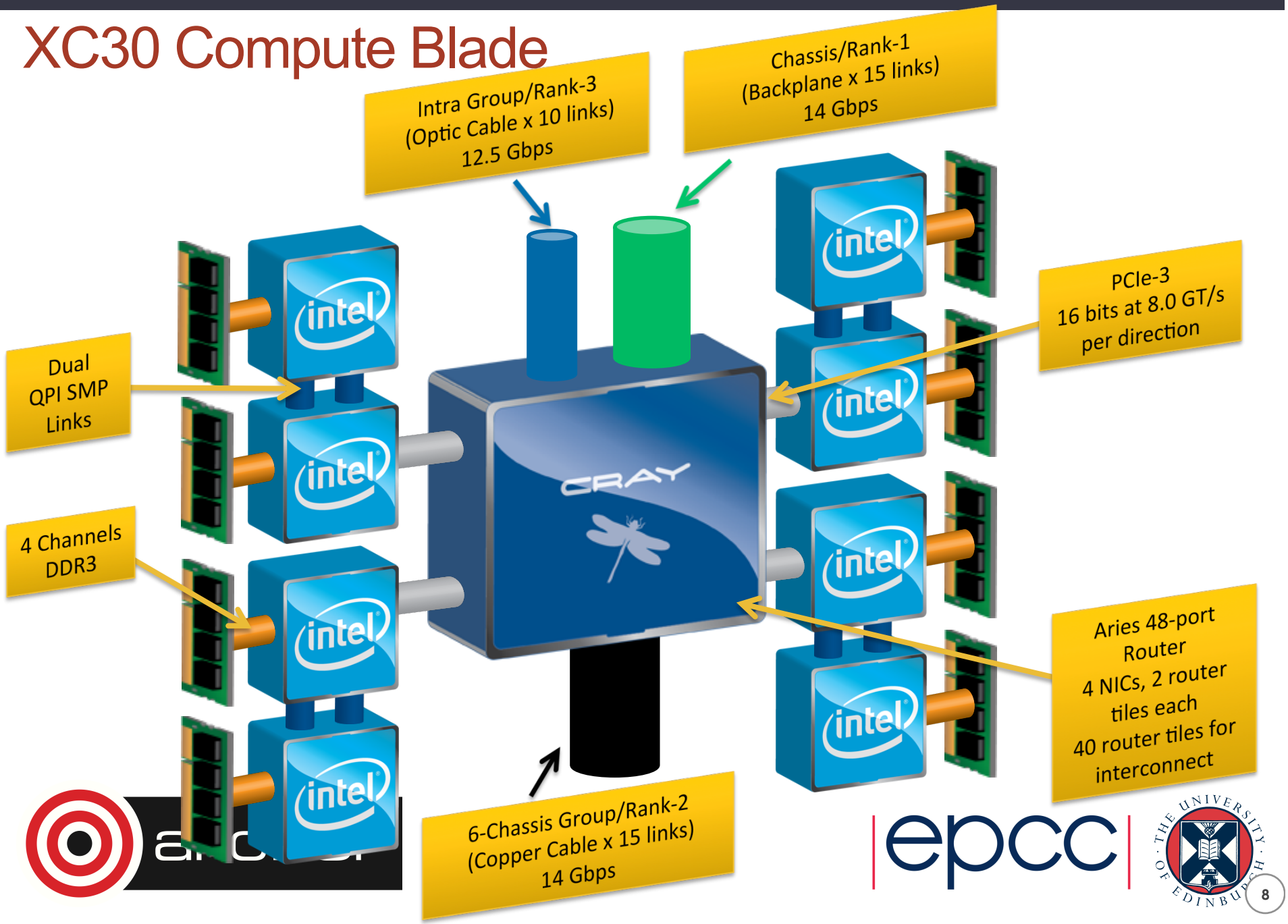
- What are the main issues your users face in exploiting HPC?
- What basic training is missing?
  - Can we all collaborate to provide this online?
- Which applications are used on different facilities?
  - Is there any commonality?
  - Where do the differences lie? Are there good reasons for these differences
- Should we try to get users into the habit of benchmarking?
  - Rather than just running as they always have

Brief ARCHER Overview

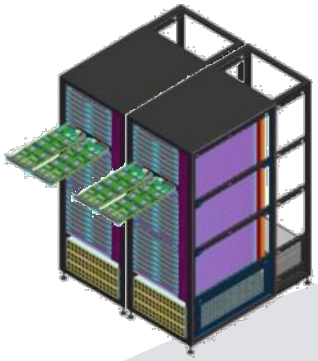
ARCHER in a nutshell

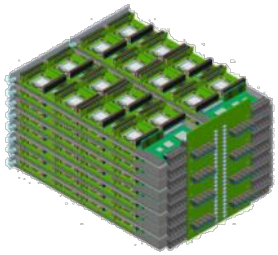
- UK National Supercomputing Service
  - £43 million 4-year project from 2013
- Cray XC30 Hardware
  - Nodes based on 2×Intel Ivy Bridge 12-core processors
  - 64GB (or 128GB) memory per node
  - 4920 nodes in total (118080 cores)
  - Linked by Cray Aries interconnect (dragonfly topology)
- Cray Application Development Environment
  - Cray, Intel, GNU Compilers (all support OpenMP)
  - Cray Parallel Libraries (including optimised MPI)
  - DDT Debugger, Cray Performance Analysis Tools

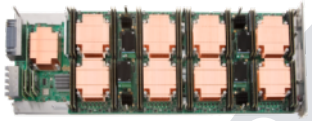
XC30 Compute Blade



ARCHER System Building Blocks







Compute Blade

4 Compute Nodes

Chassis

Rank 1 Network

16 Compute Blades

No Cables

64 Compute Nodes

Group

Rank 2 Network

Passive Electrical Network

2 Cabinets

6 Chassis

384 Compute Nodes

System

Rank 3 Network

Active Optical Network

8 Groups

3008 Compute Nodes

archer

epcc

THE UNIVERSITY OF EDINBURGH

Application Distribution

Code	Nh	% Time	Jobs	% Jobs	Users	Mean	Median
VASP	8480710	16.31	369187	32.84	260	65.54	11
Unified Model	6360508	12.24	47979	4.27	247	150.67	120
cp2k	3724270	7.16	44039	3.92	133	90.40	23
Gromacs	3162771	6.08	34115	3.03	120	48.08	22
CASTEP	2924133	5.62	233149	20.74	157	46.86	15
HiPSTAR	2033244	3.91	2342	0.21	13	465.60	427
LAMMPS	1730405	3.33	30121	2.68	79	78.06	27
ONETEP	1685671	3.24	6322	0.56	39	73.91	39
WRF	1630362	3.14	2891	0.26	28	158.98	150
NEMO	1444228	2.78	26593	2.37	34	105.27	71

40% Materials Science
23% Climate/Ocean Modelling
12% CFD
9% Biomolecular Simulation

ARCHER Allocations: kAU

- Represents performance on ARCHER across a range of benchmarks
  - Baselined to HECToR performance
  - 1 kAU produces, on average, the same amount of application performance on ARCHER as it did on HECToR
- Currently, there are 0.015 kAU per core hour
- For EPSRC and NERC, 1 kAU costs £0.56

Common Migration Issues

1. /work file system

- Only the parallel Lustre file system (/work) is mounted on the compute nodes
- All files needed during a compute job must be on this file system

2. Serial applications

- You cannot (easily) launch more than one serial process on a compute node
  - Means that 23 cores are idle
- Best solution is probably to use other resources for this type of application, but…
  - ...can use parallel Python
  - ...can use MPI bash (really!)

3. Specify nodes, not cores

- Batch script syntax requests nodes, not cores
- Can lead to expensive mistakes!
  - I wanted 200 cores but got 200 nodes (24x more expensive)
- A number of utilities available to help:
  - bolt: command-line tool to produce job submission scripts
    - https://github.com/aturner-epcc/bolt
  - checkScript: check your script for correctness and estimate kAU usage
- System also monitors for node underuse and helpdesk notifies users

4. Where is mpif90?

- Cray application development environment is quite different from clusters
  - More tightly integrated
- Unfamiliar names for compilers
  - Always use ftn, cc, CC wrapper scripts
  - Change underlying compiler by changing modules
  - Code is built statically by default
- Library paths are controlled by modules
  - Hardly ever need to specify them in Makefiles, etc.

5. Computational frameworks

- See increasing number of frameworks that tie together parallel applications
  - e.g. optimisation algorithms that launch MPI codes, look at output, modify input parameters and launch again
- Problematic to implement on Cray XC
  - Cannot launch MPI applications from compute nodes
  - Job launcher nodes cannot support compute-intensive work
  - No SSH access to compute nodes (or between compute nodes)

Discussion

Possible Topics

- What are the main issues your users face in exploiting HPC?
- What basic training is missing?
  - Can we all collaborate to provide this online?
  - Should we set up a UK computational science training repository?
- Should we try to get users into the habit of benchmarking?
  - Rather than just running as they always have
- Which applications are used on different facilities?
  - Is there any commonality?
  - Where do the differences lie?