

UCL Research Computing

Dr Owain Kenway, (@owainkenway)
UCL/ISD/RITS/[Acting] Head of Research Computing

UCL Research Computing

“Supporting UCL’s researchers with not enough money, not enough space, not enough people and not enough time”

Dr Owain Kenway, (@owainkenway)
UCL/ISD/RITS/[Acting] Head of Research Computing

- A team within Research IT Services (RITS), which is within ISD (UCL's central IT division)
- We (two teams of ~4) look after UCL's central, and UCL-hosted national services
 - RCI → “The systems team” look after the hardware, OS, are `root`
 - RCAS → “The user support team” look after users, user applications, are `ccspapp`
 - **Together** design, deliver + support all the services

- UCL is a world leading research institution :)
 - Research in almost every field :)
 - *All institutions* see IT as a cost area (i.e. minimise funding) :(

aka

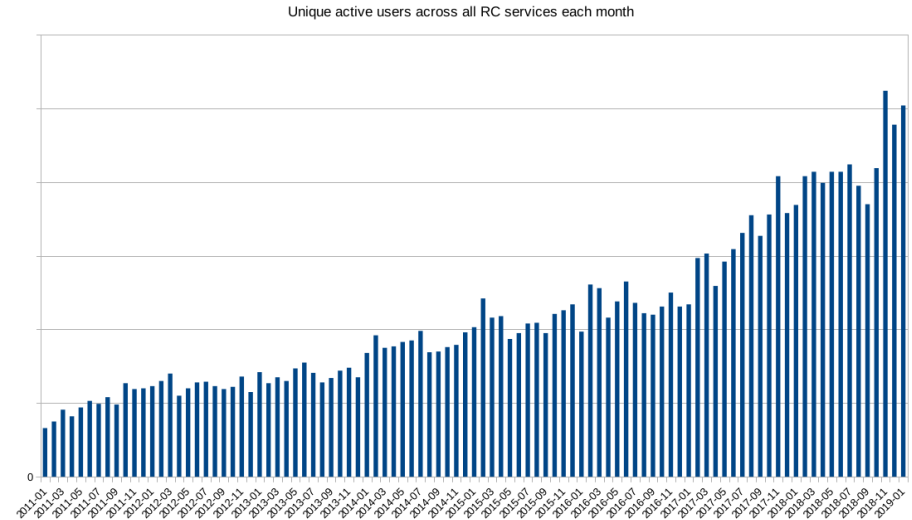
We have limited funding but must support wildly diverse needs.

(This is not just true of Research Computing but the whole of ISD!)

The Problem:

- Some users need traditional HPC (massively parallel workloads)
- Some users need HTC (thousands of independent jobs)
- Some users need a mix of the two (High Throughput High Performance Computing?)
- Some users need GPUs
- Some users need terabytes of RAM
- Some users need to use massive amounts of (temporary) storage
- Some users are HPC experts, some are novices
- Some users are developing their own code, some are using centrally installed applications
- Some users...

- Geography
 - There is no space for anything anywhere in central London
 - There is no power for anything anywhere in central London
- Money
 - IT is incredibly underfunded sector-wide
 - HPC is funded worse than that (a national problem, not a UCL one)
 - Not enough money for kit or staff
- Insatiable demand for compute from researchers



- UCL only services:

- **Grace** → High Performance Computing (HPC)
- **Myriad**, **Legion** → High Throughput Computing (HTC)
- **Aristotle** → Interactive teaching Linux service
- **DSH** → secure data storage and compute (not currently under RC control)

- National services:

- **Thomas** (Tier 2 MMM hub)
- **Michael** (Faraday Institution)

- Parallel

- Single job spans multiple nodes
- Tightly coupled parallelisation usually in MPI
- Sensitive to network performance
- Currently primarily chemistry, physics, engineering

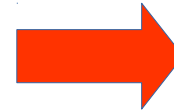
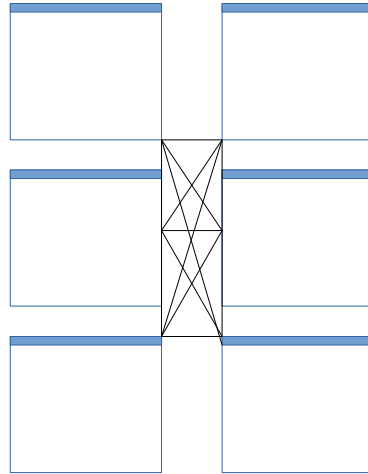
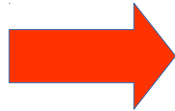
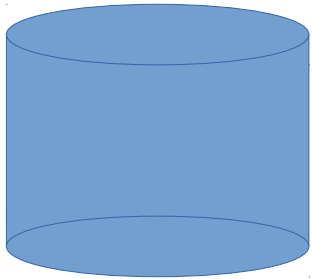
- High throughput

- Lots (tens of thousands) of independent jobs on different data
- High I/O
- Currently, primarily biosciences, physics, computer science
- In the future, digital humanities

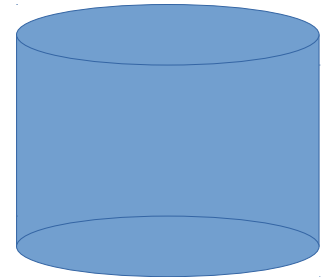
COMMON software stack across RC controlled services

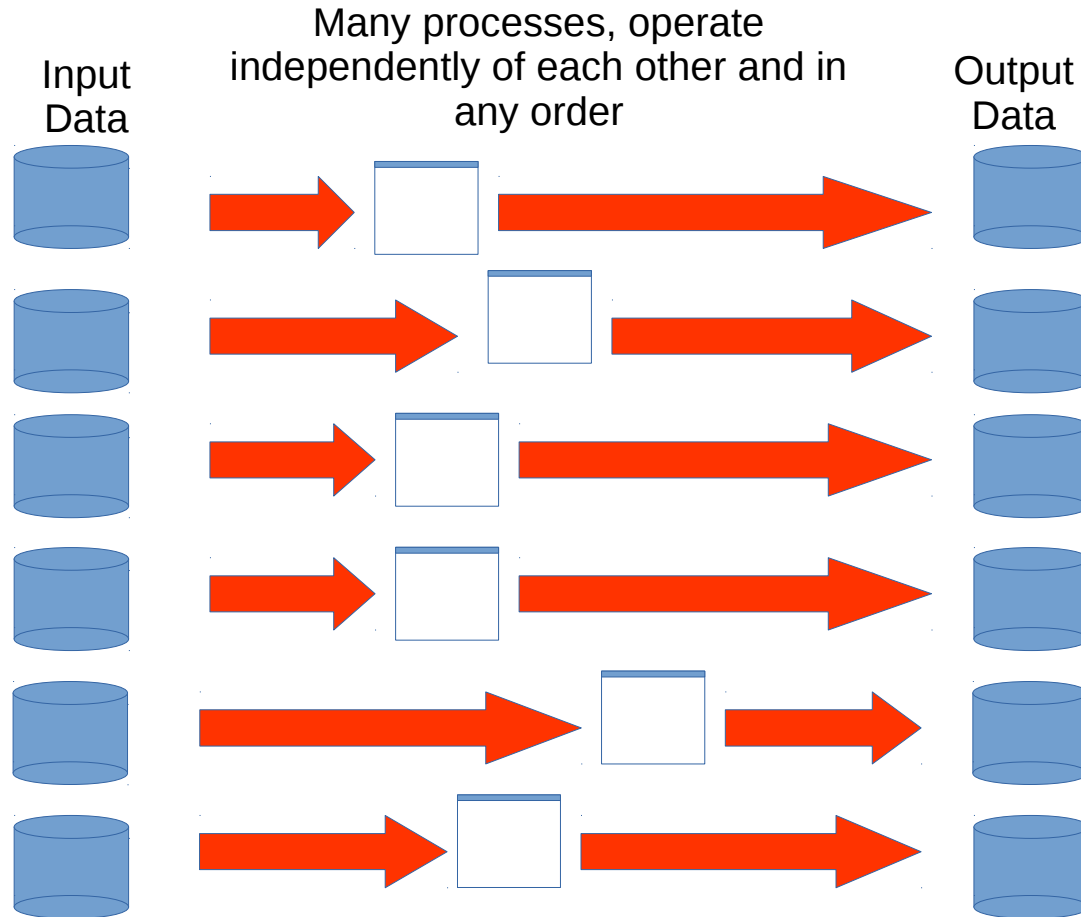
Many processes on many processors
work simultaneously + communicate
between each other

Input Data



Output Data

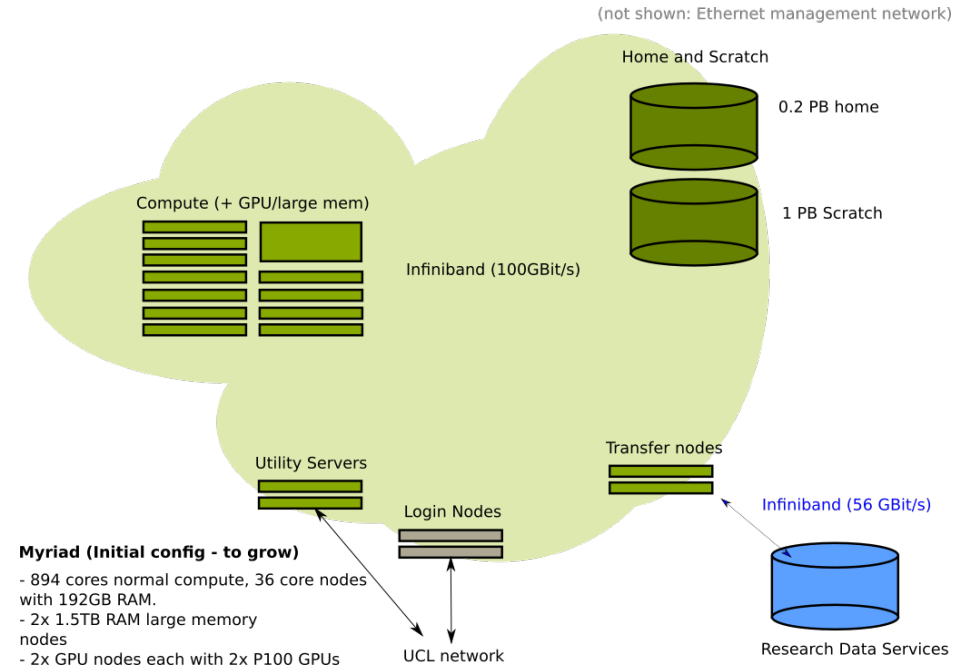


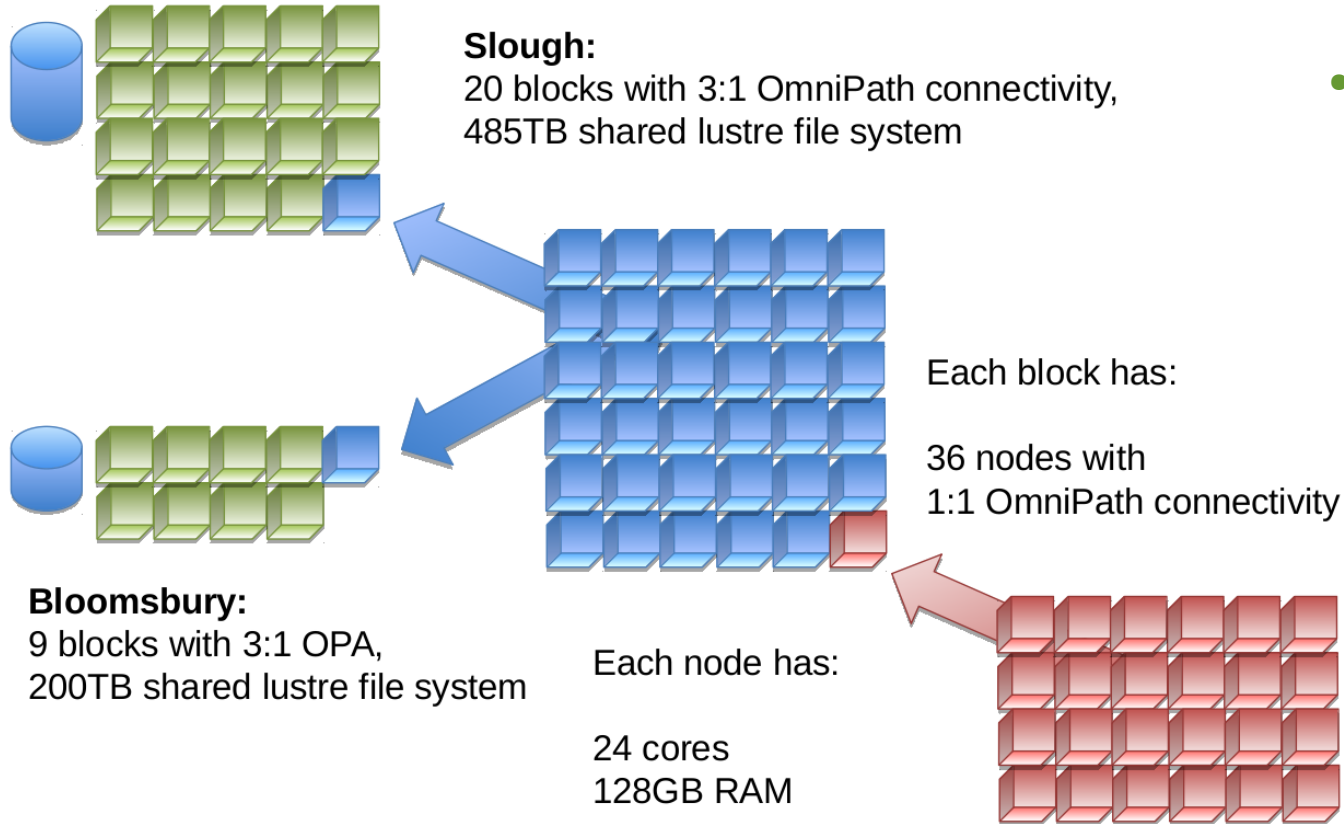


- Grace is UCL's primary HPC service.
 - OCF/Lenovo, QDR IB
 - ~11K cores
 - 16 cores/node
 - 1PB of Lustre storage



- Myriad is UCL's High Throughput/Data intensive service:
 - OCF/Lenovo, EDR IB (storage only)
 - ~1800 cores
 - 4 GPUs
 - 2 high memory nodes
 - 1 PB of Lustre Storage
- Free and paid access models

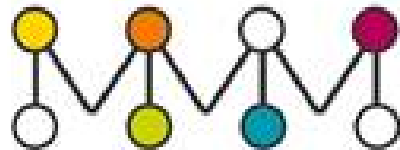




- Two national services,
 - both for specialised research
 - both “High Throughput High Performance Computing”
 - (i.e. arrays of small parallel jobs)

- Thomas (the Tier 2 MMM Hub)
 - Built off of Grace (OPA)
 - 18,000 cores
 - £4M of EPSRC funding
 - Running costs paid for by partner institutions:

Imperial College London, University of Kent, Kings College London, Oxford University, Queen Mary University of London, Queen's University Belfast, University of Southampton and **UCL**
- High Throughput High Performance Computing!
- Running for almost two years.
 - >600,000 user jobs to date
 - >2.57x10⁸ CPU hours used by user jobs
 - >29,300 years on one core
 - Upper Palaeolithic start!

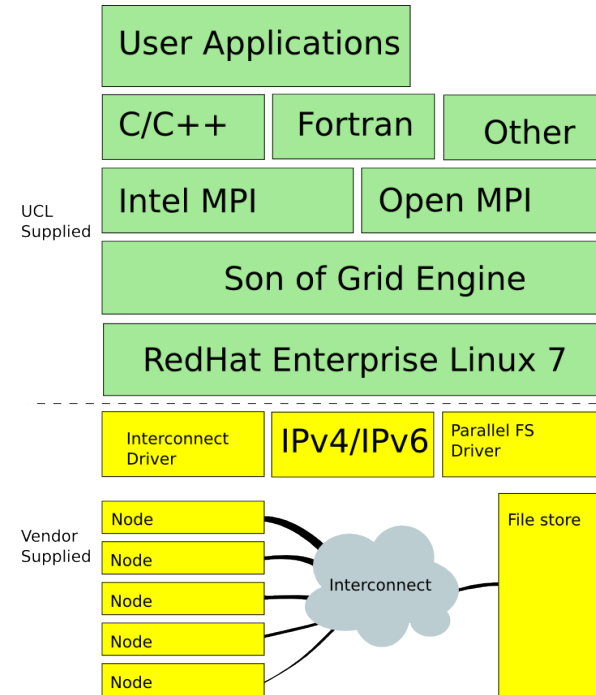


- Michael (the Faraday Institution machine)
 - Built off of Grace (OPA) in summer last year
 - 7,000 cores
 - £1M of external funding



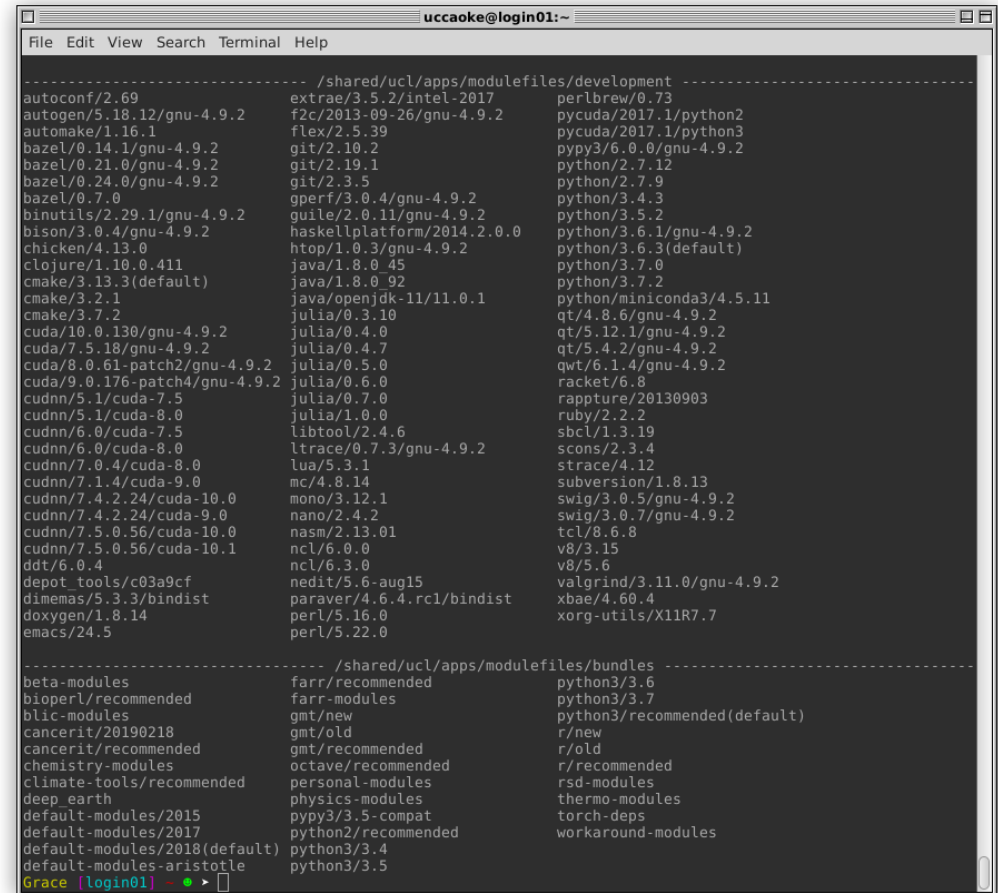
Common software stack

- Deployed across all our resources (inc Thomas + Michael)
 - ~750 user applications + development tools, presented through environment modules
 - Scripts + data from one machine can be run “seamlessly” on another
 - Same interface presented to users
 - **AUTOMATED**



Common software stack

- This is **not** a stack “just for traditional HPC users” (Fortran/C/C++)
- Supports Python (Cpython, Anaconda, PyPy), R, Julia, Perl (+ Bioperl), Java, Clojure, Common Lisp, Scheme, Mono (.Net), Lua, Go, Racket, Ruby, JavaScript, Matlab...
- ML tools like Tensorflow (GPU, MKL variants), Caffe, OpenCV...
- Allow departmental sysadmins access to install specialist applications centrally!



```
uccaoke@login01:~
File Edit View Search Terminal Help
----- /shared/ucl/apps/modulefiles/development -----
autoconf/2.69          extrae/3.5.2/intel-2017      perlbrew/0.73
autogen/5.18.12/gnu-4.9.2  f2c/2013-09-26/gnu-4.9.2    pycuda/2017.1/python2
automake/1.16.1         flex/2.5.39                 pycuda/2017.1/python3
bazel/0.14.1/gnu-4.9.2   git/2.10.2                 pypy3/6.0.0/gnu-4.9.2
bazel/0.21.0/gnu-4.9.2   git/2.19.1                 python/2.7.12
bazel/0.24.0/gnu-4.9.2   git/2.3.5                   python/2.7.9
bazel/0.7.0             gperf/3.0.4/gnu-4.9.2      python/3.4.3
binutils/2.29.1/gnu-4.9.2  guile/2.0.11/gnu-4.9.2     python/3.5.2
bison/3.0.4/gnu-4.9.2    haskellplatform/2014.2.0.0  python/3.6.1/gnu-4.9.2
chicken/4.13.0          htop/1.0.3/gnu-4.9.2       python/3.6.3(default)
clojure/1.10.0.411      java/1.8.0_45               python/3.7.0
cmake/3.13.3(default)   java/1.8.0_92               python/3.7.2
cmake/3.2.1             java/openjdk-11/11.0.1     python/miniconda3/4.5.11
cmake/3.7.2            julia/0.3.10                qt/4.8.6/gnu-4.9.2
cuda/10.0.130/gnu-4.9.2  julia/0.4.0                 qt/5.12.1/gnu-4.9.2
cuda/7.5.18/gnu-4.9.2   julia/0.4.7                 qt/5.4.2/gnu-4.9.2
cuda/8.0.61-patch2/gnu-4.9.2  julia/0.5.0                 qwt/6.1.4/gnu-4.9.2
cuda/9.0.176-patch4/gnu-4.9.2  julia/0.6.0                 racket/6.8
cudnn/5.1/cuda-7.5      julia/0.7.0                 rapture/20130903
cudnn/5.1/cuda-8.0      julia/1.0.0                 ruby/2.2.2
cudnn/6.0/cuda-7.5     libtool/2.4.6               sbcl/1.3.19
cudnn/6.0/cuda-8.0     ltrace/0.7.3/gnu-4.9.2     sconsl/2.3.4
cudnn/7.0.4/cuda-8.0   lua/5.3.1                   strace/4.12
cudnn/7.1.4/cuda-9.0   mc/4.8.14                    subversion/1.8.13
cudnn/7.4.2.24/cuda-10.0  mono/3.12.1                 swig/3.0.5/gnu-4.9.2
cudnn/7.4.2.24/cuda-9.0  nano/2.4.2                   swig/3.0.7/gnu-4.9.2
cudnn/7.5.0.56/cuda-10.0  nasm/2.13.01                 tcl/8.6.8
cudnn/7.5.0.56/cuda-10.1  ncl/6.0.0                    v8/3.15
ddt/6.0.4              ncl/6.3.0                    v8/5.6
depot_tools/c03a9cf     nedit/5.6-aug15              valgrind/3.11.0/gnu-4.9.2
dimemas/5.3.3/bindist  paraver/4.6.4.rc1/bindist    xbg/4.60.4
doxygen/1.8.14          perl/5.16.0                  xorg-utils/X11R7.7
emacs/24.5              perl/5.22.0
----- /shared/ucl/apps/modulefiles/bundles -----
beta-modules          farr/recommended            python3/3.6
bioperl/recommended  farr-modules                 python3/3.7
blic-modules          gmt/new                      python3/recommended(default)
cancerit/20190218     gmt/old                       r/new
cancerit/recommended  gmt/recommended              r/old
chemistry-modules    octave/recommended            r/recommended
climate-tools/recommended  personal-modules              rsd-modules
deep_earth            physics-modules                thermo-modules
default-modules/2015  pypy3/3.5-compat              torch-deps
default-modules/2017  python2/recommended           workaround-modules
default-modules/2018(default)  python3/3.4
default-modules-aristotle  python3/3.5
Grace [login01] ~
```


- Most of the code that builds/runs stuff on our clusters is in Github e.g.
 - Build scripts: <https://github.com/UCL-RITS/rcps-buildscripts>
 - GERun: <https://github.com/UCL/GERun>
 - ... and others

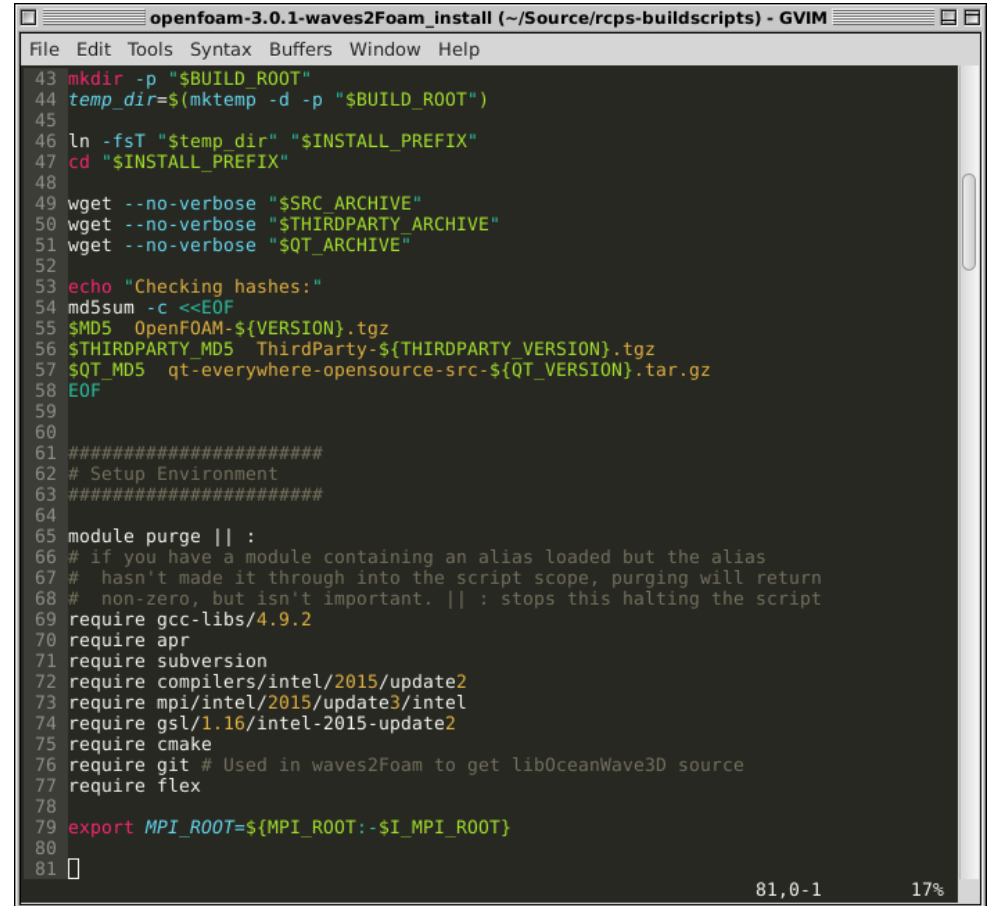
METHOD NOT APPLICATION!

- Works on our clusters but maybe no-where else!
- (relies on /shared/ucl/apps existing, not well documented etc.)

Open an issue to ask for a package to be installed.

We install software...

- From the easy:
 - \$ pip3 install numpy
- To the hard:
 - Multiple incompatible dependencies
 - Bazel
 - MPI/Cuda builds...
 - For more examples, see Kenneth Hoste's excellent FOSDEM talk "How to Make Package Managers Cry" on Youtube
<https://www.youtube.com/watch?v=NSeMIYagjIU>

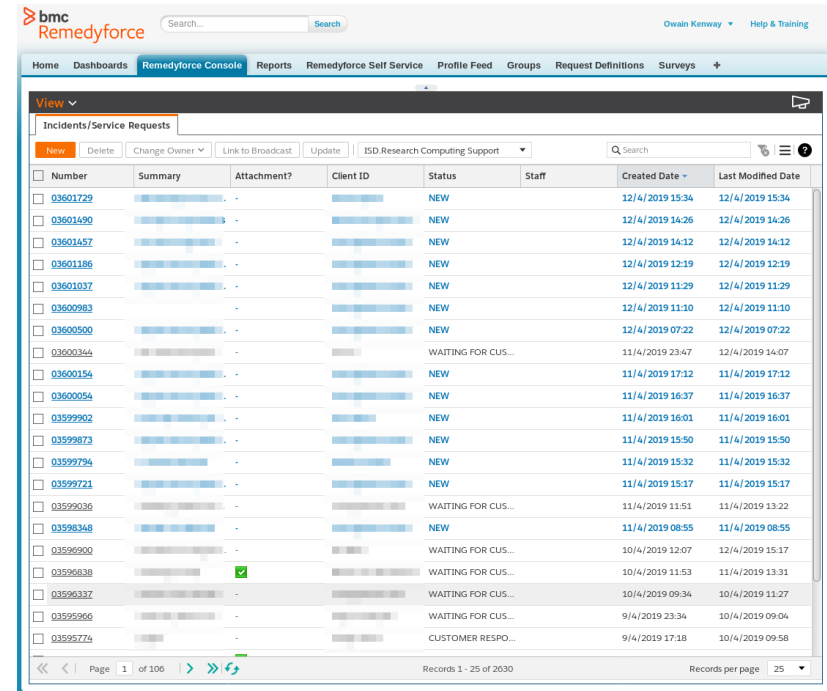


```
openfoam-3.0.1-waves2Foam_install (~/.Source/rcps-buildscripts) - GVIM
File Edit Tools Syntax Buffers Window Help
43 mkdir -p "$BUILD_ROOT"
44 temp_dir=$(mktemp -d -p "$BUILD_ROOT")
45
46 ln -fsT "$temp_dir" "$INSTALL_PREFIX"
47 cd "$INSTALL_PREFIX"
48
49 wget --no-verbose "$SRC_ARCHIVE"
50 wget --no-verbose "$THIRDPARTY_ARCHIVE"
51 wget --no-verbose "$QT_ARCHIVE"
52
53 echo "Checking hashes:"
54 md5sum -c <<EOF
55 $MD5 OpenFOAM-${VERSION}.tgz
56 $THIRDPARTY_MD5 ThirdParty-${THIRDPARTY_VERSION}.tgz
57 $QT_MD5 qt-everywhere-opensource-src-${QT_VERSION}.tar.gz
58 EOF
59
60
61 #####
62 # Setup Environment
63 #####
64
65 module purge || :
66 # if you have a module containing an alias loaded but the alias
67 # hasn't made it through into the script scope, purging will return
68 # non-zero, but isn't important. || : stops this halting the script
69 require gcc-libs/4.9.2
70 require apr
71 require subversion
72 require compilers/intel/2015/update2
73 require mpi/intel/2015/update3/intel
74 require gsl/1.16/intel-2015-update2
75 require cmake
76 require git # Used in waves2Foam to get libOceanWave3D source
77 require flex
78
79 export MPI_ROOT=${MPI_ROOT:-$I_MPI_ROOT}
80
81
```

81,0-1 17%

We answer user tickets...

- E-mail rc-support@ucl.ac.uk for help and advice (**not just for our services!**)
- Manned by the RCAS team on a rota



The screenshot displays the BMC Remedyforce console interface. The top navigation bar includes 'Home', 'Dashboards', 'Remedyforce Console', 'Reports', 'Remedyforce Self Service', 'Profile Feed', 'Groups', 'Request Definitions', and 'Surveys'. The main content area shows a table of 'Incidents/Service Requests' with columns for Number, Summary, Attachment?, Client ID, Status, Staff, Created Date, and Last Modified Date. The table contains 20 rows of data, with various statuses such as 'NEW' and 'WAITING FOR CUS...'. A search bar and several action buttons (New, Delete, Change Owner, Link to Broadcast, Update) are visible above the table.

Number	Summary	Attachment?	Client ID	Status	Staff	Created Date	Last Modified Date
03601729		-		NEW		12/4/2019 15:34	12/4/2019 15:34
03601490		-		NEW		12/4/2019 14:26	12/4/2019 14:26
03601457		-		NEW		12/4/2019 14:12	12/4/2019 14:12
03601186		-		NEW		12/4/2019 12:19	12/4/2019 12:19
03601037		-		NEW		12/4/2019 11:29	12/4/2019 11:29
03600983		-		NEW		12/4/2019 11:10	12/4/2019 11:10
03600500		-		NEW		12/4/2019 07:22	12/4/2019 07:22
03600344		-		WAITING FOR CUS...		11/4/2019 23:47	12/4/2019 14:07
03600154		-		NEW		11/4/2019 17:12	11/4/2019 17:12
03600054		-		NEW		11/4/2019 16:37	11/4/2019 16:37
03599902		-		NEW		11/4/2019 16:01	11/4/2019 16:01
03599873		-		NEW		11/4/2019 15:50	11/4/2019 15:50
03599794		-		NEW		11/4/2019 15:32	11/4/2019 15:32
03599721		-		NEW		11/4/2019 15:17	11/4/2019 15:17
03599036		-		WAITING FOR CUS...		11/4/2019 11:51	11/4/2019 13:22
03598348		-		NEW		11/4/2019 08:55	11/4/2019 08:55
03596900		-		WAITING FOR CUS...		10/4/2019 12:07	12/4/2019 15:17
03596838		✓		WAITING FOR CUS...		10/4/2019 11:53	11/4/2019 13:31
03596337		-		WAITING FOR CUS...		10/4/2019 09:34	10/4/2019 11:27
03595966		-		WAITING FOR CUS...		9/4/2019 23:34	10/4/2019 09:04
03595774		-		CUSTOMER RESPO...		9/4/2019 17:18	10/4/2019 09:58

- Grace is now more than three years old!



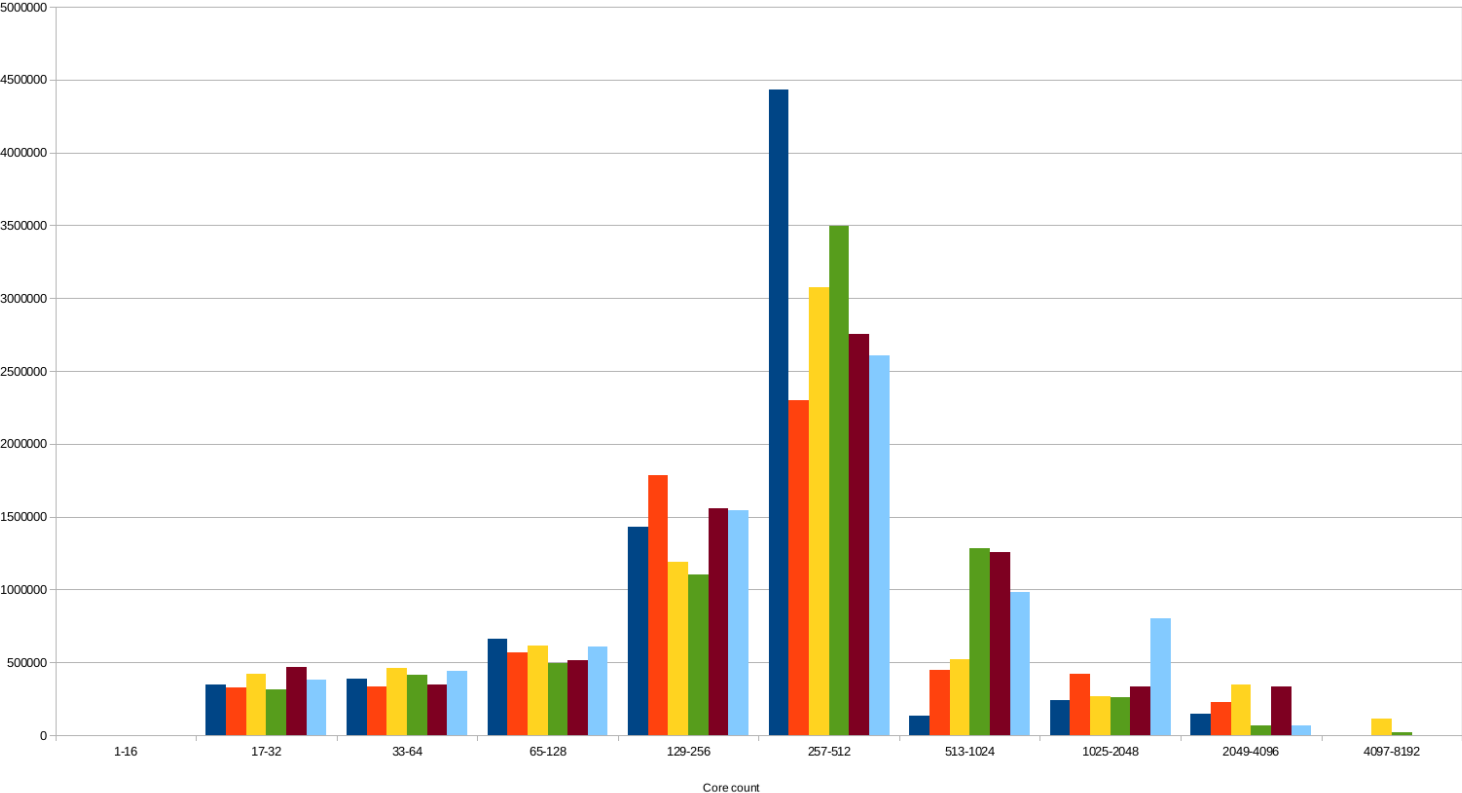
- Time to design and procure a replacement!

- Design project in 2017/18
- Procurement in 2018/19

Future developments: HPC

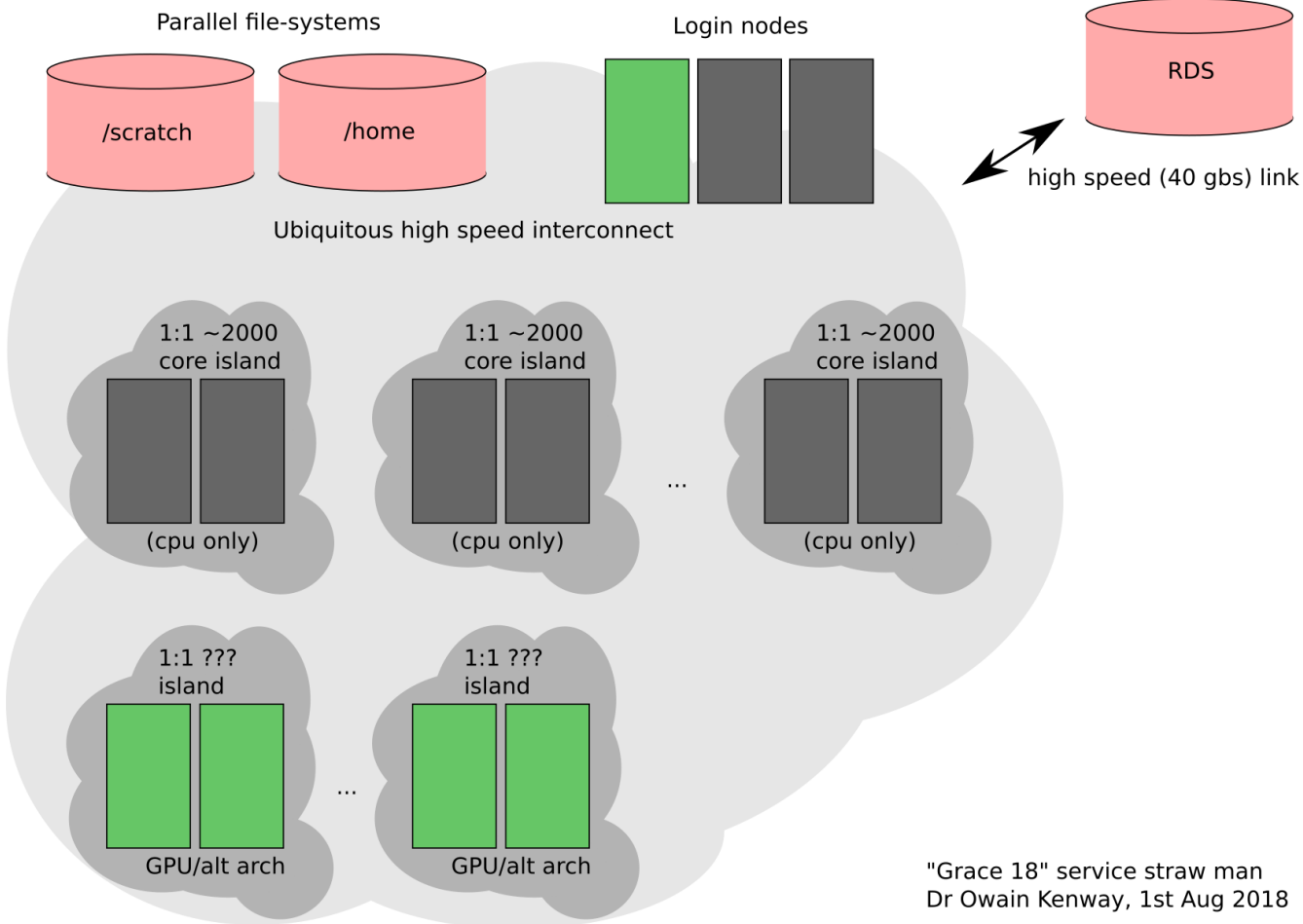


CPU Time by job size on Grace
Six month period, 2018-01 to 2018-06



Almost all time consumed is <2048 cores

Future developments: HPC

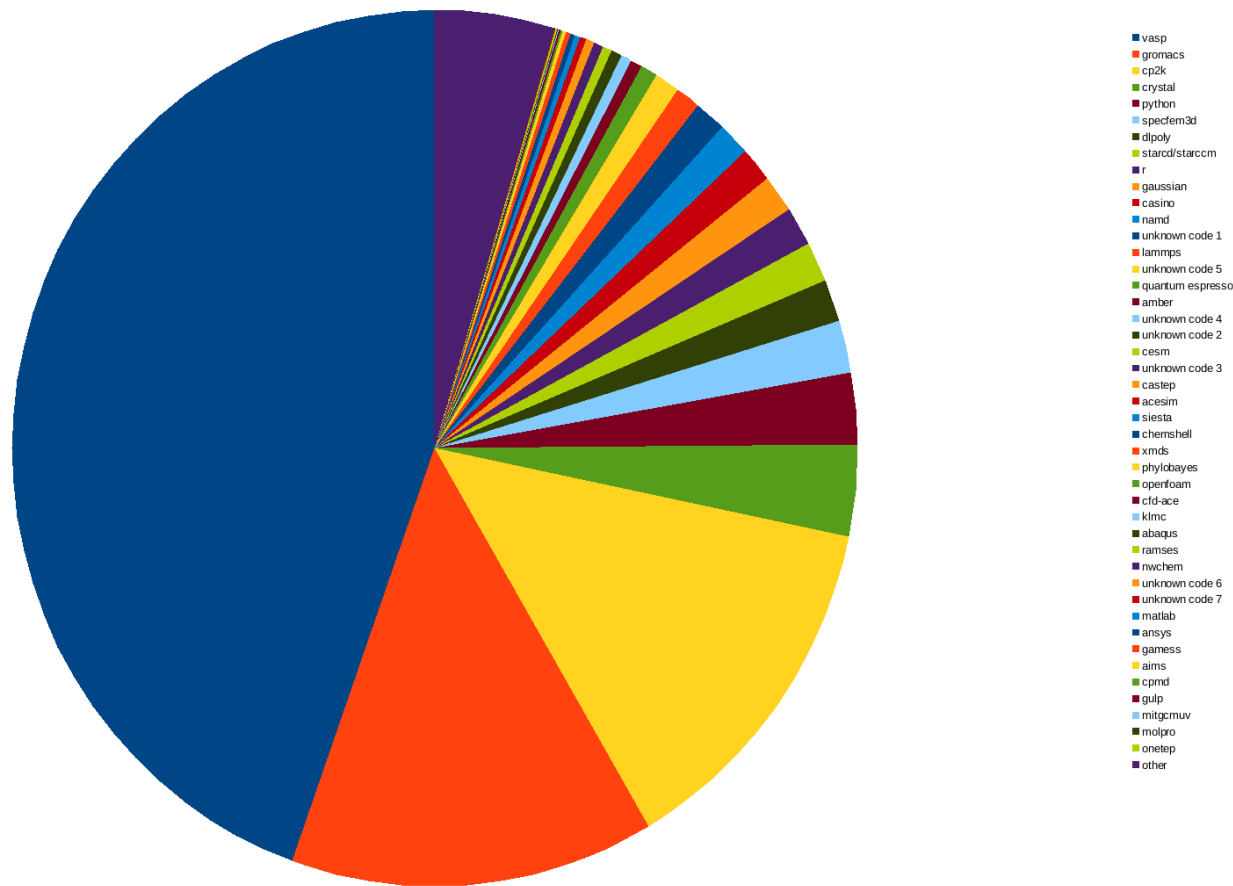


"Grace 18" service straw man
Dr Owain Kenway, 1st Aug 2018

Future developments: HPC

Application usage over six month period

(Jan 2018 to Jun 2018)



Biggest application usage is VASP, GROMACS, CP2K

- Based on application usage we have started a Github repository with standard benchmarks for this and future procurements.
- Will grow as we design benchmarks for other systems.
- Based on Archer+TeraGrid benchmarks:
 - <https://github.com/UCL-RITS/HPC-Acceptance-Tests>
- Vendors use benchmarks to prepare their bids, we use them to do acceptance tests.

**DEFINITELY A
LEARNING
EXPERIENCE!**

- Project to procure Grace's as yet unnamed replacement under way
 - Pilot machine mid-late 2019
 - Looks more like Thomas than Grace
 - Have many bids from different vendors currently being evaluated
 - (they were doing datacentre surveys today!)
 - Free and paid access models a la Myriad (~2000 core chunks!)

- File system access:
 - Presenting home directories via CIFS → people **not** using Linux will **also** be able to mount their home directories on their desktop!
(Linux users can already do this with FUSE/SSHFS)
 - Mounting Research Data Storage on login nodes
(all sorts of exciting authentication challenges!)
- Usability:
 - Work with CS to present VDI front-end
 - Other ways of accessing resource?
 - Booking system for training courses on Aristotle

- From mid 2019 we will be *collaborating with CS* in running a centrally funded R&D cluster.
- Collaboration is tentatively called RCNIC (Research Computing & Networking Innovation Centre).
- Early access to technologies for researchers e.g. FPGAs, Arm etc outside of a defined service.
 - Successful technologies will be adopted in *future* ISD service offerings
- Software collaborations (e.g. Linux VDI, our software stack, service reporting...)
- **Not just CS** – we're keen to bring in other departments that do HPC (Physics, Engineering etc.).

Thanks!

**IT'S TIME FOR
PIZZA**