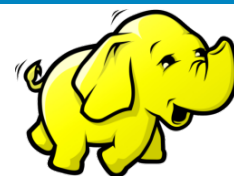


Hadoop and Universities



Strategic Goals

- CS and other science graduates come out knowing how to code with MapReduce
- Postgraduate research is done *on* and *inside* Hadoop.
- The UK & EU grids host Hadoop for PB of data and the computation
- Engagement between the ASF/Hadoop team and the Academic community

Where is Hadoop being used?

- CS: MapReduce as an algorithm
- AI: datamining (Edinburgh, Glasgow)
- Other sciences: Hadoop for data storage/analysis?

CS Teaching

- Cloudera VM
- The Eclipse plugin needs love
- Common datasets
- Re-use and adapt US coursework

Is Hadoop over-complex?

MapReduce is great with Haskell, Prolog, Erlang

Hadoop over GGF grids?

| | GOC Status (Info) | | | | | | SAM Tests (Info) | | | | ATLAS SAM (Info) | | | | CMS SAM (Info) | | | | LHCb SAM (Info) | | |
|--------------------------------|-------------------------------------|----------|----------|-----------|----------|-----------|------------------------------------|-------|--------|------|------------------------------------|----|--------|------|----------------------------------|-------|--------|------|-----------------------------------|--------|------|
| Institute | CPU Tot | CPU Free | Jobs Cur | Jobs Wait | Disk Tot | Disk Free | CE | SRMv2 | 24 Hrs | Week | CE | SE | 24 Hrs | Week | CE | SRMv2 | 24 Hrs | Week | CE | 24 Hrs | Week |
| Trinity Dublin | 772 | 632 | 115 | 0 | 172.2 | 149.7 | P | P | 100% | 90% | P | | 100% | 90% | | | 0% | 0% | P | 100% | 95% |
| Brunel | 433 | 173 | 202 | 263 | 88.7 | 39.0 | P | P | 100% | 100% | P | | 100% | 100% | P | P | 100% | 100% | P | 100% | 100% |
| Imperial HEP | 960 | 216 | 941 | 4 | 158.7 | 61.3 | P | P | 100% | 96% | P | | 100% | 99% | P | P | 100% | 84% | P | 100% | 100% |
| QMUL | 1488 | 217 | 1567 | 1333332 | 4190.1 | 3367.7 | P | P | 96% | 78% | P | | 100% | 100% | P | P | 70% | 52% | P | 100% | 100% |
| RHUL | 400 | 1 | 383 | 133 | 134.7 | 51.4 | P | P | 100% | 97% | P | | 100% | 100% | P | P | 91% | 89% | P | 100% | 100% |
| UCL CCC | 256 | 21 | 47 | 6 | 82.8 | 63.1 | P | P | 96% | 92% | F | | 0% | 12% | F | P | 0% | 0% | F | 0% | 0% |
| UCL HEP | 96 | 146 | 8 | 0 | 5.3 | 4.9 | P | P | 100% | 83% | P | | 100% | 90% | | | 0% | 0% | | 0% | 0% |
| Lancaster | 656 | 236 | 12 | 0 | 206.9 | 98.2 | P | P | 100% | 100% | P | | 100% | 100% | | | 0% | 0% | P | 100% | 100% |
| Liverpool | 790 | 8 | 404 | 295 | 223.1 | 104.1 | P | P | 100% | 100% | P | | 100% | 100% | | | 0% | 0% | P | 100% | 100% |
| Manchester | 1780 | 1553 | 117 | 0 | 124.9 | 76.8 | P | P | 100% | 99% | P | | 100% | 100% | | | 0% | 0% | P | 100% | 100% |
| Sheffield | 200 | 98 | 86 | 65 | 30.9 | 20.0 | W | P | 100% | 98% | P | | 100% | 100% | | | 0% | 0% | P | 100% | 100% |
| Durham | 336 | 0 | 1344 | 1510 | 33.2 | 29.9 | P | P | 100% | 100% | P | | 100% | 100% | | | 0% | 0% | P | 100% | 100% |
| Edinburgh | 920 | 0 | 0 | 15936 | 30.5 | 13.4 | P | P | 100% | 100% | P | | 30% | 4% | F | P | 0% | 0% | F | 0% | 26% |
| Glasgow | 1912 | 42 | 1751 | 714 | 140.5 | 65.8 | P | P | 100% | 100% | P | | 100% | 100% | P | P | 100% | 55% | P | 100% | 100% |
| Birmingham | 320 | 430 | 85 | 27 | 34.8 | 27.7 | P | P | 100% | 100% | P | | 100% | 100% | | | 0% | 0% | P | 100% | 100% |
| Bristol | 112 | 390 | 26 | 284 | 113.4 | 40.2 | P | P | 100% | 100% | P | | 100% | 100% | P | P | 100% | 99% | P | 100% | 100% |
| Cambridge | 152 | 152 | 0 | 176 | 73.2 | 43.0 | P | W | 100% | 87% | P | | 100% | 100% | | | 0% | 0% | P | 100% | 96% |
| Culham JET | 254 | 230 | 22 | 0 | 1.7 | 1.1 | P | P | 100% | 100% | P | | 100% | 100% | | | 0% | 0% | P | 100% | 100% |
| Oxford | 1152 | 153 | 552 | 4436 | 145.2 | 54.4 | P | P | 100% | 96% | P | | 100% | 100% | P | P | 100% | 78% | P | 100% | 100% |
| RAL PPD | 1284 | 1155 | 113 | 5 | 565.1 | 465.9 | P | P | 100% | 100% | P | | 100% | 100% | P | P | 100% | 97% | P | 100% | 100% |
| RAL Tier-1 | 0 | 0 | 0 | 2666664 | 3340.8 | 1585.6 | F | F | 0% | 61% | F | | 0% | 63% | F | F | 0% | 61% | P | 100% | 77% |
| Overall | 14274 | 5855 | 7775 | 4023850 | 9897.7 | 6364.4 | | | 95% | 94% | | | 87% | 89% | | | 69% | 65% | | 91% | 95% |

What can we do?

- Lecture at the local universities
- Help people set up clusters
- Offer cluster-time and datasets
- Anything else?

Postgraduate Research

- On Hadoop: new algorithms, layers on top
- On Hadoop: MR for science
- In Hadoop: scheduling, placement
- Present at ApacheCon, HUG
- Cluster time on OpenCirrus?

Steer researchers away from trouble, mentor them

A hadoop-research sub-project?

- Place to bring in researchers
- Mailing list where advanced topics are covered: graph theory, NP-completeness, physics events
- SVN repo –teach them Apache processes

UK Hadoop-in-eScience event?

Ross Gardler: "OSS Watch are putting on an open source conference in Q2 2010..."