Science & Technology Facilities Council

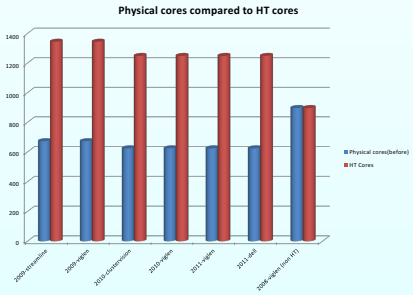
www.stfc.ac.uk

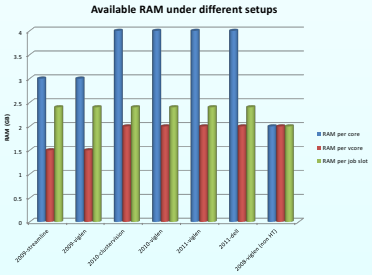
# Maximising job throughput using Hyper-Threading

Alastair Dewhurst, Dimitrios Zilaskos
RAL Tier1

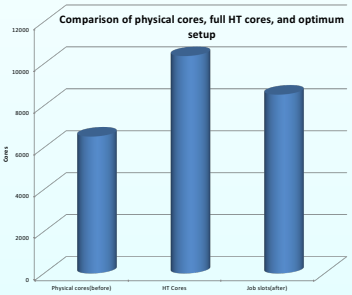
## RAL Tier1 Setup

The RAL TIER1 batch farm consists of several multicore, hyperthreading capable CPUs. Increases in the amount of memory per node combined with experiences from other sites made hyperthreading an attractive option for increasing job throughput.
RAL supports all LHC VOs, with prime users being Atlas, CMS and LHCB, and a 10% of resources is devoted to non-LHC VOs
The virtual cores provided by hyperthreading could double the batch farm capacity, however the amount of memory available in the batch nodes did not permit that. LHC jobs require more than 2GB RAM to run smoothly.



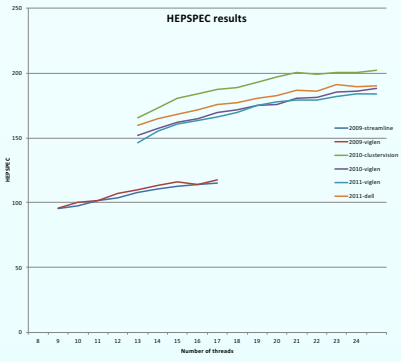



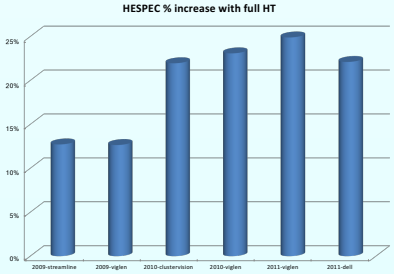

With all HT cores enabled total job slots capacity could double. In practice memory constrains resulted in an increase of 30%




## Method

- Each generation of batch farm hardware with hyperthreading capability was benchmarked with HEPSPEC, progressively increasing the number of threads up to the total number of virtual cores
- Benchmarks at that time were conducted using Scientific Linux 5. Scientific Linux 6 benchmarks were run later as the batch farm was set to be updated.
- Scientific Linux 6 performed slightly better than Scientific Linux 5. The overall trend was identical
- Power, temperature, disk I/O and batch server performance were closely monitored
- The results indicated a nearly linear increase in the hepspec scores, flattening at about 14 threads for 2 CPU 4 core nodes and 20 threads for 2 CPU 6 core nodes
- The revealed sweet spots were then configured for use in the batch farm to discover where production VO jobs would perform optimally

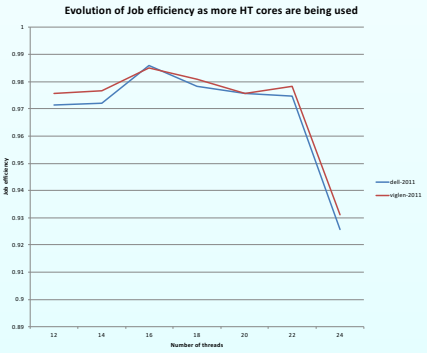



## Results

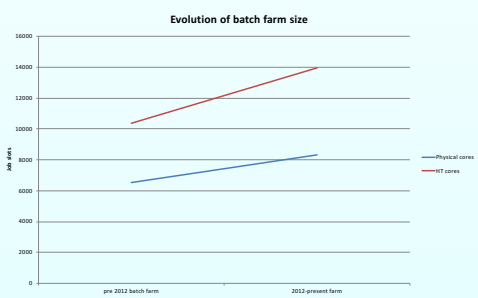
- Overall, 2000 extra job slots and 9298 extra hepspec were added in the batch farm using already available hardware
- Average job time increases as expected, but overall job throughput increased
- Network/disk/power/temperature usage did not increase in a way that could negatively affect the overall throughput or require additional maneuvers
- Batch server was able to handle the extra job slots
- **Of critical importance is the sharp drop in job efficiency as job slots approach the upper hyperthreading limit. This means that real world VO jobs would suffer if we went for full batchfarm HEPSPEC performance!**

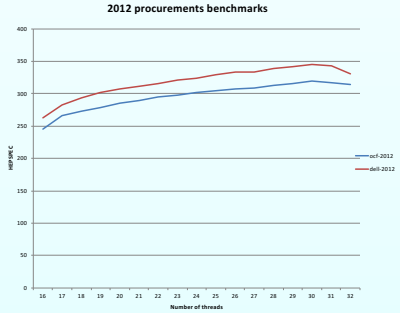
| Make | Job Slots per WN | Efficiency | Average Job Length (mins) | Standard Deviation (mins) | Number of jobs |
|---|---|---|---|---|---|
| Dell | 12 | 0.9715 | 297 | 370 | 19065 |
| Viglen | 12 | 0.9757 | 320 | 390 | 23864 |
| Dell | 14 | 0.9719 | 338 | 326 | 6118 |
| Viglen | 14 | 0.9767 | 270 | 341 | 11249 |
| Dell | 16 | 0.9859 | 343 | 254 | 6550 |
| Viglen | 16 | 0.985 | 304 | 249 | 8756 |
| Dell | 18 | 0.9781 | 377 | 390 | 5014 |
| Viglen | 18 | 0.9808 | 350 | 391 | 6263 |
| Dell | 20 | 0.9758 | 318 | 346 | 11339 |
| Viglen | 20 | 0.9756 | 260 | 285 | 11229 |
| Dell | 22 | 0.9747 | 387 | 315 | 6317 |
| Viglen | 22 | 0.9783 | 305 | 236 | 6307 |
| Dell | 24 | 0.9257 | 544 | 373 | 6650 |
| Viglen | 24 | 0.9311 | 372 | 278 | 6713 |




## Conclusions

- New procurements now take into account the hyperthreading capabilities
- For 2012, dual 8 core CPU systems go up to 32 virtual cores
- Systems were procured with 128 GB RAM in order to exploit full hyperthreading capabilities
- Dual Gigabit links, in the future single 10 GB as they became more cost effective
- So far RAID0 software raid setup has proven sufficient for disk I/O
- Performance gains so far on par with previous generations
- By spending a bit extra on RAM, we save more by buying fewer nodes
- This also saves machine room space, cables, and power




Acknowledgements: RAL Tier1 team, especially John Kelly and James Adams

GridPP
UK Computing for Particle Physics