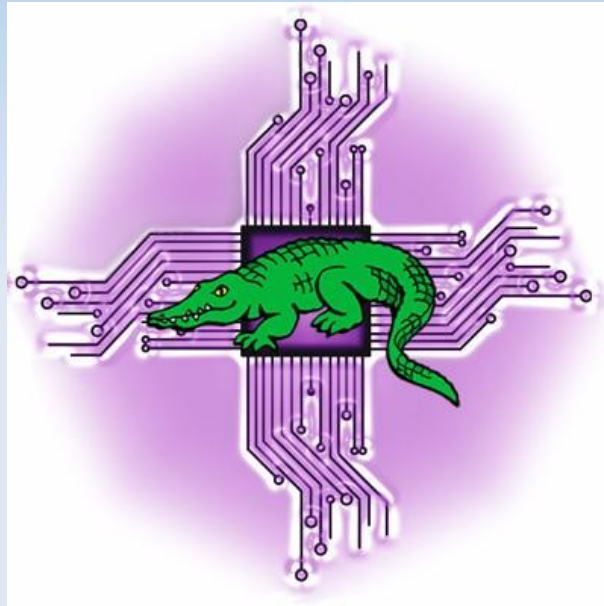


Spartan: An HPC-Cloud Hybrid



Linux Users of Victoria, September 6, 2016

lev@levlafayette.com

Clusters and Clouds

- **Tightly coupled compute clusters provide a high performance architecture multi-node multi-core task parallel computational problems. Their key feature is *performance*.**
- **Loosely coupled virtual machines in a cloud computer architectures offer shared processing single-node resources on demand. Their key feature is *flexibility*.**
- **In the past these two solutions have been developed and implemented independently.**
- **However a new hybrid system called HPC/cloud architecture developed at the University of Melbourne combines the two approaches with Linux as the core operating system, SLURM as a workload manager, and OpenStack for the cloud component.**
- **Spartan is a new architecture which combines ~~jobs and growth~~ performance and flexibility.**

Spartan is a Chimera

- The University's previous general purpose HPC system was Edward (named after the King of Wessex), the one prior to that was Alfred. However the next one would have been Æthelstan or Ælfweard!
- Spartan was originally named because of the small HPC. But it has other mythological and historical connotations. The architecture fits Sparta's citizen structure.
- This is not “HPC in the Cloud” but rather a chimera, HPC and the Cloud combined into a single system and retaining the benefits of both.

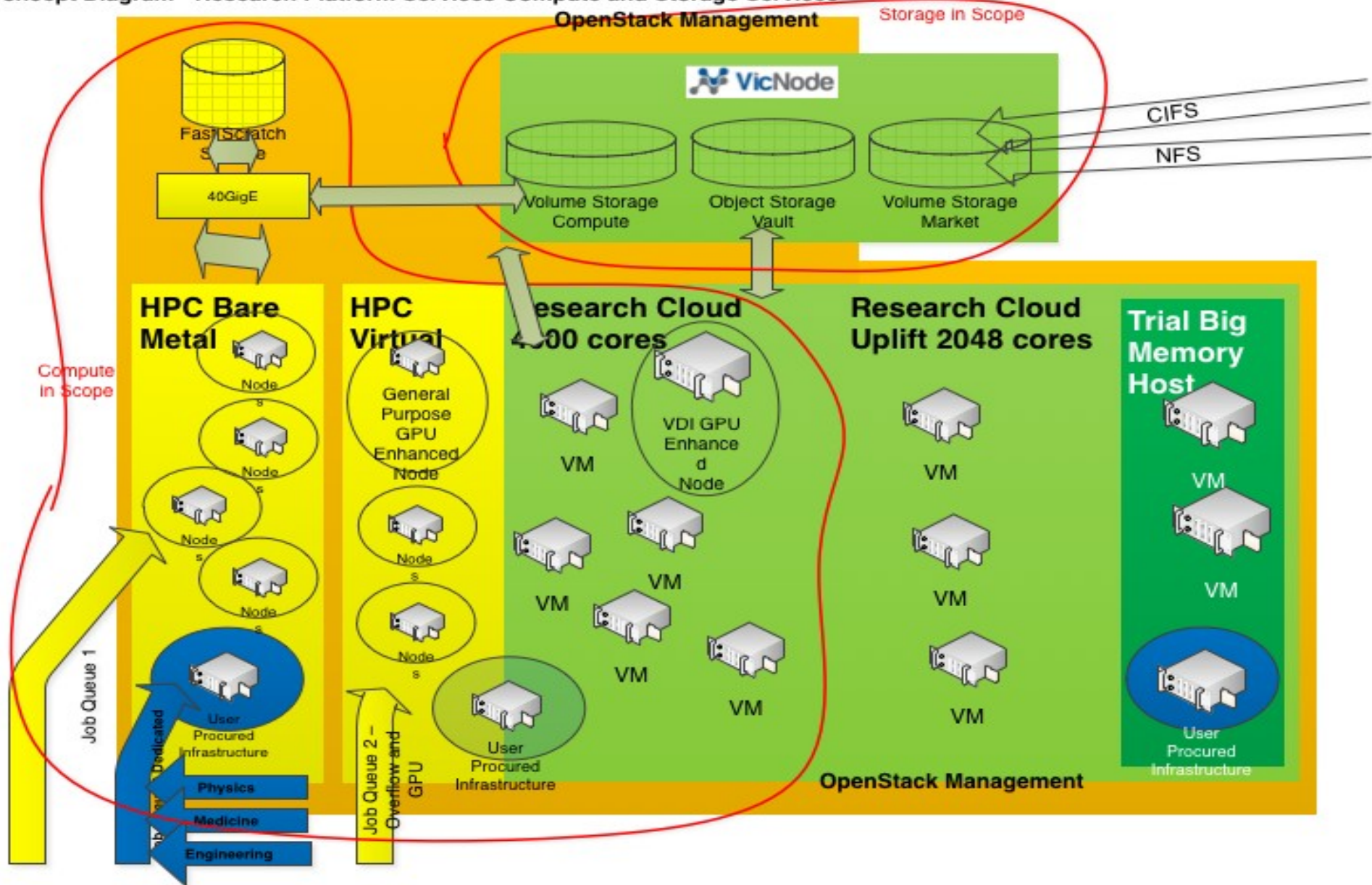


Profiling Users

- **A detailed review was conducted last year looking at the infrastructure of the Melbourne Research Cloud, High Performance Computing, and Research Data Storage Services. Single-core and low memory jobs dominated on Edward; 76.35% of jobs from Feb 9 2015 to Feb 9 2016 were single core, and 96.83% used 1-4GB of memory.**
- **Not a good use of resources, not good for people who wanted to run multinode multicore jobs. Also, not much user education. Only 38 student/days training in HPC from 2012-2014.**
- **Sparta's architecture is designed for high throughput; the multinode task parallel jobs can go on the physical nodes, and the singlenode data parallel jobs can go on the cloud nodes.**
- **The training programme has been ramped up as part of the ResPlat programme, weekly courses in Introductory HPC, transition from Edward, Advanced Linux and shell scripting, parallel programming.**

Spartan's Architecture

Concept Diagram - Research Platform Services Compute and Storage Services

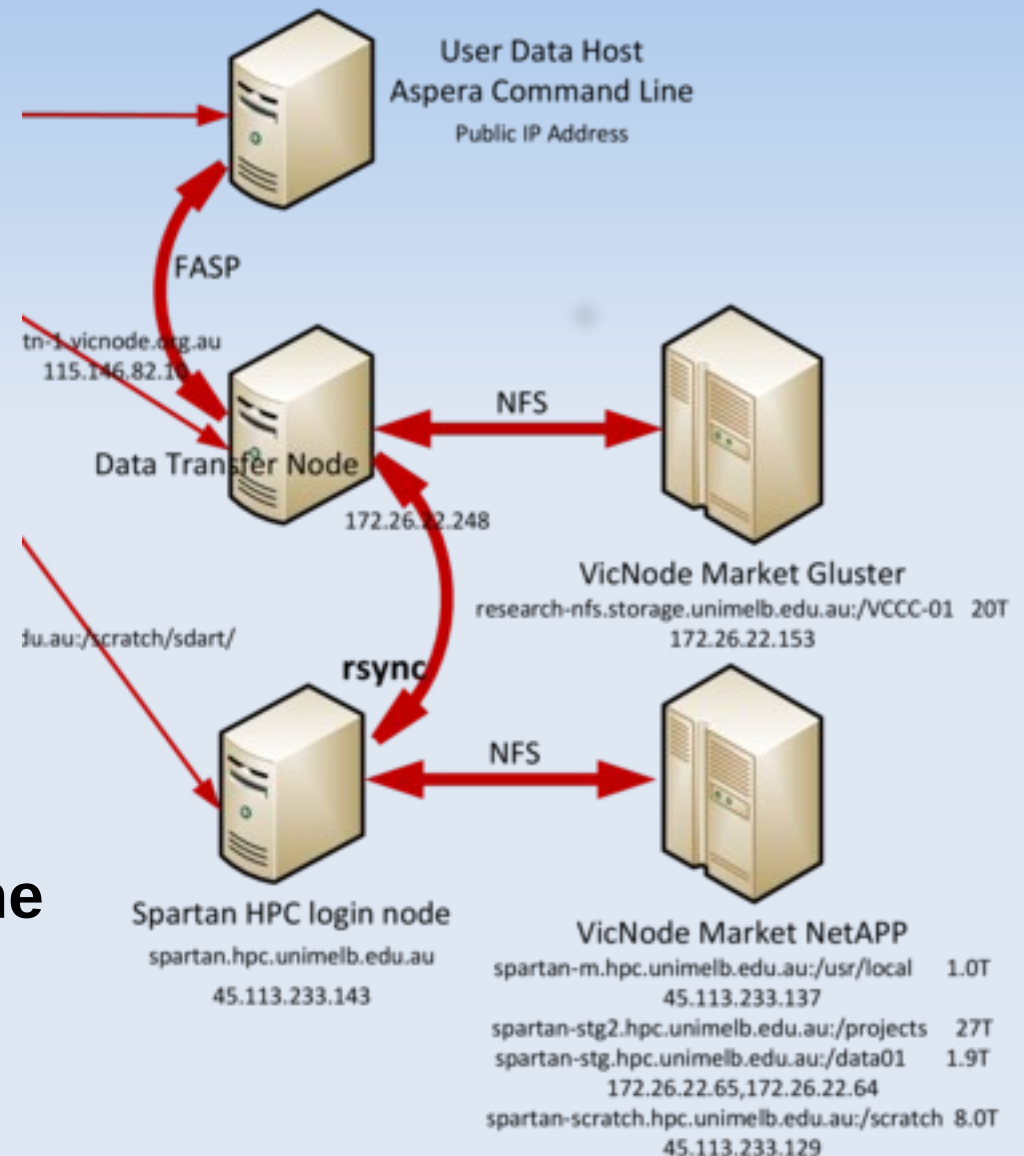


Physical Node Architecture

- **Virtual machines are used for management, login, and i/o nodes. A relatively small "physical" partition provides the HPC performance, whereas a larger "cloud" partition provides the Cloud flexibility.**
- **Spartan-m and Spartan: 16 core Intel Xeon E312xx @ 2294MHz, 64 GB RAM**
- **Spartan Bare Metal: 12 core Intel(R) Xeon(R) CPU E5-2643 v3 @ 3.40GHz, 294 GB RAM**
- **Spartan Cloud: 16 Intel Xeon E312xx @ 2294MHz, 64 GB RAM**
- **Currently relatively small 1024 cloud cores and 228 physical cores, but with the opportunity to scale easily.**
- **In comparison Edward compute nodes were 16 cores per node, AMD Opteron 6128, 2000.000 GHz, 32 GB memory, 752 cores total (plus specialist and departmental queues).**

Network Infrastructure

- The interconnect between the cloud partitions is 10gb Ethernet; with the added latency of the Neutron core service in OpenStack.
- The physical partition uses 25gb/56gb Ethernet with Mellanox SN2700 and SN2100 leaf switches with 100G connections to core. There is Fast CEPH storage from the spine, and multiple 56gb scratch storage nodes from the leaf.
- Mount points from Spartan to VicNode.

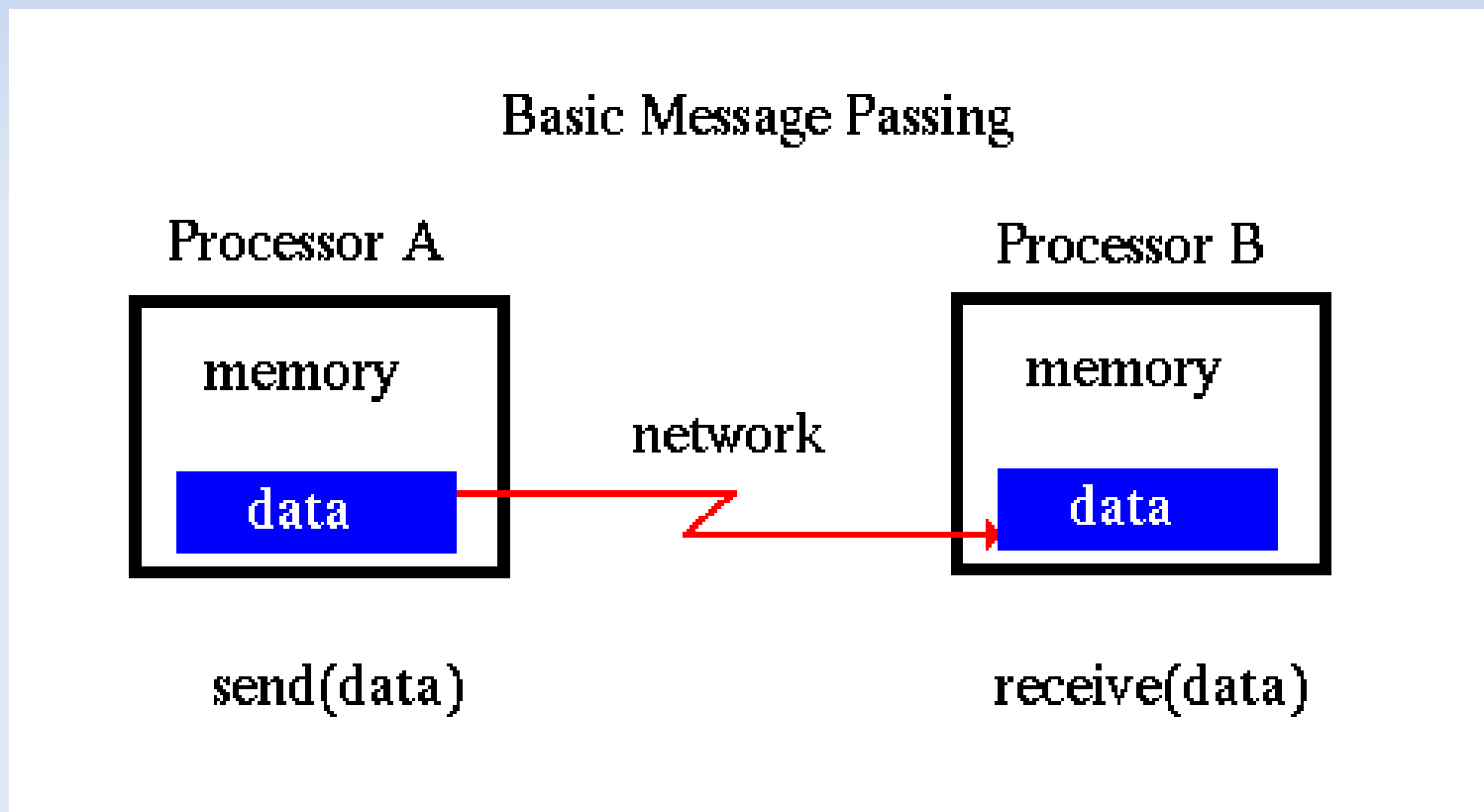


Software Infrastructure

- **Using Linux version 3.10.x and Red Hat Enterprise Linux Server release 7.2**
- **Consistency in operating system and libraries in the management, login, i/o, physical, and cloud compute nodes maintained through Git, Puppet with reviews with Gerrit.**
- **SLURM Workload Manager for resource manager and job scheduler. User and group management and reporting with Karaage.**
- **EasyBuild is used for software installations; builds from source codes with specified toolchains and dependencies. Automatically generated LMod environment module files.**
- **First jobs ran in February this year. Big test was with COMP90024 Cluster and Cloud Computing course in March - almost 10000 MPI jobs run over a single weekend by over 100 users.**

Latency Tests – Live Demo!

- Comparison with `mpi-pingpong.c` with OpenMPI/1.10.2-intel-2016.u3 on Spartan on physical and cloud partitions. (Edward is currently at 19.98 usec).



Experiences and Future

- **A couple of interesting challenges: Active Directory integration for logins, noisy neighbour effects with overcommitted cloud nodes.**
- **Introduction of cloudbursting. SLURM allows for dynamic allocation of compute nodes to a partition, elaborating from its technology to switch nodes off and on as needed.**
- **Multiple login nodes for specialist events (e.g., student classes).**
- **Incorporation of departmental specialist compute partitions, these will have different architectures to Spartan, and require recompiling of the software (that is, separate /usr/local mounts). Hekatonkheires is Spartan!**
- **This HPC/Cloud chimera architecture will become the standard for such research compute platforms in the future.**

THANKS FOR WATCHING



& LISTENING PATIENTLY