HPC-related Cloud Computing Research at the ANU

Peter Strazdins
Computer Systems Group,
School of Computer Science,
The Australian National University

Fujitsu Australia & ANU Cloud Workshop ,
ANU, 23 July 2013

(slides available from http://cs.anu.edu.au/~Peter.Strazdins/seminars)

THE AUSTRALIAN NATIONAL UNIVERSITY

1 A Dynamic Scheduling Framework for Heterogeneous Clouds

- prior work: ARRIVE-F: a open-source profiling/dynamic resource scheduling framework for (Xen) virtualized heterogeneous compute farms
  - *transparently* profiles execution of each (MPI) parallel job (using PMPI & Oprofile), and migrates jobs to more appropriate sub-clusters
  - overall throughput improvement of 25% on a 32 VM compute farm comprised of homogeneous sub-clusters
- problem: compute resources within (and across) cloud providers tend to be heterogeneous $\Rightarrow$ workloads may not be allocated appropriately
- opportunity: develop Arrive-F for clouds!
  - within a cloud provider (also generalize for use in data centers)
  - *between* cloud providers. A motivating scenario:
    - initially run and profile job on NCI/NF cloud (like the current vayu cluster)
    - migrate jobs with "undesirable' profiles to (cheap) cloud providers
    - requires a Lustre-like implementation on a cloud
- issues: generalizing for clouds, taking energy & SLA into account

2 Clouds for Large-scale Scientific Data Processing

- context: NCI hosts national IPCC climate data (2 PB NetCDF files) and water resource resources data
- typical usage scenarios
  - climate model re-analysis & evaluation (e.g. by CMIP5 framework)
  - data requests sent to special cloud hosting data
  - issues: time-skewed data, different co-ordinate systems, creation of derived data
  - scale of system: some results may be long-delayed
- challenge: need infrastructure to support development of workflows, integration with Hadoop, scheduling, (potentially) asynchronous transactions (when long delay for result)
  - evaluate applicability of the cloud model for large-scale climate data and simulations
- collaborators: NCI/NF, BOM

3 HPC Applications on Clouds (work with NCI/NF in 2012)

- comparison of systems with similar nodes (8 core Nehalems):

platform	private cloud: DCC	public cloud: EC2	premiere cluster: vayu
virtualization	VMware ESX 4.0	Xen	–
file system	NFS	NFS	Lustre
interconnect	1 GigE (dual)	10 GigE	QDR IB

- easily rsync'ed vayu's `/apps` (system-wide compilers, libraries) onto VMs to replicate software stack on clouds
- benchmark results (OSU, NPB class B, Chaste & MetUM applications):
  - OSU communication micro-benchmarks trends as per theoretical specifications ($> 10\times$ faster on vayu)
  - marked jitter on DCC (OSU) and EC2 (NAS EP) from CPU scheduling
  - scaled $\approx$ linearly on vayu to 8 nodes, but only to 2 on DCC & EC2
  - DCC particularly slow for stages using a large number of collectives
  - DCC had greater degree & higher irregularity of load imbalance (NUMA)
  - I/O stages: DCC much slower than, EC2 similar to, vayu

4 HPC Applications of Interest for RSCS (in collaboration)

- applications: performance on cloud, w/o GPGPU or Phi acceleration:
  - MetUM global atmospheric model (NF)
  - Chaste cardiac simulation
  - ANUGA tsunami propagation (MSI)
  - Self-Orienting Maps (SOMA) (CSIRO)
  - Landsat image processing (NF)
- infrastructure:
  - acceleration of RAID & ZFS backend for Lustre by GPGPU and Phi (NF)
  - acceleration of the OpenStack environment: device virtualization (NF)