Smart Connector and Scheduler on the Cloud

Iman Yusuf, Ian Thomas, Heinz Schmidt
e-Research Office, RMIT University

The 6th eResearch Australasia Conference
October 29th, 2012
The Problem

• **Research Software**: Executable code for Hybrid Monte Carlo Simulation in Physics
• **Needs**: scalable, flexible & fast computation
• **RMIT e-Research**: move some HPC users to cloud
• **Luck**: embarrassingly parallel program, not IO-bound, can be cloud-enabled (in principle)
• **Cloud** context: IaaS by NeCTAR
Requirements

• “What is cloud computing and why should I care?”
  – Maximising bang for buck
  – Faster and/or cheaper computation
  – Scalable to fit available resources
• Using familiar interfaces and tools (CLI, script)
• Packaging software environment (as VMs) for SaaS Monte Carlo Fortran code
Generalising ...

• Researchers need to
  – Repeat the whole process
  – Do the same process with slight variation
  – Do similarly for other scientific codes

• External (non-RMIT) researchers should find our solution useful, too

Smart Connector, Smart Scheduler
Smart Connector

• Connects researcher’s desktop app to cloud
• Talks to the cloud service (IaaS and PaaS) on behalf of the researcher
  – User accessibility
  – Transparent automation
  – Error Handling

• Emphasizes cloud access abstraction
  – Concrete prototype implementation exists
  – Different implementation possible
Smart Connector

• Respects the architectural style of the underlying HPC computation, e.g. Monte Carlo Simulation:
  – With stages create, setup etc
  – With constraints on stage sequences, triggers, ...
  – With predictable performance
Smart Connector

CREATE

Create Virtual Machines

Setup Virtual Machines

install dependencies, Monte Carlo program

SETUP

RUN

Run Monte Carlo Simulation

upload input data and start simulation

COLLECT

Collect Result

transfer output to user designated location

TERMINATE

Terminate Virtual Machines

kill VM instances
Example: Variation Points

Basic Smart Connector

- Run Monte Carlo Simulation
- Collect Result

Daisy-Chain Smart Connector

- Run Monte Carlo Simulation
- Collect Result
- Transform
- Check Convergence
- Collect FINAL Result

MapReduce

Combinational Logic
Smart Scheduler

Switches smart connector to Cloud provider and passes relevant parameters
Smart Scheduler

- Aware of resources (Amazon, NeCTAR)
- Chooses feasible service candidates by
  - The provided service characteristics
  - The required architectural style and resources
  - The required quality of service
- Ranks cloud service providers among candidates
  - Within above constraints
  - By cost
- Presents ranked choices to users for decision
- Switches smart connector to selected cloud provider
Smart Scheduler

- 2 VMs
- Low cost resource
- 20 VMs now
- Willing to pay
- ...
Bioscience Data Platform: Tardis in the Cloud
(NeCTAR Research Tool)
Conclusion

• Pragmatic: addresses researchers’ problems
• Familiar interfaces and concepts for researchers (CLI, towards web)
• Allows sharing of results and future expansion
• Open-source software solution
• A work in progress, part of the NeCTAR Bioscience Data Platform project
• Initial engagement to spark researcher’s interest in the potential of the research cloud
Credits

• NeCTAR funded project, No. 2179, led by Monash University
• Salvy Russo, Daniel Drumm, Chemical and Quantum Physics, School of Applied Sciences, RMIT University
• Vector clip arts
  – Creative Commons License

Questions