



Digital Institute

Intelligent Power Management over Large Clusters

Stephen McGough*, Clive Gerrard*, Paul Haldane*,
Sindre Hamlander+, Paul Robinson+, Dave Sharples*,
Dan Swan*, Stuart Wheeler+

* Newcastle University
+ Arjuna Technologies Ltd

Condor Week 2011





Digital Institute

Overview

- What is the DI?
- Background
- Motivation
- Power saving in Condor
- Intelligent Power Management
- Conclusion

What is the DI?

Society is increasingly dependent on digital technologies

energy, transport, healthcare, commerce, engineering....

How can Newcastle researchers develop and harness these technologies to shape the future for everyone's benefit?

Research is increasingly dependent on digital technologies

record, store, analyse, model, share, visualise..

How can Newcastle researchers harness these technologies to do new, innovative research?



Institute Goals

- Facilitate Internationally-Leading Research
 - innovative, exciting, pioneering
 - impact on researchers & society
- Thriving, Inter-Disciplinary Research Community



Digital Institute

Background



Digital Institute

Condor at Newcastle

- Using open access cluster computers around campus
 - ~ 1300 Computers Windows Based
 - ~ 250 Computers Linux Based
- Using Staff Desktops
 - ~ 3000 Computers (not all Condor enabled – yet!)
- Using dedicated resources
 - ~ 200 Computers in research groups around Campus
- All computers at least dual core, moving to quad
 - ~ 6000 – 10000 core cluster
- Condor originally installed 5 years ago as ‘unsupported’ service. Now we’re making it a supported service



Power Efficiency of Computers

Digital Institute

- Power Efficiency:

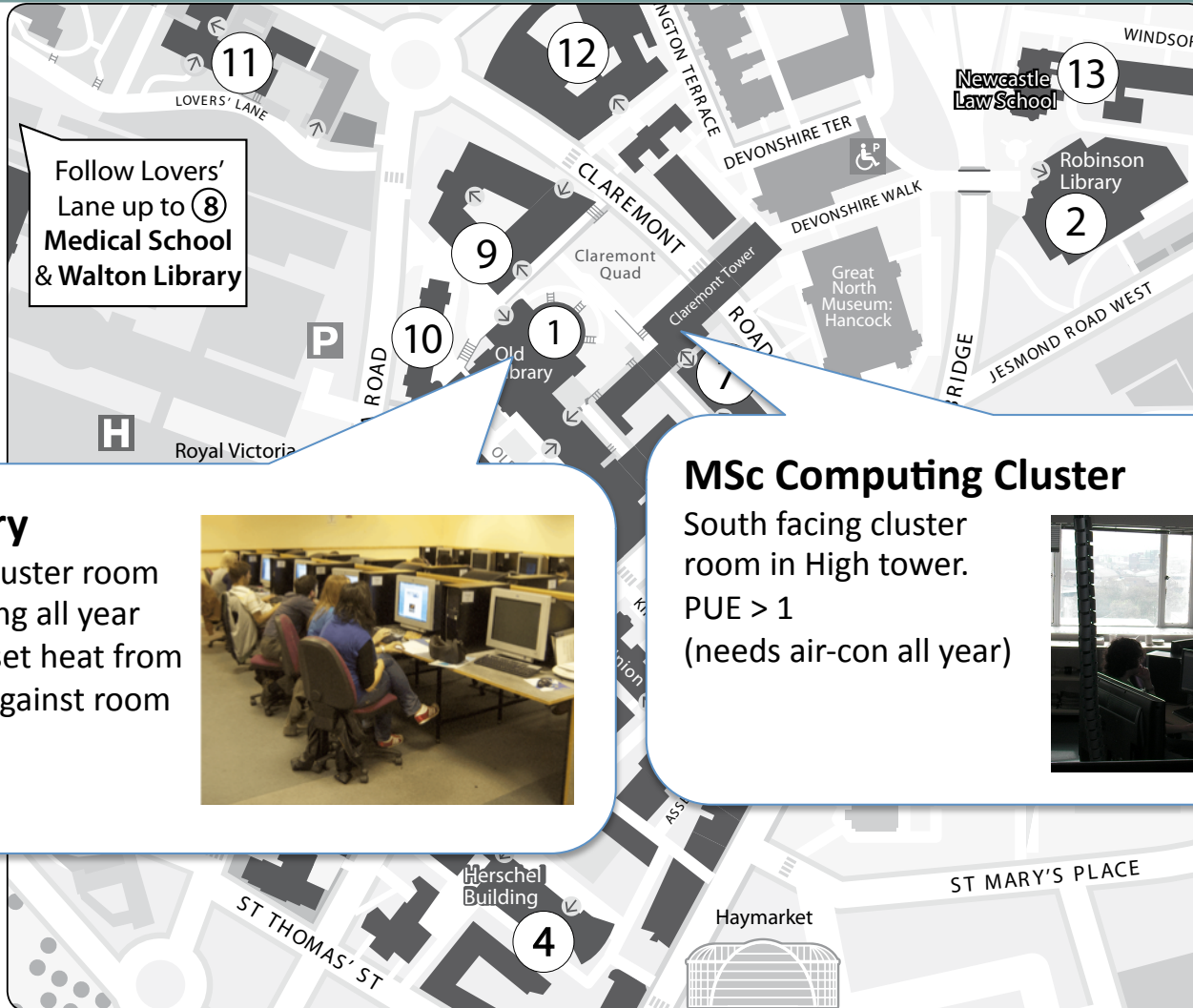
$$\text{efficiency} = \text{flops} / (\text{PUE} * \text{watts})$$

- Flops/watts – pre-determined
 - Average value used

Type	flops	Wattage
Ultraslim Desktop	941M	54
CAD Computer	1265M	87
Legacy	600M	100

- Power Usage Effectiveness (PUE) – depends on location of computer (and time)

Cluster Locations



Old Library

Basement Cluster room
Needs heating all year
PUE < 1 (offset heat from computers against room heating)



MSc Computing Cluster

South facing cluster room in High tower.
PUE > 1
(needs air-con all year)



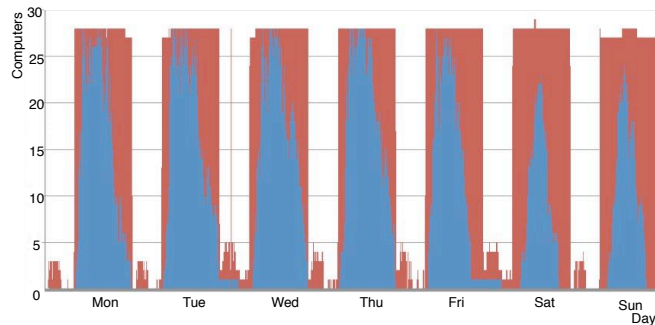


Digital Institute

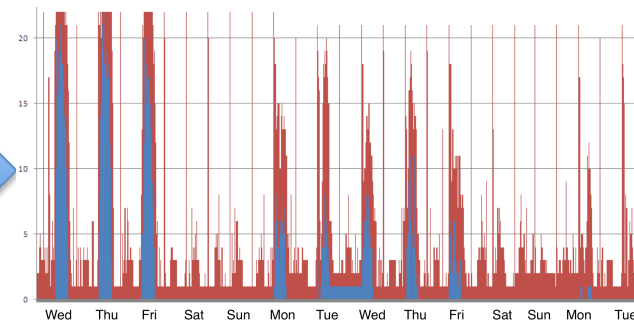
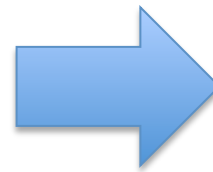
Motivation

Motivation

- University has strong desire to power down computers to save energy (money)
- If a computer is not 'working' it should be powered down.



Windows XP Cluster



Windows 7 Cluster

Motivation

- We have five years of condor logging to mine
 - Other computers were used
 - though this is the main log
 - Initial analysis looks promising...

Description	Number	%
Total number of jobs submitted to Condor	642298	100%
- Which Completed	469184	73%
- Without wasted time	415807	65%
- With wasted time	53377	8%
- Which were removed	173114	27%
- Before execution	164059	26%
- After some execution	9055	1.4%

Motivation

- Further analysis didn't look as good...

Total Time used by Condor	51 years, 29 days, 20h, 2m, 52s	100%
Total Wasted time	33 years, 129 days, 4h, 6m, 13s	65%
Total time wasted on jobs removed	26 years, 33 days, 22h, 36m, 28s	51%
Total time wasted on completed jobs	7 years, 95 days, 5h, 29m, 45s	14%
Total real job execution time	17 years, 265 days, 15h, 56m, 39s	35%

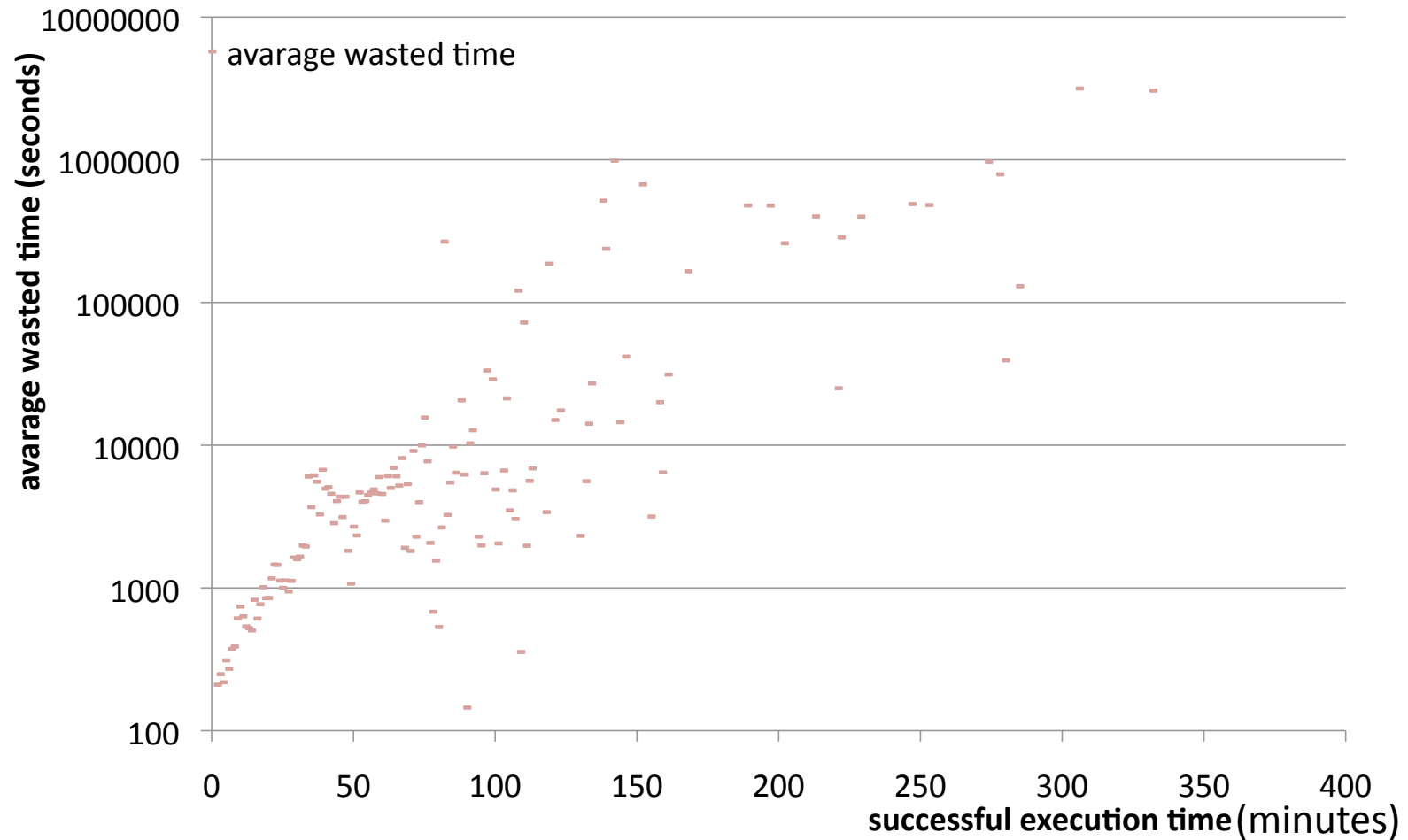
(Data from main Condor Submit computer – others exist)

- For every 1 second of useful time we require ~3 seconds through Condor



Average Wasted Time

Digital Institute



- Rank used to express preference

Rank Statement	%	Meaning
Null	99.87%	No Preference
LoadAvg	0.082%	Prefer computers with higher loads !!!
((OpSys == "WINNT51")) + ((OpSys == "LINUX") * 2)	0.047%	Prefer Linux to Windows
Memory	0.0015%	Prefer computers with more memory

- Vast majority of people don't use Rank
 - Some who do get it wrong!

Aims

- Reduce Power Consumption
 - Produce no impact on interactive machine users
 - Produce no impact on Condor users
 - Provide auditing on computing time used
 - On a cycle-scavenging Condor system.
-
- All These Things can be done using Condor
 - So What's new?



Digital Institute

Power Saving in Condor



What's Wrong With Condor?

Digital Institute

- Nothing – Condor is “too good”, but...
 - With a good administrator all this can be done
 - With plenty of time this can be done
- Put another way
 - Administrators are busy people
 - They have plenty of other tasks to do other than monitor and tweak Condor which gets along with things just fine
- Most users don't know enough Condor
 - Can't specify preferences over resources
 - Only 3 users had done this at Newcastle – and one got it wrong



Digital Institute

Power Saving in Condor

- Power rating (watts) and a PUE added to each worker description
- Rank equation needs to be added to each job submission

$$\text{Rank} = \text{flops} / (\text{PUE} * \text{watts})$$

- Automatically added by our system
- And merged with user rank - if present

Powering up and down

- We use our own computer power-down script
 - When no-one is using the computer
 - When no Condor jobs are running
- A persistent ClassAd is sent to Condor
 - Indicates computer is powered down
- Rooster can power up the computer
 - If a job description requests this
 - This calls our wakeup script



Powering down Windows

Digital Institute

- We suspend computers
- Windows can (wake up and) suspend for many reasons
 - Not all under our control
 - Anti-Virus updates, software updates
- We can catch many of these
 - But ~20% we can't
- We monitor the cluster for these
 - Monitor condor_status for nodes which don't update
 - If a node fails to update ping it
 - If it's no longer there copy it's last classad and post a fake replacement for a hibernating computer



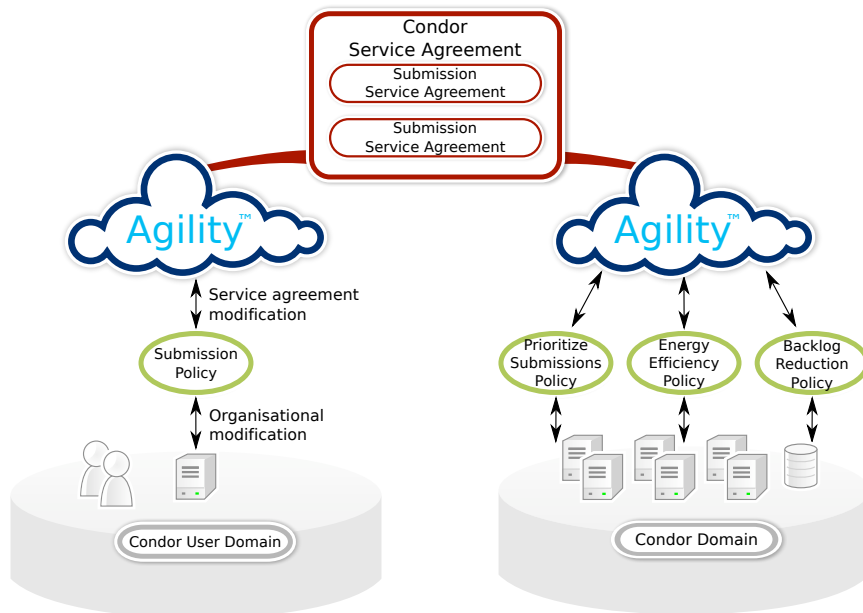
Digital Institute

Intelligent Power Management

Aims

- Reduce Power Consumption
 - Produce no impact on interactive machine users
 - Produce no impact on Condor users
 - Provide auditing on computing time used
 - On a cycle-scavenging Condor system.
-
- All These Things can be done using Condor
 - So What's new?
 - Do this all in a more automatic 'intelligent' way

- Arjuna's Agility is a federation tool
 - Often used with Cloud Computing Platforms
- Support structuring of the Cloud into federated sub-clouds through Service Agreements



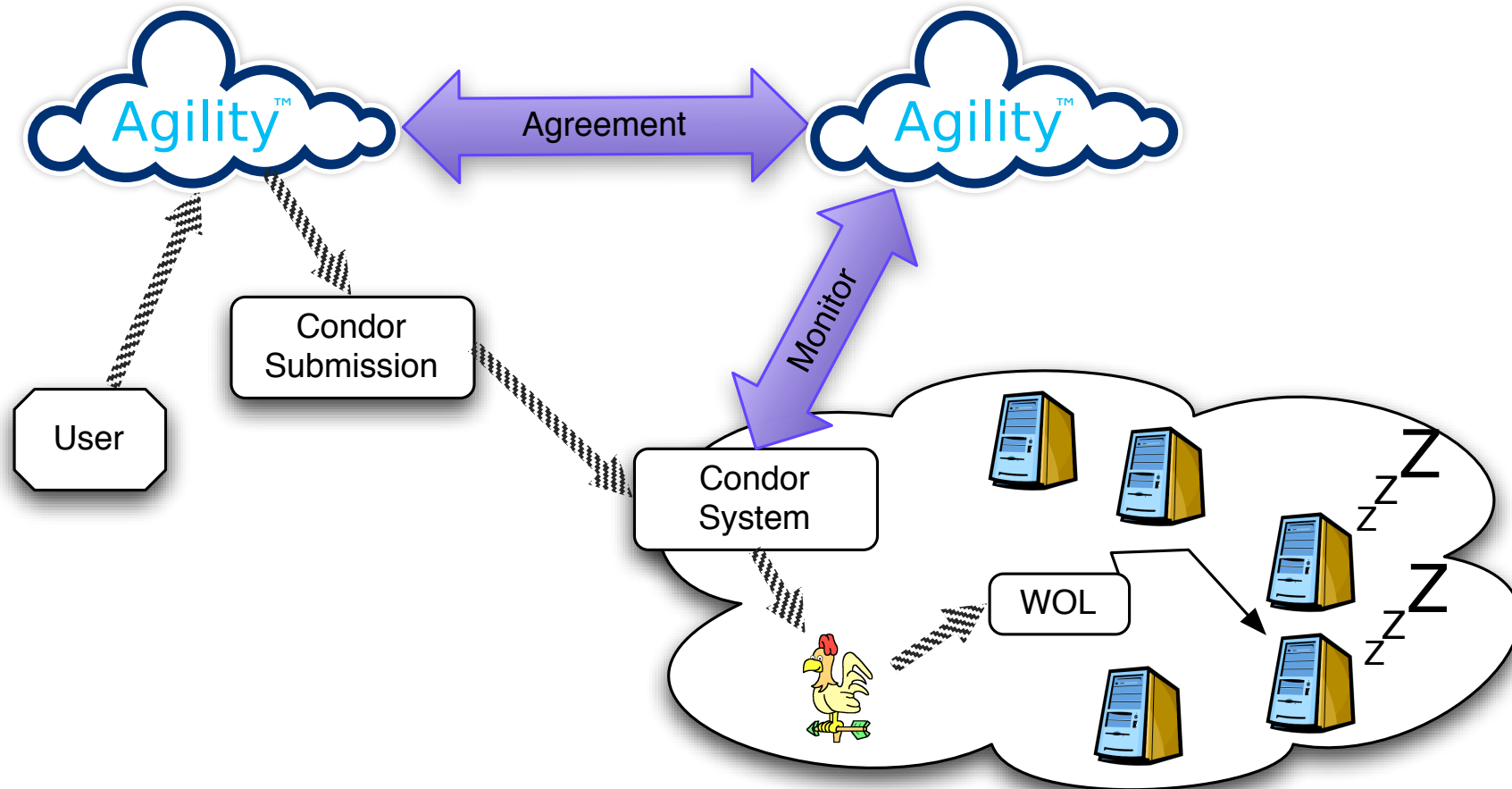
Agreements

- User configured (or implemented) policies can be installed to manage the interaction via Service Agreements

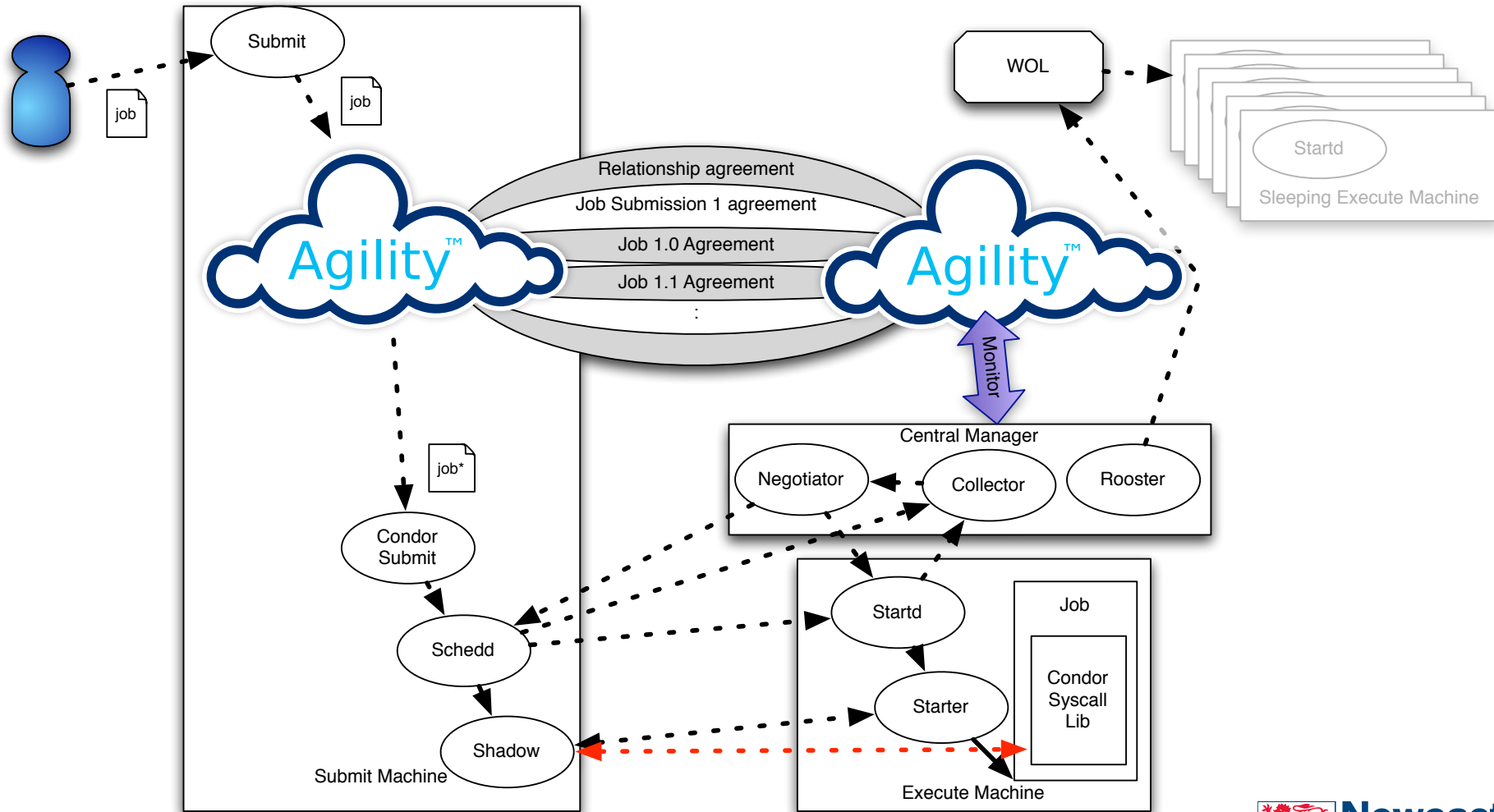
Digital Institute

- Using Agility to provide intelligent management between parts of Condor
- Adding in Rank
- Deciding if a user has the right to power up computers
- Monitoring activity and looking for anomalous behavior
- Behavior is modified by the addition of new Agility policies
 - A Service Agreement change will be accepted only if no Policy rejects it and at least one Policy accepts it
 - Modification of Agility policies can lead to modification of the condor policies and/or modification of incoming jobs
- Audit individuals' worker usage for potential billing

General Architecture



Job Submission Route





Defined Policy

Digital Institute

- Favor Energy efficient computers
 - Unless user states otherwise
- Prioritize submissions
 - Professors over PhD's
- Mark and manage Rogue jobs
 - Jobs that have executed for too long
 - Jobs that have been restarted too many times
 - Jobs incorrectly submitted
- Backlog reduction
 - Modify above policy to deal with backlogs
- Auditing
 - Provide auditing for all jobs (per –user, -group, -school, -faculty, university)

Auditing

- Agility stores all attempted job runs. We can use this to provide audits of computer use
- A job which takes n attempts to run consumes power of:

$$\sum_{i=1}^n PUE_i * watts_i(end_time_i - start_time_i)$$

- Do we include evicted job time?
- Can produce audit on user, group, school or university

- Federating between Condor Pools
- Federating with the Cloud – Cloud Bursting
- Resource profiling
 - Working out when resources are idle and for how long
- Job profiling
 - Try to estimate job lengths for users
- Dedicated workers for evictees
- Running jobs on computers with active users
- Education of users

Conclusion

- Condor is good
 - Most users aren't as good
 - System administrators are busy
 - They need tooling to automatically monitor and tweak the setup
- This along with preferring power efficient computers and turning off computers can save a lot of energy (and money)



Digital Institute

Questions?

stephen.mcgough@ncl.ac.uk