The Evolution of Cloud Computing in ATLAS

Ryan Taylor

on behalf of the ATLAS collaboration
Outline

● Cloud Usage and IaaS Resource Management
● Software Services to facilitate cloud use
● Sim@P1
● Performance Studies
● Operational Integration
  – Monitoring, Accounting
The Clouds of ATLAS

ATLAS cloud jobs (Jan. 2014 – present)
61% Single-core production
33% Multi-core production
3% User analysis
IaaS Resource Management

- HTCondor+Cloud Scheduler, VAC/VCycle, APF
- See talk 131 “HEP cloud production using the CloudScheduler/HTCondor Architecture” (C210, Tue. PM)
- Dynamic Condor slots to handle arbitrary job requirements
  - e.g. single-core, multi-core, high-mem
- uCernVM image
- Contextualization using cloud-init
- Using Glint Image Management System
  - see poster 304
Shoal
Proxy Cache “Federator”

• Build a fabric of proxy caches
  – configurationless topology
  – robust
  – scalable

• Needed to run uCernVM at scale
  – By default, DIRECT connection to closest Stratum 0/1
  – Contextualize instances to find proxy using Shoal

```
[ucernvm-begin]
CVMFS_PAC_URLS=http://shoal.heprc.uvic.ca/wpad.dat
CVMFS_HTTP_PROXY=auto
[ucernvm-end]
```

• Also use Shoal for Frontier access
  – Currently under investigation
Sim@P1

- Resource contribution similar to T1
  - 34M CPU hours, 1.1B MC events
- Used for LHC stops > 24h
- Fast automated switching via web GUI for shifters
  - TDAQ to Sim@P1: 20m (check Nova DB, start VMs)
  - Sim@P1 to TDAQ: 12m (graceful VM shutdown, update DB)
  - Emergency switch to TDAQ: 100s (immediate termination)

- See poster 169
HS06 Benchmarking Study

- Commercial clouds provide on-demand scalability
  - e.g. urgent need for beyond pledged resources
- But how cost-effective are they?
- Comparison to institutional clouds
T2 & Remote Cloud Performance Comparison

- Used Hammercloud stress tests (24 hour stream)
- Data and squid cache at grid site
  - Remote access for cloud site
    - like zero-storage processing site

Success rate similar

HC 20052434
MC12 AtlasG4_trf 17.2.2.2
• Software setup time
  – Relies on CVMFS cache and Squid proxy
  – VMs have to fill up empty cache

• Data stage-in time
  – Local vs. remote storage access

\[ (15 \pm 7) \text{ s} \quad (45 \pm 15) \text{ s} \]

\[ (11 \pm 4) \text{ s} \quad (54 \pm 20) \text{ s} \]
• Total running time
  – 1.5x longer on cloud
  – different CPUs
  – hyperthreading?
  – data & software access time not significant

• CPU efficiency equal!

• Cloud usage is efficient for this workload

• No significant performance penalty
Cloud Monitoring

- VM management becomes the responsibility of the VO
- Basic monitoring is required
  - Detect and restart problematic VMs
  - Identify “dark” resources (deployed but unusable)
  - Can identify inconsistencies in other systems through cross-checks
- Common framework for all VOs
- Implemented with Ganglia
- http://agm.cern.ch
Cloud Accounting

- Provider-side: commercial invoice for resources delivered
- Consumer-side: record resources consumed
- Need to cross-check invoice against recorded usage!

http://cloud-acc-dev.cern.ch/monitoring/ATLAS
Conclusion

- Increasing use of clouds in ATLAS Distributed Computing
- Performance characterization of commercial clouds
- More integration into operational model
  - accounting, monitoring, support
- Developing and deploying services to facilitate cloud use
Extra Material
Dynamic Federation
UGR

- Lightweight, scalable, stateless
- General-purpose, standard protocols and components
  - Could be adopted by multiple experiments
  - e.g. DataBridge, LHCb demo: http://federation.desy.de/fed/lhcb/
- Metadata plugin used to emulate Rucio directory structure
- No site action needed to join
  - HTTP endpoints extracted from AGIS with script
RACF/BNL Amazon Project

Enabled by $200k grant from Amazon to run all ATLAS workloads at large scale
Encompasses provisioning/compute, storage, networking, and ATLAS workflow.

VMs via Imagefactory and templates/profiles.
VM runtime config by cloud-init->Hiera->masterless Puppet.
Provisioning via AutoPyFactory, HTCondor-G. HTCondor batch pool.

3 EC2 regions and 12 instance types to maximize capacity. Spot market.
SRM/GridFTP EC2 instance w/ S3FS back end. One per region.
Ultimately S3 native storage endpoint. Job stage-in/out via S3.

10/100Gb peering and 10Gb DirectConnect to 3 regions via ESNet.
Data egress fees waived as long as <15% of total cost.
Event service nearing completion w/ S3 objectstore, active deletion, and EC2 merge jobs.
S3 storage support in Rucio/DDM.

2.5k node/20k core tested so far, 100k core final goal
List of Active Squids

5 active in the last 180 seconds

| #  | Hostname                               | Public IP          | Private IP      | Bytes Out | City       | Region | Country    | Latitude | Longitude | Last Received | Alive      | Verified | Access Level |
|----|----------------------------------------|--------------------|-----------------|-----------|------------|--------|------------|----------|-----------|---------------|------------|----------|--------------|
| 1  | squid-test01.gridpp.rl.ac.uk           | 130.248.183.249    |                 | 0 kB/s    | Appleton   | United Kingdom | 51.7     | -1.35     | 7s            | 42h40m43s   | ✓        | Global       |
| 2  | kraken01.westgrid.ca                   | 206.12.48.240      | 172.22.2.25     | 809 kB/s  | Vancouver  | Canada          | 49.2838  | -123.1041 | 10s           | 107h49m9s   | ✓        | Global       |
| 3  | atlascaq3.triumf.ca                    | 142.90.110.68      |                 | 0 kB/s    | Vancouver  | Canada          | 49.2756  | -123.2177 | 20s           | 168h20m3s   | ✓        | Global       |
| 4  | atlas-squid.cern.ch                    | 128.142.200.105    |                 | 0 kB/s    | Geneva     | Switzerland     | 48.1958  | 6.1481    | 22s           | 168h19m59s  | X        | Global       |
| 5  | t2software03.physics.ox.ac.uk          | 163.1.6.175        |                 | 35 kB/s   | Oxford     | United Kingdom  | 51.75    | -1.25     | 26s           | 168h18m56s  | ✓        | Global       |

© University of Victoria || Visit GitHub Project || Shoal-Server v0.7.1

PAC Interface

```javascript
function FindProxyForURL(url, host) {
    return "PROXY http://atlascaq3.triumf.ca:3128; PROXY http://kraken01.westgrid.ca:3128; PROXY http://t2software03.physics.ox.ac.uk:3128; PROXY http://squid-test01.gridpp.rl.ac.uk:3128; PROXY http://atlas-squid.cern.ch:3128; DIRECT";
}
```

JSON REST Interface

```json
{
    "proto": 0,
    "domain_access": true,
    "squid_port": 3128,
    "global_access": true,
    "verified": true,
    "last_active": 14290904480.168000000,
    "created": 14290904479.509000000,
    "external_ip": null,
    "geo_data": {
        "city": "Vancouver",
        "region_name": "BC",
        "area_code": 0,
        "time_zone": "America/Vancouver",
        "country_code": "CA",
        "country_name": "Canada",
        "postal_code": "V6Y"
    },
    "hostname": "atlascaq3.triumf.ca",
    "public_ip": "142.90.110.68",
    "private_ip": null,
    "last_received": 1026000000000000000,
    "distance": 0.002394311931114886
}
```

- [github.com/hep-gc/shoal](http://github.com/hep-gc/shoal)
- [CHEP 2013 Poster](http://shoal.heprc.uvic.ca)
http://cern.ch/go/d8Qj
CPU consumption Good Jobs in seconds (Sum: 35,171,369,721)
CERN-PROD - 43.30%

- CERN-PROD - 43.30% (15,228,057,667)
- BNL-ATLAS - 12.39% (4,358,950,286)
- AUSTRALIA-NECTAR - 5.16% (1,814,557,950)
- UKI-NORTHGRID-LANCS-HEP - 1.45% (510,576,626)
- RAL-RCG2 - 0.40% (141,790,156)
- unknown - 0.00% (242,844)

- IAAS - 24.78% (8,714,293,050)
- UKI-NORTHGRID-MAN-HEP - 9.76% (3,433,952,754)
- GRIDPP_CLOUD - 1.57% (551,154,992)
- UKI-SOUTHGRID-OX-HEP - 0.98% (343,132,582)
- UKI-GRIDPP-CLOUD-IC - 0.21% (74,660,812)
CPU consumption Good Jobs in seconds (Sum: 1,409,504,237,701)
RAL-LCG2 - 10.30%
BNL-ATLAS - 22.40%

145,196,832,559
315,756,298,652

- BNL-ATLAS - 22.40% (315,756,298,652)
- TRIUMF-LCG2 - 8.76% (123,507,958,791)
- CERN-PROD - 8.61% (121,296,702,907)
- FZK-LCG2 - 7.45% (105,012,391,024)
- NDGF-T1 - 5.56% (78,376,167,085)
- NIKHEF-ELPROD - 2.99% (42,105,052,337)
- SARA-MATRIX - 2.70% (38,051,259,244)
- RAL-LCG2 - 10.30% (145,196,832,559)
- CERN-P1 - 8.68% (122,284,923,093)
- INFN-T1 - 8.00% (112,780,682,709)
- IN2P3-CC - 6.87% (96,782,310,500)
- TAIWAN-LCG2 - 3.87% (54,496,246,718)
- PIC - 2.91% (41,009,430,947)
- RRC-KI-T1 - 0.91% (12,847,981,135)

http://cern.ch/go/HB9m