Volunteer Computing, Grid Computing and Cloud Computing: Opportunities for Synergy

Derrick Kondo
INRIA, France
Cloud Background

• Vision
  • Hide complexity of hardware and software management from a user by offering computing as a service

• Benefits
  • Pay as you go
  • Scale up, down dynamically
  • No hardware management, less software management
Platform Performance vs. Costs

What is the relationship?

BOINC + Cloud?
Clouds versus Volunteer Computing

- Performance tradeoffs
- Monetary tradeoffs
  - Client hosting
  - Server hosting
Method

- Use real performance measurements
- Exported BOINC project data
- Use real costs
  - Large/small BOINC projects (SETI@home / XtremLab)
  - Amazon Elastic Computing Cloud (EC2)
Stages of Project & Application

- Platform Construction
- Application Deployment
- Application Execution
- Application Completion
How long before I get X TeraFLOPS?

Strategy:
- Add to BOINC project list
- Press releases
- Forum Announcements
- Google Ad Sense
- Respond to users (leverage volunteers)

Can get over 20 TeraFLOPS within 6 months
How long to deploy my batch of tasks needing fast response time?

Strategy: Specify lower latency bounds [Heien et al.]

For 1000 tasks, ~10 minutes with $10^5$ hosts.
How many volunteer nodes are equivalent to 1 cloud node? 2.8 active volunteer hosts per 1 cloud node. (Total performance still orders of magnitude better)

Strategy: Use statistical prediction of availability
How long should I wait for task completion?

Strategy:
See BOINC Catalog for typical deadlines and compute/comm/mem ratios.

Median project latency bound: 9 days for 3.7 hour work unit (on 3GHz host). Ratio of lat. bound / exec time > 5.

Good success rates: 96.1% of WCG tasks met out of 227,000 tasks.
Monetary Tradeoffs

- Client hosting on cloud
- Not worth it and never will
- Server hosting on the cloud
- Possible solution
## Monthly Project Costs

<table>
<thead>
<tr>
<th>Component</th>
<th>SETI@home</th>
<th>XtremLab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salaries</td>
<td>10K for sys admins</td>
<td>5K → 1K Students</td>
</tr>
<tr>
<td>Electricity</td>
<td>90 for 6 servers</td>
<td>15</td>
</tr>
<tr>
<td>Network</td>
<td>2K for 100 Mbit</td>
<td>covered by university</td>
</tr>
<tr>
<td>Hardware</td>
<td>18K for servers, 25K for air conditioner</td>
<td>4K → 1K</td>
</tr>
<tr>
<td>Total startup</td>
<td>43K</td>
<td>4K → 1K</td>
</tr>
<tr>
<td>Total monthly</td>
<td>12K</td>
<td>5k → 1K</td>
</tr>
</tbody>
</table>
EC2 Pricing

Table 1. Pricing for EC2 Instances

<table>
<thead>
<tr>
<th>Instance Type</th>
<th>Cost/hour (USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Small</td>
<td>0.10</td>
</tr>
<tr>
<td>Standard Large</td>
<td>0.40</td>
</tr>
<tr>
<td>High-CPU</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Table 2. Pricing for EC2 Data Transfer

<table>
<thead>
<tr>
<th>Transfer Type</th>
<th>Cost/GB-Month (USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inbound transfer</td>
<td>0.10</td>
</tr>
<tr>
<td>first 10 TB</td>
<td>0.17</td>
</tr>
<tr>
<td>next 10-50TB</td>
<td>0.13</td>
</tr>
<tr>
<td>next 50-150TB</td>
<td>0.11</td>
</tr>
<tr>
<td>over 150 TB</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Table 3. Pricing for EBS

<table>
<thead>
<tr>
<th>Resource</th>
<th>Rate (USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage</td>
<td>0.10 / GB-Month</td>
</tr>
<tr>
<td>IO request</td>
<td>0.10 / million</td>
</tr>
</tbody>
</table>
What about Client Hosting on the Cloud?
Cost of Clouds versus BOINC

When is BOINC more cost effective than Clouds?

After ~13 days
Cost of Clouds versus BOINC (2)

What are total costs over time?

Within <3 days, BOINC is cheaper
Cost of Clouds versus BOINC (3)

How many months of BOINC can X months of Cloud buy me?

12 months of Cloud can pay for > 125 BOINC years. BOINC way more sustainable.
Equivalent Clouds given BOINC Costs

Given BOINC costs, what size Cloud can I buy?

- 1.75 TeraFLOPS, 7 TeraBytes
- 0.75 TeraFLOPS, 3 TeraBytes

2 orders of magnitude lower than BOINC
What about Server Hosting on the Cloud?
Number of Hosts over Time

Load variation exits w/ publicity, projects run out of work, etc. Clouds take care of server management. Potential to exploit clouds.
Server Costs on a Cloud

1.7 - 16.5 times cheaper to host on Cloud, but bandwidth is expensive.

How much to host BOINC server on cloud?

Monthly Costs of SETI@home on Cloud
(7068USD in total)  
(versus 43K + 12K/month)

Monthly Costs of XtremLab on Cloud
(303 USD in total)  
(versus {4.1}K + {5.1}K/month)
Cloud Resources with Given Budget

How big of a server can I support with given budget?

Many project servers are sustainable on the cloud.

4 Mbps, 2 TB

12 Mbps, 2 TB
Summary of Cloud Versus Volunteer Computing

• Performance tradeoffs
  • 20 BOINC TeraFLOPS within 6 months

• Monetary tradeoffs
  • Client hosting
    • After 13 days, BOINC more cost effective
  • Server hosting
    • BOINC server on cloud is cost-effective
      • Best for small-medium sized projects
      • Savings of at least 40%

• [http://mescal.imag.fr/membres/derrick.kondo/cloud_calc.xlsx](http://mescal.imag.fr/membres/derrick.kondo/cloud_calc.xlsx)
• [http://boinc.berkeley.edu/trac/wiki/CloudServer](http://boinc.berkeley.edu/trac/wiki/CloudServer)
Outline

Cloud

Grid

Volunteer Computing
Volunteer Computing (Compared to Grids)

**User**
- Not computer savy
- Cause errors (accidentaly) (or bring benefits)
- Shuts BOINC client machine on/off
- Doesn’t have or wants to avoid root access

**Software**
- Heterogeneous OS (most are Windows)
- Resource contention
- Firewalls

**Hardware**
- Heterogeneous size and type
- Low bandwidth network, firewalls
Synergy

Reasons

• Why Grids should use volunteer computing
• Most Grids jobs are embarrassingly parallel
• Order of magnitudes more compute power and storage and more pervasive
• Overloaded during peak periods
• Why volunteer computing should use Grids
  • Utilization is still low
  • Results are deterministic (relatively)
  • Performance is deterministic (relatively)

Ways

• Use storage as cache to put data closer to end nodes
• Use storage as archival
• Rerun computation on the fly
• Offload all runnable jobs to volunteer computing system
• New hybrid algorithms for “non-uniform” platforms: allow for global search on volunteers and refined local on Grids
• Increased compute/storage power
• Check results (spot checking w/ replication)
• Replication with reliable performance for failed jobs
• Support fast turnaround jobs
Bridging Volunteer Computing and Grids

- Standard API
- Servers that act as bridge
  - 3 central points of failure
  - 3 job queues: high latency
  - Scaling issues: 1 bridge for M Grids and N BOINC projects
- Pilot jobs
- VM for system checkpointing, security, and Linux support
Outline

Cloud

Grid

Volunteer Computing
Grids versus Clouds

- Clouds built for web services
- Grids built for data-intensive applications
  - Petabytes of data
  - Files used by groups of users
Issues with Cloud data management [Iamnitchi08]

- Current cost model
  - One “size” fits all (same true for comp)
- Data durability: existence
- Data availability: accessibility
- Access performance
- E.g. 1: archiving: need durability
- E.g. 2: caching: need availability, access perf.
Summary

- Cloud vs Volunteer Computing
  - Server hosting is cost-effective
- Volunteer Computing vs Grid
  - Pilot jobs seem like the ideal approach
- Cloud vs Grid
  - Clouds must provide different costs models for different qualities of service
Current and Future Work

- Bring Volunteer Computing to the Cloud
- C3: Cloud Computational Co-op
- RightScale Grant
- NFSC-ANR Grants between China and France
- Bring the Cloud to Volunteer Computing
- CloudComputing@home
- Bridging Grids and BOINC
- EDGES project (Peter Kacsuk, SZTAKI, Hungary)