UnaCloud: HPC without HPC

Harold Castro

hcastro@uniandes.edu.co

Department of Systems and Computing Engineering
Universidad de los Andes
Bogotá, Colombia
CeNAT, Mayo 7 2018
Our University Overview

Universidad de los Andes has nine schools: Business Administration, Architecture and Design, Arts and Humanities, Science, Social Sciences, Law, Economics, Engineering, and Medicine. It also has two centers that offer postgraduate programs: the Center for Research and Development in Education (CIFE) and an Interdisciplinary Center for Studies on Development (CIDER).
Our University Overview

- ** Nine schools: 582 full-time professors (67% of Ph.D.). This number increased by 50% between 2002 and 2012
- ** Students: 14K undergraduate, 4K graduate, 500 doctoral
- ** 1st Colombian university (QS World University Ranking). In Latin America, Uniandes ranks between 4th-8th.
- ** 343 international and 59 national agreements facilitate students, professors and researchers mobility in different contexts.
The Systems and Computing Engineering Department offers 1 bachelor degree program and 7 master programs. SCED hosts 3 research groups.
Focus on international collaborations
Research areas:

- Distributed computing
- **High Performance Computing (HPC)**
- Network design and security
- Information integration
- Multimedia and geo-reference
- Information systems
Difficult to attract skilled people to under skilled countries

The return process should be as smooth as possible
To use cloud computing to help stop brain drain
Researchers require large HPC/HTC during some peak periods (a project needs to be delivered, CFP deadlines, etc.).

Additionally, there are a lot of general or public campus computer labs used by students and these labs are idle most of the time.
THE PROBLEM

More than 4000 CPU cores
An alternative are Desktop Grids and Volunteer Computing Systems (DGVCS’s):

- Offer large scale computing infrastructures at low cost.
- Use inexpensive resources, most of them underutilized desktop computers.
- Interconnect thousands of computing resources available through Internet or Intranet environments: failures aware and prizes
- Are based on resources that are non-dedicated, distributed, highly heterogeneous, and part of independent administrative domains.
DGVCSSs in cloud computing environments

CernVM: LHC@Home
Cloud@Home
clouds@home
UnaCloud

Virtual machines to emulate user’s environment:

- Reaction to failures: live migration
- More concepts than implementations
- Extension to public clouds
- VM management
THE VOLUNTEERING PROBLEM

When a research group wants to use a DGVCS it regularly find that:

- They will need to **recode, modify or adapt every application** that is going to be executed on the DGVCS, for **several research groups and tens of existing applications** it is a complex process.

- The installation, configuration, maintaining and use of most DGVCSs require of people with **some/advanced skills in applications and IT infrastructures**.

- For using the **idle processing capabilities of hundreds of commodity desktops**, they will need to **configure manually every desktop computer** with the DGVCS software.
When a research group wants to use a DGVCS it regularly find that:

- Administrators of different computer labs do not want that external people modify the configurations of the physical machines.
- Most of the physical desktops (99%) machines available in computer labs have Windows operating systems.
- They would like to share easily with other research groups the idle capabilities available in computer labs using a shared model.
Virtual clusters are being executed on the same physical/shared commodity infrastructure.
Research groups can use on-demand HPC Services, sharing the same commodity infrastructure. This is achieved using an Opportunistic Infrastructure as a Service Strategy.
In this work we analyze the prospect and performance of using an opportunistic underlying infrastructure to support a Cloud Computing IaaS model. This is the main motivation and contribution of this research work.
UnaCloud is a Desktop Cloud implementation at Universidad de Los Andes.

UnaCloud is a platform that allows to create virtual clusters for scientific research in non-dedicated hardware, by harvesting idle resources in computer laboratories across a campus using virtualization.
- UnaCloud is an open source project: https://github.com/UnaCloud
- The UnaCloud client is deployed in 3 computer labs using 115 desktops computers.
- Desktops: Intel i7, 16GB, Windows 10 and GigE LAN.
- A Web user interface was used, providing a self-service model
- Different type II hypervisors: Vbox, Vmware,...
- Monitoring reports and statistics
UNACLOUD IaaS IMPLEMENTATION

**IAAS CUSTOMIZATION**
Customizable Virtual Clusters (CVCs) through 5 settings: software, hardware, quantity, location (optional) and execution time.

**IAAS DEPLOYMENT**
On demand CVC deployment and provision of necessary data to secure remote access.

**IAAS ADMINISTRATION**
VM operations such as: start, stop, restart, change execution time and monitoring.

**IAAS TRACEABILITY**
IaaS model traceability at user level with basic reports and statistics.

**PHYSICAL INFRASTRUCTURE ADMINISTRATION**
Physical machine operations such as: turn off, restart, logout and near-real time monitoring.
### UNACLOUD TESTING AND RESULTS

<table>
<thead>
<tr>
<th>APPLICATION NAME</th>
<th>INFRASTRUCTURE USED</th>
<th>CPU NUMBER</th>
<th>JOB NUMBER</th>
<th>TIME BY JOB (SEC)</th>
<th>EXECUTION TIME (DAYS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSGrid Model A</td>
<td>PC</td>
<td>2</td>
<td>150000</td>
<td>35</td>
<td>30,38</td>
</tr>
<tr>
<td></td>
<td>Chemical Eng. CVC</td>
<td>70</td>
<td>150000</td>
<td>85</td>
<td>2,11</td>
</tr>
<tr>
<td>BSGrid Model B</td>
<td>PC</td>
<td>2</td>
<td>150000</td>
<td>63</td>
<td>54,69</td>
</tr>
<tr>
<td></td>
<td>Chemical Eng. CVC</td>
<td>70</td>
<td>150000</td>
<td>111</td>
<td>2,75</td>
</tr>
<tr>
<td>HMMER</td>
<td>PC</td>
<td>2</td>
<td>4200</td>
<td>11700</td>
<td>284,40</td>
</tr>
<tr>
<td></td>
<td>Biological Science CVC</td>
<td>140</td>
<td>4200</td>
<td>12900</td>
<td>4,50</td>
</tr>
</tbody>
</table>

- Performance degradation perceived by owner users (students or administrative personal) is less than **3%**.

- The maximum overload of grid jobs executed on UnaCloud virtual machines is of **17%**.

- **New** technologies affect intrusiveness
# UNACLOUD OPPORTUNITIES AND LIMITATIONS

<table>
<thead>
<tr>
<th>REQUIRED FEATURES</th>
<th>UNACLOUD</th>
</tr>
</thead>
<tbody>
<tr>
<td>USABILITY</td>
<td>High usability Web user interfaces, which operation is almost intuitive, requiring basic IT knowledge</td>
</tr>
<tr>
<td>SELF-SERVICE</td>
<td>Unilaterally computing service provision</td>
</tr>
<tr>
<td>BROAD NETWORK ACCESS</td>
<td>Web portal available over Intranet and Internet</td>
</tr>
<tr>
<td>ON DEMAND SERVICES CUSTOMIZATION</td>
<td>On demand computing services customization, even to meet large scale computational requirements</td>
</tr>
<tr>
<td>HARDWARE MULTI TENANCY</td>
<td>Opportunistic use of idle computing resources</td>
</tr>
<tr>
<td>VIRTUALIZATION</td>
<td>On demand VM deployment through virtualization</td>
</tr>
<tr>
<td>SCALABILITY</td>
<td>Horizontal scaling model based on private clouds</td>
</tr>
<tr>
<td>INTEROPERABILITY AND LOOSE COUPLING</td>
<td>Web and service oriented architecture</td>
</tr>
<tr>
<td>EXTENSIBILITY</td>
<td>Use of open source tools, broadly diffused</td>
</tr>
<tr>
<td>DELEGATED ADMINISTRATION</td>
<td>Services to support common administration tasks</td>
</tr>
<tr>
<td>SECURITY</td>
<td>Authentication, authorization, confidentiality and non-repudiation mechanisms to secure the deployments</td>
</tr>
<tr>
<td>MEASURED SERVICE</td>
<td>Infrastructure traceability at user level with basic reports</td>
</tr>
<tr>
<td>QOS AND SLAs</td>
<td>Best effort approach due to the infrastructure availability</td>
</tr>
</tbody>
</table>
WORK IN PROGRESS

Efficiency
- VM images copy
- Enabling GPUs
- API AWS compatible
- Big Data on DCs

Energy
- VM allocation
- Reactive/Proactive VM movements

Slowdown
- CPU scheduling
- Reactive/Proactive VM movements

Reliability (QoS)
- Global snapshots
- Reactive/Proactive VM movements
Reduce interference on users’ tasks

- New class on Linux scheduler
- Avoid executing a task if it may affect the user above 15%
- Consider turbo boost, speed step, hyper-threading
**Linux scheduler:** Processor Specific Profile – Slowdown

Intel Core i7-4770 @ 3.40GHz

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0%</td>
<td>4%</td>
<td>9%</td>
<td>7%</td>
<td>2%</td>
<td>3%</td>
<td>5%</td>
<td>2%</td>
</tr>
<tr>
<td>2</td>
<td>3%</td>
<td>10%</td>
<td>12%</td>
<td>13%</td>
<td>4%</td>
<td>5%</td>
<td>4%</td>
<td>3%</td>
</tr>
<tr>
<td>3</td>
<td>4%</td>
<td>15%</td>
<td>16%</td>
<td>18%</td>
<td>8%</td>
<td>4%</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>4</td>
<td>14%</td>
<td>19%</td>
<td>14%</td>
<td>15%</td>
<td>6%</td>
<td>3%</td>
<td>4%</td>
<td>5%</td>
</tr>
<tr>
<td>5</td>
<td>16%</td>
<td>13%</td>
<td>16%</td>
<td>17%</td>
<td>5%</td>
<td>4%</td>
<td>4%</td>
<td>2%</td>
</tr>
<tr>
<td>6</td>
<td>15%</td>
<td>22%</td>
<td>13%</td>
<td>22%</td>
<td>7%</td>
<td>7%</td>
<td>5%</td>
<td>2%</td>
</tr>
<tr>
<td>7</td>
<td>16%</td>
<td>23%</td>
<td>20%</td>
<td>17%</td>
<td>11%</td>
<td>5%</td>
<td>5%</td>
<td>7%</td>
</tr>
<tr>
<td>8</td>
<td>24%</td>
<td>22%</td>
<td>19%</td>
<td>17%</td>
<td>14%</td>
<td>5%</td>
<td>5%</td>
<td>5%</td>
</tr>
</tbody>
</table>
Network elements used in the deployment of virtual machines
Improving efficiency

- To evaluate alternative protocols to TCP
- UDP, BitTorrent, proprietary: 100 images

AWS: 251s
Unacloud: 489s
Caldas I: 3402s
ISCloud: 2200s
INTEGRATING GPUs
• **Failure characterization**

- **Fallo ejecución VM**
  - $Q=1.11\times10^{-6}$
  - $\lambda=6.77\times10^{-2}$
  - $MTBF=21.1$
  - $MTTR=6.279$
  - $MTTF=14.76$

- **Fallo en Equipo**
  - $\lambda=1.17\times10^{-5}$
  - $MTTR=0.09514$

- **Fallo en el Hypervisor**
  - $\lambda=3.34\times10^{-4}$
  - $MTTR=1$

- **Fallo causado usuario**
  - $\lambda=5.88\times10^{-2}$
  - $MTTR=7$

- **La principal causa de no disponibilidad!!**

- **Apagado diario sala**
  - $\lambda=5.88\times10^{-2}$
  - $MTTR=7$

- **Apagado de usuario**
  - $\lambda=5.60\times10^{-3}$
  - $MTTR=1$
• Global snapshot

P1 is Initiator:
• Record local state S1,
• Send out markers
• Turn on recording on channels $C_{2i}$, $C_{3j}$
Questions ?!?