KVM on Grid, Shaken, Not Stirred
Cloud batch scheduler integration of KVM

Raphael Sack
Systems programmer, Intel Corporation
Co-author: Igor Ljubuncic

IT@Intel
This presentation is for informational purposes only. INTEL MAKES NO WARRANTIES, EXPRESS OR IMPLIED, IN THIS SUMMARY.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

For more complete information about performance and benchmark results, visit www.intel.com/benchmarks

Intel and the Intel logo are trademarks of Intel Corporation in the U.S. and/or other countries.

* Other names and brands may be claimed as the property of others.

Copyright © 2014, Intel Corporation. All rights reserved.
Agenda

• Introduction & background
• Overview
• Our solution
• Results & summary
• Future
Who am I?

- System Programmer, Engineering Computing
- Web and system related development
- Windows & Linux
- Car enthusiast
- Strong South African Accent :-)

PIC
2014 Intel IT Vital Statistics

>6,300 IT employees
59 global IT sites

>98,000 Intel employees\(^1\)
168 Intel sites in 65 Countries

64 Data Centers
(91 Data Centers in 2010)
80% of servers virtualized
(42% virtualized in 2010)

>147,000+ Devices
100% of laptops encrypted
100% of laptops with SSDs
>43,200 handheld devices
57 mobile applications developed

Source: Information provided by Intel IT as of Jan 2014
\(^1\)Total employee count does not include wholly owned subsidiaries that Intel IT does not directly support.
Overview of Intel chip design environment

- 40 sites globally
- 600,000 cores
- 5 PB memory
- 24 PB distributed storage
- Batch environment
- 40M+ regressions/month
So what is KVM on a Grid?
KVM on a Grid

- A comprehensive solution
- Enables rapid transformation of the operating system landscape
- Spawn any task/OS at a click of a button (or command line)
- On demand provisioning
- KVM based
- Seamless addition of virtual machines to the environment
Why do we need it?!
Operating systems

• New major release every couple of years
• Time required to verify new image
• Time required to adapt
• Time required to switch over
• Different types of customer → different time tables
• Support cycle vs. how we do things (no time to change)
What can we do about this?

• Stay put, avoid changes
• Accept no new hardware and technologies
• Leave no option for your users/customers
• Maintain several images & machines simultaneously
Ideal state

- No specific OS release dependency
- Infinite on demand capacity
- Be able to change operating systems immediately
- One (operating system) size fits all
- Go visit the Nürburgring - Nordschleife!
But seriously… we’re dealing with customers
Our solution

- Extend the batch cloud system to support virtualization
- Make virtualization as transparent as possible
- Allow users to choose the required OS or choose for them
- Both for interactive and batch/non-interactive usage
- Limit use of older OS
- Run Windows on Linux :O
- “One Cloud”
Architecture

1. User submits a task to the cloud

2. Cloud scheduler decides where to send the task

3. Considers physical or virtual machine need

4. Decides on where to land the task

5. Spawn VM

6. Task is inserted into the VM

7. User interacts transparently with the task
Architecture

Cloud agent

KVM Manager

Allocate capacity

Fetch template

Configure and spawn

Hypervisor + extra logic

Central template repository

Inject user’s task

VM
Use cases

- Software validation
- Legacy OS & tools
- Throughput vs. performance
- Urgent short-term capacity needs
- Isolation and containment of users/tasks
- Network related restrictions and limitations
Results

- Major chip design business groups adoption
- Up to 3K virtual machines concurrently spawned (on demand)
- Mixed Windows and Linux based tasks
- 5M+ tasks in the past several months
- Dynamic capacity allocation
STOP!
Number time!
## Results - Single task

<table>
<thead>
<tr>
<th></th>
<th>OS 1 Physical</th>
<th>OS 2 Physical</th>
<th>OS 1 VM</th>
<th>OS 2 VM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU*</td>
<td>1.00</td>
<td>0.80</td>
<td>1.04</td>
<td>0.81</td>
</tr>
<tr>
<td>IO</td>
<td>1.00</td>
<td>1.20</td>
<td>2.61</td>
<td>7.29</td>
</tr>
<tr>
<td>Memory Read</td>
<td>1.00</td>
<td>2.32</td>
<td>0.26</td>
<td>0.80</td>
</tr>
<tr>
<td>Memory Write</td>
<td>1.00</td>
<td>1.84</td>
<td>0.33</td>
<td>0.80</td>
</tr>
</tbody>
</table>

* Lower is better
## Results - Multiple tasks

<table>
<thead>
<tr>
<th></th>
<th>OS 2 Physical</th>
<th>OS 1 VM</th>
<th>OS 2 VM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU*</td>
<td>1.00</td>
<td>1.28</td>
<td>1.27</td>
</tr>
<tr>
<td>IO</td>
<td>1.00</td>
<td>12.9</td>
<td>21.5</td>
</tr>
<tr>
<td>Memory Read</td>
<td>1.00</td>
<td>0.19</td>
<td>0.28</td>
</tr>
<tr>
<td>Memory Write</td>
<td>1.00</td>
<td>0.14</td>
<td>0.36</td>
</tr>
</tbody>
</table>

* Lower is better
Results - Hybrid

<table>
<thead>
<tr>
<th>Test Case</th>
<th>OS 2 Physical</th>
<th>OS 2 VM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU*</td>
<td>1.00</td>
<td>1.01</td>
</tr>
<tr>
<td>IO</td>
<td>1.00</td>
<td>9.67</td>
</tr>
<tr>
<td>Memory Read</td>
<td>1.00</td>
<td>0.29</td>
</tr>
<tr>
<td>Memory Write</td>
<td>1.00</td>
<td>0.37</td>
</tr>
</tbody>
</table>

* Lower is better
Challenges

• Improving performance while minimizing cost
• Heterogeneous virtualization enclaves
• Increased complexity of implementation
• Disruptive changes to the environment
• Absolute necessity of transparency for users
• Support
Benefits

- Versatility
- Simpler and wider control of “old” images
- Significant cost avoidance
- Fully transparent
- Streamlined
- “Free”
Next steps

• Wider adoption by internal customers
• Performance optimizations and scale-out
• BI integration
• Live migration?
• Containers?
Summary

• Innovative use of existing KVM technologies
• Complex business case
• Challenging the status quo
• Increased flexibility in provisioning, deployment and OS migration
Questions?