Virtualization: The good, the bad and the ugly.

Michael Baum
thebaum@splunk.com
Splunk Inc.

Mark Cohen
Splunk Inc.

Mike Beck
Emerging Technologies Group
Sysadmin of the Year 2006
The good, the bad and the ugly.

The Good:

Virtualization can make one thing (a server) look like many things (different operating systems)

Virtualization can also make many things (storage arrays) look like one thing (a storage area network).

The goals are to
- simplify administration,
- increase utilization and
- improve scalability.

The Bad and the Ugly:

There are many different types of virtualization. Deciding which one is best can be difficult.

Virtualization creates more complicated resources.
- Resources can become over utilized and performance can vary dramatically.
- Resources are more prone to failure.
- Resources are harder to manage.
- Who will troubleshoot the virtualization when it breaks?
Server Virtualization

- Masks server resources (physical servers, processors, and operating systems).
- Users of server resources are spared the complexity of server management.
- Increases resource sharing and utilization.

Hosted versus Hypervisor Architectures

Source of diagrams: VMWare Inc.
Storage Virtualization

- Pools physical storage from multiple network storage devices into a single logical storage device.
- Managed from a central console.
- Storage virtualization is commonly used in storage area networks.
Network Virtualization

- Combines resources in a network by splitting bandwidth into independent channels, each of which can be assigned (or reassigned) to a particular server or device in real time.

- The idea is to disguise the true complexity of the network by separating it into manageable parts, much like your partitioned hard drive makes it easier to manage your files.
How we use virtualization at Splunk

- Development/QA/Customer Support across different OS and server architectures.

![Diagram showing various virtualization tools and hardware brands in use at Splunk](image-url)
The bad...

Scale creates problems. More virtual than physical machines means more complexity in the same size environment "which of my 50 web hosts is it?"

- **Monitoring.**
  - Virtual servers are constantly being created and destroyed.
  - Many tools that work with physical servers don’t work with virtual servers.
  - Physical server monitoring & application behavior discovery that rely on maps to trace business impact (e.g. can't connect errors) back to physical causes (e.g. flaky NIC) don’t keep up.

- **Patch management.**
  - Each virtual server needs to be patched even if they reside on same physical server.

- **Backup and restore.**
  - There are multiple methods exist for backing up virtual servers -- for example, as disk files or as physical machines, but each has significant limitations, particularly in retrieving files.
The ugly...

- **Troubleshooting.**
  - Finding logs, configurations, performance data and statistics can be tricky as virtual servers are constantly created and destroyed. Too many `/var/log` and too many `/etc`.

- **Corruption.**
  - Hardware support not yet robust enough to present a corrupt virtual machine from effecting others. AMD-V and Intel VT should help.

- **Workload management.**
  - Physical servers can quickly get overwhelmed. Little if any automation available.

- **Location persistence.**
  - Each virtual machine is in a different path on the SAN

- **Time clock drift.**
  - Servers tend to get more behind under heavy load. So local timestamps on different servers can't really be reliably correlated.
Quick tips and tricks.

› Centralized logging and configurations on the SAN, NAS or file system.

› Monitor for MAC / IP address conflict messages in syslog.

› Report on output of uptime command on hosted and hosting OS
  - looks like "16:45 up 1 day, 7 mins, 2 users, load averages: 0.19 0.23 0.15"
  - want to see high load averages on hosting OS (that's the point of virtualization!),
  - want to see low load averages on hosted OS (should notice / alert on trending upwards on a given virtual machine or outliers within a group of virtual machines dedicated to the same task).

› Audit re-provisioning of virtual machines to new physical hosts
  - do the logs on apps running on those machines show better/worse behavior?

› Observe changes in activity patterns on particular virtual machines.
  - trace back to re-provisioning actions that may have been the cause.