

Database Virtualization



David Fetter
Senior MTS, VMware Inc
PostgreSQL China 2011 Guangzhou

vmware®

Thanks!

- ❖ Jignesh Shah
- ❖ Staff Engineer, VMware
- ❖ Performance Expert
- ❖ Great Human Being

Content

- ❖ Virtualization
- ❖ Virtualized Resources
- ❖ VM Configuration
- ❖ Know Your * (Resources)
- ❖ Summary

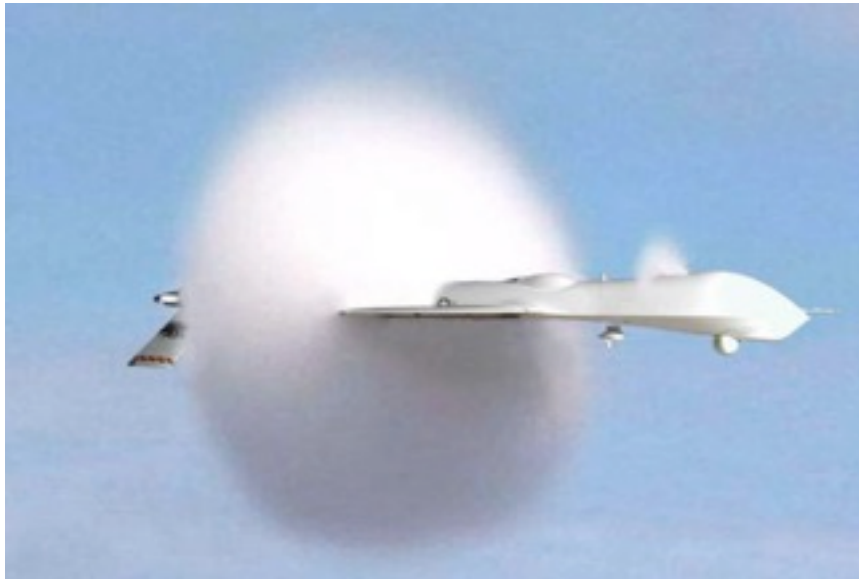
Virtualization

If running PostgreSQL on physical servers is like flying in open sky



Virtualization

Then running PostgreSQL on Virtualized Environments is like flying in clouds (Welcome to Cloud Computing)



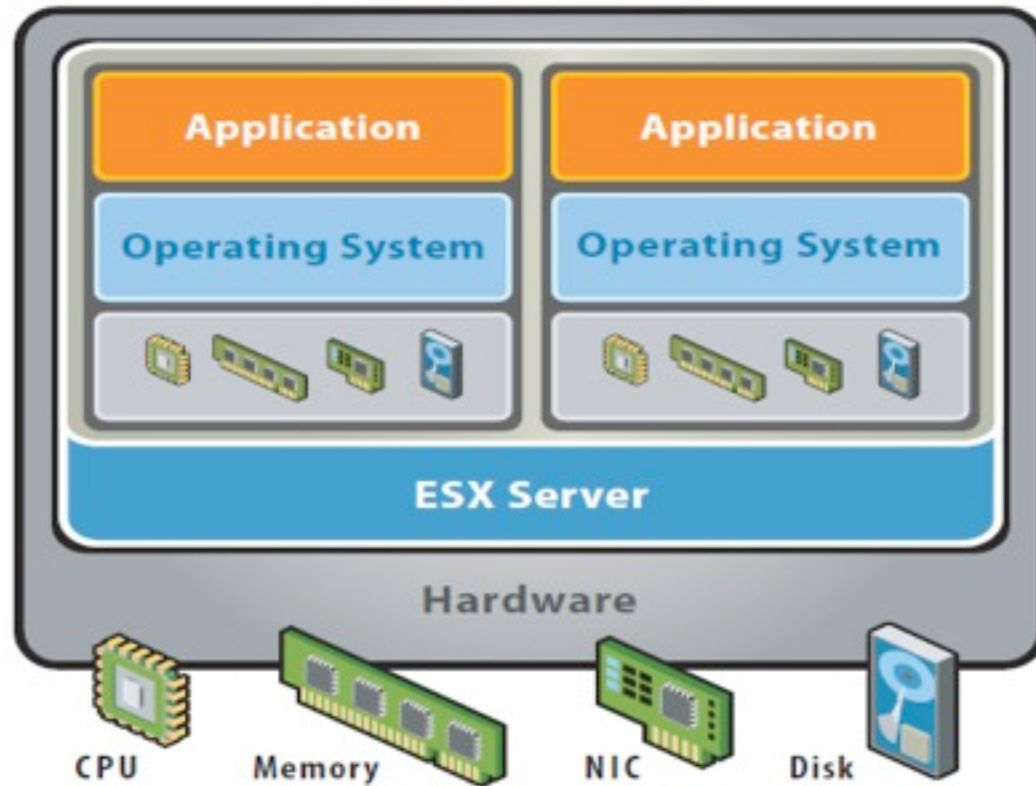
May the Force be with You! (At least GPS, Radars, etc)

What is Virtualization?

- ❖ Virtualization (according to various people)
 - Abstraction
 - Division
 - Emulation
 - Exaggeration

- ❖ Query Virtualization for information and
 - Best case: Refuses to answer
 - Worse case: Lies

Physical resources used by databases



VMware ESX Server virtualizes server storage and networking, allowing multiple applications to run in virtual machines on the same physical server.

Know Your Resources and its Administrators

- ❖ Begin to know your virtual resource layer along with physical resources
- ❖ For Physical Resources we talk with
 - System Administrator / Storage Administrator
- ❖ Now with Virtualization in the middle we also talk with
 - Virtualization Environment Administrator

Selecting VM Configuration for PostgreSQL DB Server

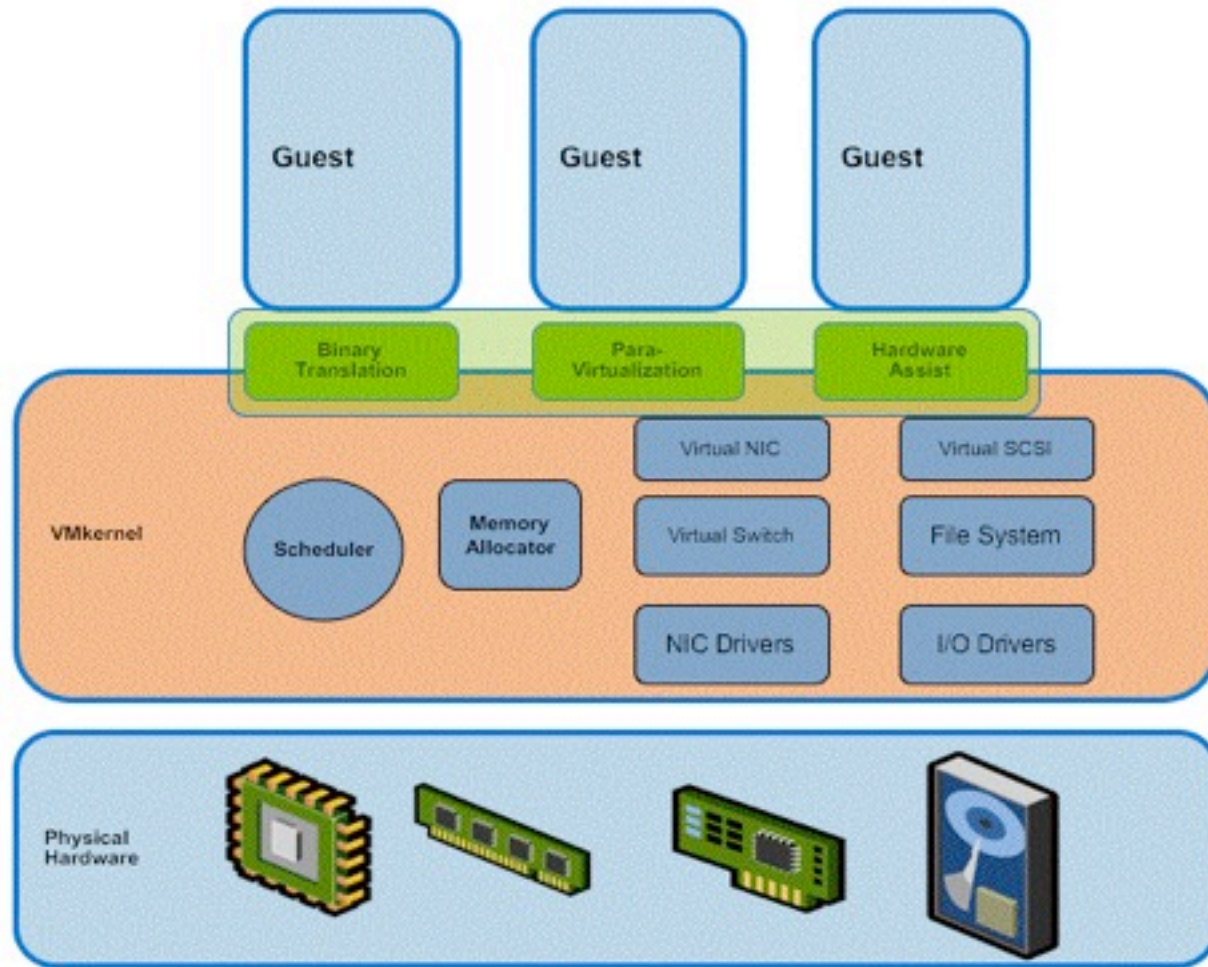
- ❖ Be good Citizens
 - Right-Size Configuration
- ❖ Try not to overcommit resources
 - Check the available Physical resources
- ❖ Separate PGDATA and WAL into separate Virtual disks
- ❖ Select Thick allocations for Virtual disks used (space fully allocated)

Selecting VM Configuration for PostgreSQL DB Server

- ❖ Use Para-Virtualized drivers/adapters wherever possible
 - For example on VMware ESX use
 - PVSCSI for SCSI disks controllers
 - VMXNET3 for Network interface controllers
- ❖ Use Hardware Assisted Virtualization technologies
 - Intel VT or AMD-V
 - Intel EPT or AMD RVI

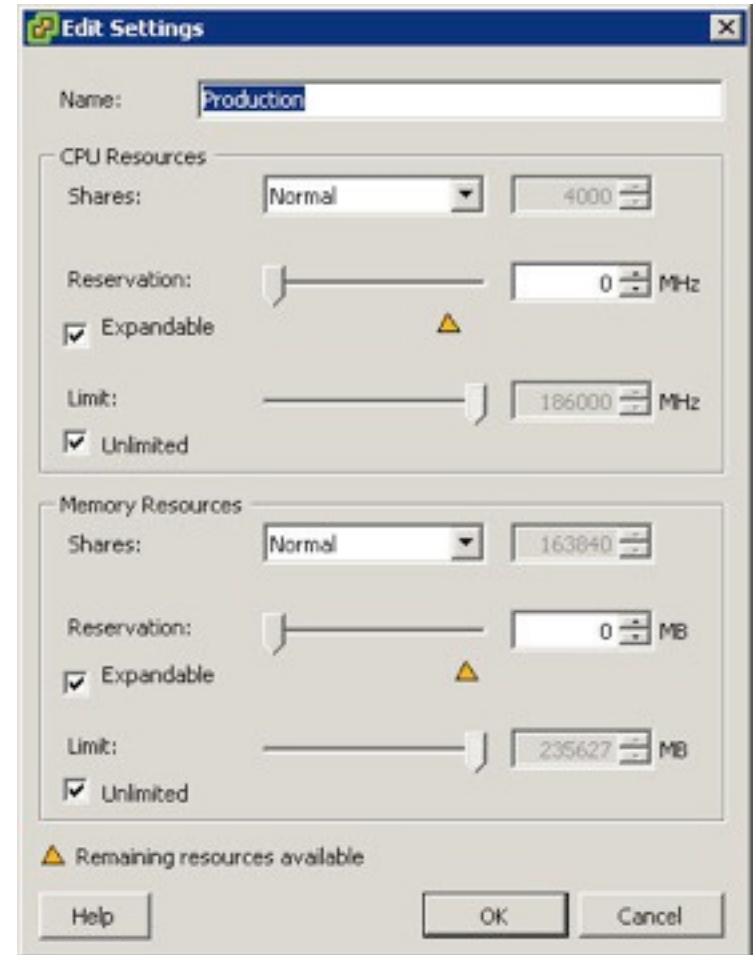
Using these PV adapters and hardware assistance features help to reduce the CPU utilization on the physical server

Selecting VM Configuration for PostgreSQL DB Server



Selecting VM Configuration for PostgreSQL DB Server

- ❖ To maintain minimum SLA use Reservations of resources
 - VMware ESX: CPU is reserved in terms of clock speed and Memory in terms of MB



Know Your WAL Sync

- ❖ Virtualization can lie about Postgres WAL sync
- ❖ For Data Integrity
 - Check with Vendor regarding your specific product.
 - WAL sync does go to the underlying disk on VMware ESX.
 - On Host OS based products it may not be true

- ❖ (People who use `fsync=off` can skip this slide)

Know Your Disks

- ❖ Thin provisioning is popular – Saves money, over provisioning
- ❖ A new write by VM on thin provisioned disks does
 - Block allocation (thick allocation) on underlying storage
 - Zero'ed out (cleaned and secured)
- ❖ So new writes can be expensive for PostgreSQL WAL segments
- ❖ Prepare the blocks in advance:
 - (On VMware ESX use EagerZeroedThick allocations using command line – not available directly via vSphere Client)
- ❖ Warm up PostgreSQL servers to create those extra WAL segments before you start measuring performance

Know Your IO

- ❖ Misaligned filesystems with underlying storage block size is a challenge to get it right
 - E.g. fdisk partitions may not align
 - Virtualization makes that more difficult since it adds another layer in between filesystem and physical storage
 - Easy way around is to use entire disks for partitions
- ❖ Monitor Storage IO latencies and keep track of it
- ❖ Use Paravirtualized adapters and drivers (e.g. VMware PVSCSI)

Know Your vMemory

- ❖ Each VM has an overhead memory on hypervisor
 - The bigger the configuration, the more the overhead increases
 - However, the relative % overhead decreases with increase in VM Memory size
- ❖ Virtual Environments often overcommit memory
 - VMware ESX uses Page Sharing, Ballooning , Swapping, Memory Compression
 - Impacts PostgreSQL bufferpool performance

Know Your vMemory

- ❖ NUMA Groups also affect Memory Latency.
 - VM does not necessary know about NUMA Groups and may not optimize for it (especially for large VMs)
 - VMware ESX does allocation based on NUMA groups for VMs
 - Impacts reads from bufferpool latency
- ❖ Memory can also be reserved by VM, (useful for minimum guarantees)

Know Your vCPU

- ❖ Guest OS in VM often tells you about physical CPU and clock speed
- ❖ Actual frequency may be different
 - Impacts SQL queries (CPU costs are higher)
- ❖ Add variable to your performance expectation equation
 - Hyperthreading (CPU costs are higher)
- ❖ CPU utilization reported by Guest OS does not include CPU overhead related to VM on physical hosts
 - Check Host CPU utilization for actual CPU usage by VM (e.g. VMware esxtop)
- ❖ Make sure Hardware Assisted CPU, MMU virtualization are being used

Know Your SMP vCPUs in VM

- ❖ Using SMP vCPU in a VM even when not needed/used causes overhead for hypervisor (specially for Legacy OS)
 - ❑ Has to handle timer interrupts
 - ❑ Maintain consistent memory view among multiple vCPU resources
 - ❑ Some Guest OS run idle threads on unused cpu power
 - ❑ Guest OS moves single threaded among the available vCPU thereby losing local cache locality
- ❖ VMware ESX generally uses paravirtualized tools to overcome these issues

Know Your Network

- ❖ Monitor network latencies
 - Virtual NICs typically connect to virtual switches which connect to physical NICs
 - Network Latencies between VMs go through virtual switches for VMs in same Physical machine
 - During Migrations, that can change and cause increased latencies from clients to machines
- ❖ Use para virtualized adapters and drivers (eg VMware VMXNET3)
- ❖ Historical data is now your friend to see if things changed

Other Workarounds in Virtualized Environments

- ❖ Time and Tide Waits for No VM
 - Historically VMs are suspended and resumed on the fly by just pausing vCPU causing Real Time and VM time to drift
 - Use either host based NTP or VM based NTP
- ❖ Here Today, Gone Tomorrow
 - Virtual Infrastructure Administrator (or automated policies) can move VM to a physical system which may have slightly different characteristics
 - (E.g. VMware DRS will move the VM with better resources)

PostgreSQL 9.0 Replication under Virtualization

- ❖ Easier to setup by cloning entire VM or clone PGDATA and WAL virtual disks
- ❖ Make sure replications are located on different physical hosts to avoid single point of failure
 - (Hint: Use VM resource policy settings to set them on different servers)
- ❖ Make sure the virtual disks from both VM are also on separate underlying storage system
- ❖ Use NTP to limit Time Drifts
- ❖ Easier to test on virtualization - Clone Standby server

Setup Guidelines Summary for Postgres

- ❖ Select VM resources to fit within NUMA Groups
- ❖ Right size vCPU (depending on hyperthreading)
- ❖ Separate PGDATA , WAL in separate file system even if they are on same physical infrastructure
- ❖ Use fully allocated Disks

Monitoring Guidelines for Postgres

- ❖ Make sure the VM resources are not heavily overcommitted
- ❖ Always check for swapping activities even when you don't expect it
 - Depending on your virtualization product, stolen memory may result in swapping
- ❖ Trust VM statistics: Verify with Host Statistics
 - (For VMWare ESX, esxtop is a great tool)

Questions / More Information

- ❖ Email: dfetter@vmware.com
- ❖ Learn more about VMware vSphere Hypervisor
 - <http://www.vmware.com/products/vsphere-hypervisor/>