SUNDE VDI Server-based Infrastructure Sizing

The number of users, intended application suite and overall performance expectation determine how powerful a host system must be in order to deliver the desired end-user experience. This section provides high-level guidelines for determining the system requirements for various numbers of Diana VDI users with various computing workloads. These guidelines can be used as a starting point for sizing your deployment – but our strongest recommendation is to thoroughly evaluate expected workloads for your Virtual Machines (VMs) and expect to make some adjustments and fine tuning as your Diana virtual desktop deployment progresses.

Understanding Virtual Desktop Workloads

Prior to deployment, you should develop use cases for the users you expect to connect to a host. An important consideration in developing use cases involves determining the number and types of applications users will need to operate. These requirements help you identify and measure the users' standard workloads. For example, you should measure the CPU, memory and storage utilization for a typical user workload in your environment. This workload data and the total expected number of users will help you determine the system requirements for your host system.

Light Workloads

- Task or knowledge workers running only 1 or 2 applications; i.e. a web browser or a billing application.
- Memory allocation of 768 MB (XP) to 1 GB (Windows 7) per VM .
- As many as 6 – 7 or more active VMs per core can be allocated.
- Storage system needs to provide roughly 30 IOPS per active user.

Medium Workloads

- Knowledge workers running multiple applications simultaneously, including Microsoft Office applications.
- Memory allocation of 1 GB (XP) to 1.25 GB (Windows 7) per VM.
- About 4 – 5 active VMs per core.
- Provide around 40 IOPS per active user.

Heavy Workloads

- Power users using scientific applications, high end graphics or software development.
- Memory allocation of 2 GB or more per VM.
- Only 3 – 4 active VMs per core.
- As much as 50 IOPS per active user.
- For best performance, reserve as many resources as needed

– create reservations in hypervisor and/or provide dual virtual CPUs per VM, even if more VMs than cores on server.

Configuring the server and storage infrastructure

CPU load of a VM is highly dependent on the workload-- VMs per Core: 3 – 7
- Typical VM load utilizes up to 15-25% of physical CPU, equal to 4-6 VMs per core
- For medium workloads, allocating 4-5 VMs per core is recommended
- 6-7 VMs may be allocated per core for lighter workloads

Disk performance is key to VM responsiveness-- VMs per Disk (15K): 4 – 6
- Both sufficient IOPS (30-50 per VM) and low latency (average read and write latency under 20 ms) required for optimal performance
- Enterprise-level 15K SAS RAID supports 4-6 VMs/drive excluding parity drives
- Lower performance disks may be used with fewer or less demanding VMs
- Either direct attached storage (DAS), or iSCSI/FC SANs may be used
- Large on-controller caches contribute greatly to storage performance
- Test SAN accordingly, actual results vary depending on VM workloads

Adequate **VM memory** ensures a positive individual user experience-- VM physical memory: 768 MB – 2 GB
- Depending on workload and VM OS, 768 MB – 2GB should be allocated
- Under-allocation of VM memory can result in Windows paging and over-burden the disk subsystem
- Although not desirable, the total physical memory on server can be over-allocated
- the hypervisor will make intelligent paging decisions for VM's
- Typically, 70-80% of VM memory will be in physical memory, the rest will be in the hypervisor swap.

VM Best Practices

Allocate plenty of memory and the hypervisor will swap out any unused DVM memory.
The hypervisor makes intelligent choices because it sees memory utilization across the entire server Try to keep VMs from paging.
Be aware of disk intensive applications – see notes under Heavy Users, above.

Storage Best Practices

Low latency is a priority - keep access latency low by over provisioning your SAN interconnect and storage infrastructure below load maximums.
Keep RAID groups between 6 – 12 drives, do not exceed 32 VMs per LUN At or near load, SAN access latency will degrade. Disk sub-system performance is the single largest factor in VM success.

Network Best Practices

Switched LAN environments are best for deploying virtual desktops and delivering the *full* desktop experience Make every effort to reduce packet loss, latency (under 10 ms round-trip), and jitter. If slow links are causing bottlenecks, use QoS. Desktops are an interactive service – only prioritize VoIP traffic above desktops. The protocol will adapt to match your network conditions.

Do I need a SAN?

The decision to use direct attached storage (local) vs. SAN (shared) storage is largely dependent on the availability requirements and the type of VM collections being used. Any VM stored on local storage is at risk of not being available if the server is unavailable due to a failure or maintenance. If this risk is acceptable, or there are alternate equivalent VMs on the other servers, then local storage can be used to reduce costs. If VMs are unique for each user and availability is a requirement, such as in the case of VMs that have been permanently assigned to users, then shared storage should be used despite higher costs.

Pooled Desktop – Local Storage

Local storage is used to reduce storage costs. Users are assigned the 1st available VM out of the pool at login time. After log out, the VM is returned to the pool. If the server goes down, a DVM from the same pool on another server can be served instead. Any active sessions when the interrupt occurs will be terminated.

Permanently Assigned – Local Storage

Local storage is used to reduce storage costs. Some availability risk. Users are tied to a specific VM. If a server becomes unavailable, the corresponding users will lose their sessions. The users will have to wait for the administrator to assign alternate desktops or wait until the server is back online.

Permanently Assigned – Shared Storage

Leverages a SAN or NAS for shared storage. Maintains high availability for users with permanently assigned desktops but at a higher cost. Users are tied to a specific DVM. If a server becomes unavailable, the corresponding VMs will be migrated to other servers in the cluster either manually or automatically.

Example of Server Sizing for Medium Workload Users

10 Users：

I7-3.4G CPU; 16 GB physical memory; 500GB 10K Hard Drive

20 Users：

I7-3.4G CPU; 32 GB physical memory; 1TGB 12K Hard Drive

30 Users：

Xeon-5600 Dual Core CPU; 48 GB physical memory; 2* 15K Drives in RAID

50 Users：

Xeon-5600 Quad Core CPU; 96 GB physical memory; 3* 15K Drives in RAID