# Virtualizing Agilent OpenLAB CDS EZChrom Edition with VMware

## Technical Overview


### Abstract

This technical overview describes the considerations, recommended configurations, and host server requirements when virtualizing an OpenLAB CDS EZChrom Edition system. It also describes performance testing which demonstrates the CPU, memory, and LAN communication impacts of virtualizing CDS components. Concluding recommendations are based on the performance data gathered using a Standard Configuration. This document assumes existing knowledge and expertise with virtualization technology.



Agilent Technologies

## Introduction

Virtualization can be an effective way to reduce IT expenses and boost laboratory efficiency. Virtualization software such as VMware creates and manages virtual machines (VMs), which run simultaneously on a single host server. Using virtualization in your OpenLAB system architecture offers several advantages:

- Time and cost associated with deploying and managing multiple physical machines is reduced because multiple machines with different operating systems and applications can run on a single virtualization host server.
- Flexible computing and application resource provisioning can increase server usage and productivity, such that more VMs can run on the same amount of hardware resources.
- Virtualization software, such as VMware, can provide an additional layer of security above and beyond that which is available in the OpenLAB software suite with permissions to control access to any VM.
- Depending on the configuration used in the laboratory, virtualization can free valuable bench space for other uses.
- Depending on the organization's operational qualification (OQ) requirements, for example if only one of many identical virtualized CDS clients must be qualified, virtualization can reduce OQ costs.

## Agilent OpenLAB CDS

OpenLAB Chromatography Data System (CDS) software is scalable from an individual workstation to a fully distributed system. This scalability provides an option to have all key functional elements of a CDS; administration, data storage, and instrument access either locally or centrally over the network.

Agilent OpenLAB CDS is offered in two versions: OpenLAB CDS ChemStation and EZChrom Editions. Most laboratories choose to standardize on one or the other and their architectural differences lead to differences in the recommended virtualization configurations.

This technical overview focuses on virtualizing OpenLAB CDS EZChrom Edition. Information about virtualizing OpenLAB CDS ChemStation Edition can be found in another Agilent publication.

Two basic configurations of OpenLAB CDS EZChrom Edition will provide the best starting point for understanding how the system components can be run on virtual machines, and examine the associated cost-performance tradeoffs.

## Standard Configuration 1 OpenLAB CDS EZChrom Edition with Agilent Instrument Controllers (AICs)

A key component of this configuration (Figure 1) is the Agilent Instrument Controller (AIC). AICs perform instrument connection and control, data acquisition, and data caching. The CDS client, the user interface for the scientist, displays sample run time information and performs data processing and reprocessing tasks.

The functions the AICs and CDS clients perform are the most important considerations when virtualizing a distributed system.

## OpenLAB CDS EZChrom Edition Workstations

OpenLAB CDS EZChrom workstations can be configured independently on separate PC's, or they can be configured to share a central storage device or content management system (Figure 2). In the Workstation configuration, instruments are connected directly to the PC's and all processing is done on the same PC. If a central data storage location is configured, appropriate network connectivity and bandwidth needs to be available as well. This configuration does not include AICs.

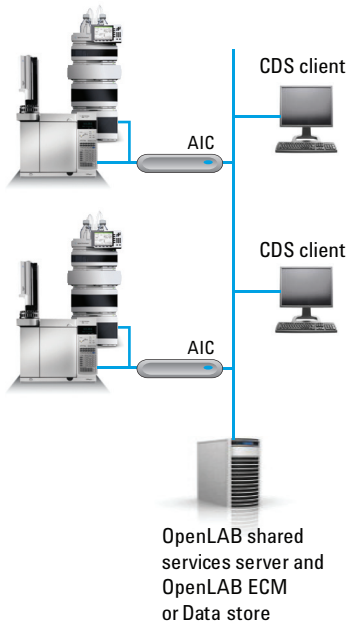


Figure 1. Distributed System with central data storage.

Figure 2. Independent Workstations.

## Virtualizing an OpenLAB CDS EZChrom Edition system with AICs

Virtualizing the CDS clients is the recommended way to use virtualization technology with OpenLAB CDS EZChrom Edition (Figure 3). The CDS client performs data reprocessing, and the physical AIC performs instrument connectivity, control, data acquisition, and data caching. ESX virtual host server resources are often idle waiting for human input on the virtual client PCs. By placing the data reprocessing load on the ESX host, these resources can be shared with other virtual machines, making the most efficient use of resources that are otherwise idle.

When virtualizing CDS clients, the key considerations are the number of clients that will be virtualized, the amount of data processing they will perform, and whether the reprocessing would be done concurrently. Figures 6A through 6D show that considerable CPU, memory, and disk resources are used for the reprocessing functions. Another consideration is how often the CDS clients will be used to view data in real time as the LAN cards on the host server will need to accommodate the real time data communication demand.

Though not usually recommended, virtualizing AIC's is also supported. Because they perform instrument connectivity, control, data acquisition, and data caching, each virtual AIC places a substantial load on the CPU, memory, disk drive, and host server LAN. They do not stand idle waiting for human interaction. For this reason, virtualizing AICs requires careful planning. The impact of virtualizing AICs on server memory, CPU, disk, and LAN use on the test system described in Figure 4 is presented in Figures 5A through 5D.

If the AICs do not regularly run instruments concurrently, it could be desirable to virtualize one or more AICs to take advantage of VMware's resource sharing capability. Be aware that virtualized AICs are connected to instruments over the network and are not able to offer data protection in the event of a network outage, negating one of their fundamental values. If the network goes down, any data collection the virtualized AIC is performing will cease. This risk can be mitigated to some degree by configuring multiple networks and multiple network connections for each virtual machine. Furthermore, the network proximity between the AIC and instruments it controls can have a direct and significant impact on performance. The greater the distance between the AIC and the instruments it controls, the slower everything will be.

When using virtual clients or AICs, connectivity is limited to instruments that support LAN connectivity. The use of virtual machines does not prevent using physical machines at the same time, especially where direct connections to instruments are either desired or required (for example, GPIB connections or critical data protection). Shifting some of the processing load to physical machines can also be advantageous if the host server is not powerful enough to run all the AICs and CDS clients.

Depending on laboratory use patterns, virtualizing an OpenLAB CDS EZChrom Edition system can offer cost savings regardless of which components are virtualized. However, the cost of a single powerful host server that can run one or more clients and/or AICs must be considered in the total cost of ownership.

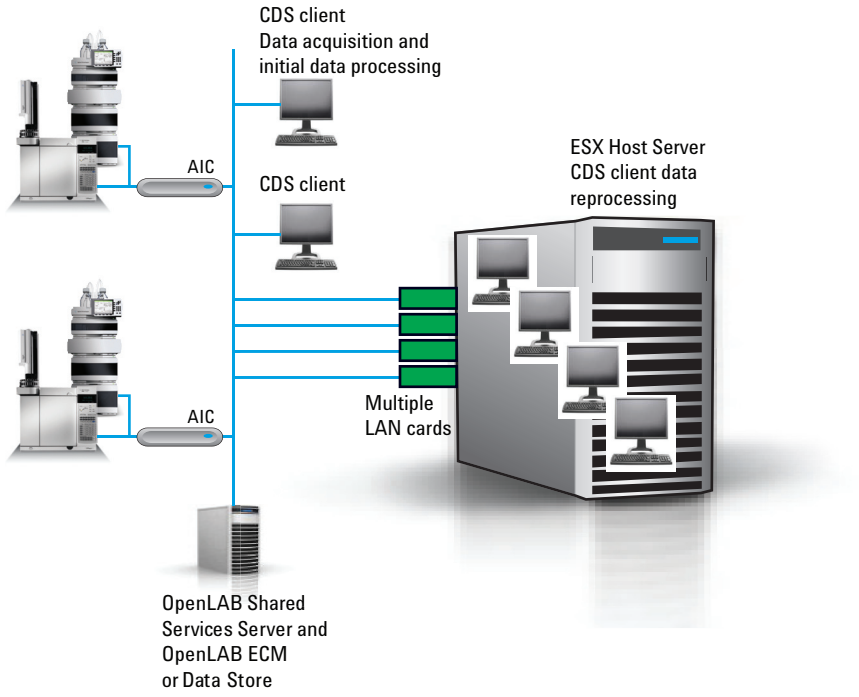


Figure 3. Recommended configuration for a virtualized OpenLAB CDS EZChrom Edition with AICs.

## Considerations when virtualizing an OpenLAB CDS EZChrom Edition distributed system

- The number of CDS clients and/or AICs that will be run on the host server. This determines the server processor, memory, and disk requirements. It also determines the amount of disk redundancy, and reserved CPU and memory needed.
- The laboratory use patterns of the virtual CDS clients and the AICs and their instruments.
- The number of instruments using the same LAN. This establishes LAN requirements.
- How often the virtual CDS clients will be used to view data in real time. This determines the LAN and host server LAN card requirements.

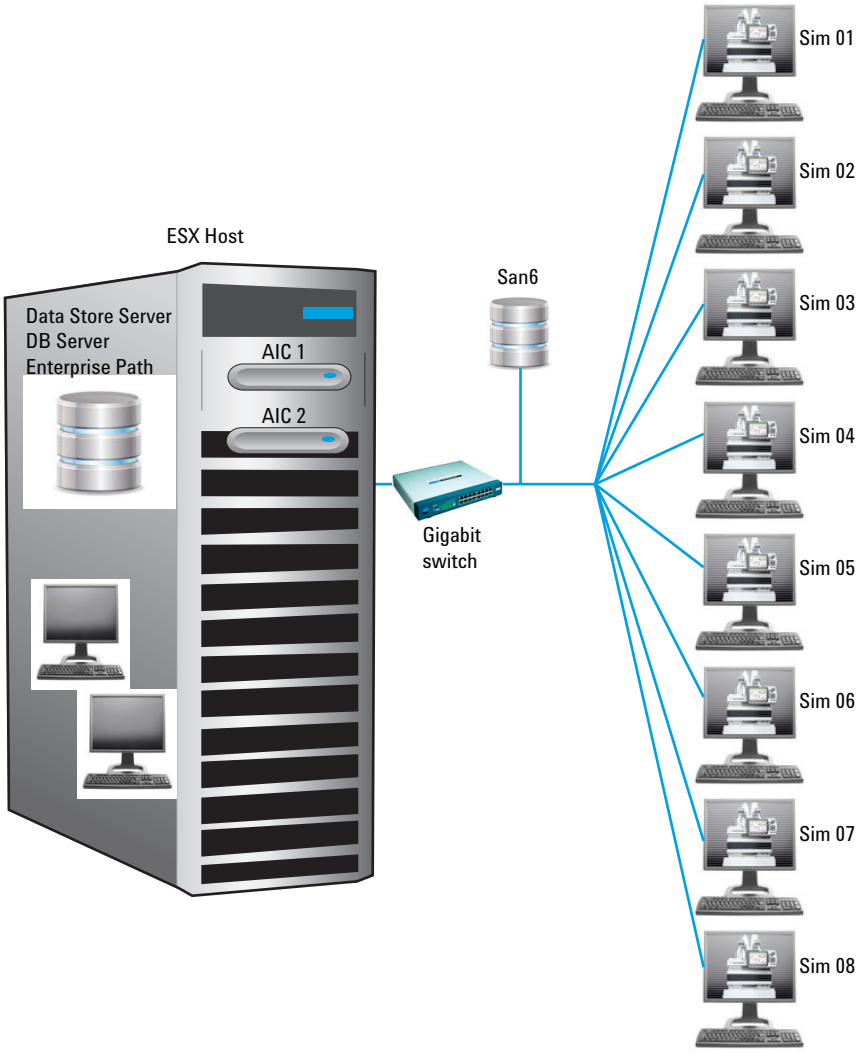


Figure 4. OpenLAB CDS EZChrom Edition system topology used to evaluate the impact on the host server and LAN when AIC and CDS clients are virtualized.

## Performance testing

To demonstrate the impact of virtualizing AICs and clients in OpenLAB CDS EZChrom Edition with AICs, the topology shown in Figure 4 was tested. Tables 1 and 2 show the specifications for this topology, including reserved CPU and memory. For communication between virtual machines, the host server was configured to use an internal virtual network.

The test system consisted of two virtualized AICs, each of which controlled four simulated Agilent liquid chromatography (LC) instruments with diode array detection (DAD), and two virtualized clients that performed data reprocessing. The simulated DADs were run at their maximum data rate. Host server CPU, memory, disk, and LAN use were recorded as the two AICs were ramped from running one to four simulated instruments each at 20 minute intervals. Sequences sufficient to track the system over a 12 hour period were submitted. Once data was available, data reprocessing tasks were launched on the CDS clients.

### Performance results

Figures 5A through 5D show the host server resource use for a single AIC running four instruments. Key observations are:

- Figure 5A shows 100% of the allocated CPU was used by the AIC. This indicates that when AICs are virtualized, the server will need to have enough processing power to accommodate the load without resource sharing.

Table 1. Virtual AIC and CDS client specifications for the system topology evaluated.

| Virtual machine | Processor | Memory | Network Interface Card (NIC) | Operating system | System drive | Data drive |
|---|---|---|---|---|---|---|
| Data store server* | Intel Xeon CPU X5690@3.47GHZ, 3.47GHZ (four processors) | 16 GB | Intel Pro/1000 MT network connection | Windows 2008 R2 SP1 | 120 GB | 256 GB |
| AIC 1 | Intel Xeon CPU X5690@3.47GHZ, 3.47GHZ (two processors) | 4 GB reserved | Intel Pro/1000 MT network connection | Windows 7 SP1 (64-bit) | 80 GB Allocated | |
| AIC 2 | Intel Xeon CPU X5690@3.47GHZ, 3.47GHZ (two processors) | 4 GB reserved | Intel Pro/1000 MT network connection | Windows 7 SP1 (32-bit) | 80 GB Allocated | |
| Client 1 | Intel Xeon CPU X5690@3.47GHZ, 3.47GHZ (two processors) | 4 GB reserved | Intel Pro/1000 MT network connection | Windows 7 SP1 (64-bit) | 80 GB Allocated | |
| Client 2 | Intel Xeon CPU X5690@3.47GHZ, 3.47GHZ (two processors) | 4 GB reserved | Intel Pro/1000 MT network connection | Windows 7 SP1 (32-bit) | 80 GB Allocated | |

* Explain absence of Data Store system

Table 2. VMware ESX host server specifications.

| | |
|---|---|
| VMware version and build | VMware 4.0 |
| Server manufacturer | HP |
| Server model | Proliant DL380 G7 |
| Number of CPU cores | 2 × 6 |
| CPU speed | 3.465 GHz |
| Number of NICs | 4 |
| Total physical memory | 703 16 MB |
| Hard disk speed(s) | HP iSCSI (SAN6) |
| Number of virtual machines running during testing (including data store) | 5 |

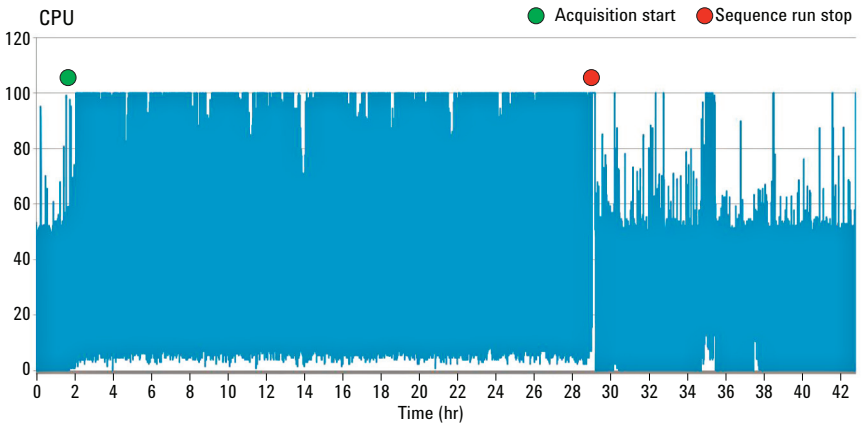


Figure 5A. CPU usage by the AIC. 100% of the allocated CPU resources were used by the virtualized AIC during data acquisition.

- Figure 5B shows approximately 50% of the memory available to the AIC was used. It is also important to note that the memory usage did not drop after the sequence was stopped. This indicates that when AICs are virtualized, significant server memory sharing should not be expected.
- Figure 5C shows the disk access (I/O) on the host server's was constantly high as data was cached to disk. Page swapping of memory to disk also contributes to the high disk I/O. To prevent performance problems resulting from all virtual machines trying to use the same disk hardware, the host server should be configured with multiple disk drives, each of which should be assigned to a different virtual machine.
- Figure 5D shows the LAN usage was constantly high as data was transferred from the instrument to the AIC over the network. LAN traffic can be a key bottleneck if all the traffic between instruments and AICs use the same LAN card.

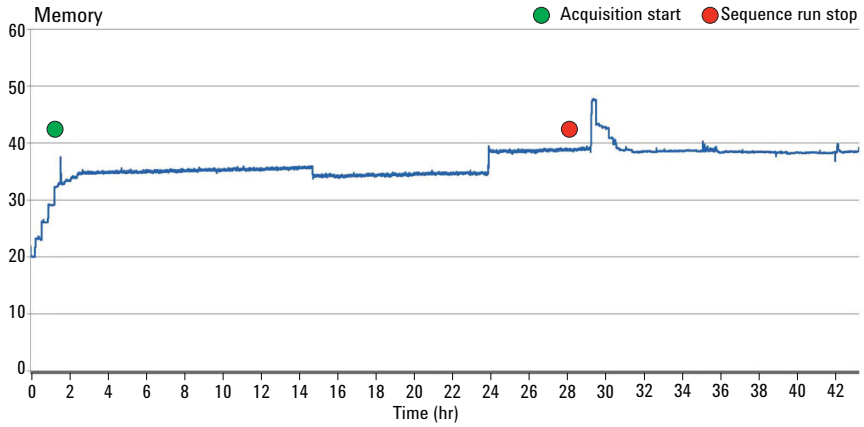


Figure 5B. Memory used by an AIC. Reserved memory is not released.

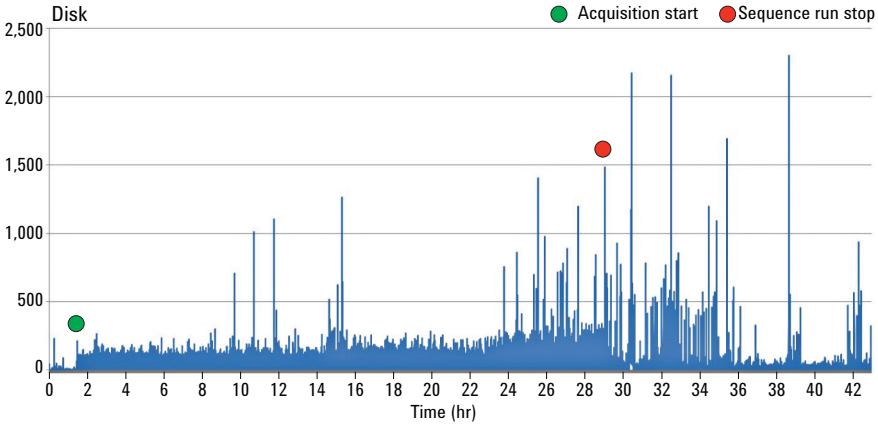


Figure 5C. Disk access (I/O). Disk access has a constant baseline but experiences spikes.

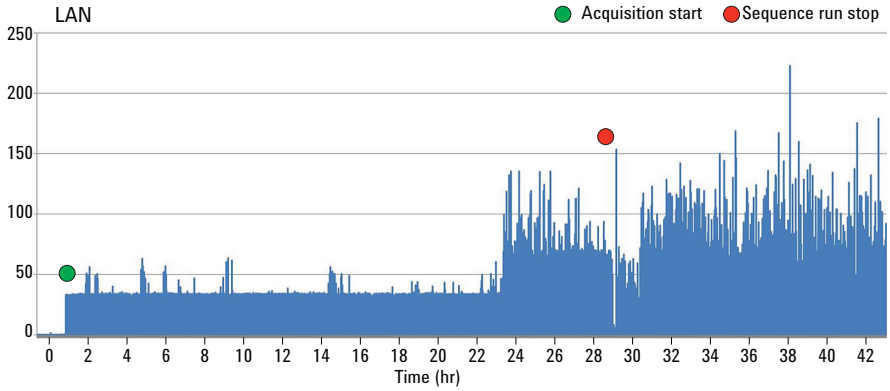


Figure 5D. LAN use by an AIC. LAN traffic can be a key bottleneck.

Figures 6A through 6D summarize host server resources used by one of the virtualized CDS clients. The CDS client was used to start the instruments on the AIC described in the previous section, and then performed three data reprocessing tasks. Key observations are:

- Figure 6A shows the CDS client used a high percentage of available CPU resources when performing reprocessing tasks. The amount of resource sharing that can be accomplished using the ESX server depends upon the amount of reprocessing that will be performed on each CDS client.
- Figure 6B shows the client did not use all available memory during reprocessing. There is more opportunity to share memory resources when running CDS clients on the ESX host server.
- Figure 6C shows that disk I/O on the CDS client has a constant baseline, with some significant spikes. To prevent all virtual machines from trying to use the same disk drive, configure the host server with multiple disk drives, each of which are assigned to a different virtual machine.
- Figure 6D shows the LAN usage was constant, but low. The client machines are in constant communication with OpenLAB Shared Services and database, but this traffic is not expected to significantly impact performance. Note that data was not displayed nor viewed in real time during this test. When used, the real-time display feature significantly increases the LAN traffic to the CDS clients.

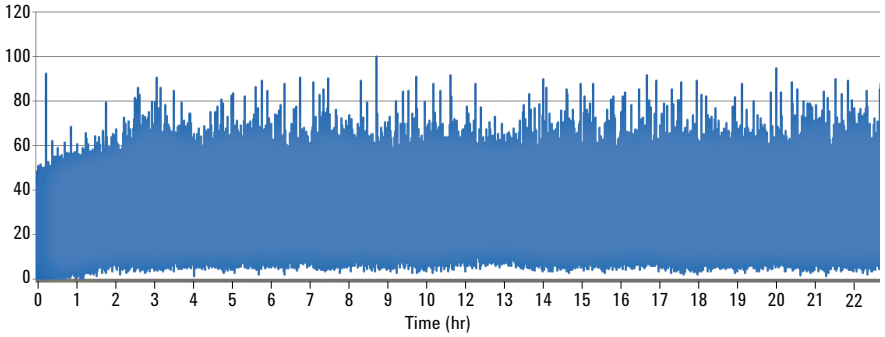


Figure 6A. The CPU usage by the CDS client was high and constant.

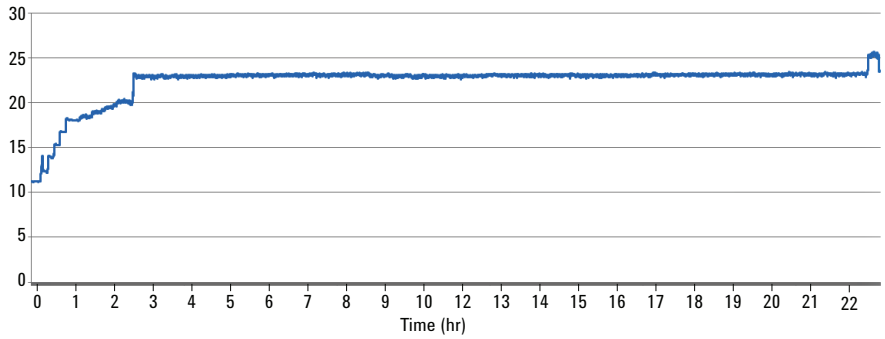


Figure 6B. Memory usage on the CDS client was constant but fairly low.

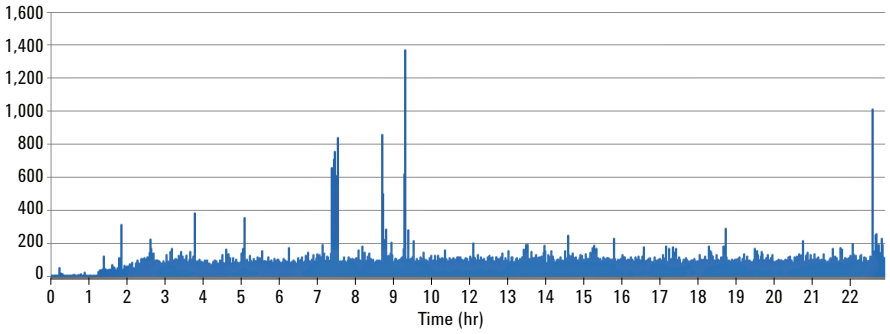


Figure 6C. Disk activity over the experimental timeframe.

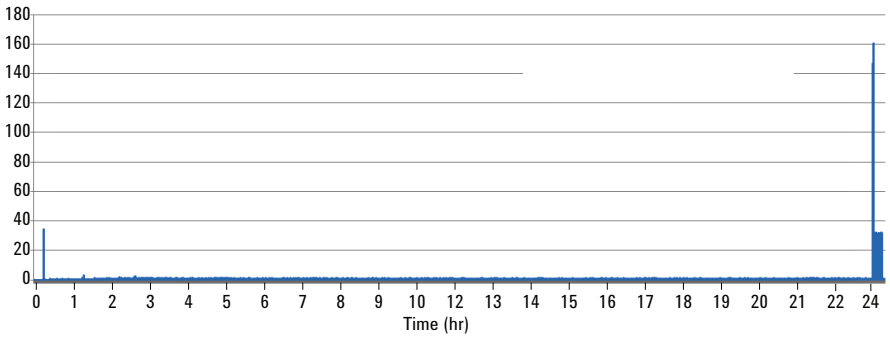


Figure 6D. LAN use

### Implications

The test results indicate that unless virtual AIC and CDS client use is very low, the ESX host server resources will become overloaded very quickly. This is true even when host CPU and memory are reserved. Because of the high level of disk and LAN use, the host should be configured with:

- Disk redundancy because sharing of disk resources between virtual machines should be minimized
- A virtual LAN for any traffic between virtual machines
- Multiple LAN cards for communication with nonvirtual machines and instruments

## Virtualizing OpenLAB CDS EZChrom Edition Workstations

In a Workstation configuration, instrument connectivity and control, data acquisition, data processing and reprocessing for each instance of the CDS happens on the Workstation. If the Workstation is a virtual machine, all of this activity is running on the ESX host server (Figure 7). It is possible to configure the Workstations to use central data storage location. The LAN communication configuration needs to accommodate the extra network traffic.

Beyond its conceptual simplicity, the ability to access and control the CDS client and associated instruments through the VMware client software can be a benefit of virtualizing OpenLAB CDS EZChrom Edition Workstations.

Though collected on a distributed system topology, the performance results shown in Figures 5 and 6 are directly applicable to a virtualized Workstation. Because the CDS clients perform all of the CPU, memory, disk and LAN intensive CDS functions, (data acquisition, data processing and data reprocessing), a server hosting multiple CDS clients will be overloaded. Because of the high level of disk and LAN use, the host should be configured with:

- Disk redundancy because sharing of disk resources between virtual machines should be minimized
- A virtual LAN for any traffic between virtual machines
- Multiple LAN cards for communication with nonvirtual machines and instruments

Virtualization does not prevent using physical machines, especially where direct connections to instruments are either desired or required. Shifting some of the processing load to physical machines is recommended if the host server is not powerful enough to run everything as a virtual machine. Reprocessing data can be very CPU intensive.

Depending on the use patterns of the virtual CDS clients and their associated instruments, virtualizing OpenLAB CDS EZChrom Edition Workstations could yield cost savings. Running the CDS software on physical machines requires reasonably powerful computers for each. If use is low, the laboratory implementing virtualization could purchase a single powerful server rather than several dedicated systems.

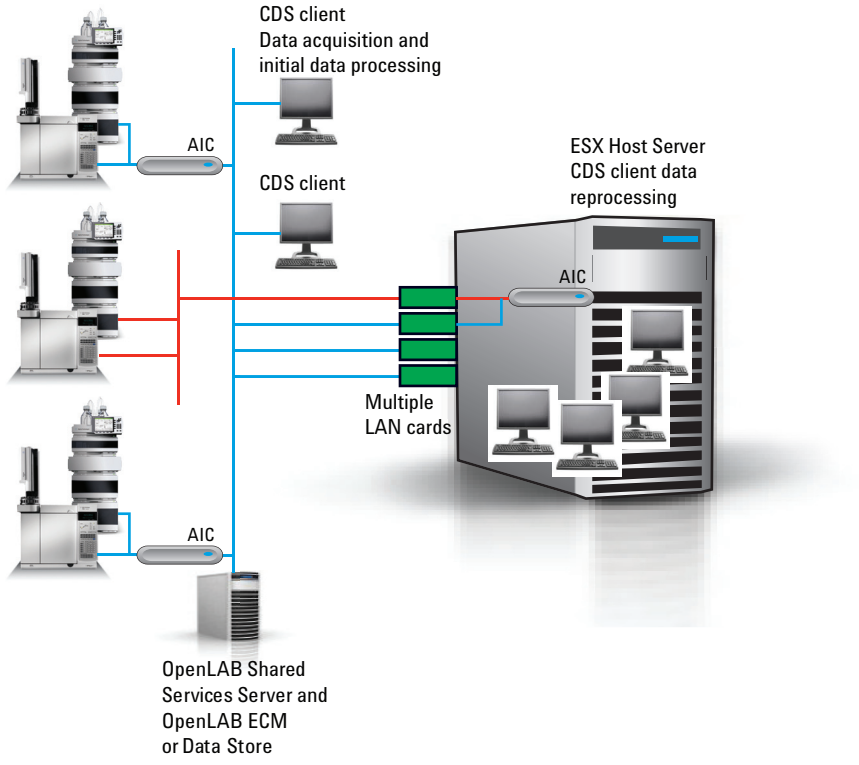


Figure 7. Virtualized OpenLAB CDS EZChrom Edition Workstations.

## Considerations when virtualizing OpenLAB CDS EZChrom Edition Workstations:

- The number of CDS clients that will be run on the host server. This determines the server processor, memory, and disk requirements. It also determines the amount of disk redundancy, reserved CPU, and reserved memory needed.
- The laboratory use patterns of the virtual CDS clients and their instruments
- How often the virtual CDS clients will be used to view data in real time. This determines the LAN and host server LAN card requirements.
- What kind of general LAN traffic and instrument LAN traffic will be happening on virtual machines. There are ways to configure multiple networks and multiple LAN cards for virtual machines in order to separate instrument data traffic from general network traffic.

## Conclusions and recommendations

Virtualization can reduce IT expenses and boost laboratory efficiency. With an OpenLAB CDS EZChrom Edition, it is possible to virtualize the CDS clients, the AICs, or a combination of both. Because of the large load an AIC places on the host server, virtualizing the CDS clients is the most desirable way to virtualize a distributed system. In this configuration, idle host server computing resources waiting for user requests can be shared with other virtual CDS clients. While an AIC can run as a virtual machine, the benefits need to be weighed against the loss of the data protection value provided by a nonvirtual AIC and the processing demands on a server that is hosting virtual machines.

OpenLAB CDS EZChrom Edition Workstations can also be virtualized. Because the CDS clients perform all of the CPU, memory, disk and LAN intensive CDS functions, (data acquisition, data processing and data reprocessing), a server hosting multiple CDS clients will be overloaded unless it is configured appropriately

Performance testing of a virtual topology showed that unless virtual AIC or CDS client use is low, the host server can become overloaded very quickly. As such, regardless of the components virtualized, the following is recommended when configuring a virtual OpenLAB EZChrom Edition system.

- Reserve host server CPU for both virtualized AICs and CDS clients.
- Reserve 50% of the host server memory for virtualized components.
- Perform frequent disk backups, heavy disk I/O puts disk drives at risk.
- Use multiple disk hardware and configure virtual machines to use different storage devices.
- Use virtual routers for communication between virtual machines.
- Configure multiple LAN cards and, if possible, separate networks for instrument connections and data traffic.
- Evaluate risk of controlling instruments over the network.
- Be aware of Network Proximity to instruments.
- Regularly monitor the ESX host server.

www.agilent.com/chem


Published in the USA, April 17, 2013
5991-2278EN

Agilent Technologies