Virtualization of the Enterprise GIS Environment

Qi Dai
Technical Lead
GIS Operation and Support
EPA National Computer Center
CSC Team for the ITS-EPA Contract
Research Triangle Park, North Carolina

Tim Richards
Geospatial Service Manager
National Computer Center
Office of Technology Operation and Planning
EPA Office of Environmental Information
Research Triangle Park, North Carolina

Presentation Outlines

- Virtualization and the EPA National Computer Center (NCC)

- The Enterprise GIS Environment at the EPA NCC

- Building an enterprise GIS environment with the virtualization technology at the EPA NCC

- Performances of the virtualized enterprise GIS environment in the EPA NCC

- Integrating the server virtualization and the desktop virtualization with the enterprise GIS environment in the EPA NCC

- Future Plan to improve the virtualized enterprise GIS environment at the EPA NCC

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

2 of 26

Virtualization and the EPA National Computer Center (NCC)

- EPA NCC virtualization history
  - Old virtualization technology: IBM Mainframe
  - New virtualization technology

- EPA NCC started to evaluate the virtualization technology four years ago.

- EPA NCC started to test and adopt the virtualization technology two years ago

- Virtualization Readiness Assessment Study for the EPA NCC

- Comprehensive Approaches for the virtualization process at the EPA NCC

- EPA agency-wide three-year server virtualization efforts

- Initiatives for green data center and green computing

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

3 of 26

Virtualization Considerations at the EPA NCC

- Enterprise Architecture Consideration

- Enterprise Network Considerations

- Storage Considerations

- Preparations for the Data Center Facility
  - Power supply impacts
  - Cooling system impacts

- Prepare for Unanticipated Consequences of Virtualization

- Other Consideration Factors

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

4 of 26

Identify Systems (x86) That Are Not Good Candidates For Virtualization

- Workloads with memory utilization greater than 16 GB
  - A more practical limit is 8 GB to enable more flexibility in supporting multiple VM's per ESX Server

- Computing power demand is greater than 3500MHz
  - Example; ((2200[MHz]*2[CPUs])*80[% utilization]=3520 MHz

- Network usage greater than 240 Mb/sec

- Physical disk I/O greater than 5000 Transfers/Sec or 50MB/Sec consistently

Source: EPA Virtualization Readiness Assessment Report by Dell Services

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

5 of 26

- Server Virtualization for the shared EPA NCC Intranet Staging Environment

- Server Virtualization for the shared EPA NCC Intranet Production Environment

- Server Virtualization for the shared EPA NCC Internet (DMZ) Production Environment

- Server Virtualization for the dedicated systems in the EPA NCC

- Server Virtualization for different types of servers including application servers and database servers

- Different types of virtualization processes:
  - IBM p570 Servers with the IBM System p virtualization technology
  - Dell PowerEdge R900/R710 with Linux OS using VMware ESX 4.0
  - Dell PowerEdge R900/R710 with Window OS using VMware ESX 4.0

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

6 of 26

IBM p570 Servers with the IBM System p Virtualization Technology

- Dynamic **LPAR** (**L**ogical **PAR**titioning) allows clients to dynamically allocate many system resources to application partitions without rebooting.

- Allow up to 16 dedicated processor partitions on a fully configured p570 system.

- PowerVM allows Micro-partitions defined as small as $1/10^{th}$ of a processor and be changed in increments as small as $1/100^{th}$ of a processor.

- Up to 160 micro-partitions may be created on a 16 core p570 system.

- VIOS (Virtual I/O Server) allows for the sharing of disk, optical devices, communications, and Fiber Channel adapters.

- Support for Multiple Shared Processor Pools and Shared Dedicated Capacity.

- Live Partition Mobility allows a partition to be relocated from one server to another while users are using application running in the partition.

POWER6™ Built on Power™

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

7 of 26

VMware vSphere 4.0 with Window OS and Linux OS

- Dell PowerEdge R900 Servers with 16-24 CPU cores and 128 GB memory

- Dell PowerEdge R710 Servers with 16-24 CPU cores and 128 GB memory

- Window 2003 Server (both 32 bit and 64 bit) Operation System

- Testing Window 2008 Server (64 bit) Operation System

- Linux Servers with Radhat Enterprise Linux version 5

- VMware vSphere 4 that includes VMware ESX. VMware vCenter, VMware vStorage, VMware HA, VMware vMotion, VMware Data Recovery, VMware vShield Zone, etc.

- Virtual Servers for hosting applications including web servers, Java applications, Microsoft .NET applications, GIS applications. ColdFusion applications, Domino applications

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

8 of 26

The VMware Virtualization in the EPA NCC

- The Intranet VMware Virtualization Environment in the EPA NCC

  ➢ 6 Dell R900 with 128 CPU cores and 128x6 GB memory

  ➢ Each Dell R900 has 8 NICs

  ➢ Currently hosting 180 virtual servers

  ➢ Currently hosting 50 VM Views (Virtual Desktops)

- The Internet VMware Virtualization Environment in the EPA NCC DMZ

  ➢ 4 Dell R900 with 96 CPU cores and 128x4 GB memory

  ➢ Each Dell R900 has 8 NICs

  ➢ Virtual servers set up in progress

- Customized VMware Virtualization for dedicated systems in the EPA NCC

  ➢ 4 Dell R710 with 96 CPU cores and 128x4 GB memory

  ➢ Each Dell R900 has 8 NICs

 ➢ Virtual servers set up in progress

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

9 of 26

The Enterprise GIS Environment at the EPA NCC

The GIS Database Environment

- ➢ Staging GIS database Servers
- ➢ Intranet Production GIS database Servers
- ➢ Internet (DMZ) Production GIS database Servers
- ➢ Hardware: IBM p570 AIX Unix OS
- ➢ RDMS Database Software: Oracle and PostgreSQL
- ➢ GIS Software: ArcSDE 9.3.1 and Oracle Spatial
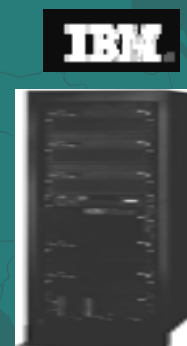
The GIS Application Environment

- ➢ Staging GIS application Servers
- ➢ Intranet Production GIS application servers
- ➢ Internet (DMZ) production GIS application servers
- ➢ Hardware: Dell PowerEdge servers and IBM Blade servers
- ➢ Window 2003 Server Operation System
- ➢ ArcGIS Server and ArcIMS 9.3.1

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

10 of 26

GIS Databases and Applications at the EPA NCC

GIS Databases in the EPA NCC

- Most EPA GIS databases are hosted as components of the overall Oracle database environment. One GIS database utilizes PostgreSQL.
- 8 national GIS databases are currently hosted on the EPA Internet and Intranet Production GIS database Servers.
- 7 national GIS databases are currently hosted on the EPA staging GIS database Servers.

39 GIS Applications in the EPA NCC
- ArcGIS Server 9.3.1, ArcIMS 9.3.1

134 Internet GIS Web Services
- geodata.epa.gov

242 Intranet GIS Web Services
- igeo.epa.gov

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

11 of 26

Top Three GIS Applications in the EPA NCC

MyEnvironment: http://www.epa.gov/myenvironment/

- Successful requests per month: 3,792,466
- Average successful requests per day: 126,418
- Data transferred per month: 22.79 gigabytes
- Average data transferred per day: 778.01 megabytes

Lead-Free Kids: http://www.leadfreekids.org/

- Successful requests per month: 3,029,090
- Average successful requests per day: 97,730
- Data transferred per month: 10.24 gigabytes
- Average data transferred per day: 362.54 megabytes

Enviromapper For Water: http://map24.epa.gov/emr/

- Successful requests per month: 1,460.249
- Average successful requests per day: 48,678
- Data transferred per month: 15.69 gigabytes
- Average data transferred per day: 491.28 megabytes

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

12 of 26

Important Procedures for Building A "Virtualized" GIS Environment at the EPA NCC

- Actively participate the development process for the overall EPA NCC Virtualization Plan.

- Full collaborations with other support team including the virtualization team, system administration team, network team, and database team.

- Ensure the "virtualized" GIS Environment as an integral part of the overall EPA NCC Virtualization Plan

- Testing GIS Software and GIS applications with virtualized servers in the staging environment first.

- Schedule the virtualization migration process for production systems during the weekend to minimize the impact.

- Conduct follow-up verification tests with GIS software and GIS applications on the virtualized GIS servers

- The GIS Server group was the first technology group to complete its virtualization migration process in the EPA NCC

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
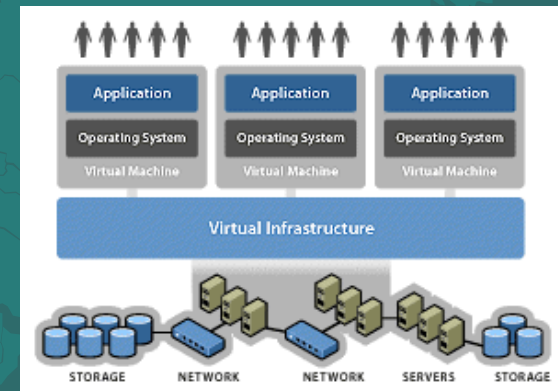July 13 – 16, 2010

13 of 26

- Start the P2V (Physical-to-Virtual) migration process with a staging GIS server that has smaller hard drives with less GIS software and GIS applications. Keep the process simple initially.

- Start the P2V migration process with staging GIS server that has large harddrives next. The migration process takes longer time to finish when dealing with large disk drives. Estimate the time for this P2V migration process with different types of GIS servers.

- For those staging GIS servers that can be shutdown for a period of times, use the "cold" P2V migration process that requires significant less times.

- For production GIS servers, use the "hot" P2V migration process where the physical server is still online during the migrating process. This process takes longer times but it reduces the downtime for the production GIS server.

- Based on estimated times for the P2V migration process with staging GIS server, set up the P2V migration process for the production GIS servers at the regular system maintenance window during the weekend.

- Keep the IP# and machine name the same after the P2V migration process. GIS software and GIS applications will work fine on the virtualized server since there is no change to the IP# and the machine name.

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

14 of 26

Direct Benefits From Virtual Servers for ArcGIS Server 9.3.1 in the EPA NCC

- Improved Manageability: eliminate and reduce planned/unplanned downtime.

- Ease of Provisioning; create and duplicate master clone very quickly with less costs

- High Availability (HA): VMware HA restart virtual server on another host with unplanned hardware failure

- Dynamic Resource allocation: allocate more resource to a virtual server dynamically when needed

- Better utilization of idle and under-utilized resources: idle and under-utilized resources from a virtual server can be shared by other virtual server

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

15 of 26

- With equivalent hardware resources at the individual server level, we did not observe performance improvement with the virtual server over the physical server.   This observation is consistent with the result from "ESRI ArcGIS Server 9.3 for VMware Infrastructure: Deployment Guide" by VMware, December 2009.

- A good SAN device provides performance improvement for virtual servers. Use of a good SAN device is strongly recommended with the virtualized GIS environment. A good SAN device allow you to best meet the I/O requirement and is often needed for leveraging advanced VMware features and capabilities.

- Virtual servers demonstrates performance improvements when they are integrated with other enterprise virtualization processes and resources such as desktop virtualization and enterprise storage system. We have observed better performances from virtual servers that utilize the enterprise SAN (Storage Area Network).

- Significant improvements from virtual servers can be achieved by integrating the virtualized GIS servers with the overall enterprise system resource such as the enterprise network devices. We can radically improve the reliability and the performance of virtual servers by combining virtualized GIS servers with advanced load balancing network devices.

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

16 of 26

GREEN CALCULATOR
Reduce Energy Cost & Environmental Impact with Virtualization

How many servers* do you plan to virtualize? **216** servers

*Calculations are based on the power consumption of a standard 2 CPU server

	Physical	Virtualized	Savings
Energy Savings:			
Annual Server & Cooling Energy Usage (kWh)	1,743,152	313,450	1,429,702
Cost Reduction:			
Physical Hardware[1]	$ 1,404,000.00	$ 270,000.00	$ 1,134,000.00
Annual Energy Cost[2]	$ 174,315.24	$ 31,345.05	$ 142,970.19

Environmental Impact:	Planting Trees	Cars off the highway [3]	Annual CO2 Emission(lbs/kg)[4]
These savings are equivalent to	4,320	324	1,917,230 lbs
			869,641 kg

[1] Assumes $6,500 per 2 CPU server
[2] Assumes $0.10/kWh, and 550 Watts per 2 CPU server
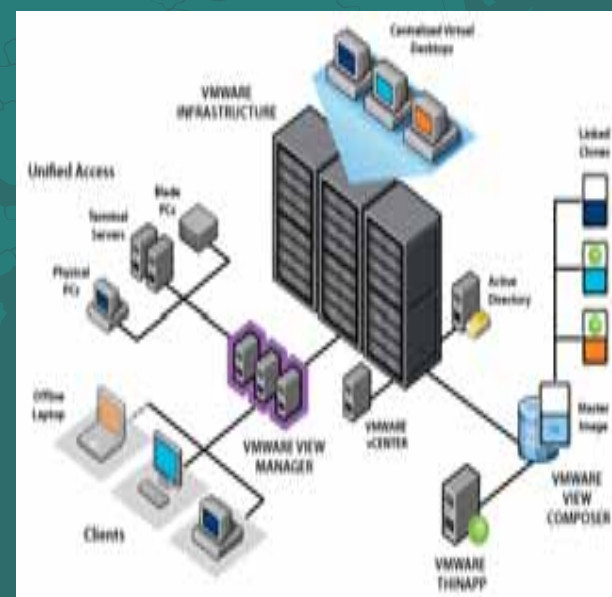[3] Assumes 12,000 miles per year and 20 mpg.
[4] Assumes 1.341 lbs CO_2 emission per kWh.

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

17 of 26

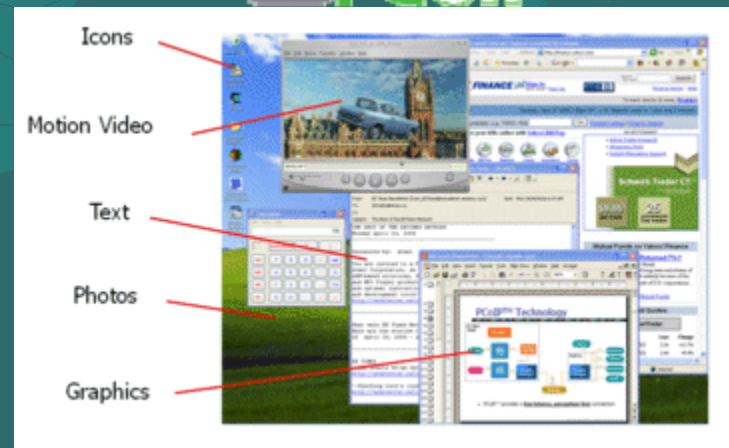# Server Virtualization and Desktop Virtualization in the EPA NCC

- Definitions of the server virtualization and the desktop virtualization.

- Some peoples consider server virtualization and desktop virtualization as two separated and unrelated processes. Server virtualization is inside the data center and desktop virtualization is outside the data center.

- There are great potential benefits for integrating server virtualization and desktop virtualization within an enterprise GIS environment.

- An enterprise GIS environment requires significant resources including hardware, software, application deployment, and security management. The challenge is to effectively deliver capabilities of the enterprise GIS environment to many users who may not be able to build an enterprise GIS environment on their own.

- The desktop virtualization technology can be used to effectively deliver capabilities of an enterprise GIS environment to many users over the wide area network (WAN).

- We have conducted various tests in the EPA NCC that utilize the desktop virtualization to deliver the capabilities of the virtualized GIS environment to many EPA users in EPA regional offices. The overall results are very positive.

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

18 of 26

Desktop Virtualization from VMware: VMware View 4.0

- Simplified Desktop management

- Automated Desktop Provisioning

- Advanced Virtual Desktop Image management with the Linked Clone

- High-performance desktop experience with PCoIP display protocol even with low bandwidth connections

- Built-in security with optional RSA SecurID for 2 factor authentication.

- High availability with no single point of failure and automatic failover

- Seamlessly access with locally attached peripheral devices such as printers, scanner and removable storage

Increase Desktop Availability

Icons	
Motion Video	
Text	
Photos	
Graphics	

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

19 of 26

Advantages of Virtualized GIS Environment in the EPA NCC

EPA NCC has adopted new network technologies with comprehensive WAN optimization solutions that deliver LAN-like application performance over the WAN in recent years

EPA NCC has implemented VMware server virtualization products and experimented VMware desktop virtualization products in recent years. These virtualization technologies can help the NCC to deliver enterprise GIS capabilities over the WAN.

EPA NCC has 12 years experiences of operating an enterprise Geospatial environment. Desktop virtualization can help EPA GIS users in other locations to utilize and leverage the vast experiences of EPA NCC support staffs and existing NCC hardware/software resources.

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

20 of 26

EPA Region 9

- VMware Views is quite user-friendly and easy to use. It provides a "Virtual XP Desktop" with the GUI that is exactly the same as a normal Window XP desktop.

- Most EPA users have reported pleasant experiences with the VMware View during the tests and would like to use it again.

- The response times for tests using data from NCC servers directly without the VMware Views are much slower than tests with VMware View consistently.

EPA Region 8

EPA WAN

- The response times for tests using data from NCC servers with VMware View are about the same as the tests using data from local regional servers.

- The performance of running GIS projects in Regional Offices using data from NCC database server with VMware Views are similar to the performance of running GIS projects in Regional Offices using data from local Regional database servers.

EPA NCC

- In few special cases, the response times for tests using data from NCC servers with VMware View are faster than the tests using data from local regional servers.

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

21 of 26

# VMware Views with the Virtualized GIS Environment in the EPA NCC

- VMware Views help to achieve significant performance improvement with GIS tests in Region 8 and Region 9 using data from NCC servers over the WAN.

- VMware Views help to demonstrates that the NCC has the ability to host GIS DR/COOP databases for EPA Regions to support various GIS projects in Regional Offices with adequate performance.

- VMware Views can meet the EPA Regional GIS DR/COOP needs in a cost-effective manner. VMware Views can be set up for many EPA users from a shared resource pool. When VMware Views are not in use, they do not consume any real resource. When EPA users access their VMware Views, the VMware system "enables" virtual desktops for its users dynamically.

- VMware View can provide significant helps to users in EPA organizations where network connections are slow and desktop machines are old.

- VMware View can be used to effectively deliver enterprise GIS capabilities over the WAN. Our RTPGIS team has been building centralized enterprise GIS resources and capabilities in the NCC for many years. The VMware View provides us a new mechanism to deliver these centralized GIS resources to EPA users at different locations with adequate performances

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

22 of 26

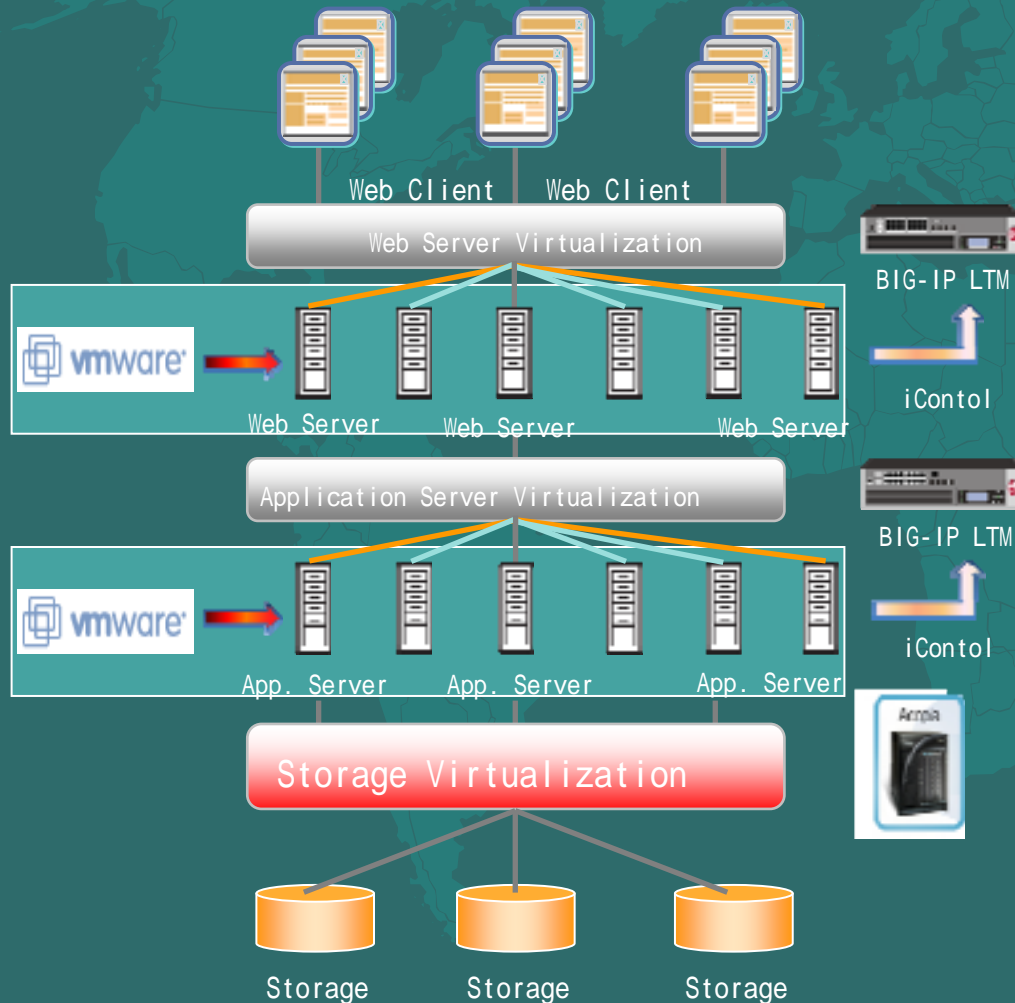# F5 BIG-IP Local Traffic Manager (LTM) and the EPA NCC Virtualization

- Local load balancing independent of blade center or subnet

- SSL offload to include FIPS 140-2 compliance from the client to the LTM
  - Between LTM and server dependent on application but with compensating controls

- SSL traffic inspection.
  - SSL packets are decrypted at the LTM allowing inspection for attacks (i.e. buffer overflows, sql injection).

- Improved certificate Management
  - Certificates be located on LTM, allowing single central group to manage digital certificates
  - Eliminate the SSL need for every server

- Load balance between any or all
  - Client and web server; Web server and application server; Application server and database server.

- Positive feedbacks from initial F5 LTM deployment
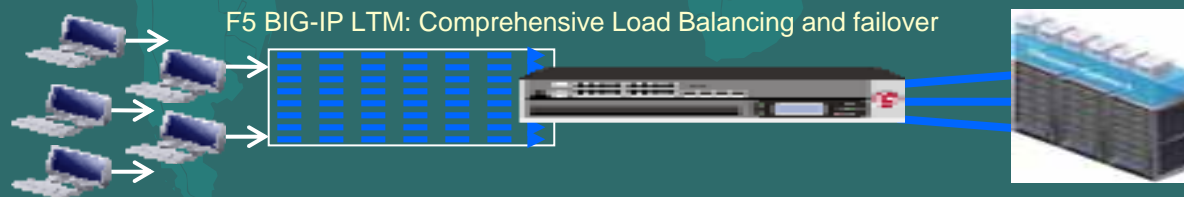
F5 BIG-IP LTM: Comprehensive Load Balancing and failover

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

23 of 26

F5 Local Traffic Manager (LTM) and VMware: Self-Turning Network (in progress)

Web Client   Web Client

Web Server Virtualization

vmware

Web Server   Web Server   Web Server

BIG-IP LTM

iContol

Application Server Virtualization

vmware

App. Server   App. Server   App. Server

BIG-IP LTM

iContol

Storage Virtualization

Acopia

Storage   Storage   Storage

1) Site traffic triples

2) VMware detect the traffic surge and starts bringing online new VM

3) VMware make an iControl API call to LTM and register the new VMs

4) LTM starts sending traffic to the new VM

5) Data center capacity increases and is capable of handling the traffic surge

6) Business continues uninterrupted

7) Traffic returns to normal

8) The process is reversed

Summary

- Integrating the virtualized GIS environment with the overall enterprise virtualization process can produce significant improvements and benefits that may not be available at the individual server level. Improvements can be maximized by combining server virtualization with other virtualization processes and enterprise resources such as desktop virtualization and advanced network devices.

- Building an enterprise GIS environment in the EPA NCC produces the following quantifiable benefits:
  - Increase agility: reduce new server provisioning time from days or weeks to minutes or hours
  - Improve availability: reduce planned maintenance windows
  - Faster and more economical recovery from data center outages
  - Improving system performance with smarter dynamic resource allocation
  - Better utilization of consolidated system resources by sharing idle or under-utilized resources
  - Significant reduction of electricity consumption, floor space usages, and cooling requirements

- At the individual server level, virtual servers may not have performance improvement over physical server with equivalent hardware resources. However, virtualization benefits at the overall enterprise level will provide enough incentives to promote widespread adoption of the virtualization technology.

- Server virtualization with the enterprise GIS environment can improve and accelerate the its transition to the Cloud Computing in the future.

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

25 of 26

Questions ?

Qi Dai
Technical Lead
GIS Operation and Support
EPA National Computer Center
CSC Team for the ITS-EPA Contract
Research Triangle park, North Carolina

Phone: 919-767-7285
Email: dai.qi@epa.gov

Tim Richards
Geospatial Service Manager
National Computer Center
Office of Technology Operation and Planning
EPA Office of Environmental Information
Research Triangle Park, North Carolina

Phone: 919-541-5307
Email: richards.tim@epa.gov

Presentation at the 2010 ESRI International User Conference
San Diego Convention Center, San Diego, Californian
July 13 – 16, 2010

26 of 26