Distr.
GENERAL

WP.6
25 April 2012

ENGLISH ONLY

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE (UNECE)
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

Meeting on the Management of Statistical Information Systems (MSIS 2012)
(Washington, DC, 21-23 May 2012)

Topic (i): Architecture

Virtualizing and Centralizing Network Infrastructure at a Decentralized Federal Statistical Agency

Contributed Paper

Prepared by Joseph L. Parsons and Elvera T. Gleaton
United States Department of Agriculture/National Agricultural Statistics Service
U.S.A.

I. Introduction

1. The mission of the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) is to provide timely, accurate, and useful statistics in service to U.S. agriculture. To do so, the National Agricultural Statistics Service (NASS) conducts hundreds of surveys every year and prepares reports covering almost every aspect of U.S. agriculture. NASS conducts its surveys through 46 field offices and seven call centers across the United States. Our main call center is also our new National Operations Center, which develops survey instruments and conducts training and frames maintenance. To be more effective and efficient in meeting its mission, NASS has migrated from its distributed local area network (LAN) environment to a centralized, virtual environment in which employees can access data, applications, and systems anywhere and anytime from their desktops, laptops, and smart phones.

2. In this paper, we discuss the motivation for NASS's virtualization effort, the business considerations that went into the decision, the technical solutions developed, the implementation process, and the challenges and opportunities that remain. The transition to a virtual environment is one of several transformational efforts at the agency. We do not focus on these other initiatives, but we discuss their relationship to the virtualization effort where appropriate. For example, NASS currently is seeking approval of a plan to consolidate resources to nine regional offices and maintain a minimal presence in the remaining field locations. Virtualization is an enabling technology to facilitate this transition.

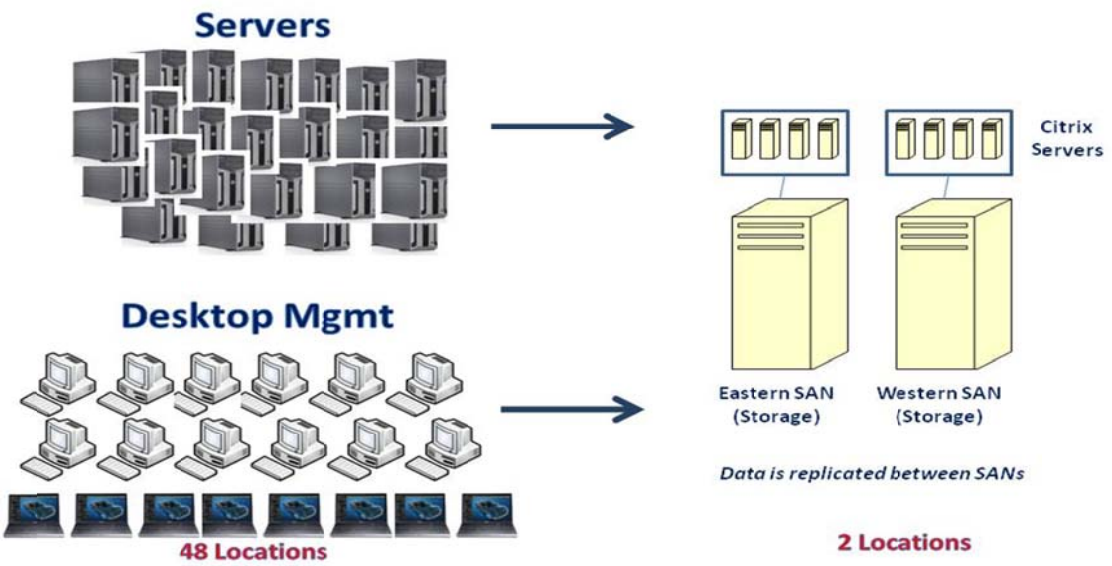
II. Motivation

3. At a high level, government-wide mandates motivated our virtualization effort. For example, since the tragedy of September 11, 2001, security has become a major concern within the federal government and has led to a movement to centralize rather than decentralize information technology (IT) systems and resources to protect them better by having to physically secure and monitor fewer facilities. More recently, concerns about energy utilization have led to directives to consolidate and centralize servers to save energy. President Obama's administration has called for federal agencies to cut energy consumption in their data centers as part of the broader effort to green federal operations. The president signed an executive order on October 5, 2009, requiring agencies to adopt best practices to manage their servers in an energy efficient manner. Prior to the virtualization initiative, NASS field office servers operated at 20–30 percent utilization; consolidating our servers offered immediate energy savings.

4. The Department of Agriculture's Chief Information Officer has directed agencies to migrate their servers to USDA data centers. Due to the need for NASS as a federal statistical agency to maintain a strong degree of independence, the Department agreed to allow NASS to continue to manage our LAN servers but required us to centralize the servers for security and energy reasons. Implementing federal mandates and initiatives requiring agencies to support telework and provide remote access to their networks also necessitated changes in how NASS operates. Virtualization can play a role in meeting all of these overarching directives.

III. Business Considerations

5. Although these larger issues underscored the need for consolidation, our agency business needs drove the decision. We wanted to use staff resources more efficiently—to remove location as a barrier to shifting work among agency offices in response to resource shortages or a business interruption. We saw value in standardizing and centrally administering the LAN environment. Our solution was designed to streamline servers, storage, desktops, software, security, and staff. It consolidated 94 physical servers across the country into 44 virtual servers in two locations. It also converted our physical desktop footprint into virtual desktops. Figure 1 depicts the impact of the LAN centralization process.



6. Virtualization technology has been around for many years. Statistical agencies commonly use virtualization to make microdata available to researchers while controlling the environment. Virtualization is rapidly transforming the IT work environment; it lets organizations run multiple virtual machines on a single

physical machine, sharing the resources of that single computer across multiple environments. With a couple of exceptions, NASS chose to virtualize the entire desktop rather than the more common technique of virtualizing individual applications to run from a server. We determined that virtualizing each individual application was not tenable because of the large number of smaller legacy applications within the agency. Depending on definitions, NASS has nearly 500 individual applications. Another NASS initiative focuses on greatly reducing this number using metadata and enterprise-level solutions. In the interim, however, we need a viable environment to support the myriad applications.

7. Consolidation and virtualization of LAN services improves the agency's operational efficiency through increased integration, streamlining, standardization, and flexibility.

(a) Increased ***integration*** allows all users access to the same Windows desktop image, making it more efficient to manage desktops centrally. With data in only two locations, NASS has a disaster recovery system in which we can immediately fail over to the other system when one system goes down for any reason. Integration also enables NASS to efficiently offer more applications. Since all applications have to reside on only a few distinct virtual desktops, NASS can develop and adopt more enterprise-level application choices, as well as create specialized virtual desktops when there is a defined business need.

(b) Increased ***streamlining*** means LAN administration can be performed by fewer full-time employees from a centralized location, enabling NASS to reduce the staff resources devoted to LAN maintenance. Desktops are managed centrally, with applications, updates, and patches installed once, and then the desktop is deployed to all users. Administrators no longer go desk-to-desk to apply updates or deploy software. This reduces the amount of duplicate data on multiple physical servers, as well as the cost of procuring replacements for those servers. Reducing the number of locations containing physical servers and electronic personally identifiable information (PII) strengthens computer and personal security.

(c) Increased ***standardization*** means all users with the same role have the same desktop. As part of standardizing our server configuration and deployments, we are also standardizing directory structures, roles, and access rights. Standardization is an important element in the failover and disaster recovery capacity the system provides.

(d) Increased ***flexibility*** means the centralized environment allows NASS staff to access applications/data from anywhere at any time. The virtual desktop can be accessed from anywhere in the world through any web browser from various types of devices, including laptops, desktops, iPads, and smart phones. This ease of access facilitates telework, continuity of operations, and disaster recovery. Users receive the same desktop remotely as they do in the office. NASS publishes over 500 reports per year, many of which have very short cycles between survey data collection and publication. For this reason, continuity of operations for every part of the country is important. As long as staff members have Internet access, any task that does not require a physical presence can be performed from anywhere. Moreover, because our data are replicated in near real-time between two sites, our disaster recovery plan provides failover capacity if an incident occurs affecting one of our physical server sites.

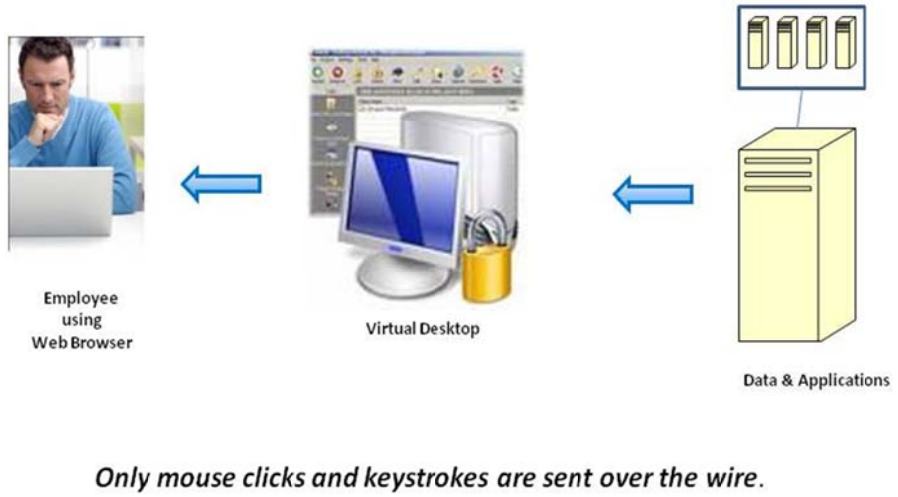
IV. Technical Solutions

8. To design and configure its centralized virtual environment, NASS engaged in significant research and technical consultation. The detailed design document runs 168 pages, and the installation and configuration documentation exceeds 60 pages. This section highlights major aspects of the effort and the specific path the agency chose.

9. The centralized solution consists of two server farms. Our field office locations have been separated into two clusters, with each cluster connected to either the eastern or western server farm. Data are continually replicated between the two server farms, which are critical pieces of the centralized system. Devices within the server farm are set up and configured as a private network. Each server farm is connected to users by four

virtual local area networks (VLANs). Figure 2 shows (right to left) how the centralized system connects to the user, who connects to the centralized environment using a web browser. A virtual Windows desktop is sent to the user's web browser from the server farm. Generally speaking, no data are sent over the connection.

Figure 2: Secured Application Delivery



10. The centralized environment consists of seven critical pieces: software, desktops, servers, telecommunications, storage, disaster recovery, and environment management.

11. **Software:** NASS uses the Citrix suite of virtualization software, which includes Citrix XenDesktop and Citrix XenServer. Microsoft Windows Server 2008, Microsoft Terminal Licenses, and Microsoft SQL Server are also part of the software solution. The software allows us to virtualize the servers, applications, and desktops. Virtual applications enable us to deliver legacy applications over the web without rewriting the application.

12. **Desktops:** The Citrix XenDesktop software enables us to deploy virtual desktops to NASS employees. Virtual desktops look exactly like a regular Windows desktop; they are generated from one main image and then distributed to NASS users. They are stored on servers in the server web farms. When NASS employees log into the NASS network, they are automatically routed to their virtual desktops. Employees access all NASS applications through a web browser. All software and applications are deployed through the virtualized desktops. The virtual desktops are read-only so users cannot install any software or devices; they can, however, change their wallpaper and minimum settings. Each of the two web farms is set up to service up to 1,300 virtual desktops.

13. With the virtualization effort, NASS has begun transitioning to thin-client machines. These machines are inexpensive and require little service while still providing full functionality because applications and processing are largely delivered from the servers. The introduction of thin-client machines has resulted in savings in our desktop refreshment program.

14. **Servers:** As noted, NASS uses two server farms. To ensure constant availability, each server farm hosts a cluster of virtualized servers across a Dell blade chassis. We use the Citrix XenServer to virtualize the servers. We consolidated the 94 physical servers into 44 virtualized servers in the two locations.

15. **Telecommunications:** We provide three levels of redundancy for the telecommunications in each served location (primary, backup, and disaster recovery). The speed and solution vary by need and location. The primary and backup redundancy is a managed service that automatically fails over as needed. In addition, we

purchased and installed BlueCoat MACH5 ProxySG WAN optimization equipment to reduce bandwidth consumption by compressing traffic and byte caching.

16. **Storage:** Our two storage area networks (SANs) are dedicated networks for storage that provide access to consolidated disks. The disk can be broken into areas (partitions or virtual disks). The SANs provide high-speed shared storage for our virtual servers and desktops. NASS selected NetApp as the SAN provider, and uses four NetApp controllers to access disks in the storage network. Each SAN has several terabytes of storage with additional shelves for expansion. This means each SAN has sufficient disk space to hold all NASS data, plus the storage for the virtual servers and virtual desktops. Our intention going forward is to coordinate the SANs for our UNIX and Windows environments.

17. **Disaster Recovery:** We have implemented redundancy in all of the critical areas, providing redundant servers, storage, databases, switches, and telecommunications. In case of failure of one server farm, we can connect users to the other server farm. If a server farm experiences a problem and is down a significant amount of time, we will fail users over to the other server farm. This is currently a manual process that involves several steps. We are in the process of automating, enhancing, and streamlining our disaster recovery and failover processes. We expect to employ load balancing that will automatically shift users to the other server farm in the event of a failure at either farm.

18. **Environment Management:** NASS purchased Solarwinds software to monitor the bandwidth utilization of the virtualization effort. In addition, we are expanding our use of this tool to monitor many appliances and other entities throughout our environments. Citrix also includes some tools for deeper monitoring and evaluation of environmental impact.

IV. Implementation Process

19. In April 2010, two NASS field offices were migrated to the centralized environment as a proof of concept. The virtualization team traveled to the field offices and monitored bandwidth, applications, and performance. The proof of concept was conducted over two months.

A. Milestones

20. Once the proof of concept was shown to be successful, NASS conducted a pilot migration with two additional field offices in July 2010. Following the four migrations, an external verification and validation of the centralized environment was performed by representatives from Citrix, Blue Coat, and Dell. The NASS team then implemented recommendations from the validation study and began the agency-wide migration in October 2010. The team encountered several additional issues during the migrations. If they were critical to operation, the problems were addressed. However, if the problems were improvements to the system, they were documented and identified in the project close-out report. The final migrations were completed in late September 2011.

21. The implementation team decided the smoothest transition to the virtual environment would be to travel to each location to perform the migration. Each transitioning office was assigned a point of contact from the migration team. The point of contact was responsible for all communication between the migration team and the transitioning office.

22. The team ensured each transitioning office met the pre-migration requirements. To proceed with the migration, each office had to have: (1) listed all of its applications in the agency’s application portfolio; (2) reviewed its files and directory structure (the team archived old files and removed particular files from the system, for example, jpg, wpd, and mp3 files); (3) upgraded its telecommunications; (4) installed WAN optimization and new router equipment; and (5) performed pre-testing.

23. The typical process was for the team to travel to the transitioning office on Sunday. The team point of contact met with office management and staff to demonstrate the new environment and answer staff questions.

On Monday evening, the team re-imaged all desktops, removing all locally installed hardware, and migrated the office data to the centralized environment. The team was in the office when the first person arrived on Tuesday morning to provide support for the desktop and virtual environment. Wednesday and Thursday were devoted to cleaning up/fixing any lingering issues, troubleshooting, and helping employees adjust to the new virtualized environment. The team then departed on Thursday.

24. Our change management process was a key to providing risk management and a vehicle for communicating change to all IT stakeholders. All changes to any of our applications or environments are submitted to the NASS Configuration Control Board (CCB) and documented within our CCB SharePoint site. During the implementation, any request for a change to the virtualized environment was submitted through the virtualization project manager.

B. Return on Investment

25. Return on investment for this transformation initiative has two major components: staffing and infrastructure (hardware and software).

26. **Staffing:** Prior to the initiative, almost all offices had a person identified to serve as the LAN administrator for the office. In fiscal year 2010, we conducted a survey of all information technology field office staff within NASS. The agency had 21 information technology specialists in our field offices spending an estimated 79 percent of their time directly on information technology activities. An additional 19 computer clerks and assistants spent an estimated 67 percent of their time on IT activities. The total salaries and employer-paid benefits for the IT-related time for these employees was approximately US$2.5 million per year. Through virtualization, much of this time can be saved or redeployed for other activities.

27. Since the virtualization initiative was implemented, some of these employees have retired or taken on other roles in IT or in another part of the organization. We estimated that we would need eight additional positions to support the virtualized environment (three positions related to server and storage, one network/telecom position, and four additional helpdesk positions). Some of these additional positions also support other initiatives that require IT support. We estimate the net effect of the additional positions as an expense of approximately US$600,000 per year, and the ongoing savings related to staffing to be on the order of US$1.8 –1.9 million per year.

28. **Infrastructure (hardware and software):** The agency IT budget plan already included planned expenses to replace file and print servers as well as annual personal computer refreshment costs; together these annual IT lifecycle costs were expected to be approximately US$3.25 million. The infrastructure savings from centralization and virtualization were calculated by subtracting the IT lifecycle costs from the project total investment cost (approximately US$2.5 million). Therefore the nonrecurring one-time life cycle savings related to infrastructure are approximately US$750,000.

V. Lessons Learned, Challenges, and Next Steps

A. Lessons Learned

29. *Deploy in stages over time.* Because the pilot migration to the Citrix environment was successful, we were able to deploy the technology quickly to other parts of the agency. A fast deployment becomes challenging with so many offices. We found it was important to maintain time in the implementation schedule to stop to fix small issues so they would not grow and escalate user frustration and learned several other important lessons.

30. *Involve and train operations and maintenance staff early.* Much of the implementation was done with contractor support and a team of key individuals. In some cases, the operational team was slow to accept the new environment, which pushed troubleshooting problems and issues on the implementation team. All key

members of the team that will ultimately support the environment need to be involved and familiar with the transition so they can effectively understand and support the new environment.

31. *Keep stakeholders up to date on the project*. Documenting frequently asked questions is useful as is communicating repeatedly with users fearful of the unknown. Migrating staff in phases and getting user acceptance from all areas is vital.

32. *Develop and implement standard operating procedures for key tasks.* We focused on right-sizing hardware and avoided the least expensive route for equipment. The technology continues to evolve and improve, so it is important to continue the analysis, design, build, test, and implement model once NASS is in the centralized environment.

33. *Develop a robust, always available operation*. With all infrastructure concentrated in two locations, any outage is a high-impact event affecting at least half the organization. Since we have operations stretching from New Hampshire on the east coast to Hawaii and calling centers that work late into the evening and weekends, our maintenance windows are tight.

B. Challenges and Next Steps

34. Switching to a virtual environment was a major change for our organization. As this document is being prepared, it is about six months since implementation was completed. At that time, all NASS groups had access to the virtual environment and the virtualization mandate had been met, but many important tasks remained. It took a period of time to optimize the environment and become proficient at troubleshooting problems.

35. One challenge was that several key members of the implementation team had built up large amounts of leave that had to be taken before the end of the calendar year. Due to an acute budget issue, contract staff were not available for several months immediately after implementation. Our operation and maintenance staff were still learning the nuances of managing the system; in the process of troubleshooting other issues, they made changes to the environment that caused stability issues. After a few months, we have worked through many of these problems and the user environment is much improved. However, we also had several slow network days and partial service outages before the system became stable again.

36. A virtualized desktop requires that Windows profiles retain their settings from one virtual session to the next. Downloading the profile to the local machine and uploading the profile when a user logs out becomes a potential point of failure or corruption. A pervasive issue with managing profiles in this manner is “profile bloat.” User profiles can become corrupted or grow to large sizes, greater than a gigabyte, which causes slowness in the system and extreme login times. NASS began having profile issues late in the implementation phase, especially for staff that used particular software or had particular settings. A more effective way of dealing with profiles is to deploy a relational database to house the registry settings and other information. We are converting to use the Appsense suite of virtualization products to manage our profile information; Appsense—along with proactive management of profiles in parts of the organization not yet converted to Appsense—has dramatically reduced login times and increased performance.

37. After the budget issue was resolved and expert contractors are again available, the NASS operations and maintenance team developed skill in handling the virtual environment and tuning the system to reduce bandwidth requirements and optimize processing. For example, after nightly maintenance but before the new business day begins, virtual machines are preloaded and ready for users to connect, thereby reducing wait times.

38. We are implementing a plan to reduce the time it takes to convert from one site to the other in the event of a disaster from the current four or more hours to mere seconds. Reducing failover time requires moving all profiles to a relational database, making configuration changes to our storage networks, and balancing the load between the two server farms.

39. Upgrading software is easy in a virtual environment but we need to implement a regimented plan of testing virtual desk images while making them quickly available to our users.

40. Some software does not perform optimally in NASS's desktop approach to virtualization. Blaise software developed by Statistics Netherlands provides an example. NASS uses Blaise for data collection and editing. A decentralized version of the software works fine because the data associated with the records enumerators will call resides on the virtual server from which they will call. However, we are migrating to a version of Blaise that utilizes a single relational database. The current architecture for that implementation of Blaise has the database and application server co-located with the server farm. A second application server resides at the western server farm. Sites accessing the eastern server farm experience no problems, but those trying to access the western server farm experience slow response times. The thin-client machines attached the western server tend to experience communication drops when using this application due to the lags. This led to hung sessions and a very frustrating user environment. Most applications in a virtualized desktop solution utilize very little bandwidth but, as configured, Blaise requires significant bandwidth. We have been able to improve performance by tuning our bandwidth optimization solution. However, the best solution is likely to virtualize the application such that all users come to a single virtualized server so that the database and the application are collocated. In Citrix, this is an implementation of Xenapp software. We will likely follow a similar path with Windows SAS software and our geospatial information software (GIS) applications. While our challenges with centralized Blaise may be unique to NASS, the challenge of tuning a virtualized environment after implementation will be common to most organizations on the path to a virtualized network.

41. Although the six months since virtualization was implemented have been challenging for our IT professionals and our users, the effort has clear value. NASS currently is seeking approval of a plan to consolidate resources from our 46 field offices to nine regional offices. We expect to maintain a minimal presence in the remaining field locations. These changes will allow the agency to operate more efficiently and consistently and produce higher-quality data products. However, it will take some time for the transition to be completed and physical moves to take place. Because of the virtualized network, we developed a way for our field offices, within a matter of weeks, to begin to function as virtual regional offices sharing work and resources.

References

Accelera Solutions. 2011. *Virtualization for Military Healthcare*. White Paper.

Citrix.com, 2010. *The Complete Solution for Enabling Bring Your Own Device (BYOD).* Santa Clara, California. **http://www.citrix.com/site/resources/dynamic/additional/BYO_completetechnicalsolution.pdf**

Haka, Fabio. 2010. *USDA-NASS XenDesktop Configuration Document.* Accelera Solutions Inc.

Nealon, Jack. 2010. *Database-Optimized, Generalized, and Modular Application Vision and Implementation Plan*. United States Department of Agriculture, Washington, D.C.