Server Consolidation

Report to the Joint Legislative Oversight Committee on Information Technology

Chris Estes

State Chief Information Officer

Office of Information Technology

December 2013

This page left blank intentionally

Contents

This page left blank intentionally

Legislative Request

A provision in the 2013-2015 budget authorized the State Chief Information Officer (State CIO) to develop an inventory of servers and server locations in State agencies and to develop a plan to consolidate agency servers in state-owned data centers.

Section 7.4(a) of Session Law 2013-360 requested a written plan for server consolidation to the Joint Legislative Oversight Committee on Information Technology and the Fiscal Research Division.

The complete text of the legislation can be found in Appendix A.

Report Focus

After passage of the legislation, the Office of Information Technology Services (ITS) collected data from Executive Branch agencies in September and October of 2013. Data from the survey was analyzed to produce a risk profile, a summary of current conditions and a plan for remediation.

The goal of this initiative is locating all servers supporting the Executive Branch into state-owned data centers with a Tier-3 rating, as defined by the Uptime Institute. A Tier-3 rating means a data center has redundant capabilities.

The objective will be the same as previous server efforts – minimizing risk. Priorities will be based on that goal and potential cost savings. The duration of the program depends on available staff and funding for one-time and recurring costs.

The broader goals include reducing locations, faster incident resolution, improved availability and reliability, and lower costs to the state.

Introduction

Background / History

Legislation in 2004 that established an initiative to reduce IT infrastructure duplication led to ITS providing core IT services to 14 agencies and commissions. Server consolidation, including increased use of virtual servers, was a key component of that effort. A virtual server is one machine that hosts multiple, isolated applications.

Financial challenges during the economic downturn reduced the scope of the effort to focus on mitigating identified risks, reducing costs and supporting changing business needs.

The most recent server consolidations occurred at the Department of Natural and Economic Resources (DENR) Green Square and a DENR site in Morehead City. At Green Square, the total number of servers was reduced by 72 percent. All production servers are now running at the ITS Western Data Center. Development and test servers operate at the Eastern Data Center and provide disaster recovery capability.

At Morehead City, the number of servers was reduced by 79 percent. All production servers are at the Western Data Center.

Current State of the Servers in the Executive Branch

The server data for this report was compiled in the survey of state agencies. Thirty agencies and commissions across the Executive Branch supplied data.

In the survey, agencies reported 5,282 servers, which are used for internal operations and to provide services to citizens, are in 396 separate locations in 125 cities across North Carolina. Almost 40 percent of the servers are in facilities that are in facilities below Tier-3 standards.

Of the servers reported, approximately 18%, or 973, are no longer under the manufacturer's warranty or a maintenance contract, increasing the risk if there is a failure.

Details of the survey can be found in Appendix B.

The tier ratings for the data center locations are derived from an assessment that was completed in 2011.

Tier Ratings and Risk

The Uptime Institute classifies data centers based on a number of criteria that result in a "Tier" ranking from 1 (lowest) to 4 (highest). Tier-3 data centers should be the minimum tier used in this initiative. Tier-3 rating specifically provides:

- Redundant capacity components and multiple, independent power distribution paths.
- Sufficient, redundant mechanical / electrical / plumbing capacity to meet the needs of the data processing systems.
- Maintenance activities and certain unplanned events can occur without interruption to the computing systems.

The seven Tier-3 locations in the Executive Branch are the ITS Eastern Data Center and Western Data Center, the Department of Health and Human Services Public Health, the Adult Corrections facility on Yonkers Road in Raleigh, the Department of Revenue, the Department of Transportation's Century Center and what was previously the Employment Security Commission Data Center on Wade Ave. Of the 5,282 physical and virtual servers reported, 3,242 (61%) are housed in Tier-3 facilities.

Tier-1 and below locations have no redundant capabilities. Power and cooling are basic. A power outage will take the site down immediately with loss of work-in-progress. Of the 5,282 physical and virtual servers reported, 1759 (33%) are in Tier-1 or below facilities. A total of 281 (6%) physical and virtual servers are in Tier-2 facilities.

Virtual server farms are generally at agency headquarters and most servers at remote locations are physical servers. Therefore, support for the components of these physical servers, including backup devices, etc., in remote locations requires dispatching staff, adding staff time and travel expense. If a server is down due to hardware issues, it remains down until staff arrives. If replacement parts are required that the staff member did not bring, the hardware remains down until the parts are procured.

Location Consolidation Opportunities

One of the goals of this Executive Branch server consolidation is to eliminate as many of the current 396 unique locations as possible. The data analysis and future design will help to quantify the locations that are candidates for closing. For each location eliminated, the risk profile for that location is eliminated as well.

The three Tier-2 locations will be assessed to determine whether it will be more economical to raise them to Tier-3 standards and move servers there, or to migrate the servers into another Tier-3 data center and eliminate the location.

During final analysis, the locations will be prioritized to eliminate the highest risks first.

Server Risk Profile

The risk profile for servers – and risk to agency operations in case of failure – is based on several criteria:

Server Warranty – If the server is not under manufacturer warranty or a 3rd-party support contract, the state must pay more for service and parts. The state typically purchases a 3-year warranty from the manufacturer with the option for one renewal. Once the server reaches a certain age, the manufacturer will no longer provide a warranty. A third-party support company may offer a support contract, but only if spare parts are still available. These 3rd-party support contracts can be more expensive than the original manufacturer warranty.

Server age – As servers age, spare parts become rare or not unavailable. The more variety the state has in server manufacturers (and therefore internal parts), the bigger this problem becomes if the servers are not replaced on a regular schedule.

If a storage device, such as a hard disk, fails on an old server and parts are no longer available, the agency's data is at risk and may be lost.

Operating Systems – Once a server's operating system is no longer supported, the manufacturer no longer produces security patches. The longer the out-of-support operating system remains running, the more susceptible the server is to viruses. This includes the application(s) that run on the server and the data that is stored on the server.

Server Manufacturers – Each manufacturer uses its own internal parts and hardware and software drivers to control those internal parts and external interfaces. Due to the differences by manufacturer (and many times by model), standardization of images, tool sets, external drivers, even standard staff support processes cannot be efficiently achieved.

Appendix B shows the servers that were reported and the number that fall into each of the risk categories above. Many of the older servers fall into multiple risk categories. Completion of the data analysis and future design will determine this aggregate risk, allowing sites and servers to be prioritized by that aggregate risk.

Server Consolidation Opportunities

Virtual Servers – A Gartner report titled *2014 Planning Guide for Private Cloud, Data Center Modernization and Desktop Transformation* said server virtualization will surpass 65% penetration in the typical enterprise (lower for small and midsize organizations) during 2014. Ultimately, the average Gartner client expects to run 80% to 90% of its servers on a virtualization platform.

The current ratio across Executive Branch is 60% physical servers, 40% virtual.

Server consolidations drive the number of virtual servers as high as possible, given each agency's business requirements. At DENR, for example, the 65% Gartner estimate has been surpassed.

Server Footprint – The design of the consolidated server environment for each agency combines similar servers and eliminates others that are no longer needed. This results in a much lower total number of servers (along with the related expense of those servers) being required to support the business. Then, augmented by virtualization, the number of physical servers is drastically reduced.

The migration process is done in a way that minimizes the level of risk due to change. Servers are copied to their new environments, tested, and then put into production. If there are issues, an agency can revert to the old infrastructure until the issues are mitigated. This is much preferable to "fork lifting" the servers to a new location, where they may not power back up, cannot be tested until they have been moved, and cannot be used as a backup if there are issues.

Roadmap

At the highest level, ITS will perform health assessments on server inventories, review Tier-3 and Tier-2 locations, gather agency current cost information, design plans by agency, complete business cases and consolidate based on priority and staff availability.

We intend to use a portion of the funds appropriated for IT Consolidation in the IT Fund for this fiscal year to complete the short-term tasks listed below. To begin the work as soon as possible, and complete it quickly, we intend to bring in supplemental staffing as needed. Supplemental staff might include experienced server analysts and finance and business analysts if the work could not be done by ITS staff.

Short term

- Complete a detailed review of the Tier-3 data centers, to determine how much additional volume they can support.
- Based on the above, make executive-level decisions on which Tier-3 agency data centers will become shared data centers.
- Review the three Tier-2 locations and determine whether it is more economical to upgrade them to Tier-3 or consolidate the servers into another Tier 3 facility and close them.
- Complete server health assessments.
- Coordinate with the Office of State Budget and Management to work with each agency to identify current costs for server infrastructure.
- Compile the final matrix of the server inventories, health summaries, location summaries, risk profile, and resulting priority ranking.
- Complete business cases by agency and submit to OSBM for comment and approval.
- Assess need for additional appropriation of one-time and recurring costs and submit requests.

Middle term

- Begin server consolidations in priority sequence, based on staffing and funding availability.
- Follow process for project approval and monitoring by the State CIO's Enterprise Project Management Office.

Long term

- Continue the effort until all are complete. Timeliness to complete the initiative will depend heavily on (a) available technical resources, (b) one-time project investment funds, (c) recurring costs being built into agency budgets ahead of time, and (d) executive-level support throughout the program.

Issues & Cost

Cost drivers

Cost drivers, which will guide the speed at which the duplication can be eliminated, break down into two areas.

One-time costs – These include new virtual and physical servers, project management time and possibly temporary technical staff to allow work in multiple agencies simultaneously. This would also include any data center upgrade costs, should the decision be made to upgrade any of the Tier-2 data centers to Tier-3.

Recurring charges – The design for each agency will specify the number of servers, whether they are virtual or physical and their use, such as applications, database, webs, or file storage. The design will also specify the amount of data storage that is required, backup requirements and any additional hardware needed for disaster recovery. These details will allow OSBM and the agencies time to adequately budget.

Cost recovery for any Tier-3 agency data centers that will be used by multiple agencies also must be considered. The costs of operating the two ITS data centers are covered by rates charged to users. Currently, the Tier-3 data centers are used only by the agencies that own them, so some cost-sharing mechanism would be needed.

Appendix A

The text of Section 7.4(a) of Session Law 2013-360 follows.

SECTION 7.4.(a*)* Server Inventory. – The State Chief Information Officer (State CIO) shall develop an inventory of servers and server locations in State agencies. Based on this inventory, the State CIO shall develop a plan to consolidate agency servers in State-owned data centers. By November 1, 2013, the State CIO shall provide a written plan for accomplishing this to the Joint Legislative Oversight Committee on Information Technology and the Fiscal Research Division.

Appendix B

Server locations

- Servers are housed and running in a total of 396 unique locations.
- Seven (7) of these locations are rated Tier-3 data centers.
- Three (3) of these locations are rated Tier-2 data centers.
- Remaining 386 locations are Tier-1 or below.
- These 396 locations are in 125 cities across the state.
- 54 of the 125 cities have multiple Executive Branch agencies, each with their own location(s), in the same city.

Server Metrics:

- Total servers reported = 5,282
    - Physical Servers = 3,179 (60%)
    - Virtual Servers = 2,103 (40%)

- Servers in Tier-3 data center locations = 3,242 (61%)
- Servers in Tier-2 and below locations = 2,040 (39%)

- Physical Servers under Manufacturer's Warranty or Maintenance Contract = 2,159 (68%)
- Physical Servers not under Manufacturer's Warranty or Maintenance Contract = 973 (31%)
- Physical Servers where Warranty Situation is Unknown = 47 (1%)

- Physical Servers that are 0-2 years old = 639 (20%)
- Physical Servers that are 2-5 years old = 1,118 (35%)
- Physical Servers that are 5-10 years old = 1,346 (43%)
- Physical Servers that are greater than 10 years old = 76 (2%)
    - Oldest physical server reported as still running in Production = 15.7 years old

- Number of manufacturers of the current 3,179 physical servers = 11

- Number of servers running supported server Operating Systems = 3,769 (71%)
- Number of servers running out-of-support Operating Systems = 1,320 (25%)
- Number of PCs with PC Operating Systems acting as servers = 193 (4%)

- Applications running on these 5,282 Executive Branch servers, per the EPMO APM tool:
    - Statewide Critical = 124 (Highest Criticality)
    - Department (Agency) Critical = 163 (Second Highest Criticality)