Agilysys

Commonwealth of Virginia
Virginia Information Technologies Agency (VITA)

Audit of Northrop Grumman's Performance Related to the DMX-3 Outage and Associated Infrastructure

2/15/2011

Table of Contents

Definition of Terms Used in Document

BIA	Business impact analysis is an assessment of an outage's impact to business operations.
CESC	Commonwealth Enterprise Solutions Center is the primary data center located Chester, VA.
CI	Configuration item is a record in the configuration database that describes hardware or software in the enterprise computing environment.
CIA	Comprehensive Infrastructure Agreement is the contract between Northrop Grumman and the Commonwealth of Virginia to provide IT services.
Clones	Clones are full copies of disks containing data.
CMDB	Configuration management database is a database that contains all relevant information about the components of the information system used in an organization's IT services and the relationships between those components
Database	A database is a collection of information that is organized so that it can easily be accessed, managed, and updated. Databases can be classified according to types of content: bibliographic, full-text, numeric, and images.
Data Corruption	Data corruption as it pertains to this document is data that appears to be available such as a Microsoft Word file but the file is unusable due to problems within the file system or disk drive.
DLCI	Driver's License Central Issuance is the application supporting the issuing of driver's licenses in the Commonwealth of Virginia.
Data Loss	Data Loss as it is applied to in this document is data that was lost and is not recoverable from tape or other media due to data corruption or another event such as file deletion.
DMX-3	The DMX-3 is an EMC enterprise, best-of-breed storage array.
DR	Disaster Recovery is the process, policies and procedures related to preparing for the recovery or continuation of technology infrastructure that is critical to an organization after a natural or human induced disaster.
EMC	EMC Corporation is a manufacturer of enterprise best of breed storage arrays, subcontracted by Northrop Grumman
EMC, Ionix Control Center	Control Center is EMC's storage resource management software solution that simplifies and automates discovery, monitoring, reporting, planning, and provisioning in large, complex environments.
Global Catalog	The Global Catalog is the mechanism used by the enterprise backup application to track all backup and restore activity.
Hard Error	Hard errors are an error in a computer system that is caused by the failure of a memory chip. The solution to a hard error is to replace the memory chip or module entirely.
HP OpenView	Hewlett Packard (HP) OpenView is the HP Network and System Management and Monitoring Software used to manage complex data center environments.
ITIL	Information Technology Information Library is a set of concepts and practices for managing Information Technology (IT) services.

ITSCM	Information Technology Service Continuity Management is the management of services and processes as they apply to business continuance operations and is a component of ITIL.
ITSM	Information Technology Service Management is a discipline for managing IT systems, centered on the customer's perspective of IT's contribution to the business.
RCA	Root Cause Analysis as it pertains to this document is a document that identifies the root cause of an incident or problem.
RMAN	Recovery Manager is the Oracle native tool for backing up and recovering an Oracle database.
SAN	Storage area network is a type of computer architecture in which remote computer storage devices (such as disk arrays, tape libraries, and optical jukeboxes) are attached to servers in such a way that the devices appear as if they are locally attached to the operating system.
Snapshot	A Snapshot is a picture of data at a point in time.
Soft Error	Soft errors are errors that occur in a memory system that changes an instruction in a program or a data value. A soft error will not damage a system's hardware.
SQL	Structured Query Language is a standard language used in Microsoft databases.
SRDF	Symmetrix Remote Data Facility is a process used to replicate data from a local storage array to a remote storage array.
SWESC	Southwest Enterprise Solutions Center is the data center used for disaster recovery of the CESC data center.
Symantec NetBackup Ops Center Analytics	OpsCenter Analytics, formerly Backup Reporter, helps enhance backup and archive operations and verify service level compliance by using more in-depth policy and schedule information, and align backup and archiving with the business.
VDSS	Virginia Department of Social Services.
VITA	Virginia Information Technology Agency.

Table 1: Definition of Terminology

Executive Summary

Agilysys provides professional information technology services including, but not limited to, enterprise architecture and high availability, infrastructure optimization, storage and backup management, identity management and business continuity for Fortune 50, 500 and mid-tier customers located in the financial, telecommunication, service provider, health care, education, government and manufacturing sectors. Within these technologies Agilysys maintains a Professional Services organization comprised of individuals that have extensive industry experience and maintain certifications including, but not limited to, Cisco, EMC, HDS, HP, IBM, ITIL, Microsoft, Oracle, Project Management Institute, SNIA, Symantec and VMWare.

In this audit, Agilysys does not distinguish between state and local government service providers and other industry service providers. In the professional opinion of Agilysys, this is because the best practices for technologies typically deployed in IT service implementations remain the same regardless of sector. Even though the exact implementations may differ, the same best practices are applied.

Review Process

This audit was prepared by Agilysys following a three-month review and is based upon data collected during interviews with Northrop Grumman, EMC, several state agencies, and an analysis of backup, storage, server, database and monitoring systems.

After completing the research for this audit, Agilysys provided more than one opportunity for staff from the Virginia Information Technologies Agency (VITA), the Joint Legislative Audit and Review Commission, and Northrop Grumman to review the document and provide technical feedback. As part of this process, both VITA and Northrop Grumman were given the opportunity to provide additional documentation and to submit formal, written responses in the form of a letter (Appendix A).

Key Findings

The information included in this audit covers many aspects of the current technical architecture, operational processes, incident responses, data storage, data recovery, disaster recovery and management of the reviewed environment. This audit details several areas where, in the professional opinion of Agilysys, Northrop Grumman failed to meet industry best practices commonly used by top tier service providers with which Agilysys is familiar. Solutions to issues that have been observed and reported range from simple to complex and are reviewed in detail later within the document. In other instances, however, Northrop Grumman is using best-of-breed practices that meet or exceed industry best practices.

The review of the Root Cause Analysis (RCA) provided by Northrop Grumman indicated that Northrop Grumman did not meet contractual requirements pertaining to delivery and content of an RCA. In addition, this audit uncovered missing information that Agilysys would have expected to be in a complete RCA. Although Northrop Grumman's RCA provided detailed information on the cause of the incident and subsequent actions to recover operations, it failed to provide key data points on the cause of the memory board failure and any corrective actions taken to mitigate a future failure. Information regarding the root cause of the memory board failures was provided to Agilysys during the course of the audit as a separate document. As a result of the audit, Agilysys has identified the following key points:

- Human error during the memory board replacement process resulted in the incurred extended outage (*Root Cause Analysis and Predictive Analysis*).
- The dual memory board failure was reported by EMC to be caused by an electrical over stress condition at the component level. The reason for the over stress is not known (*Root Cause Analysis and Predictive Analysis*).
- The loss of backup data was attributed to corruption of a primary element of the enterprise backup and recovery system (*Root Cause Analysis and Predictive Analysis*).
- In the professional opinion of Agilysys, a gap in the Information Technology Service Continuity Management (ITSCM) risk management processes contributed to the spread of data corruption and contributed to an eighteen (18) hour delay in return to service (*Disaster Recovery Infrastructure and Processes)*.
- The storage environment has not fully implemented key management tools found in enterprise implementations of this size and scope (*Storage Management, Data Backup and Recovery/Restore Services*).
- The use of a multi-step process to backup and recover Oracle databases, instead of a centralized process, contributed to a perceived loss of data and increased time to restore data (*Storage Management,Data Backup and Recovery/Restore Services*).
- The monitoring system lacked dependency data that would have helped identify the scope of affected systems during the outage (*Incident and Problem Management Services*).
- Recovery testing of backup data is done only twice yearly during the scheduled Disaster Recovery testing. In the opinion of Agilysys this is not a sufficient backup and data restoration test model (*Incident and Problem Management Services*).
- The incident ticket information provided to Agilysys by Northrop Grumman did not contain any data indicating that Northrop Grumman database support personnel had opened any initial incident tickets pertaining to database issues prior to the reporting of the outage by state agency database staff. This suggests that Oracle Enterprise Manager is under-utilized (*Monitoring and Proactive Management*).

In an interview conducted with the EMC engineer who made the decision to replace memory board zero (0) first, it was stated that the decision to replace memory board zero (0) first was based on prior experience. During the initial troubleshooting of reviewing the log files, there were some uncorrectable (hard) errors observed on memory board (1) and correctable (soft) errors on both memory boards zero (0) and one (1). Both memory boards (0) and (1) were showing correctable error counts as being at maximum. As part of standard troubleshooting procedures, the engineer reset the counter to observe the frequency of the errors being generated in real time. During this time no uncorrectable errors were experienced and memory board zero (0) was posting correctable errors faster than memory board one (1). Further status views of the global memory continued to show memory board zero (0) logging correctable errors faster than memory board (1). When asked directly why the engineer determined to replace memory board zero (0) first, which had no uncorrectable errors as opposed to memory board one (1), which did show that uncorrectable errors had been logged at some time in the past, the engineer responded, that due to the rate at which memory board zero (0) was logging errors, prior experience indicated that it was only a matter of time before memory board zero (0) would begin to log uncorrectable errors and that was the deciding factor in his decision to replace memory board zero (0) first.

Based on interviews conducted with EMC and reviews of the data provided to Agilysys by Northrop Grumman and EMC, after the DMX-3 dialed home multiple EMC engineers reviewed the DMX logs and status conditions. EMC determined that errors were being logged on a complimentary pair of memory boards. These boards, memory board one (1) and memory board zero (0), were both exhibiting error conditions. Memory board one (1) had posted hard uncorrectable errors as well as correctable (soft) errors and memory board zero (0) was posting correctable errors. After an additional review by EMC engineering, it was determined that memory board zero (0) should be replaced first. The decision to replace memory board zero (0) before memory board one (1) resulted in

data corruption across multiple critical systems. EMC's own RCA of the incident stated, that "*the initial determination to replace memory board 0 first did not take into account the uncorrectable events that had posted on board 1"* and *"Based on extensive post-incident analysis, EMC has determined that replacing memory board 1 first would have prevented any issues during the replacement activity itself."*

The dual memory board failure of memory board zero (0) and memory board one (1) was attributed to an Electrical Over Stress (EOS) condition at the component level. EMC completed extensive testing of the failed memory boards. The memory boards were then sent to the component vendor for analysis. The component manufacturer concluded that the memory board failure was attributed to an EOS condition experienced by both boards but no indication has been provided regarding when this EOS condition occurred or its cause.

The loss of backup data was attributed to corruption of the Global Catalogs, a primary element of the enterprise backup and recovery system. The Global Catalogs are used to track backup and restore functions within the enterprise backup system and are stored on the DMX-3. The Global Catalog corruption was also caused by human error during replacement of the memory board. This corruption, and the subsequent recovery steps implemented, resulted in data backups not being available from August 25th, 2010 through August 28th, 2010. Although procedures were in place to protect the Global Catalogs by replicating them to SWESC, as well as maintaining tape-based copies, the failure to suspend SRDF before the maintenance event allowed the corrupted data to be replicated to the SWESC location, thus corrupting the disks that contained the copies of the Global Catalog in the SWESC location.

In the professional opinion of Agilysys, a gap in the Information Technology Service Continuity Management (ITSCM) processes contributed to the spread of data corruption and contributed to an eighteen (18) hour delay in return to service. SRDF was not suspended prior to the memory board replacement process, which negatively impacted the data recovery procedures and propagated corruption to the secondary copy of disks in the SWESC disaster recovery center, which contained the enterprise backup Global Catalog. EMC does not have an official best practice regarding whether SRDF or TimeFinder clones/snapshots should be suspended during maintenance. Instead, Northrop Grumman is responsible for managing risk when using the SRDF process and evaluating the impact on the business of suspending or retaining replication during actions that pose a higher risk to the environment. According to the risk management process associated with ITSCM, as the service owner Northrop Grumman should evaluate their processes to avoid "the impact of failure (perceived or actual) through public or commercial embarrassment and financial loss" (Office of Government Commerce published ITIL manual "Continual Service Improvement"). In situations where risk to data consistency will be introduced to an environment, and "Point in Time" copies of the replicated data do not exist in the secondary site to mitigate a corruption event, it is the professional opinion of Agilysys that a top tier provider best practice is to suspend replication. During maintenance events in which a higher risk of data corruption exists than during routine maintenance events, such as a simultaneous error on two memory boards, Agilysys recommends the creation of a procedure to suspend data replication prior to such elevated risk maintenance actions. Related procedures should have been part of the documented maintenance and management procedures created by Northrop Grumman for the DMX-3 when it first entered service.

The storage environment has not fully implemented key management tools that Agilysys would expect to find in enterprise implementations of this size and scope. Key to any operational environment of this size is the implementation and adequate use of proactive monitoring and capacity planning tools to properly plan, monitor, and design capacity upgrades and identify future staffing needs. During data collection and interviews conducted with the Northrop Grumman storage support team, it was observed that EMC's Control Center storage management tool was deployed but in a limited fashion. It is the professional opinion of Agilysys that the management and reporting toolset as deployed for the storage management environment does not follow ITIL-based best practices, as described under the ITIL Service Operation, Operational Health model.

The multi-step process used to backup and recover Oracle databases contributed a perception of data loss and increased time to restore data. Instead of a centralized process, database backups were performed in a three-stage process involving Northrop Grumman and state agency database staff. In the professional opinion of Agilysys this is not a viable backup methodology for use in enterprise class data centers containing mission critical data.

The monitoring system lacked dependency data that would have helped identify the scope of affected systems during the outage. Based on the information provided to, and reviewed by Agilysys, it is the professional opinion of Agilysys that the monitoring system does not provide adequate database and server dependency information and does not represent a best practice implementation of monitoring framework implementations in comparison to top tier service providers with which Agilysys is familiar.

Data restore testing is essential to an understanding of the impacts of recovery efforts on the backup architecture, the time needed to complete recovery, and the coordination of recovery and application teams. Although the data restore testing process was completed twice yearly during scheduled Disaster Recovery testing, and Northrop Grumman exercises restore procedures on an incident basis, the twice yearly data restore testing only accounts for data that belongs to state agencies that subscribe to the Disaster Recovery service. Data belonging to state agencies that do not subscribe to Disaster Recovery services is not tested. It is the professional opinion of Agilysys that data restore testing from backup, testing full data restores should be implemented monthly, using random samplings of data from across the entire enterprise. This would provide accurate restore timelines and verify restore procedures by providing baseline data that could be used to accurately predict restore times and adjust restore documentation if necessary.

There was no data in the incident ticket information provided by Northrop Grumman indicating that Northrop Grumman database support personnel had opened any initial incident tickets pertaining to database issues at the start of the outage. Initial tickets were opened by state agencies. It is the professional opinion of Agilysys that this indicates a lack of proactive database monitoring or misconfigured notification rules within the implementation of Oracle Enterprise Manager that monitors the Oracle database environment.

Introduction

Agilysys provides professional information technology services including, but not limited to, enterprise architecture and high availability, infrastructure optimization, storage and backup management, identity management and business continuity for Fortune 50, 500 and mid-tier customers located in the financial, telecommunication, service provider, health care, education, government and manufacturing sectors. Within these technologies Agilysys maintains a Professional Services organization comprised of individuals that have extensive industry experience and maintain certifications including, but not limited to, Cisco, EMC, HDS, HP, IBM, ITIL, Microsoft, Oracle, Project Management Institute, SNIA, Symantec and VMWare. In this document, Agilysys does not distinguish between state and local government service providers and other industry service providers. In the professional opinion of Agilysys, this is because the best practices for technologies typically deployed in IT service implementations remain the same regardless of sector. Even though the exact implementations may differ, the same best practices are applied.

Review Process

On October 22nd, 2010 VITA contracted with Agilysys Inc. to provide a comprehensive operational audit of system failures that occurred on the afternoon of August 25th 2010 and of Northrop Grumman's response. This failure caused an extended loss of service to 26 of 89 state agencies for more than a week.

This audit was prepared by Agilysys following a three-month review and is based upon data collected during interviews with Northrop Grumman, EMC, several state agencies, and an analysis of backup, storage, server, database and monitoring systems

After completing the research for this audit, Agilysys provided more than one opportunity for staff from the Virginia Information Technologies Agency (VITA), the Joint Legislative Audit and Review Commission, and Northrop Grumman to review the document and provide technical feedback. As part of this process, both VITA and Northrop Grumman were given the opportunity to provide additional documentation and to submit formal, written responses in the form of a letter (Appendix A).

Key Findings

The information included in this audit covers many aspects of the current technical architecture, operational processes, incident responses, data storage, data recovery, disaster recovery and management of the reviewed environment. This audit details several areas where, in the professional opinion of Agilysys, Northrop Grumman failed to meet industry best practices which are commonly used by top tier service providers with which Agilysys is familiar. Solutions to issues that have been observed and reported range from simple to complex and are reviewed in detail later within the document. In other instances, however, Northrop Grumman is using best-of-breed practices that meet or exceed industry best practices.

This document details the findings of the audit. Two themes recurring throughout the audit are that in the professional opinion of Agilysys the largest causes of the delay in data recovery efforts were the failure to stop the data replication mechanism (SRDF) used for disaster recovery, and the failure to use remote or local "Point in Time" clones/snapshots of data on critical application and infrastructure servers connected to the DMX-3 that failed. The use of "Point in Time" clones/snapshots in the remote site should have been an obvious means of protecting against this event or another type of data corruption event that could propagate corrupted data to the remote SWESC recovery site. The frequency with which these concerns appear throughout the document reflects the importance of addressing these issues in such a highly complex and inter-dependent enterprise environment.

It is the professional opinion of Agilysys that steps need to be taken to address these issues to minimize the potential impact of any future outage, and that these steps would result in an across-the-board improvement in Northrop Grumman's operational performance. Agilysys does not have information indicating whether Northrop Grumman detailed this obvious risk to state agencies when state agencies chose specific disaster recovery services.

Throughout the document, the term best practice is used to compare the technologies implemented by Northrop Grumman to industry best practices. This determination has been made by Agilysys based upon its extensive observations of the architectures, processes and methodologies used by top tier industry service providers and enterprise architectures.

Data points referenced throughout the document are from:

- Northrop Grumman, EMC "Root Cause Analysis"
- Symantec, The Alchemy Solutions Group "Business Value Analysis Market Research Report 2009"
- EMC "Managing Storage – Trends, Challenges and Options 2010-2011"
- Computer Economics "IT Staffing Ratios 2010, Benchmarking Metrics and Analysis for 15 Key IT Functions"
- Information Technology Infrastructure Library (ITIL)
- Office of Government Commerce published ITIL manuals "Continual Service Improvement, Service Design, Service Operations, Service Strategy, Service Transition"

Data was gathered by interviewing key technical and managerial staff at Northrop Grumman, Northrop Grumman subcontractors, VITA and state agencies, a review of documentation provided to Agilysys by Northrop Grumman, and the implementation of data collection tools specific to technology implementations. The following items were reviewed under these methods:

- Server Services
- Data Storage
- Backup and Recovery Services
- Disaster Recovery ("DR") /DR declaration strategy
- Network
- Data Center
- Technical Architecture
- Operational Process and Procedure
- Operational Incident Response and Resolution
- Staffing Levels and Skill Sets
- Information Technology Infrastructure Library

In accordance with the Statement of Work (SOW) agreed to by Agilysys the review of the technologies mentioned above was done in concert with a review of the services and service levels defined in the CIA.

The following are key dates and times associated with the failure of the DMX-3. Because the presentation of information in this document reflects the order in which tasks were described in the Statement of Work for this audit, and not necessarily the chronological order of events, Agilysys encourages the reader to refer to this timeline when reading this document.

- August 25th, 10:08 AM – EMC DMX-3 reports errors on memory board 0 and memory board 1.
- August 25th, 1:27 PM – EMC informs Northrop Grumman of intent to replace memory board zero (0).
- August 25th, 2:53 PM – Memory board replacement failure.

- August 25th, 2:55 PM – Database errors observed by state agency application staff.
- August 25th, 3:25 PM – Database errors and server outages reported to Northrop Grumman Help Desk by state agency application staff. Northrop Grumman then informs EMC.
- August 25th, 10:00 PM – Decision made by EMC and Northrop Grumman to pursue online data recovery procedures that do not take DMX-3 offline.
- August 26th, 12:01 AM – Online recovery of the DMX-3 begins.
- August 26th, 7:30 PM – DMX-3 shutdown process begins (offline recovery efforts).
- August 27th 12:35 AM – DMX-3 is returned to service by EMC.
- August 27th, 7:30 AM – Rebuild of the enterprise backup/recovery system primary server begins.
- August 27th, 6:30 PM – Enterprise backup and recovery system is returned to service by Northrop Grumman and data recovery/restoration begins.
- August 29th, 7:41 AM – Completed reading backup tapes created after 08/23/2010, backup catalog was closed and all backup catalog metadata restored
- September 15th, 5:56 AM - Incident ticket was closed with all data restoration efforts complete and ninety-nine percent (99.9%) of all data recovered. Remaining 0.1% (Department of Motor Vehicles' DLCI data) sent to external provider for recovery.
- October 22nd, 9:00 PM – Recovery of corrupted data complete in all agencies.

Detailed Review

The following sections detail the audit of the root cause analysis and predictive analysis, disaster recovery architecture, data backup/restore services, predictive analysis methods, incident management services and data center management. At the conclusion of each section, recommendations are listed to provide process or technology improvements if it was determined that a section warranted improvement.

Root Cause Analysis

Root Cause Analysis

It is the professional opinion of Agilysys that Northrop Grumman's Root Cause Analysis did not meet contractual requirements and also did not reflect best practices because it did not detail the root cause of the memory board failure within the RCA and lacked any information regarding corrective actions, communications, incident management, or processes associated with the failure of the DMX-3, and related recovery efforts.

Agilysys received and reviewed the Root Cause Analysis provided by Northrop Grumman. Agilysys reviewed section 3.13.1 of the CIA which requires Northrop Grumman to describe *"in detail the cause of, and procedure for correcting, such failure and providing the Commonwealth with reasonable evidence that such failure will not recur,"* and Appendix 1 to Schedule 3.3 *"Root Cause Analysis Services are the activities required to develop, implement, and maintain a Root Cause Analysis ("RCA") process and perform the activities required to diagnose, analyze, recommend, and take corrective measures to prevent recurring Problems and/or trends."* Based on the information provided within the Northrop Grumman RCA, it is the professional opinion of Agilysys that Northrop Grumman failed to meet the contractual requirements set forth in Section 3.13.1 governing RCA data requirements. The following section elaborates on why, in the opinion of Agilysys, Northrop Grumman did not meet those requirements.

Findings

In an interview conducted with the EMC engineer who made the decision to replace memory board zero (0) first it was stated that the decision to replace memory board zero (0) first was based on prior experience. During the initial troubleshooting of reviewing the log files, there were some hard errors observed on memory board (1) and soft errors on both memory boards zero (0) and one (1). Both memory boards (0) and (1) were showing soft error counts as being at maximum. As part of standard troubleshooting procedures, the engineer reset the counter to observe the frequency of the errors being generated in real time. During this time no hard errors were experienced and memory board zero (0) was posting soft errors faster than memory board one (1). Further status views of the global memory continued to show memory board zero (0) logging soft errors faster than memory board (1). When asked directly why the engineer determined to replace memory board zero (0) first, which had no hard errors as opposed to memory board one (1), which did show that hard errors had been logged at some time in the past, the engineer responded, that due to the rate at which memory board zero (0) was logging errors, prior experience indicated that it was only a matter of time before memory board zero (0) would begin to log hard errors and that was the deciding factor in his decision to replace memory board zero (0) first.

Based on interviews conducted with EMC and reviews of the data provided to Agilysys by Northrop Grumman and EMC, after the DMX-3 dialed home multiple EMC engineers reviewed the DMX logs and status conditions. EMC determined that errors were being logged on a complimentary pair of memory boards. These boards, memory board one (1) and memory board zero (0), were both exhibiting error conditions. Memory board one (1) was posting hard uncorrectable errors and memory board zero (0) was posting soft errors. After an additional review

by EMC engineering, it was determined that memory board zero (0) should be replaced first. The decision to replace memory board zero (0) before memory board one (1) resulted in data corruption across multiple critical systems. Based on the data provided and a review of the situation, Agilysys cannot conclude if the decision to replace memory board one (1), instead of memory board zero (0), should have been apparent to EMC engineering. EMC's own RCA of the incident stated, that "*the initial determination to replace memory board 0 first did not take into account the uncorrectable events that had posted on board 1"* and *"Based on extensive post-incident analysis, EMC has determined that replacing memory board 1 first would have prevented any issues during the replacement activity itself."*

The Northrop Grumman RCA Failed to Meet Three Contractual Requirements. Northrop Grumman did not provide the RCA within the contractually-required ten day period. In addition, the RCA did not provide reasonable evidence that the failure would not recur. Lastly, the information detailing the actual root cause of the failure of the memory boards was not available in the original RCA, but was provided as a separate document upon Agilysys's request during the course of the audit. The failed memory boards were tested by EMC in three different testing scenarios. The scenarios subjected the memory boards to varying degrees of high and low stress in an operational environment, to a manufacturing environmental stress test, and to testing in an environment mirroring the configuration of the DMX-3 located in the CESC data center. In all of the tests, Hard Single Bit errors were encountered on memory board one (1) and Soft Single Bit errors were encountered on memory board zero (0). After EMC completed its testing, the memory boards were sent to the chip manufacturer for component removal and analysis. According to relevant documents reviewed by Agilysys, the chip manufacturer concluded that the failure of the memory board components on memory board one (1) and zero (0) was caused by DC degradation due to an Electrical Over Stress ("EOS") condition. EOS is defined as an electrical stimulus (event) outside the operational range of a semiconductor component. However, in documents provided to Agilysys by Northrop Grumman, there was no data available to pinpoint the time that the EOS condition occurred and the relevant documents did not elaborate on the type of EOS that caused the semiconductor failure on the memory boards.

It Is the Professional Opinion of Agilysys that Best Practices Regarding RCAs Were Also Not Met. It was observed by Agilysys that the RCA lacked any information regarding communications, incident management, or processes associated with the failure of the DMX-3, and related recovery efforts. It is the professional opinion of Agilysys that a complete RCA would include data regarding improvements to the above mentioned items, and a continual review of these items should be part of an overall risk management process. In contrast, Northrop Grumman has utilized these review procedures in their remediation of disaster recovery testing issues, as indicated in documentation provided to Agilysys, that outlined the results of the last disaster recover test. Northrop Grumman's review of these procedures to address shortcomings found during disaster recovery testing raises the question as to why the review of these types of procedures was not included in an RCA of a critical event.

Communications to state agencies regarding data availability was poor but the RCA does not indicate how this should be addressed. During the restoration process, the status of "orphaned" data was not effectively communicated. Orphaned backup tapes began to be imported back into the enterprise backup and recovery system once it was returned to service on August 27th, 2010 at 6:30PM. Based on information provided in interviews with Northrop Grumman backup staff, orphaned tapes were imported from August 27th, 2010 through August 29th, 2010. As importing of the orphaned tapes began, it was known by Northrop Grumman that until all orphaned tapes had been read back into the enterprise backup system, the first available date from which data could be restored was August 23rd, 2010. During interviews with Northrop Grumman backup support personnel, Agilysys was informed, that this information was relayed to the Incident Management team when the enterprise backup system was returned to service and began restore activities on August 27th, 2010 at 6:30PM. Subsequent interviews with state agencies indicate they did not receive this information in a formal fashion.

Data was lost due to corruption but the RCA does not indicate how this should be prevented. The corruption incurred by the DMX-3 failure affected multiple servers attached to the DMX-3 array. The last consistent backup

of data that was available from the enterprise backup and recovery system, prior to the recovery of the orphaned backup tapes, was from August 23rd, 2010. Once all orphaned tapes were imported back into the enterprise backup and recovery system (by August 29th, 2010), the most recent consistent backup for systems affected by the DMX-3 failure was August 24th 2010. Any data recorded to disks after completion of the August 24th, 2010 backup, and which was affected by data corruption due to the DMX-3 failure was lost. At present, "Point in Time" clone/snapshot implementations are not deployed as part of the overall IT Services Continuity Management model, except in the Mainframe environment. It is the professional opinion of Agilysys that, the lack of a "Point in Time" clone/snapshot implementation on the DMX-3 that suffered the failure facilitated an environment in which greater data loss could occur in critical application and infrastructure servers.

Recommendation

1. Review RCA procedures to ensure pertinent information, not available at time the RCA is submitted, is subsequently provided in a timely manner and without the need for an additional request by the Commonwealth or its auditors.

IT Service Continuity Management and Disaster Recovery

IT Service Continuity Management

It is the professional opinion of Agilysys that Northrop Grumman's IT Service Continuity Management was not best practice because the risk management process did not evaluate processes to avoid "the impact of failure (perceived or actual) through public or commercial embarrassment and financial loss" (Office of Government Commerce published ITIL manual "Continual Service Improvement").

Findings

IT Service Continuity Management (ITSCM) addresses risk that could cause a sudden and serious impact that would immediately threaten the continuity of the business. Threats typically addressed, but not limited to, by ITSCM include loss, damage or denial to key infrastructure services, non-performance by critical providers, and loss or corruption of key information. IT Service Continuity Management services are dictated by what the customer, as the consumer of the services, has defined as their requirements in the service level agreement (SLA) executed with the provider.

It is the professional opinion of Agilysys that there has been a departure from industry best practices in the ITSCM model as implemented by Northrop Grumman in the primary and recovery data centers. Specifically, the procedures used to manage use of the data replication mechanism (SRDF) do not account for a data corruption event when remote point in time copies do not exist. Moreover, based on implementations by the industry and top tier service providers with which Agilysys is familiar, the implementation of remote "Point in Time" clones/copies is also key to successful data recovery operations during a disaster or during a service disruption (non-disaster) event. It is the professional opinion of Agilysys that the risk of data corruption posed by this departure should have been obvious to Northrop Grumman before the DMX-3 outage. In addition, as the DMX-3 began experiencing an increasing rate of errors on both redundant memory boards, which indicated a need to replace both boards, it should have been obvious to Northrop Grumman that SRDF should have been stopped prior to the memory board replacement procedure because of the unusual nature of the problem and increased risk to the environment.

If corruption of data is incurred at a primary site location (CESC) and replicated to the recovery site (SWESC), prior to the shutdown of the replication service (SRDF), then the data in the recovery site is corrupt. In that event, the use of "Point in Time" clones/copies for critical applications and infrastructure service servers would provide a picture of the data as it was at a point in time before corruption. Implementation of "Point in Time" clones/copies in a schedule that satisfies the business needs of the customers would provide a consistent, recoverable copy of data for service restoration in the event of data corruption, either in the CESC or SWESC locations. It is the professional opinion of Agilysys that Northrop Grumman did not meet ITSCM best practices by failing to address data corruption threats to IT Service Continuity, not evaluating processes to avoid "the impact of failure (perceived or actual) through public or commercial embarrassment and financial loss" (Office of Government Commerce published ITIL manual "Continual Service Improvement") during the risk management process, and that the implementation of processes such as remote "Point in Time" clones/snapshots to protect against data corruption events should have been obvious during previous normal operations. As mentioned earlier Agilysys does not have information indicating whether Northrop Grumman communicated the obvious risk of the possibility of compromising the remote copy of data, inherent with data replication without "Point in Time" clones/snapshots to state agencies, when state agencies selected specific disaster recovery services. It is also the opinion of Agilysys that Northrop Grumman was aware of this risk to the environment due to the implementation of "Point in Time" technology deployed in the mainframe environment to mitigate this same risk.

Recommendations

2. Implement "Point in Time" clones/copies for critical infrastructure and application servers, where appropriate.
3. In the absence of "Point in Time" copies, develop a process to suspend SRDF operations during times of increased risk to data integrity in the primary site.

Disaster Recovery Services

It is the professional opinion of Agilysys that Northrop Grumman's Disaster Recovery Services met best practices because all items that were expected in the design of a Disaster Recovery program as in comparison to industry top tier providers and enterprise Disaster Recovery Programs were present.

Findings

Disaster Recovery Services are a key component of ITSCM processes. Northrop Grumman provides comprehensive disaster recovery services for the Commonwealth as part of the CIA. Based on documentation provided to Agilysys by Northrop Grumman, and data collected during interviews with Disaster Recovery Program staff, it was noted that Northrop Grumman maintained detailed documentation, processes, test procedures, remediation action plans, and an appropriate approach to program management pertaining to the Disaster Recovery implementation as per industry best practices.

The Disaster Recovery Service implemented as per the CIA is an "all or nothing" approach, and Agilysys agrees that its use would not have been appropriate in this situation. As required by the CIA, under a Disaster Recovery scenario all network, server, storage, monitoring and other IT service related services must transfer (failover) to the SWESC data center together. No single service or agency application can failover without all others. The option to failover to the SWESC location was presented by Northrop Grumman to VITA's Chief Information Officer during the initial recovery efforts but that option was not exercised because it would have increased the time required to recover operations by virtue of the requirement that all systems, even normally-functioning systems, failover to SWESC.

Agilysys has reviewed the disaster recovery procedures and agrees with the decision to not declare a disaster because recovery from the SWESC data center following a failover of all systems would have increased the time required to resume normal operations because failure to halt the data replication process (SRDF) resulted in the replication of corrupt data to SWESC. This raises a concern about the reliability of disaster recovery services in this or other scenarios where SRDF is not suspended prior to failover, because in this kind of situation the effects of data corruption, in the absence of "Point in Time" copies, would not have allowed Northrop Grumman to fully rely upon the data at the SWESC location.

Storage Management, Data Backup and Recovery/Restore Services

Storage Management and Backup Services

It is the professional opinion of Agilysys that Northrop Grumman's Storage Management and Backup Services was not best practice because there is no documented or practiced process to suspend SRDF in the absence of "Point in Time" copies in the primary or secondary site, a lack of project management processes, and the use of a three stage backup process for Oracle databases instead of a centralized approach.

Findings

The following section details the cause for extended restore times during the August 25th, 2010 outage, the report of corrupt backup tapes, placement of backup data and how it relates to extended restore times, and a review of processes, procedures, architecture and people as it pertains to backup/restore services and the preservation of Commonwealth data. Key findings pointed out in the following section detail the lack of implemented functionality in the DMX-3 supporting open systems such as "Point in Time" copies (like those already in use on the mainframe) and the lack of procedures to prevent data corruption that spread to the disaster recovery site.

Failure to Suspend Data Replication and Lack of Point-in-Time Backup Copies in The Remote SWESC Location Increased Restoration Time. On August 25, 2010, the SRDF process was replicating data from CESC to SWESC as the memory boards were being replaced on the DMX-3. The EMC DMX-3 is designed to have maintenance performed during normal business hours, as each component is redundant. However, two (2) redundant memory boards were streaming errors in a very rare condition and subsequent human error during the DMX-3 maintenance procedure (previously described in the RCA review section of this document) resulted in the corruption of state agency data. In addition to corruption to state agency data, corruption occurred on disks containing the Global Catalog which tracks and indexes all data backed-up on the Northrop Grumman enterprise backup and recovery system.

Due to the effects of data corruption, once EMC finished the repair of the DMX-3 and handed it back to Northrop Grumman for use, but before Northrop Grumman could begin to restore state agency data, Northrop Grumman had to restore the Global Catalog to insure the integrity of enterprise backup and recovery system. This accounts for an initial delay of eighteen (18) hours before agency data could begin to be restored (between just after midnight on August 27th, 2010 when EMC declared the DMX-3 repaired and 6:30 PM August 27th, 2010). As mentioned previously in this document, it is the professional opinion of Agilysys that Northrop Grumman should have requested that in the absence of remote "Point in Time" copies, and the increased risk to the environment during a dual memory board error condition, that SRDF be suspended during this maintenance to preserve the data at the disaster recovery site, and it is also the professional opinion of Agilysys that the failure to suspend this process failed to meet best practice for data replication methods implemented at top tier service providers with which Agilysys is familiar.

It is the professional opinion of Agilysys that if SRDF replication had been suspended prior to memory board replacement, then once the problem in the EMC DMX-3 was fixed Northrop Grumman could have incrementally restored the disks containing the Global Catalog and the disks supporting applications that subscribed to the tier one data replication service on the DMX-3 from the secondary copy of data, thereby reducing the time to recover the data that had been identified as corrupted and minimizing data loss to minutes prior to the corruption event. Furthermore, it is the professional opinion of Agilysys that Northrop Grumman failed to meet their obligation to provide backup services across all service environments, in accordance with Addendum 1 Appendix 1 to Schedule 3.3 of the CIA in which Northrop Grumman agrees to "*Reduce backup windows through the use of*

TimeFinder and SnapView software, which support near instant creation of point-in-time disk copies that can be used to create back-up tapes for local data recovery and support ***quick resumption of production processing*** (emphasis added)." In the local site "Point in Time" clones/snapshots could have been used to quickly restore systems affected by the data corruption to the time of the last snap, quickly resuming production services once the DMX-3 was operational. This technology was not in use on August 25th, 2010 on the DMX-3 in the SWESC location that was hosting the secondary replicated copies of data or on the DMX-3 in CESC that suffered the failure. Northrop Grumman states that, in its opinion, it is not obligated to use local or remote TimeFinder clone/copies for any systems other than the mainframe system as part of its responsibilities under the Comprehensive Infrastructure Agreement.

Lack of Updated Information on Software Versions Hinders Planning and Does Not Reflect Accepted Project Management Practices. The RCA provided by Northrop Grumman stated that the presence of an older version of enterprise backup software on the dedicated media server supporting the Driver's License Central Issuance (DLCI) application for the Department of Motor Vehicles (DMV), attributed to recovery errors and a delay in recovery efforts. The older software was upgraded on August 29th, 2010 and this resolved restore issues on the client. During interviews with Northrop Grumman backup support personnel it was observed by Agilysys that no formal upgrade schedule had been in place for upgrading the enterprise backup environment prior to this audit and the failure of the DMX-3. Without a formal schedule to follow, Agilysys was unable to evaluate if the server supporting the DLCI application had been scheduled for upgrade to the latest version prior to the failure. It is the professional opinion of Agilysys that the lack of a formal schedule reflects a failure to implement project management best practices for a project of this size.

Use of Multi-Step, Decentralized Process for Oracle Database Backups Increased Time to Restore Data for Six (6) of Twenty-Nine (29) Databases That Were Restored From Tape. In the current environment, which was in place during the outage, database backups are performed via a three-stage process that involves two groups of Northrop Grumman personnel and state agency staff. It is important to note that the process for backing-up a database involves two components: a database and the associated log files (which are generated after a full backup is complete). To explain, the first step in the Northrop Grumman backup model for Oracle databases involves the backup of the database. This process is configured and automated by Northrop Grumman database support and the database is written to a predetermined area of disk on the Tier 1 DMX-3 (the equipment that suffered the outage). Second, Northrop Grumman backup support then backs-up the database during the next scheduled backup window, using the enterprise backup application to save the database to tape. Third, the backup of the database log files is scheduled by the state agency application owners, the logs are written to disk, and then are backed up in the same fashion as the database files.

The process for recovering Oracle databases depends upon the availability of both the database and the log files. When a database is restored, each log file must be restored in sequence to recover to the desired time or data point. Based on interviews with Northrop Grumman personnel, the scheduling of the scripted Oracle backup of log files varies from agency to agency and it appears that the variation among agencies accounted for some missing data. It is important to understand that a database requires two components to complete a restore: a recent full backup and the associated log files (which are generated after the full backup was complete). With that stated, in some cases agencies had decided to retain log files on some servers for only two (2) days. As a result, some data could not be restored in a timely manner. For example, If an agency requests that a database be restored on Thursday, but the last full backup had been done on Sunday, then the logs from Monday would be missing. This would require that the missing data be retrieved from tape, thus increasing the time required to restore the data. Furthermore, in all instances it does not appear that Northrop Grumman backup personnel had daily insight into whether the Oracle RMAN backups to disk completed successfully, nor did the state agency personnel have daily insight into whether the enterprise backup and recovery system successfully backed up the disk-based RMAN backups and log files.

In the opinion of Agilysys, this multi-step, decentralized process resulted in a convoluted and lengthened recovery process, is not viable in enterprise class data centers containing mission critical data, and does not reflect a best practice implementation of top tier service providers with which Agilysys is familiar. Instead, Agilysys has observed that top tier service providers use centralized enterprise backup agents, which provide the most streamlined approach and is the preferred implementation of Oracle database backups in enterprises of this size.

Adequacy of Staffing Resources Should Be Reviewed. In initial documentation provided to Agilysys by Northrop Grumman it was stated that Northrop Grumman maintained three (3) dedicated personnel that managed the entire enterprise backup and recovery system at CESC, and Avamar remote backup system for the CESC location. In updated documentation and clarification of roles and responsibilities received by Agilysys, Northrop Grumman outlines a total of nine (9) staffing resources. Two of these resources are dedicated to roles at two state agencies and seven (7) resources manage the enterprise backup environment and the Avamar backup environment located at CESC as well as share responsibilities for legacy backup applications still existing at state agencies. The enterprise backup environment back up capacity per day at CESC is currently at approximately thirty five (35) terabytes. The Avamar environment consists of eight hundred and three (803) clients and is currently consuming sixty (60) terabytes. The Alchemy Solutions Group "Business Value Analysis Market Research Report 2009" indicates that one backup administrator can handle up to 39 TB of backup. However the study does not show the complexity of the environment being studied but rather reflects the average of data being backed up by the respondents to the survey. Based on the current supported backup architectures and ongoing state agency transformation activities, Agilysys recommends that Northrop Grumman evaluate the current work load associated with the backup staff personnel to effectively determine if the current staffing model is efficient.

Enacted Service Level Agreement Governing Data Restoration Does Not Adhere to Industry Best Practices. The agreed to SLA 1.21 "Cross Functional Restore Services Restore Requests for production data in CESC/SWESC" in the CIA states that data restoration activities will start within four (4) hours of a request. This apparently was not the case during the outage due to the unavailability of the backup system. In addition, the contract documentation reviewed by Agilysys does not mention a deadline to complete full recovery, which is not an industry best practice Service Level Agreement (SLA) implementation. Instead, it is an industry best practice to include a recovery point objective and recovery time objective as part of almost every data recovery and disaster recovery SLA. Recovery times should be agreed upon by the customer, and agreed to and validated by the service provider, based on restore testing and the restore technologies implemented in the environment.

Recommendations

4. As part of an overall review of backup/restore policies and procedures, implement full backup restore testing procedures that take into account random samplings of data and simultaneous restores on a monthly basis. Backup staff, state agency application personnel, database support personnel and server support personnel should be engaged during recovery testing.
5. As part of improving the efficiency of the backup/recovery methods, deploy the enterprise backup and recovery system agent for Oracle databases as a best practice. This action will significantly simplify, centralize and streamline data backup/recovery operations, thus removing the identified points for error in the current process.
6. As part of an overall process to review internal project management, implement schedule tracking and implementation of project control processes for infrastructure projects.
7. As part of ongoing backup staff evaluations and review of the current and projected backup requirements, evaluate the current project load and day to day responsibilities of the backup staff assigned to the enterprise backup and Avamar systems to evaluate staffing levels.

Storage Architecture and Capabilities

It is the professional opinion of Agilysys that Northrop Grumman's Storage Architecture and Capabilities was not best practice because of the limited implementation of performance, predictive analysis, proactive monitoring and management software for the DMX-3. These capabilities are required in Appendix 1 to Schedule 3.3 of the CIA which requires Northrop Grumman to "*Continually monitor IT resource usage to enable proactive identification of capacity and performance issues."*

Findings

The DMX-3 Storage and Available Features Are Appropriate for Use at CESC. Agilysys reviewed the Northrop Grumman storage environment and associated data and interviewed storage support staff, focusing on the DMX-3. The DMX-3 storage array is considered to be "best of breed" in enterprise class storage. Agilysys has observed this storage deployed by numerous top tier service providers and the use of these types of storage devices is appropriate for use at tier three (3) datacenters like CESC.

The EMC DMX-3 has enterprise class features such as remote replication and "point in time" copies. Symmetrix Remote Data Facility ("SRDF") is used to replicate data from a local storage array to a remote storage array. In the event of a disaster in the local site, the remote site data can be brought online and used. EMC TimeFinder "Point in Time" clone/snapshot allows users to create point in time copies of data volumes within a single EMC array. Deploying TimeFinder volumes in the local array can vastly improve the local recovery time of critical data volumes in the event of corruption. When TimeFinder is coupled with SRDF, an additional level of protection is introduced, because the remote data site (SWESC) then has a "Point in Time" copy in the event data corruption is propagated to the remote site. Local TimeFinder volumes and remote "golden copies" could also be leveraged for backend processing such as backup jobs. The use of "golden copies" on the DMX-3 at the disaster recovery site would serve as a point in time copy in the event of data corruption at the primary site, as was experienced on August 25th, 2010.

Proactive Monitoring Does Not Meet Best Practices. Key to any operational environment of this size is the implementation of proactive monitoring and capacity planning tools to properly plan, monitor, and design capacity upgrades and future staffing needs. During data collection, and interviews conducted with the storage support team, it was observed that Northrop Grumman deployed EMC's Ionix Control Center storage management tool in a limited fashion. The environment has been in place for more than four (4) years, and by this time one would expect to see greater use and implementation of an intelligent toolset in an environment of this size and complexity. It is the professional opinion of Agilysys that the limited implementation of the tools in the storage management environment have had a negative impact on day-to-day performance monitoring, root cause analysis, operations, operational response, capacity planning and incident management of the storage system. It is also the opinion of Agilysys that the absence of even the basic functions of EMC's Control Center results in a non-compliance with industry best practice of implementation of ITIL under the ITIL Service Operation, Operational Health model.

Recommendations

8. As part of the ongoing effort to deploy EMC Control Center management and monitoring functions, deploy EMC Ionix Performance Manager to provide error correlation, performance monitoring and root cause analysis capabilities in the storage environment.
9. As part of an overall review of business continuance services and processes, implement EMC TimeFinder "Point in Time" clones/snapshots to mitigate the adverse effect on critical infrastructure systems. Develop

a clone/snapshot synchronization schedule based on a Business impact analysis of the enterprise backup and recovery system.

10. As part of an overall review of business continuance processes implement a process to suspend SRDF prior to maintenance events that could have an adverse effect on data consistency. The process needs to identify the most probable scenarios where maintenance or unplanned component failure could introduce data corruption, while evaluating the impact of suspending data replication services during the maintenance or outage window.

EMC DMX-3 Predictive Analysis

It is the professional opinion of Agilysys that EMC's DMX-3 Predictive Analysis was above best practice because of the design of the monitoring and alerting functions of the "Call Home" feature of the DMX-3 alerting EMC to events on the DMX-3. Even though errors occurred, the Predictive Analysis features worked as designed by alerting EMC to the need for corrective action.

Findings

The DMX-3 is a best of breed enterprise storage array, deployed in some of the most demanding storage environments throughout enterprise class data centers. The EMC DMX-3 dialed home to EMC as soon as the errors were detected within the array. This dial home feature is a feature of best of breed hardware and worked as implemented, utilizing error reporting and predictive analysis features of the array architecture.

Based on the collection of data from the DMX-3, interviews conducted with Northrop Grumman storage support staff and EMC support personnel, Agilysys has concluded that the maintenance of the DMX-3 was consistent with best practices, as observed in other similar enterprise top tier service providers.

Backup Architecture and Capabilities

It is the professional opinion of Agilysys that Northrop Grumman's Backup Architecture and Capabilities were within best practice with some minor exceptions because the architecture scales to meet the needs of the environment, employs technologies to stream line and accelerate backups and is in line with top tier provider and enterprise backup implementations.

Findings

Enterprise Backup and Recovery System Is Generally In Line with Top Tier Data Center Providers But Not Uniformly Implemented. Northrop Grumman has implemented an enterprise backup and recovery system addressing the backup and recovery needs of the data center. The enterprise backup and recovery system is comprised of local backup services, remote backup services and de-duplication services. The Northrop Grumman enterprise backup and recovery system implementation is in line with top tier data center providers. The design is robust with multiple servers that can simultaneously write data directly to a bank of tape drives. All daily backups were performed using a mix of disk staging, direct write to tape and direct write to tape with multiple streams to a single tape drive. All network backups were performed over a private network on a staggered basis in order to provide adequate throughput and a balance across the network. Standard network backups had multiplexed depths of no more than four (4) streams per tape drive, where applicable. Servers with more than 500GB per backup were designated as SAN media servers with dedicated tape drive units to accommodate the required bandwidth to complete backups within the specified backup windows. In the opinion of Agilysys, this was a good practice. However, it was observed that this practice is not strictly adhered to or practiced uniformly throughout the enterprise. But overall, it is the professional opinion of Agilysys that based on the data provided by Northrop

Grumman, and in comparison to industry best practices, that this system is adequate to handle the size and data capacity of the existing data center.

Recommendations

11. Review process to change a client to a SAN Media Server and standardize implementation.

Incident Management and Recovery Services

Recovery from DMX-3 Failure

It is the professional opinion of Agilysys that Northrop Grumman's Recovery from DMX-3 Failure was not best practice because as mentioned previously in this document there was no process in place to suspend SRDF in the absence of remote "Point in Time" copies of the replicated data.

Findings

The EMC DMX-3 dialed home as soon as the errors were incurred within the array and continued to operate. Within four hours after the DMX-3 initiated the first call home, EMC dispatched an on-site engineer to replace the failing global memory boards on the affected array. At 1:27 PM on August 25th, 2010, Northrop Grumman was informed of the failure and EMC's intent to replace the memory boards. It is the professional opinion of Agilysys that the decision to replace the memory boards during the production day is within best practices implemented by top tier service providers. When errors indicate a need to replace a failing hardware component, the component should be replaced, to avoid the risk of compromising the hardware.

During the memory board replacement, uncorrectable errors were incurred resulting in the corruption of data. Approximately forty (40) minutes after the memory board replacement failure, Northrop Grumman reported server errors to EMC and engaged EMC engineering. Based on information that the majority of systems that were connected to the affected DMX-3 were not incurring error conditions or data corruption, a decision was made at 10:00 PM on August 25th, 2010 by EMC and Northrop Grumman to pursue online data recovery procedures. The desired effect of this decision was to minimize the impact to the overall environment by not incurring an entire array outage if it could be avoided. A process was run at this point to return the DMX-3 to a stable state and recover the data corrupted during the memory board replacement. The data recovery process was partially successful in reducing the amount of corruption and a decision was made to pursue offline recovery procedures, which required a shutdown of the DMX-3. Based on information provided to Agilysys by Northrop Grumman on agency-reported errors incurred on servers connected to the DMX-3, Agilysys agrees with the logical steps taken to try to recover the most amounts of data, with the least amount of impact to the larger, perceived functioning server environment connected to the DMX-3.

However, as Mentioned Previously in This Document the SRDF Data Replication Process Was Not Suspended Prior to the Memory Board Replacement Process, Which Negatively Impacted the Data Recovery Procedures. In situations where risk to data consistency will be introduced to an environment, there is an absence of remote "Point in Time" copies of the replicated data, in the opinion of Agilysys it is a top tier provider best practice to suspend replication. Northrop Grumman and not EMC must decide when to suspend SRDF and clones/snapshots since Northrop Grumman, and not EMC, has detailed knowledge of the resulting impact to the business environment and impact of data change rates on the network links to the remote recovery site when resuming replication, when SRDF is suspended. It is also the opinion of Agilysys that the practice of suspending disaster replication in the absence of remote "Point in Time" copies of the replicated data, prior to select maintenance events that incur a greater risk to data integrity and business operations, is a basic procedure and should have been part of the maintenance and management procedures created for the DMX-3 upon service entry of SRDF.

Recommendations

12. As part of the published maintenance, management and procedures governing the DMX-3, ensure processes are implemented to suspend SRDF prior to maintenance events that could have an adverse effect on data consistency. The process needs to identify the most probable scenarios where maintenance or unplanned component failure could introduce data corruption, while evaluating the business impact of suspending data replication services during the maintenance or outage window. (See item 3.)

Recovery from Data Corruption

It is the professional opinion of Agilysys that Northrop Grumman's Recovery from Data Corruption did not meet best practices because remote "Point in Time" copies were not in use and a component of the monitoring system does not report on critical dependency data.

Findings

Data Restoration Time Would Have Been Reduced if Point-in-Time Copies Had Been Used. The enterprise backup and recovery system data resides on a tier one disk in a highly available environment that is replicated to the SWESC location for disaster recovery. If the backup system data residing on the DMX-3 had been using the remote or local "Point in Time" snapshot technology available on the DMX-3, then it is the professional opinion of Agilysys that the disks containing the global catalogs responsible for tracking all backup data could have been incrementally restored from the most recent snapshot/clone upon availability of the array, thus significantly shortening the recovery time of the enterprise backup system. Agilysys cannot provide an exact recovery time since this would have to come from a tested procedure within a lab environment with recorded data, and this data was not collected during data collection. Common practice by top tier providers with which Agilysys is familiar integrates "Point in Time" copies for infrastructure critical and business critical applications in enterprise environments of this size, due to the greater impact of data corruption and outage events in the environment. It is the professional opinion of Agilysys that Northrop Grumman failed to implement processes such as "Point in Time" clones/snapshots to protect the enterprise backup and recovery system and mitigate the effects of data corruption scenarios. In addition, it is the professional opinion of Agilysys that Northrop Grumman did not meet the standards set forth in the CIA to use *"industry best practices and methods to avoid, prevent and mitigate any material adverse effect on the Systems or the continuity and quality of the Services being provided to the Commonwealth"* (§3.1.2 of the CIA) and Addendum to Appendix 1 of Schedule 3.3 of the CIA in which Northrop Grumman agrees to *"Reduce back-up windows through the use of TimeFinder and SnapView software, which support near instant creation of point-in-time disk copies that can be used to create back-up tapes for local data recovery and support quick resumption of production processing."* Northrop Grumman states that, in its opinion, it is not obligated to use local or remote TimeFinder clone/copies for any systems other than the mainframe system as part of its responsibilities under the Comprehensive Infrastructure Agreement.

The Scope of the Impact to Systems Would Have Been More Readily Apparent if the Configuration Management Database Maintained Dependency Information. The current implementation of monitoring and incident reporting lacks key error correlation information about application and hardware inter-dependencies. It is the professional opinion of Agilysys that the lack of dependency information hindered Northrop Grumman's ability to properly assess the impact of data corruption on the environment.

Northrop Grumman uses three primary tools to monitor databases and hardware. The HP OpenView monitoring system deployed by Northrop Grumman reports on the basic performance metrics of system availability, processor utilization, memory utilization, and file system utilization for server systems. However, the

implementation of HP OpenView does not report on database level errors generated by the two major deployed databases, Microsoft SQL and Oracle. Specifically, the application agents are deployed but the alerts are suppressed at the HP OpenView consoles. Oracle implementations are managed and monitored by Oracle Enterprise Manager and SQL implementations are managed and monitored through decentralized management tools.

The current implementation of monitoring and incident reporting lacks key error correlation information about application and hardware inter-dependencies. Instead, the error correlation maintained in the configuration management database merely goes one (1) level deep, to hardware only, and does not provide enough information to properly assess the impact of a hardware or network outage. The lack of information on database and hardware inter-dependencies appears to have hindered Northrop Grumman's ability to assess the impact that the DMX-3 related errors had on other dependent environments. To account for these inter-dependencies, top tier service provider monitoring architectures take into account dependencies and error correlation. Based on the information provided and reviewed by Agilysys, it is the professional opinion of Agilysys that the configuration management database does not provide adequate application and server dependency information in comparison to monitoring frameworks at top tier service providers with which Agilysys is familiar.

Recommendations

13. In accordance with Addendum 1 to Appendix 1 of Schedule 3.3, implement EMC TimeFinder clone/snapshot technology to protect critical systems that the overall architecture is highly reliant upon, such as the enterprise backup and recovery system primary servers. Implementing "Point in Time" snapshot technology decreases the data recovery to minutes from hours thus allowing for a streamlined return to operational readiness. (See item 2.)
14. Update Configuration Item ("CI") relationships to reflect application and server dependencies in the configuration management database. Defining application dependencies creates an environment that provides more data at the onset of an outage as to what systems are directly affected.

Staffing

Agilysys cannot conclude if Northrop Grumman's staffing for the server environment is sufficient, but finds that the number of servers per administrator in Northrop Grumman's environment is higher than the industry average. The server support organization is spread across multiple support groups servicing agencies located in the capital region, remote agencies, the CESC primary data center and the SWESC secondary data center. Clarification of server responsibilities across the organizational groups for server support personnel was not clear enough for Agilysys to reach a conclusion.

Findings

Based on data from "IT Staffing Ratios 2010, Benchmarking Metrics and Analysis for 15 Key IT Functions" by Computer Economics, the database, storage, server and backup staff ratios fall within industry staffing norms. Based on organizational charts and relevant information provided by Northrop Grumman, the central server support groups located at CESC and the remote server support groups are comprised of a total of eighty one (81) server support personnel. In some remote locations, members of End User Services provide support for server services, utilizing 12 full time resources per year. In its decentralized administration model, Northrop Grumman states that all eighty one (81) resources can remotely manage the entire enterprise server environment in addition to their primary roles as server support for specific areas such as the capital region, CESC, SWESC and other remote agencies.

Taking these remote resources into account, the server per administrator average in the current operating environment is 55 servers per administrator. (This average is based on 4443 servers supported throughout the Commonwealth by the server support teams.) In contrast, the mean server per administrator average in the industry in a blended operating system environment is 32 servers per administrator.

Although it is not clear if Northrop Grumman's staffing for the server environment is sufficient, given the difference from the industry average an assessment of staffing ratios is needed. This assessment is needed because if the server support environment is not staffed appropriately for daily server operations, then the staffing for this or other incidents that require resources from the server support group may not be appropriate as well.

Recommendation

15. Conduct an evaluation of server support staff project, incident response and day to day activities to determine if the current staffing ratios are appropriate for the supported enterprise environment to include all remote sites.

Skill Sets

It is the opinion of Agilysys that Northrop Grumman's Skill Sets do not meet best practices because of a lack of commonly found Microsoft certifications in the Windows Server Operations.

Findings

During the interview process Agilysys observed that skill sets ranged from junior to senior level. This is by design and conforms to norms expected in the industry. However, even though industry vendor certifications are not the only measure by which skills are assessed, among Northrop Grumman's staff expected industry certifications are not common across the board.

Northrop Grumman uses three server support teams to support the server infrastructure. Northrop Grumman has provided information showing approximately twenty (20) Microsoft Certified System Engineers (MCSE) and ten (10) Microsoft Certified Professionals in the enterprise server support environment. All resources that have an MCSE also have an MCP by default since the MCP is a required step in attaining the MCSE. Northrop Grumman has stated that although a resource may be assigned to a support group servicing state agencies located in the capital region, the same resource is utilized for server support activities throughout the enterprise by use of remote access capabilities.

These data indicate that of the 81 server support personnel, approximately twelve percent (12%) of the server support staff are certified Microsoft Certified Professionals (MCP) and approximately twenty five percent (25%) are MCSEs. As mentioned previously Northrop Grumman has stated that they leverage expertise from across the server support groups, to provide support to the greater enterprise Windows environment. It is the opinion of Agilysys based on observations of other similar enterprise environments that one would expect that 25 percent (25%) of server support personnel would be MCSEs, as Northrop Grumman appears to have, but that fifty percent (50%) of the personnel would be MCPs.

Recommendation

16. Implement a certification tracking program with targets to meet industry averages, which tie training requirements to certifications commonly found in the industry.

Recovery and Restoration of Data, Systems and Databases

It is the professional opinion of Agilysys that Northrop Grumman's processes for the Recovery and Restoration of Data, Systems and Databases did not meet best practices because of the lack of remote "Point in Time" copies of the primary replicated data, and the minimal implementation of backup data recovery testing in the CESC location.

Findings

Need for Point-in-Time Copies of Global Catalog Should Have Been Evident Prior to Failure of the DMX-3. By design, replication mechanisms like SRDF replicate data with no regard to what the data is. As a result, if corruption occurs in the primary site the replication mechanism will replicate the corruption to the secondary site. It is the professional opinion of Agilysys that this risk of corruption is well known throughout the industry, and that implementation of a remote "Point in Time" copy of the data that is replicated to the remote site is an industry best practice to protect against a data corruption event. After EMC repaired the DMX-3 and returned it to service, the RCA indicates a gap of eighteen (18) hours before Northrop Grumman commenced data restoration (between just after midnight on August 27th, 2010 when EMC declared the DMX-3 repaired and 6:30 PM August 27th, 2010). As noted earlier in this document under Storage Management and Backup Services, due to the effects of data corruption Northrop Grumman had to restore the Global Catalog to insure the integrity of the enterprise backup and recovery system before it could begin to restore state agency data. The RCA notes that corruption of the enterprise backup system's Global Catalog was related to the DMX-3 memory replacement errors.

Procedures were in place to protect the Global Catalogs by replicating them to SWESC, as well as maintaining tape-based copies. However, the failure to suspend SRDF when there was an absence of remote "Point in Time" copies of the primary replicated data, in combination with the failure caused by replacement of the incorrect memory board, allowed the corrupted data to be replicated to SWESC thus overwriting the backup copies of the Global Catalog. To further complicate issues, the tape copy process for the Global Catalog backup was running during the incurred data corruption event, rendering those copies unusable. In the professional opinion of Agilysys, the implementation of remote data replication without remote "Point in Time " copies for data consistency does not meet industry best practices. In fact, the lack of local or remote "Point in Time" copies implemented for the Global Catalog highlighted the lack of an effective business impact assessment for critical infrastructure systems, which is a standard part of ITSCM. It is also the opinion of Agilysys that the lack of "Point in Time" clones/snapshots should have been identified as a risk to the Global Catalog data consistency protection prior to the failure of the DMX-3.

Once the enterprise backup and recovery system was restored to an operational state at approximately 6:30 PM on August 27th, 2010, data restoration operations commenced.

The recovery of file-based services and Microsoft SQL services did not incur any noticeable breaks in process, and thus there were not any delays. This is mostly attributed to the simplicity associated with these types of restores, the resiliency of the Microsoft SQL application and the intuitive nature of the Microsoft SQL backup and restore process.

In contrast, longer recovery times experienced during some database restorations were attributed to data corruption at the disk level. As a result, a single database restoration might have been tried multiple times before trying an older backup set. Agilysys was informed during interviews that there were no corrupt backups on the tapes but rather the data restoration attempts were being made to volumes (parts of the disks) on the DMX-3 that were experiencing hard to detect low-level file system inconsistency errors, due to the corruption caused by the DMX-3 failure. Once Northrop Grumman personnel identified the low-level format issue and performed corrective

actions, state agencies were able to successfully recover the databases from the restored data. It is the professional opinion of Agilysys that corruption at the disk level, which masked itself as corruption of databases, made it extremely difficult to diagnose the cause of the errors. The corruption of disk locations containing log files required application owners to rely on the most recent consistent backup, from August 23rd, 2010, for log file and database recovery.

Recovery Testing of Back Up Data Needs to Be More Frequent and a Random Sampling of Data Should Include All Agencies. The recovery testing process is completed twice yearly during scheduled Disaster Recovery testing. However, recovery testing completed during scheduled Disaster Recovery testing does not test the restoration of backup data for those state agencies that do not subscribe to Disaster Recovery Services. It is the professional opinion of Agilysys, that recovery testing of random backup data for the entire enterprise should be scheduled monthly in the CESC location. Recovery testing is essential to understand the impacts on the backup architecture, time to recovery and the coordination of recovery and application teams.

Recommendations

17. Implement EMC TimeFinder "Point in Time" clones/snapshots to protect critical open systems infrastructure servers providing a path to a quick, streamlined recovery (See item 2.)
18. As part of an overall review of backup/restore policies and procedures, implement full backup restoration testing procedures that take into account random samplings of data from across the enterprise and simultaneous restores. Backup staff, state agency application personnel, database support personnel and server support personnel should be engaged during recovery testing. (See item 4.)
19. As part of improving the efficiency of the backup/recovery methods, deploy the enterprise backup and recovery system agent for Oracle databases as a standard backup mechanism. This action will significantly simplify, centralize and streamline data backup/recovery operations, thus removing the identified points for error in the current process. (See item 5.)

Data Center Environment and Management

Data Center Facilities Fault Tolerance and Redundancy

It is the professional opinion of Agilysys that Northrop Grumman's Data Center Facilities Fault Tolerance and Redundancy met standards because all items reviewed pertaining to data center design and functions met Tier III standards.

Findings

The facilities conform to expected Tier Three (3) redundancy in construction, function and management. The mechanical and electrical equipment selection for the sites was superb and showcases some of the dedicated elements of the design.

Power to the facility is supplied by two different utility services. In addition the site has more than adequate power generation facilities and UPS ride-through protection. True to tier 3 tenets the diesel generators are supplied from two fuel tanks through two fuel rails. Water is supplied through a primary municipal source as well as a backup well source. In addition the site is equipped with a pumped well water source used for cooling tower makeup and to supply the deionized water system. Cooling is accomplished first by a pair of externally-mounted evaporative cooling towers, which supply a pair of advanced Freon-based chiller units. The system is assisted by pairs of forwarding pumps that appear to be logically obligated to each chiller. Humidity is maintained by an ultrasonic system incorporated into the air ducting and supplied by the deionized water system. In the professional opinion of Agilysys, while vendor documentation for this site is adequate, it is not assembled in a manner conducive to general use or interpretation. Much of this is handed off to a contract maintenance provider.

Confirmed on mechanical drawings M-701 and M-112, both cooling tower bays were piped to a common header, which is below grade in concrete. Should either cooling tower become contaminated, both will be contaminated in short order. The cooling water system was not capable of operating as two (2) segregated systems should either become contaminated.

Recommendations

20. To carry redundancy to an improved mission-critical reliability the supply pumps should be duplicated and configured as a primary/secondary 100%X100% pair and piped individually to each respective chiller. The same should be considered for the condenser-side feed pumps. Instead of a common header supplying the chillers/condensers they would be independent with only a bypass valve needed to bridge the systems during emergency operating conditions. This assumes full manual control of the pumps and chillers would be available.

Data Center Facilities Operations

It is the professional opinion of Agilysys that Northrop Grumman's Data Center Facilities Operations cannot be fully evaluated since Agilysys cannot adequately assess the performance of the contracted firm, Lee Technologies.

Findings

Facility operation and management is adequate and fits the higher-level tenets of a Tier 3 data center. NG appears to have a very capable facilities management staff. However, since most of the maintenance is scheduled by and performed by a contracted firm, it is not possible to adequately assess the maintenance team effort overall. Review of Lee Technologies procedures and service logs revealed a generic service life program.

Core Network Redundancy

It is the professional opinion of Agilysys that Northrop Grumman's Core Network Redundancy was above best practice because each layer of the network met expected redundancy best practices implemented in highly available environments.

Findings

The hierarchical three-tiered design used by Northrop Grumman has become the preferred architecture for most networks. The three-tiered architecture is comprised of an Access layer, Distribution Layer, and Core Layer. The Access Layer directly connects the host agency servers and access layer switches are also connected to a Distribution Layer via layer 2 802.1Q trunks. The Distribution Layer will have a number of access switches connected to it and these switches are deployed in pairs for system redundancy. The Distribution Layer switches are then connected to a core layer switch via layer 3. The Core Layer is a pair of switches deployed for redundancy and supports layer 3 as well as layer 2 functions.

It is the professional opinion of Agilysys that the implemented core network was well architected and provides a best of breed environment. The wide area network (WAN) edge was used to interface to the Verizon Multi-Protocol Label Switching (MPLS) cloud and provided access to the Internet for the CESC data center. The WAN routers implemented the virtual routing and forwarding (VRF) for each agency to meet the requirement of security segmentation in the cloud. The Firewall Service Modules (FWSM) implemented the virtual firewall contexts and provided a security boundary between the agencies in the data center.

Core Network Operational Practices

It is the professional opinion of Agilysys that Northrop Grumman's Core Network Operational Practices were not best practice because there is no formal schedule in place to test core network redundancy.

Findings

During interviews and data collection by Agilysys it was observed that a formal schedule is not currently in place to test core network internal redundancy. It is the professional opinion of Agilysys this does not reflect a best practice implementation of ITSCM.

Initial documentation on network support personnel detailed ten (10) tier one personnel, eight (8) tier two personnel and six (6) tier three personnel. Additional documentation provided to Agilysys outlined a much larger organization supporting the network environment to include subcontractors. This information results in an average of ninety point six (90.6) network devices per network support resource, compared to an industry mean average of forty-eight (48), as per Computer Economics "IT Staffing Ratios 2010, Benchmarking Metrics and Analysis for 15 Key IT Functions". In the professional opinion of Agilysys, a detailed study into the project requirements, incident response and day to day activities of the network operations staff is needed to conclusively determine if the current staffing model is adequate for the support of the environment.

Recommendations

21. As part of a review of ITIL practices within the environment, implement and document a network redundancy testing schedule, testing all redundant components a minimum of annually but preferably every six (6) months.
22. As part of a staff productivity review, assess current network staff workloads and productivity to develop a plan if required to expand the current network staff.

Monitoring and Proactive Management

Server Monitoring

It is the professional opinion of Agilysys that Northrop Grumman's Server Monitoring does not meet best practices for maintaining a stable infrastructure because server monitoring is implemented without any application dependency information, historical trending is ineffective, and the current notification process does not have any clear requirements as to when or if to notify state agency application owners of any outage. In the professional opinion of Agilysys, it is a top tier provider best practice to have adequate information on the inter-dependencies of servers and applications so that Northrop Grumman can inform state agencies when their applications have been affected by server-related problems.

Findings

HP OpenView Is a Best of Breed Monitoring System But Northrop Grumman's Implementation Does Not Reflect Top Tier Provider Practices. The HP OpenView software used by Northrop Grumman is a best of breed industry standard monitoring and alerting framework. The software is very advanced and has many configurable options and modules allowing it to be custom tailored to any organization. Based on data collected during interviews with Northrop Grumman's HP OpenView Staff, Agilysys observed that the current implementation of OpenView within the infrastructure reported on network equipment alerts, basic Windows server alerts, and some UNIX server alerts.

However, the current HP OpenView implementation lacks key error correlation information that would indicate how problems involving a server affect applications dependent on that server. Northrop Grumman's configuration management database lacks sufficient information on server and application dependencies that would provide information on how an application is affected by a hardware failure. As a result, when a ticket is opened in response to an HP OpenView alert Northrop Grumman does not have enough information to inform state agencies that an application has been affected by instability in the operating environment. Instead, the current error correlation merely goes one (1) level deep to hardware only and does not provide enough information to properly assess the impact of a hardware or network outage. As has been discussed previously in this document, top tier service provider monitoring architectures take into account the implementation of dependencies and error correlation.

In addition, historical reporting on the metrics collected is available for only forty five (45) days. In a complex enterprise environment such as is managed by Northrop Grumman, trending analysis needs to take into account month end, quarter end, and year end activities to properly track issues associated with providing a stable server environment during periods of increased activity. It is the professional opinion of Agilysys that without historical reporting capabilities outside of forty five (45) days within the alerting environment it is nearly impossible to review any error trending analysis and identify or report on any consistent problems affecting the operations of servers within the environment.

Based on the information provided, it is the professional opinion of Agilysys that the monitoring system does not provide adequate application and server dependency information and is not a best practice implementation of a monitoring environment compared to top tier service providers with which Agilysys is familiar.

Current Server Outage Notification Process Has Gaps. During interviews with state agency personnel, Agilysys observed that there are gaps in the communication process for incident notification and alerting pertaining to server outages. In some cases application owners have not been informed or aware of server availability issues during weekend and production hours that directly affected dependent applications. In addition, the current notification process does not have any clear requirements as to when or if to notify application owners

of any outage. This lack of a policy would explain why no notification was sent to the state agency application owner at the start of the DMX-3 outage. Moreover, without any knowledge of server and application dependency reported from the HP OpenView environment and Configuration Management Database (CMDB) it would be impossible for the operator generating the ticket for the server availability issue to understand the resulting impact. It is the professional opinion of Agilysys that although Northrop Grumman's notification, escalation and ITIL processes are well documented, the lack of an official process for notifications to be sent to the state agency application owner is not best practice. It is also the professional opinion of Agilysys, that the flaw residing in the lack of dependency information currently available in the CMDB does not ensure a stable operating environment..

Recommendations

23. As part of a continual process to improve the monitoring and alerting infrastructure, implement historical reporting capabilities that allow for data trending longer than forty five (45) days to proactively monitor, track and implement remediation processes to avoid performance issues in the environment.
24. Update Configuration Item ("CI") relationships to reflect application and server dependencies in the configuration management database. Defining application dependencies creates an environment that provides more data at the onset of an outage as to what systems are directly affected. (See item 16.)
25. As part of ITIL review and process improvement implement a review of the current infrastructure alert and escalation process.

Database Monitoring

It is the professional opinion of Agilysys that Northrop Grumman's Database Monitoring did not meet best practices because the enterprise monitoring system HP OpenView does not monitor the complex multi-version Microsoft SQL database environment and there is an apparent lack of monitoring or a rules misconfiguration within the implementation of Oracle Enterprise Manager.

Findings

Microsoft SQL Databases Lack Best Practice Central Enterprise Monitoring and Management Tools. During interviews of Northrop Grumman database support staff and a review of data provided to Agilysys by Northrop Grumman it was observed that enterprise-monitoring tools in the SQL environment were not implemented. Currently Microsoft SQL 2008 Standard Management Studio is deployed but that version is only capable of managing single server instances. In contrast Microsoft SQL 2008 Enterprise Management Studio, which is not deployed, has built in multi-server administration support enabling one console for all SQL management activity. It is the professional opinion of Agilysys that the decentralized management model currently implemented, due to the version of the tools deployed and the complexity of managing versions of SQL 2000, 2005 and 2008, creates an environment that adds undue overhead to the administration of the SQL environment.

In an environment that supports four hundred eighty-two (482) Microsoft SQL databases with varying versions from SQL 2000 through SQL 2008, it is not best practice to have the issuance of current notifications of database health and errors to SQL database administrators be implemented in a decentralized model, and depend upon the use of contact information in the email specific destination properties of the database. Any state agency application owner with database administration privileges has the ability to alter the parameters for email alert destinations of the database, thus changing the original email recipient alert destination. During interviews with SQL administration staff it was noted that this condition had been experienced in the daily operations of the SQL environment. It is the professional opinion of Agilysys that the lack of centralized enterprise monitoring tools in the environment presents a situation that affects the operational proficiency of the supported SQL environment. It is also the opinion of Agilysys that this is not an implementation indicative of an enterprise of this complexity and size and therefore does not reflect a best practice implementation of SQL monitoring tools in comparison to top tier service providers with which Agilysys is familiar.

Oracle Databases Are Managed by Oracle Enterprise Manager But Its Use Appears Limited. During interviews conducted by Agilysys of Northrop Grumman database support staff and based upon a review of data provided to Agilysys by Northrop Grumman, it was observed that the Oracle database environment consisted of approximately one hundred sixty-eight (168) Oracle databases managed by Oracle Enterprise Manager and that the monitoring of the environment by the HP OpenView monitoring framework was suppressed. Although Northrop Grumman has deployed these tools it does not appear that they have fully employed them. This is illustrated by Northrop Grumman's response to the outage. Agilysys observed that the timeline and incident ticket data provided by Northrop Grumman indicated that server and database errors were reported to the Help Desk by state agency users and application owners starting at approximately 3:25 PM on August 25th, 2010. However, time line data provided by state agencies to Agilysys indicated that database errors were observed half an hour earlier, at 2:55 PM, on August 25th, 2010, but there was no data in the master incident ticket information for the outage, or any tickets associated with the outage provided to Agilysys by Northrop Grumman to indicate that Northrop Grumman database support personnel had opened any initial incident tickets pertaining to database issues. It is the professional opinion of Agilysys that this highlights a lack of proactive database monitoring, or a misconfiguration of rules governing alerting within Oracle Enterprise Manager in the Oracle database environment by Northrop Grumman, and that this practice does not reflect best practices for a monitoring environment compared to top tier service providers with which Agilysys is familiar.

Recommendations

26. As part of a review of the SQL management environment investigate and deploy Microsoft SQL enterprise monitoring tool within HP OpenView centralizing monitoring of the SQL environment.

Conclusion

The information included in this audit covers many aspects of the reviewed environment, including computing, storage, data recovery, monitoring, and management. Implementation of the recommended actions provides the groundwork for creating a more agile, proactive environment that can respond to critical incidents in a timely and efficient manner. The recommended actions within each topic should be reviewed and an assessment of the scope of work to implement such actions be created.

Human error during the memory board replacement process resulted in the incurred extended outage. Both memory boards zero (0) and one (1) were reporting errors prior to the memory board replacement. Memory board one (1) was reporting hard errors and memory board zero (0) was reporting soft errors. As stated in EMC's RCA, "*the initial determination to replace memory board 0 first did not take into account the uncorrectable events that had posted on board 1"* and *"Based on extensive post-incident analysis, EMC has determined that replacing memory board 1 first would have prevented any issues during the replacement activity itself."*

During the interviews and in a review of the data, two issues kept presenting themselves. The first is that in the professional opinion of Agilysys, the absence of a process for determining the conditions in which the data replication mechanism (SRDF) should be suspended allowed this process to continue to run even though maintenance was being performed in an environment of unusual risk. As mentioned in different sections throughout this document, it is the professional opinion of Agilysys that the lack of proactive planning regarding when to suspend the SRDF replication mechanism in the absence of remote "Point in Time" copies of data was the cause of data corruption experienced in the SWESC recovery site. If replication had been stopped prior to the hardware replacement, incremental restores of data for customers subscribing to the tier one replication service could have been completed from the SWESC location, reducing recovery times and streamlining recovery actions. Northrop Grumman indicates that a process is in place to assess if the SRDF replication mechanism needs to be stopped in the event of a possible corruption event, but it does not appear that a full impact analysis has been completed to identify events that would require the stopping of SRDF, or that documented procedures have been provided to support staff. This exercise should be completed to avoid the issues experienced during the August 25th event.

The corruption of the Global Catalog and other critical databases highlights the second issue. Namely, a lack of data protection in key environments. Although the mainframe environment uses "Point in Time" snapshot/clone copies to recover from data corruption/disruption events, this process is not used in the open systems environment. The question arises as to why the same level of criticality has not been assigned to other key applications that the enterprise relies upon. A business impact analysis review should be implemented in conjunction with VITA and state agencies to reassess the recovery time and recovery point objectives needed for key data. Recovery metrics should be based on business criticality, revenue loss, and how long a disruption of service can be absorbed by the business.

The lack of active monitoring of the environment also raises a concern. It is the professional opinion of Agilysys that the configuration management database and OpenView architecture is not at the maturity level expected at this point in its lifecycle, and that the degree and kind of current reporting on detailed application and system dependencies available during events, which is needed to enable a stable environment, is inadequate. It was also noted that historical reporting for error trending is only available for forty five (45) days. Furthermore, there is no official process as to when and if to notify state agency application owners when a system outage is observed. During interviews with Northrop Grumman staff and sub-contractors it was stated, that there are efforts underway to implement projects to improve upon historical trending and reporting, and add additional application dependency information to the configuration management database, but none are currently initiated. It is the professional opinion of Agilysys, that these two issues should have been part of the initial design and requirements.

Currently the storage management tools are available but installed in a limited fashion. The environment has been in place for more than four (4) years, and by this time one would expect to see greater use and implementation of an intelligent toolset in an environment of this size and complexity. Control Center and related performance monitoring and capacity planning tools such as Ionix Control Center Performance Manager should be implemented in a more complete fashion in the environment.

The current multi-step backup process for Oracle databases introduces points in the process for human error and data integrity issues. The backup process for Oracle databases should be reviewed and an agent-based approach, that integrates with the enterprise backup system and that centralizes and streamlines the backup process should be implemented as a standard.

Restore testing of data is documented and completed twice yearly during scheduled disaster recovery testing. Instead, data restoration testing should be completed monthly using a random sampling of data from the entire enterprise, including those state agencies that do not subscribe to the disaster recovery services. This process will provide insight into whether recovery procedures need to be modified and provide metrics on the average time to restore data. The process should be documented and updated to reflect the changing environment.

In closing, although a substantial amount of transformation occurred in 2010, Northrop Grumman should have anticipated the environment it would need to support when it designed its processes. It is the professional opinion of Agilysys that many of the deficiencies outlined in this document represent an insufficient degree of self-governance towards continuous process improvement and the management of risk in the environment. In order for organizations to successfully provide quality IT services to their customers, processes pertaining to every aspect of the delivery of service to the customer must be reevaluated and modified to manage the risk to the environment. This continual process of improvement enables organizations to take advantage of technological changes, develop required individual skills and organizational competencies, and manage growth.

Appendix A

Formal Responses from Northrop Grumman and the Virginia Information Technologies Agency

After completing the research for this audit, Agilysys provided more than one opportunity for staff from the Virginia Information Technologies Agency (VITA), the Joint Legislative Audit and Review Commission, and Northrop Grumman to review the document and provide technical feedback.

As part of this process, both Northrop Grumman and VITA were given the opportunity to provide additional documentation and to submit formal, written responses in the form of a letter. Those letters are included in this appendix.

Northrop Grumman Corporation
Information Systems
Civil Systems Division
Chesterfield Enterprise Solutions Center
11751 Meadowville Lane
Chester, VA 23836
(804) 416-6555
sam.abbate@ngc.com

February 11, 2011

Dear Mr. Nixon and Mr. Tittermary:

Northrop Grumman thanks you for the time and effort spent preparing this Independent Assessment of the unprecedented events that led to the August 2010 infrastructure failure and the recovery process. Our goal since the first moment of the outage has been to learn from this experience and develop and implement better ways to mitigate and manage risks. Many new procedures, policies and safeguards have already been established. We agree with many of the recommendations of the Independent Assessment. We are ready to engage in discussions about how they could be best implemented. Many require the involvement and support of the agencies Northrop Grumman serves, so they will have an important role to play in considering some of the changes.

Northrop Grumman disagrees with the premise that industry best practice evaluations can be applied without regard to industry segment. A number of service areas have been inappropriately evaluated as less than industry best practice based upon unsubstantiated opinions, rather than through application of documented and verifiable industry norms as required by our contract. The report also fails to recognize instances where Service Level Agreements supersede industry best practices as standards for performance.

Some of the most important recommendations are matters for the Commonwealth's policy makers to consider, for example establishing a common definition of critical data and determining what protective measures should be undertaken. Also, the report observes and is critical of the fact that typical enterprise approaches are not always applied in the program. These observations may be fundamentally correct, but individual agencies, not VITA or Northrop Grumman, retain the ability to deviate from centralized enterprise standards.

There are also some portions of the Independent Assessment which are factually incorrect, such as the length of time the environment has been in place, which is fundamental to establishing reasonable expectations of its maturity. The report also misreads the contract's 10-day period to "commence performing a Root Cause Analysis" as the deadline for its submission.

As with any technical analysis of this type, experts will disagree over some conclusions and recommendations. Northrop Grumman disagrees with some of the contents, particularly where the context of the contract, including definitions of roles and responsibilities, has not been adequately considered. The report fails to take into account the dynamics of the working relationship between VITA, Northrop Grumman and the Commonwealth's agencies. For

example, all of the recommendations related to business impact assessments are clearly the responsibility of the application owners--the agencies. Those areas of the report, where it is unclear that the findings and recommendations are the sole responsibility of Northrop Grumman, deserve additional analysis.

We remain committed to this partnership and are dedicated to fully meeting our obligations with regard to the quality of the Commonwealth's infrastructure and the services we deliver. We look forward to working together as we continue, within the framework of our contract, to devise improvements and solutions that are in the best interest of the Commonwealth and its citizens.

Sincerely,

Samuel A. Abbate
Vice President
VITA IT Partnership

COMMONWEALTH of VIRGINIA

Virginia Information Technologies Agency
11751 Meadowville Lane
Chester, Virginia 23836-6315
(804) 416-6100

Samuel A. Nixon, Jr.
Chief Information Officer
E-mail: cio@vita.virginia.gov

TDD VOICE -TEL. NO.
711

February 14, 2011

Mr. Glen Tittermary
Director
Joint Legislative and Audit Review Commission
General Assembly Building, Suite 1100
Richmond, Virginia 23219

Dear Mr. Tittermary:

Thank you for the opportunity to comment on the report provided to the Commonwealth by Agilysys Technology Solutions Group, LLC ("Agilysys"), *Audit of Northrop Grumman's Performance Related to the DMX-3 Outage and Associated Infrastructure*. The report represents nearly three months of extensive research and analysis by Agilysys. I appreciate the thoroughness and professionalism Agilysys exhibited in conducting the audit and authoring this report.

I believe the audit findings contained in the report confirm the Commonwealth's basic understanding of the circumstances that led to the computer outage and subsequent delays in recovery and complete restoration of agency operations. Specifically, I am pleased that the report provides a list of measures written by industry experts that need to be taken to reduce the possibility of a similar failure in the future. In this light, Agilysys' report will help the Commonwealth hold Northrop Grumman accountable for future performance of its duties on behalf of the Commonwealth and its citizens.

The IT infrastructure program with Northrop Grumman is a complex, yet critically needed program for the Commonwealth. VITA, Northrop Grumman and our customer agencies have overcome significant obstacles and challenges. Considerable progress has been made to consolidate, centralize, and standardize Virginia's infrastructure into a cohesive, secure platform across the enterprise of state government. All but a few state agencies have been transformed. But as this report indicates, the work to ensure that the Commonwealth enjoys the benefits of transformation is ongoing, and requires constant discipline and effort by both the Commonwealth and Northrop Grumman.

Again, I thank you for the opportunity to respond to this report.

Cordially,

Samuel A. Nixon, Jr.