94.4.1 NSLDS II Technical Architecture Reassessment

Mid-Tier and Mainframe Alternatives

December 10, 2002

Outline

I. Purpose

II. Executive Summary

III. Configuration Assumptions

IV. NSLDS II – Proposed Architecture Alternatives

V. Planned NSLDS II Environment Context

   a. Alternative 2 – Mainframe (zSeries z900)

   b. Alternative 4 – Mid-tier (clustered IBM P690)

   c. Alternative 5 (Original) – Mid-tier (clustered IBM P660)

VI. Comparison Highlights

VII. ROM Pricing and Total Cost of Ownership

VIII. Findings

   a. Cost

   b. Technical

IX. Appendix A – Capabilities Comparison

X. Appendix B – Disk Space Capacity Analysis

XI. Appendix C – Capacity Summary

XII. Appendix D – Deferred Items

Purpose

- Perform a review of an NSLDS II Mainframe alternative, to demonstrate whether the mainframe is a viable architecture to support NSLDS II requirements as defined in the NSLDS II preliminary design (Deliverable 94.2.2).

- Provide a mainframe configuration specification that CSC will use to create a Rough Order of Magnitude (ROM) estimate for operations costs (Pricing will be used for comparison only).

Executive Summary

- A team of FSA, Modernization Partner, IBM, CSC and Raytheon personnel was assembled.
- In addition to the zSeries Mainframe (Alternative 2) and the p660 Mid-Tier (Alternative 5) , the p690 Mid-Tier (Alternative 4) was also included in this analysis.
- After a high-level, week-long analysis it was determined that each of the configurations sufficiently meets the previously agreed upon NSLDS II Requirements for 2003. However in scaling the p660 architecture (Alternative 5), CSC cannot guarantee meeting the required SLA of 99.7%.
- The five-year Total Cost of Ownership (TCO), including DB2 Licensing and Maintenance as well as Operations Costs (GFI), is the major differentiator. A pricing breakdown by platform is provided.
- Summary highlights and findings along with the complete comparison are included.

Configuration Assumptions

Processing capacity is defined in fundamentally different ways for mainframe and mid-tier architectures. Therefore, the team used the following assumptions and logic to arrive at comparable benchmarks for the different platform configurations.

Mainframe:

- Mainframe capacity is measured in Millions of Instructions per Second (MIPS).
- The current NSLDS system MIPS of 361 was used as the baseline for processing needs required by the EDW portion of the NSLDS II solution.
- The MIPS requirement for the Data Mart portion of the NSLDS II solution equals the MIPS required by the EDW (361), for a total FY02 base MIPS requirement of 722*.

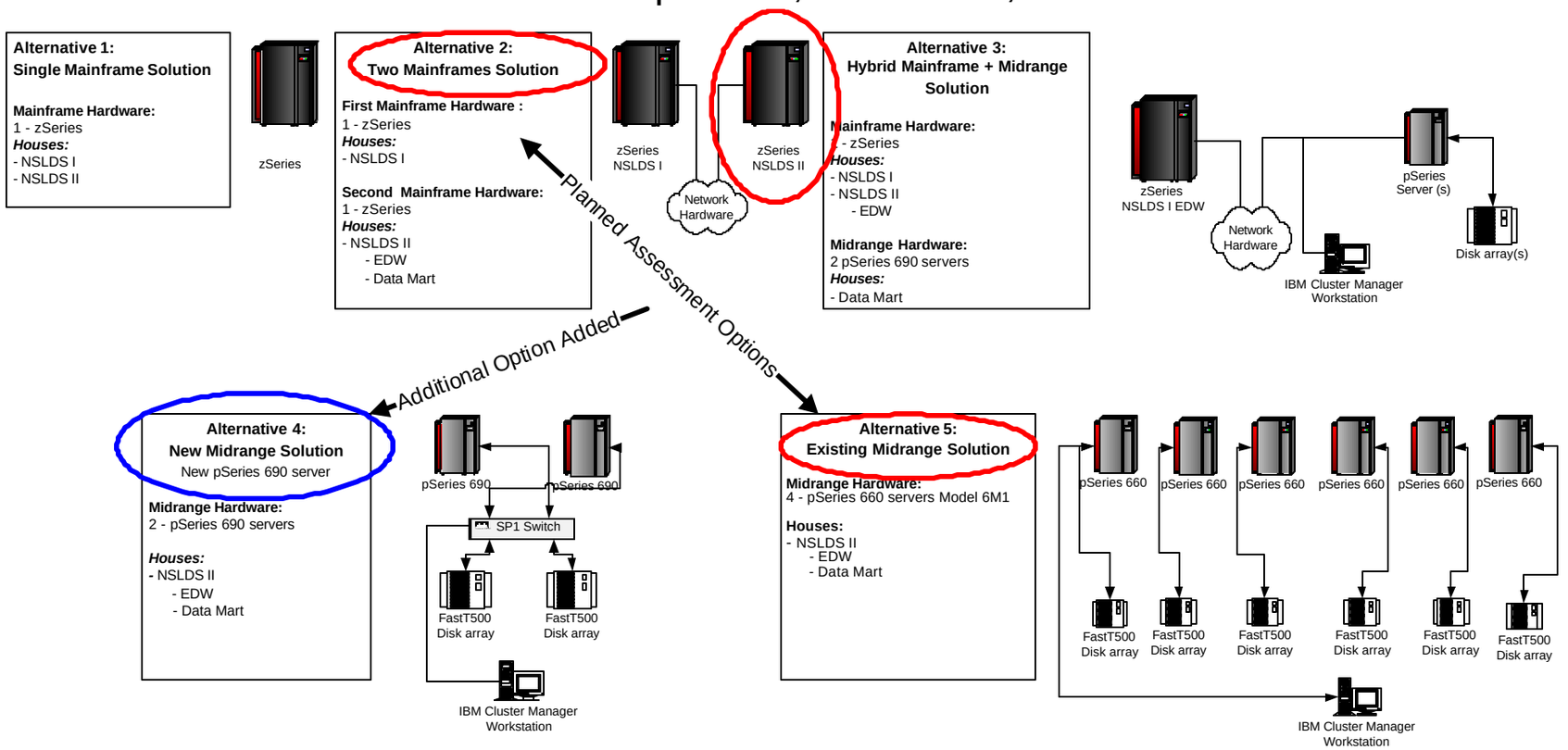
Mid Tier:

- Mid-tier capacity is measured in Transactions per Minute (TPM).
- The batch load processing will drive peak TPMs for the NSLDS II system.
- At peak 88 million rows during a single batch window, resulting in 490,000 TPMs* are required.

** See Appendix C for MIPS /TPM Calculations and Growth Projection*

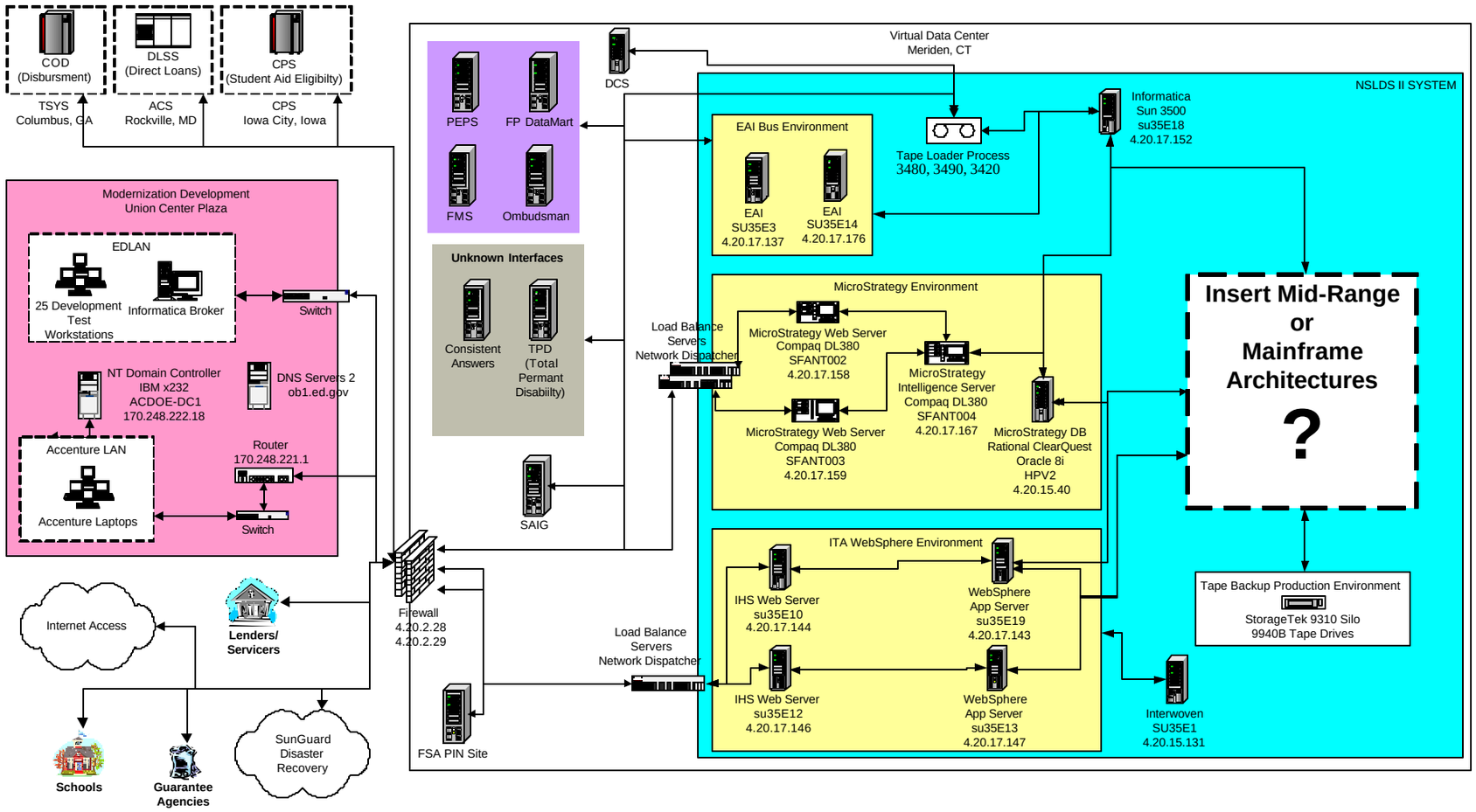
NSLDS II - Proposed Architecture Alternatives

For this work effort, options 2 and 5 were originally proposed. Option 4 was added due to market advancement of the pSeries, model 690, over the last 6 months.



Planned NSLDS II Environment Context

Either DB2 hardware platform will leverage the existing FSA Architectural Standards for Web, Reporting, and ETL (Extract, Transform, and Load of Data).



Alternative 2 – Mainframe (zSeries z900)

The NSLDS II Mainframe Configuration for 2003 is well within the maximum capacity of a Z900 Model 104.

NSLDS II 2003 Mainframe Capacity Requirements

2003 MIPS = 787

2003 Memory = 10 GB

2003 DASD = 10 TB

Maximum Single zSeries Scalability

Z900 Maximum MIPS = 3192

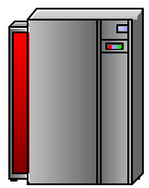
Z900 Maximum Memory = 64 GB

Z900 Maximum DASD = (Dependent on External Array)

(XP-512 DASD, or an HDS 9960)

Production, Dev and Test Topology

(Multiple LPARs for Production, Dev, Test1, Test2, Test3, Training/ UA, Maintenance)



zSeries z900
Model 104

MIPS = Millions of Instructions per Second

TB = Terabytes

DASD = Direct Access Storage Device

LPAR = Logical Partition

Alternative 4 – Mid-tier (clustered IBM P690's)

The NSLDS II Mid Range Configuration for 2003 is within the maximum capacity of two p690 servers.

NSLDS II 2003 Mid Tier Capacity Requirements

2003 TPM = 545,000

2003 Memory = 96 GB

2003 DASD = 15 TB

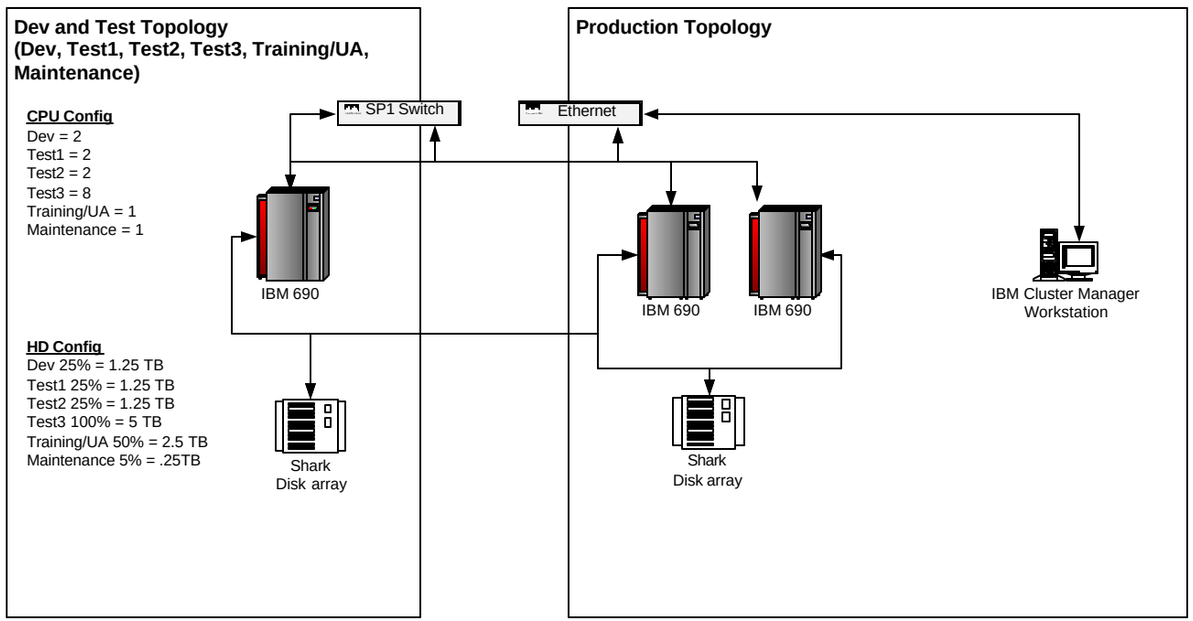
Maximum Mid Tier Dual Clustered p690 Scalability

Maximum TPMs = 403,000 x 2 = 806,000

Maximum Memory = 256 GB x 2 = 512 GB

Maximum DASD = (Dependent on External Array)

(Shark Array maximum = 55 TB)



Alternative 5 (Original) – Mid-tier (clustered IBM P660's)

The NSLDS II Mid Tier Configuration for 2003 is within the maximum capacity of multiple p660 6M1 servers.

NSLDS II 2003 Mid Tier Capacity Requirements

2003 TPM = 545,000

2003 Memory = 96 GB

2003 DASD = 15 TB

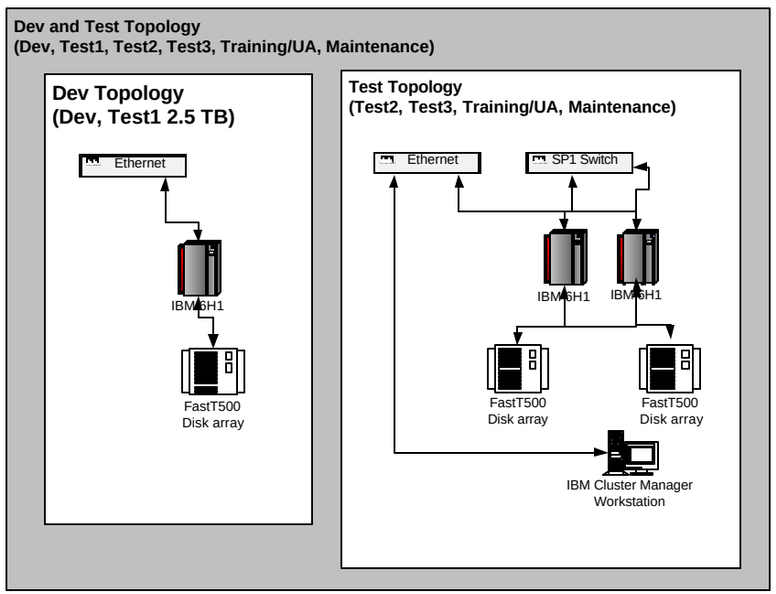
Maximum Mid Tier 6 Way Clustered p660 6M1 Scalability

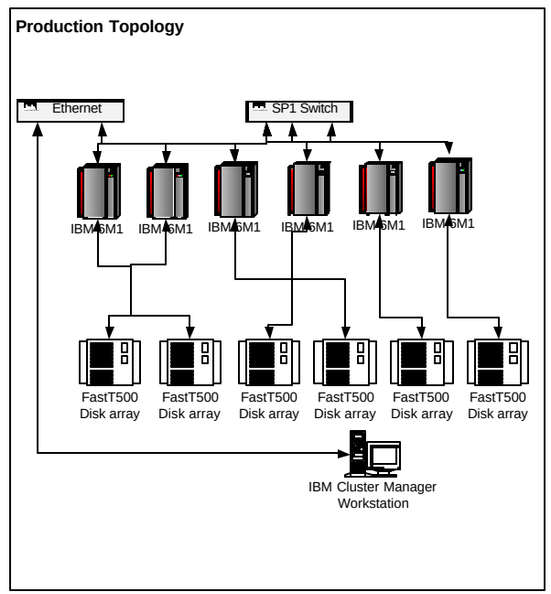
Maximum TPMs = 105,000 x 6 = 630,000

Maximum Memory = 64 GB x 6 = 384 GB

Maximum DASD = (Dependent on External Array)

(FastT500 Array maximum = 14 TB x 6 = 84 TB)





Comparison Highlights

Each platform received a passing mark for each of the 20 items compared in this analysis. Of the 20*, these 7 are the items for which there was some material difference in capability between the Mainframe and Mid-Tier architecture.

Requirement	Definition	Mainframe	p690 Mid Tier	p660 Mid Tier
High Availability	99.7% Availability outside normal maintenance	+ Capable of meeting a 99.99% availability SLA	+ Capable of meeting a 99.9% availability SLA	Capable of meeting the 99.7% availability SLA in first year.
Scalability	Ability of the solution to scale in size by a factor of +/- 30% per year.	+ Smallest initial footprint (physical size). Capable of scaling without footprint increase.	+ Larger initial footprint than mainframe. Capable of scaling without footprint increase.	CSC cannot guarantee the SLA when scaled. Also largest initial footprint. Incapable of scaling without footprint increase
Interoperability	Ability to integrate with the existing tape loaders and virtual tape servers.	Supports required interoperability	- Alternative approach required to support interoperability.	- Alternative approach required to support interoperability.
Database Capabilities – EDW	Ability of the solution to support a 3rd normal form EDW	+ Inherently better I/O performance for EDW (transactional) functions. Workload Manager Tool available to optimize load.	Effectively supports EDW (transactional) functions	Effectively supports EDW (transactional) functions
Database Capabilities – Mart	Ability of the solution to support a dimensional data mart.	Effectively supports Mart (dimensional) functions	+ Greater query performance for Mart (dimensional) functions	+ Greater query performance for Mart (dimensional) functions

Plus (+) indicates requirement exceeded. **Minus (-) indicates additional work will have to be done to meet the requirement.**

*** The full comparison is included as Appendix A of this deliverable.**

ROM Pricing and Total Cost of Ownership

Below is the “apples-to-apples” comparison for TCO using the Government Furnished Hardware Operations Costs and IBM Provided DB2 Licensing and Maintenance Costs. TCO of the Mainframe is significantly higher than that of either Mid-Tier solution.

FY03 – FY07 (Cost in Millions)			
Platform	DB2	Operations	TCO
p660	$5.3	$28.7	$34.0
p690	$3.6	$36.9	$40.5
Mainframe	Included in Ops	$102.0	$102.0

Please See Appendix A for a complete explanation of costs broken down by year for each of the three Alternatives

Findings - Cost

Each platform utilizes a different software mix and hardware configuration to meet the analyzed NSLDS II requirements. In the end, TCO provided the biggest differentiator as a single point of comparison across the alternatives.

Cost Findings:

- TCO for the Mainframe is 2.5 times higher than the p690 and 3 times higher than the p660. These costs represent a combination of GSA (DB2) and GFI (Operations) costs and are for comparison purposes only.

- It is also expected that a substantial discount could be negotiated on both the Mid-Tier Server Hardware and DB2 software, driving the TCO even lower for the Mid-Tier alternatives.

- There is no known expectation for substantial negotiated discounts regarding the mainframe alternative.

Findings - Technical

While the analysis found that both solutions were capable of meeting all of the previously determined NSLDS II requirements, each architecture platform had different advantages.

Mainframe Solution:

- Since the EDW transaction environment runs more efficiently on the mainframe platform, there should be greater processing capability for transactional updates and loads run against the NSLDS II EDW.
- Proven interoperability with the existing VTS/ATL tape store.
- Exceeds high availability requirement. Capable of meeting SLA up to 99.99%.

Mid Tier solution:

- Since the data mart environment runs more efficiently on the mid tier platform, there should be greater performance for business intelligence processes executed in the NSLDS II data mart.

Appendix A – Capabilities Comparison

Requirement	Definition	Mainframe z900 zSeries running DB2 on zOS	Mid-Tier clustered p690 running DB2 EEE on AIX	Mid-Tier clustered p660 running DB2 EEE on AIX
Capacity	Ability of the solution to accommodate the Transactions Per Minute (TPM), memory and data storage requirements of NSLDS II.	The mainframe configuration (700 MIPS) created during the assessment is sized to handle the current NSLDS II design. While additional internal hardware changes would be necessary, this configuration of the z900 is capable of scaling to over 3000 MIPS within the same physical footprint. In addition, DB2 Estimator, a tool offered for DB2 on the zSeries mainframe, can be used throughout the life cycle for capacity analysis, performance analyst, and "what if" modeling. The group agreed upon a peak NSLDS II I/O requirement of 500 MB per second. As configured, the Mainframe has the highest I/O bandwidth; a Z900 can handle 24 GB per second per machine.	The p690 configuration created during the assessment is sized to handle the current NSLDS II design. Similar to the mainframe, internal capacity growth is available within this configuration and physical footprint, however the processing ceiling is lower than the z900 mainframe. The group agreed upon a peak NSLDS II I/O requirement of 500 MB per second. As configured, this p690 Mid-Tier solution has an I/O bandwidth of 8 GB per second per machine totaling 16 GB for production and 8 GB for non-production).	The p660 configuration created during the assessment is sized to handle the current NSLDS II design. Unlike the mainframe and Mid-Tier p690 configuration, internal capacity growth is not available within this configuration and physical footprint. In addition, capacity upgrades are more complex on a p660. (the need for additional servers / arrays is greater as capacity grows). The group agreed upon a peak NSLDS II I/O requirement of 500 MB per second. As configured, this p660 Mid-Tier solution has an I/O bandwidth of 1 GB per second per machine totaling 6 GB for production and 3 GB for non-production).
High Availability Analysis (SLA) Meets SLA	Ability of the solution to provide access to the production system at least 99.7% (VDC Priority 2) of the time outside allowable scheduled maintenance windows. During the course of the analysis it was pointed out that the existing NSLDS Production system is maintained with a 99.9% availability. During the course of reengineering this VDC Priority and corresponding SLA should be reexamined to determine is the nature and quantity of the workload merits this 99.9% availability requirement.	Will meet a 99.9% availability requirement and is configurable up to 99.99% availability.	While the p690 architecture will meet the 99.9% availability. This SLA requires the implementation of a software based failover configuration called High Availability Clustered Multiprocessing (HACMP). This HACMP configuration is a component of the pricing model.	The p660 architecture is capable of meeting the 99.7% requirement, however CSC was not comfortable stating that it would support an SLA of 99.9% availability on the clustered p660 environment. Further research would have to be done to determine that possibility.
Disaster Recovery	Ability of the solution to be restored fully.	Disaster Recovery on the mainframe architecture is the least complex and lowest cost of all proposed architectures.	Disaster Recovery on the p690 environment is less complex and less costly than on the p660 architecture, but more complex than on the mainframe zSeries architecture.	Higher complexity and greater cost than each of the other architecture solutions.

Appendix A – Capabilities Comparison

Requirement	Definition	Mainframe z900 zSeries running DB2 on zOS	Mid-Tier clustered p690 running DB2 EEE on AIX	Mid-Tier clustered p660 running DB2 EEE on AIX
Scalability	Ability of the solution to scale in size by a factor of +/- 30% per year. Scalability includes the hardware, software and connectivity necessary to support the solution. (If processes are moved off of the backend solution, the scalability could go down).	All solutions are scalable to meet requirements and the five year projected capacity growth (see Appendices A, B and C). The mainframe architecture starts with the smallest physical footprint, and is scalable using upgrades or additions which are internal to the physical mainframe box. This means that there would be no increase in physical footprint, or fees associated with physical footprint increases, as the architecture scales.	All solutions are scalable to meet requirements and the five year projected capacity growth (see Appendices A, B and C). The p690 architecture starts with the physical footprint that is slightly larger than the mainframe, and like the mainframe it is also scalable using upgrades or additions which are internal to the physical boxes. However, to match the maximum capacity of the mainframe additional boxes would need to be added to the cluster, meaning an increase in physical footprint and fees associated with that physical footprint increase.	All solutions are scalable to meet requirements and the five year projected capacity growth (see Appendices A, B and C). The p660 architecture starts with the physical footprint that is much larger than the mainframe and larger than the p690. Unlike the other two architectures, it is not scalable using upgrades or additions which are internal to the physical boxes. To match the maximum capacity of either of the other configurations, additional boxes would need to be added to the cluster, meaning an increase in physical footprint and fees associated with that physical footprint increase.
Hardware				
Software		The ability to scale the DB2 software is less granular and more rigid on the zSeries mainframe platform. Since DB2 is priced on a per MIP basis on the mainframe, and MIP increases are in "chunks", DB2 pricing follows this "chunk" model as well.	The ability to scale the DB2 software is more granular on the mid-tier platform. Since DB2 is priced on a per CPU basis on the mid-tier, as CPUs are added, DB2 pricing follows this small increment model as well.	The ability to scale the DB2 software is more granular on the mid-tier platform. Since DB2 is priced on a per CPU basis on the mid-tier, as CPUs are added, DB2 pricing follows this small increment model as well.
Network Connectivity	Ability of the solution to integrate with the existing FSA VDC network architecture (standard gigabit server network).	The mainframe configuration has fewer physical connections (One Server) to maintain and therefore inherently fewer opportunities for network issues.	The p690 configuration has more physical connections (three servers) than the mainframe (one server) to maintain and therefore inherently more opportunities for network issues. It does however have fewer physical connections to maintain than the p660 configuration (nine servers) and therefore inherently fewer opportunities for network issues.	The p660 configuration is the most complex in terms of physical connections (Nine Servers). This number of connections and servers also dictates that a separate network be created to maintain the environment, creating the most opportunities for network issues of the environments compared in this study.
Security	Ability of the solution to provide Login, Authentication, Operating System, Web compatible and Database level security.	Security on the mainframe configuration would make use of the existing RACF product and would require no re-design or recoding effort to implement in the new zOS NSLDS II environment.	The p690 configuration would make use of the RACF model from the existing NSLDS, however implementation of this model would require some re-design and recoding to implement it in the mid-tier p690 AIX NSLDS II environment.	The p660 configuration would make use of the RACF model from the existing NSLDS, however implementation of this model would require some re-design and recoding to implement it in the mid-tier p690 AIX NSLDS II environment.

Appendix A – Capabilities Comparison

Requirement	Definition	Mainframe z900 zSeries running DB2 on zOS	Mid-Tier clustered p690 running DB2 EEE on AIX	Mid-Tier clustered p660 running DB2 EEE on AIX
Replication	Ability of the solution to support frequent (a minimum of one time per hour) replication of data from the 3rd Normal Form data warehouse to the dimensional data mart.	No clear advantages	No clear advantages	No clear advantages
Backup	Ability of the solution to support the nightly incremental backup of data in the data warehouse and weekly full backup of the production file system. Currently, Saturday maintenance is 6 pm to 2 am and Sunday's maintenance window is 6 pm to 3 am.	There is a performance advantage as the data on the database is already compressed for backups. This results in less physical tape required for data backup.	Since the data in the database is not stored in a compressed format, the backup to tape will require more physical tape storage.	Since the data in the database is not stored in a compressed format, the backup to tape will require more physical tape storage.
Interoperability	Ability of the solution to integrate with the existing tape loaders and virtual tape servers.	This architecture will provide support for all required tape drives without the need for a work around solution.	In order for this environment to support all required tape drives, a work around for the 3480 and 3420 tape drives will have to be implemented. This work around could be to switch trading partners to electronic format or use the CPS system as a pass through.	In order for this environment to support all required tape drives, a work around for the 3480 and 3420 tape drives will have to be implemented. This work around could be to switch trading partners to electronic format or use the CPS system as a pass through.
IBM 3490, 3480 and 3420 tape loaders				
IBM 3494 Automated Tape Library (ATL) Data server with Magstar Virtual Tape Servers (VTS) and 3480s 3490s or 3490-Es tapes.		Mainframe is compatible with the VTS / ATL architecture in place currently.	Further research needs to be done to determine the p690 compatibility with the VTS/ATL architecture.	Further research needs to be done to determine the p660 compatibility with the VTS/ATL architecture.
Web interaction	Ability of the solution to support functionality of the NSLDS Financial Aid Professionals (FAP) and Student Access websites as well as the migration of legacy CICS "green screen" functionality to a web-based format for interaction with NSLDS II.	Web interaction to the mid-tier ITA architecture will require the compression and decompression of data as it flows from the mid-tier ITA web architecture to the mainframe DB2 architecture. MIPS were allocated in the configuration to perform this function.	No compression or uncompression is needed for data reads or writes between the ITA and DB2 architecture therefore no additional capacity was configured for this processing.	No compression or uncompression is needed for data reads or writes between the ITA and DB2 architecture therefore no additional capacity was configured for this processing.
Batch timing	Ability of the solution to support a batch update window of not longer than 9: 00 PM Eastern to 6:00 am Eastern.	No clear advantages	No clear advantages	No clear advantages

Appendix A – Capabilities Comparison

Requirement	Definition	Mainframe z900 zSeries running DB2 on zOS	Mid-Tier clustered p690 running DB2 EEE on AIX	Mid-Tier clustered p660 running DB2 EEE on AIX
Procedures	Ability of the solution to support the execution of DB2 stored procedures.	No clear advantages	No clear advantages	No clear advantages
AMF				
CDR				
Federal Receivables				
LPIF				
Payment Reasonability				
Presceening				
Postscreening				
Student Transfer Monitoring				
FSA Architecture Compatibility	Overall ability of the solution to integrate with the existing FSA Integrated Technical Architecture environment and predicted future direction.	No clear advantages	No clear advantages	No clear advantages
MicroStrategy				
Informatica				
ITA Web Architecture (IHS/WAS/Interwoven)				
EAI Bus (MQSeries)				
Rational Suite				
Database Capability	Ability of the solution to support a 3rd normal form enterprise wide database and a dimensional data mart.	The mainframe will have inherently better I/O performance for a normalized operational data store, like the NSLDS II EDW, prior to any focused tuning or optimization activities.	The p690 configuration is sized to sufficiently meet the requirements of the NSLDS II EDW. However, in order to achieve optimal performance more tuning and optimization activities will be required than on the mainframe alternative solution.	The p660 configuration is sized to sufficiently meet the requirements of the NSLDS II EDW. However, in order to achieve optimal performance more tuning and optimization activities will be required than on the mainframe alternative solution.
EDW				
Data Mart		The mainframe configuration is sized to sufficiently meet the requirements of the NSLDS II data mart. However, in order to achieve optimal performance more tuning and optimization activities will be required than on the mid-tier alternative solutions.	The mid-tier will have inherently better I/O performance for a non-normalized dimensional data mart, like the NSLDS II data mart, prior to any focused tuning or optimization activities.	The mid-tier will have inherently better I/O performance for a non-normalized dimensional data mart, like the NSLDS II data mart, prior to any focused tuning or optimization activities.

Appendix A – Capabilities Comparison

This price comparison utilizes the Government Furnished Operations Costs and IBM provided DB2 Costs for each alternative platform.

Alternative Platform	FY03	FY04	FY05	FY06	FY07	Non Discounted 5 Year Prices
DB2 Costs						
Mid-Tier p660s (Revised)	$1,925,712	$459,988	$741,754	$856,449	$1,269,924	$5,253,828
Mid-Tier p690s	$1,283,808	$306,659	$576,925	$679,258	$793,696	$3,640,346
Mainframe zSeries z900 (Model 104)	Included in Operations Cost					
Operations Costs						
Mid-Tier p660s (Revised)	$4,564,320	$4,707,360	$6,203,040	$6,060,000	$7,126,560	$28,661,280
Mid-Tier p690s	$5,854,123	$6,438,080	$7,198,862	$7,961,237	$9,445,000	$36,897,302
Mainframe zSeries z900 (Model 104)+	$15,845,347	$16,966,715	$19,131,243	$22,657,304	$27,431,091	$102,031,700
Total Cost of Ownership						
Mid-Tier p660s (Revised)	$6,490,032	$5,167,348	$6,944,794	$6,916,449	$8,396,484	$33,915,108
Mid-Tier p690s	$7,137,931	$6,744,738	$7,775,787	$8,640,495	$10,238,696	$40,537,648
Mainframe zSeries z900 (Model 104)	$15,845,347	$16,966,715	$19,131,243	$22,657,304	$27,431,091	$102,031,700
+ Cost of DB2 Included						

Appendix B – Disk Space Details Mainframe

Time	NEW NSLDS II Potential MAINFRAME Production (EDW + Data Mart)							
Year	(S) = S + S*W Potential NSLDS II EDW Production DB Capacity Compressed	(T) = S Potential NSLDS II DataMart Capacity Compressed	(U) = U + U*W Potentail NSLDS II System Capacity and Temporary Capacity	(V) = V + V*W Interface and Procedure Files *5	(W) Potential NSLDS II DB Production Growth Rates *3	(X) = S + T + U + V Potential NLSDS II Total Production Projections (TB)		
2003	1.00	1.00	0.50	0.50	0.15	3		
2004	1.15	1.15	0.58	0.58	0.15	3		
2005	1.32	1.32	0.66	0.66	0.30	4		
2006	1.72	1.72	0.86	0.86	0.30	5		
2007	2.24	2.24	1.12	1.12	0.30	7		

Time	NEW NSLDS II MAINFRAME TOTAL Capacity (Development, Test1, Test2, Test3, Training/UA, Maintenance, and Production)							
Year	(1) Dev and Test Growth Rates *6	(2) = .25*X Dev	(3) = .25*X Test1	(4) = .25 * X Test2	(5) = X Test3	(6) = .50 * X Training / UA	(7) = .05 * X Maintenance	(8) = 1+2+3+4+5+6+7 Total for NSLDS II
2003	0.00	0.75	0.75	0.75	3.00	1.50	0.15	10
2004	0.00	0.86	0.86	0.86	3.45	1.73	0.17	11
2005	0.00	0.99	0.99	0.99	3.97	1.98	0.20	13
2006	0.00	1.29	1.29	1.29	5.16	2.58	0.26	17
2007	0.00	1.68	1.68	1.68	6.71	3.35	0.34	22

Appendix B – Disk Space Capacity Summary

Year	Production Environment ONLY		ALL Environments (Prod, Dev, Test1, Test2, Test3, Training/UA,	
	Total NSLDS II MAINFRAME Compressed (TB)	Potential NSLDS II MID_RANGE (TB)	Potential NSLDS II MAINFRAME Compressed (TB)	Potential NSLDS II MID_RANGE (TB)
2003	3	5	10	15
2004	3	6	11	20
2005	4	7	13	23
2006	5	9	17	29
2007	7	12	22	38

Appendix B – Disk Space Assumptions

- The NSLDS II Data Mart contains roughly 80% of the same tables that the EDW has and will approximate the same size of the EDW.
- The NSLDS II Enterprise Data Warehouse is a close approximation of the existing NSLDS database. Please see Data Mart Sizings. Raytheon has verified that of the 1 TB of database space allocated for NSLDS production 45% of this is compressed (or approximating 1.81 TB 1/.55 of disk space will be needed for NSLDS II production if compression is not used).
- The mainframe database environments are smaller than the mid-range environments are due to compression. Cliff Clemens the DBA from Raytheon states that the database is "running approximately 45% compression in all of the environments. It is noteworthy to mention that compression is at the tablespace level and some of the tables are compressed at 80% compression and some are compressed at 15%. Overall, the compression rate is about 45% which is typical and expected on IBM mainframes. It is noteworthy to mention that the compression is throughout the machine's data path (e.g. memory, channels, buffers, CPU, etc.) and not just on DASD. The sizes of non-production environments are not very significant (compressed or not)."
- CSC has verified that there is an automatic tape library with roughly 25,682 "live" tapes on NSLP. This represents roughly 1,284 TB as each tape is 50GB tapes that are neither deleted nor in scratch status. Jim Synard from Raytheon states the following: "Interface files are kept for four months on tape - not DASD." Additionally, "this space is not specific to interface files. It houses current interface files, user query results, sort space, utility space, program libraries, JCL libraries, and data sets set aside for analysis, i.e. just about everything other than the database, development, and testing space... Also, for capacity planning purposes, we have found that this space requirement has tracked well with the size of the database. So, if the database was projected to increase by 5%, the VDC would automatically add 5% to this non-database space as well."
- Raytheon has verified that the automatic tape library is growing at approximately the same rate as the NSLDS database.
- Raytheon has verified that there is roughly 10GB of data is written to the database a month. This is just an estimate which would approximate to about 120GB per year. Most of the data does not come from the web since the data comes from the Data Providers via Tape and the Network. There is a small percentage, probably less than 5%, of data that comes from the web. This translates to roughly 25 MB a day is written from the web site.
- Once the system is rolled out, there will be a 30% increase in capacity due to increased web usage and additional requirements being added to the system. (Note that much of the new functionality has been on hold due to the expectation of NSLDS II.)

Appendix B – Disk Space Assumptions Continued

- The Test2 environment will be able to replicate at least one full production environments for performance testing, regression testing, and system testing activities. It will be twice the size of production. A second test environment will be a subset of the production database size.
- Number of Databases
  - Development will be 25% capacity of Production (not 100% of the structure)
  - Test1 (Assembly & System) will be 25% capacity of Production (not 100% of the structure)
  - Test2 (Assembly & System) will be 25% capacity of Production (not 100% of the structure)
  - Test3 (Performance Testing, Conversion) will be 100% capacity of Production (100% of the structure)
  - Training / UA will be 50% of Production (100% of the structure)
  - Maintenance 5% of Production (100% of the structure)
  - Production 100%
- The growth rate flattens out over time as the system matures and less space is needed for development and the test environments.
- The Test3 environment for conversion and performance testing is temporary and will contract and expand.
- In 2005, NSLDS II is deployed and the growth rate jumps to 30%.
- RAID 5 is used for both Mainframe and Mid-Tier configurations. The capacity calculations are for usable space only. Additional space is allocated on pricing from the vendor to account for RAID 5 and hot spare requirements.

Appendix B – Disk Space Details Mid Tier

Time	NEW NSLDS II Mid- Range Production (EDW + Data Mart)								
Year	(H) = H + H*(L) Potential NSLDS II EDW Production DB Capacity Uncompressed	(I) = H Potential NSLDS II DataMart Size	(J) = J + J*L Potentail NSLDS II System Capacity and Temporary Capacity	(K) = K + K*L Interface and Procedure Files	(L) Potential NSLDS II DB Production Growth Rates	(M) = H + I + J + K Potential NLSDS II Total Production Projections (TB)			
2003	1.82	1.82	0.50	0.50	0.15	**5**			
2004	2.09	2.72	0.58	0.58	0.15	**6**			
2005	2.40	3.13	0.66	0.66	0.30	**7**			
2006	3.13	4.06	0.86	0.86	0.30	**9**			
2007	4.06	5.28	1.12	1.12	0.30	**12**			

Time	NEW NSLDS II Mid - Range TOTAL Capacity (Development, Test1, Test2, Test3, Training / UA, Maintenance and Production)							
Year	(N) Dev and Test Growth Rates	(O) = .25* M Dev	(P) = .25* M Test1	(Q) = .25 * M Test2	(R) = M Test3	(10) = .50 * M Training / UA	(11) = .05 * M Maintenance	(12) = N+O+P+Q+R+10+11+M Total for NSLDS II
2003	0.00	1.16	1.16	1.16	**4.64**	2.32	0.23	15
2004	0.00	1.49	1.49	1.49	**5.96**	2.98	0.30	20
2005	0.00	1.71	1.71	1.71	**6.85**	3.43	0.34	23
2006	0.00	2.23	2.23	2.23	**8.91**	4.45	0.45	29
2007	0.00	2.90	2.90	2.90	**11.58**	5.79	0.58	38

Appendix C – Mainframe Capacity Summary

NSLDS II Mainframe Capacity Numbers - MIPS

Time	MIPS Analysis				
Year	Production	MIPS Growth Rate	MIPS with Growth Rate	End of Year (Increase due to data mart doubled)	DASD Growth Rate
2002	361	0.00	361	722	0.15
2003	361	0.09	393	787	0.15
2004	393	0.09	429	858	0.15
*2005	429	0.18	506	1012	0.30
2006	506	0.18	597	1194	0.30
2007	597	0.18	705	1409	0.30

Time	Memory Analysis		
Year	Memory	Growth Rate	Total Memory
2003	10	0.00	10
2004	10	0.00	10
*2005	10	0.10	11
2006	11	0.10	12
2007	12	0.10	13

Time	DASD Analysis
Year	All Environments
2003	10
2004	11
*2005	13
2006	17
2007	22

The group agreed upon a peak NSLDS II I/O requirement of 500 MB per second. As configured, a Z900 can handle 24 GB per second.

*** Potential NSLDS II Deployment Date**

Appendix C – Mid Tier Capacity Summary

NSLDS II Mid Tier Capacity Numbers - TPM

Time	p660 TPM and Server Analysis						p660 CPU Analysis			p690 CPU Analysis		
Year	Production Beginning of Year (TPM)	Growth Rate for TPMs	End of Year Production TPM with Growth Rate	*Production Number of p660s - 6 M1 (105,000 TPMs per p660 - 6M1)	Number of p660s- 6H1 for other Environments	Total Number of p660s (6M1 and 6 H1)	Total Number of CPUs for Production p660 6M1 Environment (must be a multiple of 8)	*Number of CPUs for other Environments	*Total Number of CPUs	Total Number of CPUs for Production p690 Environment (2/3 the number for six p660 6M1 and must be a factor of 8)	Number of CPUs for other Environments	*Total Number of CPUs
*2003	500,000	9%	545,000	6	3	9	48	24	72	32	16	48
2004	545,000	9%	594,050	6	3	9	48	24	72	32	16	48
2005	594,050	18%	700,979	8	3	11	56	24	80	40	16	56
2006	700,979	18%	827,155	8	3	11	64	24	88	48	16	64
2007	827,155	18%	976,043	10	3	13	80	24	104	56	16	72

*CPUs must be an Even Number

Number of TPMs per p660 6M2 with 8 CPUs = 105,000

Number of TPMS per CPU = 13,125

Assume Maximum Number of CPUs for Dev and Test (8 CPUs per box)

The IBM's BISizer modeling tool indicates that the relative processing power of a p690 to six p660 6M1 servers is 2/3. Therefore, this sizing factor was used in determing CPU conifiguration and growth for the p690.

Time	Memory Analysis p660				
Year	Max Number of CPUs for Six p660 6M1 Servers for Production	Max Number of CPUs for Three p660 6H1 Servers for Dev/Test	Total Memory for Six p660 6M1 Servers for Production (2GB per CPU)	Total Memory for Three p660 6H1 Servers for Dev/Test (2GB per CPU)	Total Memory p660 Architecture
2003	48	24	96	48	144
2004	48	24	98	48	146
2005	56	24	114	48	162
2006	64	24	128	48	176
2007	80	24	162	48	210

Time	Memory Analysis p690				
Year	Max Number of CPUs for Two p690 for Production	Max Number of CPUs for One p690 for Dev/Test	Total Memory for Two p690 Servers for Production (2GB per CPU)	Total Memory for One p690 for Dev/Test (2GB per CPU)	Total Memory for the p690 Architecture
2003	32	16	64	32	96
2004	32	16	64	32	96
2005	40	16	80	32	112
2006	48	16	96	32	128
2007	56	16	112	32	144

The group agreed upon a peak NSLDS II I/O requirement of 500 MB per second. As configured, a p690 has an I/O bandwidth of 8 GB per second. With two machines, I/O approximates 16 GB per sec for production.

The group agreed upon a peak NSLDS II I/O requirement of 500 MB per second. As configured, a p660 has an I/O bandwidth of 1 GB per second. With six machines, I/O approximates 6 GB for production.

Appendix D - Deferred Items

- The impact to the FSA ITA environment of re-platforming the NSLDS web sites to Java,and moving ETL functionality to Informatica and reporting capability to MicroStrategy, was not explored in this analysis.
- There is a processing cost associated with transferring data to and from the Mainframe when using the ITA based web site, reporting structure, and Informatica tools. This analysis assumes that a MIPS increase of 10%is required to account for compression and decompression. A more thorough analysis may yield an increase or decrease in this need.
- It is possible to run IHS, WAS, ETL, and MicroStrategy Reporting on a single mainframe solution. This analysis did not examine this configuration and has focused only on housing the database on the mainframe or mid-tier.
- While total MIPS for the mainframe configuration was established, a more thorough analysis of the MIPS breakdown for each environment may yield a different distribution than was assumed for these 793 MIPS.
- Regardless of platform, the HSM tape store usage and capacity requires more detailed analysis to determine its impact on operational costs.
- The backup strategy for the mid-range servers will be engineered depending on the final solution. CSC needs to determine the appropriates standard for backing up the system.
- The conversion strategy will depend upon the platform selected. This architecture will be determined at a future time.
- Transfer Monitoring Policy – issues regarding the monitoring of school/student data by other schools and external parties were raised and should be reviewed by the policy team. (Jeff Baker's group).