16 Appendix F - Capacity Planning for NSLDS II

Please note that capacity estimates are still being reviewed and may changed based on different assumptions and requirements.

NSLDS II Assumptions

Number	Assumption
1	The NSLDS II Enterprise Data Warehouse is a close approximation of the existing NSLDS database. Plesae see Data Mart Sizings.
2	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. Please see the Data Mart Sizing Spreadsheet for additional detail.
3	The mainframe database environments are a fraction of what the mid-range environments are due to compression and smaller sample sizes. 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)."
4	Raytheon 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."
5	Raytheon has verified that the automatic tape library is growing at approximately the same rate as the NSLDS database.
6	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).

Number	Assumption
7	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.
8	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.
9	The growth rate flattens out over time as the system matures and less space is needed for development and the test environments.
10	The development environment is half of the production environment (compressed) as only a subset of data will be needed.
11	The growth rate for Development and Test1 environments will assume to be 10% as the environments are a subset of production.
12	The Test1 environment is equivalent to the development environment.
13	The Test2 environment will be able to replicate at least two full production environments for performance testing, regression testing, and system testing activities. It will be twice the size of production.

NSLDS II Capacity Summary

	Production Environment ONLY		ALL Environments (Production, UA, Testing, Development)	
Year	Total NSLDS MAINFRAME Compressed (TB)	Potential NLSDS II MID_RANGE (TB)	Potential NLSDS II MAINFRAME Compressed (TB)	Potential NSLDS II MID_RANGE (TB)
2003	1.95	4.45	11.60	17.82
2004	2.24	5.72	13.20	20.42
2005	2.58	6.75	15.01	22.92
2006	2.97	8.77	18.47	24.94
2007	3.41	11.41	22.95	27.58

NSLDS II Capacity Details

Time	EXISTING NSLDS Production System						
Year	(A) Current NSLDS Production DB Capacity Compressed at 45% (TB)	(B) = (A)/.55 Current NSLDS Production DB Capacity Uncompressed (TB)	(C) = C + C*E System and Tempory Capacity Compressed (TB)	(D) = D + D*E Interface and Procedure Files Compressed	(E) Current NSLDS Growth Rate 15% a Year	(F) = (B) + (C) + (D) Total NSLDS Production Capacity Not Compressed (Not Counting Development, Test, or Training)	(G) = (A) + (C) + (D) Total NSLDS Production Capacity Compressed (Not Counting Development, Test, or Training)
2003	0.95	1.73	0.50	0.50	0.15	**2.73**	**1.95**
2004	1.09	1.99	0.58	0.58	0.15	**3.14**	**2.24**
2005	1.26	2.28	0.66	0.66	0.15	**3.61**	**2.58**
2006	1.44	2.63	0.76	0.76	0.15	**4.15**	**2.97**
2007	1.66	3.02	0.87	0.87	0.15	**4.77**	**3.41**

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	1.73	1.73	1.00	1.00	0.15	**4.45**
2004	1.99	2.58	1.15	1.15	0.15	**5.72**
2005	2.28	2.97	1.50	1.50	0.30	**6.75**
2006	2.97	3.86	1.94	1.94	0.30	**8.77**
2007	3.86	5.02	2.53	2.53	0.30	**11.41**

Time	NEW NSLDS II Mid - Range TOTAL Capacity (Development, Test1, Test2 (training), and Production)				
Year	(N) Dev and Test Growth Rates	(O) Dev Capacity	(P) Test1 Capacity	(Q) = M*2 Test2 Capacity	(R) = M + O + P + Q Total for NSLDS II
2003	0.10	2.23	2.23	8.91	**17.82**
2004	0.10	2.45	2.45	9.80	**20.42**
2005	0.10	2.70	2.70	10.78	**22.92**
2006	0.00	2.70	2.70	10.78	**24.94**
2007	0.00	2.70	2.70	10.78	**27.58**

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	0.95	0.95	0.50	0.50	0.15	**2.90**
2004	1.09	1.09	0.58	0.58	0.15	**3.34**
2005	1.26	1.26	0.66	0.66	0.30	**3.84**
2006	1.63	1.63	0.86	0.86	0.30	**4.99**
2007	2.12	2.12	1.12	1.12	0.30	**6.48**

Time	NEW NSLDS II MAINFRAME TOTAL Capacity (Development, Test1, Test2 (training), and Production)				
Year	(1) Dev and Test Growth Rates *6	(2) Dev Capacity *7	(3) Test1 Capacity *8	(4) = Y*2 Test2 Capacity *9	(5) = M + O + P + Q Total for NSLDS II
2003	0.10	1.45	1.45	5.80	**11.60**
2004	0.10	1.60	1.60	6.67	**13.20**
2005	0.10	1.75	1.75	7.67	**15.01**
2006	0.00	1.75	1.75	9.97	**18.47**
2007	0.00	1.75	1.75	12.96	**22.95**