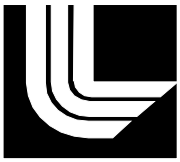

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-466881

US-VISIT Independent Verification and Validation Project:

Test Bed Establishment Report

N. W. Jensen, J. D. Gansemer

January 2011

Technical Report

submitted to the

U.S. Department of Homeland Security

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

US-VISIT Independent Verification and Validation Project:

Test Bed Establishment Report

submitted to the
U.S. Department of Homeland Security

LLNL-TR-466881

January 2011
Lawrence Livermore National Laboratory

Executive Summary

This document describes the computational and data systems available at the Lawrence Livermore National Laboratory for use on the US-VISIT Independent Verification and Validation (IV&V) project. This system – composed of data, software and hardware – is designed to be as close as a representation of the operational ADIS system as is required to verify and validate US-VISIT methodologies. It is not required to reproduce the computational capabilities of the enterprise-class operational system.

During FY10, the test bed was simplified from the FY09 version by reducing the number of database host computers from three to one, significantly reducing the maintenance and overhead while simultaneously increasing system throughput. During the current performance period, a database transfer was performed as a set of Data Pump Export files. The previous RMAN backup from 2007 required the availability of an AIX system which is not required when using data pump. Due to efficiencies in the new system and process, loading of the database refresh was able to be accomplished in a much shorter time frame than was previously required. The FY10 Oracle Test Bed now consists of a single Linux platform hosting two Oracle databases including the 2007 copy as well as the October 2010 refresh.

Table of Contents

1 Data and Cyber Security

1.1 US-VISIT Data

LLNL possesses two copies of the US-VISIT (ADIS) database utilized for verification and validation:

- The original US-VISIT data which contains the critical tables from the beginning of US-VISIT operations through November 2007 based on an RMAN backup provided by the sponsor organization and is approximately 4.5TB in size.
- A refresh of the ADIS database from August of 2010 (referred to as A2010) which is based on US-VISIT operations through late August 2010. The source database from the A2010 copy is approximately 7TB in size.

1.2 Data Security and Access Control

The computer hardware hosting the US-VISIT data at LLNL is physically located in a secure computing facility within LLNL's "Q" area, and only individuals with "Q" access (Top Secret clearances) are allowed into this building. Physical access to the computer operations room is further controlled by keycards with unique PINs, and unescorted access is limited to the System Administrators (SA's) who maintain the systems within the room.

The system is on LLNL's Yellow network. Network access to the Yellow network is limited to computers on site or accessed via an LLNL VPN account. All UNIX level access to the machines is controlled through individual username/password access. Currently only the System Administrator, the LLNL US-VISIT development team and the Database Administrator have accounts on the system.

The project teams access the database using individual Oracle userid/passwords. They access the data via Oracle's SQL*Net, typically using SQL*Developer or in-house java routines.

All project individuals are current on their US-VISIT privacy training.

2 Computer Environment

For FY10, the computer environment was simplified, reducing the number of hosts to one, saving both time and dollars. All of the AIX hosts were eliminated, leaving just the single Linux host. A new hardware RAID device was added and some of the older portable storage was retasked.

Hardware configuration:

Model:	HP ProLiant DL 180 G5 Server
Chassis:	rack mount
CPU:	2x Dual Core Intel Xeon E5205
Memory:	16GB total
Storage:	2x 146GB RAID1, 10x 750GB RAID5, 4x600G RAID10, 12x600G RAID5 (new)
Optical Storage:	N/A
Graphics:	N/A
OS:	RedHat 64bit Linux
Swap Space:	24GB

Local File Systems:

Filesystem	Size	Blocks Used	Available Blocks	Capacity Used	Description
/u01	75G	28G	46G	38%	SATA – RAID1
/u02	6.1T	3.4T	2.7T	56%	SATA – RAID5
/u03	10T	3.1T	7.0T	31%	Portable Storage
/u04	1.2T	314G	826G	28%	SAS – RAID10
/u05	6.6T	2.1T	4.3T	33%	SAS –RAID5

The original system has 12 internal disk drives. The operating system, Oracle software and a few critical database files are on /u01. The file partition /u01 is on a pair of 75G drives that are configured as RAID1. The remaining drives are all 750G drives. Eleven drives were used in a RAID5 configuration to create /u02. The remaining 750G drive is a hot spare.

This year an external hardware RAID system was added, a Winchester SX-3418R Controller with 16 x 600GB - 15KRPM Hitachi drives (HUS156060VLS600). In the FY09 analysis, it was determined that the majority of the queries used within the match algorithm packages were Oracle index and single table block accesses. To optimize performance in this context, the new disks are SAS drives spinning at 15,000 rpm drives, which are better suited for this type of work.

2.1 ADIS – FY09 Database

The database used for FY09 testing, based on the data transfer of November 2007, is still in place and available. The memory footprint was scaled down and the database now runs on the older, less efficient SATA disks. When this database is not needed it is shut down, allowing the newer database to fully utilize all the system resources.

The FY09 database is available for testing and gives the ability to compare performance based on improvements to the ADIS algorithms between 2007 and 2010.

2.2 A2010 – The Current Test Database

This database copy was created in late August 2010 and is based on a set of Oracle Data Pump Export files. The data physically arrived at LLNL in November 2010. The Data Pump Extract format greatly simplified the transfer process, making it possible to directly load the data into the Oracle 11g Release2 (11.2) Linux database. There was no need to load it onto an AIX system and then transfer the tables over to the Linux system as was required by the RMAN based backup performed in 2007.

A new database was created, putting files with high write activity onto the new SAS RAID10 partition with tablespaces for each of the larger tables on the SAS RAID 5 partition. Having only one mount point means that there is no need to create separate tablespaces for the indexes. Having each large table in a single tablespace allows transfer of a table to a different database utilizing Oracle's Transportable Tablespace technology.

Minor data compression was achieved, reducing the footprint of the database and improving the efficiency. The original tables will never be updated so there is no reason to leave any percent free (pctfree), saving approximately 10%. After the data is loaded, the tablespaces are changed to READ ONLY.

The system went live at the beginning of November, 2010.

2.3 Portable Storage

The only component of the Test Bed that has not been mentioned is the Portable Storage. This is the unit that was shipped to the DOJ data center and it still contains the original Oracle Data Pump Files. In addition to these files, the 10TB drive is being used for storage of the RMAN backups from both databases.

Appendix A – Test Bed Establishment Timeline

Following is a high-level outline of activities related to hardware procurement, data transfer, hardware/software installation and data installation for the 3 US-VISIT systems at LLNL.

August 2007 Network Attached Storage Devices (NAS) devices were procured for the purpose of data transfer of the 4.5TB database hosted at DOJ.

September 2007 Due to privacy concerns, it was determined that data encryption of the databases backup was required in order to ship the data from the DOJ datacenter to LLNL. GnuPG was chosen after consultation with US-VISIT privacy officials. Test files were sent between LLNL and DOJ staff to test the encryption/decryption procedure.

January 2008 US-VISIT data arrived at LLNL from the DOJ datacenter and decryption of the 5TB data transfer is initiated.

February 2008 Initial decryption of the data completes. The process of importing the data into the LLNL system begins, but is unsuccessful due to hardware incompatibilities. Several days of testing suggests that an IBM AIX based machine is required to host the RMAN backup files. An IBM p620 was retasked for this purpose, but it is quickly determined that this system was underpowered as an operational system although provides proof of concept for resolving the data import issues. A new IBM p505 server is ordered

March 2008 Corruption of the encrypted files is discovered requiring a second decryption of files. Work continues on the retasked IBM p620 and Oracle 9i loaded.

April 2008 The new IBM p505 arrived without an OS installed due to miscommunication between the LLNL procurement department and the vendor.

May 2008 The OS is installed on the IBM p505. Attempts to install Oracle 9i ran into java and NFS issues.

June 2008 FY07 US-VISIT funding expires, requiring team members to be assigned to other projects.

October 2008 FY08 funding resumes, however, because of the 4 month lapse in funding, original project staff are unavailable. A new team is identified and brought on board including a computer scientist, a statistician and a database administrator. Final work on the IBM p505 resumes. Work with US-VISIT data is on hold pending privacy training of new staff.

November 2008 Privacy training for new project staff is completed and they are able to work with the data. Databases on both IBM machines are operational and performance testing begins. Due to instability issues with NFS mounting of the NAS to the IBM p505, performance is determined to be insufficient to support project scope and schedule. Order placed for a HP PC with ~6TB local RAID storage.

December 2008 New hardware arrives and software loaded. Data transfer and conversion (endian) started.

January 2009 New database available for the teams use by the end of the first week of January 2009

July 2010 NAS delivered to DOJ Data Center to facilitate transfer of database update

October 2010 NAS chassis and disks containing database refresh received at LLNL. Data stored as Oracle Data Pump Export.

November 2010 Database refresh online and available for use at LLNL.