

---

# National Patient Information Reporting System: National Data Warehouse

---

## **NDW Data Mart Test Environment**

### **Technical Guide**

Version 1.1

April 2008



Department of Health and  
Human Services

Indian Health Service

Office of Information  
Technology (OIT)

---

# Contents

<b>Version Control</b> .....	<b>iii</b>
<b>Overview</b> .....	<b>1</b>
<b>Design Parameters</b> .....	<b>1</b>
<b>System Environment</b> .....	<b>2</b>
<b>Security</b> .....	<b>3</b>
<b>Test Environment Data Mart Design</b> .....	<b>4</b>
Structure .....	5
Extract/Import Processes .....	6
Federation .....	7
Backups.....	8

## Version Control

Version	Date	Notes
1.0	September 2007	FY07 Contract Deliverable (D1.12.9) <b>Note:</b> Appendix C - Test Environment Data Mart Column Detail is a separate document, in PDF format. Accepted on Nov 7, 2007
1.1	April 2008	Annual Review - no content changes, appendices removed, and Appendix A: Test Environment Data Mart Schemas and Tables / Views / Nicknames, Appendix B: Test Environment Data Mart Physical Models now separate documents; Appendix C: Test Environment Data Mart Column Detail now available at NDW Metadata web site, as noted in text. FY08 Bridge Contract Deliverable (D1.7.8) approved 4/29/08.

## Overview

The NDW Data Mart Test environment (RTE66 database) contains all of the target and reference tables in the National Data Warehouse (NDW). The Data Mart Test environment allows an authorized user to evaluate the type and structure of data, and to become familiar with the capabilities of the NDW, before accessing the General Data Mart or requesting a new data mart. It further allows a user to become familiar with appropriate and efficient methods of retrieving data from a relational database.

The NDW Data Mart Test environment is essentially a mirror of selected data from the NDW Production environment, and is available for use by properly approved users who are experienced in reviewing this type of data. The data is a selected subset of the NDW Production environment for evaluation purposes only. From time to time, it may become necessary to test specialized subsets of data, modified tables, and/or views as users develop conceptualized approaches to new data marts.

### **IMPORTANT**

The Data Mart Test environment contains a copy of **LIVE** production data.

No scrambling, cleansing, encryption; or other methodology is used to disguise patient identifiable data.

## Design Parameters

- The Data Mart Test environment exists on a separate server from the NDW Production environment.
- The data in the Data Mart Test environment database is a copy of live production data. It does not use any scrambling, cleansing, encryption, or other methodologies to disguise patient identifiable data.
- All data within the Data Mart Test environment database is usually refreshed monthly. (as stipulated in the *Service Level Agreement - NDW Data Mart Test Environment*).
- The NDW Program Manager, working in collaboration with the NDW data owner, will specify to the contractor those who will be granted access to the Data Mart Test environment, as well as the time period during which they will be granted access.

- Only authorized users will be allowed access to the Data Mart Test environment.
- Access controls are administered that allow users to see (query) data that is appropriate to their authorized level of access.
- Security controls commensurate with those in the NDW Production database and adhering to IHS standards, as outlined in separate security documents, are enforced.
- Availability to users is maximized (restricted only to the time it is shut down for data refresh, system maintenance, etc.). This includes 24/7 availability, although support is limited to normal business hours (refer to *Service Level Agreement - NDW Data Mart Test Environment*).
- The Test Data Mart is enterprise compliant, to allow various environments to access the database, including ODBC, JDBC, OLE, CLI. These are some of the environments and protocols that may be used to access this data mart, depending on user needs, access protocols, and environments.
- SAS DB2 connect software has been installed on the server that houses the Data Mart Test environment database as provided by the Indian Health Service.

## System Environment

This is the physical environment of the NDW Data Mart Test environment:

Server:	BILBO, 64bit
Database:	RTE66
Version	DB2 Version 8.2.8 (64 bit)
Partitions	2
AIX Version:	5.2
FTP Address	198.45.1.8
System Access:	Enterprise compliant to allow various environments to access the database, including ODBC, JDBC, OLE, CLI.

For a detailed list of Data Mart Test environment related schemas and tables, see the following documents:

- *NDW Schemas and Tables/Views/Nicknames*
- *Reference Tables*

See the *NDW Physical Models* document for corresponding models.

Detailed descriptions of data elements are available at the IHS Meta Data internet web site: <http://www.ihs.gov/CIO/scb/metadata/>

## Security

- The Data Mart Test environment is limited to authorized personnel.
- The NDW Program Manager, working in collaboration with the NDW data owner, will specify to the contractor those who will be granted access to the Data Mart Test environment, as well as the time period during which they will be granted access.
- Access controls are administered that allow users to see (query) data that is appropriate to their authorized level of access.
- Only authorized users will be allowed access to the Data Mart Test environment.
- Security controls commensurate with those for a transactional, non query based database and adhering to IHS standards, as outlined in separate security documents, are enforced.

## Data Mart Test Environment Design

The following figure displays the schemas of the Data Mart Test environment.

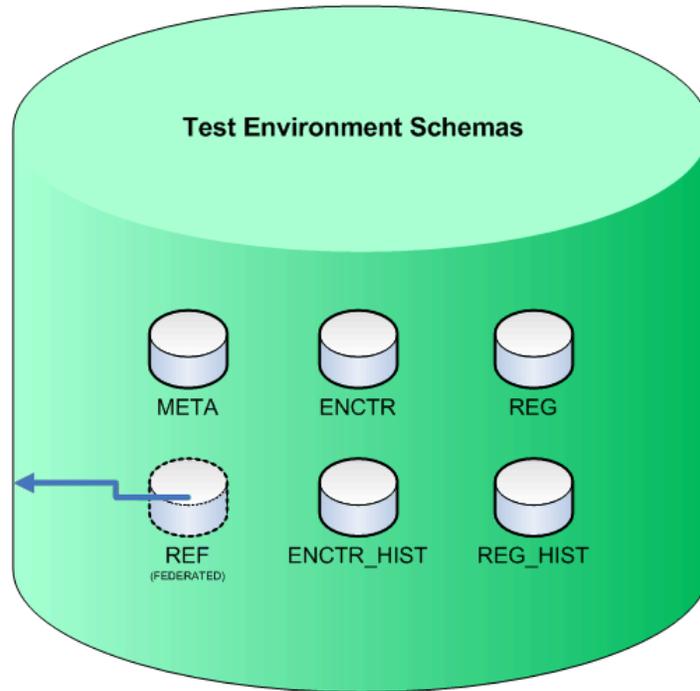


Figure 1. Data Mart Test Environment Schemas

Each schema is typically assigned its own pairing of tablespaces. The Data Mart Test environment is multi-partitioned into two nodes to maximize performance. A node is logical and physical partition within the database that enhances performance by promoting CPU parallelism and opportunistic behavior.

For a detailed list of NDW related schemas and tables, see the following documents:

- *NDW Schemas and Tables/Views/Nicknames*
- *Reference Tables*

## Structure

These are the main structure types associated with the Data Mart Test environment.

### Table

A grouping of data having the same construct; usually, but not always, keyed and indexed. Target tables are considered the factual data about the NDW. All other tables exist in support of the Target tables (Registration and Encounter schemas).

### Indexes

A physical construct to allow rapid access to data in tables.

### Tablespaces/Normal

These are actual physical storage area used by tables, and temporary tables. Each tablespace has its own memory buffer pool assigned to minimize contention. A tablespace is associated with a buffer pool (allocated memory) and inherits certain characteristics from the buffer pool, in particular, the page size. Tablespaces can either be system managed or database managed. System managed tablespaces grow automatically; database managed tables must be pre-allocated but are faster. All normal tablespaces are database managed.

## Tablespaces/System

These are actual physical storage area used by views, sorts, joins, and internal operations. This storage is both pre-allocated and dynamically allocated. A tablespace has its own memory buffer pool assigned to minimize contention. A tablespace is associated with a buffer pool (allocated memory) and inherits certain characteristics from the buffer pool, in particular, the page size. Tablespaces can be either system managed or database managed. System managed tablespaces grow automatically. Database managed must be pre-allocated, but are faster. Most system tablespaces are database managed.

## Extract/Import Processes

To ensure that marts are refreshed with a proper snapshot of the database, all extract/import processes involve at least the following.

- For Encounter (ENCTR) related data, a snapshot of either the desired ENCTRSS\_ID or EXPORT\_ID is stored in a temporary table. Extracts from all ENCTR schema tables are made from this temporary table. This process allows continued processing of incoming files while eliminating partial file extracts to a mart.
- A similar process is used for Registration (REG) related data.
- All extracts use a “with ur” (dirty read) operation to avoid conflicts with transactional processing.

Marts directly supported through an ETL/TEL process include the Data Mart Test environment (RTE66), General Data Mart (HOLLYWD), Data Quality Mart (WILDHRSE), Export Tracking Mart (WILDHRSE), NDW Computation Mart (WILDHRSE), and the User Population/Workload Reporting Mart (TATONKA).

Extracts are performed from within the NDW Production environment to flat files for loading to the various marts.

The extract processes utilized for the data transfer are of the following types:

1. EL (Extract, Load). The tables inside NDW are extracted into a flat file for subsequent load into another mart.
2. TEL (Transform, Extract, Load)

- a) A variation of the EL process, TEL builds a flat file from a view or MQT (Materialized Query Table) that pre-exists by performing calculations or transformations with or without joins and/or OLAP (On Line Analytical Processing).
- b) Builds a temporary table in SANDIA with calculations and transforms (with or without joins and/or OLAP), and then extracts these to another mart for load.
- 3) ETL (Extract, Transform, Load) is an extract process, then a transform at the receiving process, prior to or during the load. This is often done with third party software.
- 4) TTEL (Transform/Transform, Extract, Load) builds a temporary table in the database with calculations or transforms (with or without joins and/or OLAP), updates the table, then extracts these to another mart for load; then potentially, there is further updating of the table in the mart.

Most extract/import processes are controlled by cron scripts, and most processes are re-entrant processes. Re-entrant processes are processes that can be restarted from a failure point, if they fail to run for some reason such as a conflict, resource, or storage issue. Processes not currently controlled by cron or not re-entrant are being modified to comply with this behavior.

All extract processes are SQL based processes with simple AIX shell script wrappers.

## Federation

A federated process is a DB2 process of connecting other databases or data sources through defined connections called cservers or Distributed Relational Database Architecture (DRDA) on an enterprise system. DRDA is a set of protocols, or rules, that enable a user to access distributed data regardless of where it physically resides. It provides an open, robust heterogeneous distributed database environment. DRDA provides methods of coordinating communication among distributed locations. This allows applications to access multiple remote tables at various locations and to have them appear to the end user as if they were a logical whole.

A distinction should be made, however, between the architecture and the implementation. DRDA describes the architecture for distributed data and nothing more. It defines the rules for accessing the distributed data, but it does not provide the actual application programming interfaces (APIs) to perform the access. So DRDA is not an actual program, but is more like the specifications for a program.

Additional information on federation available at this website:

<http://www.redbooks.ibm.com/abstracts/sg244249.html>

DB2 is a DRDA-compliant RDBMS product, that is, it follows the DRDA specifications. DRDA is supported and certified by The Open Group only on DB2 6.5 and higher, Informix 11 and higher, and Oracle Gateway for DB2.

Reference Tables are currently federated through a nickname. The Federation, from the Temecula database, is instantaneous. However, due to a limitation in DB2 8.1.5, Federated tables can not be read directly except by the Instance owner, requiring a separate update process by a daily Instance owner cron job, to refresh physical reference tables. This step will be eliminated in a future update.

## **Backups**

Backups are not performed on the Data Mart Test environment, since this data mart is neither critical nor transactional. If the mart becomes corrupted or is lost, it can be reconstructed using the extract/import process defined in the “Extract/Import Processes” section.