KDOT's Spatially Enabled Data Warehouse

Paul Bayless - KDOT Data Warehouse Manager

and

Bill Schuman - GeoDecisions Project Manager
Goals of the Session

- Describe what a data warehouse is and why it is of value over a simple database or repository.
- Familiarize session attendees with data warehousing terminology.
- Provide an example of GIS and LRS being used in data warehouse staging processes.
- Explain of how the data warehouse is used by GIS-enabled applications.
Simple Data Warehouse Concept

- General Maintenance Systems
  - Crash Records
  - Bridge
  - Pavement
  - Financial

- General Decision Support
  - Information Portal
  - Dash Board
  - Specific Analysis Systems
  - Pavement Lifecycle Analysis

Data Warehouse

Derived Data Values from Integrated Data
Data Warehouse Data Flow

On-Line Transaction Processing (OLTP) Environment

Data Maintenance Databases
- Pavement
- Crashes
- Environmental
- Financial
- Rail

Data Preparation/ Staging “ETL” Processes

On-Line Analytical Processing (OLAP) Environment

Data Warehouse Environment

Integrated Data Warehouse

KANSAS DEPARTMENT OF TRANSPORTATION

GeoDecisions®
OLTP vs. OLAP

**OLTP**
- Represents system where data is maintained
- Typically the data is normalized in order to simplify data maintenance
- Users are experts on the data and collection
- Focused single business need / reporting

**OLAP**
- Represents system where data is queried
- Typically data is denormalized in order to simplify data analysis and understanding
- Users are business experts that need information
- Focused on enterprise business analysis
User Interfaces to Warehouse

Human Resources Dashboard - Benefits
Last Updated on 12/31/06

EMPLOYEE CLAIMS
2005 Actual 2006 Actual 2006 Budget

EMPLOYEE COMPENSATION CLAIMS

LEAVE OF ABSENCE

KANSAS
DEPARTMENT OF TRANSPORTATION
User Interfaces to Warehouse
A Data Warehouse is...

- A query-able/researchable source of data
  - Enterprise data integration
  - Data structured for query, not maintenance

- A corporate effort
- Centrally coordinated
- Arranged around the corporate data model
- Built iteratively
- Structurally sound
- Optimized for high performance analysis
A Data Warehouse contains...

- Back-end raw material for decision support
- Technology suited for high volumes of data
- Centralized data which represents corporate understanding of business domain
- Data stored at lowest level of granularity
- Optimized for high performance query-based analysis
- Copious historical data
- Metadata, metadata, metadata
Getting into the details...
DW Development Lifecycle (Ralph Kimball Methodology)

- Business Requirement Definition
- Technical Architecture Design
- Dimensional Modeling
- Physical Design
- Product Selection & Installation
- Data Staging Design & Development
- End-User Application Specification
- End-User Application Development
- Deployment
- Maintenance and Growth
- Project Management
What is Data Staging?

Extract, Transform, and Load (ETL)
Data Warehouse Staging

On-Line Transaction Processing (OLTP) Environment

Data Maintenance Databases

- Crashes
- Environmental
- Financial
- Rail

Data Preparation/ Staging “ETL” Processes

- “Extract” Data from Maintenance System
- “Transform” Data and Quality Check
- “Load” Data into Data Warehouse

On-Line Analytical Processing (OLAP) Environment

Data Warehouse Environment

Integrated Data Warehouse
Data Staging - Traditional

**Extracting**
- Reading and copying the needed data to a data environment for further work.

**Transforming**
- Clean data (spelling, spaces, data conflicts)
- Additional purging of proprietary data
- Combining data (text equivalents for codes, denormalizing data, etc.)
- Creating surrogate keys
- Spatial processing / geometry creation
Data Staging - Traditional

**Loading**
- Indexing the data for performance.
- Quality assurance checks.
- Formal publication with other data marts.
ETL is Different for Spatial Data

- Traditional staging tools haven’t supported “geometry” fields
- Geometry fields in the database have some unique database indexing needs
- GIS is traditionally feature/table based, data warehouse models are star or snowflake data structures
- Geometry fields not traditionally considered part of a multi-dimension OLAP (MOLAP) cube
Staging - Spatial/Linear Data
(All traditional sub-processes plus...)

- **Extracting**
  - Validating projections, datum, units, 2D/3D coordinates, and/or linear referencing methods (LRM) between business data and the base map.

- **Transforming**
  - Converting LRM to common location method (including temporal matching).
  - Creating geometries from business data location.
    - Geocode coordinates or addresses
    - Dynamic segmentation of linear/point location

- **Loading**
  - Spatial indexing data
Example of Spatial Staging - Source Data

Pavement Marking Database (PMDB)

- OLTP System
- No map - only manual data entry/truck entry
- Linear reference or GPS location data

![Image of Pavement Marking Database (PMDB) data entry interface]
### Example of Spatial Staging Extract Process

#### Location in Statewide Reference Post LRM
- Users see it in the field
- Reference posts are tied to LRS in CANSYS
- LRM “published” yearly and loaded to Oracle data warehouse for transformation procedures

### Pavement Marking Stripe Placement Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>County</td>
<td>Shawnee</td>
</tr>
<tr>
<td>District</td>
<td>1</td>
</tr>
<tr>
<td>Road Name</td>
<td>1070-EB(S)</td>
</tr>
<tr>
<td>Begin Reference Post</td>
<td></td>
</tr>
<tr>
<td>End Reference Post</td>
<td></td>
</tr>
</tbody>
</table>

### Pavement Marking Reflectivity Values

![Data warehouse chart with reflectivity values]
Example of Spatial Staging
Transform/Load Processes

- **Stripe Attribute Values**
  - County: Shamrock
  - District: 1
  - Road Name: 1074-EB(S)
  - Begin Reference Post: 500
  - End Reference Post: 1500

- **Reflectivity Attribute Values**

- **GIS Striping Project with Reflectivity Summary**

- **Extract Processes**

- **Data Warehouse (Staging)**
  - OWB, Oracle Spatial, and Oracle LRS Transform Processes

- **Load Process**

- **Data Warehouse (Production)**

- **Data Mart / Views for GIS Features (Production)**
Example of Spatial Staging
Spatial Transformation
Example of Spatial Staging
Spatial Transformation Process - Roads

- Base map loaded to staging environment

SQL> INSERT INTO lines VALUES ( attribute_1, ..., attribute_n, mdsys.sdo_geometry ( 2, null, null, mdsys.sdo_elem_info_array (1,2,1), mdsys.sdo_ordinate_array ( 10,10, 20,25, 30,10, 40,10)) );
Example of Spatial Staging
Spatial Transformation Process - LRS

- Identify route and mileage fields in the centerline attributes (format and units).
- Create “routes” in the Oracle structure using Oracle LRS.
- Dynamically segment pavement marking locations along road centerline / routes.
- Populate geometry in pavement marking project star schema.
Example of Spatial Staging
Spatial Transformation Process - Geocode

- Identify latitude and longitude fields in the retro-reflectivity attributes (format and units).
- Create geometries for the pavement retro-reflectivity locations.

<table>
<thead>
<tr>
<th>COORD ROUTE LRM KEY</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOCATION NM</td>
</tr>
<tr>
<td>NETWORK DIRECTION</td>
</tr>
<tr>
<td>BEGIN LATITUDE</td>
</tr>
<tr>
<td>END LATITUDE</td>
</tr>
<tr>
<td>END LONGITUDE</td>
</tr>
<tr>
<td>BEGIN LONGITUDE</td>
</tr>
<tr>
<td>LOCATION ID</td>
</tr>
<tr>
<td>PROJECTION</td>
</tr>
<tr>
<td>LOCATION TYPE</td>
</tr>
<tr>
<td>LRS GEOMETRY</td>
</tr>
<tr>
<td>RPOST LRM NETWORK YEAR</td>
</tr>
</tbody>
</table>
Example of Spatial Staging
Spatial Transformation Process - Aggregation

- Identify retro-reflectivity locations along a pavement marking stripe (spatial or LRS).
- Averages the retro-reflectivity along a pavement marking stripe (considers different colors, materials, etc.).
GIS Enabling the Dimensional Model

Create materialized views in Oracle

- Generate “feature-level” granularity
- Create a relational looking table for the GIS
- Populate SDO information (projection, datum, units, etc.)
- Create spatial indexing

Ensure materialized view is defined as a feature layer

- Populates GeoMedia metadata tables
User Interfaces to Warehouse
User Interfaces to Warehouse
Summary

- Important to differentiate the modeling of data for maintenance versus query/analysis.
- Data warehousing often makes use of dimensional versus relational modeling.
- Spatial data is not always a priority of the off-the-shelf data staging tools.
- Data warehouse staging processes are ideal for generating geometry from attribute information.
- Spatial and LRS capabilities of databases starting to make formal spatial ETL processes more efficient.