Data Vault Modeling & Methodology

Business Side and Introduction
Introduction / Expectations

• Who’s attending?
• What do you expect to learn from this presentation?
• What is your background (in data warehousing)?
• What is your experience with data modeling?
What is a Data Vault?

- **Data Vault Model**
  - Detail oriented
  - Historical traceability
  - Uniquely linked set of normalized tables
  - Supports one or more functional areas of business

- **Data Vault Methodology**
  - CMMI Level 5 Project Plan
  - Risk, Governance, Versioning
  - Peer Reviews, Release Cycles
  - Repeatable, Consistent, Optimized
  - Complete with Best Practices for BI/DW
Who’s Using It?
What Does One Look Like?

Elements:
- Hub
- Link
- Satellite

Hub = List of Unique Business Keys
Link = List of Relationships, Associations
Satellites = Descriptive Data
What are the Issues Today?

Business...
- Changes Frequently
- Needs Accountability
- Demands Auditability
- Has No Visibility
- Wants More Control

IT...
- Takes Too Long
- Is Over-budget
- Too Complex
- Can’t Sustain Growth
EXAMINING THE ISSUES

Business to IT to EDW to BI – Gap Analysis
Gap Analysis, Incorrect Graph

Business Process says…
We are running “X”

GAP#1
(perception based)

Business User says…
I think we are running “K”
and Collecting “O”

GAP#2
(Data & Process Based)

Business User finds out:
Gee, I have to FIND the GAP, and
Now I have to close them all???
I CAN’T RECONCILE MY
BUSINESS PROCESSES TO MY
WAREHOUSE!

Operational System says…
We are collecting “Y”

GAP #3
One-Way Data!
(Changes to Data by
Business Rules make it
Impossible to reconcile!)

History in Warehouse
says: “1” – 5 yrs ago
“G” – 2 yrs ago
“X” – last year
“Y” – last month
“Z” – now…
Gap Analysis, Triangulation

Business Process says…
We are running “X”

Raw Data Vault
History in Warehouse says:
“1” – 5 yrs ago
“G” – 2 yrs ago
“X” – last year
“Y” – last month
“Z” – now…

Operational System says…
We are collecting “Y”

GAP #2
(Data & Process Based)

GAP #3, COVERED!
New: Two Way Data
(Data is RAW on the way in to the EDW, and therefore can by synchronized with the Operational Systems)
ETL Automation

We *should* be able to reduce the ETL workload *by at least 80%* through *auto-generation* of base-line ETL processing!

In order to succeed, we *MUST HAVE* the following:

- Common Design Patterns based on **TARGET** data models/table types
- Business Rules moved *downstream* of the data warehouse
- Generic Templates for each **pattern** that we are interested in
- Automatic Impact Analysis and re-generation (based on new changes)

**Does such a solution exist?**

**Yes!!!** (see me after the presentation…)

*But it requires a RAW Data Warehouse view of the world*

(Data Vault OR Star Schema based)
Why business drove the need to build Data Vault Models

DATA VAULT IN 5 EASY STEPS
The Environment

- SEI/CMMI Level 5 Push
- Lean Initiatives
- Cycle Time Reduction
- KPA’s/KPI’s

7 Sectors Of Business

53 Independent Companies
Each managed by their own P&L

Enterprise Info Systems (Charge Number Required to do business)

Me…

$\$ Project Funding Levels
Directed through specific companies to fund Specific IT teams
The Challenge

Build an Enterprise Data Warehouse collecting manufacturing, planning, rework, contracts, HR, and financial information from different sectors, different companies.

The Details:
- 3 FTE’s: 1 project manager/data architect/ETL coder, 1 SME (Cobol expert), 1 DBA
- 1 PT: Data Modeler
- 15 business users (middle management)
- 1 PMO office – assigned PMO officer
- Can’t get time from IT employees without a Charge Number
- 23 disparate source systems
- 500+ business reports
- 150+ ad-hoc users
- 1000 blue-collar workers receiving reports daily

Timeline:
- 6 months start to finish, phase I production release
- Person-Hours estimated on project: 3000
- Person-hours Actually consumed: 3250
The Requirements

Government Contractor

**Commercial (Unclassified)**
*Internal + External Communication*

- Real-Time (*latency in seconds*)
- Batch (*multi-arrival across time zones and international boundaries*)

National Security Agency

**Classified Systems**
*(Private/Protected)*

- Inherit 100% of Changes (*both data & model & loading processes*) Without disturbing existing history
- Expandable with Classified Information (*bolt-on solution*)
- Protected / Secured

- **EDW Must Be:** Accountable, Auditable, Compliant (*SOX, HIPPA, Gov’t Regulations*)
- **Project Level:** Repeatable, Consistent, Measurable, Automatable
Step 1: Load & Design DV

Government Contractor

Commercial (Unclassified)
Internal + External Communication

Classified Systems
(Private/Protected)

Real-Time Feeds
(transactions)

Batch Load
(may affect entire model)

LEGEND:
• H = Hub
• L = Link
• S = Satellite

✓ Real-Time (latency in seconds)
✓ Batch (multi-arrival across time zones and international boundaries)
✓ BOTH LOAD TYPES ACCOMODATED TO THE SAME MODEL AT THE SAME TIME
Step 2: Inherit DV in Classified

Government Contractor

Commercial (Unclassified)
Internal + External Communication

- Replicate: Data Model, Data Sets, ETL, SQL, Loading Processes, BI Reports

- Inherit: Future changes to unclassified world without disturbing HISTORY in Classified Systems

National Security Agency

Classified Systems
(Private/Protected)
Step 3: Extend DV In Classified

Commercial (Unclassified)
Internal + External Communication

Classified Systems
(Private/Protected)

Expandable! Add new elements in the classified world while maintaining ALL elements inherited from unclassified world

• ZERO Re-Engineering Impact
Step 4: Extend DV In Unclassified

Government Contractor

Commercial (Unclassified)
Internal + External Communication

National Security Agency

Classified Systems
(Private/Protected)

• ZERO RE-ENGINEERING TO ABSORB BUSINESS CHANGES & NEW SYSTEMS

✓ Inherit: Future changes to unclassified world without disturbing HISTORY in Classified Systems
Step 5: Secure DV in Classified

**Government Contractor**

**Commercial (Unclassified)**

*Internal + External Communication*

- Data Flows 1 way in
- Machine is shut-down, IP reset, machine re-booted

**National Security Agency**

**Classified Systems**

*(Private/Protected)*

- Allow electronic transfer from low privacy to high privacy, one-way push only, protect high-side assets

Separate Secured Systems
Accountability and Auditability take FRONT SEAT!

PARADIGM SHIFTS IN BUSINESS LOGIC
**EDW: Old Way**

**Corporate Fraud Accountability** Title XI consists of seven sections. Section 1101 recommends a name for this title as “Corporate Fraud Accountability Act of 2002”. It identifies corporate fraud and *records tampering* as *criminal offenses* and joins those offenses to specific penalties. It also revises sentencing guidelines and strengthens their penalties. This enables the SEC to temporarily freeze large or unusual payments.

Are changes to data *ON THE WAY IN* to the EDW equivalent to *records tampering*?
EDW: New Way

Compliant and Auditable System Of Record!!

1. Implement a Raw Data Vault Data Warehouse
2. Move the business rules "downstream"

SOX Compliance (for DATA only) is MET!
BUSINESS CONCLUSIONS

Why and when to use a Data Vault, When NOT to use a Data Vault
When NOT To use a Data Vault

If What you Have Is Working For you...

DON’T CHANGE IT!
Why Use a Data Vault?

• Need Auditability and Accountability
• Require Real-Time Data Flow mixed with Batch
• Have massive amounts of data to manage
• Need Quicker Turn-Around Time
  – (to implement business changes, deliver marts)
• Must have flexible absorption of new systems data
• Deliver on-time & in-budget
Objective Of Data Vault

Achieve a HORIZONTAL View of the business (cross-lines of business)

Business Keys
Span / Cross
Lines of Business
Questions?

Dan Linstedt
President, Empowered Holdings, LLC
http://EmpoweredHoldings.com
http://DanLinstedt.com
Tel: +1 802-524-8566
E-Mail: danLinstedt@gmail.com

SERVICES:
• Consulting
• Assessments
• Product Selection Scorecards
• Architecture / Design
• Mentoring and Workshops (training)