AGILE DATA MODELING WITH DATA VAULT
Scot J Reagin, CBIP, CDVDM

- Master Builder, Sensible Data Integrations Inc.  
  www.SensibleDataIntegrations.com
- TDWI Certified Business Intelligence Professional
- Board Member, Colorado TDWI http://tdwichapters.org/colorado
- Certified Data Vault Data Modeler http://geneseeacademy.com/
- scot@SensibleDataIntegrations.com
- 303.378.4834
Overview:
- Agility in Business Intelligence
- Data Vault Data Modeling

Activity:
- Updating an EDW Modeled in 3NF
- Updating an EDW with Data Vault
Value is created by quickly responding to changing business needs
  • But…Data needs to be persistent and auditable
The EDW needs to store all data of INTEREST to the business
  • But…the time and cost to add new data needs to be managed
A long-term VISION exists for the organization’s information needs
  • But…that vision will evolve over time in ways we cannot currently predict
- Culture
- Inflexibility of Methodology
- Lack of Executive Support
- The EDW
  - In BI the EDW is often the largest single barrier
    - Long update cycles
    - Inflexible Design
    - Significant Cost of Updates

BARRIERS TO AGILITY IN BI
3rd Normal Form
- Optimal for Operational Systems
- Heavily used in ‘traditional’ EDW
- Minimizes data storage for relatively static data sets

Dimensional Modeling
- Optimal for Data Marts
- Favors query performance over storage efficiency
- **Note: the Data Mart and EDW are not the same thing!**

Data Vault
- Leading form of Ensemble Modeling
- Optimal for EDW
- Favors flexibility over storage efficiency and simplicity
- Simplifies and Optimizes Data Query for Analysis and Reporting
- Abstracts the User/Application from the EDW

DATA MART
3rd NORMAL FORM
DATA VAULT MODEL COMPONENTS
DATA VAULT MODEL EXAMPLE
MANAGING CHANGE

- Schema Changes
  - Additional Attributes
  - Changes in Data Grain

- Auditing and Historical Data
  - Retain only current values (Type I) or Historical values (Type II)

- Changes in Source Systems
  - Legacy Systems retired but data needs to remain
  - New systems have different schemas and/or data types

- ETL Updates
- Testing/Validation
LOE
- Schema Updates
- Data Migration
- ETL Updates
- Testing

Scenario #1
- Add Customer Type
- Oops, and Sub Type

Scenario #2
- Customer has 3 Addresses

Type II?

DISCUSSION
DISCUSSION

LOE

- Schema Updates
- Data Migration
- ETL Updates
- Testing

Scenario #1

- Add Customer Type
- Oops, and Sub Type

Scenario #2

- Customer has 3 Addresses

Type II?
SUMMARY OF DATA VAULT MODELING

- Allows for but does not require schema re-engineering
- Automatically supports data versioning for Auditing and Historical analysis
- More Joins so more objects in the schema but...  
  - Data Mart abstracts this from Users and Applications
  - Far greater flexibility (**Agility**!)
- More objects in schema = more objects in ETL  
  - This is a balance: less ETL maintenance of existing processes but more individual processes
- Testing required only on newly created objects


▶ http://www.geneseeacademy.org/

▶ https://www.youtube.com/watch?v=QbBmYMaQFec

MORE INFORMATION
Scot J Reagin
scot@SensibleDataIntegrations.com
303.378.4834

THANK YOU