Data Vault

DW Automation

Classification Matrix

Data Vault Automation Conference 2011

Hans Hultgren

gohansgo
Welcome

• Overview of Data Warehouse Automation
• Scope of the Classification Matrix
• Classification Criteria
• Automation Categories
• The Automation Matrix
• Applying the Matrix
• Operational Applications support business processes. Typically this implies the support and partial automation of components of a particular business process.

• In addition to software, business processes are also supported by methodologies, frameworks, specialized techniques, and also forms, templates and checklists.

• Together these form a pool of tools and techniques that support certain aspects of these business processes.
• With Data Warehousing and Business Intelligence another pool of tools and techniques exists to support particular aspects of these programs.
• In fact these tools and techniques are vast and varied – each addressing some combination of DWBI activities.
• To limit these tools to some degree, this presentation will focus mainly on Enterprise Data Warehousing and in particular those that utilize data vault modeling.
As mentioned, the focus is on Enterprise Data Warehousing:

* Integrated
* Subject Oriented
* Non-Volatile
* Auditable
* Time-Variant
* Adaptable

...and also utilizing data vault modeling:

* Atomic Level
* Business Key Based
* All Data
* Business Aligned
* Traceable
* Hub/Link/Sat

Plus all forms of automation tools and techniques:

* Software Tools
* Code Generators
* Common Models
* Methodologies
* Templates
* Documentation
* Frameworks
* Shells
* Metadata
Classification Criteria

• To begin working with the automation matrix we must consider – and understand – the various classification criteria.

• Effectively (in simple terms) this means that we look at different ways of thinking about these tools and techniques.

• As you will find, when a certain classification criteria is presented, and you begin to think about that criteria in context, the meaning becomes clear.
Classification Criteria

• For example, consider the following classification criteria:
  - Templates for ETL
  - Support for Data Modeling
  - Generation of Mappings
  - Automation of Testing

• For each one, consider them in the context of some of the tools and techniques presented earlier today.

  This process of contemplating criteria in the context of particular tools and techniques is the purpose of the automation matrix
Classification Criteria

• Notice in the prior examples there are two parts to each classification criteria...
  
  Templates for ETL
  Support for Data Modeling
  Generation of Mappings
  Automation of Testing

• The **LEFT** side items are tool or technique **Features**

• **The RIGHT side items are DWBI Functions**
Combinations of these Classification Criteria help us to form sets of **Automation Categories**

While there are some obvious ones to consider (ETL code generators, DWBI program methodologies, Model and Integration Templates, etc.) we are also able to assemble a custom set of criteria for our own automation category.
The Automation Matrix

- Header Section
- Main Matrix
- Profile

<table>
<thead>
<tr>
<th>FEATURES</th>
<th>Manage</th>
<th>Support</th>
<th>Structure</th>
<th>Organize</th>
<th>Automate</th>
<th>Generate</th>
<th>Template</th>
<th>Patterns</th>
<th>Document</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scoping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Requirements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visualization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information Modeling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Modeling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creating Databases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic Alignment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mapping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transform Rules/Logic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profiling, Data Quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Build ETL/ELT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Documentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deals with</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary Value / Uses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWBI Program Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Approach</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

www.GeneseeAcademy.com
The Automation Matrix

• Header Section

Data Vault Data Warehouse Automation Matrix

<table>
<thead>
<tr>
<th>Name:</th>
<th>Date:</th>
</tr>
</thead>
</table>

- Tool/Application
- Methodology
- Framework

• Capture name, note and date

• Categorize based on
  - Tool / Application - Software tool, application, template, shell, etc.
  - Methodology - PM, program, management, governance, etc.
  - Framework - Overall comprehensive end-to-end components
The Automation Matrix

- **Main Matrix**

<table>
<thead>
<tr>
<th>FEATURES</th>
<th>Manage</th>
<th>Support</th>
<th>Structure</th>
<th>Organize</th>
<th>Automate</th>
<th>Generate</th>
<th>Templates</th>
<th>Patterns</th>
<th>Document</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scoping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Requirements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visualization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information Modeling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Modeling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creating Databases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic Alignment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mapping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transform Rules/Logic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profiling, Data Quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Build ETL/ELT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Documentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Sets of Features and Functions – the classification criteria**
<table>
<thead>
<tr>
<th>FEATURES</th>
<th>Manage</th>
<th>Support</th>
<th>Structure</th>
<th>Organize</th>
<th>Automate</th>
<th>Generate</th>
<th>Templates</th>
<th>Patterns</th>
<th>Document</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scoping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Requirements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visualization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information Modeling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Modeling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creating Databases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic Alignment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mapping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transform Rules/Logic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profiling, Data Quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Build ETL/ELT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Documentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Features

- Manage  Assists in the management of this function
- Support  Directly supports the function itself
- Structure  Provides structure and structural components
- Organize  Helps to organize the function
- Automate  Automates components of the function
- Generate  Actual generation of artifacts related to the function
- Templates  Templates to provide consistency & to expedite
- Patterns  Design, architectural, and software patterns
- Document  Creates or provides documentation related to function
- Test  Helps with testing related to this function
Functions

- Scoping
- Requirements
- Analysis
- Design
- Visualization
- Information Modeling
- Data Modeling
- Creating Databases
- Semantic Alignment

- Mapping
- Integration
- Transform Rules/Logic
- Profiling, Data Quality
- Build ETL/ELT
- Testing
- Metadata
- Documentation
The Automation Matrix

- **Profile**

<table>
<thead>
<tr>
<th>Deals with</th>
<th>Data &amp; Data Models</th>
<th>Meaning &amp; Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Value / Uses</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Process Improvement</td>
<td>Design Automation</td>
</tr>
<tr>
<td></td>
<td>Code Generation</td>
<td>Documentation</td>
</tr>
<tr>
<td></td>
<td>Prototyping</td>
<td>DW Agility</td>
</tr>
<tr>
<td>DWBI Program Type</td>
<td>EDW - Large</td>
<td>EDW - SMB</td>
</tr>
<tr>
<td></td>
<td>ODS</td>
<td>Data Marts</td>
</tr>
<tr>
<td></td>
<td>Analytical Applications</td>
<td>Data Integration</td>
</tr>
<tr>
<td>Overall Approach</td>
<td>Data Driven</td>
<td>Source - Centric</td>
</tr>
<tr>
<td></td>
<td>Model Driven</td>
<td>Requirements - Centric</td>
</tr>
<tr>
<td></td>
<td>Semantics Driven</td>
<td>Enterprise - Centric</td>
</tr>
</tbody>
</table>

- Considers the scope of what the tools and techniques support, the primary value proposition and uses, what type of DWBI program is supported, and the overall approach for the DWBI program.
Questions?

Hans@GeneseeAcademy.com

gohansgo

HansHultgren.WordPress.com

HansHultgren

DataVaultAcademy
Architectural Layers
### Data Vault Data Warehouse Automation Matrix

**Name:** RapidACE  
**Date:** Sep 22, 2011  
**Note:** This analysis for version 1.0

**FUNCTIONS**
- Scoping
- Requirements
- Analysis
- Design
- Visualization
- Information Modeling
- Data Modeling
- Creating Databases
- Semantic Alignment
- Mapping
- Integration
- Transform Rules/Logic
- Profiling, Data Quality
- Build ETL/ELT
- Testing
- Metadata
- Documentation

**FEATURES**
- Manage
- Support
- Structure
- Organize
- Automate
- Generate
- Templates
- Patterns
- Document
- Test

**Deals with**
- Data & Data Models
- Meaning & Semantics
- Process Improvement
- Design Automation
- Documentation
- DW Agility
- Prototyping
- EDW - Large
- EDW - SMB
- ODS
- Data Marts
- Data Integration
- Analytical Applications
- Overall Approach
- Data Driven
- Model Driven
- Semantics Driven
- Source - Centric
- Requirements - Centric
- Enterprise - Centric

---

**Name:** Wherescape RED  
**Date:** Sep 22, 2011  
**Note:** This analysis is for RED, not for 3D.

**FUNCTIONS**
- Scoping
- Requirements
- Analysis
- Design
- Visualization
- Information Modeling
- Data Modeling
- Creating Databases
- Semantic Alignment
- Mapping
- Integration
- Transform Rules/Logic
- Profiling, Data Quality
- Build ETL/ELT
- Testing
- Metadata
- Documentation

**FEATURES**
- Manage
- Support
- Structure
- Organize
- Automate
- Generate
- Templates
- Patterns
- Document
- Test

**Deals with**
- Data & Data Models
- Meaning & Semantics
- Process Improvement
- Design Automation
- Documentation
- DW Agility
- Prototyping
- EDW - Large
- EDW - SMB
- ODS
- Data Marts
- Data Integration
- Analytical Applications
- Overall Approach
- Data Driven
- Model Driven
- Semantics Driven
- Source - Centric
- Requirements - Centric
- Enterprise - Centric
## Data Vault Data Warehouse Automation Matrix

### Features

<table>
<thead>
<tr>
<th>Name</th>
<th>Date</th>
<th>Tool/Application</th>
<th>Methodology</th>
<th>Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quiro</td>
<td>Sep 22, 2011</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Functions

- Scoping
- Requirements
- Analysis
- Design
- Visualization
- Information Modeling
- Data Modeling
- Creating Databases
- Semantic Alignment
- Mapping
- Integration
- Transform Rules/Logic
- Profiling, Data Quality
- Build ETL/ELT
- Testing
- Metadata
- Documentation

### Deals with

- Data & Data Models
- Meaning & Semantics
- Design Automation
- Documentation
- DW Agility
- EDW - Large
- ODS
- EDW - SMB
- Data Marts
- Data Integration
- Analytical Applications
- Source- Centric
- Requirements-Centric
- Model Driven
- Semantics Driven
- Enterprise - Centric

### Primary Value / Uses

- Process Improvement
- Code Generation
- Prototyping

### DWBI Program Type

- EDW - Large
- ODS
- Analytical Applications

### Overall Approach

- Data Driven
- Model Driven
- Semantics Driven