Agile Data Warehousing

Chris Galfi
Project Manager

Brian Zachow
Data Architect

COUNTRY Financial
IT Projects are too slow
IT Projects cost too much money
I never get what I expected
There must be a better way!
Review of Waterfall

- Analysis
- Design
- Build
- Test
Agile brings a new mindset

- Quick iterations
- Remove waste
- Frequent feedback to deliver the right product
- Deliver most important features first
- Scrum empowers the team
Agile (Scrum)

Sprint 1
- Plan
- Build
- Test
- Review

Sprint 2
- Plan
- Build
- Test
- Review

Sprint 3
- Plan
- Build
- Test
- Review

Deploy
- Increment

Product Backlog

[Diagram showing the Agile (Scrum) process with sprints and associated activities]
The Scrum Team

- Stakeholders
- Business Owner
- Scrum Master
- Product Owner
- SME
Task Board
Why Agile Data Warehousing

• Information consumption is constantly changing
• Business needs time to learn how the warehouse can help them
• Allows us to deliver value early!
# Pipeline Construction

<table>
<thead>
<tr>
<th>User Story</th>
<th>Sprint 1</th>
<th>Sprint 2</th>
<th>Sprint 3</th>
<th>Sprint 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Analysis</td>
<td>Code</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>Analysis</td>
<td>Code</td>
<td>Test</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td>Analysis</td>
<td>Code</td>
</tr>
</tbody>
</table>

Hughes, Ralph. *Agile Data Warehousing Project Management*. Elsevier Inc. 2013
As a Marketing Analyst I would like to understand what products customers are buying by their geographic region, so I can create better marketing campaigns.

The warehouse needs customer address data from the CRM system to further complete the marketing story to understand what products customers are buying by geographic region.

The warehouse needs product data from the point of sale system to further complete the marketing story to understand what products customers are buying by geographic region.
Agile Data Modeling
Traditional Data Modeling Option 1

• Build a conceptual data model
• A logical data model is created
• A physical data model is created
• Development begins!
Traditional Data Modeling
Option 2

- A conceptual/logical & physical data model is purchased
- A training event occurs to learn the data model, mapping process and customization process
- Analysis begins by mapping the business model to the purchased data model
- Development begins! (sometimes by the Company that sold you the model!)
Data Modeling for Agile

- Flexible/Adaptable
- Accept that the business model changes over time (and so should your model!)
- Have a Strategy because “You do not know what you do not know”
Modeling Goals

- Think Things, Details & Relationships
- Isolate Foreign Keys (they just cause trouble)
- Be prepared to add information about an entity or create a new relationship (including a new source of that information)
- Minimize impact to the ETL process
- Known as Data Vault or HyperNormalization
Big Data!!!!

• 3 V’s? (Volume, Velocity & Variety)

• Unstructured Data

• Unknown Benefits/Opportunities

• Our Business has a desire to have access to the information faster
Big Data Dilemma

• It is costly to store ALL of the data in a traditional database

• The data must be structured to load into the database

• ETL (Extraction, Transformation & Load) must be built to move the data

• The Business may never use it
Fishing at the Lake of Big Data

What if?
How Unstructured Data Fits in Warehousing

- Feeds the warehouse team business requirements and those requirements can evolve
- Allows business to learn about data and prototype quickly
- Pooling data allows us to always go back if the business needs more
Traditional Architecture
Big Data EDW Architecture
Success to date

- Deliver five major entities in three months with historical loads and daily updates
- Development team is self-sufficient in organizing their own work
- Success is being recognized
Questions?