AGILE DATA WAREHOUSING / STATE OF THE ART FOR 2016

Adaptive Data Modeling Techniques

MAXIMIZING THE WORK NOT DONE

Ralph Hughes, MA, PMP, CSM
ralph.hughes@ceregenics.com

Presenter’s Background

Ralph Hughes

- 30 years solutions architecture, ETL & BI development
- MA, PMP, CSM
- Author of three agile methods books
- Member of DW/BI advisory boards and best practices panel
- Frequent keynote speaker & instructor DWBI conferences

Website: www.Ceregenics.com
Email: ralph.hughes@ceregenics.com
www.linkedin.com/in/ralphhughesadw
Twitter: @ceregenics
Today's Topics

*Agile = quick & continuous delivery of value to the customer*

Agile EDW achieves this goal through:

- “Surface Solutions”
- End-User Hadoop
- Document data stores
- Hyper modeling
  - Hyper normalization
  - Hyper generalization
- Agile value cycle

Agile EDW Works Fabulously

- Major Healthcare Clinic (2014)
- Ceregenics joins an agile team that wasn’t getting traction during Iteration 7
- Best practices accelerates project by 3x to 10x, depending upon units of measure

First 6 iterations with a Scrum master only, no AEDW best practices
(Formerly) Outrageous Statements

The problem of enterprise data warehousing has been solved:

- If you’re not using 80/20 specifications, you’re taking 5x longer to get started
- If you’re coding by hand, you’re wasting 90% of your programming
- If you’re thinking only RDBMS, you’re building at least 3X more than you need
- Without automated testing, you’re missing over half of the defects

Presenters Must Remain Tool-Agnostic
Today’s Topics

*Agile = quick & continuous delivery of value to the customer*

Agile EDW achieves this goal through:

- “Surface Solutions”
- End-User Hadoop
- Document data stores
- Hyper modeling
  - Hyper normalization
  - Hyper generalization
- Agile value cycle

---

EDW is Like Building 5+ Applications at Once

*Mini BI universes / frameworks*

- Each architectural layer has different purpose and constraints
- Why approach them all with the same techniques and tools?
- Provisional value available long before application layer, so why wait?
Surface Solutions Possible Even with RDBMS

Today’s Topics

Agile = quick & continuous delivery of value to the customer

Agile EDW achieves this goal through:

- “Surface Solutions”
- End-User Hadoop
- Document data stores
- Hyper modeling
  - Hyper normalization
  - Hyper generalization
- Agile value cycle
Option 3: Agile Big Data

Schema-on-Read at Facebook

Evolving Surface Solutions Using Hadoop
Today’s Topics

Agile = quick & continuous delivery of value to the customer

Agile EDW achieves this goal through:

- “Surface Solutions”
- End-User Hadoop
- Document data stores
- Hyper modeling
  - Hyper normalization
  - Hyper generalization
- Agile value cycle

Option 4: Document Database

Increasingly More Big Data are in XML and Json Documents

Document data stores will let us explore and build quick solutions with very little programming
Powerful Search with Little Programming
Out-of-the-Box Delivery Tool – Requires only changing the query text and the indexing of the documents

Google-style query
Weighted elements

Registry of Patient Registries (Prototype)

Data Facets
- Age Bracket
  - < 10
  - 10-19
  - 20s
  - 30-45
  - 46-60
  - 60+
- Data Source
  - Anesthesia
  - Cardiology
  - Immunology
  - Pharmacy

Data Facets
- anesthetic implanted atrial edema OR autoimmune

Sort by: relevance

Atrial defibulator implant: 8/1/14; Medronic ADU-2193; copper leads; ICD-10-CM I50.9 diagnosis for reimbursement purposes; billable; date of discharge August 13, 2014; require the use of: anesthetics: inhaled agents nitrous oxide, Sevoflurane; transfusion-linked arterial edema; new-stroke ...

11/21/14 re-implantation service; ventral left atrium; presenting ICD-10-CM I50.21 non-reimbursed; discharged 12/1/2014; general anesthesia: Halothane and Methoxyflurane (inhalation), Diazepam (intravenous); significant-delay; autoimmune hepatitis: ...

Skills needed:
- Some HTML
- A little CSS
- Some XML or Json
- Some Xquery / XPath or Java script

Today’s Topics

Agile = quick & continuous delivery of value to the customer

Agile EDW achieves this goal through:
- “Surface Solutions”
- End-User Hadoop
- Document data stores
- Hyper modeling
  - Hyper normalization
  - Hyper generalization
- Agile value cycle

Ceregenics proprietary information
Step 1: Identify Business Keys

- **Customer**
  - Namebook ID

- **Order**
  - Order Nbr

- **CRM Agent**
  - Agent ID

- **Manufacturer**
  - Mfr Name

- **Package**
  - Package Code

- **Installer**
  - Installer Name

- **BK_Dealership**
  - Installer Name

- **BK_Manufacturer**
  - Manufacturer Name

- **BK_Package**
  - Package ID

---

Step 2: Create M-M Links

- **BK_CRM Agent**
  - CRM Agent ID

- **BK_Customer**
  - Namebook ID

- **BK_Installer**
  - Installer Name

- **BK_Dealership**
  - Installer Name

- **BK_Manufacturer**
  - Manufacturer Name

- **BK_Package**
  - Package ID

---

Ceregenics proprietary information
Step 3: Add Attributes

HNF Makes Re-Usable ETL Straightforward
Parameter-Driven ETL

- **Load_BK** (target, source, natural key column list)
- **Load_Link** (target, source 1, src 1 natural key cols, source 2, src 2 natural key cols)
- **Load_Attribs** (target, source, exclude column list)

"Cookie-cutter ETL"

---

**Today’s Topics**

*Agile = quick & continuous delivery of value to the customer*

Agile EDW achieves this goal through:

- “Surface Solutions”
- End-User Hadoop
- Document data stores
- Hyper modeling
  - Hyper normalization
  - Hyper generalization
- Agile value cycle
From HNF to HGF

0 of 5

Convert to Metadata to Distinguish Instances

1 of 5
“Fold” the Model to Eliminate Separate Tables

Combine Tables with Equivalent Function

The Model

The Data
Allow Re-Classifications of Instances

The Model

The Data

Completely Temporal Data Warehouse

Ceregenics proprietary information
Model Objects Map to Meta Data Entities

1. "Product Type" class exists
2. "Product" class exists
3. Product rolls up to Product Category

Computer-Assisted ETL Programming

- Parts Data Mgt
  - Extract 150213
  - Product Nbr
  - Product Name
  - Packaging Code
  - UPC
  - Product Type Nbr
  - Prod Type Name
  - Extraction Date

- Add/Mod Instance
  - PRODUCT
    - Product Nbr
    - Product Name
    - Packaging Code
    - UPC
    - PRODUCT TYPE
    - Trans Date
Automation Surrounds Us

- Computers build our goods...
- ...and our planes...
- ...will soon drive our cars.

Why are we still building data warehouses by hand?

Henry Ford Considers a Tesla

Where's the carburetor?
... the transmission?
...the radiator?
I can see all kinds of problems with this car. What a hoax!
Why are we still building data warehouses by hand?

- Don’t understand what’s going on under the hood
- Don’t want to give up data modeling
- Don’t want to be the first one to do it
- Too much invested in 1990s technology
- Don’t want to eliminate people’s jobs
- Don’t want to eliminate my department
- Staff can’t handle learning another tool

Tools for the Business Opportunity Cycle

1. Grab (More) Data
2. Google-Style Search
3. Raw Data Vault
4. Add Integration Cross Lists
5. Promote & Support Usage
6. Publish Certified Solutions
Automated Monitoring for Faster Requirements

1. Data Virtualization Server
2. Adaptive Master Data
3. Enterprise BI Package
4. Document Data Package

Tools for the Business Opportunity Cycle

1. Hadoop
2. Document Data Package
3. Data Warehouse Generator
4. Citizen Data Scientist Tool
5. XQuery Configuration
6. Data Warehouse Generator
Find Your Voice & Help Others to Find Theirs
– Stephen Covey’s “Eighth Habit”

- Call for contributors
- Write a chapter, sidebar, or a section
- Focus: theory & case histories that blend
  - Disciplined agile & EDW methods
  - Hadoop, M/R, Spark
  - Textual and triple data stores
  - Empowering citizen data scientist

Long Design will Still Delay Value

Traditional Methods

| Project Definition | Database Design | ETL Coding | Release Review |

Agile Approach + Productivity Tools

- “Surface Solutions”
- End-User Hadoop
- Document data stores
- Hyper normalization
- Hyper generalization
- Agile value cycle
Hyper normalization: //www.youtube.com/watch?v=3QOSOeN8vcY
Hyper generalization: //www.youtube.com/watch?v=aNtUoVkeq_Q