What is a Data Warehouse?

“A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data used in an organization”
Characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject-oriented</td>
<td>Data available on an entire subject area</td>
<td>Visits, Orders, Respiratory, etc.</td>
</tr>
<tr>
<td>Integrated</td>
<td>Data from multiple sources integrated in a single place</td>
<td>Orders data integrated across Cerner, Epic</td>
</tr>
<tr>
<td>Time-variant</td>
<td>Data from a specific time-period</td>
<td>Nursing documentation data in the only available from 2007 onwards</td>
</tr>
<tr>
<td>Non-volatile</td>
<td>New data constantly added, but old data is not deleted</td>
<td>Financial records from two retired systems (Allegra &amp; IDX) and current (Epic)</td>
</tr>
</tbody>
</table>

Enterprise Data Warehouse

An Enterprise Data Warehouse (EDW) contains all the relevant current and historical data for the entire organization.

Data-marts

A data mart is a logical subunit of a data warehouse which contains data related to a single subject area (e.g. Orders), department/service line (e.g. Cardiovascular), business unit, etc.

High Level Overview

Brief History of Systems

Transaction Systems
- 1993-1999 OAIS (inpatient)
- 1995-1999 Epic Care Ambulatory
- 1995-2000 Allegra
- 1995-2010 IDX
- 1999-2003 e-Chart (inpatient)
- 2003-2014 Cerner (inpatient)
- 2007 electronic nursing documentation
- 2009 computerized provider order entry
- 2010-date Epic for Business
- 2014-date Epic OneChart Inpatient

Analytic Systems
- 1990 HL7 integration engine
- 1993 Distinct data store populated using HL7
- 1995 Cognos business intelligence (BI) suite
- 2000 integrated financial data (batch & HL7)
- 2002 Subject-oriented data-marts
- 2005 Corporate Radar web-based tools for reporting and dashboards
- 2011 SAP Business Objects Enterprise Business Intelligence

How is an EDW different?

<table>
<thead>
<tr>
<th>Transactional Database (EMR)</th>
<th>Analytical Database (EDW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimized for simple, single transaction queries.</td>
<td>Optimized for complex queries on patient populations.</td>
</tr>
<tr>
<td>Data from a single system.</td>
<td>Data from multiple source systems.</td>
</tr>
<tr>
<td>Highly normalized data structures.</td>
<td>De-normalized or dimensional data structures.</td>
</tr>
<tr>
<td>Raw data on transactions.</td>
<td>Details + aggregate data.</td>
</tr>
<tr>
<td>Used by clinical, financial and other application end-users.</td>
<td>Used by analytical users, researchers, etc.</td>
</tr>
</tbody>
</table>
Types of Schemas

Data Warehousing Approaches

<table>
<thead>
<tr>
<th>Inmon</th>
<th>Kimball</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top-down: first common relational model, then data-marts, then integrate</td>
<td>Bottom-up: first several data-marts, then integrate</td>
</tr>
<tr>
<td>Relational model – 3rd Normal Form (3NF)</td>
<td>Dimensional model – Star schema and Snowflake</td>
</tr>
<tr>
<td>Time to fully build an EDW and then develop data-marts</td>
<td>Faster development cycle – data-marts built first</td>
</tr>
<tr>
<td>Higher consistency across the EDW</td>
<td>Potential for silos in separate data-marts</td>
</tr>
<tr>
<td>Most granular data for reporting as the need arises.</td>
<td>May require additional development if a dimension was not planned initially</td>
</tr>
</tbody>
</table>

The Data Integration Challenge (1)

- Malus domestica
- Fruit
- Sweet
- Things you eat
- Made in California
- On sale at Smiths (expires 09/13/2014)

The Data Integration Challenge (2)

- Three common attributes of these objects
- Three textural descriptions
- Three numeric properties
- Three uses of each object
- Three date/time properties

The Data Integration Challenge (3)

- Malus domestica
- Fruit
- Sweet
- Things you eat
- Made in California
- On sale at Smiths (expires 09/13/2014)
- Citrus sinensis
- Fruit
- Citrus fruit
- OJ (no, not Simpson, sweet)
- Tangy
- Things you eat
- Made in Florida
- Volkswagen Beetle
- Year
- Trim
- License plate
- VIN #?
- Things you drive
- Made in Germany

Data Integration Considerations (1)

- Like Vs. Unlike attributes
- Old Vs. New sources/attributes
- Necessary Vs. Unnecessary attributes
- Feasible Vs. Unfeasible approaches to integration
Data Integration Considerations (2)

• What is the intended use?
• How soon does it need to be deployed?
• How will the users access the data?
• How reusable are these data for users across the organization?

Example of a Snowflake Model

EDW & UUHC Infrastructure

• Tier-1 system
• Multiple copies – redundancy
  – Downtown Data Center (primary)
  – Hot Site Data Center (failover)
  – Disaster Recovery Site Data Center (scheduled)
• Multiple nodes – Oracle RAC
  – 2 Rack-mount Servers (768 GB RAM, 16 proc. EA)
  – 4 Blade Servers (256 GB RAM, 12 proc. EA)
What is Metadata?

“Metadata is all the information in the data warehouse environment that is not the actual data itself.”

Metadata Contributors

- EMR Developers
- Workflow Engineers
- EMR Implementers
- Clinical Users
- EDW Architects
- Analysts

30

Metadata Example (1)
Defining Data Quality

• “Fitness for use”
  – Operations
  – Decision-making
  – Planning, etc.

Assessing Quality

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness</td>
<td>Does the data provide a complete picture of events in a given domain?</td>
<td>Visits data-mart - complete picture of all patient/visit-level events.</td>
</tr>
<tr>
<td>Accuracy</td>
<td>How well do the data reflect entities and relationships in the real world?</td>
<td>UPDRS process in the EMR - modeled in the Orders data-mart in terms of medication and lab orders, other orders, lab results, medication administrations, etc.</td>
</tr>
<tr>
<td>Timeliness</td>
<td>Are the data available at the time needed, without delay?</td>
<td>HL7 feed instead of nightly extract for medication orders</td>
</tr>
<tr>
<td>Syntax correctness</td>
<td>Are the data captured and stored in a well-conceived data model?</td>
<td>Determining the most suitable data model (See 3NF, dimensional, etc.)</td>
</tr>
<tr>
<td>Semantic correctness</td>
<td>Are the data encoded using proper healthcare terms?</td>
<td>Determining if NDCs were used in documenting medication dispense, administration and billing</td>
</tr>
</tbody>
</table>

Monitoring Completeness (1)

Monitoring Completeness (2)
Novel Applications & Integration

Cerner – Allegra & Cerner – Epic

Methods

- Enterprise Business Intelligence – Business Objects
- Warthog Targeted Chart Review Tool
- Direct access to the database – SQL Navigator, Toad, ODBC, etc.
- Legacy tools – Corporate Radar
- Other custom tools and applications

The “Clinical Universe”
Drag Objects to Include

Add Filter Objects

Example Report

A Report With Drill-down

Drill down to individual study

Warthog: Initial Cohort