Methodological Standpoint: Inmon Vs Kimball

Aravind Kumar Balasubramaniam

February 19, 2015
Abstract

This article defines Data warehousing and its basic concepts and describes the methodological standpoint between two influential data warehousing experts Bill Inmon and Ralph Kimball by providing the identical attributes, contradictions, influential factors favoring Inmon and Kimball approach with a couple of real-time executed projects following some of the guidelines to determine the best approach based on the referenced papers.
Introduction

Data Warehouse: An Information Technology system used for reporting and data analysis which has centralized repository having the data integrated from one or more related or unrelated sources. They store current and historical data which is effectively used by business users or senior management for creating trending & other reports.

**Inmon’s Definition:** “A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management’s decision making process.” (Inmon 2005)

**Kimball’s Definition:** “A data warehouse is a copy of transaction data specifically structured for query and analysis.” (Kimball & Ross 2002)

Inmon’s Data warehouse: A centralized repository of an enterprise spanning across all lines of business and subject areas containing integrated data from disparate sources.

Kimball’s Data mart: A specific subject oriented or pertaining to individual business line that are built using dimensional modeling. (Abramson 2004)

Methodology Comparison

An organization should choose data warehouse design and its maintenance tools which are commercially available. All the market available tools may not be compatible with each other and may not suit for all development methodologies. But in a holistic view they are based out of only two models created by either Inmon or Kimball.
Inmon’s Methodology:

Bill Inmon defines data warehouse as a centralized repository and advocates a top-down development approach that adapts relational database tools (Extract Transform Load - ETL) to the development needs of an enterprise wide data warehouse which extracts data from various sources and stores into 'atomic' data at the lowest level of detail passing through staging or landing area. Data marts are created only after the complete data warehouse has been created. So data warehouse is at the center of the Corporate Information Factory (CIF), which provides a logical frame work for delivering business intelligence. (Breslin 2004)

![Figure 1 The Inmon Warehouse model (Abramson 2004)](image)

Inmon divides the overall database environment of the organization into four levels:

1. Operational
2. Atomic data warehouse
3. Departmental
4. Individual
The last three levels comprise the data warehouse. The first level contains data from legacy and other transaction processing systems. This level supports the day-to-day operation of the organization; in other words, the first level supports all transaction processing. From the operational systems, data is extensively manipulated and then moved to the atomic data warehouse. (Inmon 2005)

Let’s see an example from Inmon illustrating the difference between Operational data and data stored in atomic data warehouse

**Example 1**

**Entity = Customer**  
Attribute = Customers credit rating

**Operational system database:**  
Attributes (loan balances, outstanding balance, address, etc.) in a SINGLE record.

**Atomic data warehouse:**  
Credit history by year with one record/year.

**Department level: (Dropping irrelevant information)**  
For (e.g.)  
Credit department: Customer address (With a "flag" to indicate change of address)  
Marketing department: All customer-identifying data except zip code.

(Inmon 2005)

The above example illustrates each department might hold data summarized based on its needs. But Inmon’s architecture makes certain that all data is consistent as all department data comes from atomic data warehouse.

**Kimball’s Methodology:**

Ralph Kimball, suggests a bottom-up approach that uses dimensional modeling. Rather than building a single enterprise wide database, data marts are created first (Data extracted from various sources then transformed and loaded into dimensions and facts passing through
staging or landing area). These provide a thin view into the organizational data, and as and when required they can be combined into larger data warehouse. Enterprise wide cohesion is accomplished by using another Kimball innovation, a data bus standard. (Breslin 2004)

![Diagram of the Kimball's Data Lifecycle](image)

**Figure 2 The Kimball's Data Lifecycle** (Abramson 2004)

Dimensional Modeling: It begins with tables rather than entity-attribute data models such as ERDs. These tables are dimensions (lookup tables) and fact tables. Fact table contains the measures of interest. For e.g. sales amount. This measure is stored in fact table with appropriate granularity. While dimension tables contains attributes of the measures in the fact tables. For e.g. product, store, time dimensions stores the product, store, time (month) attributes which enables to aggregate the measures in fact with multi-dimensional view.

For e.g. total amount of sales over a particular period in a store for a product.

Four-step dimensional design process:

1. Select the business process.
2. Declare the grain.
3. Choose the dimensions.
4. Identify the facts. (Kimball & Ross 2002)

Discussion

Identical attributes between Inmon and Kimball:

a. Usage of time-stamped data.
   i. Kimball calls the time attribute the “date dimension;” (As per Kimball date key in a date dimension is an artificial key that defines a conformed dimension.)
   ii. Inmon calls it the “time element.” (As per Inmon date attributes either in different normalized tables or simply calculated at the time of user query.)

b. Extract transform and load process.
   i. Data is extracted from operational databases, transformed to meet the data warehousing standards and loaded into centralized Data warehouse (Inmon’s) or Data mart (Kimball’s)

c. Data attributes and query results. (Breslin 2004; Rifaie et al. 2008)

Contradictions between Inmon Vs Kimball:

<table>
<thead>
<tr>
<th>Description</th>
<th>Inmon</th>
<th>Kimball</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approach</td>
<td>Top-down</td>
<td>Bottom-up</td>
</tr>
<tr>
<td>Data alignment</td>
<td>Subject or data driven</td>
<td>Process oriented</td>
</tr>
<tr>
<td>Complexity</td>
<td>Very complex</td>
<td>Fairly simple</td>
</tr>
<tr>
<td>Structural design</td>
<td>Enterprise wide (atomic) data warehouse feeds departmental database</td>
<td>Data marts model an individual business process and enterprise consistency achieved through data bus and confirmed dimensions.</td>
</tr>
</tbody>
</table>
Redundancy  | Redundancy is regulated to great extent, data is mostly in 3NF | Data is often redundant
---|---|---
Flexibility  | This approach is comparatively rigid | This approach is flexible design and build
Primary audience  | Delivers a technical solution to IT Professionals | Delivers a solution for Business or End users to directly query data in reasonable response time

*Table 1: Contradictions between Inmon Vs Kimball* (Breslin 2004)(Senapati & Anil Kumar 2010)

**Influential Factors:**

A data warehouse is designed based on the objective, nature of business, time span and affordable budget of an enterprise.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Inmon</th>
<th>Kimball</th>
</tr>
</thead>
<tbody>
<tr>
<td>System design &amp; build</td>
<td>Time consuming</td>
<td>Comparatively less time</td>
</tr>
<tr>
<td>Cost to business</td>
<td>Initially high cost and subsequent projects and maintenance costs are reduced</td>
<td>Initial cost is low but subsequent phase and maintenance will cost the same</td>
</tr>
<tr>
<td>Skill Requirement</td>
<td>Specialized team with high technical skills</td>
<td>Generalized team</td>
</tr>
<tr>
<td>Data integration</td>
<td>Enterprise wide</td>
<td>Individual business wide</td>
</tr>
<tr>
<td>Persistency of data</td>
<td>High rate of change in source system</td>
<td>Relatively stable source system</td>
</tr>
<tr>
<td>Change in design</td>
<td>They do not change the design with every change in business instead changes are accommodated within the existing model.</td>
<td>Local optimization is frequent and focus is on quick win.</td>
</tr>
</tbody>
</table>

*Table 2: Influential factors favoring Inmon or Kimball* (Rifaie et al. 2008; George 2012)
Paradigm over few sectors:

1. Marketing domain: This is a specialized domain looking precisely for a particular area of information for its business. Hence, Kimball’s data mart approach is suitable.

   **A real time e.g. on GTE Airfone, Inc.**

   "In the past, we implemented new marketing promotions that weren't very well targeted, and [so] we didn't get much pay back," - Shekar Vengarai, Senior advisory system engineer at GTE Airfone.

   "The air traveler requires specialized marketing programs, and to achieve success, we need access to accurate, timely data" (Weston 1998)

   GTE decided to go with data mart design as it precisely needed to process its customers credit card information extracted from vendors and cross reference with their profile information such as average income, age levels for its marketing promotions which is not limited to process the dirty data to monitor the quality of service by collecting the phone calls that didn't go through properly.

   For e.g. A particular plane has a high number of incomplete calls, which could signal a problem in system.

   (Weston 1998)

2. Insurance domain: It is essential to acquire a holistic picture with respect to individual clients, groups, history of claims, mortality rate, agents etc. All the subjects are inter-related and hence suited for Inmon's data warehouse approach.

   **A real time e.g. on Samsung Life Insurance**

   Samsung had been initially using 10 non-compatible data storage systems. With the demand for integrated information which would allow:
   - Field-workers to perform their own analysis resulting in competence increase.
   - Common dashboard to access highly compatible data.
   - Reducing the time span of report generation.

   (Rostek 2009)
Critical Evaluation

Some of the essential guide lines to choose the right methodology are based on the various factors like business objective - short term and long term, technical expertise, cost to operate, quantitative and qualitative analysis of data sources.

There is a misnomer about bottom-up approach of Kimball methodology typically viewed as quick and dirty – focused on the needs of a single department rather than enterprise. But following the "enterprise bus matrix" and conformed dimension these data marts explicitly embrace an enterprise point of view.

The summarized, departmental data marts supported by a large, normalized data warehouse providing access to atomic data are avoided in Kimball methodology since centralized and normalized view of data warehouse is responsible for many of the failures in terms of business intelligence application as well as exposing snow flaked or normalized table directly to end users is strongly discouraged in order to reduce complexity. (Ross 2003; Becker 2007)
References


Citing Internet Sources

