

Practical uses for a Data Warehouse: Part Two

Marketing

Marketing is like the weather, everybody talks about it but nobody does much about it. While this is not exactly true, certainly more can be done for marketing.

One of the greatest challenges for a marketing program is accessing existing customer's data from across all the various systems within a company. Many of these systems were developed independently years ago and they are not integrated today. These host systems are termed 'silos of information'. For example, it is not uncommon for a single individual's data to be found across many separate systems represented quite differently. *Best's Law states that if data is in two places, it will be inconsistent.* As a result, marketing may create and maintain yet another customer database or worse yet buy data from the secondary market. Think of that, buying data about your existing customers because you can't get it right! This isn't very cost effective or efficient.

A second challenge is the expense in time and resources involved with integrating all the disparate silos of information. There are three ways of accomplishing data integration using data warehouse (DW) concepts and methods. They are:

1. Top Down.

- Gain approval for a multi-year project extending across all aspects of the business
- Start with an enterprise data model by identifying all subject areas (For example, line of business such as clothing/hardware/appliances, or organizationally such as manufacturing, Human Resources, finance, or accounting, etc.)
- Identify all the entities (products, customers, locations, etc) and their respective attributes
- Identify the relationships between the entities
- Create universally agreed upon definitions for the entities, attributes, and relationships
- Create a data dictionary with everything about everything
- Make the data model available to all parts of the business as THE standard and require all new development to use the newly created entities, attributes, relationships, and definitions.

2. Bottom Up.

- Do everything in Top Down, but focus narrowly on a single subject area
- Tie it all together later when you have the budget (which is never)

3. Hybrid Development

- Do everything in Bottom Up, but instead of delaying integration include it as a requirement from the very start
- Start the enterprise data model with the first subject area defining and making available to subsequent projects everything they have created and learned.

Practical uses for a Data Warehouse: Part Two Marketing

- Create a line item on the budget for integration, maintenance, and coordination of all models into a common enterprise data model, data dictionary, data quality initiative, system architecture, and Zackman¹ diagram.

The careful reader will immediately identify the hot spots in attempting a data integration project. They are resources, time, and commitment. Sorry, there is no magic here. No one can make the pain go away with a new acronym or buzzword. Even the most accomplished project manager with all the certifications in the world can not put five pounds of peanuts into a three-pound bag. Lots of peanuts will spill over and be wasted. Project management is a discipline of the possible, not the improbable.

Why would anyone attempt an integration project? The answer lies in this simple question. Do you want to be another **Wal-Mart** or another **Kmart**?

Now the good news, an integration project using DW concepts and methods is affordable, flexible, and grows with you. You and your staff become the gurus in your own business. While outside experts may still be occasionally required, you can manage them as any other project resource. Analogy may be helpful here. Think of a working DW as your public library. Many library customers gain access to the books, magazines, and other periodicals using the Dewey Decimal System. While Mr. Dewey is now long dead, his efforts continue to be used by librarians and library customers around the world. Your DW project can follow this model too.

A couple of additional points are very important here. First, data warehouses/marts are architecture and not product or the 'technology du jour'. To the point, successful DW projects have been completed using Oracle, DB2, Teradata, Sybase, Informix, Cache, Unix, Linux, Windows, MVS, VSAM, COBAL, Smalltalk, Visual Basic, C++, Java and a wide variety of software tools. You name it and you can find it in the DW environment. The particular vendor/technology/tool is not as important as the commitment of resources, time, process, and most of all a solid data model.

Second, an integration project is not a single, one-time project. Much ballyhoo has been made in the press about the failure rate of DW projects. The press has it all wrong. DW projects are ongoing efforts much like continuous process improvement, six sigma, and other long term initiatives. Essentially, it is the formalization of a process that will be with your forever. It is a change in the corporate culture. To the point, has anyone seriously suggested to you lately to abandon your quality initiatives?

So now that we have separated some of the chaff from the wheat, let us continue with how DW concepts and methods can be used for Marketing Projects.

For the purpose of this example, we will assume:

- Two host production systems, Host A and Host B. They will remain as data sources and they contain similar and overlapping customer information. (Redundant data)

Practical uses for a Data Warehouse: Part Two Marketing

- The marketing department needs data from both systems to be readily available and understandable. (Integration project)
- The agony and pain incurred by requesting another IT project will be short-lived compared to the benefits. (Appropriate project sponsor and well defined business requirements)

Review the tables below.

Name	Address	Phone	Purchases	Age
J. B. Smith	123 S. Main Street, Your Town, MI 49201	6167891234	10	35

Host A

Fname	Lname	Address	City	State	Zip	Marital Status	Gender
J	Smith	123 S. Main St.	Your Town	MI	49201-4371	G	7

Host B

Let us analyze the data available in the two tables above. As a human reader, you can gain a lot of information by looking at both tables at once. However, some mysteries remain. They are:

- Are 'J.B Smith' and 'J. Smith' the same person?
- What does Purchases = 10 represent?
- What does Marital Status = g represent?
- For heavens sake, how many genders can there be, Gender = 7?

First of all, if you are every going to build/maintain a durable and flexible enterprise data model to support data integration you must follow some simple axioms of sound data model design. That is only atomic data should be found in a cell. Atomic data is data that represents only one concept or thing. For example, Phone 6167891234 is NOT atomic. Review the table below for a durable and flexible design.

Contact Phone Area Code	Contact Phone Exchange	Contact Phone Number
616	789	1234

Atomic Data

Now let us solve all the mysteries we can.

- How can we be sure if 'J.B.Smith' and 'J. Smith' are the same person? Well, quite frankly we can't. This is an example of 'dirty data' and/or lack of vision by the host system designer. Typically, host systems were built with a very narrow scope. If the project was not going to use it, then it was not included. In a perfect world the host systems would be required to 'clean up' their data, but this seldom happens. A more realistic option would be to interact with the customer and request the clarification. This updated and validated data would add value to the marketing system.

Practical uses for a Data Warehouse: Part Two Marketing

In addition, if we do the best we can and create durable and flexible DWs, we can augment our efforts and ‘clean up’ addresses and phone numbers by using outside vendors. For example, by referencing the United States Postal Service (USPS) ‘change of address’ or Ameritech’s ‘working phone numbers’ databases, we can make huge dents into improving data quality. This is real added value.

- B. By referring to the Host A’s documentation and talking to the system owner, we discover that ‘10’ represents the number of times the customer ordered from our company. Asking the Subject Matter Experts (SME) is exactly the process by which the DW project identifies, defines, and implements data from the host systems. This is something that certainly can be done with your own staff after some training.
- C. These are system application codes. Someone in the distant past created a code and hopefully gave it a meaningful description. But beware, one of the biggest problems in older host systems is ‘code reuse’. In some organizations, it has become an art form. For example, a description changes over time and/or it gets re-used for an entirely different purpose based on the current business need (previous description a = applesauce, current description a = single). The absolute worse case scenario is to maintain the description outside the host system, possibly in paper notebooks.

Review the imaginary code table below for the current descriptions.

Code	Description
a	Single
b	Divorced
g	Married
1	Male
7	Female

Host Code Table

- D. Refer to C above.

By combining the data from Host systems A & B, we can now infer more about our mystery customer.

We now know our customer is a female, but we are not sure of her first or middle names yet. We suspect she is a frequent customer, but without a time frame we are not quite sure. We could compare her number to other customers in Host A to get a better idea. We certainly know she is married and she is 35 years old. Hmmm, based on the year 2003 this means she was born in 1968, which means she probably graduated from high school between 1984 and 1986. She might be a customer would like to know more about.

Now if this doesn’t boil your marketing blood, I know of little else that will. And, it is not rocket science.

Practical uses for a Data Warehouse: Part Two Marketing

Bibliography

1. Zackman, John. "A Framework for Information Architecture." IBM Systems Journal, Vol.26, No. 3. (IBM Publication G321-5298).