



OPENING THE DATA WAREHOUSE

Cliff Brandon
October 2009

OPENING THE DATA WAREHOUSE

By: Cliff Brandon 10/06/2009

TABLE OF CONTENTS

FORWARD.....	4
EXECUTIVE SUMMARY	4
INTRODUCTION	5
THE GENERAL VALUE PROPOSITION	5
MOVING FROM SERVICE to PRODUCT	5
A FEW NOTES... ..	7
AUDIENCE	7
OTHER DOCUMENTS.....	7
OPEN SOURCE and OPEN DATA WAREHOUSE.....	8
CURRENT STATE OF DW and 'OPEN' DW TECHNOLOGY.....	8
WHAT IS ODWM?	8
INDUSTRY 'SLICES'	9
WHY 'OPEN'?	9
BUILT TO FAIL.....	10
THE MARKET	11
DOES THE MIDMARKET NEED DATA WAREHOUSING?.....	11
DATA PORTABILITY	11
WHY WILL ODWM WORK?	12
WINNERS	12
Open Data Warehouse Developers	12
Owners	12
Developers	12
Tool Providers	13
Source Providers	13
Service Providers.....	13
LOOSERS.....	13
CONCLUSION	14

COPYRIGHTS and CAVEATS

Copyright © 2009 by Automated Systems Alliance, Inc. All rights reserved worldwide.

This document and its preceding drafts are intended for general distribution without charge. No portion of this work or its derivatives may be used in whole or in part in any other document or work product without the express written permission of the author.

No liability, patent or otherwise is assumed with respect to the information herein. Nor is any liability assumed for damages resulting from the use or misuse of the information contained herein.

Any opinions expressed herein are those of the author and are not meant to represent the position of any other company, organization or professional association.

This is your last chance. After this, there is no turning back. You take the blue pill - the story ends, you wake up in your bed and believe whatever you want to believe. You take the red pill - you stay in Wonderland and I show you how deep the rabbit-hole goes. Morpheus – The Matrix 1999

FORWARD

In many respects this paper is a summary of conversations I have had with colleagues since 2003 and is a time capsule from the bottom of a new S-curve. From here, it's hard to tell how steep the curve will be or how far it will go.

EXECUTIVE SUMMARY

Data warehouses, as deployed today, are services masquerading as custom products. Data warehouses are better deployed as 'open' products with attendant services and supporting commercial products. This basic idea can change the face of data warehousing and create frontier opportunities for a new set of entrepreneurs in this technology space.

Making a data warehouse an 'open' product reduces a consumer's initial construction cost and mitigates risk through rapid deployment. This approach allows data warehouse and other co-dependent technologies (Business Intelligence, Analytics, etc.) to be more attractive to a new and largely untouched midmarket (50M-2B).

This frontier opportunity has low barriers to entry which will make the proposition attractive to entrepreneurs interested in 'owning' an industry 'slice' (eg. health care - clinical records, small box retail – marketing, banking – loan evaluation, etc.). Commercial ecologies designed to support and enhance each of these industry slices offer additional new market opportunities to developers. Availability of talent that can take advantage of this perishable opportunity is high due to current market conditions and will be very attractive to current technology providers who understand the economic benefit of becoming technology owners. Said another way; this approach will create a lot of new financial winners.

There are technical, social, and economic reasons that will move data warehousing and, perhaps all data models toward 'open' solutions. This move toward Open solutions will create new product and service markets and will have the effect of making data warehouse and business intelligence technology available to a broader audience.

INTRODUCTION

In the last couple of years 'Open Source' products have emerged in the Data Warehouse/Business Intelligence space. There are now products representing every aspect of the stack (source, ETL, Data Bases, metadata architectures, Business Intelligence, Analytics, etc.) except one.

Until recently, no one has delivered a data warehouse architecture in the open source space (ASA delivered the Open project Control Data Warehouse in March of 2009). This paper explores the importance and consequences of delivering this last element to the 'Open' DW/BI stack and to the midmarket.

THE GENERAL VALUE PROPOSITION

Let's face it, free is a great value proposition. However, 'open' doesn't necessarily mean free. In this case what doesn't have a cost or fee associated with it is the data warehouse architecture. Providing an architecture reduces overall cost and risk by reducing design and implementation time. It does not eliminate all costs associated with standing-up a data warehouse.

Each Open source data warehouse instance has the potential to create a local ecology of products and services specific to the Open data warehouse theme or 'slice'. This ecology would include familiar tools for reporting and analytics as well as some new tools or package solutions (see section on winners) that did not previously exist.

To make these changes to the current custom data warehouse model requires a change in thinking about what data warehouses are; not in the technical aspects of data warehouses but in what they should contain and how they should be deployed. This lateral shift in thinking about data warehouses will be disruptive to the current industry and processes. This disruption will provide a number of new opportunities for new participants in the untapped middle market.

MOVING FROM SERVICE to PRODUCT

Current data warehouses are services not products. Many that have a data warehouse in place probably think they 'have' a product. If you look at the costs of development, implementation, and maintenance however, you will find that most of your costs are for labor. What's been purchased is a service not a product.

Service, costs, and belief in the custom nature of data warehouses are endlessly entwined. It's one of the primary reasons data warehouses have been so resistant to commoditization. If data warehouses are turned into actual products they become more easily commoditized. If they become 'Open' their value proposition changes dramatically.

If data warehouses begin to transition to Open products in the near future, they will make this transition in an environment where:

- 1) The underlying data warehouse concepts are very mature...to the point of being stagnant,
- 2) There is a vibrant market for data warehouses, business intelligence (BI), and analytics technology,
- 3) There are 'Open Source' product representatives for almost every aspect of the BI/Data warehouse stack except architectures (yet),
- 4) There are low barriers to entry for developing data warehouses and potential for economic gain for doing so,
- 5) There are resources (people) that, because of the economic downturn, can take advantage of and drive various aspects of Open data warehouse development,
- 6) Hardware is no longer a serious consideration for the vast number of data warehouses and,
- 7) Vendors that have been feeding at the top of the economic pyramid have begun looking down-market for new sales opportunities due to the current economic conditions.

It's the perfect storm. The opening of data warehousing, whatever it winds up looking like, represents a new opportunity for a mature concept. It probably won't be a new technological breakthrough. It's more likely going to be a lateral shift in thinking about the technology and how it can be applied.

Where does this go? What does it look like when we get there? The truth is...I don't know. The end result may not even be called data warehousing and this commoditization may not be restricted to data warehouse architectures.

The commoditization of the mature, entrenched, billion dollar data warehouse industry is going to upset a few people. Most change of any significance does. Industry entrenchment will ultimately have to contend with low barriers to entry, frontier economic benefit, and a very large, untapped midmarket.

A FEW NOTES...

This paper does not address the mechanics of building 'Open' type data warehouses. It focuses on the 'why' question. A separate paper addressing the construction of Open data warehouses was written to address the 'how' issues (see the 'Other Documents' section below).

An instance of an Open data warehouse has been constructed for Project Control (PCDW). It is hoped that this will not only provide a functional model but also provide a useful roadmap for other model developers (see <http://www.asaservices.biz>).

Finally, to eliminate confusion, an Open data warehouse should be thought of as a container for data; a data 'sandwich bag' if you will. Using an Open data warehouse does not mean that your data must also be available as an 'open' resource. Likewise, data is not provided with Open models. Open data warehouse providers supply the 'bag' you provide, maintain, consume and benefit from the sandwich.

AUDIENCE

This document is meant for a fairly broad audience. It is directed at those who want to understand and build on an emerging market or re-emerging market as the case might be. Some familiarity with the Business Intelligence/Data Warehouse market is assumed. Deep knowledge of the technology behind Business Intelligence/Data Warehousing is not required.

OTHER DOCUMENTS

There are several other related documents. All are available at www.asaservices.biz

The Open Data Warehouse Model (ODWM): Basic instructions for how to build an Open 'style' data warehouse,
Introduction to Project Data Warehouses: Paper describing data warehouses for Project Control,
The Project Control Data Warehouse (PCDW): An instance of the ODWM for Project Control, and other related information.

ABREVIATIONS

These abbreviations may appear in this document:

ASA – Automated Systems Alliance, Inc.

BI – Business Intelligence

DW – Data Warehouse

PCDW – Project Control Data Warehouse

ODWM – Open Data Warehouse Model

OPEN SOURCE and OPEN DATA WAREHOUSE

The phrase 'Open data warehouse' is used throughout this paper to identify and create a distinction between an approach to product licensing that follows the basic tenants of the 'Open Source' and products that have an actual license that has been approved by the Open Source Licensing Committee (Open Source).

CURRENT STATE OF DW and 'OPEN' DW TECHNOLOGY

As of the initial release of this document, I'm are not aware of any literature concerning the design of 'Open Source' or 'Open Style' types of data warehouse architectures except those which have been created by ASA.

ASA released the first 'Open Style' architecture for Project Control in March of 2009 and is currently working with its partners to complete the first completely 'Open' DW/BI project stack.

There is a considerable, mature body of publically available literature on data warehouses, business intelligence, data modeling, analytics and related topics. This document does not significantly build on that body of knowledge.

WHAT IS ODWM?

The Open Data Warehouse Model (ODWM) provides a framework for delivering an Open data warehouse. ODWM is a way of thinking about delivering a generalized data warehouse model that is tuned to a specific industry 'slice'. ODWM is not a data warehouse architecture itself.

The central tenants of ODWM are:

1. There is an organization responsible for licensing, distributing, and maintaining the Open data warehouse model.
2. The people that build the generalized Open warehouse have to have deep knowledge of the industry 'slice' they are constructing and engage a community in developing the Open data warehouse ecology.
3. The Open data warehouse is platform agnostic (works on most OS and database platforms).
4. For a specific industry 'slice', a generalized Open data warehouse can be built that will service the information needs of that industry 'slice'.
5. The data warehouses tend to be big and wide. Hundreds of potential dimension tables with thousands of fields form a 'foundation' set of information. OLAP cubes and data marts can be constructed from the foundation as required by the business.
6. The generalized Open data warehouse must be stable enough to permit use of commercial ETL bridges yet allow for interpretation through aggregation and by other means.
7. The generalized Open data warehouse must allow for integration of multiple data sources of different granularity and should, in some fashion, accommodate slowly changing dimensions.

INDUSTRY 'SLICES'

Most data warehouses have a focus that is both industry specific and focuses on a specific portion of the business. For instance, you probably won't find an automotive data warehouse but you will find that there are Automotive Marketing data warehouses. An Automotive Marketing data warehouse could be considered a candidate industry 'slice'.

The Project Control Data Warehouse (PCDW) is an instance of ODWM. As an example of the Open data warehouse approach, it is somewhat of an exception in that its scope is pan-industrial. This is possible because projects have similar goals regardless of industry. I suspect there are a few but not many pan-industrial 'slices'.

WHY 'OPEN'?

I could spend time telling you that that there is a great social good that is to be had through adopting or developing Open data warehouse solutions. Please feel free to construct your own, should you feel so inclined, and insert it [here]. The fact is, Open data warehousing is a solid business proposition.

There are two important business ideas that will drive Open data warehousing: 1) Open does not mean free, and 2) Economic opportunity will drive adoption.

Open does not mean free. Creating an Open product from a service does not eliminate the service; it changes the service. Open data warehouses are architectures or, if you will, containers for data. To make them functional requires creating ETL processes (to load data), attaching Business Intelligence and analytic systems (to report on and evaluate content), creating security 'fences', engaging training, etc. In this Open data warehouse model, service is no longer focused on creating the data warehouse (the custom model) but on making it functional. Not everyone will need or want services or products from external providers but some portion of the adopters will. Because this offering targets the midmarket, this 'some' could be quite large.

Economic opportunity will drive adoption. Many of the service elements identified above can be made into products (we can identify at least two new product markets). These support products are not required to be Open Source and they don't eliminate service; they make services more efficient. Economic participation from a number of independent, loosely linked parties will be economically attractive to developers.

BUILT TO FAIL

It takes time to build a custom data warehouse; often a year or more. The problem is, many businesses change far more rapidly than that. There are now large companies that are re-budgeting quarterly in order to respond more rapidly to business change. How likely is it that any 'final' rendition of a data warehouse, as originally specified, is going to meet the organizational need in a year or even six months after you begin building it?

One way to approach the problem is to build a very large, generalized 'foundation' architecture that can accommodate and interrelate a lot of domain data. This allows for current and future flexibility. OLAP cubes and data marts can be built from foundation content to satisfy specific and or high performance needs.

Who wants to begin a multi-million dollar project where the success is dependent on the business standing still? We need Open data warehousing to reduce the risk of project failure and premature obsolescence.

THE MARKET

Open data warehousing reduces the initial cost and risk of data warehouse adoption making it a far more palatable solution for the midmarket. While there may be some Open data warehouse adoption at the top of the economic pyramid, this technology will likely see its greatest and best use in the much larger, untapped, midmarket.

DOES THE MIDMARKET NEED DATA WAREHOUSING?

Organizations in the midmarket have the same problems as those at the top of the market. Solutions and value that can be provided through a data warehouse and reporting stack may largely be the same. The need for a data warehouse will largely be determined on organizational need for information dissemination, reporting, and analytics.

DATA PORTABILITY

The rise of SaaS solutions and the 'Cloud' along with vendor lock-in has surfaced another potential market for the Open data warehouse.

SaaS vendors have been known to lock-in customers contractually by 'holding' their data. One reason given for not releasing data back to a customer is that it would expose the SaaS vendor's proprietary database. This is actually a fair complaint.

SaaS and other application databases are structurally very different from data warehouse architectures. As such, a data warehouse could be used as an intermediate container for SaaS data. To avoid another form of vendor lock-in, this solution is best implemented by using an Open data warehouse.

Some SaaS vendors probably aren't going to like this solution because it lowers the barriers to replacement and it removes the chief complaint against data transfer. However, if customers insist on this type of functionality it would quickly separate out the SaaS providers interested in providing service from those benefiting primarily from vendor lock-in. Smarter SaaS providers will understand this is a mechanism for attracting new customers as well as providing new, enhanced reporting solutions.

WHY WILL ODWM WORK?

There will be a lot of winners. To be clear, winners as its being used here, means financial winners. Most Open Source Software (OSS) systems have an OSS provider and a single type of consumer. The ODWM has a broader group of potential participants that can benefit from each individual model instance. Because of its open nature, each of these groups can benefit financially without necessitating interdependence.

WINNERS

Each of the winner groups are profiled below. Owners get low cost, rapid entry into a data warehouses they can extend. Developers get to create/sell new ETL/BI products in a new market. 'Source' vendors can solve reporting problems and advance new ways to compete. Consultants get a bigger market for their services. Individuals can participate by creating new open data warehouses using their deep industry knowledge. These people have garages.

Open Data Warehouse Developers

These are the individuals or organizations that decide to develop an Open data warehouse for a specific industry 'slice' in which they have deep domainal expertise. Hopefully this paper has given you an understanding of what it would mean to own a domain standard. So, stake out your territory and claim your 'slice'. It's a perishable opportunity.

Owners

These are the consumers of the Open data warehouse solutions. In addition to the normal reporting and analytic advantages provided by a data warehouse, reduced cost and lower risk are the most obvious advantages of the Open data warehouse solution for this group. Data portability for SaaS (Cloud) solutions is another potential, emerging market for these products (see discussion above).

Developers

These are organizations or individuals that build and resell data warehouse tools. There are two new product markets created by Open data warehouses: ETL bridges, and Reporting/Analytic packages.

Neither of these has been marketable before because, to this point, data warehouses have been custom affairs.

Tool Providers

These are organizations that have existing products (commercial and Open Source) that can be attached to Open data warehouse solutions. Many of these products were originally designed to work with data warehouses. This restricted their sales range to the upper part of the market where data warehousing has been most extensively used. Making data warehousing available to the midmarket via the Open data warehouse concept exposes these tools to a new and larger market.

Source Providers

These are organizations that have applications that hold 'source data' that can be consumed by the data warehouse (ERP systems, CRM, etc.). The Open warehouse concept can address numerous reporting challenges for this group and make their product more competitive. Source providers can also build and sell ETL bridges from their product and attract new customers by making it easier to move to their platform.

Service Providers

These are organizations or individuals providing Open data warehouse services, training, and support. The Open data warehouse model, because it focuses on industry 'slices' favors the specialized (industry focused) consultancy over the larger professional service providers. Engagements are shorter and less risky than the longer (1 year+ custom) affairs.

LOSERS

Creating something new often means disrupting or destroying something that already exists. The most obvious loser in this will be the larger professional service (PS) organizations with data warehouse practices.

Long term (1 year +) contracts where 90%+ of the revenue is generated by labor will disappear with the Open data warehouse model. Since the risk is reduced, more, smaller consultancies and contractors can provide these services and compete with the larger PS

organizations. This may also have the affect of further reducing associated labor costs.

CONCLUSION

I built my first 'official' data warehouse almost 15 years ago. Last year I walked in a data warehouse class and was shocked to find that almost nothing had changed in the 10 years since I had last had training on the subject. I probably should have known. Software technology has to be pretty crusty for hardware to grow up around it.

What's interesting is that even though the data warehousing approach is stagnant, the industry is quite vibrant with endless incremental rifts and repeating themes. Stuck and vibrant is an interesting combination that you just don't see that often seen in the technology world.

I've always found it troubling that the DW/BI world has focused most of its attention on the top of the economic pyramid. I'm not a technologic socialist. I just see much more opportunity a bit further down-market. The great irony is, many organizations that can afford DW/BI are the ones who can least use actionable information from these systems. Often these organizations are simply too big to respond fast enough to the information provided to be effective. Smaller, more agile organizations that could respond to actionable information simply can't afford the cost and risk of the current DW/BI systems.

Current economic conditions have turned off the spigot at the top of the economic pyramid. DW/BI product and service providers alike have begun looking to sell 'down-market'. However, using the same expensive custom DW/BI model in the midmarket isn't going to work. Simply put, labor, not tools, drives the cost of the DW/BI proposition. The only way to penetrate the midmarket is to reduce the cost and risk of data warehousing. I believe that Open data warehousing will do both.

In the end, the ideas in this paper are not going to Open up data warehousing. If data warehousing is going to be Opened, then the collective 'we' will do it. I'm here to point out that there is a considerable amount of technology 'brush' piled in the data warehouse gully, the winds are blowing, and that matches exist.