

## Escape from Data Jail: Getting business value out of your data warehouse

Monica Woolmer, Catapult BI, (Formally Formation Data Pty Ltd)

Does your organisation have data but struggle with providing effective reports and/or access for analysis?

Does your organisation have many reports available but struggle with utilising the data in the support of decisions?

Has previous success with the data warehouse not turned into expected future successes?

This paper highlights insights into how to unlock the data and provide greater business value. The target audience includes both business users and those with technical responsibilities.

The topics covered include:

- Data Warehouse Value
- What is Data Jail?
- Types of Data Jail
- Keys to unlocking data
- Approaches and Examples
- Lessons Learned

### Data Warehouse Value

Let's start with a discussion of data warehouse value. Value is obtained when the business / organisation takes **action**. Action is taken based on **decisions**. Decisions require information and information requires appropriate **data**.

Caution! Do not confuse information with data. If a report is generating many pages of detailed data that does not mean there is information that can be gleaned in support of a decision. Ask yourself, "What about this report supports a decision? What about this report makes it actionable?" Is the report merely a dump of raw data or is it organised to provide the key pieces of information required by the targeted audience?

## What is Data Jail

Data Jail occurs when the organisation struggles in turning data into information. Even a successful data warehouse can experience data jail as the data warehouse expands to include other areas of the business or with an increase in the sophistication of use (e.g. Active/Real-Time Data Warehousing or support for advanced analytics and data mining).

There are many types of data jail and I am going to focus in on three types that I have had direct experience with at different organisations:

- Unsuitable user tools
- Inappropriate data format
- Insufficient data quality.

## Unsuitable User Tools

Does your data warehouse support different types of users, each with an appropriate tool? Ranges of data warehouse usage go from reporting to analysis to advanced data mining. Even within the category of reporting there can be different classes of users and usage including scorecards, dashboards, summary and detailed reports as well as alerts. So if you are treating all the users the same, then some of them are experiencing data jail.

Query inflexibility is another type of data jail. Different types of report and analysis usage need different types of queries. Queries for advanced analytics are very different than queries used to monitor and alert. Multi-dimensional analysis is a useful method of providing users a platform for business analysis, however if there is no ability to go deeper than the planned dimensions allow, for example to find out the specific customers behind a given number, then the overall effectiveness of the environment is reduced.

Finally, what about providing support outside the typical realm of business intelligence? Application support can range from supporting scoring applications to moving into the real-time space of direct updates and queries from a CRM application. If your data warehouse was designed and tuned exclusively to support decision-support types of usage then you could experience data jail and struggle with adding in the support for applications.

## **Inappropriate Data Format**

The various types of users and usage of the data warehouse will require differing data formats. If all users were expected to write SQL queries against the relational model, many of them will find the process difficult and therefore not use it as often or to the full potential.

Lack of consistency is another type of data jail. All of us have heard the desired target of “Single version of the truth”. The reality is most organisations have at least some type of reporting off both the data warehouse and the source system. If it is not clear as to when each system is to be used, or if the results achieved are inconsistent, then there will be uncertainty and a natural lack of trust in one or both systems. Even worse for the data warehouse is if there are two applications, cubes or data marts built from the data warehouse that are inconsistent with each other!

A less obvious way of locking up the data is when business transformation rules must be applied on the way out. That means that each query and report that utilises the information must then include the business rule. As the usage grows, so also is the likelihood of getting it wrong. What about a future change of the business rule? It would be difficult at best to impossible for most organisations to ensure they have made the change in every query and report impacted.

## **Insufficient Data Quality**

The last of the three types of data jail covered in this paper deals with data quality. I specifically chose the term insufficient since data does not have to be perfect to be of use to the organisation. Most marketing areas can use data that is “directionally accurate,” however in contrast financial data stored in the data warehouse is generally expected to match what is reported by the general ledger. If the data is not fit for purpose, if the quality is not up to the standard required by the target audience, then there is an issue.

### **Keys to unlocking your data**

Now that you have an understanding of the concept of data jail, next we will look at some keys to unlocking your data. The keys include:

- Key #1: Enable / Empower business users

- Key #2: Focus on the data
- Key #3: Focus on data quality
- Key #4: Take advantage of the power of your database

## Key #1: Enable / Empower business users

As discussed previously the varying types of users and usages of the data warehouse need different types of tools. The requirements of the tool need to match the intended use. Tools will be needed to support all types of usage including reporting, analytics, exploration, data mining as well as application support.

You just cannot leave it at providing tools, as appropriate skilling is also a key. Skilling needs to cover the spectrum from business to technical. A data warehouse user within a business area needs an understanding of the tool, an understanding of the data as well as a business understanding of the target for the data, the appropriate types of questions to ask. One of the marketing departments I previously worked with provided an “Introduction to direct marketing” course that included the strategies and goals behind direct marketing along with how to use the established query tool to achieve the desired outcomes.

**Appropriate skilling – technical and business**

<b>Business Skills</b>		<b>Technical Skills</b>
------------------------	--	-------------------------

- Develop and encourage a spectrum of skills across the organisation
- Include education on the business itself as well as the data

An often overlooked aspect is to plan for growth over time (i.e. growth in terms of the number and types of users as well as an increase in user sophistication). As a business area starts to utilise the simple reports and queries they will then want to move to basic exploration and business analysis. Advanced users will be interested in data mining and statistical analysis. Finally organisations can grow to the sophistication of wanting to utilise the data warehouse in real-time to support application access. The key here is to plan ahead and realise that the users will continue to demand more of the data warehouse.

My final topic within this key of empowering business users is to not 'over grant' or 'under grant' access. By over grant access I mean giving users access to more sophisticated tools or a broader view of the data warehouse than they need to satisfy their business needs. By contrast, artificially limiting access to the full relational model to advanced data miners will limit their ability to draw out previously unknown relationship contained within the data itself.

## **Key #2: Focus on the data**

To start, plan to build an Enterprise Data Warehouse. Build it incrementally over time, however if you focus in each time on the fact that the end goal is an Enterprise Data Warehouse, then the decisions that are made within each project will take the goal into consideration. This will lessen the risk of implementing pieces that might have to be redone later when additional subject areas or users are added.

In the same manner, build the data warehouse from detailed data. The detailed data will provide the greatest amount of flexibility over time to provide new applications, support new users, etc. Keep in mind that the amount of history required is a business decision, not a technical one. Use a relational model as the base as this provides greater flexibility to support a variety of different types of request. There will be the need to de-normalise off the relational base. This includes creating a presentation layer - providing specific subject views of the underlying relational data or presenting the data via a dimensional model that works well for many of the BI tools. Finally, there may be key summary tables that are built to capture business-rule driven summarisations (as opposed to straight aggregations).

## **Key #3: Focus on data quality**

I previously mentioned insufficient data quality as a type of data jail. The key is to establish a data quality program. Begin with identifying foundational principles such as "Data Quality begins at the source." Next define what 'fit for purpose' means for your organisation in general and for the specific business area. Establish thresholds for monitoring the quality of the data being loaded.

An approach for testing and validation of the data should be developed prior to any data load being implemented. Use the established approach for testing prior to deployment as part of acceptance testing and for the ongoing monitoring of data quality. Plan to measure and make improvements over time focussing in on priority areas that align to business priorities.

---

## Key #4: Take advantage of the power of your database

All the databases have advanced features for tuning and optimising a data warehouse. It is important to understand and utilise all that the database has to offer. Following are a few examples of key features of Teradata (a powerful data warehouse management tool) that help to unlock your data.

One of the basic features in Teradata is views. The power of views is that they appear the same as tables to third-party tools. This enables the creation of a presentation layer that hides the complexity of the relational layer or presents it in a different format like a dimensional model. The presentation layer simplifies report development and enables reports to be written by business users, which increases self sufficiency.

Indexing is another key feature. This includes basic indexing such as establishing the appropriate primary index as it is utilised as the hash-key as well introducing appropriate secondary indices when required. Teradata also supports several advanced indexing techniques including:

- Partitioned Primary Index (PPI) - Horizontal partitioning of a table. Rows with the same partition value are physically grouped together.
- Join Index (JI)– An index that consists of the keys required to join tables together. The index is maintained by the database and improves query performance.
- Aggregate Join Index (AJI) – Similar to the join index, as the name suggests, the Aggregate Join Index allows for specification of aggregations (Sum or Count). A Sum AJI contains a hidden column containing the row count, so that AVERAGE can be calculated from the index. AJIs are useful for presenting the data in a multi-dimensional manner.

Another table-related feature within Teradata is Table Re-distribution, also known as a Reversal Table. Maintained by the database, this is a copy of a table physically stored via an alternative primary index (hash key). The cost of the physical storage needs to be compared against the benefits for queries to determine when this feature should be utilised.

Data warehouse systems include features for workload management and these must be utilised effectively to ensure adequate performance service levels by the different types of users. Teradata Active Systems Management tool provides mechanisms for both “gate keeping”, identifying users and their specific queries to determine if their request can run now

or must wait until later as well as “priority management”, assigning queries to priority groups and monitoring to ensure they do not exceed set thresholds for that group. Effective workload management is a key to having an efficiently tuned system and provides users with more consistent experience over time.

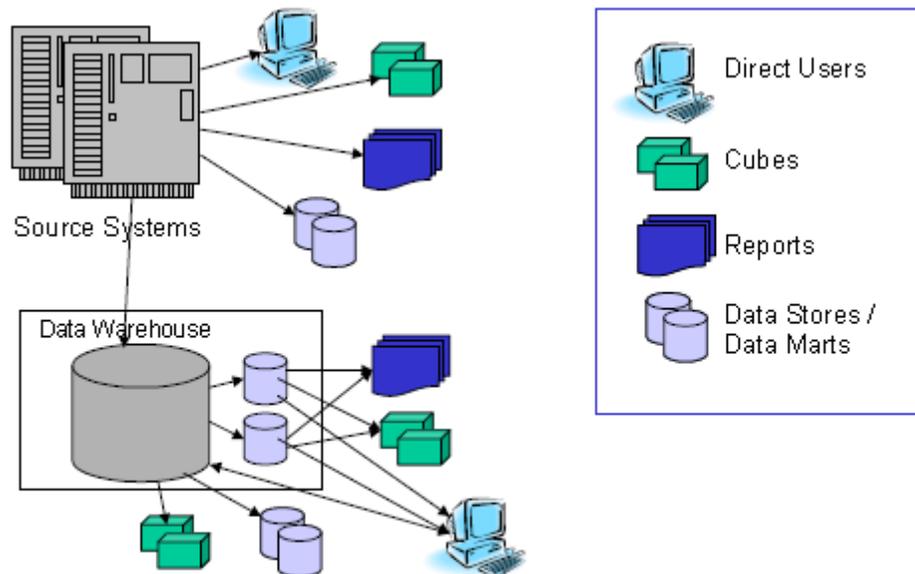
Finally, take advantage of the materials produced by the vendor, including manuals, research papers, webcasts, podcasts and formal courses.

### Approaches and Examples

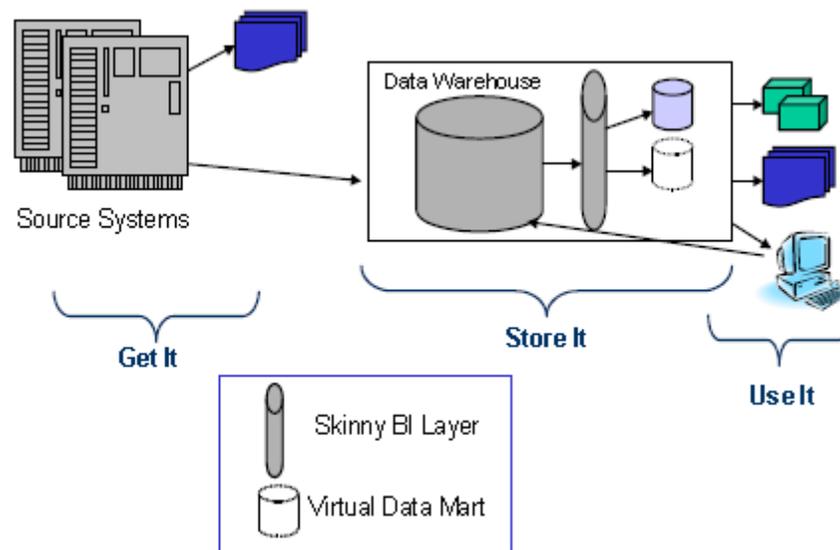
Let’s switch gears and walk through a few examples.

#### Example #1

The first example had many direct users, cubes, reports along with data stores and data marts, which were built directly from the source systems as well as a data warehouse. This led to inconsistent results being produced as well as confusion as to exactly when to go to the data warehouse or to the source system.



To start with a strategy was developed and communicated. It had a simple message behind it, “Get It – Store It – Use It.”

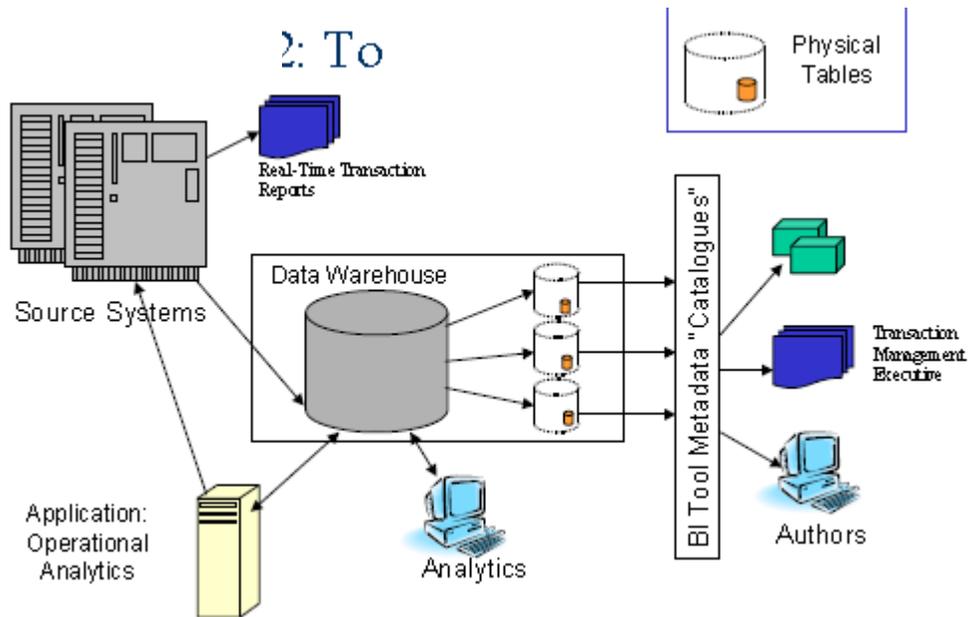


The data warehouse was established as the go-to point for direct users, cubes and reports. The only exception was the continued to use targeted transactional reports off the source system. These reports complemented the set of reports available in the data warehouse.

The architecture introduced a Skinny BI Layer and Virtual Data Marts. The skinny BI layer held common rules-driven transformations. For example, the main data warehouse contained the base financial data as represented in the general ledger and the skinny BI layer contained the rules for a consistent method for applying returns and back orders. A virtual data mart is equivalent to the presentation layer mentioned previously. These were a series of views used to provide a business area view into the data warehouse.

## Example #2

The second example was very similar to the first. Direct users, cubes, reports along with data stores and data marts were built directly from the source systems as well as a data warehouse, however the usage was more distributed and therefore the number of cubes and data stores increased.

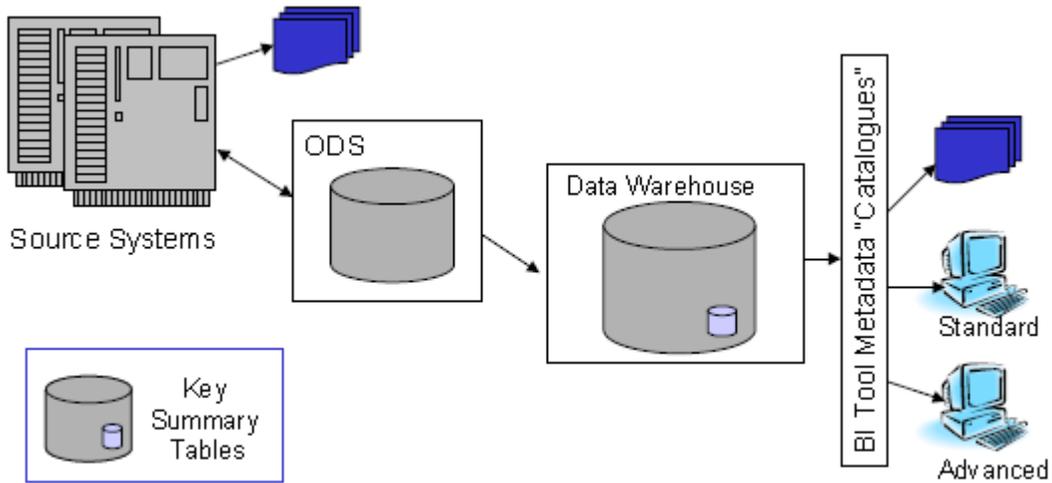


The target environment was established based on an Enterprise Reporting Model and associated BI Architecture, which supported the continuum from reporting to analysis to advanced data mining. "

The data warehouse was built as a relational model based on an enterprise logical data model. A series of data marts were designed using a dimensional model and were a combination of views and physical tables where warranted.

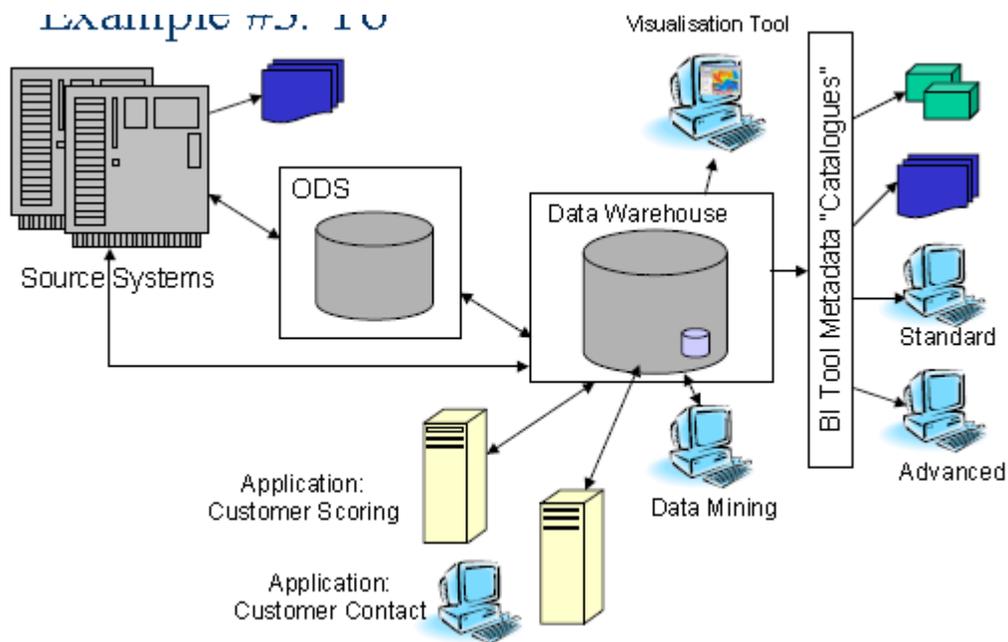
### Example #3

The third example is an interesting example in that there was not the disparate usage as seen in the first two examples.



A centralised data warehouse was being effectively utilised. The data warehouse was built utilising a relational model with key summary tables. The issue experienced here was there was only a limited toolset available.

An expansion and evolution of the data warehouse was implemented through a series of projects.



The features added included:

- Supporting direct feeds to and from the source systems
- Supporting an application for customer scoring
- Supporting real-time updates and interaction to web-based CRM tool
- Introducing Cubes for business analysis
- Introducing visualisation tools

## Common concepts

There are some common concepts behind the three examples:

**Establish the system of record** – Meaning the recognised source of the data. This may be an operational source system and the data warehouse needs to match the source. The data warehouse itself can be the system of record for example, as the application platform or as the historical system of record where the data has been archived off the source system.

□ **Shared data warehouse layer** – A relational data model as the base provides the flexibility to support a variety of users and types of usage.

**Dependent over independent data marts** – Dependent data marts are those built from the data warehouse. These are preferred as the data warehouse can help ensure consistency, for example in applying business-rule driven summaries that can be utilised by the different data marts.

**Dependent data marts moving towards virtual** – Moving towards virtual means implementing the data marts as views and advanced indices instead of or in addition to physical tables.

**Business Intelligence or presentation layer** – Representing the underlying relational model in a more business friendly and BI tool friendly manner.

**Different classes of users / usage** – Supporting the spectrum of users and types of usage from reporting to business analysis to advanced statistical analysis and data mining.

**Support Users and Applications** – Supporting users via the appropriate tool for the target usage as well as providing direct read/write support for applications.

## Lessons Learnt

In closing, I wanted to share a few of the key lessons learned along the way.

## Separate functional concepts from technical implementations

Examples of separation are:

- Presentation Layer (functional) vs. Data layer (technical)
- Data Mart design (functional) vs. Physical data marts and virtual data marts (technical)
- Transformation Rules (functional) vs. transformation methods (ETL, ELT, EAI) (technical)

## Materialise only where needed

Materialise is the process of physically implementing a view or table. There will be valid business reasons for materialising key tables and views, including managing business rule driven summarisations. It is important to look at the cost-benefit analysis. The cost of maintaining the physical implementation must be less than the benefit achieved by the users. Benefits can include a reduction in CPU usage and improved query speed / throughput.

Where the database supports materialised views these should be chosen over maintaining a physical table as the views are automatically maintained by the database and integrity to the source table is guaranteed.

## Use portal technology

Use portal technology as a common access point for bringing different tools together. Portal technology enables the introduction of user roles. Each of the user roles should be given access to the appropriate tools as well as the appropriate content.

## Manage environments

Environment must be managed! Plan for mixed workloads and establish Service Level Agreements by application / usage types.

## **User driven content to be able to be added**

Enable user-driven content to be added back to the data warehouse (e.g. providing the results of statistical analysis models). Ensure that this process is controlled. A good practice is to not let the user update the main tables directly, rather allow the user to load to a staging area and then the production ETL process validates and then loads the data.

## **Reality = Evolution not Revolution.**

This is especially true for existing data warehouses where existing users must continue to be supported while changes are being made to improve the overall environment for the organisation. The culture of the organisation plays a role, for example, in determining how fast or slow changes can be introduced.

## **Remember it is not all technical.**

As stated by Rob Armstrong in Teradata Magazine (March 2007), "Being successful in data warehousing and driving true business intelligence (BI) is more about aligning politics, culture and processes."

Whether you are a business user or have technical responsibility, look for opportunities to unlock the data to provide greater business value. I hope you have read something that causes you to nod in agreement as well as paused to make you think.

© Dialog Pty Ltd ABN 16 010 089 175 - Ph: 1800 DIALOG [dialogit@dialog.com.au](mailto:dialogit@dialog.com.au)

This work is copyright. Except as permitted under the Copyright Act 1968 (Aust Commonwealth), no part of this publication may be reproduced by any process, electronic or otherwise, without the specific written permission of the copyright owner. Neither may information be stored electronically in any form whatsoever without such permission. This document does not constitute advice and Dialog recommends that readers exercise their own care, skill and diligence with respect to the use, interpretation and reliance on this document. Trademarks mentioned are the property of the owner.