

---

**Santiago Castro**  
is senior consultant at Vine Cube,  
a joint venture between Web  
Synergies in Singapore ([www.websynergies.biz](http://www.websynergies.biz)) and TAH in the  
United Kingdom ([www.tah.co.uk](http://www.tah.co.uk)). The combination of Web  
Synergies' technical expertise  
and TAH's business intelligence  
domain expertise offers wider  
choices to organizations wishing  
to extract maximum value from  
their technology investments and  
the data within and around their  
business.

**Keywords:** big data, data  
management, business intelligence

**Realizing value is  
critical**

**Change in the storage  
and analysis paradigm**

**Defining value before  
storage**

Santiago Castro  
Vine Cube (UK), 30 Factory Road,  
Upton Industrial Estate, Poole,  
Dorset, BH16 5SL, UK.  
Tel: +44 (0)751 525 2877  
E-mail: [santiagocastro@yahoo.com](mailto:santiagocastro@yahoo.com)

# Opinion Piece

## Optimizing your data management for big data

*Santiago Castro*

### Abstract

With the growth of big data, there is a risk of data management costs increasing before the business has seen any value from these new sources. This article outlines an approach to data management optimization that increases the volume of data being analysed, while raising the likelihood of realizing its value and reducing the risks of building a full-scale data warehouse.

*Journal of Direct, Data and Digital Marketing Practice* (2014) **16**, 15–18.  
doi:10.1057/dddmp.2014.46

How many times have you heard about the promises of big data? How many times have you heard that, with the exponential increase in volume, velocity, variety and veracity of data (known as 'the four Vs'), insights and decisions will improve like never before? However, in order to get there, one must also recognize the challenges of big data, most importantly how to gain value from it. If you like, without the fifth V — value — all the rest are just big challenges.

Big data is real, but is yet to realize its promise. The recent exponential growth of data makes it harder and harder for organizations to generate the value they need out of the other four Vs of big data. Therefore, then, how to realize that value? A change of paradigm is required. Previously, the storage of data used to happen in one place and computation and analysis in another. Owing to the exponential growth of the above-mentioned four Vs, this cannot continue to be the case. Storage, computation and analysis all need to be more flexible and they all need to happen in the same logical location.

In addition, the cost of keeping data is mis-matched when compared with the value it provides to the business. The question now is 'How can I extract value from the data?' not 'How can I keep it?' Understanding that value needs to happen first, before even considering whether it is worth keeping and, eventually, investing in storing it. Not only is keeping data costly but also, in general, its value tends to decrease with time. In sum, an important challenge with big data is that data tends to be less relevant with time, while at the same time more expensive to keep in the long term. Hence, with big data, the value question has become the priority.

**Ask first, invest later**

However, there is hope. With powerful data discovery and visualization tools like Tableau, QlikView or Necto (Panorama Software), the cost of asking questions has decreased. These tools allow us (without even needing to write any code) to access, explore and analyse data easily for many different sources outside the traditional data warehouse. Therefore, it is now possible and more efficient to start interacting with data by asking business questions first, before investing in a data warehouse.

**Lowering the cost of curiosity**

Playing with data is no longer restricted exclusively to IT roles. At the same time, lowering the cost of ‘curiosity’ for business users stimulates innovation, realizing efficiencies in the business and supporting self-service. Furthermore, Analytics and machine learning can now happen on the source system, without needing to migrate and generate additional costs unless their value has been proven to justify the costs.

**No need for data marts**

Moreover, the previously mentioned data discovery and visualization tools also now make it possible to look at data without needing extraction, transformation and loading (ETL) processes. With the addition of user-friendly data blending, enrichment and processing tools, such as Alteryx, business departments like marketing, customer insight and operations are empowered to get the specific data they require to perform their analysis with no need to know querying languages, like SQL, or write any code and with no delays — as fast as drag and drop — instead of waiting for IT to build another data mart. The advantage of this approach is that it allows business users flexibility and the ability to add new data sources in order to prove their relative worth before needing to commit IT resources in order to bring these into a traditional data warehouse environment. It provides, if you will, an agile and innovative proving ground for data and analytics that can be drawn from, in order to allow the more costly data warehouse build process to concentrate on gathering only the data that truly adds business value.

**Single version of the truth**

Subsequently, powerful data integration tools, such as WhereScape Red, allow you to bring proven high-value source systems and data into the data warehouse very quickly (in weeks, as opposed to months). These tools have considerably reduced the time and effort needed to structure data, automatically generating information flow diagrams and documentation. Here is where a robust methodology is needed (we call it ‘Time2Value’) in order to ensure the best use of these tools is made and to guarantee that only one version of the truth is produced and maintained in a centralized data warehouse.

**Larger window of data for analysis**

With this approach, the importance of the data warehouse is not being questioned. On the contrary, we believe that a data warehouse is good at delivering performance improvements, safety, consistency and one version of the truth and at limiting risks. However, before data is introduced to this formalized and scheduled environment, it is now possible to interact with much more raw data in what we call the ‘data hub’ (see Figure 1), still producing value in a quick and economic way and only spending additional resources to bring more structure to the high-value data that is worth storing. The change in context is that new data visualization tools, combined with faster data integration tools, allow optimal combination and

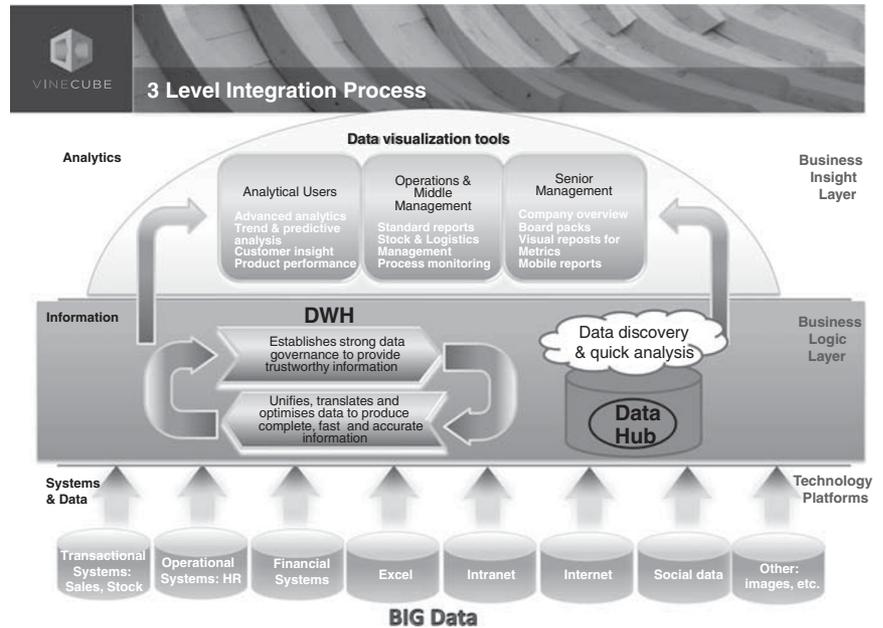


Figure 1: New data flows between data warehouse and data hub

exploitation of both the data warehouse and this new data hub, dynamically enlarging the ‘window’ of data that can be accessed in the analysis layer.

**Focus on the three business values**

Generating extraction, loading and transformation (ELT) processes instead of ETL is very easy and, as a result, data can be available immediately, before there is even a need to transform and process it. Therefore, the data warehouse can be used more efficiently, focusing exclusively on gathering and processing data that has a proven and repeatable value. Data visualization and analytics tools can then be plugged into both the data hub (where value from all sorts of big data is realized quickly at a lower cost) and the data warehouse (where queries can retrieve structured, high-value data worth storing). This speeds up the process, making it more economic and flexible. This approach not only brings agility, but also offers operational efficiency (doing it cheaper and faster) combined with an information advantage (quicker exploration and analysis of a wider view of information), while focusing on the three main business value variables: less cost, lower risk and more opportunity.

**Optimize before migration**

All this opens the opportunity for real data management optimization, which can now truly be based on an upfront understanding of what data is viable for migration and what can be justified from a cost-versus-value point of view, and, as a result, enables the construction of a real business case for the data warehouse before committing to the investment. A data management optimization process also means lowering risks by starting on the data hub side with the identification of the most valuable targets for migration first, before proceeding with the steps of profiling, prioritizing and migrating into the data warehouse.

**IT as custodians of  
management  
information (MI)**

In conclusion, a change in context is prompting a change in paradigm. Until now, there has been relative stability in the way management information (MI) or business intelligence has been developed and maintained by the IT department and delivered to the business. In addition, it was often believed that MI could only be delivered to the business user by IT utilizing strong development governance and IT-centric change control, maintaining the data processes and being custodians of the data.

**Shift to business user  
tools**

However, new tools (as referenced earlier) and methodologies (business discovery, business user self-service, Time2Value) are challenging IT's control. This challenge comes from the emergence of ELT instead of ETL, demand for real-time in-memory analytics and business discovery provided by business user tools that are the equal to the heavy-duty tools that IT has been using. Given the new tools available, if the enterprise adopts our suggested approach to data management optimization, the following competitive advantages can be realized:

- Lower costs of data warehouse management;
- More flexibility and agility for an improved quality of service;
- Enlarged window of analysis and optimized data warehouse capacity for a higher return on investment;
- Real opportunity to interact and retain all (big) data for analysis.