

How to swallow an elephant

(or Productising analytic infrastructure, if you prefer)

Carriers, in common with other large businesses, have huge amounts of data to analyse in a short space of time. Here, Charles Garry, an independent consultant and former vice president and director with META Group's Technology Research Services organisation, explains how productising the process and creating the data warehouse appliance can reduce cost and increase efficiency

Computer scientists have spent decades considering the problem of how to effectively process ever-increasing amounts of data. The answer has always been somewhat akin to the question of how one would eat an elephant (small bites please).



Charles Garry,
an independent
consultant

In other words, overcome the big problem – a huge amount of data to analyse in a short amount of time – by breaking it down to a large number of small jobs running in parallel. For example, assume you had an acre of grass to mow. If you had ten people and ten lawn mowers attacking the problem you could accomplish the job in one-tenth of the time.

This is essentially the theory behind what is known as massive parallel processing and how today's largest corporate data warehouses are capable of handling the increasing volume of corporate data. All of this sounds relatively simple and straightforward; however, massively parallel computing infrastructures are both expensive to buy and complex (for which read expensive) to manage.

To make matters worse, organisations of all sizes are becoming increasingly dependent on data analytics. Indeed since 2000, data warehouse/business analytic infrastructure has become or is in the process of becoming a business critical application for most companies.

What secrets does the data hold? Companies have always searched for better ways to understand their customers, and anticipate their needs. They have longed to

improve the speed and accuracy of operational decision-making. In short, they wanted to know all the secrets hidden within the massive amounts of ever-increasing data volumes.

While the desire to improve the analysis and timeliness of an organisation's data has been felt for over 20 years, the practical capability to do so has eluded all but the largest IT shops. However, powerful trends have been impacting the data warehousing space over the past few years. These trends are creating a convergence of an organisation's historical desire to derive value from the data, with the opportunity – and, more importantly, the capability – to address the growing demand for business analytics with a simpler, cost effective approach.

Now that most IT organisations have implemented the large ERP (enterprise resource planning) packages and have Web-enabled most key customer applications, the focus of attention has moved towards data warehousing and analytics. Technology innovation continues to drive down the costs associated with server processing power and storage capacity (33% annually).

Software licensing costs are also beginning to be impacted by these trends as well as the growing influence that open source software is having on commercial software licensing and pricing. Greater computing capacity at a lower cost equals an opportunity to redefine what "big" means with respect to a data warehouse or data mart. Multi-terabyte sized analytic stores will be the norm, not the exception. ☛

So processing power is getting cheaper but organisations are chewing up that capacity as fast as it is available. By continuing to innovate how data is used, or even creating new classifications of data (such as sub-transactional data) organisations will continue to stress traditional analytic infrastructure. How can the complexity issue be addressed?

The data warehouse appliance

Since the concept of the data warehouse was first introduced, end users have wanted a solution that was less complex. Many end users wish they could simply purchase a data warehouse the way they purchase a payroll application. Unfortunately, business analytic needs are constantly evolving, making productisation of the warehouse



difficult. Even the word 'evolving' is inaccurate in the context of an organisation's business analytic needs, as it implies constant but slow moving change. The reality is that analytic needs within an organisation change very rapidly. Additionally, demands for immediate tactical analysis versus longer-term strategic analysis make analytic infrastructures inherently complex.

Innovative vendors are now emerging to attack warehouse complexity by taking advantage of many of the previously mentioned trends in hardware and software. While delivering a packaged data warehouse might be impractical, complexity can be addressed through the productisation of a warehouse or data mart's underlying infrastructure.

The data warehouse appliance combines the price/performance of Intel based processors, open source software and low cost disk storage in a single cabinet. The combination is purpose-built to handle analysis against terabytes of data quickly and simply. By using a massive number of CPUs, these data warehouse appliances are uniquely designed to eat the elephant that is a multi-terabyte analytic data store.

The market for data warehouse appliances is growing quickly. For example, Netezza, a vendor pushing the data warehouse appliance trend, offers its Netezza NPS system which scales from less than 1 terabyte of user data up to as much as 27 terabytes of user data. Other vendors are already rushing to market with similar solutions and users are buying. But why are large companies willing to take a flyer on such a new trend?

Total cost of ownership: The key differentiator

Total cost of ownership (TCO) is a major, top-of-mind issue for virtually every IT organisation today. Defining what TCO consists of can be ambiguous at times for many organisations. It can be defined as the

initial purchase price for the solution plus how long it takes for the vendor to deliver an acceptable working production environment.

Then we add the cost of maintaining or sustaining a well performing stable environment. It is this third piece that often comprises as much as 80% of the TCO for an application. This portion consists primarily of personnel costs to monitor and tune the system.

Since appliances are built specifically to address large analytic workloads, the time-to-value piece of the TCO equation is rather simple. Time-to-value is an extremely important metric because it directly drives an organisation's return on investment (ROI) for the warehouse or mart environment.

Some early adopters of the Netezza appliance have reported provisioning times of four hours to get a working sustainable analytic environment versus four weeks to do the same thing with an Oracle/Sun/EMC infrastructure. More importantly, the performance was 10-50 times faster on the appliance.

Clearly non-appliance vendors such as Teradata and IBM have also demonstrated

Since the concept of the data warehouse was first introduced, end users have wanted a solution that was less complex.


good time to value. They are total solution providers with well-defined configurable units that they can deliver quickly, based on their extensive warehousing experience and deep knowledge of their reference infrastructure's capabilities. However, IBM and Teradata are often used for enterprise-wide strategic initiatives that typically require customised solutions and professional services.

Right tool for the job

The data warehouse appliance may be used in the future for the same strategic purposes, but today there is so much demand for tactical and operational analysis that must be done quickly. Users should strongly consider using the right tool for the right job. So for example, if you are a telco that needs to query 18 billion call detail records (CDRs) daily to stay current with billings, the job is operational in its timeliness but fundamentally analytic in nature. The data warehouse appliance is able to accomplish this task in minutes versus hours.

The data warehouse appliance also shines relative to its traditional warehouse

infrastructure brethren in the area of maintenance. Appliances are 'load and go' environments. Since they brute force the data efficiently with a high ratio of disk to processor, creating a massively parallel query engine in a box, they don't require indexing. More importantly, they don't require any specific physical database design or hints to make the database optimiser use indexes designed so painstakingly by a DBA. So, organisations spend the bulk of their time actually querying data, not tuning the database to query the data. What a concept!

With the demand for data analysis increasing, IT organisations must look for the proper tools to address the fast changing needs of their business user clientele. While the data warehouse appliance may not be the same thing as a data warehouse in a box (or should we say cabinet) it does simplify the underlying analytic infrastructure. While no tool yet addresses the entire spectrum of analytic needs, the data warehouse appliance model is sure to be an option that most IT organisations will want in their analytic toolbox. 

The data warehouse appliance may be used in the future for the same strategic purposes, but today there is so much demand for tactical and operational analysis that must be done quickly.



2nd ANNUAL

Telecoms BUSINESS INTELLIGENCE

Wednesday 28th – Thursday 29th September 2005

Le Meridien Piccadilly, Central London

It's possible...

make more profit from your existing customers!

Hear presentations, case studies and personal experiences from the major telecommunications companies around the globe, such as Vodafone, O₂, SBC and British Telecom at **ViB events' 2nd Annual Telecoms Business Intelligence conference**

- Learn from experts in the global telecommunications industry
- Ask questions and listen to panel discussions
- Discover solutions to difficulties you face today
- Prepare for the challenges of tomorrow

For further information visit www.businessintelligence-events.com or email us at events@vibevents.com

Please quote reference: VNP

Senior Sponsor

SYBASE


Lead Sponsor

IBM

Sponsors

 Agilent Technologies

NETEZZA

 TRILLIUM SOFTWARE

A ViB events conference www.vibevents.com