



# Intelligent Data

Dave Kloc at Netezza explains how the pharmaceutical industry can satisfy its need for data analysis by identifying and employing data warehouse appliances

In an extremely competitive business environment, it goes without saying that it is absolutely critical for pharmaceutical organisations to bring new products to market quickly and cost-effectively. The average drug consumes eight to 12 years of time and resources from initial testing to market launch, and during that time it is impossible to get a 'single view' of each product. The volume of biological data is

doubling every six to nine months, structured and unstructured data is stored across disparate data marts, and inconsistent terminology prevents industry-wide data collaboration. The future success of pharmaceutical companies will rely strongly on how quickly and easily they can harness massive amounts of data to make intelligent business decisions.

## DATA INTEGRATION AND ANALYSIS

While biological data volumes are exploding, traditional data management and retrieval methods cannot keep up with advances in research and discovery. Despite the fact that the industry needs to utilise ever-growing amounts of data, companies are still coming up against a technology roadblock when they try to



analyse data or utilise business intelligence using their current infrastructure of general-purpose server, storage and databases. These older architectures were not designed to handle terabytes of data or complex analyses; the resulting patchwork of technology makes querying their data unnecessarily arduous. Gone are the days when arrays of flat files distributed across countless computers, storage facilities and operating systems could be used to manage the discovery and validation processes.

Data integration and analysis require data that is organised, available, readily searchable and easily queried. In order to store and re-analyse data over years of discovery and drug development, it is critical to provide a powerful analytic system that can warehouse and integrate massive amounts of laboratory and clinical data from disparate sources, performing fast ad hoc analyses while remaining easy to use.

Now that most IT organisations have implemented large ERP 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. 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, while processing power is getting cheaper, organisations are consuming that capacity as soon as it is available. By continuing to innovate how data is used (demanding increasingly

complex queries), or even creating new classifications of data (such as sub-transactional data) organisations will continue to stress traditional analytic infrastructures. So, how can these issues be addressed?

**ENTER THE DATA WAREHOUSE APPLIANCE**

Since the concept of the data warehouse was first introduced, end-users have called for a straightforward, easy-to-use solution. Many organisations wish that 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 analytic business needs, as it implies constant but slow-moving change. The reality is that analytic needs within an organisation change extremely quickly.

Innovative vendors have emerged to attack warehouse complexity by taking advantage of many of the previously mentioned trends in hardware and software. Some systems architecturally integrate server, storage and database into

a single, powerful and scalable appliance to process queries faster, and to facilitate complex data analysis that was not previously possible. These systems are purpose-built to handle analysis against terabytes of data quickly and simply. By leveraging an innovative, massively parallel streaming architecture, these data warehouse appliances are uniquely designed to 'eat the elephant' that is a multi-terabyte analytic data store. This enables on-demand analytics on medical claims for pharmaceutical companies for example, and during drug discovery processes for life sciences organisations. Faster data delivery to these industry professionals leads to better decision-making and more efficient drug discovery processes overall. The market for data warehouse appliances is growing quickly because customers are realising the benefits of the appliance approach. But why are large companies willing to take a risk on these appliances?

As an example, one company recently selected a data warehouse appliance system to provide on-demand analytics that enhance delivery of key market data to pharmaceutical



[www.samedanltd.com](http://www.samedanltd.com)



**Total cost of ownership (TCO) is a fundamental issue for virtually every IT organisation today. Defining what TCO consists of can at times be ambiguous for many organisations. It can be defined as the initial purchase price for the solution, plus the length of time it takes for the vendor to deliver an acceptable working production environment. The maintenance cost of sustaining a well-performing stable environment is then added. It is this third piece that often comprises as much as 80 per cent of the TCO for an application.**

industry professionals. It replaced an existing Oracle data warehouse solution that was unable to keep pace with their data requirement needs and subsequently led to an increase in new business opportunities.

Among the challenges the company was experiencing with its previous data warehouse solution was an inability to complete large database queries in an acceptable timeframe, increasing maintenance fees and the need to hire more database administrators to fine-tune the system to achieve tangible results.

#### About the author



Dave Kloc joined the Netezza Corporation as General Manager for Europe in September 2001 with the responsibility for all sales and operations activity in Europe, the Middle East and Africa. He is responsible for some of Netezza's very first customer sales worldwide, including telco giant Orange UK. Prior to that, he served an eight-year tenure as Director of Channel Sales at Hyperion Solutions. He also enjoyed five years at Arbor Software and three years as a consultant at Arthur Anderson.  
Email: dkloc@netezza.com

The data warehouse appliance they put into practice replaced the old system's server, operating system, database and disk storage systems and became the core component of a new grid-processing environment, which can scale to meet the needs of thousands of concurrent users, while also quickly analysing and segmenting data at unprecedented speeds. The company's NPS-powered grid system runs more efficiently and processes more than 40 queries – a number that was not previously executable. Time-sensitive sales information requests were converted from days to minutes and administration was reduced significantly – a testament to the appliance's ability to implement an immediate business impact.

#### TOTAL COST OF OWNERSHIP: A KEY DIFFERENTIATOR

Total cost of ownership (TCO) is a fundamental issue for virtually every IT organisation today. Defining what TCO consists of can at times be ambiguous for many organisations. It can be defined as the initial purchase price for the solution, plus the length of time it takes for the vendor to deliver an acceptable working production environment. The maintenance cost of sustaining a well-performing stable environment is then added. It is this third piece that often comprises as much as 80 per cent of the TCO for an application. This portion consists primarily of personnel costs which are needed to monitor and tune the system.

Since appliances are built specifically to analyse large terabytes of data very quickly, the time-to-value piece of the

TCO equation is rather simple. Time-to-value is an extremely important metric, due to the fact that it is the driving force behind an organisation's return on investment (ROI) for the warehouse or mart environment. One customer who replaced their existing, traditional infrastructure, saw a marked improvement with provisioning times of four hours to get a working sustainable analytic environment, as opposed to four weeks to do the same thing in their previous environment. More importantly, once implemented the performance of the system was 10 to 100 times faster on the appliance.

#### CONCLUSION

The data warehouse appliance outshines its traditional warehouse infrastructure brethren in the area of maintenance. Appliances are 'load and go' environments. Since they use brute force to achieve data efficiency with a high ratio of disk-to-processor, creating a massively parallel query engine in a box, they don't require tuning, indexing or the large amounts of database administrator time that is typically required of traditional data warehouse and analytic systems. Organisations spend the bulk of their time actually querying data, not tuning the database to query the data. It is evident that with an increasing demand for data analysis, IT organisations must look for the proper tools to address the fast-changing needs of their business user clientele.