kognitio

Competitive advantage from data

Lowering the Total Cost of Ownership (TCO) of Data Warehousing

If Gordon Moore's law of performance improvement and cost reduction applies to processing power, why hasn't it worked for data warehousing?

Kognitio provides solutions to business problems that require acquisition, rationalization and analysis of large and/or complex data

www.kognitio.com

The Kognitio Technology and Data Center has been at the forefront of the design and development of massively parallel technologies for more than fifteen years.

The center has over a hundred man-years of development experience and is unique in its latest development of Kognitio WX_2, a high performance analytical database that serves as a robust data warehouse platform

Contents

Traditional data warehouses struggle to provide the information required in timescales that are acceptable

Introduction

Moore's Law effectively states that processor power doubles every two years. This exponential growth in computing performance has resulted in a technology explosion, with the majority of the population having access to computers and the web. A side effect has been the dramatic increase in the amount of data captured by companies, which is doubling every year.

There is considerable value embedded within the data that companies capture, about their business and their clients. Bringing this data together from disparate sources into a central data warehouse allows complex queries to be performed in order to understand how the business is performing, how clients are behaving and to measure the impact of different initiatives such as marketing campaigns.

Challenge

With huge volumes of data and complex queries being demanded, the traditional data warehouse struggles to provide the information required in timescales that are acceptable. This has resulted in the development of data-marts that provide summarized data for analysis. However, these data marts are restrictive in what they can provide, and multiple data marts are often needed to meet the different needs of the organization. All of this adds to the cost of maintaining the data environment and the time required to react to the ever-changing requirements of the business. These marts may have been developed as a result of considered best practice at the time, but to quote Bill Inmon:

> *'Now, like Rip van Winkle, we wake up a decade or so later and where is the world? First off, the data warehouse is no longer a subject that people debate, or least the subject of debate has changed. Instead of asking 'should we build a data warehouse' people are asking 'how do we build a data warehouse?'. Enough people have been burned by the cluster of data-marts being called a data warehouse, but in reality not many people are fooled into thinking that they don't need a real data warehouse today. At about the fifth or sixth data mart, the organization steps back and realizes that a cluster of data marts is not giving them what they need.'*
>
> **Inmon: Business Intelligence Network**

In addition to the pre-disposition to develop data marts, there is now a pervasive use of tools deployed solely to improve performance, not just accessibility.

The solution must be to make the data warehouse perform

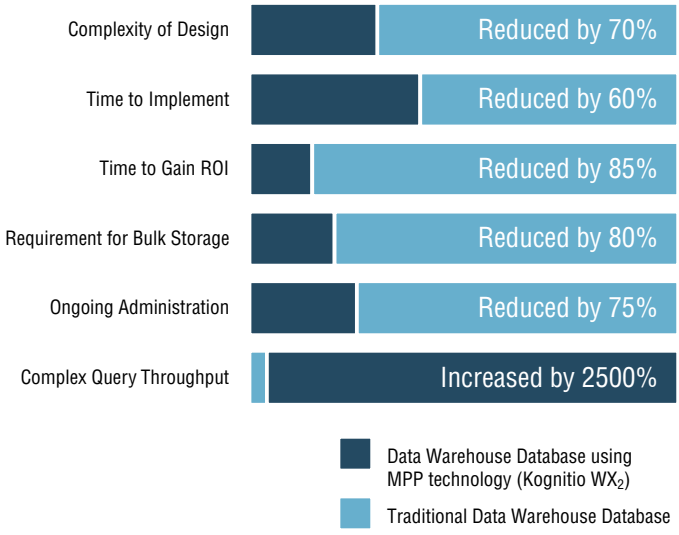
But if the costs and timescales for mart development and maintenance are prohibitive, and fast access is not being provided to granular level data, what is the solution?

Making the data warehouse perform

The solution must be to make the data warehouse perform. But how can this be achieved if data volumes are increasing at twice the speed of processing performance? To make matters worse, most database architectures do not have linear scalability, with query performance being unpredictable as data volumes grow.

In certain circumstances, performance improvements for database queries can be improved by the use of indices. However, these indices must be built and maintained, which creates a significant overhead for the loading of data into the database. With only small time-windows often being available for data loads, this overhead in building indices can often be prohibitive.

Another means of improving query performance is the use of specialist software that captures and optimizes queries run against the database. Although these can produce excellent results, they achieve percentages of improvement not magnitudes.



MPP technology offers considerable advantages in terms of time, speed and ease of use

So, many of the solutions to query performance involve creating more and more layers of software and data (in the form of indices and marts) on top of the data warehouse. This only provides short-term, costly solutions. A solution is needed that addresses the fundamental issues that prevent queries from being run directly against the data warehouse.

If we ignore the limitations imposed by many of the traditional database technologies, the main bottlenecks to query performance are disk access and processing performance.

With Moore's Law suggesting that processing performance is doubling every two years, then significant performance improvements can only be gained by the use of multiple processors.

In much the same way that processors are improving, there have also been considerable cost, performance and capacity improvements in memory (RAM).

Combining the use of multiple CPU's, large volumes of RAM and a compatible database architecture allows complex queries to be run against large volume data warehouses.

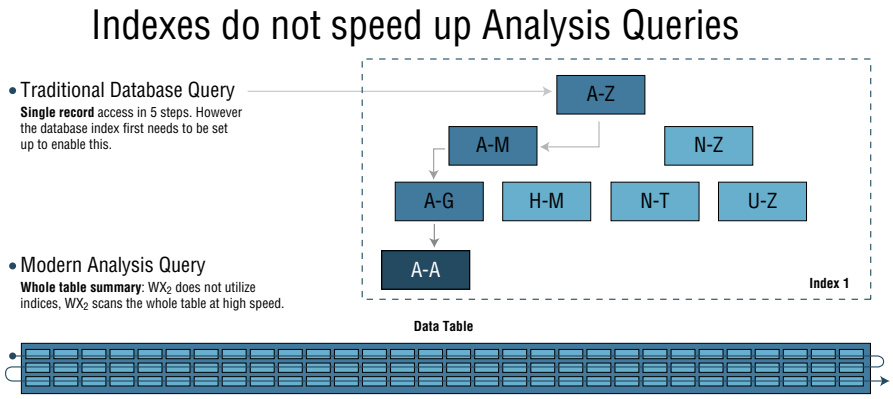
Within the data warehousing arena, this technology has been available for years, but only as costly and highly proprietary hardware and software.

Massively Parallel Processing Technologies

This is all changing, with the evolving Massively Parallel Processing (MPP) technologies being made available on low cost, non-proprietary blade-server hardware from companies such as HP, IBM and Siemens. This provides access to multiple CPU's, large volumes of RAM and the ability to scale up (or down) as volumes of data and queries change.

Companies can now create a generic data warehouse able to satisfy a huge spectrum of demands within an industry sector. The new, non-proprietary, MPP architectures are, on average, 30 times faster than today's warehouses, thereby alleviating any performance losses inherent in the use of a generic model as opposed to a specific, customized, data mart. The overall performance will still be magnitudes faster and, without the overheads of mart design, index maintenance and performance monitoring, the TCO will be much lower.

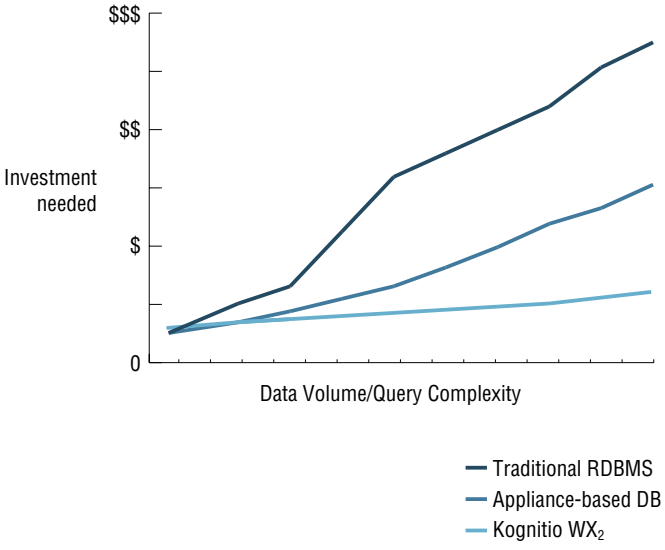
The availability of low-cost blade server architectures and maturing MPP software technologies meet cost-driven business requirements



WX_2 scans 190 million records per second per node (blade)!
Why waste resources on setting up and maintaining

This generic approach takes advantage of three major breakthrough components.

The first is the ability to truly process data in parallel starting from a single node through to hundreds of nodes seamlessly. The second is the ability to load data at rates of up to 600GB per hour. In many cases, this can reduce an eight-hour loading period to less than three hours. The third is the availability of blade server technology that allows nodes to be created from industry-standard processing "building blocks," which can be re-configured and re-deployed at will.



As data volumes increase, MPP's linear scalability allows complex data analysis with a minimum of investment

Conclusion

The driver for this inflection point in today's market is the combination of cost-driven business requirements, the availability of low-cost blade architectures, and maturing MPP software technologies. With true linear scalability, parallel processing means that data warehouse performance is no longer constrained by Moore's Law, and that costs to provide the information desperately sought by today's business will no longer continue to escalate.

So how special are these parallel architectures? Each one is different, but a visit to **www.kognitio.com** will provide a view of one of the leading technologies.

These architectures are available today and need to be considered in the long-term strategy of any company looking to be competitive in the next decade.

kognitio
Competitive advantage from data

About Kognitio

Kognitio is an innovative, technology-rich company, providing leading-edge solutions to business problems that require the acquisition, rationalization and analysis of large or complex data.

Kognitio's offering is centred around three areas: WX_2, the fastest and most scalable analytical database on the market, DaaS and data migration expertise. All three areas are complemented by an extensive professional services team helping businesses to gain a competitive advantage from their data.

With its industry-leading analytical database offering, WX_2, Kognitio is able to rapidly turn a company's raw data into valuable business insight, empowering its customers to realize comprehensive answers to critical business questions.

Kognitio's DaaS model allows its customers to focus on running their businesses and increasing their bottom line. By also adopting Kognitio's outsourced approach, customers are able to reduce start-up time and costs, as well as avoid expensive product acquisition costs.

With a strong specialization in the insurance (life and pensions) as well as financial services, Kognitio's data migration services help companies to provide lower risk and lower cost solutions to companies that are rationalizing and consolidating operational platforms.

Kognitio delivers competitive advantage to clients in various industries, including telecommunications, retail, the financial sector, leisure, hospitality and utilities.

About Kognitio WX_2

Kognitio WX_2 is the most powerful and scalable analytical database in the industry. It enables organizations to query, in detail, vast amounts of granular data in seconds. The software-only solution uses high-speed, Massively Parallel Processing (MPP) technology to deliver an extremely fast data mart/warehouse platform to organizations seeking to gain intelligence from their data.

Kognitio WX_2 runs on low-cost non-proprietary, industry-standard hardware, does not use indices or data partitions and can be scaled to handle hundreds of terabytes of data with performance that delivers answers in real time. This technology delivers the most comprehensive, cost-effective Business Intelligence database platform in the industry and offers delivery mechanisms from pure product through to fully managed services.

More than fifteen years of development have been focused on providing and refining the best tool for corporate Business Intelligence users to freely engage with ever-increasing volumes of data and/or disparate sets of data. Kognitio WX_2 enables business to work its data harder: to take more benefit out; in shorter times scales; with considerably less effort; and without the need for a complex large-scale IT installation. Tests have shown Kognitio WX_2 to run up to 60 times faster than typical databases and at a lower cost of ownership when compared to the lifecycle cost of other solutions.