



---

January 2008

## There Must Be A Better Way: Building Quick, Inexpensive, and Flexible BI Environments

Claudia Imhoff, Ph.D.

---

## Executive Summary

In today's fast moving business environment, it is clear that Business Intelligence (BI) is a mandatory part of any company's decision making environment. The question then becomes how to build this environment in a way that keeps up with ever-changing business requirements, new and large numbers of users, all forms of analytics, and increasing volumes of data.

Forward-thinking CIOs understand the changeable nature of their BI environments but may be stymied about how to build flexible BI architectures without "breaking the bank". Fortunately there are BI vendors specializing in new database technologies that can handle many of the pressures for quick, inexpensive and flexible BI architectures.

These new databases use a completely new approach to storing data. Instead of the traditional row-by-row mechanism of most RDBMSs, these innovative databases store data in a column-by-column fashion. The benefits for analytics from this change are enormous.

However, simply storing data in a columnar fashion is not enough to result in the vast increase in performance achieved by some of these new databases. Compression along with intelligent de-compression techniques add to the performance increase.

But even this is not enough. In determining which column-based database to use, you must also look into the maintenance burden on the IT staff. Look for technologies that no longer require the creation and backbreaking maintenance of indices or partitions.

Finally, look for a database technology that truly does seamlessly scale. Its loading timeframes should not be affected by increasing volumes of data or by increasing numbers of tables.

## Introduction

Business Intelligence (BI) started more than a decade ago as home-grown, nice-to-have, decision support mechanisms, mostly used by statisticians and financial analysts. The ability to analyze month-over-month financials or study past market performances was on the list of needed BI information – just not at the top.

We've come a long way since then. BI is now a mandatory part of the enterprise IT environment. Its usage has spread to every corner of the organization, and to every level of the organization chart. It is now a mission-critical part of operations, supporting not only traditional analytics but also daily operational decision making, supporting rapid fraud detection mechanisms, risk mitigation analytics, behavioral and market predictions, etc.

With these significant requirements come massive changes to the underlying infrastructures sustaining BI environments. These changes include:

- Significantly increasing volumes of data. Now that BI has permeated the business, the amount of data that must be collected to satisfy these needs is enormous. And it is not just more of the same data. Click stream data from websites, RFID tracking data, and other new sources have greatly added to the volumes needed for analytics. Further increasing storage requirements, many business users require that the data be collected several times during the day for operational BI purposes. It is not unusual for an enterprise data warehouse to contain many terabytes or even a petabyte, of historical data for analytical and reporting purposes.
- Fast response times for queries and analyses. With the advent of operational BI or decision support for front line decision makers, the response time for queries must match those commonly found in other operational systems. That is, less than 2 seconds. An already angry customer will not take kindly to waiting while the rep “pulls up” their lifetime value score or next-best product offer. This means that BI environments must differentiate between traditional strategic or even tactical queries from those of an operational nature. This “mixed workload” capability is at the heart of the new

technologies and data warehouse appliances built specifically for BI.

- Seamless scalability. Most data warehouses don't start in the terabyte storage range, but they tend to get there quickly. This means that the hardware and software must scale up, with minimal impact on both the IT staff and the existing technological environment. Going from 1 TB to 100 TB should be as painless and trouble-free as going from 1 GB to 1 TB. Again this is the focus of the new data warehouse appliance vendors.
- Much broader audiences. BI started out supporting business analysts, statisticians, advanced market researchers and others having a need for and education in analytical capabilities. Over the years, more and more employees throughout the enterprise began to see the value of BI but did not necessarily have the skills or education in its usage. BI vendors had to change their interfaces, making utilization much simpler to understand, navigation easier to perform, and comprehension effortless through visualization techniques. From an infrastructure standpoint, it meant the technology had to support massive increases in the numbers of business users, while still supporting the traditional audience of sophisticated analysts.
- More innovative analytics by business users. As the usage increased, so did the complexity of the analytics. The evolution of BI utilization in most organizations proceeds from the creation of simple reports, to time series comparisons, to complex models of fraud detection and risk mitigation, to intricate predictive analyses of customer and market behaviors. The time frame of the data for these analytics also changed. Daily, weekly and monthly snapshots of data were no longer sufficient. Enterprises required BI support for intraday or operational decision making. Operational systems are now being fitted with streaming and embedded analytics to help front line workers make better decisions throughout the day.

Savvy CIOs understand these changes and are examining their BI environments to determine the role that innovative technologies can play in the new world order of analytics. They understand that a new reality has come to their businesses and that reality is called change. They recognize that their technologies must support the new pressures

on their enterprises for rapid change to satisfy the BI stakeholders.

## The Stakeholders

The first two – and probably most important – stakeholder groups in any corporation's BI initiative are the business users and the executives of the enterprise. Unfortunately, many users are, by and large, frustrated with their BI environments. There are several reasons for this dissatisfaction:

1. They can't get the data they need – either because they forgot to ask for it during the requirements gathering stage or because their needs changed after the initial implementation. Unfortunately, as every executive knows, the economics of today's competitive markets are anything but stable. Unless the company can change directions nimbly, it may be doomed. And its ability to move swiftly and in the right direction is very dependent on its executives and analysts having BI systems that can detect, analyze, and report on substantial changes in the enterprise's environment quickly.
2. Their queries run too slowly – especially as they require more and more real time or near-real time BI support for operational decision making. Also, their queries tend to get much more complicated as they use their BI environment more and more.
3. It takes BI implementers too long to get applications up and running – the technology may be cumbersome or inflexible causing slow implementation timeframes. Business people work in a fluid environment that demands flexibility in terms of the reports, analytics, or data they need to make good decisions. They must have a BI environment that gives them the ability to change – on the fly – the set of analytics to suit a particular situation. Agility is the key to a successful BI environment.

The next group of stakeholders consists of the BI implementers themselves. They are frustrated, too, by a number of problems such as:

1. Sudden and unexpected increases in the volumes of data required – this typically occurs when the time frame for data snapshots changes, say from weekly to daily or daily to intraday. Unless the planning for the technological infrastructure included a roadmap to accommodate these massive increases, the implementers can find themselves out of luck, and searching in vain for a solution.
2. Untimely loading of data – implementers using traditional databases often find the load utilities overwhelmed by the massive

increases in needed data with the result that data cannot be loaded in the timeframe allotted.

3. Constantly changing data models and schemas – as mentioned, change in a BI environment is inevitable, yet traditional BI approaches and technologies do not easily support changing data models or schemas. This change is reflected in the workloads of the DBAs who are constantly trying to tune the environment, force it to perform better, by tweaking the data schema. It is also reflected in the constant addition of more data, more tables, more indices, etc., all in the attempt to improve performance.
4. Difficulty handling mixed workload processing – with the increase in operational BI, BI technologies must now handle both the traditional forms of BI queries (those requiring large numbers of records and complicated processing) and the more operationally focused ones (those requiring one or a few records and simple processing) with suitable response time for each type of query.

The final group of stakeholders in a BI environment consists of the CIOs or VPs of IT, responsible for the overall integrity and suitability of the technology supporting BI. They face the following challenges:

1. Business demands to get things done as inexpensively and quickly as possible – IT executives are constantly under pressure to reduce the costs of all IT investments including BI. They also feel the stress of keeping up with the rapid and significant business changes. They understand that the business mandate and therefore the IT holy grail of “build it as fast and as cheaply as you can”.
2. The difficulty of determining business unit charge back for BI – much of the BI environment is considered “overhead” by the business. For example, the business may not see that ETL and data quality processing are necessary for the “goodness” of the BI environment. All they want to pay for is their particular analytical application without regard to how the data got in the application. Someone must pay for these activities that provide a consistent and reliable BI environment though.
3. The balance between building BI fast versus building it to last – this is the age-old balancing act between implementing an environment that supports the enterprise versus satisfying an urgent need of a department at the expense of reusability or consistency in the environment. For example, it is difficult for a CEO to know when or what his company will be required to produce in terms of data and reports supporting new or changing governmental regulations. The

inability to prove that the corporation is in compliance with these regulations can have severe ramifications from large monetary fines to jail sentences for the executives. That is guaranteed to get the CEO's attention quickly. And you can be sure that the CEO will turn to his or her CIO for the critical evidence that the company has performed its responsibilities appropriately. The CIO had better be ready for these requests by having a world class BI environment set up.

So what does the CIO look for to ensure the BI environment can handle the onslaught of changes? First and foremost, there must be flexibility in the BI technical architecture. This means that new users and analytics can be accommodated with minimal impact on the existing performance for existing applications and users. Secondly the environment must be easy to maintain and enhance. The new data and new database entities should be added easily without disruption to the existing schema. And a third consideration is the partnership between the business and IT. Both groups must understand the need for solid infrastructure if they expect to get reliable and consistent BI applications.

These considerations mean that the CIO cannot solve today's business problems with yesterday's thinking and technologies. IT executives must look to the innovative technologies being developed today for their BI solutions.

## How Innovative BI Database Vendors Can Help

One area that has been remarkably innovative is database technology. A fundamentally different approach from traditional RDBMS data storage comes from the column-based databases. Traditional RDBMS databases store data row-by-row which may be fine for transaction processing but is cumbersome or inefficient for the high data volume, fast response time, massively analytical world of BI. So what should you know about these new database technologies?

1. Column-based databases store and access data very differently. As the name implies, these technologies store data vertically in table columns rather than in horizontal rows. Why is this important? By putting similar data together, column-based databases reduce the time to read the disk. This becomes a major factor when

executing large-scale queries such as those typically done in a BI environment – a significant advantage over traditional databases. By storing the data in this fashion, only the data related to a specific query is accessed – all other columns are eliminated from the query.

2. A second feature related to their unique storage mechanism is the performance-boosting feature of very aggressive data compression. Because of the column orientation, a much more efficient form of data compression occurs through the automatic built-in intelligence because the compression can be optimized for a particular data type found within each column. Compression has a number of advantages but first and foremost is the reduction in data volume which, in turn, has a positive impact on disk access time. Because data compression reduces I/O activities due to the close proximity of the data, query performance for BI is again greatly enhanced.
3. But compression alone is not enough though. Obviously for a BI query to be resolved, the data must be de-compressed for the analysis. Performance would be impacted if all data had to be de-compressed just to run a single query. Therefore, another necessary feature is that a column-based database has the built-in intelligence allowing it to discern only the bits of data must be de-compressed and ignore all others. This is a major factor in the overall improved performance these innovative databases deliver.
4. But there is more to this story. Column-based databases are a good start and compression is a major advance but some column-based databases still require the creation of indices to ensure their good performance. As we mentioned above, much of the maintenance heartaches come from the constant tuning of indices. But do these indices solve the problems associated with ad hoc queries or unplanned queries? No, unfortunately they do not. The column-based databases that require physical data models will find it difficult to keep up, just as their traditional database brethren do. The creation of new indices and partitioning of the data soon become daily nightmares for DBAs trying to keep up with ever-changing requirements.

Ideally, a BI environment should have a column-based database that has some form of automated index replacement. By not requiring indices for performance, the technology is able to support fast response times for all forms of queries without being

under the constant control or maintenance of a DBA.

5. Finally, the column-based database you choose should have a profile where load times remain constant regardless of table size. It should also have query times that also remain constant regardless of table size. The bottom line is that the technology must be seamlessly scalable.

Are there other advantages to using a new column-based database that has the above characteristics? The answer is a resounding yes. Here are just a few additional advantages they offer to CIOs and BI implementers:

- A significantly time-consuming activity in any BI project involves the development of the physical data model, that is, the creation of indices, partitioning of the data, creation of materialized views, etc. These are done to improve the query performance of traditional databases and it is also the area most impacted by business user changes and unplanned queries. Because of the column orientation and compression capabilities of these databases there is no need to create physical data models. They do not require constant monitoring to develop appropriate indices – there are no indices required. Database partitioning and materialized views are also not needed to improve performance. These new databases get significant performance from a variety and combination of techniques including compression, column-orientation and alternatives to indices that are design specifically for large data set handling; narrowing down the data required in milliseconds.
- The lack of physical modeling also leads to a great boost to productivity and reduced maintenance. This comes from a reduced need to understand exactly what queries the user is going to ask. In traditional BI environments, DBAs will often hear complaints from their business users about queries not running or running too slowly. The DBA would then try to redesign and re-implement the right physical structures to overcome these query problems. With these new databases, a DBA can focus on enriching the data and enabling the business users to perform queries that were perhaps unthought-of before due to design limitations. With minimal effort and no restructuring of the data, DBAs can support their users as they move into more complex, more sophisticated forms of queries.

- A third advantage comes from the reduced volumes of data required. Many BI implementers are constrained by the very size of the database and therefore tend to be retrospect when it comes to bringing in “nice to have” data for the business users. Unfortunately this eliminates many of the innovative uses of the data as well by the business users. With a compression-oriented database, the implementers can bring in just about anything and everything the business users may want in terms of data – nice-to-have data in addition to must-have data.
- Another aspect that may be less obvious is that the improved query performance means you can store just detailed data, not aggregated or summarized data. Traditional databases must create aggregations and summarizations as a means to improve their performance. However, these processes add even more to data volumes and to overall maintenance since these algorithms and calculations often change. With the compression-based BI databases, aggregations and summarizations are only needed if it is mandatory to have consistent calculations and reliable numbers, that is, it is more of a convenience and an ease of use factor rather than a mandatory process.
- Star schema designs in traditional databases require that business users declare all queries they are likely to run so that the appropriate dimensions and facts may be brought together. Each query run must fit within a single star schema, thus eliminating the ability to ask ad hoc or unplanned queries. For the new compression-based databases, there is no need to create arbitrary star schema designs in order to improve performance. This is a great advantage to the business users, freeing them to ask any and all questions with no constraints. It is also a great maintenance boon since the implementers do not have to constantly redesign star schemas every time a new query is suggested.

## Wrap Up

It should be obvious that the problems that the three major stakeholders had with traditional BI environments can be greatly reduced if not eliminated by bringing in the new and innovative BI databases. The frustrations of the business users – their inability to get at the data they need, the slowness of their query responses and the long turn-around

time to get enhancements made to their environments – can all be mitigated by new technology. No longer do they need to constrain their requirements for data due to unwieldy data volumes. Performance is significantly improved, and enhancements and maintenance to the existing environment can be completed in a timelier manner.

The dissatisfaction of the BI implementers with the BI environment is alleviated through the introduction of newer, better technology such as these new BI databases. Compression technology makes the worry about data volumes, load times, and mixed workloads a thing of the past. The fact that physical data models are not needed not only speeds up development and maintenance but also diminishes the concern with constant changes to the environment.

The last stakeholder, the IT executive, now has a viable option to traditional technologies to help reduce costs of the BI environment and increase the speed with which BI projects can be completed. The pressure on these executives to constantly do more with less is lessened at least for this aspect of IT. And they can be assured that a BI environment using these new BI technologies will last – the architecture will not be broken by too much data, too many users, too burdensome maintenance, or too high a cost. They can rest assured that their BI environment is state-of-the-art and built to last.

### **About Infobright**

Infobright is taking a radical new approach to data warehousing. Infobright's Brighthouse analytic data warehouse software is designed to answer some of the most complex, detailed questions involving massive amounts of information, and answer those questions today, not next week. By fundamentally reinventing data warehouse architecture, Brighthouse "works smarter, not harder." Brighthouse achieves breakthrough performance, while eliminating the need for indices, data partitions, and other physical data structures, as well as the need to deploy large hardware systems. By delivering the most powerful and usable analytic data warehouse, Infobright is enabling companies to bring a whole new level of information intelligence to business operations.

For information, please call 416.596.2483 x225 or visit us online at [www.infobright.com](http://www.infobright.com).