

Ten Reasons Traditional Business Intelligence Approaches Fail to Deliver Value... and Simple Alternatives That Work

Abstract

This paper discusses 10 ways traditional Business Intelligence (BI) approaches fail to deliver value in today's environment, and some simple alternatives that work. The goal of Business Intelligence is to provide right-time information to management and function as a decision support mechanism. Traditional Business Intelligence techniques have fallen short of this goal for one major reason: they are designed exclusively for very large scale deployments involving dozens or even hundreds of systems. Most databases today, however, reside on individual systems used by SMBs (Small and Medium-sized Business) or departments of larger organizations.

Summary

Traditionally, the Reporting and Analysis functions of Business Intelligence were deployed in large enterprises by first combining data from dozens or hundreds of systems.

- The extracted and filtered data was transformed into operational data warehouses by ETL (Extract, Transform and Load) processes.
- Then, techniques such as OLAP cubes were used to enhance the data storage to deliver better analytical performance.

The benefits, however, come at a steep price.

- The performance gains are overshadowed by the enormous complexity and additional delays imposed by an ETL process.
- The complexity of configuring the hardware and software involved in ETL and Data Warehouses requires skills that are often far beyond the reach of most IT organizations.
- Specialized consultants must be utilized to perform ETL tasks and set up the Data Warehouses that few decision-makers will use.
- Additionally, any in-house database experts must undergo extensive training on proprietary products for the information resources to be kept up to date.

Alternatives to these time-consuming and expensive approaches exist, and we discuss some of these better methods in the paper below.

Background

Traditional Business Intelligence approaches evolved out of the need for some of the world's largest organizations to combine information from hundreds or thousands of different sources. Many of these data sources were not in relational type safe platforms. Therefore most of this data would need to be converted into more usable formats. The data came from systems that were generally not capable of even simple summary calculations without custom code. Until a few years ago, the storage requirements of these systems would require a high capacity data warehouse running on large database clusters that used dozens or hundreds of disk drives for storage.

ETL and OLAP: A legacy approach

- **Extract Transformation and Load (ETL)** technologies Extracted data from these native sources and combine them into a single data store called a data warehouse. The data is often restructured and re-organized as it is migrated. These types of tools run on a periodic basis which is usually once a day.
- **Online Analytical Processing (OLAP)** engines were used to then allow for fast analysis across many dimensions. These proprietary platforms restructure data in order to improve performance and add support for dimensional hierarchies.

Database Deployments Today

Today, nearly all business applications rely on a relational database from a limited number of vendors. Rather than being limited to the worlds largest organizations, nearly all departments now have more than one system. These databases cannot only fit into a single hard disk, but many can fit into a single server's memory. Most users are now ordinary business users that need basic reporting and analysis from a single database. Instead of needing to do multi-dimensional analysis across many systems, they require basic summary reports in the form of grids and graphs. Additionally, users expect the results to be real time because they often want to see the update results in progress.

Problems with traditional approaches

While ETL and OLAP cubes still make sense for enterprise-wide analysis of a large number of systems, most user requests are simple reports out of an individual system. The explosion in the number of databases globally makes it nearly impossible to integrate them all. Furthermore, ETL and OLAP technologies are not real time and are very difficult to keep up to date without database experts, who often have many higher priority objectives. As a result, many data warehouses take so long to launch, they are obsolete by the time they are complete. More importantly, updating OLAP cubes is so involved that the reports rarely keep up with the pace of business.

1. Data Warehouses are Complex to Setup and Very Difficult to keep Updated

Creating a traditional data warehouse involves leveraging ETL technologies to migrate data into a special data store which is then analyzed by Online Analytical Processing (OLAP) tools designed for fast multi-dimensional data analysis.

Historically, the primary reason for a data warehouse is that the infrastructure used for transactional processing could not support additional load or execute analytical queries fast enough. These tools required significantly more expertise than most database administrators, SQL experts or software engineers have.

As a result, organizations rely on specialized consultants or the capacity of their most senior technical experts to develop these assets. Once a data warehouse is built, the real challenge becomes keeping the data warehouse up to date as the nature of the business changes because technical resources that are fit to the task have very limited availability.

Alternatives

Rapid advances in storage technology, the sharp decrease in the price of memory and storage has made it possible to do reporting and analysis on standard database servers rather than specialized OLAP technologies. Such databases conform to open standards like SQL which are already familiar to most database administrators and software engineers. The primary reason behind this is that system memory, which tends to be multiple orders of magnitude faster than disk based storage is now sufficient to hold the working set of most databases.

- **Use database views instead of ETL and OLAP Cubes.** Since most of today's databases are relational, the views can be quickly added that facilitates reporting and analysis. This approach is always real-time and can provide similar performance results on modern hardware. A key advantage of using Views is that most database experts can create them without learning new skills.
- **Upgrade to hardware that can fit the working set in memory.** The working set, or the part of the database used for reporting and analysis can now fit entirely in a server's memory. A server with 4 gigabytes of memory can now be purchased for just over \$1000. Since system's memory is over 1000 times faster than the fastest hard drives, this speed increase allows for capabilities never before possible. Another possible option is to use a Solid State Disk(SSD) or flash memory device to store the database.
- **Introduce flexible end-user reporting and analysis capabilities.** Sophisticated multi-dimensional analysis requires database experts to setup. However, business users looking to create basic charts and grids can do so themselves with Excel over ODBC, export to CSV or with next generation web-based reporting technologies like Izenda Reports.

2. Expertise in Traditional OLAP-based Technologies is Costly and Very Difficult To Find

Designing data warehouses, MDX expressions and OLAP cubes requires levels of knowledge and experience that most database developers simply do not have. While most developers can learn how to use the actual design tools, understanding how to apply best practices to effectively construct and maintain effective decision support systems requires a much higher level understanding of business process optimization.

Effectively designing cubes requires

- Expertise of a database developer as well as a business analyst. Very few technical experts understand both aspects of data analysis. As a result, implementations get delayed and live deployments fail to adapt to new business conditions.

The solution

- Is to use alternative concepts like views which are familiar to most database experts and to allow the business user to instantly create and customize business reports through an intuitive browser-based interface.

Alternatives

- **Utilize relational database capabilities.** While standard SQL expressions cannot do true

multi dimensional analysis very well, most business users will only understand two or three dimensions at a time. Existing operational databases can be enhanced with special views to facilitate this level of reporting and analysis. As long as the working set of data can fit in the system memory, most aggregate queries will take about a second to run on a modern server.

- **Empower business users to create and customize reports.** Business users today are increasingly comfortable with doing basic analysis on a small number of variables. The analysis, however, needs to adapt to business conditions so that reporting and analysis is more dynamic than just providing fixed, canned reports. Additionally, most users need to be able to access these capabilities from a secure browser-based interface without having to request special access from IT. While most business users are initially reluctant to build reports from scratch, they are very comfortable with being able to instantly customize stock reports and save their own personal variations.

3. C-Level Executives Need to Drill-Down Into Low-Level Details and Operational Managers Also Benefit from Snapshots

Conventional wisdom suggests that top level executives need scorecards and dashboards where lower level managers need more detailed reports. In reality, executives need to drill into details and operational managers can benefit from having a snapshot. Traditional platforms make creating such snapshots reports so costly that the expense could only be justified for the highest tiers of the organization.

While vendors emphasize the flash and sizzle that their dashboards offer, these tools are only actionable if the organization has a clear understanding of their key performance indicators and the ability to measure them in real time. Not all levels of an organization will initially have a clear understanding of what the key performance indicators actually are. The ideal indicators can not always be measured or retrieved easily while critical aspects of the business may not be measured at all. Sometimes they are sensitive in nature and not something that can be distributed to all employees. Organizations can also have hundreds or thousands of measurements without a clear understanding of how to limit or combine these into a small number of key factors.

Alternatives

- **Build initial balanced scorecards using low cost methods that are more flexible and collaborative before investing in flash and sizzle.** Online spreadsheets like Google Docs or Zoho Office in addition to content management systems like Sharepoint may be used to create a "shared" scorecard in the form of a spreadsheet. This approach can be utilized at a low cost while the real KPIs get developed and refined. Organizations will achieve better results by investing in flash and sizzle *after* this process is complete. Organizations with smaller budgets can even continue to use these low-cost options.
- **Provide everyone in the organization with flexible end-user reporting and analysis capabilities that do not require IT involvement.** When the dashboards or scorecards indicate that something is wrong, the next step is to investigate. Executives as well as lower level managers need to interactively explore operational data to gain insight into why an element of the business is not performing. Traditionally, reliance on database experts has made this process frustrating for both business users and IT. Users often need a new report right after getting the results of the last one. The solution is to provide reporting and analysis capabilities that can be instantly customized so users that need to calculate or group data in a different way can do so instantly and independently.

4. OLAP Engines Require Adding and Maintaining Additional Infrastructure.

Traditionally, relational databases have been optimized for high volume transactional operations rather than reporting and analysis. In order to analyze large quantities of data it was necessary to use ETL to copy data on a periodic basis to a data warehouse. A cubes engine designed for aggregate analysis rather than transactional performance could be used to quickly summarize data. Such deployments require adding additional servers and purchasing additional software. More importantly, those tasked with maintaining and updating the system would need to learn additional skills as the ETL and cube modules generally have very different administrative tools and conventions than a relational database.

Alternatives

- **Use views on the existing transactional database to facilitate reporting and analysis.** While this would not have been possible for multi-gigabyte databases a few short years ago, advances in hardware and cost reductions in memory have made it possible for a transactional database engine to perform sub-second queries on very large data sets. A query that may have taken 20 minutes a decade ago on a gigabyte of data can now run in less than a second on modern hardware. While this approach will not work on data sets that are significantly larger than system memory, most operational databases today have a working set that is smaller than system memory.
- **Browser-Based Applications do not Integrate Well With Modules that Require a Download.** The majority of new applications developed over the last several years have been web applications. While developing web applications often requires additional effort and many limitations, organizations have chosen this path because browser-based applications are significantly easier to deploy. Today's security landscape involves many security and privacy concerns in addition to the threat of mal-ware and viruses causing problems to users system.

A web application running in a browser has the following advantages

- Running the application will not risk compromising the user's machine.
- No additional firewall or administrative privileges need to be granted to the user's machine to install and run the application.
- Users can access the application instantly, without the additional task or running an install process or getting permissions.

Disadvantages of Database Software that runs on End-User Machines

- Installation requires administrator privileges be given to the user or IT staff with sufficient permissions must do the install. Most users do not know the proper way to install new software for their environment. Software programs often fail to run due to conflicts with other programs on the same system.
- Software that accesses a database contains connection strings that can be used to damage or extract the database.
- When the system needs to be upgraded or reconfigured, every single end-user machine must be manually updated as well.

- Since most applications are transitioning to web access, organizations have also setup their security policies to work well with web applications.
- Most reporting and analysis tools adhere to protocols like ODBC from the client-server era. Such protocols do not work with modern security infrastructure. Firewalls on both the user's machine and the network they are on must be specially configured to allow these protocols to operate properly. These types of protocols can also be risky because users or hackers that gain access to their computers can potentially extract from or damage the database.
- Even if permissions can be set via a VPN, a desktop tool requires an involved and sometimes risky install process that the user may require IT to complete.

Alternatives

- **Select reporting technologies that work exclusively over a browser.** While nearly all reporting and analysis products now support the web, they are generally limited to allowing reports created and published only by an experienced database expert. While web publishing is useful, it renders the database expert a bottleneck and prevents users for interactively exploring data because getting a report updated may take days or weeks.

5. Traditional Reporting and Analysis Tools Do Not Work with Hosted Applications, Service Oriented Architecture or Multi-Tenant Environments.

Browser-based applications are hosted on a web server. Applications can support four different ways to host applications.

- **On-Site.** Customers with security requirements that mandate the data should not leave their corporate network may host an application on a web server from their network. The actual application may be available only to users on that network, or to anyone on the web.
- **Off-Site.** The web and database servers run on a network outside the organization's networks. Each customer still has dedicated servers, but they are physically on a different network.
- **Multi-Tenant** Multi tenant systems take advantage of economies of scale by placing data for many customers on the same web and database servers. Sophisticated security measures must be added to ensure that no customer accidentally sees data from another. This approach greatly simplifies setup and maintenance because it avoids the need to configure an entire software stack for each customer.
- **Multi-Tenant with Database Isolation.** This hybrid approach involves using a shared server environment but isolating each customer's data into an individual isolated database.

In all of these scenarios, the database server is securely configured so that only the web server can access it. Making a database server public significantly increases the risk of hacker attacks or data theft. Consequently, traditional reporting and analysis tools rely on direct connections. Additionally, multi-tenant environments need reporting capabilities that prevent users from accessing records that do not belong to them. The traditional all-or-nothing model desktop reporting tools are fundamentally incompatible with multi-tenant security.

Alternatives

- **Give each customer their own database server and VPN Access.** Giving each customer or tenant his own server would allow each to securely use existing reporting tools across a VPN connection to the database. There may be issues to doing this from a security perspective as there is potential risk of data loss, theft or privacy and compliance violations.
- **Provide each customer with a web-based reporting environment that supports true multi-tenant security.** Solutions like Izenda Reports allow for true multi-tenant security down to the record level. Even users building reports from scratch will not have access to secure fields or records for tenants they do not have access to.

6. Extract Transform and Load (ETL) Can Introduce Data Lag Which Can Slow A Business Process by Several Days

While ETL approaches tend to work well when systems are small and the data model is simple, delays can reach several days as these systems evolve. The inherent problem is that an ETL process essentially migrates the entire database every time it runs. As the database grows and experiences more load, the process can experience an exponential increase in run time. While it is possible to only move updated records, doing so can introduce duplicate records and other data consistency issues.

The problem is compounded when multiple ETL processes that run on different schedules are used by multiple interdependent systems. If a process does not complete before the start of another, it can add 24 hours to the data lag. When data is streaming through multiple systems using ETL, a simple update that should be instantaneous can take a week. Anyone who has dealt with a customer service representative who does not have updated records in "their system" has experienced the result.

Alternatives

- **Use near real-time data publication and subscription.** Modern databases include sophisticated publication capabilities which allow data updates to be transferred in near real-time without data integrity problems.
- **Use Notification Services.** Notification services can inform a system of individual record updates. The notified system can then conditionally retrieve updated data depending on the nature of the notification.
- **Create views that connect directly to other systems.** Modern databases include features like OPENQUERY and cross-database SELECT that allow instantaneous retrieval. Views that utilize these capabilities properly can pull required data just in time and avoid having to transfer every record. Additionally, these approaches can be combined with summarization for situations where only a summary of the original data is required.

7. Organizations Fail To Anticipate the Volume of Reports Required

Implementers that ask the user community what reports will be required generally get little feedback. This is because new reports are required as new opportunities and challenges arise, and these usually cannot be anticipated. Neither the user population, the executives nor the implementation team has a

crystal ball that can help them understand what reports will be needed.

The reporting requests reach a high volume when the new system is live and achieves high utilization. The dozen or so reports originally requested quickly turn into hundreds or thousands of requests over the first few years of a new system's life. By then, consultants involved in the project have often moved on to other projects. Even if an IT organization has the technical expertise to maintain and update reporting and analytics, they are often overwhelmed with many other responsibilities. As a consequence, IT gets even more backlogged and users can not get customized reports in a timely manner.

Alternatives

- **Utilize platforms that the company has sufficient expertise in for reporting and analysis.** While technologies capable of true multi-dimensional analysis have many benefits, very few IT organizations have the in-house expertise that can update and maintain these types of systems. As an alternative, standard relational databases may be used to drive the majority of reporting and analysis. Expertise in SQL and relational databases is much easier to find.
- **Utilize flexible end-user reporting capabilities which allow report creation without IT involvement.** While using relational approaches for the majority of reports will make it significantly easier for IT organizations to keep reporting systems up to date, the real solution is to empower end-users to customize their own reports without IT involvement. Most users are already comfortable doing basic analytics using spreadsheets. While spreadsheets are flexible, there are many security and collaborative issues that argue against their use – for example, emailing spreadsheets that may become out of date, incomplete, etc. The ideal solution is to use a flexible web-based solution like Izenda Reports that securely allows any user to customize or create reports with little training, on a real time basis, using up to date information directly from the database.

8. Too Many Dimensions Can Quickly Overwhelm Business Users

Financial analysts and people from engineering or scientific backgrounds are comfortable slicing and dicing data across several dimensions simultaneously. However, most users are challenged when dealing with data that contains more than two or three dimensional hierarchies and a single measure.

An example of a dimensional hierarchy would be

- Quarters and years or cities and states. Years and quarters are essentially the same dimension, but are broken down into hierarchies.
- A report that shows time across the top, geography on the left side and a measure like revenue inside the cells is generally the level at which most business users are comfortable.
- An additional measure like unit volume may be added inside the cells. A third dimension like product category may be added by using multiple table or charts, but essentially three dimensions and two measures is the limit of what can be easily visualized by most people.

Users are most comfortable with 2D visualizations that involve a single measure summarized across two dimensions or hierarchies. Adding more makes the data difficult to understand and makes it difficult to print.

Alternatives

- **Present All Summaries a Two Dimensional Format.** A two dimensional summary can be displayed in a grid with one dimension on the top and one on the side. It can also be presented in charts easily.
- **Provide Drill-Down capabilities to users.** Rather than trying to force too much data into a single visualization, it is possible to make the visualizations interactive. Additional dimensions can be provided to users when they click on specific elements that they want more information on. These drill-down visualizations should also be limited to two dimensions.

9. Users Expect Real-Time Reporting and Analysis on Real-Time Applications

Batch processing was once the norm. Today employees and the customers they serve expect and demand real-time access to information. A salesperson that just added a new opportunity to their system expects the reports to reflect this immediately. Technologies that rely on moving data with ETL tools or pre-processing cannot deliver true real-time reporting and analysis.

Alternatives

- **Invest in systems and infrastructure that deliver near real time performance.** Most ETL techniques perform periodic batch processing to migrate data, some techniques are available that update the data store on a per transaction basis. These types of systems may be difficult to setup and still have ETL overhead, but spread it out over many transactions.
- **Build reporting and analysis capabilities into your transactional database with specialized views and indexes.** With modern hardware, most reports can be executed by a transactional database in a manner of seconds. Because the working set of most databases now fits in system memory, cubes engines are not required for fast performance when dealing with systems containing several gigabytes of data or less. To achieve good performance, fields which are indexed, sorted or joined must be indexed. Fortunately, tools like Index Tuning Wizard automate the bulk of the indexing process.

10. Advanced Analysis and Flashy Visualizations Are Not Actionable Without Proper Business Modeling and User Training

Sizzle and flash are important elements of creating visualizations that drive high performance organizations. Unfortunately, many organizations invest in dashboards and scorecards without first developing the Key Performance Indicators (KPIs) that drive them. Even if an organization knows what is important, it is not always measurable in an automated fashion. KPIs should be actionable, easy to understand and most importantly something that can be measured automatically. When these models are not developed first, the most sophisticated analysis or presentation will go unused.

A great example of an actionable metric with a simple visualization is a weather forecast. Powerful supercomputers are used to perform very sophisticated analysis of enormous datasets. The results are then visualized using single values for temperature, for wind and humidity. Visual icons like illustrations of the sun, clouds, wind and rain are then used to make the information easier to consume. This works very well because we all have a lifetime of experience and education on what these things mean. The experience of going outside and correlating that with the temperature metric results in an

intuitive correlation between the two. Additionally, experience with the accuracy of the forecast in the past teaches us what degree of trust we should place on the predictions.

Another example is the speedometer on a car. The values have significant meaning in terms of what actions are appropriate and which ones are possible. Our understanding comes from a many years of experiencing the speedometer as a driver or passenger as well as during driver training.

When users first encounter a balanced scorecard or forecast, the same level of exposure and experience is not generally there. This is especially true for any metric that results from a sophisticated analysis like a regression or data mining. The metrics must be explained and experienced before they gain actionable value.

Alternatives

- **Provide end-users with flexible reporting capabilities so they can drill-into the source behind the metrics.** Some users may not initially trust the results coming from a new analytical system. Indeed, new systems are sometimes plagued with calculations that need to be tweaked. Empowering the users to dive into the analysis with tools that let them customize their reports will build trust and credibility and allow the users to contribute to refining the analysis.
- **Avoid forecasting methods that are too hard to understand.** Forecasts involve sophisticated techniques such as doing a regression and removing seasonality from the data. While these methods can be much more accurate than traditional ways to forecast, they are so complex they may be ignored rather than refined and utilized. Simple models based on quarterly and annual growth rates are easy to understand and can be reproduced in a spreadsheet by most users.
- **Securely provide visibility into all data rather than trying to force it into an oversimplified visualization.** While the idea that running a business can be as easy as driving a car has been effectively sold, the reality is that things are never that simple. Operating a business is more akin to a mechanic trying to keep an old car in operation. A mechanic needs lots of tools and the ability to look into every aspect of the engine, just as users need to delve into much of the data, not just a set of gauges that summarize it.

###