



Six Key Considerations When Selecting BI for Big Data

By Wayne W. Eckerson

January 2017



atscale

Research Sponsored by AtScale

About the Author



Wayne Eckerson has been a thought leader in the business intelligence and analytics field since the early 1990s. He is a sought-after consultant, noted speaker, and expert educator who thinks critically, writes clearly, and presents persuasively about complex topics. Eckerson has conducted many groundbreaking research studies, chaired numerous conferences, and written two widely read books on performance dashboards and analytics.

Eckerson is founder and principal consultant of Eckerson Group, a research and consulting firm that helps business and analytics leaders use data and technology to drive better insights and actions. His firm helps companies develop strategies and road maps that maximize their investment in data and analytics.

About Eckerson Group

Eckerson Group is a research and consulting firm that serves business analytics leaders and helps them use data and technology to drive better insights and actions. Our research analysts each have more than 20 years of experience in the field and are uniquely qualified to help business and technical leaders optimize their investments in business intelligence, advanced analytics, and data management. Our consultants help clients develop forward-thinking strategies and roadmaps using the latest technologies and techniques.



Analyzing Big Data: What Really Matters?

If you lead an analytics or business intelligence (BI) program, the impact of big data is palpable. Your business users want to analyze large volumes of multi-structured and fast-moving data. Power users want to build their own reports without waiting days, weeks, or months for the information technology (IT) department to do it for them. They want faster, better, and cheaper tools to access data and generate insights. You need to act fast.

You know all too well that your current BI environment is inadequate to meet the need. But you also recognize that new data analytics platforms and technologies are in various stages of maturity, and any changes you make will be disruptive and costly. Therefore, your goal is to find new technologies and approaches that maximize the upside and minimize risk.

This paper is designed to help you navigate the shoals of selecting a BI tool for a big data environment. There are many things to consider. For instance, how do you:

- Query terabytes or petabytes of data with near instantaneous response times?
- Load, store, and manage both summary and detailed views of data?
- Empower users to explore data without creating data silos?
- Create a standard view of data that adapts quickly to changing requirements?
- Simplify a complex data architecture with multiple data processing engines?
- Leverage the talents of the IT department without creating bottlenecks?

The answers to these questions are not straightforward. They involve a mix of people, process, and technology. This guide provides six action items to guide your efforts in selecting an appropriate BI tool for a big data environment. It's designed to help you ignore the hype surrounding big data and focus on what's really important.



1 Don't Forget Your Primary Users

Most business users have simple information requirements. This fact is almost too obvious, but many BI and analytics leaders neglect it in their rush to embrace big data and new analytics tools. The big data movement has been largely focused on meeting the needs of a small and exclusive group of users: data scientists and data analysts who are hired to access and analyze data and need true self service. They comprise perhaps 5% to 10% of your employees, but they're the ones technology vendors have pursued intensively for the past five years. Don't fall into the same trap!

The reality is that most business users don't want or need extraction, exploration and preparation. They simply want to view and analyze data tailored to their role and the areas they manage. Fundamentally, they want a dimensional dashboard or report that graphically displays top-level performance indicators and lets them point and click to explore anomalies and trends at the speed of thought.

Such tailored, interactive dashboards that provide three clicks to any data with near instantaneous response times are the bread and butter of BI. When designed properly, they meet a majority of information needs for 90% of your business users (i.e., data consumers and explorers) and replace dozens if not hundreds of reports. The dashboards also significantly improve planning, decision making, and worker productivity. Every workgroup, department, and executive team needs them. Don't forget this!

KEY TO SUCCESS: Build a tailored, dimensional dashboard for each department that provides three clicks to any data and answers a majority of business questions for 90% of your business users.

2 Standardize What Matters

As a consultant, I've helped many organizations select BI and analytics tools. And I've had endless discussions with business analytics leaders about tools standards and portfolios. These are important discussions, but they're almost irrelevant. Today, business unit heads buy the BI tools they want, and corporate IT departments can do little to stop them. In my experience, standardizing tools is the wrong argument to have with business users.

Rather than focus on tools, BI leaders should devote their time and energy to standardizing what really matters: data definitions and data models. BI leaders should work with business unit heads to standardize metrics, dimensions, and other data objects used across functional boundaries. They should make sure the business builds these definitions into local BI and analytics tools and applications and uses them when reporting results to corporate entities.

Single Business View. Without proper data standards, organizations are Towers of Babel where everyone talks but no one communicates. Business analytics tools compound this problem because each implements its own semantic layer—a business-oriented model of data. Each model is built by different people at different times using different data. It's no surprise, then, that executives can't get straight answers to simple questions.

To solve this, corporate IT departments should create a universal semantic layer that can be accessed by every individual using any BI tool or application via any data access method (e.g., SQL, MDX, DAX). That way, everyone gets the same answer to the same question and no one distrusts data or reports created by others. Although a universal semantic layer won't satisfy every data inquiry since models trail new data requirements, it will meet the needs of most business users most of the time.

Exploration Environment. For the remaining inquiries, organizations need to deploy an exploration environment geared to data analysts and data scientists that leverages the same universal semantic layer. These analytic sandboxes enable these power users to mash together data and analyze it in a free-form and unfettered manner. But the organization must also apply governance processes to these environments to prevent them from becoming data silos.

KEY TO SUCCESS: Create a universal semantic layer accessible by any business user, BI tool, or application via a variety of data access methods and build analytical sandboxes for unfettered exploration and analysis.

3 Don't Move or Duplicate Data

Today, the classic information supply chain for analytics has quite few hops in it—from source to staging to hub to data mart. And that doesn't count the need to move or replicate data for desktop discovery tools. The result is a complex, costly, and error-prone data architecture and a proliferation of data extracts that undermine data consistency.

In the age of big data with terabyte and petabyte repositories, it doesn't make sense to move data. There is simply too much of it, and batch windows are shrinking. Consequently, we're quickly moving to a world in which data architects no longer have to move data to process it. The best architecture for minimizing risk and cost is to process data in place.

Many organizations are doing this by building data lakes on Hadoop and other big data platforms. This approach brings the user to the data instead of moving the data to the user (or BI tool or desktop.) With Hadoop, companies load detailed data at low cost and process it in place using built-in engines such as SQL, OLAP, transactions, and graphing. Other specialized big data platforms, such as Teradata, Amazon

Redshift, and and Google BigQuery obviate the need to create data marts or data extracts that feed in-memory engines of BI tools. The trick then is to provide suitable query performance against these consolidated repositories of data. (See #4 below.)

If data is scattered across multiple databases and can't be combined into a single repository, organizations may need to adopt a federated approach. Here, a data architect might link together data sets from different sources into a virtual cube that users query. Or the BI tool might query each database separately, join the data on the fly using push-down optimizations and in-tool conversions, and display results to the business user. Either way, the data stays put. Unfortunately, a federated approach taxes query performance more than any other type of data architecture.

KEY TO SUCCESS: To simplify data architectures and eliminate data extracts, consolidate data in a big data repository and query it there. Bring business users to the data, not data to business users.

4 Don't Sacrifice Scalability or Performance

In my experience, BI tools are designed for either scalability or performance. A scalable tool queries all the data, but struggles to provide sub-second response times. Conversely, a performance-oriented query tool provides consistently fast queries, but only against a subset of data within an in-memory cache or database managed by the tool.

Most BI tools have devised workarounds in an attempt to optimize both scalability and performance. Some workarounds are cumbersome, while others create additional problems. For example, BI tools with in-memory engines require the IT department to create numerous data extracts that users download into server or desktop tools. These extracts create a proliferation of data sets that are hard to manage and give users a fragmented view of enterprise data.

A better approach is first to optimize scalability by using big data platforms to process queries and then optimize performance using dynamically generated views of commonly requested data. If designed properly, the business views ensure users experience consistently fast queries while still giving them access to detailed data. A scalability-first approach also often uses native drivers to squeeze every ounce of performance from the underlying data platform.

KEY TO SUCCESS: In the era of big data, don't create extracts to drive performance. Rather, query the entire data and use dynamically generated views to deliver fast performance.

5 Select Open Tools That Work with Many Platforms

The bane of BI has been the closed nature of the tools. Most BI tools are islands unto themselves. One BI tool can't query another BI tool's semantic layer, reports, or embedded databases. No wonder BI tool proliferation is such a problem!

To avoid creating BI islands, select tools that are designed for openness and integration. BI tools should not compete with other tools for turf inside an organization. Rather, they should work together. If a business unit has modeled its data environment, other tools should be able to query that model without having to recreate it in another tool. It's rare for BI tools to be so open, but there are new solutions that have made this functionality a priority. The best BI solutions complement what already exists and don't require organizations to rip and replace products.

In addition, the BI tools should be able to query any big data platform, ideally with native connectors. This includes relational, Hadoop, and NoSQL platforms both on premises and in the cloud. This degree of open connectivity is critical to avoid getting locked into a single data platform, vendor, and style of processing.

KEY TO SUCCESS: Select an open BI solution that complements other BI tools and runs against various big data platforms.

6 Forge a Strong Business-IT Partnership

In the age of self-service, business managers have been quick to write off the IT department. But the pendulum is swinging back. Business units often find it difficult to scale, secure, and staff their own data and reporting environments while keeping up with user demand for data and reports. Some business managers see the deleterious impact of data silos and understand the need to work with their IT teams. They discover the hard way that the path to self-service requires standards and architecture supplied and managed by IT.

The IT department excels at creating standards and architectures that serve the long-term needs of the enterprise. But to succeed in the new world of empowered business users, IT must function as a facilitator, not a dictator.

Specifically, the IT department needs to work collaboratively with the business to establish data standards and a universal semantic layer. (See success factor #2 above.) It also needs to design and manage the big data environment that houses both detail data and aggregate views. (See #5 above.) It can then help select and implement a BI solution that optimizes both scalability and performance without moving data. (See #3

and #4 above.) Finally, IT can create, if required, rich, interactive, dimensional dashboards valued by a majority of business users. (See #1 above.)

The IT department is an integral player in the delivery of big data analytics. It plays a key role in each of the success factors outlined in this report. However, in the current era of self-service, the IT department needs to understand and accept its role as a facilitator. At the same time, the business has to acknowledge that IT plays a key role and work with it. Bridging the gulf between business and IT is a perpetual challenge, and one that organizations must overcome to master big data analytics.

KEY TO SUCCESS: Both IT and the business must work together to optimize a big data analytics environment. IT must recognize that its role is to facilitate, not dictate, while the business must recognize the importance of IT in setting up and governing a self-service environment.



Finding Success

Succeeding with big data is no different than any other technology-driven project. It takes knowledge about how to harness people, processes, and technology to achieve a business goal. It also requires a good team, a healthy budget, and strong sponsorship. For big data analytics projects, follow this checklist of the six success factors to consider for BI technologies on big data, outlined in this guide:

Success Factor	Key to Success
Don't forget your primary users.	Build a tailored, dimensional dashboard for each department that provides three clicks to any data and answers a majority of business questions for 90% of your business users.
Standardize what matters.	Create a universal semantic layer accessible by any business user, BI tool, or application via a variety of data access methods and build analytical sandboxes for unfettered exploration and analysis.
Don't move or duplicate data.	To simplify data architectures and eliminate data extracts, consolidate data in a big data repository and query it there. Bring business users to the data, not data to business users.
Don't sacrifice scalability or performance.	In the era of big data, don't create extracts to drive performance. Rather, query the entire data and use dynamically generated views to deliver fast performance.
Support open tools that work with many platforms.	Select an open BI solution that complements other BI tools and runs against various big data platforms.
Forge a strong business-IT partnership.	Both IT and the business must work together to optimize a big data analytics environment. IT must recognize that its role is to facilitate, not dictate, while the business must recognize the importance of IT in setting up and governing a self-service environment.

If you follow these simple action items, you can make great strides in helping your organization modernize its analytics environment to meet business demands for big data, greater self-service and autonomy, and fast query performance. We hope this paper will help guide your thinking and shape your decisions during the journey toward big data nirvana.



Need help with your business analytics or data management and governance strategy?
 Want to learn about the latest business analytics and big data tools and trends?
 Check out **Eckerson Group** research and consulting services

**Company Bio:**

AtScale makes BI work on Big Data. With AtScale, business users get interactive and multi-dimensional analysis capabilities, directly on Big Data, at maximum speed, using the tools they already know, own and love – from Microsoft Excel to Tableau Software to QlikView. Built by Big Data veterans from Yahoo!, Google and Oracle, AtScale is already enabling the BI on Big Data revolution at major corporations across healthcare, telecommunications, retail and online industries. To see how AtScale can help you, go to www.atscale.com/try.