FIVE TECHNOLOGIES THAT CAN IMPROVE YOUR REPORTING AND ANALYTICS ARCHITECTURE—WITHOUT BREAKING THE BANK

BUSINESS CONSULTANTS

DEEP TECHNOLOGISTS

westMONROE

A WEST MONROE WHITE PAPER

FIVE TECHNOLOGIES THAT CAN IMPROVE YOUR REPORTING AND ANALYTICS ARCHITECTURE—WITHOUT BREAKING THE BANK

At one time, business intelligence and data warehousing platforms provided the primary avenue for using an organization's data to drive performance and advantage. The high cost of building these platforms, however, prevented many smaller and mid-sized organizations from deriving all the benefits available to their larger competitors.

Today, various technologies are "disrupting" the traditional reporting landscape and analytics architecture. Not only that; they are accessible to organizations of all sizes—enabling them to reap benefits ranging from lower costs to faster and more effective analysis.

This article examines five such technologies that offer significant business benefits on their own—but may be even more powerful when applied together.

A WEST MONROE WHITE PAPER

FIVE TECHNOLOGIES THAT CAN IMPROVE YOUR REPORTING AND ANALYTICS ARCHITECTURE—WITHOUT BREAKING THE BANK

TODAY'S CUTTING-EDGE DATA-MANAGEMENT TECHNOLOGIES AREN'T JUST FOR THOSE WITH BIG BUDGETS

Becoming "data driven" is a strategic imperative for organizations of all sizes. Not surprisingly, companies are spending more than ever before on data platforms.

In years past, cutting-edge tools that enabled organizational insight through advanced monitoring, reporting, and analytics capabilities were largely the domain of larger organizations that could afford them. Today, a variety of technologies—some of which have existed for several years or more—are changing the reporting landscape and analytics architecture. The good news is that this "shake up" is rewarding companies of all sizes with benefits that range from lower costs to expedited development to faster analysis.

Following are five specific technologies that we see shaping the future of analytics and reporting.

HADOOP DATA LAKES

Technical overview: A data lake is a concept for gathering and storing raw data in a massive repository until needed, at which point that data is read into a schema that is useful for analysis. Hadoop has been around for decades and in the media more recently as the "must-have" big data technology; however, there is still much confusion about what it does. In the simplest terms, Hadoop is a way of storing data on a file system. While every database stores its data on a file system, Hadoop does so with many small files instead of a few larger files. Unlike traditional relational database management systems (RDBMSs), Hadoop also supports semi-structured and unstructured data (e.g., text blobs, images, videos, etc.) natively. The biggest difference in Hadoop, though, is a design that distributes data across many servers in a cluster—helping Hadoop write large volumes of data quickly and to be fault tolerant.

Business impact: Hadoop allows you to ***store vast amounts of data inexpensively*** until you identify a need for that data or until certain events occur that then make that data actionable. Hadoop also ***lowers the barrier to capturing and storing new data sources***. For example, if you want to mine Twitter data related to your company but aren't sure yet how you will use it, having a technology like a Hadoop data lake in place makes it easier to justify capturing and storing that new data until you are ready to use it.

JSON DATA TYPES

Technical overview: After living in the realm of JavaScript, the JavaScript Object Notation (JSON) has been picked up by a number of languages and, perhaps more significantly, databases.

In the past, XML was great for creating semi-structured data sets, but working with it wasn't exactly easy. What is different this time around is that database platforms like MySQL and Postgres have enabled native JSON data types. This means three things: 1) the database can verify a valid JSON body upon insert, thus taking that validation out of the application or data transformation layer; 2) the JSON body can be stored in a binary format instead of simply a text object that allows for optimized storage; and 3) native JSON function allows you to extract parts of the JSON body without having to flatten the entire object. This simplifies usage of this data type inside of queries because developers don't have to write an application layer. It also simplifies writing of data transformation logic.

Business impact: It is easy to keep data objects from applications (e.g., a webservers) in a relatively raw format, pick out just the parts that are of concern for reporting, and leave the rest of the data intact for another day. There is no need to throw away precious transaction data. If you identify another attribute for reporting that is present in the JSON data, it is much easier to change the data transformation to incorporate it.

JSON creates a ***flexible structure***. Many reporting solutions are notoriously brittle when it comes to source data changes. Perhaps a vendor sends a spreadsheet with 10 columns today but adds an 11th column tomorrow. Some solutions will break until a there is a code change. If you serialize your data into JSON, as long as the 10 existing column headings don't change, you can store the new data. And only once your business has a need for that 11th column do developers need to make a code change.

Finally, JSON is ***easier to work with*** from a development perspective; thus, code changes are simpler to make. This means a ***faster turnaround*** time for source data changes.

IN-MEMORY COLUMN-ORIENTED DATABASE MANAGEMENT SYSTEMS (DBMS)

Technical overview: Unlike a traditional row-oriented DBMS, which stores each record serially in a file page, column-oriented DBMS stores each column in its own file page. To retrieve a given value with row-oriented storage, the DBMS must be able to read across all fields regardless of the fields to be used. With column-oriented storage, columns that aren't relevant to the query results aren't accessed, improving read performance. Additionally, because each file page contains only one data type, this allows higher compression and requires less data transmission, which increases speed. Adding new columns is also fast, since you provision new pages for new columns instead of reshuffling existing data.

Organizations adopt this technology largely to improve query performance, but it also changes the data modeling used in reporting and analytics. For example, the process of building reporting data transformation can be simplified—namely by removing the need to create surrogate keys in a star-schema.

Business impact: The first and most obvious benefit of column-oriented storage is faster queries and faster reports—in our experience, query performance can be more than 100 times faster. With query results often returning sub-second, report iteration and experimentation are easier to cycle through without settling for "good enough."

Column-oriented storage also makes it possible to write a query or use a reporting tool without having to worry about joining tables. This makes it ***easier for novice reporting application users*** to use the data. The key to fast performance is to select only the fields required and to use some filters. Tools like Tableau work well with this technology, making the process as simple as selecting a table source.

Report model changes may no longer require extensive data warehouse projects, making them ***less costly and time consuming***. In the past, adding a field or dimension to a model could take anywhere from six to 12 weeks. The ease of adding new columns and simplified data modeling greatly reduce that timeline.

Column-oriented storage also may enable you to ***defer investments in super-fast storage*** such as SSD SAN. While such investments will not be a waste of money, the query performance improvements delivered by column-oriented storage solutions may justify deferring additional storage purchases.

Finally, the reduction in report model complexity translates to up to 50-percent ***faster data transmission development***. Without the overhead of multiple transformation steps to accommodate each new source, developers can add new fields or new sources very quickly.

RESOURCE-SENSITIVE DESIGN PATTERNS

Technical overview: With a fixed (non-cloud) data center, you want to use your infrastructure as completely as possible without requiring more. A server that sits idle is a depreciating asset, while running out of CPU and memory regularly will degrade performance. Typically, environments are sized to accommodate some number of peak periods and some amount of idle usage. Certain cloud offerings can provision additional nodes or environment resources dynamically to react to peaks. This is helpful—but resource-sensitive design patterns can be useful, too, when you are able to fork off parallel processes.

In traditional data warehouse modeling, an orchestration usually drives most of the packages, and the system finishes when it finishes. On the other hand, design patterns that embed script tasks to check system conditions can help spawn new processes when the system is below a specified threshold (e.g., 90-percent CPU). Often, this can be done in parallel, and having logic that checks CPU usage will allow a solution that starts new processes until usage reaches a desired maximum utilization. If the CPU is above threshold, the process waits a few minutes and tries again. This controls the degree of parallelism to adjust system conditions dynamically.

Business impact: Design patterns that account for your resources will ensure the ***shortest loading time possible*** while making the most of your hardware investment. In addition, cloud-based solutions can offer ***greater resource control and lower costs***, since all systems do not need to be turned on all of the time. Turning off nodes will often save money, while turning on nodes can yield higher performance. Resource-sensitive design patterns provide automated control of this resource spend.

MATURATION OF OPEN SOURCE VISUALIZATION

Technical overview: Several emerging open-source solutions make visually attractive reporting and visualization capabilities attainable without substantial investment. One such example is D3, a set of libraries and examples that use JavaScript to create interactive visuals viewable from a web page. While these tools offer limited ad hoc analysis, they still address many reporting needs. As open-source tools, these often include examples that developers have shared for the purposes of collaboration. In fact, their ease of use and low cost can raise the question of whether you need packaged reporting tools at all. To be clear, packaged reporting tools are mission critical for larger organizations, but the price barrier can be limiting for many smaller organizations.

Business impact: While the capabilities don't compare to packaged reporting tools like Tableau, the low price point makes glossy visualization ***attainable for even the smallest organizations and/or those with limited budgets***.

SIGNIFICANT ON THEIR OWN, BUT EVEN BETTER TOGETHER

While traditional business intelligence and data warehousing methods remain critical to today's enterprises, there are also many emerging "disruptive" technologies that provide point solutions for diverse needs at a low cost. The five technologies described above are just a few examples of solutions that are disrupting traditional approaches and offering attractive benefits—without the substantial investments of time and dollars traditionally required to enhance reporting and analysis.

And the best part of all: We have found that these particular technologies, when combined, can provide powerful capabilities for consuming most any type of data and delivering reporting tables.

West Monroe Partners' Rapid Analytics Platform harnesses the five technologies described above to enhance the reporting capabilities of today's data-intensive organizations. To see the Rapid Analytics Platform benefits in action, please contact Dennis Donald (ddonald@westmonroepartners.com) or Vadim Orlov (vorlov@westmonroepartners.com).