USING BIG DATA AND ANALYTICS TO UNLOCK INSIGHTS

Robert Bradfield
Director
Dell EMC Enterprise Marketing

ABSTRACT

This white paper explains the different types of analytics and the different challenges at each level, and highlights Dell EMC data solutions designed to address the challenges. These select solutions, which incorporate technologies from Dell EMC and its partners, address just a few of the virtually infinite use cases for analytics.

October, 2016

TABLE OF CONTENTS

EXECUTIVE SUMMARY

In today's dynamic world, new technologies are emerging to allow businesses to better compete and to assist organizations as they grow and strive to serve their customers in more efficient and effective ways. These efforts center on the ability to unlock insights and extract value from data, and organizations around the world are turning to advanced data analytics capabilities, many that were originally developed for big data.

At Dell EMC, we understand — perhaps better than anyone — that data analytics must be viewed as a complete ecosystem that requires effective alignment of business processes with advanced analytics software, cutting-edge data integration and robust infrastructures, with each element having a critical role in the delivery of timely and accurate insights to better enable decision making.

It is important to note that analytics are not just the purview of large businesses, but that any organization — of any size, any industry or any sector — can benefit from improved analytics. Because every organization is unique, the mission of each organization or department may require different degrees of analytical rigor, but regardless of your specific requirements, improving or evolving your analytics capabilities can help you to drive fundamental organizational transformation, deliver greater operational efficiency, or both.

In this paper, we look briefly at the types of analytics and different challenges at each level, as well as highlighting three Dell EMC data solutions. These select solutions address only a few of the virtually infinite use cases for analytics and often feature our partnerships with Intel, Cloudera, Hortonworks, Microsoft, SAP, Syncsort and others.

WHAT ARE ANALYTICS?

In this context, analytics is the discovery and interpretation of meaningful patterns within data used by organizations to describe, predict and improve performance. Essentially, analytics can be grouped into four broad categories, based on the answers that each can give:

- "What happened?"
- "Why did it happen?"
- "What will happen?"
- "How can it be made to happen?"

WHAT HAPPENED?

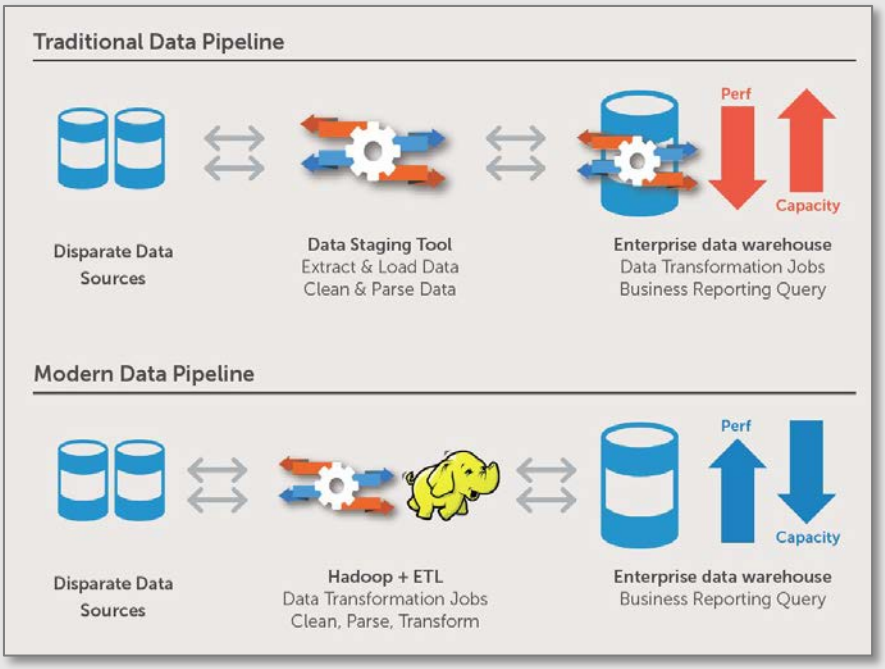
By far the most straightforward category is "What happened?" This reactive and historically focused category is often called descriptive analytics, with correspondingly simple and relatively forthright answers. For example, this category would include sales reports that show how much of which product was sold in a given month, or how many students with a particular major were enrolled in a given semester. Insights are not particularly deep, but often involve the merging of different data tables or sources, even data transformation, within databases or data warehouses in order to answer the question.

As business environments become increasingly digital, these data transformation challenges are magnified due to exponential data growth. This avalanche of data means even relatively straightforward data transformation takes longer and longer, and organizations are challenged to deliver even the most basic reporting in a timely manner. This is coupled with increasingly sophisticated users looking beyond traditional standardized reports in their search for insights, leading to more customized queries or specialized reporting.

Because of these challenges, improvements in descriptive analytics are primarily concerned with operational efficiency, such as containing or reducing the ever-expanding costs of data transformation, or delivering expanded or customized reports and queries faster.

SOLUTION SUMMARY: OFFLOADING ETL PROCESSES

The Dell EMC | Cloudera | Syncsort Data Warehouse Optimization–ETL Offload Reference Architecture (RA) provides a blueprint to help your organization build an environment to augment your enterprise data warehouse. The RA provides the architecture, beginning from bare-metal hardware, for running Extract-Transform-Load (ETL) jobs in Cloudera Enterprise with Syncsort DMX-h software.



Many customers have a skills-set gap when it comes to using Hadoop for ETL in their environments. They don't have time to build up expertise in Hadoop. The software components of this reference architecture help you address this challenge.

The Dell EMC solution makes it easy to build and deploy ETL jobs in Hadoop. Syncsort's high-performance ETL software enables your users to maximize the benefits of MapReduce without compromising on the capabilities and ease of use of conventional ETL tools. Syncsort software enables faster time to value by reducing the need to develop expertise on Pig, Hive and Sqoop, technologies that are essential for creating ETL jobs in MapReduce.

WHY DID IT HAPPEN?

The next area, diagnostic analytics, is where most people begin to uncover the real benefits of analytics. Both operational efficiency and transformative insights can come from effective diagnostic analytics, as the focus moves beyond basic reporting and static queries into a deeper understanding of causation.

Improvement to existing diagnostic abilities typically involves efforts to better "slice and dice" existing data, but it can also involve even broader integration of additional data sources, increasing the scope of information available. This may require bringing in unstructured data from social media for a better sense of customer sentiment to explain sales results, or perhaps incorporating sensor or Internet of Things (IoT) data for deeper insights into manufacturing processes.

For example, a simple sales report that indicates twice as many units of a given product were sold than was usual begins to take into account other information that can point to seasonality or advertising activities as the reason for the change. Or perhaps a manufacturer can now acquire the ability to trace back a product defect to a particular lot of components, a malfunctioning machine or a poorly trained associate. Knowing why something happens is the first step toward truly understanding how to fix an issue or seize opportunities.

As organizations reach this point in their analytics maturity, unforeseen challenges may occur of their own creation. We often see departments such as marketing or finance directly accessing data and creating their own unofficial business management systems in an effort to understand causation, and it is at this point that efforts at analytical innovation can actually limit long-term effectiveness. With so many seeking to better their individual initiatives or departmental missions through data-driven insights, multiple approaches can emerge in an effort to perform similar or even identical analysis. Since even basic key performance indicators (KPIs) can be calculated in different ways, this can create conflict and debate over results, instead of unified agreement and action on insights.

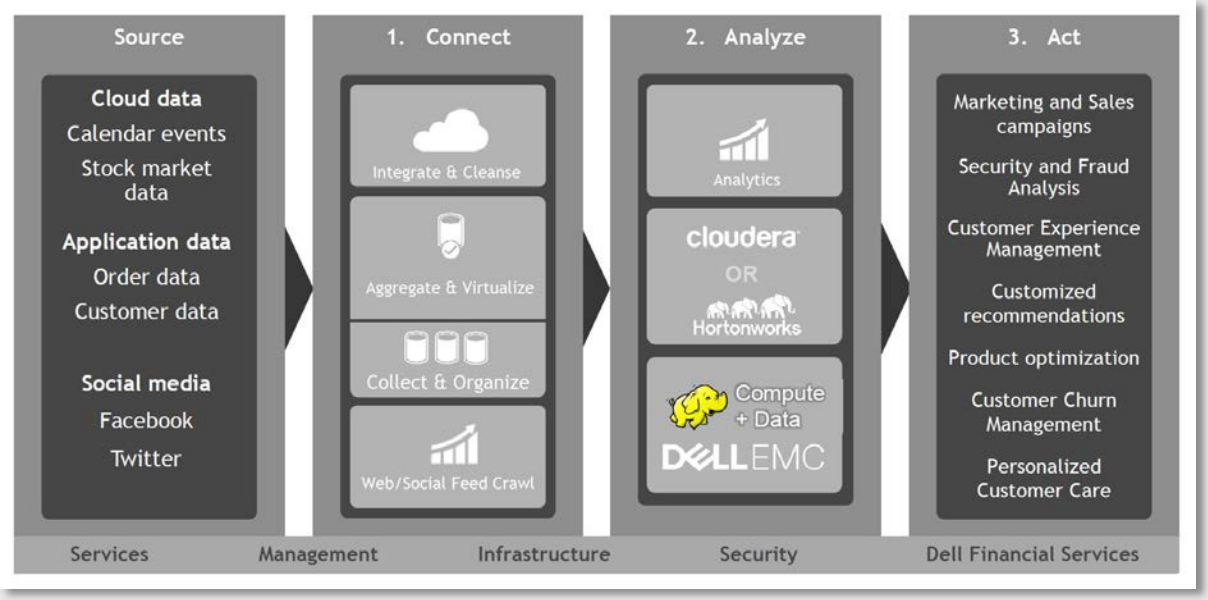
This fragmented approach can also lead to the development of “shadow IT” — business users essentially running their own data marts, from desktop systems or even personal servers — which causes a strain on IT to support unplanned and often rogue equipment, and more importantly, this can have a catastrophic impact on data integrity, with data warehouses struggling to keep data current in these often redundant, unofficial data marts. To combat this, it is critical that there be complete alignment between IT and the business regarding goals, expectations, resource planning, and investment requirements to support a holistic analytics infrastructure.

SOLUTION SUMMARY: ENABLING A 360-DEGREE CUSTOMER VIEW

The Dell EMC Big Data Marketing Analytics Solution brings together all the technology components your organization needs to gain a 360-degree view of your customers. This end-to-end solution allows you to manage, capture, integrate, and analyze massive amounts of structured, unstructured, and semi-structured data.

Integration of off-premises and on-premises data enables the synchronization of information across various data stores, so it’s ready for analysis. You can then leverage smart algorithms to derive actionable information from masses of data from disparate systems and sources. Better still, this isn’t a monolithic solution that forces you to replace your existing marketing analytics tools, and this isn’t a propriety solution that locks you into a particular platform. In fact, you can deploy the modular components individually to meet targeted needs, or you can work with Dell EMC to roll out an end-to-end solution tailored to your unique requirements.

In an age of intense competition for customers, you can’t afford to simply trust your instincts when it comes to efforts to retain your existing customer base. You need to enrich your natural understanding of customer retention with insights from big data. The Dell EMC Marketing Analytics Solution helps you get there today. It brings together a unique mix of technology components to help your marketing team gain a 360-degree view of your customers and work proactively to maintain their loyalty.



WHAT WILL HAPPEN?

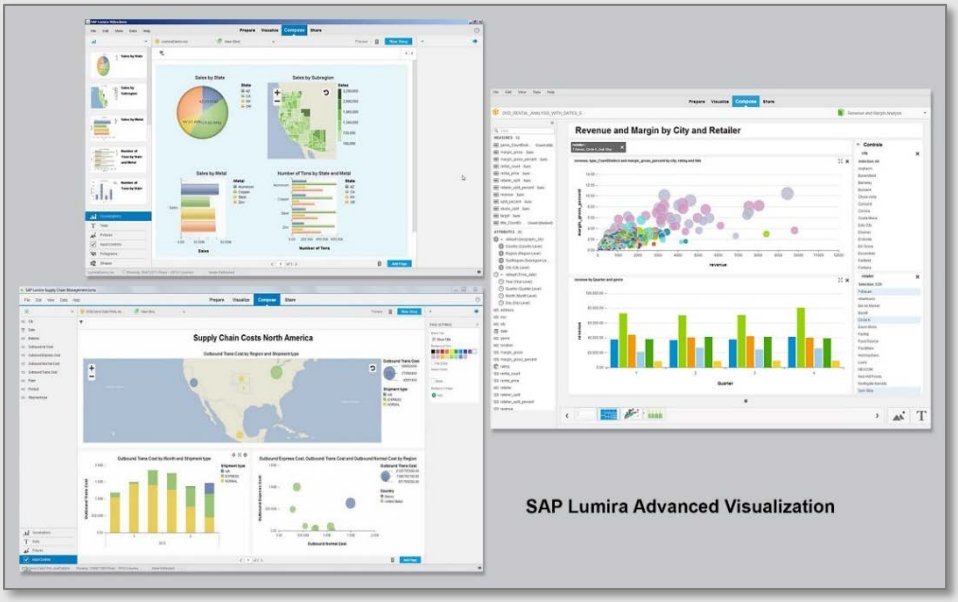
Predicting future results — predictive analytics — introduces the concept of causation and the discovery of statistical patterns that can forecast future risks and opportunities with acceptable levels of reliability. In a sense, it's simply taking diagnostic analytics to its logical next step, but to be reliable, statistical models must be continually validated and adjusted for accuracy, which tends to require the analysis of exponentially large amounts and types of data from such diverse sources as sensor or Internet of Things (IoT) data, genomic studies, unstructured data from social media, or rich content like video or photographs.

For example, some hospitals use predictive analytics in clinical systems that incorporate medical histories with real-time treatment data to determine which patients would be at greater risk for infection, and the financial services industry uses it extensively for such tasks as credit scoring, where continually adjusted credit models are used to determine lending rates. Predictive analytics can be extremely valuable in uncovering the deeper insights needed for fundamental organizational transformation, but can require significant resources and ongoing commitment, as predictive models are constantly evolving, often in real time.

SOLUTION SUMMARY: MAKING ADVANCE ANALYTICS AVAILABLE TO EVERYONE

Dell EMC and SAP have a long history of partnership in bringing data analysis solutions to businesses and agencies of all sizes. The Dell EMC Validated System for SAP HANA Edge offers blazing in-memory analytics, just like the Dell EMC Validated Systems for SAP HANA Enterprise Edition, but is engineered and sized specifically for midsize and small businesses, organizations and agencies, at a much lower entry investment. The latest evolution in real-time analytics platforms, this appliance enables report consolidation and self-service analytics to dramatically reduce the costs associated with business insights, while at the same time improving overall reporting services levels, finally placing near real-time analysis of data within easy reach.

SAP HANA, Edge Edition can pull data from multiple sources, load in-memory, and enable real-time predictive analytics for direct use by non-technical departmental users. Combined with embedded business logic that replaces separate staging and query platforms, the Dell EMC Validated System for SAP HANA Edge reduces data and infrastructure sprawl by aggregating and eliminating disparate machines that IT must manage and support, and the costs of business intelligence (BI) are reduced via data centralization, report standardization and the elimination of end-user support requirements. Focusing explicitly on data analysis, SAP HANA, Edge Edition is the perfect software platform for data marts, development, and real-time in-memory analysis for smaller organizations or even divisions of larger companies.



Dell EMC is proud to be the first to offer this easy-to-deploy and easy-to-manage system that finally makes advanced analytics accessible to everyone. Delivered with best-in-class Dell EMC hardware, as well as business adoption services that include integration services for database, migration, training, and business adoption workshops, the Dell EMC Validated System for SAP HANA Edge is ready to run upon delivery. Featuring built-in reporting logic, point-and-click predictive analytics, and SAP Lumira advanced visualization software that can quickly and easily translate reams of data into heat maps, scatter charts, bar graphs, or other pre-built and integrated libraries of visualization reports, the Dell EMC Validated System for SAP HANA Edge finally makes this power affordable to any business that wants to cut through the data deluge, no matter what size.

HOW CAN IT BE MADE TO HAPPEN?

This ultimate expression of analytics is sometimes called prescriptive analytics, which couples the understanding of complex systems and activities found in predictive analytics with recommendations or even automated execution of predefined steps to mitigate risk or take advantage of future opportunities. In other words, anticipating what will happen, when it will happen, and why it will happen, and then effectively determining or automating responses to rapidly address the revealed opportunities or risks.

This adds another layer of capabilities and additional resource demands, as prescriptive analytics must continually take in new data from a variety of structured and unstructured sources in order to improve prediction accuracy and refine the understanding of the implications of each decision option, which must be weighed against business rules and then pushed immediately back into the business applications that run the company.

Historically, the significant resources demanded by prescriptive analytics, both human and computational, made this type of analytics achievable only by the largest organizations who could afford the massive investments. However, the entry costs of technologies that enable real-time data collection, such as in-memory computing infrastructures, continue to decline. As the adoption of these infrastructures grows, enterprise applications are simplifying the alignment of best-response logic with these real-time insights, making prescriptive analytics more accessible than ever.

SUMMARY

Together we have looked at different categories of analytics and a brief glance at just a few of the many Dell EMC solutions designed for assisting you with addressing your data analytics challenges. Not every organization requires the same level of analytical rigor, and no two organizations are coming from exactly the same place or have exactly the same goals. Your organization may simply require enhancements to your data infrastructure to allow for greater operational efficiency or more timely responses, or it may need more intensive integration of new data sources to allow for deeper insights necessary for fundamental business transformation to better serve customers or address competitive pressures.

Regardless of where you are in your personal data analytics journey or which type of analytics your organization needs to best accomplish its unique mission, Dell EMC has the expertise necessary to assist you to improve or evolve your capabilities.