Enterprise Strategy Group by TechTarget

SHOWCASE

Seizing the Data Lakehouse Opportunity with Dell Validated Design

Date: January 2023 **Author:** Mike Leone, Senior Analyst

ABSTRACT: As data continues to expand at dizzying rates in both volume and complexity, many organizations have sought a new solution that combines the best qualities of data warehouses and data lakes—a data lakehouse. But it's easy to get caught up in technical snags and confusing details without the right blueprint for designing, building, and deploying a data lakehouse. Dell Validated Design for Analytics—Data Lakehouse helps simplify a data lakehouse implementation without compromising performance and cost efficiency.

Introduction

Data analytics is undeniably at the heart of many organizations' digital transformation initiatives. For one thing, analytics fuels how organizations make better decisions, faster. Analytics drives an endless range of business goals, from anticipating customer requirements and building a stronger competitive position to determining the ideal selling strategy to enhance market share, profitability, and customer loyalty. But there are many things standing in the way of organizations looking to use data analytics to accomplish their key goals and put them on an improved business footing. As such, organizations are looking at new data architectures like the data lakehouse to unify and simplify data for the wider business.

The Complexity of Data Environments

Complexity begins with the never-ending challenge of just keeping up with the massive, unrestrained growth of data in all ways—volume, variety, and velocity. Data takes many forms, and the mountainous growth of data volumes in unstructured and semi-structured formats makes it extremely difficult for technical users, such as data scientists and data-savvy business users, to locate and access the data they need. Making it even more challenging is the fact that it is harder than ever to discern where the real value of that data lies and how best to put it to work.

The challenges in creating powerful, insightful analytics in an increasingly confusing and complex environment fall into three categories:

- Generating sufficient value from data depends upon many technical and business factors, including several that are difficult to anticipate, control, or even understand without prior exposure to the data and its related workloads.
- Complexity is choking the pace of innovation. That complexity can be technological (the rapid rate of improvements in areas like compute, storage, data warehouses, and artificial intelligence) or business process-related.
- Data security becomes harder to attain and ensure over time as data volumes grow and spread. Cybersecurity threats are multiplying in number and complexity, and hackers' sophistication has grown beyond the capabilities of most organizations' already-outnumbered security teams and resources.

As a result, the essential task of analyzing data for measurable, actionable insights is a huge challenge due to the sheer volume of data and the difficulties in locating and extracting it into a usable format. This often results in the creation of additional data silos, each with its own data and each designed for what its architects felt was a highly unique requirement. Traditional solutions, such as data warehouses and, more recently, data lakes, were initially helpful, but they have since morphed into their own data silos, as they are typically built and deployed on separate foundations. Perhaps even more problematic is that earlier data warehouses—and even many first-generation data lakes—aren't designed to accommodate increasingly valuable artificial intelligence/machine learning tools, thus leaving "money on the table" when it comes to gaining valuable data insights.

This dual-approach solution also taxes already-stretched in-house IT administrators to support data scientists' desire to get a complete, real-time view of all relevant data needed to inform transformative insights. The resulting complexity stifles, rather than promotes, innovation.

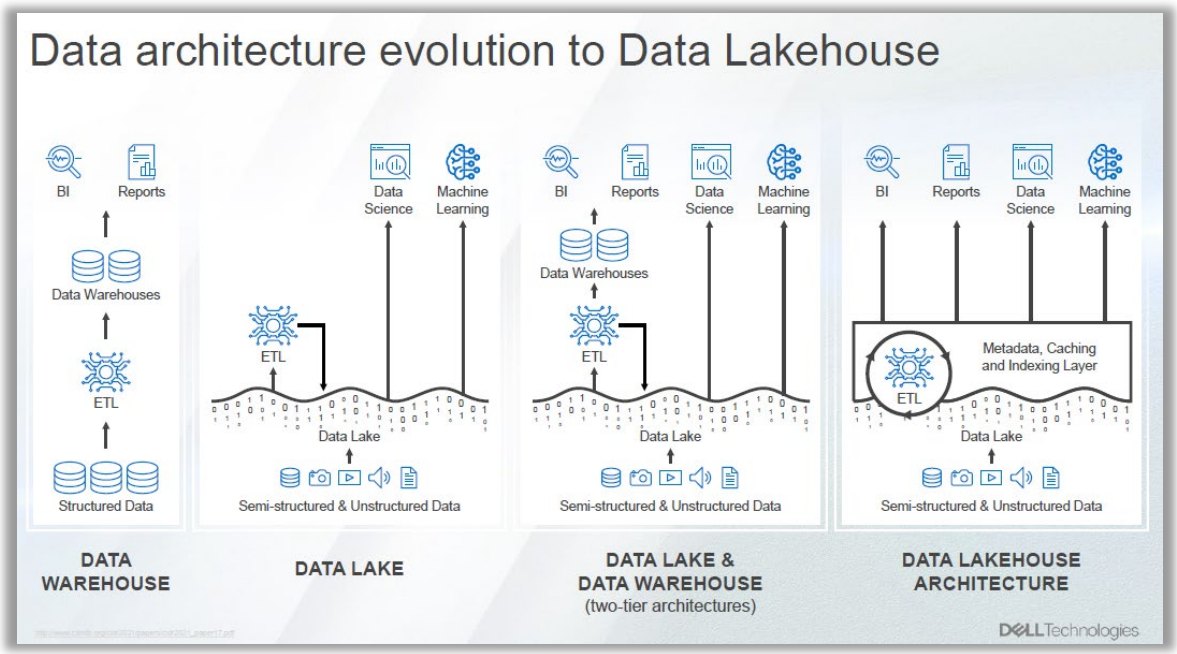
This has driven many organizations to look for newer, more flexible, and more modern solutions designed to overcome these problems. This is where a data lakehouse comes into play.

Benefits of a Data Lakehouse

Data lakehouses have sprung up fairly recently, but they have quickly demonstrated their worth in several ways when compared to previous data architectures with data warehouses and data lakes.

First, data lakehouses create more value—and do it more quickly and efficiently—from the hordes of data organizations are faced with. The self-service user model for data lakehouses improves accessibility, particularly for business users that lack the technical skills or experience of data scientists. They also provide on-demand tools for interactive queries, as well as integrated techniques to optimize performance and eliminate bottlenecks. This creates tremendous economic value; customer research from Dell Technologies indicates that those organizations leveraging data lakehouses have generated more than $60 million in increased economic benefits over the past three years, while also saving even more money in cost avoidance with fraud detection.

Figure 1. The Data Architecture Evolution to the Data Lakehouse



Source: Dell Technologies

Second, data lakehouses achieve the critical goal of simplifying the data landscape. This is done through a number of carefully engineered design features, such as faster configuration and integration, a single, unified platform to break down and eliminate data silos, and the availability of all data regardless of the data's location, whether in on-premises data centers or in a colocation facility.

Finally, data lakehouses do more to protect and secure data to prevent problems and to ensure availability of the right data at the right time for the right requirements. For instance, it is far easier to protect and secure a single data platform (the data lakehouse), compared to more frequently seen, silo-based architectures. Security is further enhanced through embedded access control and audit logging, which also improves the organization's ability to demonstrate compliance and achieve proper data governance. Also, data lakehouses improve overall data quality by limiting ETL transfers between the data lake and the data warehouse, as typically was done in earlier data analytics architectures.

Dell Validated Designs for Analytics—Data Lakehouse

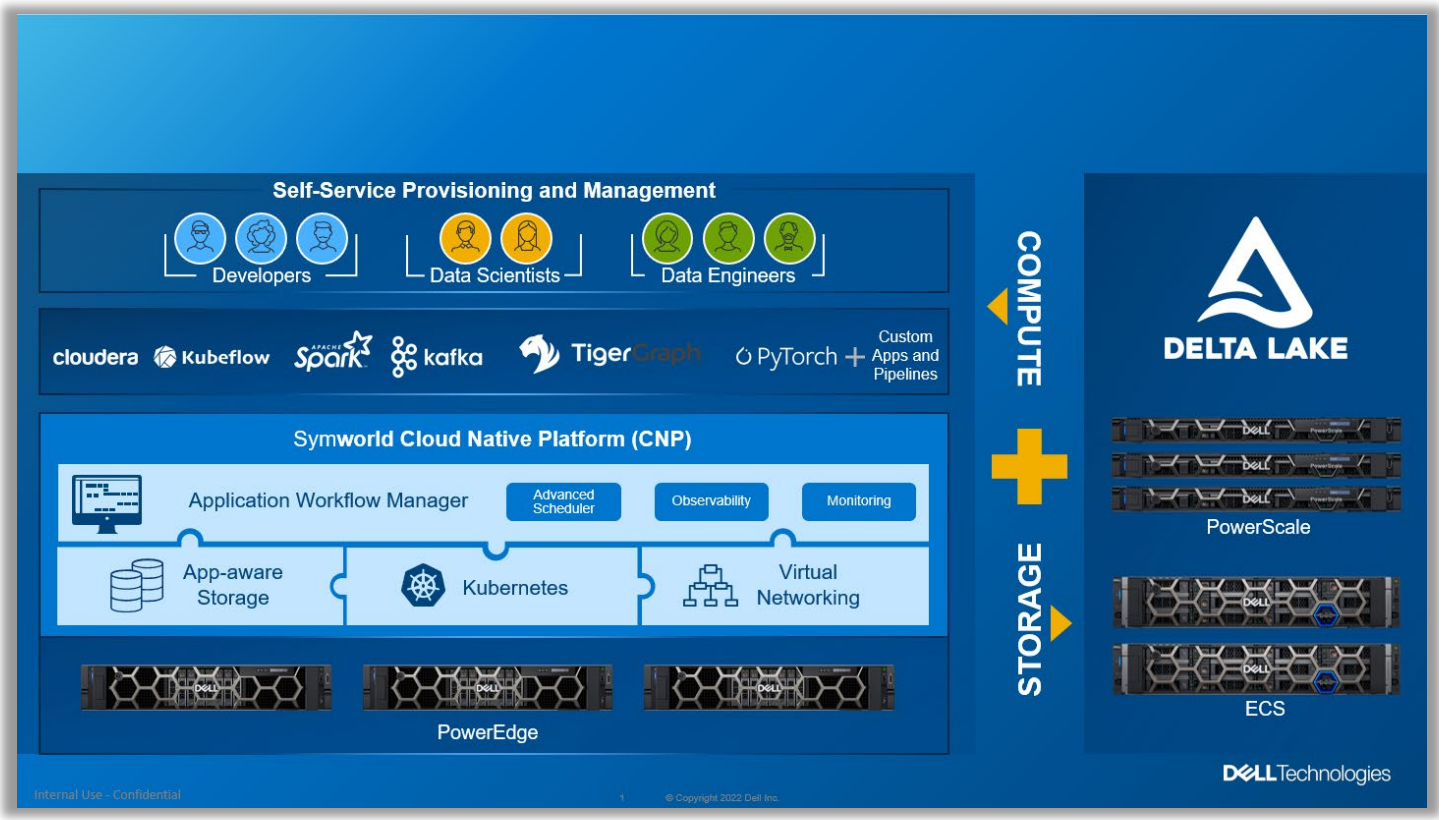
Dell Technologies has created a modernized solution called Dell Validated Designs for Analytics—Data Lakehouse. This solution is targeted at both on-premises data centers and colocation deployments that need optimized, sustained performance, high availability, efficient deployment, and robust security.

This solution offers an on-demand, self-service capability for a range of different types of users, from sophisticated data scientists deploying their own clusters without IT involvement to business users who just want something that works out of the box so they can begin to uncover insights quickly and generate tangible business benefits.

The Dell solution is based on open-source technology and offers a full set of integrated services and support for customers. It is designed to provide modern, efficient, and scalable data management, combining powerful data analytics capabilities traditionally seen in data warehouses with the cost efficiency and agility typically embodied in data lakes, all while storing all types of data (structured, semi-structured, and unstructured data) in a uniform way.

Dell Validated Designs for Analytics—Data Lakehouse supports both business intelligence and artificial intelligence/machine learning. Its open architecture design includes affordable, highly scalable storage, and it breaks down the often-prevalent data silos of older architectures into a single, unified data format that anticipates, spots, and remediates performance bottlenecks that frustrate users and add to data management complexity. Simplifying the data architecture benefits organizations by making it easier and faster to identity and extract new insights that drive business innovation. Importantly, it does so securely because of the native security tools engineered into the solution.

Figure 2. Dell Validated Design for Analytics



Source: Dell Technologies

Finally, the solution is available through multiple Dell-validated bundles for different use cases, due to Dell's wide and deep technology partner ecosystem.

Easing the Process and Improving Deployment Efficiency With Dell Technical Services

Planning, building, and deploying a data lakehouse takes time, experience, and expertise. In-house organizations, often stretched to the max simply handling day-to-day tasks and other, often-termed "keeping the lights on" activities, need the help of a proven technology partner to get the most benefit from their data lakehouse.

Dell Technical Services fills out the Dell data lakehouse solution with its ability to deploy and support the full hardware technology stack, software solutions and tools, process innovation, and business and technical consulting. Dell's proven expertise, combined with the company's substantial R&D and support resources, helps drive innovation and long-term business and economic value. Dell Technical Services also provides a deep bench of skills and experience to design, build, and deploy the right solution for each organization, rather than an off-the-shelf, cookie-cutter approach.

This Dell team brings to bear a full array of technologies, services, process knowledge, implementation skills, and consulting on how best to derive an optimized return on investment. Together, these capabilities help organizations identify the right outcomes, determine how to achieve them, put together the solution to make it happen, and build the right metrics to chart progress toward the desired outcomes.

Dell also provides options to help organizations choose the best way to pay for the solution, including subscription models, traditional CapEx purchasing and leasing, or as a managed service. It also allocates "resident experts" to work alongside in-house teams to ensure everything is running smoothly and to identify new opportunities to create additional value from current and future data.

The Bigger Truth

As organizations increasingly look for simpler, more efficient, yet still powerful solutions for advanced analytics amid growing data complexity, data lakehouses have become an attractive, modernized alternative to data warehouses and data lakes. But the proper planning, build-out, and deployment of a data lakehouse requires a combination of the right hardware, software, and services with an experienced eye toward how to adapt that solution for state-of-the-art analytics.

Dell Validated Designs for Analytics—Data Lakehouse offers organizations a purpose-built solution to help them gain greater value from their data in all their devices and data sources. Tying together data from the data center to the edge to the cloud, including the full range of mobile platforms and internet of things devices, the Dell solution helps organizations realize the much-promised benefits of a data lakehouse.

When combined with the technical and consulting skills of Dell Technical Services, the solution helps organizations leverage data wherever it resides in an efficient, secure, and manageable approach.

Enterprise Strategy Group by TechTarget

Enterprise Strategy Group is an integrated technology analysis, research, and strategy firm that provides market intelligence, actionable insight, and go-to-market content services to the global IT community.

www.esg-global.com contact@esg-global.com 508.482.0188