DELLTechnologies

Unleash the Combined Power of Your Data

Run analytics and ML on a single data lakehouse that enables you to create value from data while simplifying your data landscape and protecting data everywhere.

Data Lakehouse: Defined

A data lakehouse combines the flexibility, cost-efficiency and scale of a data lake with the quality, performance, security and governance of a data warehouse, enabling direct access to a single source of trusted data for analytics, AI and other data-intensive workloads.

With the Dell Validated Design for Analytics — Data Lakehouse, you can capture and use all types of data — structured, unstructured and semi-structured — in an open data lake with the data quality, performance, security and governance you expect from a data warehouse along with self-service, on-demand flexibility.

The Data Lakehouse, developed jointly with leading partners, enables you to evolve data management from siloed, rigid, costly and slow to unified systems that enable business intelligence (BI), analytics, real-time data applications, data science and machine learning (ML). Consisting of Dell PowerEdge servers, Dell PowerScale, Dell ECS Object Storage and Dell PowerSwitch networking and powered by Apache® Spark® and Kafka® with Delta Lake technologies and Robin Cloud Native Platform (CNP), this solution is designed to help you harness more data to transform insights across your organization.

Provide rapid, direct access to trusted data.

In the data-driven era, you need to be able to generate value from all of your data capital, from edge to multiple clouds to core. But the distributed nature of data can make that complex and costly — setting up barriers to insight and innovation. Traditional data management systems, like data warehouses, have been used for decades to store structured data and make it available for analytics. However, data warehouses aren't set up to handle the increasing variety of data — text, images, video, Internet of things (IoT) and so on — nor can they support artificial intelligence (AI) and ML algorithms that require direct access to data.

Adding a data lake promised to help solve these issues, by enabling enterprises to capture and process all types of data — structured, unstructured and semi-structured — more flexibly and cost-effectively than traditional data warehouses. Today, many organizations use a data lake in tandem with a data warehouse — storing data in the lake and then copying it to the warehouse to make it accessible to users — but this adds to the complexity and cost of the analytics landscape.

To compete in the digital era, your organization needs new solutions that evolve data management from siloed, rigid, costly and slow to unified systems that enable analytics and AI with speed, scalability and confidence. The Data Lakehouse supports BI, analytics, real-time data applications, data science and ML in one, providing rapid, direct access to trusted data for the data scientists, business analysts and others who need data to drive business value.

Learn more

Key benefits

- **Create value from data** — Better performance and data quality and control give you the power to create more value from data.
- **Simplify your data landscape** — Provide a single source for all data and eliminate the need for separate systems to serve real-time data applications.
- **Protect and secure your data** — Fine-grained security along with advanced data management and governance add reliability and quality to your data lake.

Technical specifications

Validated Design for Analytics — Data Lakehouse takes a unified approach to data-driven use cases from analytics to AI, storing enterprise data in an open data lake with the quality, performance, security and governance of a data warehouse. This eliminates the need to copy or move data between environments and provides a unified source of reliable, quality data.

Consolidating data management into a data lakehouse enables support of analytics, AI/ML and other data-driven applications, improving efficiency and lowering costs and complexity. Robin CNP provides an open-source Kubernetes® platform that delivers cloud-native agility for data-centric applications.

Master nodes	Worker nodes	Networking	Storage	Software	Kubernetes Platform
3x minimum of PowerEdge R650	4x minimum of PowerEdge R750 (NVIDIA GPU optional)	2x minimum of PowerSwitch S5248F-ON	• 3x minimum of PowerScale H5600 • 5x minimum of ECS EX500 Enterprise Object Storage	• Apache Spark • Apache Kafka • Delta Lake • Parquet • NVIDIA AI Enterprise (optional)	Robin CNP

Validated Design for Analytics — Data Lakehouse helps you harness the benefits of analytics and AI on optimized solutions designed from the ground up to unlock the value of your data capital, simply and securely. Contact your Dell Technologies representative to find out more today.

DELLTechnologies