Dell Open Data Lakehouse

Unleashing the power of AI begins with a Robust Data Foundation

Today's Situation

Capitalizing on AI begins with Data

Organizations are adopting a forward-thinking strategy, treating AI and Data as a product, to lead in innovation and value creation. In the digital age, data is a strategic asset, and AI is transformative. Maximizing AI's potential begins with a robust data foundation. It enables AI to thrive, ensuring ongoing model improvement. Investing in data ensures immediate effectiveness and long-term innovation.

The Data Challenge

In the rapidly evolving landscape of data management, organizations are grappling with a multitude of challenges including:

- **People and Skills:** Data teams spend only 22% of their time innovating. Traditional processes, delays, and skills shortages hinder progress.
- **Data Complexities:** A recurring challenge is the significant time invested by data scientists and engineers in obtaining and prepping data for AI workflows.
- **Expansion of data:** The proliferation of data, including multiple copies, across growing multicloud environments is hampering organizations' ability to fully leverage AI for business outcomes.
- **Infrastructure:** Managing data involves complex tasks, from pipelines to security compliance. Agility in infrastructure management is crucial amidst a complex tool landscape.

Why a Data Lakehouse, Why Now

For many organizations, addressing the challenges sited above involves a transformation to create a cohesive adaptable data ecosystem.

**Traditional Data Stack:** Today's data landscape has evolved to blend traditional data components: Data Warehouses and Data Lakes. Data Warehouses excel in structured data processing, while Data Lakes handle unstructured data with flexibility and scalability. This dual approach combines agility and governance but often leaves a gap in data agility and cost efficiency. Challenges like data integration, accessibility, and insight derivation across diverse data types can hinder decision-making and increase operational costs.

**Modern Data Stack:** Data Lakehouse architectures have emerged as a powerful solution that bring the best of both worlds i.e. Data Warehouses and Data Lakes. A Data Lakehouse provides:

- **Unified Data Stack:** Reduces data movements, simplifying the data landscape and cutting costs.
- **Enhanced Data Warehouse Capabilities:** Enables transaction management and time travel features, akin to traditional data warehouses.
• **Scalability and Performance**: Separates storage and compute architecture, ensuring scalability and performance.

• **Open Data Formats**: Leverages open formats like Iceberg and Delta Lake, reducing vendor lock-in and offering flexibility.

• **Unified Governance**: Provides consistent governance and access control, ensuring data is available to the right users.

• **Data Source Compatibility**: Access and process data from various sources, including RDBMS, data warehouses, and data lakes, across on-premises, cloud, and edge environments.

• **Data Product Creation**: Facilitates the development of data products, promoting data discovery and reuse across the organization.

**Dell Open Data Lakehouse**

You can now partner with Dell Technologies to start your data modernization journey. At Dell, we recognize that every enterprise is at a different stage in their digital transformation journey, and we are committed to meeting you precisely where you are.

**Logical View**

Per the illustration below, the Dell Open Data Lakehouse sets the foundation for a modern data stack and simplifies the data landscape by minimizing data movements and consolidating data silos.

![Open Data Lakehouse Diagram](image)

Our customer-centric approach ensures that your IT solutions are not only tailored to your current needs but also primed for future advancements.

**Key capabilities include:**

- **Analytics and AI/ML Workloads**: Dell partners with an open ISV Ecosystem empowering customer to plug into their existing tools to meet the varying demands across their enterprise. This includes tools for data processing, BI reporting, AIML Ops and many more.

- **Open Data Lakehouse**: Dell’s Lakehouse engine is powered by Starburst, the leading data analytics platform that allows users to activate data wherever it resides in their enterprise.

Get ready to unleash the power of AI with Dell Open Data Lakehouse Solution

© 2023 Dell Inc. or its subsidiaries.
The Data Lakehouse streamlines data operations by enabling in-place discovery and access to data across the enterprise, whether on-premises or in the cloud. It supports data ingestion from a variety of sources, allowing data to be materialized within the lakehouse for further processing and transformation, ensuring data is well-prepared for analysis and decision-making.

- **Multicloud Data Sources:** The Dell Data Lakehouse open source software and open architecture enables data federation for joining data across various sources and/or moving a data from as events, costs and needs change. It promotes data sharing through curated data product via connectors like Starburst Stargate.

- **Dell Storage:** Dell Object and File Storage provides a versatile data store for structured, semi-structured like CSV, XML and JSON, and unstructured data like images, .pdfs, audio, videos. Dell object storage supports read/write consistency within and across sites for handling concurrent, atomic transactions with the choice of two open table formats, Delta Lake and Iceberg. Both use parquet and Iceberg can also support Avro or ORC.

- **Dell Compute:** Dell PowerEdge computing and networking are finely tuned for scaling and high-performance handling of diverse data workloads. Effortless management of Kubernetes distributions like RedHat OpenShift, Rancher and Symcloud empowers seamless deployment and lifecycle management of the lakehouse stack at an enterprise level.

**Dell Validated Designs**

Dell Validated Designs offer substantial value by providing pre-tested, reliable solutions that streamline IT infrastructure deployment. These well-documented blueprints reduce the risk of errors, optimize system performance, and accelerate time-to-value, making them a valuable asset for efficient and successful IT projects.

**Powered by Starburst together with leading open source technologies, designs are built on leading Kubernetes platforms including RedHat OpenShift container platform, SUSE Rancher and Symcloud Native Platform.**

Get ready to unleash the power of AI with Dell Open Data Lakehouse Solution
© 2023 Dell Inc. or its subsidiaries.
Dell essential compute, network and storage power Dell Open Data Lakehouse validated designs include;

<table>
<thead>
<tr>
<th>Component</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dell Compute</td>
<td>The Dell PowerEdge server infrastructure delivers essential compute, memory, and storage resources vital for running customer workloads effectively. With a broad range of PowerEdge server configurations available, these recommendations cater to a diverse set of workloads commonly encountered in a Lakehouse architecture, ensuring versatile support for various data operations and analysis.</td>
</tr>
<tr>
<td>Dell Storage</td>
<td>The Lakehouse storage system has maximum flexibility leveraging Dell PowerScale with HDFS protocol or utilizing Dell ECS and Dell ObjectScale for object storage through the S3 protocol. The choice of configuration and scale is tailored to the specific workloads and the volume of data that needs support.</td>
</tr>
<tr>
<td>Dell Network</td>
<td>Built on the foundation of Dell PowerSwitch and SmartFabric technologies, the network is crafted to fulfill the demands of a high-performance and scalable cluster. It offers both redundancy and seamless access to comprehensive management capabilities</td>
</tr>
</tbody>
</table>

Why partner with Dell

Dell Technologies offers a suite of resources that bring immense value to organizations embarking on their digital transformation journeys.

**Dell Executive Briefings:** Tailored sessions that enable leaders to engage directly with experts and explore solutions aligning with specific business needs.

**Dell Customer Solution Centers:** Hands-on environments where organizations can test-drive technologies, fine-tune configurations, and validate strategies before implementation, reducing risks and enhancing success. Contact your account team to submit an engagement.

**Dell Specialists for Data Management and Analytics Solutions:** Dell Technologies has a dedicated team of specialists and experts who focus on data management and analytics solutions. These experts provide tailored guidance and support to organizations, helping them optimize data infrastructure, implement data governance policies, and harness data analytics for valuable insights.

**Dell Global Footprint:** A worldwide presence that guarantees support and scalability, facilitating seamless technology deployment and management on a global scale, making it a trusted partner for organizations worldwide.

**Sizing Guidance:** For more information, including sizing guidance, technical questions, or sales assistance, email analytics.assist@dell.com, or contact your Dell Technologies or authorized partner sales representative.