El Data Lakehouse se creó cómo una arquitectura abierta que combina los beneficios de un Data Lake con el poder analítico y los controles de una Data warehouse (almacén de datos) construido sobre un Data Lake (lago de datos).

El Data Lakehouse puede almacenar datos de todos los tipos y convertirse en una única fuente confiable de verdad, proporcionar acceso directo a Data Lakehouse de IA, cómo la plataforma Databricks Lakehouse que ofrece varias características:

* Soporte de transacciones: Incluidas transacciones ácidas para interacciones de lectura y escritura simultaneas.
* Aplicación de esquemas y gobernanza para la integridad de los datos y auditorias solidas.
* Necesidades de gobernanza de datos para respaldar la regulación de la provacidad.
* Soporte BI: De métricas de uso de datos para reducir latencia entre la obtención de datos y la extracción de conocimientos.
* Adicional el Data Lakehouse ofrece almacenamiento desacoplado de la computación, lo que significa que cada uno opera en sus propios clústeres lo que les permite escalar de forma independiente para satisfacer necesidades específicas.

Formatos de almacenamiento abierto cómo Apache Parquet, que son abiertos y estandarizados para que una variedad de herramientas y motores puedan acceder a los datos de manera directa y eficiente.

* Soporte para diversos tipos de datos para que una empresa pueda almacenar, refinar, analizar y acceder a datos semiestructurados, estructurados y no estructurados en una sola ubicación.
* Soporte para diversas cargas de trabajo, lo que permite una variedad de cargas de trabajo, cómo el aprendizaje automático de ciencia de datos y el análisis SQL para utilizar el mismo repositorio de datos.
* La transmisión de un extremo a otro para informes en tiempo real elimina la necesidad de un sistema separado dedicado a aplicaciones de datos en tiempo real.

El Data Lakehouse respalda el trabajo de los analistas, científicos e ingenieros de datos, todo en un solo lugar ya que es la versión modernizada de un data warehouse que proporciona todos los beneficios y características sin comprometer la flexibilidad y profundidad de un Data Lake.

