THE INSIDER

# Key Considerations for Moving to a Data Lakehouse

Is your data warehouse struggling to keep up with diverse data types and sources? Are the added costs piling on? If so, it may be time to consider a data lakehouse.

A lot of our clients have been asking about what's involved with making a move from a traditional data warehouse to a more modern data lakehouse. Although it can be a bit involved, we simplified it for you with some key considerations below.

**Tracey Doyle**
Analytics8 CMO

# Key Considerations When Moving from a Data Warehouse to a Data Lakehouse

There are many advantages you gain from moving to a data lakehouse, but to fully realize data lakehouse benefits, our data experts made a list of 6 crucial factors to consider and how to address them.

- **Decouple storage and compute:** Separating storage and compute in a data lakehouse provides cost and performance benefits and flexibility, but it's important to experiment with compute and monitor cost to establish resiliency in these areas.
- **Expect structured, semi-structured, and unstructured data:** A data lakehouse allows developers to work with diverse data types, making it easier to incorporate new data sources and expand into new markets or entities.
- **Support polyglot (several different languages):** A data lakehouse enables the entire data team to use different tools and languages on a single platform, but best practices for when to use python or SQL are necessary.
- **Work with files vs. a database:** Curating data within a data lakehouse requires careful extraction and integration from the source to prevent data swamp. Starting with one or two workloads and building a curated “raw” layer will help here.
- **Choose your open-source file format:** Choosing a workload to migrate and considering retention policies, data size, and end goals is crucial when optimizing a data lakehouse with file partitioning and table formats like DELTA LAKE and ICEBERG.

- **Stop copying data everywhere:** A data lakehouse's multi-hop architecture reduces data duplication, but strong data governance and access controls are essential for sensitive data. Consider both security and access when transferring data to the lakehouse platform.

## What About Challenges?

The move from a data warehouse to a data lakehouse presents consistent challenges. The following are some of those issues and pro tips to overcome them:

**Configuration:** Migrating to a data lakehouse architecture requires more configuration than a traditional cloud data warehouse.

> ⇒ Pro Tip: Invest in experienced cloud architects or data and analytics consulting firms to configure your data lakehouse architecture. Also, fully understand your workloads and data storage requirements to optimize cluster and engine configurations.

**Access Control:** This can be a challenge for organizations without clear security processes.

> ⇒ Pro Tip: Develop a comprehensive security plan that clearly defines access controls and policies for cluster and engine creation and usage.

**Compute/Engine Start:** Although serverless compute is reducing this pain point, this can still be frustrating.

> ⇒ Pro Tip: Set realistic expectations for users and automate the spin-up process as much as possible to manage frustration and optimize cost.

**Data Quality:** Data quality can be a challenge due to the variety of formats in which data can be stored in a data lakehouse.

> ⇒ Pro Tip: Establish a data governance framework that outlines clear policies and procedures on how to properly curate and manage data, and invest in tools and training to enable your data team.

**Cost:** Improper configuration of compute resources can result in unnecessary costs.

> ⇒ Pro Tip: Monitor and optimize compute usage. This could involve using serverless compute or right-sizing compute

resources based on workload demands, with the help of cost management tools and processes.

**Get more info about moving to a data lakehouse**

## More on Data Warehouse, Data Lake, & Data Lakehouse

No matter where you're at in your data architecture journey, we have guides that will help you get the most out of your existing data stack.

- Data Warehousing Guide: from strategy to implementation
- Data Lakehouse Guide: what makes it different and how to fit it into your existing architecture
- Video: How a data lake, data warehouse, and a data lakehouse differ

**Watch Tony Dahlager, Managing Director, break down these concepts**

# Enjoying this content?

Send this newsletter to someone or explore more content like this.

**Check out recent issues of The Insider**

Analytics8

*Transform your business with data.*

**Subscribe to our monthly newsletter**

**The Insider** includes our latest insights, tips, and advice to transform your business with data.