

Transforming Retail Legacy Systems with a Data Operating System



Companies are spending more than they think maintaining legacy systems, but what's the alternative? Offloading them? If the thought of burning bridges with legacy systems made you sweat, there is a way to integrate these systems into your new stack without dramatically increasing (already) creeping costs or risking losing your historical data. You need a new way to think about your data warehouses and data lakes—one that addresses how these systems are evolving over time.

Hidden costs of maintaining legacy systems

If legacy systems are basically doing their jobs, it's challenging to justify the cost of upgrading. However, it's time to take a close look at what legacy systems actually cost.

Legacy systems can cost organizations hundreds of millions of dollars to maintain. Each year, those costs grow an average of 15% and, for many companies, make up a large portion of their technology budget. Enterprises end up deep in the weeds managing these data systems with no way out and no way to relieve the burden. However, uncovering hidden costs is a step toward freeing data to achieve its true potential and stopping the budget-bleed. Hidden costs include the following:

IT department backlogs

Data should be in motion. When IT spends all its time trying to keep warehouses and lakes from becoming swamps (and liabilities), they don't have as much time to build new tools to ensure the company can use all that data. They can't innovate, and they can't explore new ways to reduce silos.

Integrations, upgrades, and technical debt

Imagine that a retail company purchases a competitor. It gains valuable data but an outdated warehouse. Now, to understand their customer purchase history, they have to spend money to upgrade away from the original, disorganized framework. Technical debt—inconsistencies and incompatibilities as new components come into play—builds up over time and makes it challenging to integrate new, necessary tools.

Downtime and missed opportunities

Legacy systems can experience downtime thanks to outdated hardware and software. This downtime could cost companies an accurate view of inventory, leading to overspending in purchases and increased storage costs. It could lead to vulnerabilities as others try to manage extracting data around the outage. It also costs companies in missed business opportunities because data is not responsive.



Exploring Legacy Systems: Data Lake vs Data Warehouse

Most retail operations have a combination of legacy warehouses and lakes storing historical data from multiple sources. Digital transformation means connecting each of these systems to functional pipelines and upgrading the architecture to make them both accessible to stakeholders.

Creating a new processing system—an operational layer, if you will—requires an understanding of the differences between these systems.

- **Processing:** Warehouses typically use ETLs (extract, transform, and load), and lakes use ELTs (extract, load, and transform). Managing these different processing requirements may call for an entirely new type of processing.
- **Analysis:** Warehouses contain structured data already scrubbed and ready for use. This lends itself to operational use. On the other hand, lakes contain unstructured data in their original form, making them suitable for experimental queries. Solutions that blend these components provide a holistic view of the company and its customers.
- **Compute and storage:** Lakes decouple the compute and storage layers while warehousing integrates both. An operational layer could streamline the complex process of integrating both data sources.
- **Cost:** Warehouses can bring more rigidity and thus greater costs. While lakes are more affordable, the technologies are not as mature. However, both can bring significant ROI with the right integrations and architecture.
- **Purpose:** Warehouses provide preformatted information suited for business users while lakes offer a flexible repository suited for data scientists. Blending these legacy systems will require an operational layer that can reduce complexity while still allowing complex queries.

Upgrading legacy systems using a data operating system

A data operating system provides the connective tissue to unite warehouses and lakes. It can simplify data pipelines and ensure that both business users and data science teams can access and query data on their own terms.

Upgrading warehouse and lakes should provide:

- **Consistency:** Administration can determine governance, data comes in usable forms, and pipelines generate automatically with transparency.
- **Accessibility:** All users within the enterprise can leverage data to make decisions—whether the interface is command-line for data engineers or automated using machine learning for business users.
- **Multiple uses:** Whether rigidly defined like warehouses or open and undesignated like lakes, an operating system creates a connection between storage systems with many different uses.

DataOS provides these things for enterprises currently wrestling with their legacy systems. It can integrate warehouses and lakes, all applications, and tools to create a playground for multiple user types. It removes complexity to future proof your data—all sources, all types, in one place.

A data operating system enables companies to take advantage of agile methodology despite challenges from legacy systems. Download our paper **“Data Lakes 101: Making the Most of Data Lakes through Agile Methods”** to find out how DataOS can increase speed and innovation while ensuring governance and security remain intact.

BY E. WALLACE



Transforming Retail Legacy Systems with a Data Operating System

© 2022 The Modern Data Company. All trademarks are properties of their respective owners.

The Modern Data Company
306 Cambridge Ave
Palo Alto, CA 94306
[TheModernDataCompany.com](https://www.TheModernDataCompany.com)
info@TMDC.io