Understanding the True Cost of Data Debt

Modern

Technology moves fast. Sometimes solutions to big challenges already exist, but more often, a problem appears before a solution. Companies must then take creative measures to "fix" technology challenges, leaving them with temporary solutions that quickly obsolesce. You can't blame companies for playing the cards they're given, but now data debt is costing companies more than they think, even when solutions seem to be working…for now.

What Causes Data Debt?

Data debt is similar to technical debt. It's the combined cost of continuous reworking and troubleshooting, where each new solution stage creates further challenges. Over time, teams spend more and more time trying to fix things instead of gathering value.

Data, unlike technology, seems simple on the surface. Companies gather and analyze data, use it to answer questions, and predict or understand behaviors. However, the data landscape is deceptively complicated.

Data debt is caused by:

- **Lack of comprehensive data strategy**: Companies have focused on point solutions to create a data ecosystem because there was no holistic solution in place that could integrate every data source. Maintaining a clear strategy for gathering, storing, and labeling data is challenging.
- **Limited resources**: Data management has always been resource intensive, but not all organizations can maintain a full data team. Without suitable resources for company-wide data management, it's easier to fall behind.

- **Inefficient data governance**: Companies relying on manual strategies to govern data access may not be able to make data available when it's needed most. In addition, making copies of data each time it's necessary for a query leaves potential vulnerabilities unaddressed.
- **Rapid growth**: When organizations grow quickly, managing the increased data volumes can become more difficult. In addition, pressure to get things done quickly can lead to patches and workarounds that can cause problems down the road.

Animesh Kundera, co-founder and CTO of The Modern Data Company, distills these ideas down to "the four horsemen of data debt." According to him, companies of all types and sizes grapple with the following:

- Complex architecture and no accountability
- A gap between raw data and insights for business teams
- Fragile data pipelines
- Delayed delivery with no sense of cohesion

As a result, companies that aren't technology and data forward are incapable of realizing the full value of their data. No matter how much data they collect, what tools

they buy, and what talent they manage to lure away from companies like Google and Facebook (if they can), data remains a source of powerful but hidden potential.

What Is Your Data Debt Costing You?

According to data expert John Ladley, there are four quadrants for where businesses are when it comes to accumulating and understanding their data debt:

- Data illiteracy (or immaturity): Those at the beginning stages of data thinking
- Resistance: Those without a long-term vision for data governance
- Realization: Those who are just beginning to understand the rising costs of data debt after the fact
- Acknowledgement: Those who acknowledge their existing data debt and plan to right the ship or have factored in those costs

But what are these costs? Let's look at a few scenarios.

Untrustworthy Data
One result of data debt is a lack of data quality due to ineffective governance. An organization may strive toward data-driven decision-making, but if it cannot trust its data, it won't succeed. This is particularly common at the data illiteracy tier because the company will continue to replicate errors in data with each successive use.

For example, imagine an organization wants to target a new market for services based on purchasing data from the area for the last three years. They aggregate their own data plus data from partners and task marketing to create personalized advertising campaigns.

Unfortunately, the marketing team doesn't have a complete picture of these potential customers because of inconsistent data. The marketing campaigns don't have the expected ROI, and the company loses some of this potential market share to a competitor.

Data Swamps
It's not just the business side that's suffering. It's the result of "vague data modeling and suboptimal storage mechanisms," according to Kundera. At the data realization stage, companies are highly susceptible to this cost because they're at risk of implementing fancy solutions with no real long-term strategy.

Data swamps are difficult to query and complicated to maintain. They emerge because enterprises are trying to manage data from multiple sources with no overarching plan and because of silos that make sharing between departments very difficult. It prevents collaboration between business teams and IT on coherent business models necessary for decision-making.

Data swamps cost a lot in the short term because business teams must wait a long time for answers to queries and for requested pipelines to be built. It's challenging to make data-driven decisions with any kind of timeliness. The long-term cost is a spiral toward greater chaos in the data swamp as IT teams become overwhelmed trying to monitor and maintain databases and existing pipelines while working on new ones. The longer the swamp persists, the worse it gets.

High Soft Costs
It's not just money that companies need to consider. Data debt also costs companies resources, time, and effort. When companies are data illiterate or resistant, they might attempt to follow governance policies but frequently make exceptions or ignore them entirely. This keeps processing costs and resources too high.

For example, imagine a company invests heavily in building an experienced data science team. However, that team discovers that fragile pipelines require constant reworking, putting most of this team's knowledge and expertise toward backward-looking tasks like recovery and troubleshooting.

They could be capable of innovation in strategic tasks that skyrocket the data's value, but that company will never know — as long as they refuse to acknowledge their data debt.

What to Do About Data Debt

Companies can't run from data debt forever; eventually, all debts come due. Companies need a way to integrate all data tools and sources into the data stack in a composable yet stable way. In addition, business users must have support to explore data through a governed, self-service portal and build stable pipelines for everyday decision-making.

Once this happens, IT teams can shift the majority of labor and resources from maintenance and troubleshooting tasks to higher-order activities. This allows companies to take advantage of IT expertise to build more complex data models that move the business needle forward.

DataOS is the world's first data operating system. It's designed with business users in mind to provide self-service dashboards and drag-and-drop engineering. Administrators can govern data access through attribute-based controls, and IT users can get behind the scenes to build the apps and tools the company needs for big data processing.

Find out how DataOS can help you chip away at your data debt.

Schedule a demo →

BY E. WALLACE