



# HOW TO AVOID DROWNING IN YOUR DATA LAKE

BY DARRYL DEMOS AND KAUSHIK DEKA

**D**ata and analytics are sparking innovation at all levels within banks, but they are also breathing new life into an age-old corporate war over ownership.

The data lake is often at the center of the debate. As a result, business leaders need a better understanding of its value and technology leaders need to better manage expectations to avoid

over-promising and under-delivering.

If unchecked, data lakes can turn into a money pit, dividing the business and technology teams that battle for control of data and analytics.

A data lake is enabled by low-cost storage for raw data in a “natural state”, creating the potential for high-speed data processing and a vision for democratized analytics. The full promise of the data lake is to enable predictive modeling and real-time analytics at scale, driven by a richer and granular data model. The models and insights can be deployed with an agile environment to produce results that weren’t previously possible with a traditional warehouse architecture and development methodology.

Structured properly, not only does the data lake provide a cheap and scalable data repository often hosted on Hadoop Distributed File System (HDFS), but it can also be embedded with a significant computational layer to support intense workloads like machine-learning models and streaming analytics. Companies like Netflix, for example, have been able to leverage Hadoop on the cloud to serve tens of millions of customers with streaming services.

Banks are rushing to dump files into their lake, telling users that they are opening the data for individual self-service.

What’s not to like? All data are in one place, easily accessible with promise of massive processing power, especially to handle customer analytics at scale.

Not so fast. Like many new technology investments, the vision of the future and the timing of reality are often out of whack. In the 1990s, customer relationship management (CRM) platforms were all the rage, with predictions that they would revolutionize the industry in year or two. But back-end problems and other stumbling blocks delayed the projects, some of which took a decade to reach their stated goals.

As a result, Novantas believes that business leaders need a better understanding of the value of a lake and its role in a data and analytical ecosystem. CDOs and CTOs, meanwhile, need to have a firmer grasp on the nuances of their technology selections and temper expectations to avoid over-promising

and under-delivering. And both groups need to be focused on use cases and agile development to make the right kind of short-versus-longterm trade-off.

Based on formal and informal surveys with business managers, CTOs and CDOs, Novantas has identified the most common data-lake pitfalls that can cause significant disenchantment if not addressed:

1. Data put into the lake are not as complete or well-documented as central data teams claim. These are holes that, discovered later, can be difficult to go back and plug. Serious business problems can develop, eroding faith in the promise of a new vision.
2. If poorly architected, processing capacity may initially appear to be slower than traditional warehouses. For banks that are spending millions of dollars to move from an enterprise data warehouse to an enterprise data lake, an under-utilized data lake can quickly sour expectations.
3. New skills are needed to work in a distributed computing environment like Hadoop and to develop analytics that can leverage its computing horsepower. The engineers who build data and machine-learning pipelines and the scientists who harness the data lake to develop predictive models have skills that are hard to find and may not have much banking domain expertise. Existing analysts may not be trained for the new technologies.
4. The people who want funding for the analytic arms race may make too many promises and heighten expectations, setting up everyone for failure. Skeptical business

people may be less willing to assert authority in the early stages because all the analytical promises sound great and the technology is so new.

One of the first ways to avoid such stumbling blocks is to make sure the two groups (business unit and technology leaders) are fully-informed at the outset about the potential of the data lake and the possible problems that may arise when onboarding the technology and filling the lake.

Technology leaders need to take the initiative. While money is needed to fund new initiatives, it does no one any good to over-promise. Successful initiatives can start with smaller, targeted programs that can grow in size and sophistication.

The data lake is a big investment, both in time and dollars. Be prepared for the reality that certain other businesses and functions may come first, delaying the benefits that a data lake can provide. On the flip side, a bank may need to scale back other investments to make way for the data lake.

Business leaders and technology experts must understand that a data lake is not an immediate panacea. Introduced correctly, it can be a mechanism that treats data as a strategic asset and adds significant operational leverage. But that will take time, talent and a well-defined technology roadmap aligned with clear business objectives. ■



*Darryl Demos*  
EVP, New York  
[ddemos@novantas.com](mailto:ddemos@novantas.com)



*Kaushik Deka*  
Director, New York  
[kdeka@novantas.com](mailto:kdeka@novantas.com)

