

DATA LAKES IN THE CLOUD

Fishing out competitive advantage and enhancing your ROI

Does your organization create oceans of data? Now consider how that's stored and curated; you can (hopefully) be certain it's stored, but is it being stored in a way that gives you an edge in business? That's an important question to ask; before getting to that let's take a step back.

Technology is ever moving and advancing; it's difficult to keep up with the now, let alone look ahead to where it's going. That said, it's important to understand how technology can help your business. Even though it can be confusing, especially in an analogy-heavy industry, there is a lot to be said for leveraging existing technology to give your enterprise a competitive advantage, and to prepare for future technology by streamlining your existing datasets. Before we dive into that (sorry), let's consider the history of data storage and creation and where we are today.

A brief history of data storage

Back in the olden days of 1980, the first one gigabyte hard drive was created by IBM; this (for the time) high-capacity drive weighed a quarter ton and cost a handsome \$40,000. Your smartphone is probably feeling pretty smug right now, as even a modest one will have at least 32x that storage and weigh in at less than a half pound. Even in this century, hearkening all the way back to 2005, a 500gb mechanical desktop hard drive was considered [enormous](#). Needless to say, data storage and creation capacity go hand in hand. Data storage capacity has gone parabolic, and so has data generation. The amount of information created has surpassed all forecasts, with over [40 trillion gigabytes of data](#) generated in 2020 alone.

Humanity generates 2.5 billion gigabytes of data each day. Businesses alone generate huge amounts of data both internally and externally. The variety of data has significantly changed over the last couple of years, especially with project lifecycle applications and business process tools. Data from current projects as well as old project data for legal and regulatory purposes results in copious amounts of information. One of the biggest challenges faced by organizations is storing the structured and unstructured data in a usable format where data-driven insights can be generated. As shocking as it sounds, [99.5% of the data collected is never used or analyzed](#).

The rise and rise of analytics

The idea that so much information goes unanalyzed is surprising and yet not at the same time. There's almost too much information generated to make sense of, but there is a way to find a signal in all that noise. Indeed, the science of studying data (Data Science) is a [field unto itself](#). This study of data has led to ways of analyzing data, ways of addressing data more efficiently, and big data structures. With advances in computing power, it has become easier to address even unsorted data to gain meaningful insight. In other words, rather than analysis paralysis, companies are digging into their own data and creating their own gold. [Analytics](#), or the scientific analysis of existing and ongoing data, makes it possible [to approach real time analysis and recommendations](#). Progressive companies are relying more and more on analytics to make decisions that enable them to gain a competitive advantage, all within a [wide range](#) of business applications.

Optimizing Apps & Infrastructure to Drive Business Value



Gathering data is important and ensuring that it's usable is even more important. Enabling you and your organization to make use of the data you create – that is hugely important. That said, without thinking about storage, organization, and addressing of data, making any sense out of what you've got is going to be a challenge. That's where different methods of data storage come into play.

Brace yourselves: The water analogies are coming

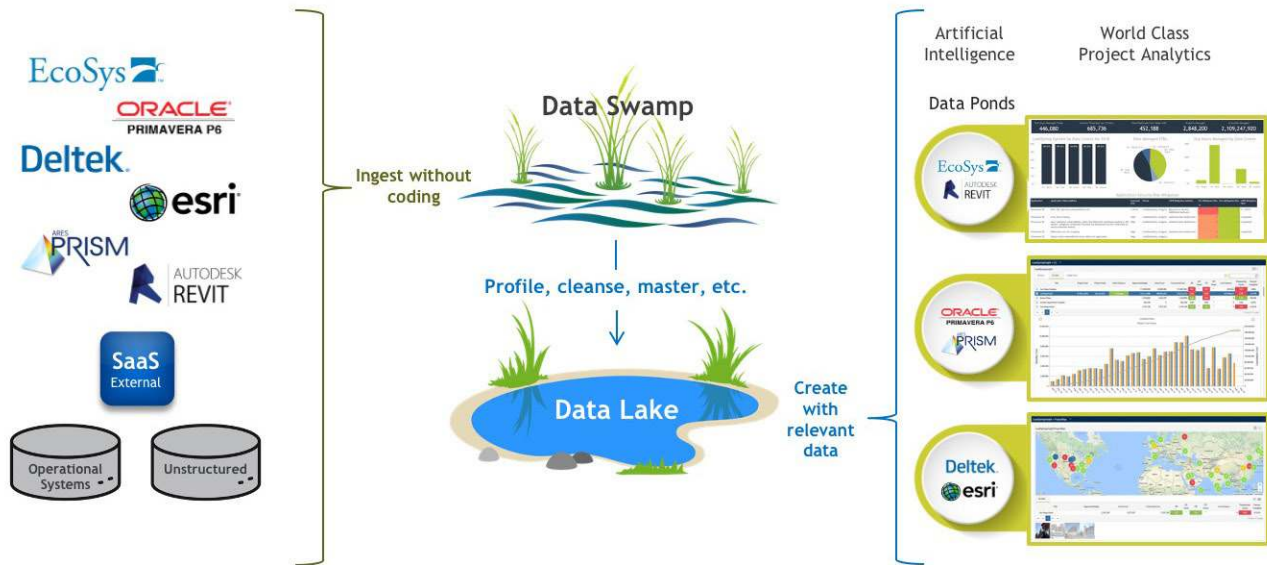
With such a flood of data (sorry!), it can be extremely difficult to get the right data in the right format so that executives and business analysts can generate actionable insights. Some data structures that are commonly utilized are data lakes and data ponds.

Data lakes

A data lake is a central repository that enables organizations to store pretty much any data type at any scale in its native format. It can be searched / queried when properly cleaned and configured. When not properly set up, the data lake is instead a data swamp; the data swamp is an unorganized, potentially unusable data dump that is best avoided.

Data ponds

Data ponds are subsets of data lakes that are raw data in its native format. Often intended to be added to/ integrated with a data lake later on.



Do it yourself data lake: What would it take?

To host your own big data structure (be it warehouse or lake or pond) will take considerable time and resources.

1. IT infrastructure

You need to determine the amount of on-premise hardware that will be required to create a sufficient data lake that can store all different types of data. If your organization or IT department doesn't specialize in this area of expertise, this initial hurdle could be a large one. This is where a partner like LoadSpring could come in handy...but on to the task at hand.

2. Integrate data from all sources

You need to integrate all organizational data from different sources and platforms such as clouds, SaaS-based applications, on-premise data sources and more.

3. Build complex data pipelines

You need to build complex data pipelines that are responsible for transmitting large amounts of data from multiple sources, which in turn is intended to be stored in a structured data lake.

4. Scalability

Scalability is a huge challenge for on-premise data lakes as you will need to manually configure your servers to support scalability (going bigger will cost more money).

5. Action planning

You need to determine what business intelligence tools you require to make actionable improvements that will provide the desired return on investment for your data lake. If you don't have a roadmap with well-defined goals, you may end up creating a data swamp with no active management, leaving stakeholders with unactionable data.

Why enterprises are turning to Cloud-based data lakes

Cloud-based data lakes are hosted in the cloud as a central repository with quick and easy access to all your data and project analysis. Cloud-based data lakes allow you to scale as-you-go and leverage the BI tools and analytics offered by your cloud solution provider.

Cloud-based data lakes have several key benefits over on-premise data lakes including:

1. Availability

Cloud data lakes have up to 99.99% uptime which keeps teams productive and avoid the nightmare of potential data loss or security exposure.

2. Speed

Cloud data lakes are constantly upgraded and maintained with the latest hardware technology and the highest performance for critical project apps.

3. Agility and scaling

Businesses can automatically and immediately scale their capacity up or down to meet demands, improving their ability to adapt to changing landscapes.

4. Trusted data security

Cloud solution providers modernize your cloud security strategy and keep your sensitive data safe. It also has a business strategy component by limiting access to information based on job roles ensured by a stringent authentication process.

Your dream lake hosted by your dream Cloud Partner

To achieve the highest state of well-being, you need to look higher. Choose the right Cloud Partner to build your custom Cloud Data storage; you'll want one that:

1. Starts with the end in mind

We will help you leverage the "Art of Possibility" to define the desired insights and analytics you'd like to have to drive your business decision support process.

2. Identifies data sources

With your help, we'll derive from the desired business intelligence what applications, data sets, and sources are required to support the same. This can be existing on-premise applications, SaaS offerings, mobile apps, contractor apps, and LoadSpring-hosted applications. We will then evaluate the location and type of data and best "go-forward" strategy to collate into a common data lake based on your defined business requirements.

3. Sets up your data lake solution

We can ensure that your storage needs are safe, secure, and ready for you to configure. LoadSpring helps reduce the headaches that are associated with wrangling huge amounts of data and enables you to reach into that data and pull out KPIs that will improve your ROI.

We can help you navigate

A cloud data lake in the care of the right Cloud solution provider empowers organizations to manage storage, processing, big data handling, and analytics while maintaining the security, integrity, data regulations, compliance, and governance standards that are required to deal with big data.

Finding the right partner with the combined technology and experience in Cloud project management and solutions is critical for building the best data lake for enterprises.

Now, Cloud-based data lakes and ponds are the most innovative way to access, visualize, and analyze big data, anytime and anywhere. LoadSpring's experience and expertise can help put you in the analytics driver's seat.

Learn more

Regardless of the difficulty or complexity, we've got you covered. Our project intelligence systems enable "data lake" critical, real-time KPI's needed for faster and accurate decision making and improved ROI.

If you're interested in finding out more about Cloud-based data lakes, or if you're actively looking to build out your Cloud enabled data structure, give us a call at +1 978.685.9715! Or, you can fill out this [form](#).



+1.978.685.9715 www.loadspring.com