# Inspire

# The Value of Data Lakes in Healthcare IT

By Michael Gill

As the number of healthcare data sources continues to grow, and the variety of data formats and delivery methods continues to expand, a robust data management strategy is now more important than ever. By leveraging a data lake, you can empower your organization to tackle these challenges in a simple an organized manner.

## What is a Data Lake?

A data lake is a repository that holds a vast amount of raw data in its native (structured or unstructured) format until the data is needed. Storing data in its native format enables you to accommodate any future schema requirements or design changes.

A data lake is ideal when valuable data sources are dispersed among on-premises data centers, software providers, partners, third-party data providers, or public datasets. A data lake offers a foundation for storing on-premises, third-party, and public datasets at low prices and high performance.

Additionally, an array of descriptive, predictive, and realtime agile analytics built on this foundation can help meet a company's most important business needs, such as forecasting service delivery and service utilization patterns, evaluating effectiveness and cost of service delivery, and analyzing financial performance comparisons against estimated expenses to support federal and/or state reporting.

## How is a Data Lake different than a Data Warehouse?

It's been said that a data warehouse is like "a store of bottled water - cleansed and packaged and structured for easy consumption - the data lake is a large body of water in a more natural state. The contents of the data lake stream in from a source to fill the lake, and various users of the lake can come to examine, dive in, or take samples."

That's why many healthcare organizations are shifting to a data lake architecture. A data lake is an architectural approach that allows you to store massive amounts of data into a central location, so that it's readily available to be categorized, processed, analyzed and consumed by diverse groups within an organization. Since data can be stored as-is, there is no need to convert it to a predefined schema and you no longer need to know what questions you want to ask of your data beforehand.

A data lake can have all kinds of data, structured or unstructured, and offers the agility to reconfigure the underlying schema on the fly. The same cannot be said for data warehouses. Additionally, the raw data stored in data lakes is never lost - it is stored in its original format for further analytics and processing.

A data lake can actually complement and extend your existing data warehouse. If you're already using a data warehouse, or are looking to implement one, a data lake can be used as a source for both structured and unstructured data.

## Understanding the functions of a Data Lake

1. **Data Submission**
   The first function of a data lake is to receive large amounts of data. This data can be submitted to the data lake as either batch uploads or streaming data.

2. **Data Processing**
   The data lake then validates the data, adds any necessary metadata, and indexes the data accordingly.

3. **Data Management**
   Next, the data is transformed and aggregated into a format that is ready to be stored long term, and made available to the analytics applications.

4. **Searching**
   The data lake catalogs the indexed data and provides a query tool for interacting with the data in a read-only manner.

5. **Publishing**
   The final function of a data lake is to publish the analyzed data in meaningful charts and graphs via a BI tool or data visualizer.

## What the emergence of Data Lakes means for Healthcare IT

By 2015, 96% of non-federal acute care hospitals had adopted electronic health records, compared to only 71.9% in 2011. As the Healthcare industry has transitioned into a world of fully-electronic records, including fee-for-service and value-based care, the amount of data being accumulated has skyrocketed. This data comes in many forms, but can primarily be categorized as either clinical data, or claims data. Claims data coming from the payers is typically consistent and well-structured, whereas clinical data coming from providers may come in many formats.

A scalable data lake can accommodate massive amounts of data in its raw form. With the help of strategic analytic applications and actionable insights from actuaries and data scientists, this data can be easily consumed and leveraged by providers and payers in many ways, including:

- Effectively transitioning to value-based care
- Projecting and planning for growth
- Improving cost effectiveness
- Reducing readmissions and improving quality of care

As the amount of available data continues to grow in the healthcare industry, a flexible, cost-effective solution such as a data lake becomes a major difference-maker in positioning a healthcare organization for success.

## About the Author:

Michael Gill is a Partner and Chief Technology Officer at Axene Health Partners, LLC and is based in AHP's Temecula, CA office. Michael can be reached at (951) 239-3070 or michael.gill@axenehp.com.

Any views or opinions presented in this article are solely those of the author and do not necessarily represent those of the company. AHP accepts no liability for the content of this article, or for the consequences of any actions taken on the basis of the information provided, unless that information is subsequently confirmed in writing.



# axene health partners
## HEALTH ACTUARIES & CONSULTANTS