



Don't Get Bogged Down: How to Keep Your Data Lake from Turning into a Data Swamp

By Anand Kumar

What is the key to driving digital innovation with data analytics—and where should healthcare leaders begin?

It's often said that "data is the new oil."

If that's true—and many think it is—life sciences organizations have a growing wealth and variety of the commodity at their disposal to advance the research and analytics essential to their success. No longer dependent on structured data like text, life sciences organizations are increasingly making use of "poly-structured" data like images, audio and video files, email, sensor data, and even social media data to aid development of new tests, drugs, and other innovative healthcare weaponry.

Indeed, the sheer amount and kinds of data being collected, stored and used by the life sciences fields are exploding thanks to the elasticity of the cloud, where nearly infinite levels of data from nearly limitless sources can be stored in so-called "data lakes."

The side effect of data lakes, however, is that they greatly increase the burden on researchers and data scientists. Unlike data warehouses, which store mostly structured information readily available for research and analysis, data lakes contain raw, unprocessed data that must be curated, cleaned, prepared and transformed (refined, to put it another way) into forms useful for research and analysis.

Much to the chagrin of data scientists, this process is tedious and time consuming. A 2017 study found that data scientists spend nearly half their time cleaning and preparing data before they can begin working with it. Not surprisingly, the same study found that data scientists find these aspects of their job the most onerous.

“ A 2017 study found that data scientists spend nearly half their time cleaning and preparing data before they can begin working with it. ”

DataEz: An End-to-End Solution

Fortunately, there's a solution: DataEz from Healthcare Triangle, a robust, end-to-end, software-as-a service(SaaS) platform that automates the curating, cleaning, preparation, and transformation of poly-structured data. In doing so, it reduces a process that can take months down to weeks at much lower cost than current practices.

DataEz even fully self-catalogues all data and metadata for faster, easier access by current and future researchers. This is especially critical for life sciences organizations, as it enables researchers to access all data, ranging from entire studies to one-minute snippets of audio or video. Indeed, without fast and easy access to vital data, data lakes can essentially turn into data swamps, bogging down research and analytics teams.

As a SaaS platform that can be up and running within half a day, DataEz is cost effective; it eliminates the considerable expense of self-implementation, feeding and operation of data lakes. The platform also remains current with the offerings of major cloud vendors, providing on-demand the latest tools for better and faster research and analytics. For those organizations that want to operate and retain control over their own data lakes, DataEz can be installed and implemented on site in as little as eight weeks.

Moreover, together with CloudEz from Healthcare Triangle, our cloud-based solution for comprehensive data management and security, the DataEz platform is an affordable solution for life science enterprises of all sizes—billion-dollar companies, small- and medium-sized businesses, and start-ups—to maximize their use of the cloud.

At a time when speed and efficiency in data collection and processing are critical to success and competition in the life sciences, a platform like DataEz takes a vital but important process off the shoulders of data scientists so they can put their focus where it truly belongs: on research and analytics.

You can almost hear them breathe a sigh of relief.



HEALTHCARE
TRIANGLE

Reinforcing Healthcare Progress™

Request an
Assessment

info@healthcaretriangle.com
(888) 706-0310



Anand Kumar
Senior Vice president, Healthcare Triangle.

Healthcare Triangle, Inc.™ (HCTI), based in Pleasanton, Calif., reinforces healthcare progress through breakthrough technology. HCTI achieves HITRUST Certification for Cloud and Data Platform (CaDP) to manage risks. We support healthcare and life sciences organizations improve health outcomes by enabling the adoption of new technologies, data enlightenment, business agility, and accelerating the value of their IT investments. HC/LS turn to HCTI for expertise in cloud transformation, security and compliance, data lifecycle management, and clinical/business performance optimization.

©2022 Healthcare Triangle Inc.

All rights reserved. All other registered trademarks or trademarks are property of their respective owners.
DataEZ