

DATA ANALYTICS SPECIAL

FEB - 23 - 2018 CIOREVIEW.COM

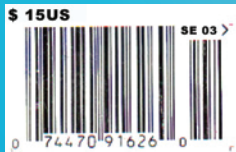
CIO Review

The Navigator for Enterprise Solutions

CATALYTIC
DATA SCIENCE

SPEARHEADING
INNOVATION IN
LIFE SCIENCES

Scott Sacane
Founder & CEO



CIO REVIEW
44790, S Grimmer Blvd.
#202, Fremont, CA-94538

COVER STORY



Scott Sacane
Founder & CEO

CATALYTIC DATA SCIENCE

SPEARHEADING INNOVATION IN LIFE SCIENCES

By Syed Haseeb Ahmed

Growing up as a sports enthusiast, Scott Sacane, a molecular biology scientist, was smitten by the way a player-coach could help their team achieve success. He believes that being in the game while advising the players in the moment creates a unique perspective that often leads to successful outcomes. Donning the role of a player-coach at Catalytic Data Science, Sacane brings the same enthusiasm, commitment, and the positive attitude to transform how life sciences companies find, analyze and share information. “I certainly don’t have all the answers and could never pretend to be an expert in the engineering domain, but I have operated in the life sciences industry for three decades, and I know the problems well,” he says. With 30 years of rich experience in the life sciences industry, Sacane is clear on what he wants to achieve with his venture. “At Catalytic Data Science, we want to connect researchers to the vast world of

life sciences data, information and analytics through a single, highly integrated cloud-based platform.”

Let’s face it: the life sciences industry is at a fairly nascent stage when it comes to exploring and discovering the enormous world of data with analytics. “Data is everywhere in the life sciences industry, and researchers are struggling to make sense of it all,” says Sacane. In this scenario, Catalytic Data Science is setting a new paradigm through continued innovation in the construction and implementation of cutting edge information technologies that enable researchers and life sciences companies to more efficiently derive unique insights from all this data. A typical life science researcher often needs access to a broad range of digital resources, from scientific publications and patents to research protocols, regulatory documents, different types of research data, analytic applications to process data, workflows, study reports, and shared resources that are a part of collaboration and conversations. In most cases today, all of these resources live in different, unconnected places that are hard to find and nearly impossible to use at scale. The challenges in sharing research data and related information between

applications and across individuals leads to time inefficiency, and more importantly, loss of key information. “However, modern technology stacks—purpose-built for the life sciences—can fundamentally transform the way data is turned into knowledge and the way it is monetized,” says Sacane.

Catalytic Data Science is a silver bullet that brings all of the digital resources that a life sciences company needs into one integrated, interoperable and secure environment. The company has devised an integrated knowledge management and informatics platform that congregates digital assets, workflows, communication tools and people, enabling researchers to complete a task, advance a workflow or make a more informed decision in a rapid fashion. The Catalytic platform assists researchers to not only perform their job but also instantly communicate, share information, answer questions and collaborate with colleagues. With such capabilities, the Catalytic platform streamlines the research workflow, fueling life sciences innovation and drug discovery that can save countless lives.

USHERING LIFE SCIENCES R&D IN THE CONTAINERIZATION ERA

“Scalable, cloud-based computing environments, microservices-based architectures, container images, orchestration strategies and the application of machine learning algorithms facilitate construction of new data consumption platforms that can enhance the efficiency and productivity of life sciences companies,” explains Sacane. At the intersection of innovation in life sciences and information technology, Catalytic’s ‘Corporate Platform’—built on AWS virtual private cloud (VPC)—is designed to converge the parallel innovations and provide value to organizations in ways not possible before. Now, the team can work across diverse datasets and avail the analytics application of choice for generating key insights into experimental data. The platform performs the complex task of ingesting, processing, normalizing and in-



Modern technology stacks, purpose-built for the life sciences, can fundamentally transform the way data is turned into knowledge and the way it is monetized



Scott Sacane
Founder & CEO

dexing data so that diverse data types are findable, interoperable and readily usable when researchers need it.

Catalytic leverages containers to help life sciences companies avoid the need for building new research environments for different types of analytics. By deploying necessary analytic algorithms including their dependencies in a single software container, Catalytic empowers the life sciences industry to flexibly use data analytics. Utilizing Docker platform, they help clients to build highly scalable and dynamic infrastructures. Coupled with the power of the cloud, Catalytic empowers life sciences companies to build complex infrastructure running container-orchestrated systems that can handle ever-increasing amounts of data and an ever-growing diversity of analytics. The adoption of microservices architecture helps life sciences companies to leverage different types of analytics that they are comfortable running today in a plug and play fashion. In addition, scientists can explore new ground by pipelining different analytic applications in ways not possible before. “With the Catalytic platform, life sciences companies can build custom architecture and compose the workflows that fit their research needs accordingly,” adds Sacane.

For example, one of Catalytic’s early customers from the life sciences industry—a company that was generating large genomic datasets—was struggling to leverage the datasets for multiple uses. The client had a hard time running different analytics effectively on a genomic data set to identify variants. After having several brainstorming sessions with scientists and researchers working in the client’s environment, Catalytic understood the different data types coming from the genomic sequencers and built an ingestion pipeline facilitating the flow of data into their VPC. Soon after, the R&D team at Catalytic was able to provide the client with the analytics of choice on their cloud platform. Once the client’s research team logged into the system the next morning, they had access to all the data they were

The screenshot shows the Catalytic search results page for the query "gene upregulation pancreatic ductal adenocarcinoma". The interface includes a search bar at the top with the query entered, a "SAVE AS WORKFLOW" button, and navigation tabs for Dashboard, Library, Folders, Analytics, Workflows, Article, Search Results, and Org Datasets. Below the search bar, there are filters for "year published" (any time, since 2017, since 2016, since 2013, custom range...) and "gene" (kdn: 557, ac: 473, akt1: 452, tp53: 449, more...). The search results are displayed in a table with columns for Title, Author, Year, Journal, and Library. The first result is "Pancreatic Cancer, A Mis-interpret of the Epigenetic Language" by Iguchi Eriko, Safgren... (2016, Yale J Biol Med, Open Access). The second result is "Overexpression of miR-221 mimics in PDAC cells caused an increase in migration and an increase in MMP2 and MMP9 gene expression [157]. miR-221 was also identified in a meta-analysis of miRNAs in PDAC as upregulated in seven studies with an average fold change of 6.7 [158]." by Wang Jun, Dumartin... (2017, Sci Rep, Open Access). The third result is "Splice variants as novel targets in pancreatic ductal adenocarcinoma" by Wang Jun, Dumartin... (2017, Sci Rep, Open Access). The fourth result is "Inhibition of renalase expression and signaling has antitumor activity in pancreatic cancer" by Guo Xiaojia, Hollande... (2016, Sci Rep, Open Access). The fifth result is "RNLS gene expression was increased in pancreatic ductal adenocarcinoma cell (PDAC) lines with KRAS2 mutations (MlaPaCa2 and Panc1) compared to those with wild type KRAS2, such as BxPC3 (Supplement Fig. 35)." by Liu Pingping, Yang Ha... (2017, Sci Rep, Open Access). The sixth result is "The lncRNA MALAT1 acts as a competing endogenous RNA to regulate KRAS expression by sponging miR-217 in pancreatic ductal adenocarcinoma" by Liu Pingping, Yang Ha... (2017, Sci Rep, Open Access). The seventh result is "Several studies have reported that MALAT1 displays increased expression in a variety of tumours 17-21, 41, in this study, we mined publically available". The interface also shows "30 Rows | Showing the top 1000 of 1238 results" and "Page 1 of 34".

looking for—to analyze, reanalyze, and repurpose according to their needs.

“We employ domain experts from the life sciences industry and highly talented software engineers—both residing under the same roof—and maintain constant communication between them. This approach allows us to transform customer’s request into a comprehensive R&D solution quickly,” says Sacane. Aware of the mission-critical nature of life sciences data, Catalytic also has a Chief Information Security Officer (CISO) in place—an expert in handling life sciences data—who ensures enterprise-grade security and data encryption while in transit and at rest.

STREAMLINING LIFE SCIENCES COLLABORATION

In the life sciences industry, collaborations can be complex, worth hundreds of millions of dollars and spanning many years, demanding an IT infrastructure that serves the terms and conditions outlined in the collaboration agreements. With Catalytic’s Alliance Platform, life sciences companies and research agencies can rapidly spin-up a dedicated and secure environment configured to serve the collaboration’s need from an R&D, data ownership and governance perspective.


Within hours of being notified, Catalytic will create an ‘Alliance

Platform’ and appoint administrators at the respective companies to set authorization, permission, auditing and archiving policies that fit their demands. Employees at the respective companies are onboarded, and the Alliance Platform is pre-configured to hold the content, data, analytics and collaborative work that will be generated. The highly scalable platform proves to be an effective tool to cut costs and meet compliance complexities before, during, and after the collaboration. When the collaboration comes to an end, Catalytic disseminates the content, data, and other resources to each partner as instructed and winds down the infrastructure. Because none of the resources ever lived on either company’s infrastructure, it’s as clean and easy to wind down as it is to spin up.

With the first release of their Catalytic platform in 2017, Catalytic Data Science has integrated with Box Inc., a file sharing and content management company. With this integration, Catalytic provides search and access to not just scientific data but also other critical enterprise documents, spreadsheets, and any other file format that Box handles today. Being able to find these files from a single search interface and loading these files into the Catalytic platform improves a number of important company workflows. Firmly believing in the



We employ domain experts from the life sciences industry and highly talented software engineers—both residing under the same roof—and maintain constant communication between them

value offered by container systems and microservices architecture, the company will continue to build momentum and offer more standalone analytics as container images. “Pipelining of analytic services will begin to emerge as a powerful way to create custom workflows by enabling researchers to combine and automate multiple analytic applications,” says Sacane. Catalytic Data Science ensures these trends are manifested as part of their portfolio of offerings for the industry. All in all, the company is well positioned to help life sciences companies improve the outcomes of their research and innovation and take the right step toward enriching lives. 

CIOReview

The Navigator for Enterprise Solutions

DATA ANALYTICS SPECIAL

FEB - 23 - 2018

CIOREVIEW.COM

20 Most Promising Data Analytics Solution Providers - 2018

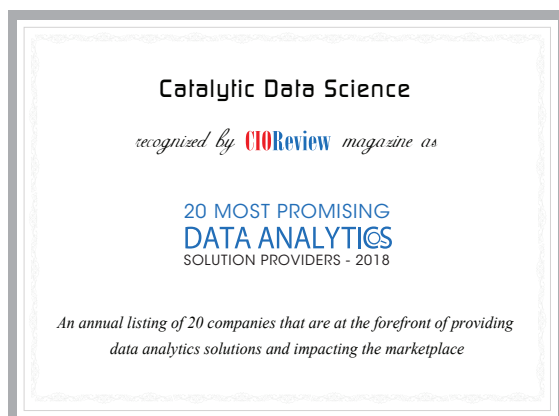
Companies are often coming up with creative ways to enhance customer experience. Thus, adaptation to the staggering amount of data accumulation has provided the academic and the economy with unprecedented knowledge of population patterns. But the data accretion is vast in its quantitative measures and is susceptible to duplications, glitches, and factual errors. Additionally, on the qualitative front, the talent required for data clearing is scarce in their availability along with the upsurge in the use of storage space. Security is another major issue for data mining companies as big data is often vulnerable to cyber breaches, ethical hacking, and malware attacks that lead to loss of trust in data.

As 2018 commences, companies are far more motivated to adopt ambitious analytics trends to amplify the worth of web experience in day-to-day lives. Small and medium-sized companies are increasingly more enthusiastic about reaping the benefits of data analytics and are willing to outsource

their analytics requirement to data companies specializing in it. Down the line, this year EU General Data Protection Regulation will come into effect and open up new opportunities for data scientists to explore.

In conclusion, the data analytics landscape is flooded with myriad solutions and thus zeroing in on the apt one remains an uphill task for a CIO. To tread on the right path, CIOReview Magazine has become the torch bearer in the data analytics landscape. Our distinguished selection panel, comprising CEOs, CIOs, VCs, industry analysts and the editorial board of CIOReview Magazine narrowed the excellent providers that exhibit competence in delivering data analytics solutions.

We have considered the vendor's ability in building solutions and services that can effectively yet economically account for advanced and effective data analytics offerings, keeping in mind the factor of time-focused delivery. We present to you CIOReview's "20 Most Promising Data Analytics Solution Providers - 2018."



Company:
Catalytic Data Science

Description:
Integrated knowledge management and informatics platform that uses configurable architecture and composable workflows for life sciences innovation

Key Person:
Scott Sacane
Founder & CEO

Website:
catalyticds.com