

# Data Science for Biotech:

How Related Sciences  
Achieved Breakthroughs  
in Drug Discovery

**VERY**

# Revolutionizing Drug Development

If you've ever been prescribed a new medication during a doctor's visit, it's possible you were able to pick it up from the pharmacy that very same day. Perhaps you've even had medication delivered right to your doorstep—the pinnacle of the patient experience in the age of modern medicine.

On the surface, getting new drugs into patients' hands seems as quick and easy as ordering takeout. But what patients don't see is the long, slow journey that those medications undergo to arrive at their doorstep in the first place.

According to McKinsey research published last year, it takes an average of 12 years for a new drug to hit the market, from candidate nomination to launch. This languid and agonizing process signals an urgent need for a revolutionized drug development approach to accelerate preclinical development, trial processes, and discovery.

That's exactly what Related Sciences is here to do.



## Client Background

As a leader in the biotech sector, Related Sciences has been at the forefront of investment and discovery of valuable new medicines since 2019. Leveraging innovations from the past 25+ years in biology, engineering, and machine learning, Related Sciences is building a scalable, data-driven AI/ML (Artificial Intelligence & Machine Learning) platform. The aim is to translate promising basic science into valuable drug candidates with optimal safety and efficacy profiles for patients.

At the core of Related Sciences' operations is its proprietary AI/ML platform, which systematically identifies the most promising new drug discovery opportunities from millions of possibilities. Coupled with a unique decentralized R&D model, Related Sciences brings together data scientists from around the world to collaborate on these opportunities. This approach allows the company to efficiently invest in an evergreen internal pipeline of new medicines with exceptional clinical and economic promise.

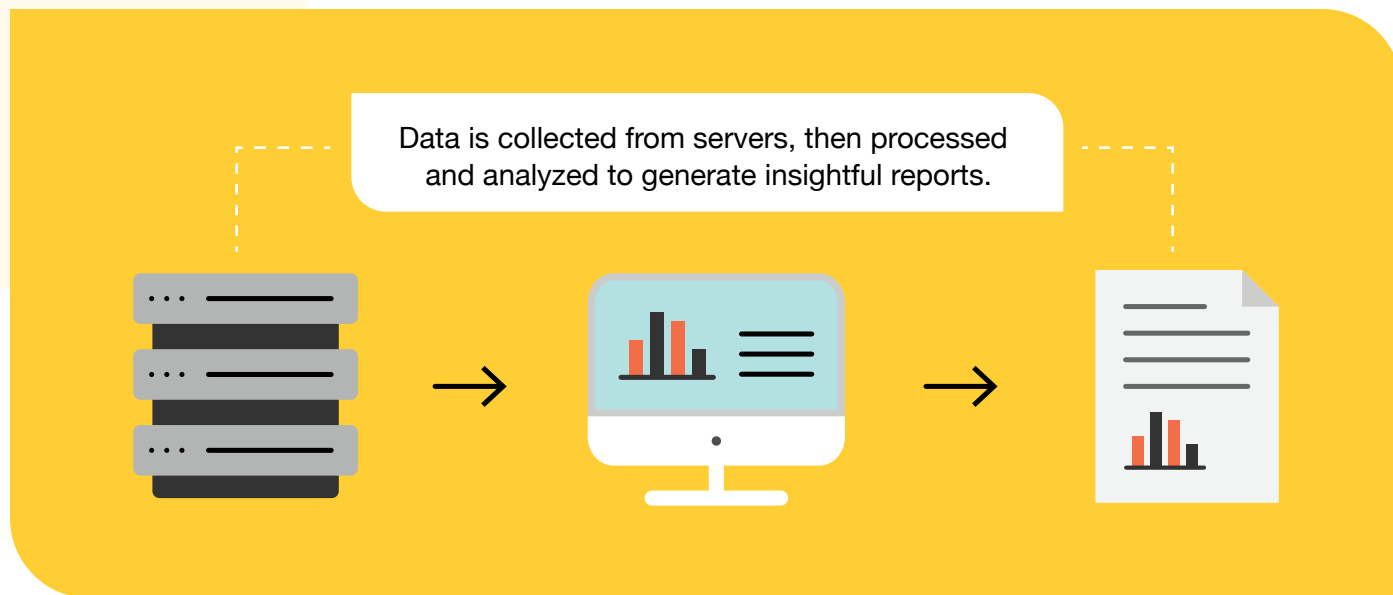
# The Problem

The challenge faced by Eric Czech, Head of Data Science at Related Sciences, was evident: selecting the most promising opportunities for creating new medicines. Given the high costs and time-intensive nature of research and development in this field, there was a pressing need to enhance decision-making processes using data-driven approaches.

In order to innovate and drive drug discovery forward, the team at Related Sciences recognized the critical need to access top-notch engineers to enhance its AI/ML platform, Facets. However, they faced another hurdle: accessing and effectively managing an external set of high-quality senior engineers proved challenging. To overcome this obstacle, Related Sciences sought a strategic partnership that could provide a curated pool of talent and managed experience working with engineers with diverse backgrounds.

Related Sciences faced challenges in accessing and effectively managing an external supplementary set of senior engineers with AI/ML experience capable of working seamlessly with their internal team.

Understanding the necessity for a partner with established expertise in Data Science and Machine Learning to develop this new solution, Related Sciences turned to Very. Together, we collaboratively addressed these challenges, developing a data-driven program to improve the likelihood of success in identifying successful disease treatments.



# Why Very?

Amidst the search for engineering talent, Related Sciences partnered with Very for its expertise in delivering high-quality outputs and effectively integrating with internal teams. Recognizing the need for a curated solution, we rose to the challenge, quickly adapting to the unique requirements of Related Sciences. By providing valuable code and seamlessly integrating with Related Sciences' internal staff, we demonstrated our capability to address tough challenges and deliver results within a short period of ramp-up.

While Related Sciences specializes in scientific products, particularly in domains like genetics, life sciences, and biomedicine, we don't claim to be experts in all those fields. However, our ability to collaborate effectively with domain experts and rapidly acquire the necessary knowledge sets us apart. Leveraging this adaptability and our solid foundation in the full data science stack, we were able to contribute meaningfully and adapt quickly to the industry's highly specialized and technical nature.

## Choosing Very

Our team's proven expertise and experience stood out from other vendors due to our dedicated data science practice and our proven ability to meet the high expectations of our clients.

**We needed a team that could deliver - and excel - in the full stack of data solutions, from data engineering to machine learning and data visualization. The Very team had all of the capabilities we required, and were able to seamlessly deliver on every task we threw their way, regardless of what it was.**



**Eric Czech**  
Partner, Head of Data  
Science, Related Sciences.

# Plan of Action

Our team implemented an end-to-end data science product, encompassing data ingestion, processing, and storage, as well as algorithm development and evaluation. We focused on proper versioning, scalability, and reliability throughout the entire data science stack.

1



## End-to-End Data Science Stack

To address Related Sciences' need for a data-driven program to improve the identification of solutions for various illnesses, our team developed a full end-to-end data science stack. This comprehensive approach involved every aspect of the data lifecycle, from data ingestion through ETL processes to data processing and storage on the Google Cloud Platform (GCP) using tools like BigQuery. This ensured scalability and efficiency in handling large volumes of biomedical data. Additionally, we paid meticulous attention to data wrangling, modeling, and model evaluation to guarantee the robustness and efficacy of our predictive models.

2



## Modeling and Model Refinement

Our team actively contributed to the development and optimization of machine learning and ranking algorithms that form the core of Related Sciences' offerings. This involved refining existing models, enhancing model evaluation processes, and aligning models with the specific requirements of Related Sciences' biomedical research.

3



## Data Pipeline and Versioning

A critical component of the project was establishing a robust data pipeline and versioning system. We utilized Spark for ETL pipelines, facilitating seamless data flow and ensuring consistency, scalability, and integrity throughout the project lifecycle. We used a variety of additional tools for tracking changes, version control, and managing large datasets, which are essential for maintaining the reliability and accuracy of the predictive models we developed.

4



## Comprehensive Data Science Services

In addition to the above data science tasks, our team provided a range of comprehensive services tailored to the needs of Related Sciences. This included conducting ad hoc analyses, offering assistance with dashboarding, and optimizing queries for enhanced performance. By offering a holistic suite of services, we ensured that Related Sciences had access to the necessary support and expertise at every stage of the project, which allowed for continuous refinement and optimization, ensuring that the solutions provided were always aligned with their evolving needs and goals.

5



## Advanced Technologies and Practices

Our data science team leveraged cutting-edge tools and practices to address the challenges faced by Related Sciences. This included utilizing advanced technologies like GPT and large language models (LLM) for natural language processing, enabling more sophisticated analysis of biomedical data. By staying at the forefront of technological innovation, we demonstrated our commitment to delivering solutions that were effective, forward-thinking, and adaptable to emerging trends in the biotech sector.

6



## Focus on Best Practices

Throughout the project, we maintained a steadfast focus on adhering to best practices in data science and engineering. This included rigorous quality assurance processes, adherence to industry standards, and continuous optimization of workflows and methodologies. By prioritizing best practices, we ensured that the solutions delivered to Related Sciences met the highest quality, efficiency, and reliability standards.

# Results

Together, our team and Related Sciences achieved significant milestones and delivered impactful outcomes. Over the course of six months, our team successfully enhanced one of the core ranking abilities by 62%, marking a notable advancement in Related Sciences' capabilities. This milestone allows Related Sciences to more confidently prioritize drug ranking for its own trials.

## Data: The Backbone of IoT

Data isn't just a byproduct of innovation; it's the core product driving technological advancements and decision-making. Whether it's curing diseases or optimizing IoT devices, data serves as the foundation for solving complex problems – it's the what that delivers the why.

With the depth and breadth of our expertise, we ensured that Related Sciences was set up for success from the beginning, just as we do with all of our clients. Managing and organizing data effectively is crucial to avoid pitfalls like skyrocketing cloud costs and ensure the success of business cases. Ultimately, expertise in data engineering is essential for extracting value from IoT projects, where data wrangling forms the foundation for informed decision-making and driving meaningful outcomes.

## Project Team

Building a platform that harnesses data insights to drive strategic decisions demanded meticulous planning, expertise in data science and AI methodologies, and effective project management throughout the process.

With this in mind, we enlisted its team of experts, which included:

- **1 Senior Technical Program Manager:** Responsible for coordinating project activities and ensuring alignment with stakeholder objectives.
- **1 Senior Data Engineer:** Responsible for implementing the data ingestion and pipelining.
- **1 Senior Data Scientist:** Responsible for the design of robust modeling and analytical frameworks.

## Data Science and IoT:

Data science expertise is essential for unlocking value across IoT projects as data forms the bedrock for solving complex challenges and driving meaningful outcomes.

## Last Thoughts

Very isn't just another technology provider; we're a trusted partner committed to driving innovation and delivering transformative outcomes across industries. Whether you're revolutionizing healthcare, optimizing manufacturing processes, or pioneering new solutions in the IoT space, we're here to help you achieve your goals.

If you're ready to embrace the power of your data with data science and need a dedicated partner to help you unlock the full potential of advanced technologies like AI, predictive modeling, and more, [get in touch](#). Let's bring your data-driven vision into reality together.

[BOOK A MEETING](#)