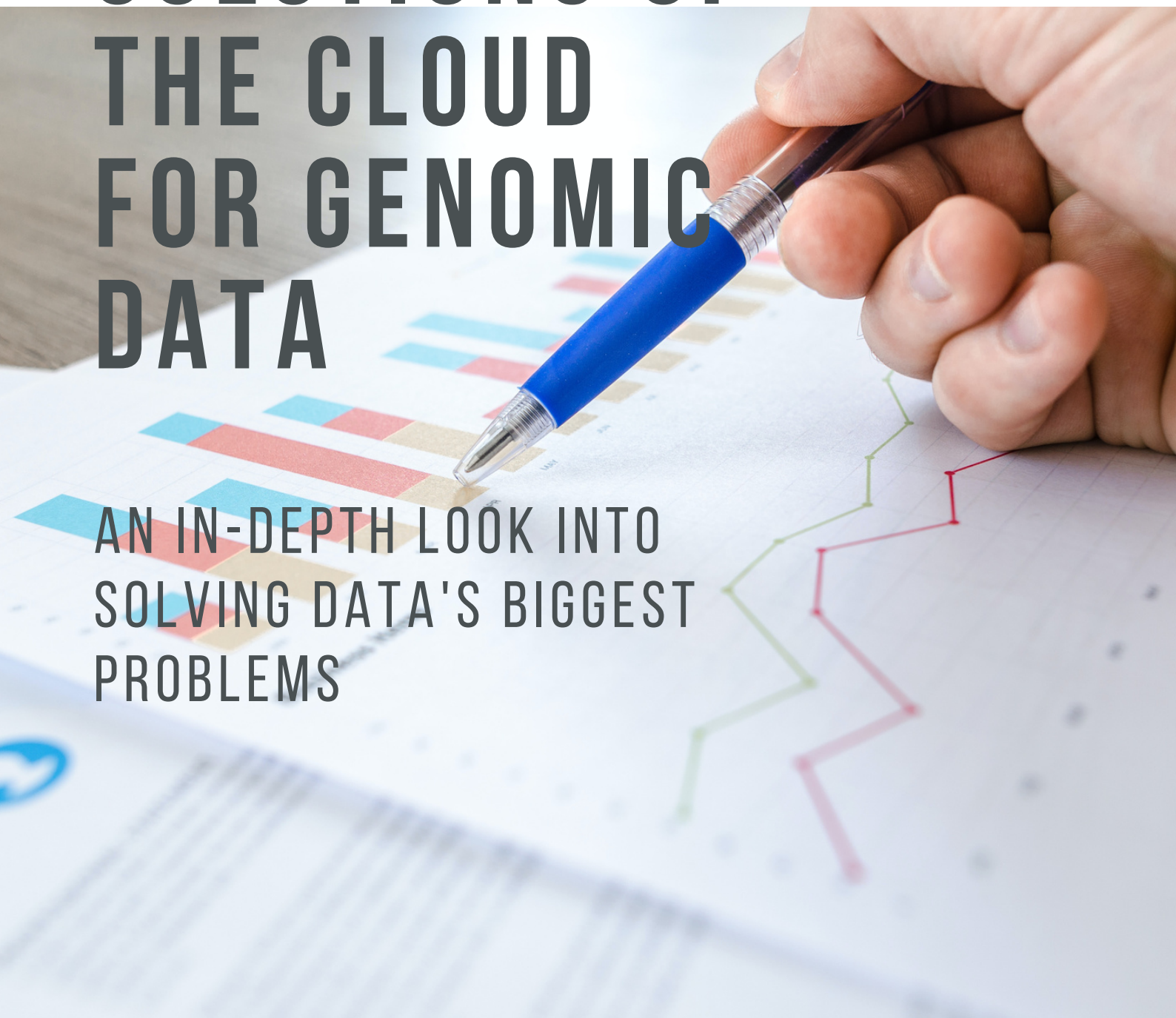
THE CHALLENGES AND SOLUTIONS OF THE CLOUD FOR GENOMIC DATA



AN IN-DEPTH LOOK INTO SOLVING DATA'S BIGGEST PROBLEMS

01

INTRODUCTION

Perhaps the greatest quest completed by humanity was the Human Genome Project. From 1990 to 2003, a dedicated team of international researchers mapped out the grand design of homo sapiens. This initial project opened the frontier of next-generation sequencing, which revolutionized genomic research. Advances in genomics are growing at an unprecedented rate.

These advances have led to a more personalized approach to medical care. Medications can be tailored to individuals based on that person's response to specific drugs. The secrets of genetic disorders can be unlocked, treated, and in some cases, even prevented.

However, as genomics continues to progress, the biggest challenge it faces is big data management, and within that data management, how to maintain compliance of privacy for sensitive data. A project looking at multiple genomes can easily use petabytes of data.[1] The cost of the first Human Genome Project was around 2.8 billion dollars. The cost to generate a high-quality genome sequence as of 2015 was approximately 1,500 dollars.

A significant contributor to the reduction of expenses was the advancements in data storage capabilities. Although cloud technology was not designed with science in mind as its primary customer, it has offered new ideas for arising challenges.

02

GLOBAL COLLABORATION

As of only a year ago, The National Library of Medicine (NLM) began moving its Sequence Read Archive (SRA) onto the cloud. Previously, archived data was not user-friendly. It was a time-consuming task of having sufficient storage to download large data sets and then re-analyze the data from the beginning.

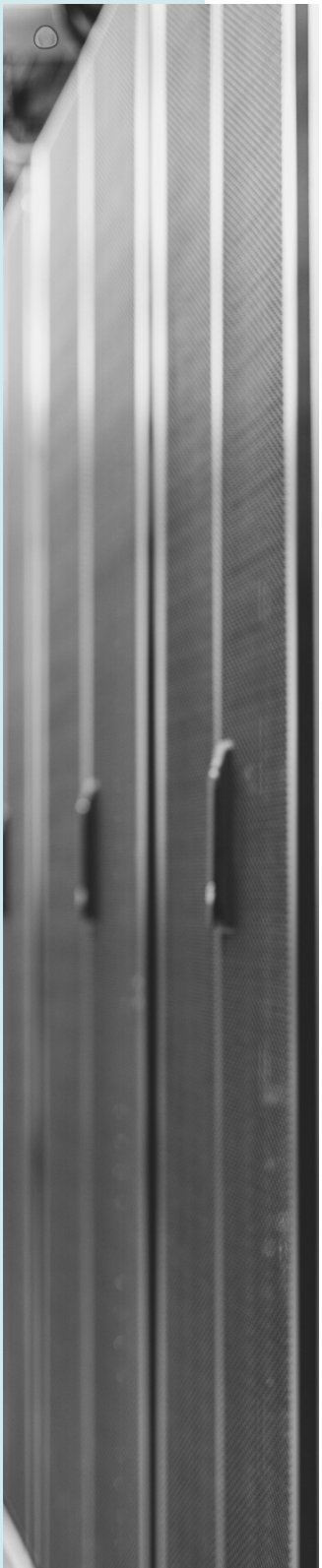
This complicated process led to relevant data going unutilized. The SRA has more than 26 petabytes of genomic data and is expanding.[3] Although still in progress, this initiative is a gigantic leap towards international collaboration, allowing data to be accessible and reusable for the biomedical science field. Such enormous partnerships will only continue to fuel the quick expansion of genomic research and innovation.

The cloud is a technological solution to allow for international scientific collaboration. Never before could such global participation take place. Cloud technology provides a platform for genomic projects, with billions of data points each, to be run in parallel and compared. These projects are also able to be stored and searchable on the cloud for current and future researchers' use, allowing researchers to arrive at solutions at a much quicker rate. Cloud technology ensures that no data will slip through the cracks or go unused.

03

CLOUD ELASTICITY

> *Employing elastic cloud computing allows researchers to get to their answers quickly and move on to the next. That is where one of the most popular benefits of cloud technology comes into play, elasticity.*



The first human genome project produced a massive amount of data, and that was just one project. However, with the commercialization of genomic research over the past two decades, it is estimated that two billion human genomes could be sequenced by 2025.

This amount of data generated has surpassed the human capacity to analyze it. A primary challenge in analyzing large data sets is the time it takes to complete the computer analysis, which can take days and weeks to perform on local computer resources.[4]

Employing elastic cloud computing allows researchers to get to their answers quickly and move on to the next. That is where one of the most popular benefits of cloud technology comes into play, elasticity.

Users no longer must predict the scale of a project's data and match it to a local computer cluster. Which also frees the user from the purchase and maintenance of computer hardware. Instead of users adjusting their data to fit the usage, the usage is adjusted to fit the data.

04

PRIVACY CHALLENGES

When patient data is involved, security is always going to be a concern. There can be risks associated with storing data on the cloud. However, it is estimated that 99% of cloud security failures are the customer's fault.[5] This risk comes from a user not properly taking advantage of the numerous security features cloud technology can offer and implementing the necessary procedures to ensure proper use of the cloud is performed.

Patient privacy is vital, so to avoid issues such as genetic discrimination and information misuse. Many laws have been put into place to do just this. New challenges are presented with the increasing amount of electronic patient information (ePHI) and how to keep it secure on multiple tech platforms. Cloud technology is continually tracking the evolving regulatory and legislative environment, so proper tools will always be in place to meet compliance standards.

The user is still in control of the content and how it is secured with data encryption, access, and logging features. Although it is the user's responsibility to comply with all laws in place to protect ePHI, cloud technology offers powerful tools to ensure that compliance and privacy are met.

Users have the advantage of designing a customizable security environment to meet their compliance needs. Tools at the user's disposal include:

- Strong password policies
- Appropriate user permissions
- Firewall and network segmentation
- Encrypting content

With the cloud, security is every bit as paramount as any other benefit associated with the technology. With so many security features, users genuinely have the tools and assistance needed to meet the rigorous and essential compliance standards, so patient privacy and anonymity are secured.

05

EXPENSE CONCERNS

As with any business project, the expense can often be the final deciding factor. The decision between on-premise or cloud-based infrastructure is a challenging one that requires a good deal of knowledge of the benefits and risks associated with each option.

However, cloud technology is becoming mainstream. It is estimated that cloud technology will grow 17% in 2020.[6] The Cloud is becoming the best option for most companies because of the cost benefits. It allows for an economical alternative to the local storage of data.

Rather than supplying resources that may or may not be used, users pay only for what is needed. As cloud technology advances, costs continue to decline. Many suppliers offer innovative pricing plans and low cost per GB prices, and as the market continues to progress, so does the competition of providers, continuing to drive down costs.

Scalability allows for the pricing plan to fit the project, so no resources are wasted. This benefit is particularly valuable to genomic data as it is hard to predict the scale of many projects as they often produce billions of data points.

06

CONCLUSION

The genomics revolution is upon us. It will change the face of biomedicine as we know it. Previous concerns for the cloud now dwindle as the technology continues to advance and offers innovative answers to these concerns.

Properly applied, cloud technology provides the solutions needed to meet the many challenges arising in the vast field of next-generation sequencing. As these big data problems are solved on the cloud, personalized medicine and drug discoveries will only continue to accelerate.

As each industry continues to grow, so will the technology needed to answer biomedicine's most significant questions. All of which can be answered with an economic and privacy protection mindset in the cloud.

07

SOURCES

1. Dove, E., Joly, Y., Tassé, A. et al. Genomic cloud computing: legal and ethical points to consider. Eur J Hum Genet 23, 1271–1278 (2015). https://doi.org/10.1038/ejhg.2014.196

2. "The Cost of Sequencing a Human Genome," National Human Genome Research Institute, https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost

3. "NLM is Moving its Sequence Read Archive (SRA) to the Cloud," National Library of Medicine, https://www.nlm.nih.gov/news/NLM_Moves_SRA_Cloud.html

4.Big data: Astronomical or genomical? Stephens Z.D., Lee S.Y., Faghri F., Campbell R.H., Zhai C., Efron M.J., Iyer R. Robinson G.E. (2015) PLoS Biology, 13 (7)

5."Is the Cloud Secure?" Gartner, https://www.gartner.com/smarterwithgartner/is-the-cloudsecure/

6."Gartner Forecasts Worldwide Public Cloud Revenue to Grow 17% in 202." Gartner, https://www.gartner.com/en/newsroom/press-releases/2019-11-13-gartner-forecastsworldwide-public-cloud-revenue-to-grow-17-percent-in-2020