CASE STUDY

How Quanam and Genlives Streamlined the Next-Generation of Genomic Sequencing, Saving Time and Saving Lives

Founded in Uruguay in 1978, Quanam is a multi-Latin federation of firms specializing in consulting and management, professional services, communication and change management. Priding itself on innovation and knowledge, Quanam's broad range of clients vary from banking and government agencies to financial institutions and even genetic laboratories. Comprising a 400-strong team of engineers, analysts, economists, managers, accountants and statisticians, Quanam uses state-of-the-art solutions and tools to solve modern day global problems.

Partnering with GenLives — a genomic sequencing company based in Montevideo — Quanam set itself on a course to streamline next-generation sequencing (NGS) analysis for clinical geneticists, with a view to allow clinicians to make more accurate diagnoses in shorter timeframes. With the average human genome capable of generating around 300 gigabytes of information, Quanam was on the lookout for a big data platform that could withstand demanding storage and computing requirements to process multiple complex algorithms.

KEY HIGHLIGHTS

- Reduced DNA sequencing time from days to hours
- Pathogenic mutations can be detected in much shorter timeframes
- Process of diagnosis shortened from an average of 5 years down to eight/ten weeks
- Created a model which allows clinicians to make more informed decision as more diagnoses are made

FINDING PATHOGENIC MUTATIONS: A NEEDLE IN A HAYSTACK

Next-generation sequencing is a high throughput DNA sequencing technique, which in recent years has been applied across multiple fields, including human genomics. The genome of a single person is extremely complex and during sequencing can deliver up to four million mutations. With each single mutation being responsible for defining individual physical characteristics — such as the type of hair or the color of the eyes — Quanam was finding it difficult to isolate the information responsible for causing pathogenic conditions.

Relying on conventional methods, a single clinical geneticist had to sift through millions of existing academic papers, medical databases and literature to manually assess potentially malignant conditions. With researchers spending hours of valuable time reading and cross-referencing a variety of sources to isolate relevant mutations, it was becoming unsustainable for the organization to produce timely insights that would help doctors and clinicians make a diagnosis.

Medical literature features millions of contributions from researchers around the globe. With an estimated 300 new medical papers released every day — equaling to more than a 100,000 a year — Quanam and GenLives realized that human input alone would never match the speed and accuracy that automation could bring. In a field where diagnoses for rare diseases can take years, and a time-saving solution can potentially be life-saving, the aim was to prevent a bottleneck in the system and speed up the diagnosis and treatment of a patient.

> "We're convinced Hortonworks is a great option, especially in providing a proven infrastructure for both data-at-rest and data-in-motion"
>
> **Fernando Lopez**
> *Quanam*

AUTOMATING THE CLASSIFICATION OF GENOMIC VARIANTS

Having recognized the potential of big data analytics, Quanam had established a big data division within its business unit. It was meant to address not only the challenges brought by speed and volume of data, but also capitalize on the opportunities brought by artificial intelligence and cloud computing.

Quanam chose to partner with Hortonworks as a result of its open source nature and its unique range of services. Through Hortonworks Data Platform (HDP®), Hortonworks offered the best example of a community-driven tool stack for Quanam to analyze and store large volumes of data at scale. Newer developments such as Spark and NoSQL meanwhile allowed Quanam to address a large list of now-computing-affordable problems.

The proposed solution was twofold, aimed at streamlining the analysis of DNA sequencing and improving the speed of data mining, whilst also automating the process of correlating medical literature for relevant references to potential pathogenicity.

Quanam developed a semantic knowledge application to perform several natural language processing tasks aimed at identifying relevant entities and relationships across complex datasets. Hortonworks Data Platform offered a foundation on which several machine learning algorithms could be run, whose intermediate results could be either queried in Apache HIVE or exported to Apache Solr for indexing and exploring.

The application also featured an IBM Watson-powered chatbot, allowing users to perform queries across different languages to that of the resources in the database. From there, it automatically generates a report for the geneticist.

MOLECULAR-BASED DIAGNOSIS

Complete with visualizations, powerful linguistic tools and heuristics, clinical geneticists can now make more informed decisions on the pathogenicity of genome variants. The platform streamlines and empowers next-generation sequencing analysis, and more broadly, democratizes the availability of genomics analysis.

The process of pathogenicity data mining and the analysis pipeline for DNA sequencing now takes only a matter of hours, rather than an entire day. Meanwhile, the now pre-enriched literature database uses algorithms to establish connections across millions of resources, helping to build a clearer picture of which genomic variant is pathogenic — an invaluable tool for clinical geneticists developing diagnostic methods and potential treatments.

"*We're convinced Hortonworks is a great option, especially in providing a proven infrastructure for both data at rest and data in motion*" said Fernando Lopez, Quanam. "*Hortonworks has helped us to effectively streamline and empower next-generation sequencing. The potential implications of this are significant. Pathogenic mutations can now be detected in much shorter timeframes, meaning that patients get an informed diagnosis in a matter of weeks, as opposed to months or even years. What's more, the process is now constantly improving – with each new diagnosis, those findings can be re-used in later cases.*"

So far, more than 50 patients have been analyzed with the new platform. With traditional methods unable to effectively cross-correlate the sheer amount of medical knowledge available worldwide in a short timeframe, Quanam and Genlives have been able to shorten the process of diagnosis from an average of 5 years down to eight/ten weeks. With over 400 million patients globally suffering from rare conditions, the potential of molecular-based diagnoses promises to save lives.

Quanam is currently involved with several other projects, which rely on the analysis of data-in-motion to produce timely insights. Specifically, Quanam is working on an educational use case which leverages Hortonworks Data Flow (HDF) to ingest data from a number of laptop devices such as watching logs and activity tracking.

The next steps for Quanam and GenLives' genetics project is to develop governance, relying on governance components like Atlas and Ranger in order to accommodate a multi-tenant platform of different laboratories with their own data sets.

About Hortonworks

Hortonworks is a leading provider of enterprise-grade, global data management platforms, services and solutions that deliver actionable intelligence from any type of data for over half of the Fortune 100. Hortonworks is committed to driving innovation in open source communities, providing unique value to enterprise customers. Along with its partners, Hortonworks provides technology, expertise and support so that enterprise customers can adopt a modern data architecture. For more information, visit hortonworks.com.

Contact

For further information,
visit hortonworks.com

+1 408 675-0983
+1 855 8-HORTON
INTL: +44 (0) 20 3826 1405

HORTONWORKS

07/2018