Datameer
stay curious

Using Big Data Analytics to Create Better Outcomes for Cancer Patients

dkfz.
GERMAN
CANCER RESEARCH CENTER
IN THE HELMHOLTZ ASSOCIATION

The Problem

Cancer is responsible for the early deaths of millions of people worldwide each year. In Germany alone, more than 450,000 people are diagnosed with cancer annually. Because every cancer is unique, cancer diagnosis is complicated, and treatment outcomes vary hugely from patient to patient.

Thankfully, organizations like DKFZ, the largest biomedical research Institute in Germany, are working to understand the mechanisms of cancer, identify risk factors, and find new ways to prevent people from getting cancer. A key focus of DKFZ's medical researchers is genomic data research. By analyzing human genomes, they can identify DNA problems which are the root cause of cancer in an individual — information that can be used to personalize cancer treatment — as well as understand the genetic evolution of disease and monitor patient responses to different treatments to better understand their efficacy.

The Analytical Challenge

Due to the massive volumes of genomic data involved in this research, DKFZ faced huge challenges on the data and analytics front. The human genome contains 3 billion base pairs. To analyze cells at a genetic level to identify root causes and indicators, they needed to analyze several tissue samples from several thousand patients, focusing on 900,000

genomic base pair positions out of 200 that are related to cancer. With oversampling factors up to 100, and the equivalent of 200 GB of data per patient, their analytic systems were overwhelmed by many petabytes of data. Even when they tried using high performance clusters to speed things up, analyzing an entire patient data set took weeks and even months to complete before they could determine the next set of questions data analysis could answer. These huge bottlenecks greatly slowed research and frustrated staff.

To address this issue, they tried using data reduction methods, which did help shorten analytic time frames to between 24 and 48 hours per patient. But the loss of information came at a high cost. It was hard to know which data was inconsequential and thus could be eliminated from the data set without negatively impacting outcomes and resulting in data misinterpretation. Moreover, they can only parallelize by patient (in other words, only analyze one patient data set at a time). The researchers simply didn't have time to wait this long given the several thousand patients that needed to be included in the study.

DKFZ wanted a way to analyze an entire patient data set at once – simultaneously, quickly, and in parallel with other patient data sets so they could arrive at actionable results in a shorter time frame. The goal was to simplify and accelerate analysis in order to analyze complete data sets quickly and efficiently.

Datameer at Work

DKFZ collaborated with Fujitsu to deploy the Fujitsu Prime Flex Integrated System for Hadoop, a powerful and scalable server cluster that uses Datameer's analytic platform for parallel processing power and incredibly fast, actionable analytics.

With Datameer, researchers can analyze the complete, raw genomic data sets of multiple patients in parallel, along with patient data records, detail selection data, and reference genome data. Analysis on the complete, unreduced data set can be performed by block of data and by patient. Using this new parallelization technique the analysis on the complete data set can be completed between five and twenty minutes. By comparisons, the analysis on the greatly reduced data set used to take 24-48 hours to complete.

Equally important, Datameer eliminates the risk of data degradation by minimizing the number of processing steps used in analyses and the need for data movements between systems.

Results

Using Datameer, DKFZ can now analyze 10 TB of raw data per day – the equivalent of 140 billion records looking at 900,000 thousand positions in each genome. They can analyze complete data sets in minutes, eliminating the need to reduce data and risk missing out on key insights. Vastly faster processing enables the DKFZ to more quickly identify specific, optimal cancer therapies for each patient, as well as further their overall research on correlations between cancer and genetics.

Because DKFZ can analyze complete, raw data sets (instead of reduced data sets), they can uncover significant new findings in each patient, by analyzing Exons (the section of DNA that contains the protein coding instructions) and Exon gaps, or Introns (the material in DNA that breaks apart the Exons). These findings can identify where extra intronic material (gaps) has been spliced into the genes creating specific problems in the human body that can be targeted with personalized treatments. The Exons and Exon gaps simply couldn't be detected using prior analytic methods.

Through this analysis, powered by Datameer, DKFZ is helping patients fight cancer every day, and is actively on the path to finding a cure.

Datameer's ability to dramatically reduce the time to analyze genomes and identify granular aspects such as Exon gaps in the analysis can help revolutionize the medical field. This creates tremendous opportunities for healthcare, biotechnology and government research organizations to apply genomic research to a wide variety of fields, and can help the entire healthcare industry apply personalized treatments that can radically change patient care.

FREE TRIAL
datameer.com/free-trial

TWITTER
@Datameer

LINKEDIN
linkedin.com/company/datameer

SAN FRANCISCO 1550 Bryant Street, Suite 490 • San Francisco, CA 94103 USA • Tel: +1 415 817 9558 • Fax: +1 415 814 1243
NEW YORK 9 East 19th Street, 5th floor • New York, NY 10003 USA • Tel: +1 646 586 5526
HALLE Datameer GmbH • Große Ulrichstraße 7 – 9 • 06108 Halle (Saale), Germany • Tel: +49 345 2795030
SINGAPORE Datameer Singapore Pte Ltd • 03-20 Galaxis 1, Fusionopolis Place • Singapore 138522, Singapore • Tel: +65 6809 1157
HONG KONG 12/F International Commerce Centre, 1 Austin Road, West Kowloon, Hong Kong • Tel: +852 2824 8646