Academic Research

CDFD accelerates research into rare diseases

The Centre for DNA Fingerprinting and Diagnostics (CDFD)

With data volumes growing, the CDFD established a new high-performance computing cluster powered by Lenovo ThinkSystem and GOAST solutions to help researchers accelerate cutting-edge bioinformatics research.

1

Who is the CDFD?

Established in 1996, The Centre for DNA Fingerprinting and Diagnostics (CDFD) is an autonomous research and educational organization that is primarily funded by India's Department of Biotechnology, Ministry of Science and Technology. Headquartered in Hyderabad, the CDFD provides powerful computational resources to support scientific discoveries across multiple domains, including research on DNA fingerprinting and forensic, disease diagnostics, genetic disorders, fungal pathogenicity, infectious diseases, epigenetics, cancer genomics, plant pathogen interactions, and more.

The CDFD also runs the National Genomics Core, which provides cutting-edge genomics sequencing and data analysis solutions to academic and government institutions.

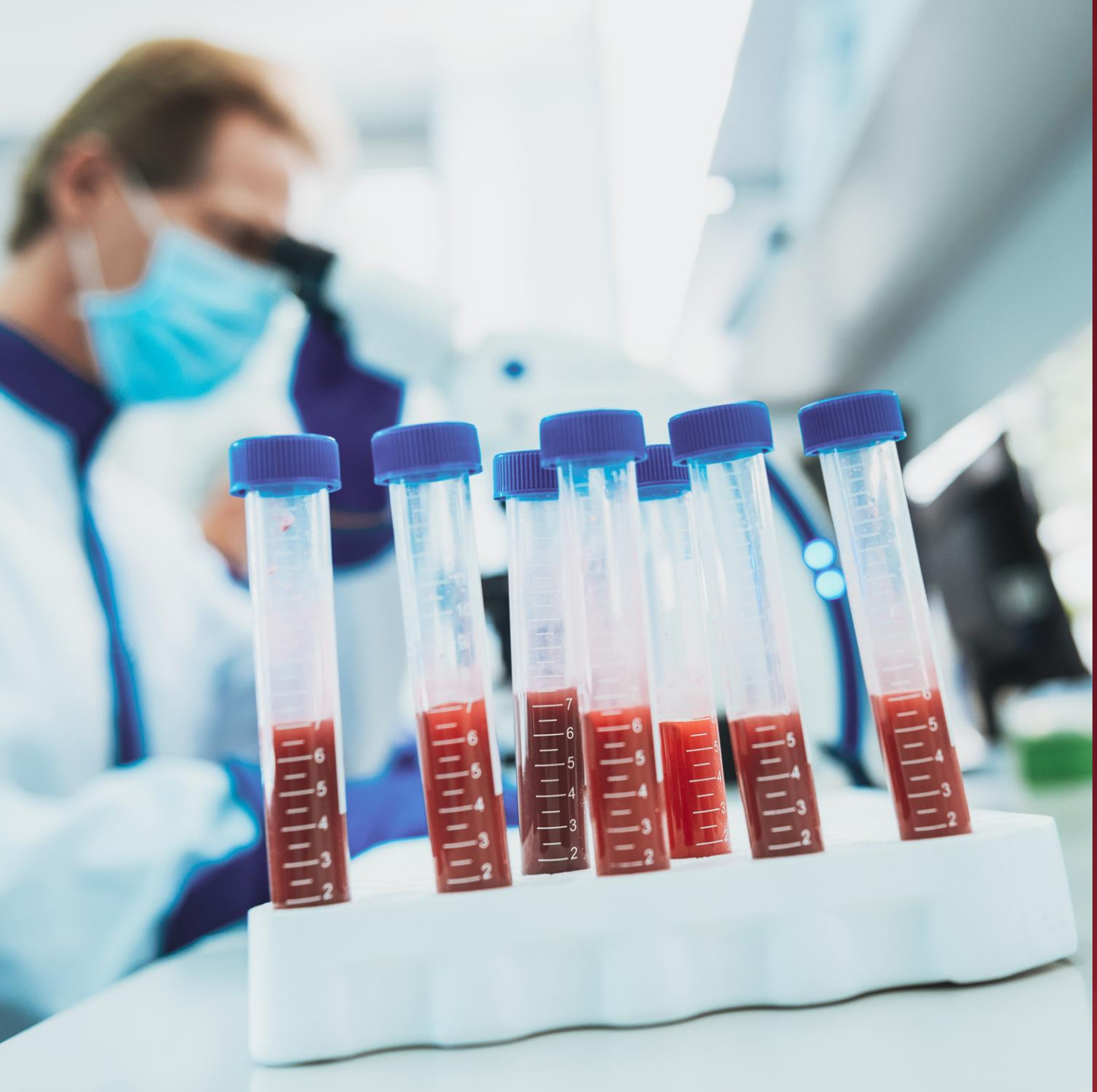
2

The Challenge

Recent advances in genome sequencing and computer modelling have spurred a new wave of scientific research. However, these advanced research techniques require powerful compute, RAM, and storage resources to support rapid, accurate, and effective data analysis.

As one of India's leading academic research institutions, the CDFD provides the country's scientific community with the high-performance computing (HPC) and genomics sequencing tools they need to understand rare diseases, discover new medicines, improve public health, and more.

In 2022, the CDFD launched the Mission Program on Pediatric Rare Genetic Disorders (PRaGeD), a pan-India initiative funded by the Department of Biotechnology, to tackle pediatric rare genetic diseases through diagnosis, novel gene discovery, counselling, and therapy development. The program aims to sequence and analyze the genomes and exomes of approximately 10,000 patients in India by 2026. However, relying on its traditional HPC environment for Genome Analysis Toolkit (GATK) jobs would have prevented the CDFD from meeting this ambitious target.



“Genomics and bioinformatics are rapidly evolving, which means we must constantly enhance and improve the computational resources that we can offer to researchers. As a publicly funded institution, we want to invest in innovation while also still ensuring maximum value to the taxpayer. With our genomics workloads increasing, we decided to look for a cost-effective way to level up our HPC capabilities.”

Dr Ashwin Dalal
Staff Scientist, CDFD

Investing in innovation

To support a higher volume of projects and drive more complex research, the CDFD worked with Lenovo Services to build a brand new HPC cluster using a combination of Lenovo ThinkSystem SD530, SR650 V2, SR665, and SR950 servers alongside DDN storage devices.

To accelerate genomics workloads, the CDFD also deployed the Lenovo Genomics Optimization and Scalability Tool (GOAST)—Lenovo's tailored genomics solution designed to enable faster, more responsive mapping and whole-genome sequencing.

Hardware

Lenovo ThinkSystem SD530
Lenovo ThinkSystem SR650 V2
Lenovo ThinkSystem SR665
Lenovo ThinkSystem SR950
DDN Storage 7990 with Lustre
Infiniband Switch10G Network Switch

Software

Red Hat Enterprise Linux

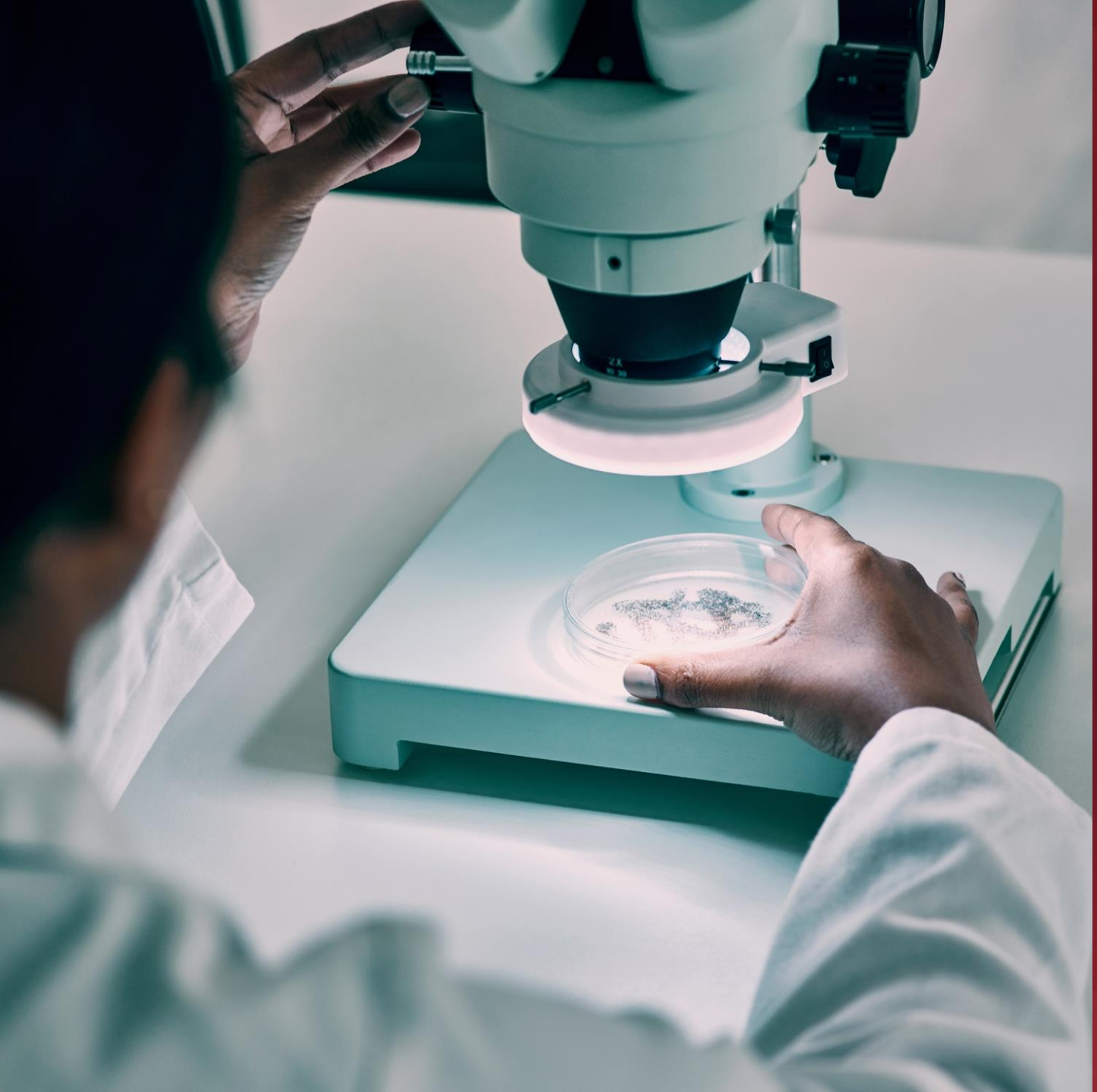
Services

Lenovo Consulting Services

Ensuring rapid deployment

The CDFD worked with Lenovo to deploy its brand new HPC cluster in the middle of the global COVID-19 pandemic, which presented significant supply-chain challenges for the organization.

“At the time we were implementing our new HPC environment, there weas a global shortage of hardware,” explains Dr. Ajay Kumar Mahato, Staff Scientist at the CDFD. “Despite this, Lenovo went above and beyond to deliver the hardware on time and worked with us to ensure that the Lenovo servers were shipped from a site outside of Asia to meet the specific procurement requirements of the CDFD. What’s more, with Lenovo support we were able to install our new HPC cluster in less than three days, which is much faster than we originally anticipated.”



> “When I learned about the Lenovo GOAST solution, it was clear that it would help us to process more genomics samples much faster—and with the added benefit that the underlying infrastructure could be used for other HPC workloads when it is not needed to support genomics-based research.”

Dr Ajay Kumar Mahato
Staff Scientist, CDFD

3

Results

Since expanding its HPC capabilities with Lenovo solutions and accelerating its genomics workloads with GOAST, the CDFD can now take on a larger number of research projects and process them in parallel. Researchers now get faster access to genomic sequencing and analytics data, helping them to make new scientific breakthroughs.

At the same time, the CDFD has been able to make its HPC resources more widely available to academics and students by offering remote HPC capabilities through a dedicated virtual private network—helping more researchers and students to gain hands-on experience in bioinformatics.

- Accelerates genomics workloads
- Makes HPC resources more accessible
- Helps the CDFD support more research initiatives

Supporting new discoveries

Already, the CDFD's newly enhanced HPC environment is supporting new breakthroughs that will help scientists deepen their understanding across multiple fields. In one recent project, the CDFD used the Lenovo GOAST solution to decode the genome of a chicken native to India; work that will help to uncover fresh insight into optimal feeding strategies for farmers.

"During this recent project, Lenovo GOAST helped to speed up the decoding of the genome of a unique breed of chicken, approximately 1.21 gigabase pairs (Gbp) in total," says Dr Ajay Kumar Mahato. "Not only did the GOAST solution make decoding the genome much quicker, but it also made secondary and tertiary analysis much faster. Being able to accelerate bioinformatics in this way ultimately frees up more time for the CDFD to drive more research initiatives."

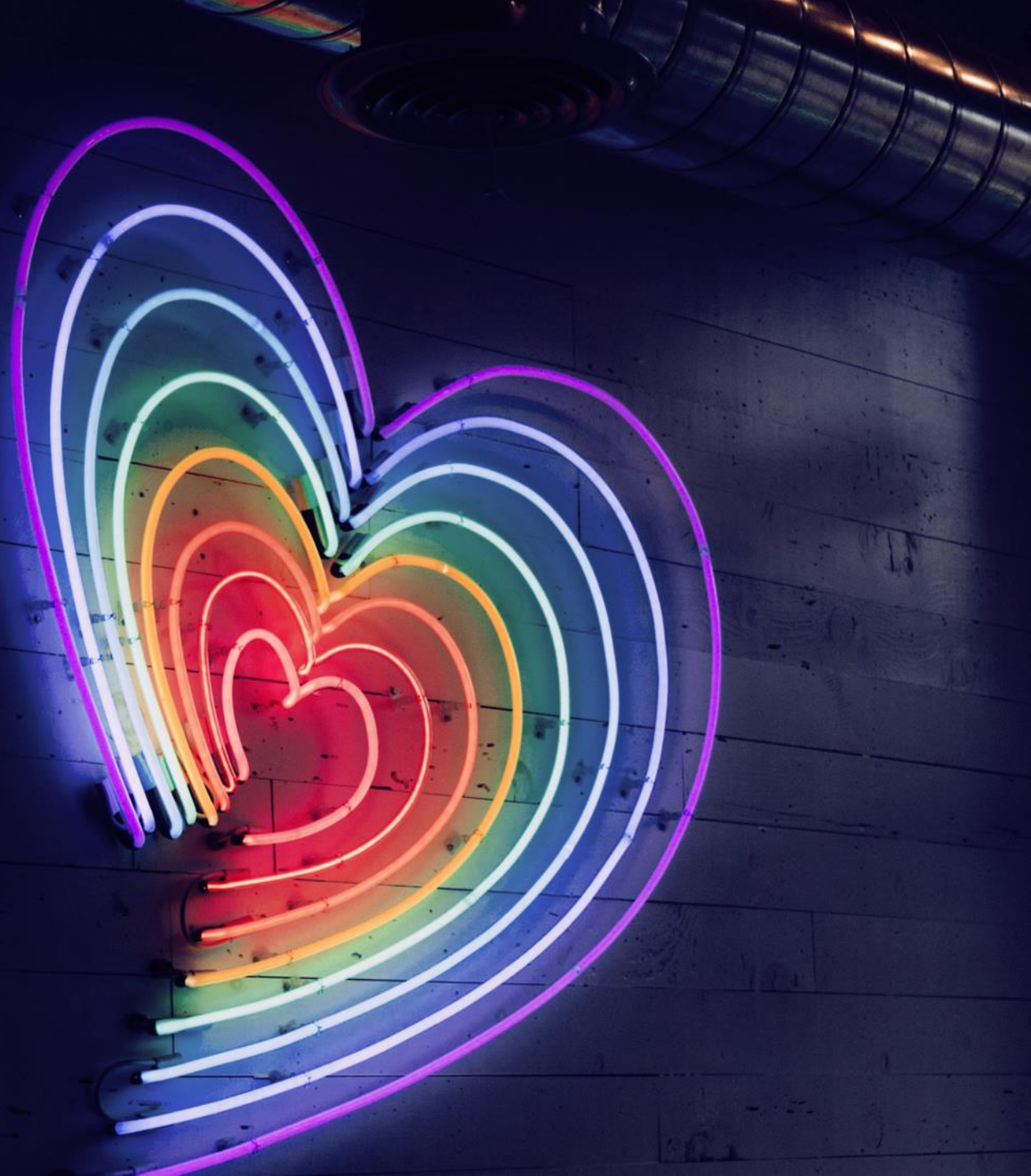
> “We have only just scratched the surface of what is possible with our new HPC capabilities. With powerful HPC hardware and bioinformatics software from Lenovo, we are now on track to reach our ambitious goal of sequencing and analyzing the exomes and genomes of approximately 10,000 patients with rare diseases in India, while also supporting a broad range of research programs in parallel.”

Dr Ajay Kumar Mahato
Staff Scientist, CDFD

Why Lenovo?

To ensure that it provides scientists with the very best resources, the CDFD looked for a technology vendor with a proven track record of success in HPC.

“As a PhD student and a researcher, I have worked with many different vendors of HPC vendors and technology partners,” says Dr Ajay Mahato. “Lenovo, by far, offers the best support in my experience. If we ever encountered any issues, the local Lenovo team was always very responsive and happy to help us solve them.”



How can researchers accelerate genomic data processing?

The CDFD built a new HPC cluster with Lenovo—enabling it to take on more research projects and get sequencing results faster.

Explore Lenovo HPC solutions