



Dell Genomic Data Analysis Platform

Accelerate discovery and insights through optimised infrastructure and support

Amazing work is being done today in the healthcare and life sciences. With the advent of genome sequencing researchers now have a critical gateway to understanding the underlying molecular pathways for disease. And for some diseases, every hour closer to discovery and results can mean the difference between life and death.

However, many factors are hindering discovery and causing inefficiencies. Genomic processing requires immense computational power and storage. Many researchers, lacking adequate offsite resources, are faced with the challenge of designing and deploying their own infrastructure with little IT expertise or support. Add to this complex integration and performance tuning of the environment—which can take months and result in low system usage and delays in key projects.

A complete, end-to-end genomic processing infrastructure

Partnering closely with our customers at the forefront of genomics research, we designed the Dell Genomic Data Analysis Platform to help accelerate time to insight in a diverse range of fields including drug design, cancer research, agriculture, biofuels and forensics.

This solution provides a flexible, high performance computing environment architected to deliver high-throughput and fast turnaround of genomic workflows. Sequenced data is written from Next Generation Sequencing (NGS) machines directly into the system's computational scratch space for processing; output data and user files are network-accessed by researchers for further investigation

The Dell Difference

Dell's comprehensive high performance computing and data management solutions, domain expertise and technical know-how help genomic research organisations in healthcare, government, universities and industrial labs accelerate their research and engineering to:

- Concentrate resources on R&D, not complex infrastructure
- Optimise genomic pipelines for quicker results
- Overcome obstacles to mainstream product viability
- Identify treatments in clinically relevant timeframes
- Enable cost-effective bioinformatics centers
- Maintain compliance and protect confidential data using secure, in-house resources

Learn more at Dell.co.uk/hpc



Dell's complete, end-to-end genomic processing infrastructure

Designed to meet the needs of genomic research centres and core facilities, the Dell Genomic Data Analysis Platform is a complete high performance computing and data management solution architected to deliver high-throughput and fast turnaround of genomic workflows. Sequenced data is written from NGS machines directly into the system's computational scratch space for processing across up to 512 computing cores; output data and user files are network-accessed by researchers for further investigation.



Open, integrated HPC infrastructure

- Scalable Linux-based 32 node cluster, 1.5TB large memory node (optional)
- Lustre® parallel file and Network File Storage (NFS) systems to optimise data
- Standard workload management, development tools
- Simple connectivity with next generation sequencing (NGS) systems



Efficiency and high performance

- Up to 9.4 Tflops peak, 1.5 TBs RAM, and 540TBs in a single 42U chassis¹
- Best-in-class compute density and energy efficiency
- Standard 10Gb Ethernet or InfiniBand® cluster interconnects



Architected for genomic workloads

- Architected and optimised by Dell genomic HPC experts
- Industry-leading performance in genomes per day, genomes per watt
- Cost-effective genomic data collection and analysis



World class HPC expertise

- Available in Dell-recommended configurations to achieve optimal performance and reliability
- Built on proven Dell products: latest generation compute, storage and networking
- One source for solution design, delivery and support

Accelerate time to insight

Deliver fast, efficient results

The Dell Genomic Data Analysis Platform is a complete, integrated infrastructure platform of modular building blocks specifically optimised for high performance genomic analysis.

It easily integrates with standard genomic sequencing systems. Parallelized genomic workloads reduce complex workflows from days to just hours, enabling processing of up to 37 genomes per day and an energy efficiency of 7.42 kWh/genome.²

Simplify your HPC infrastructure

Make it simpler and more cost-effective to deploy and manage your analytical processes through a highly flexible framework of scalable HPC technology and services. Overcome the challenges to collect, store and analyse growing genomic data sets, and increase focus on research, not on managing complex infrastructure.

The right system and implementation, proper services and support must be designed to make efficient research workflows and processes that get results faster. With Dell's optimised and pretested configurations, you can reduce lengthy implementation timelines from months to weeks.

Maximise productivity without compromising on performance

Built on an open, standards-based architecture, our flexible analytics platform is tuned for performance, density, efficiency and cost—enabling research labs to reduce complex workflows processing times from days to hours.

Integrated, energy efficient Dell hardware provides the right balance of compute density and energy efficiency to help deliver superior performance per watt for lower ongoing operational costs.

¹ Performance metrics above based on May 2013 internal Dell HPC lab benchmark testing (SW: bwa, GATK, bcio-nextgen Ref. Genome: GRCh37 Input reads: 212M, ~10X coverage), confirming 1) 37 genomes per day throughput on a Genomic Data Analysis Platform configuration across 480 cores 2) confirming 37 genomes per day throughput compared to 2.01 genomes per day across 16 cores. 3) the complete analysis of a genome in 6.8 hours across 96 cores 4) total energy consumption of 227.77 kWh when running 30 concurrent whole genome analyses in parallel, resulting in 7.42 kWh/genome. 5) Confirmed sustained 8.26 Tflops, 88% of a theoretical maximum of 9.4 Tflops, based on May 2013 internal Dell HPC lab benchmark testing. Actual performance will vary based on configuration, usage and manufacturing variability.

² The Dell Genomic Data Analysis Solution can process up to 37 genomes per day / 259 genomes per week. Results based on additional internal Dell HPC lab benchmark testing completed in July 2013 confirming 37 genomes per day throughput on a Dell Genomic Data Analysis Platform configuration across 480 cores. Actual performance will vary based on configuration, usage and manufacturing variability.

The Dell Genomic Data Analysis Platform can process genomes with an energy efficiency rate of less than 7.5 Kilowatt-hours per genome processed. Results based on additional internal Dell HPC lab benchmark testing completed in July 2013, confirming energy consumption of 227.77 kilowatt hours (kWh) when running 30 concurrent genome analyses, resulting in 7.42 kWh/genome (227.77 kWh/30 genomes). Actual performance will vary based on configuration, usage and manufacturing variability. Dell Ad# 13001995.

Reader services

If you do not wish to receive this marketing communication from Dell in the future, please let us know by sending us your name, company name and postal address:
Post: Dell – Suppressions, PO Box 59,
ROSS ON WYE, HR9 7ZS, United Kingdom
Email: Delluk@clientmail.eu.com
Fax: 0800 3283071



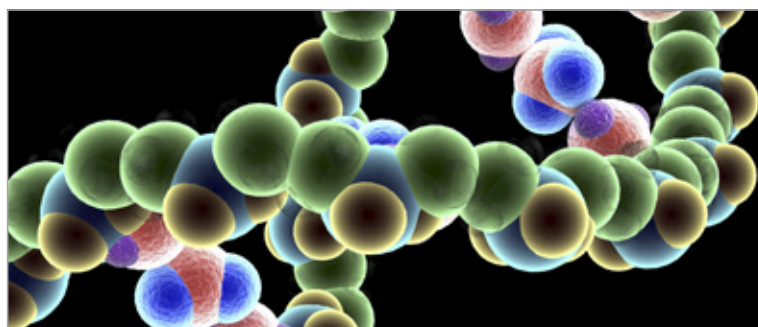
Enable bioinformatics innovation by maximising productivity

Dell's Genomic Data Analysis Platform is designed to help you achieve timely results and insights faster, through quick deployment, simple management, and high performance, backed by Dell's dedicated solutions, support and services teams.

Through insightful partnership and collaboration, Dell is committed to help our life sciences and research customers dedicate more time to science and engineering, enabling new discoveries, novel innovations, and timely trials and treatments.

"We can now process a patient's DNA in less than a day. That's not just high-performance computing; that's life-saving computing."

Dr. Harold "Skip" Garner
Director, Medical Informatics and Systems Division
Virginia Bioinformatics Institute



Finding a cure for Pediatric Neuroblastoma

Research challenge

Getting to results in a clinically relevant amount of time for an aggressive form of childhood cancer. The next-generation genome sequencing technologies that the Translational Genomics Research Institute (TGen) uses for a neuroblastoma trial generate a tremendous volume of data.

Obstacles

- Whole genome data sets are on the order of a TB or more in size
- Huge data volume – 100% increase every year for the past 3 years
- Genome data requires parallel processing for timely results

Results

- Fast turnaround times – up to 37 genomes (30x) processed per day
- Increased computational performance by 1,200%
- Compressed genomic analysis pipeline from 7-8 days to <1 day
- Reduced mapping and analysis from months to days and cost from millions to thousands of \$
- Net: faster time to treatment and better odds for positive outcomes

"With diseases like neuroblastoma, hours matter. Our new Dell HPC cluster allows us to do the processing we need to get a meaningful result in a clinically relevant amount of time."

Jason Corneveaux
Bioinformatician Translational Genomics
Research Institute (TGen)

Read the entire TGen case study at Dell.com/casestudies.