

Singapore scientists design novel genome sequencing data compression method

May 30 2012

Hitachi and Data Storage Institute (DSI), a research institute of the Agency for Science, Technology and Research (A*STAR) are devising a data compression technique to tackle the increasing volume of genome sequencing data generated by the healthcare and biomedical industry. As the volume of such data has been forecasted to double annually, the collaborators aim to develop a more efficient data storage technology that will compress genome sequencing data more effectively than existing methods. This is an extension of an earlier partnership, where Hitachi and DSI researchers discovered the pattern of typical genome data transactions that would enable current storage systems to function optimally.

Genome sequencing is a data intensive process and high-powered machines are required to decipher the order of deoxyribonucleic acid (commonly known as DNA) nucleotide bases – Adenine (A), Cytosine (C), Guanine (G), and Thymine (T) that consist within a DNA molecule. A human genome of an individual contains over three billion of these genetic letters and occupies up to 725 MB of uncompressed data. The data multiplies when it is replicated, processed and shared globally among researchers for more experiments which can amount to terabytes of data. Scientists and medical practitioners rely on genome sequence to decode the string of letters and gain a clearer understanding of the human anatomy, how genes interact and affect the growth and development of an organism. This in turn helps identify the causes of common genetic disorders. For instance, sequencing the genes of tumour cells can aid doctors in their study of mutations and differentiate

cancerous cells from normal tissues, enabling them to prescribe appropriate drugs that will treat the affected tumours more accurately.

With such tangible medical benefits compounded by the advancement of high throughput sequencers, the use of genetic analysing tool is becoming more widespread and is likely to lead to an overwhelming increase in the velocity, volume and variety of genome data being created. This trend poses significant challenges for data centres to provide high performance storage systems and fast retrieval of large genomic data files. The exponential growth of genome sequencing data will also place pressures on current data centres, slowing down performance levels and creating massive demands for larger hard disk space. Other factors that will drive cost up include the high energy consumption required to power the data centres and the operating cost of maintaining the infrastructure.

In a bid to address the current computational and scalability limitations, DSI researchers were commissioned to study how genome sequencing data is optimised by researchers from Genome Institute of Singapore (GIS), another A*STAR research institute. Research into the characteristics of genome data revealed that existing data compression methods are unlikely to manage current workloads due to inefficiencies and heavy demands for larger memory storage. Building on the collective insights from this earlier project collaboration, Hitachi and DSI are now working towards perfecting the shortfalls identified in current data storage models to design an innovative genome data compression method reduce data storage capacity needs, quicken decompression speeds and lower storage costs.

“By raising compression capacity, we can envision smaller genome sequencing facilities to handle petabytes of data in a year compared to current terabytes levels which are mostly restricted to large genome sequencing centres due to storage limitations. DSI will continue to play a

pivotal role in enabling new storage technologies for the biomedical research and healthcare industry to accelerate research findings and discoveries,” said Dr Pantelis Alexopoulos, DSI’s Executive Director.

“We are delighted to continue our long-standing partnership with DSI in the research field of networked storage. As the industry leader in storage technology and bioinformatics software solutions, I am confident that the outcome of this collaboration will lead to more innovative solutions that could potentially be one of Hitachi’s future areas of business expansion,” said Mr Makoto Nagashima, Managing Director of Hitachi Asia Ltd.

Provided by Agency for Science, Technology and Research (A*STAR), Singapore

Citation: Singapore scientists design novel genome sequencing data compression method (2012, May 30) retrieved 21 June 2024 from <https://phys.org/news/2012-05-singapore-scientists-genome-sequencing-compression.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.