GenomeStack™

SQREAM TECHNOLOGIES

GenomeStack™ is a first of its kind platform, developed specifically for bioinformaticians, representing a paradigm shift for the Genomic Research Industry. GenomeStack™ provides researchers with the ability to store and effectively query massive amounts of genomic data (BAM/VCF/fasta/fastq files etc.) on a columnar, big data SQL database – delivering insights up to **100 times faster** than any other leading solution available on the market.

Advancing Research by Years

Next-generation sequencing produces enormous amounts of data. By loading and analyzing these massive quantities of genomic data in an efficient, ultra-fast and cost-effective way, bioinformaticians are enabled to rapidly extract knowledge that has been unattainable thus far speed- and scale-wise. With GenomeStack™, researchers are now able to quickly discover patterns among multiple genomic samples and reach informative insights, way ahead of their estimated timelines. Finalization of research phases may be advanced by months and even years – depending on the size of the data. With SQream, genomic research and personalized medicine, can truly be progressed.

Preloading Databases

GenomeStack™ enables pulling and preloading of content from various databases such as UCSC, 1000 Genomes and dbSNP, while adapting it to the database's structured SQL tables. This leads to a transformed workflow for bioinformaticians – with a hassle-free, automated, quick analysis process.

AUTOMATION OF RESEARCH WORKFLOW

Key Benefits

- 100X faster insights
- Single-point-of-knowledge
- GPU technology:
  - » Decompression - 50X faster than with CPU
  - » Sorting – 10X faster than with CPU
  - » Analyzing capabilities of up to Petabyte-scale on a single rack
- Fast ROI; Low TCO
- Quick, hassle-free plug&play implementation
- Rapid NAS dataset management
- Easy and intuitive usage (SQL)
- Real-Time accurate, reproducible results
- Low at resources
- Maximization of existing core IT investments

A First-Time Ever Presented Ability: Effective Data Modeling for SQL Querying of BAMs

The GenomeStack™ database replaces the thus far dominant way of storing and analyzing sequenced data, involving an old-school, non-flexible file-based, highly time- and resources consuming complex manual/scripted process. In this traditional file-based storing process, each query requires transformation of compressed BAM data into its readable SAM representation (a time consuming process), as well as execution of a pre-existing script, for extracting the desired data for further analysis and observation. Furthermore, if the research calls for a comparison of a certain chromosome position(s) between multiple samples, this complex and lengthy process needs to be repeated for each sample, prior to the comparison.

With GenomeStack™, post-aligned sequencing data from multiple samples is stored in an ultra-fast and highly scalable database, enabling the querying of a large number of samples simultaneously. Thus, an analysis of 50 BAM files with 30 Billion rows of aligned sequencing data, can be analyzed in under a minute, while providing a detailed drill-down from all participating samples to the single read level, in order to review the nucleotides distribution on a specific chromosome position.

Focusing on Research Instead of Information Technology

As the digital revolution gains pace, and Genome Research institutions need access to rapidly-scaling constantly-generated multiple-sources data, bioinformaticians are faced with a growing challenge related to the storage and management of their data. GenomeStack™ provides bioinformaticians with the ability to store and quickly analyze large sets of data, with a peace of mind knowing that the database in use won't collapse when processing massive quantities of datasets. A hassle-free, easy-to-use and implement solution, that will enable them to focus on the genome research instead of the computational environment.

With GenomeStack™, no data modeling is needed, no new DBA skills are required, and no new and expensive development and usage skills need to be learned/acquired. Up to 100TB of raw data can be stored and queried in a standard 2U server internal storage (disks), with growth potential to petabyte-size data through the use of an external storage. Third party ETL and BI/visualization tools can be connected via standard ODBC/JDBC interfaces.

Breakthrough GPU-Based Technology in Big Data

GenomeStack™ combines the best technologies of both aspects of the Genomic Research – NGS (Next-Generation Sequencing) and a GPU-based database (GPU -

BREAKING THROUGH SPEED-AND SCALABILITY BARRIERS

Graphics Processing Unit). GPUs are typically used for efficient and fast image and video rendering on screens, such as those used by the movie and gaming industries. With SQream DB™ - the underlying Big Data database of GenomeStack™, the GPU technology is used in a revolutionary different fashion that delivers vast compute power to data processing. This results in a major power boost, enabling research institutions to reach their research insights up to 100 times faster when compared with their file-based traditional computation methods.

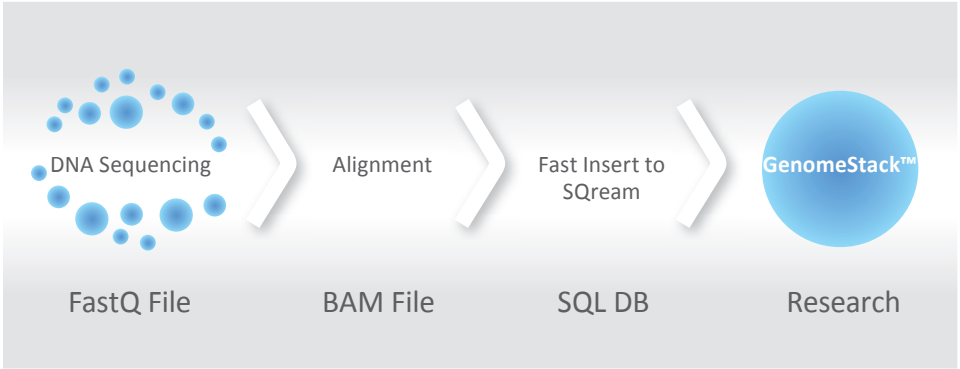
The power of GPU when encapsulated in commodity servers delivers significant savings in hardware, power consumption, cooling systems and floor space. When combined with the SQream database, there are also significant savings in storage, especially when compared to other storage solutions that require full floor-to-ceiling racks.

With GenomStack, bioinformaticians are now able to enjoy the strengths of a high-speed GPU-based columnar database, leveraged by a high-capacity storage management system, with transparent genome-specific compression capabilities.

The Process

With GenomeStack™, research queries are transformed into relational (SQL) queries that in turn are translated into highly parallel execution algorithms. Researchers upload their sample data into the GenomeStack™ database, together with data originating from an unlimited number of other databases, commonly used by bioinformaticians (i.e. 1,000 Genome/dbSNP/UCSC).

GenomeStack™ delivers accurate, correlated, deterministic, annotated, real-time results. Computational SNP results are reproducible. The results may always be transformed back into BAM file format (or be downloaded into CSV or txt format), for further manipulation by other programs if and when required. Analyses may be scheduled to be launched automatically, or manually upon request.



Key Features

- Massive Scalability (Petabytes)
- Operational ease: ANSI SQL support
- Columnar, massive parallel computing
- Structured; Raw data
- 8PB NAS storage capacities (compressed)
- In-line CDR-specific transparent compression
- Integrates enterprise-wide
- Supports JDBC; ODBC; .NET APIs
- Seamless support for analytical applications, BI environments and ETL tools
- Allows interoperability and native integration with other Genome software solutions

Key Values

Increased reach for researchers:

- Simultaneous searches in multiple genomes
- Fast, parallel GPU-based database for structuring genomic data
- 50X faster compression without any CPU overhead
- Up to 100 times faster searches
- Real-time results

Instant database from day one:

- Option to prepopulate the database with a 1,000 genome/dbSNP/UCSC databases
- Option to get more datasets preloaded

About SQream Technologies

SQream Technologies provides organizations with the most rapid, cost-effective, petabyte-scale big data analytics SQL database available on the market today. With SQream, organizations are able to get the answers they are looking for, quickly, and gain significant industry leadership advantage. SQream Technologies introduces the first patent-pending innovative technology that boosts analytics performance through massive parallel computing, using a GPU-based technology (Graphic Processing Unit). This revolutionary technology delivers up to 100 times faster big data analytics than any other key market player, with scalability capabilities surpassing existing database analytics by orders of magnitude – representing a new era for the Genome, Telecom Cyber, Finance and IoT industries.

SQREAM TECHNOLOGIES

For more information about GenomeStack™, visit www.sqream.com or call +972.3.544.4871