

Event: 17th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB)

Presentation Title: Your large data: query, process, share – in no time

Date: Monday, June 29

Time: 2:45 p.m.

Room: C8

Author: Tomasz Mikołajczyk

The challenges . Efficient management of large data and databases, ranging in terabytes and soon petabytes, is the key pain in all sectors of Bioinformatics today.

- How to design your data management, so that queries can be executed in an instant?
- What techniques can be applied to an existing database that underperforms?
- How to share the data in fully controlled yet flexible fashion?

Case study . Last year we were approached by a commercial company with large production data sets. The multiterabyte database had hundreds of millions of rows. Data mining took 50 minutes on average. Today, they process the same data almost in real time, while 5 minute delay is the worst case.

How was this possible? GridwiseTech is a knowledge company, an independent specialist in scalability. We provide technical integrations and help decision makers crystallize their vision on the IT infrastructure. GridwiseTech comes with Momentum™, the systematic approach to building scalable and/or distributed data management systems that will improve the current state of affairs and guarantee future-proof system efficiency.

Momentum™ comprises methodologies and software for approaching data oriented problems at both the I/O and data processing levels. Momentum™ facilitates the systematic deployment of architectural patterns that result in highly scalable and manageable systems. Such patterns include: data caching, database clustering, distribution, grid/cloud, storage distribution, and more.

Scale out and speed up. Momentum™ Framework is a set of software components to create efficient data management architectures.. For example, to provision data in close-to-real-time, a relational database layer may have an object oriented data grid overlaid. This is just one of many patterns that Momentum includes. In the result, the application delivers results orders of magnitude faster than if connected directly to database.

Share data in an instant . Momentum™ AdHoc allows to securely yet easily share data, applications and other resources within a distributed infrastructure. Giving or blocking an access to your data to a co-worker is as easy as a mouse drag. No technical skills are needed, and – most importantly – no need to call your administrator. AdHoc works well inside a corporation as well in collaborative academic projects.

AdHoc dynamically create, modify, store, and remove access policies on-demand. A policy can be created by linking system users together and granting them access to multiple, geographically dispersed resources in real-time.

AdHoc is currently used for the highly distributed ViroLab infrastructure. ViroLab, a 7th Framework project, is a consortium of 11 hospitals and universities. ViroLab operates a distributed virtual laboratory for HIV research. It allows researchers to assess the efficacy of drugs that treat HIV infected patients based on their patient records stored in 6 separated databases. Available data, connected computational resources and repositories storing experiments are used by researchers and clinicians to examine HIV genetic mutations. AdHoc allowed for the efficient sharing and processing of this data, enabling effective prevention, diagnosis, and treatment of diseases.