ANALYZE GENOMES
by Hasso-Plattner-Institut

Enterprise Platform and Integration Concepts:
Research Group of Prof. Dr. Hasso Plattner

Real-time Analysis of Big Medical Data

MOTIVATION

The vision of the human genome project was born in the early 1980s. One decade later, it was officially started in the U.S. in 1990. Another decade later, a first draft of the humane genome was announced in 2000. In the same period, costs for computer hardware dropped and capacities of main memory and storage systems underwent an exponential growth. Today, DNA sequencing and genome analysis are turned into reality, e.g. in cancer treatments, analyzing gigabytes up to terabytes of data.
With the upcoming trend of e-Health, i.e. supporting medical processes by electronic devices and communications, the amount of data produced keeps growing also in other fields of the medical sector. Patient records and clinical documentations are digitalized, self-tracking fitness devices and applications are getting more popular.
However, analysis of this kind of data nowadays is a time-consuming, manual task or even not possible at all because it lacks appropriate tools. Data management and analysis comes with various challenges, such as huge storage requirements, traditional scanning algorithms are based on reading sequences of characters from files, processing of operational data in databases is only rarely considered, parallelization of processing, etc.

GOAL

Building on our long-lasting experience in applying in-memory technology to selected enterprise challenges, we also focus on processing and analyzing of scientific data sets in real-time. In particular, the applicability of in-memory technology for analysis of medical data will be evaluated. Proof of concept prototypes will be engineered and showed to real-world users in the course of this project.

EXTERNAL PARTNER

The project team will have frequent contact with experts of our cooperation partners:
- Charité - Universitätsmedizin Berlin, and
- SAP AG, Walldorf.

Thus, trips to the headquarters of either both companies are very likely.

CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN

SAP®

ANALYZE GENOMES
by Hasso-Plattner-Institut

SETTING

The project team will work on latest server hardware, in-memory and multi-core technology provided by the "In-Memory Research Laboratory" at our group and the HPI's "Future SOC Lab". Both laboratories build the foundation for HPI's in-memory technology activities. Due to our cooperation with hardware and software vendors, we are able to access high-end hard- and software before it is available for the public market. For example, SAP's in-memory database "SAP HANA", which is optimized for enterprise data management, will be used as technology foundation.

SKILLS

Our external partners provide the required real-world biological and technical input for this project. Thus, we expect you to work with interdisciplinary experts from our project partners. Due to the expected intensive work with database technology, a passed exam in at least one database technology or equivalent lecture is favorable. Furthermore, knowledge in working with either or all of the following development languages is helpful: C++, Python, L, R, Bash, SQL.
In the course of the project, you will be equipped with knowledge about the foundation of in-memory technology and biological skills. We also provide introductions to further technologies, such as SQL, SQLScript, L, and R. However, it might be necessary to investigate additional technologies, which also requires you to dive deep into new areas by yourself or as team.
You will have contact to experts from Charité and SAP to gain additional insights in their work.

WHAT YOU CAN EXPECT FROM US

We will provide you introductions to the relevant fields of research, e.g. genomics or sensor data, and with hands-on experiences, e.g. in in-memory database technology. For that, you will have access to latest server hardware. You will obtain insights in specific software development processes as well as project
management and self-organization methods. Furthermore, you will interact with experts and partner in the corresponding fields.

TEAM STRUCTURE AND KICKOFF

The team will consist of 6-8 students. The project kickoff is scheduled for Tue, Oct 21, 2014.

CONTACT

Please feel free to contact us at "Hasso Plattner High Tech Park" at August-Bebel-Str. 88 or via e-mail.

- Cindy Fähnrich (cindy.faehnrich@hpi.de), room V-1.02
- Dr. Matthieu-P. Schapranow (schapranow@hpi.de), room V-1.01