ANALYZE GENOMES
by Hasso-Plattner-Institut

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

Global Medical Knowledge

Motivation

Selecting the best treatment option for patients suffering from rare diseases, e.g. autoimmune diseases, is still a challenging task for clinicians. Discovery of scientific innovations lacks instant access to patient data, e.g. to test research hypotheses. Current medical knowledge about rare diseases is young and fragmented while biomarkers for meaningful decision taking are not verified, yet.

The aim of precision medicine is to incorporate all available data, e.g. from other patients, historic cases or biobanks, to derive a meaningful decision based on statistical analyses. Medical doctors should be able to access all details relevant for the treatment of a current case, e.g. from similar historic and worldwide patient cases, in a way that guarantees the privacy of each individual. Researchers should be able to test their hypotheses on linked data sources to improve scientific findings. For the first time, the medical documentation would be used to improve decision taking for a current patient and not only for accounting and billing purposes. However, it remains unclear how to identify, link, and explore relevant data from distributed data sources, e.g. across hospitals, countries, and continents, in a timely manner.

Goal

Let us assume a cooperation project between two medical centers that requires the exchange of pseudonymized data for a subset of patients suffering from a specific cancer disease to find the optimal treatment decision. Today, medical data is stored in distributed, heterogenous databases of each of the medical centers. We incorporate latest in-memory database technology to combine data from these individual systems enabling precision medicine. Therefore, we need to define a) a hierarchy of distributed database systems and b) how to connect them. The master database node needs to have access to specific subsets of data of individual partners. It distributes incoming queries to the corresponding databases for further processing and merges individual results from sub databases to form the complete response. Subsets of private database content can be sharable between project partners in such a setup while full control about the complete data set remains local. As a result, all cooperation partners are able to perform analyses on the shared data set only. This approach incorporates no changes of the individual database landscape to enable exchange of data.

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 clinical data will be evaluated. Proof of concept prototypes will be engineered and showed to potential end users in the course of this project.

External Project Partners

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

What we expect from you

We are looking for students, who are motivated to adapt to new research area, such as in-memory database technology and analysis of medical data. You should be hands-on experienced in using at least one programming/developing language, preferable C++ or Python, and one database query language, preferable SQL. Furthermore, a strong understanding of database concepts is beneficial. We expect you to have strong expertise in applying modeling techniques, such as UML or FMC, to exchange knowledge and design decisions. You should be flexible to work on top of existing tools and software and to extend it with your contributions. In addition, you should have communication abilities to collaborate with team and chair members as well as external cooperation partners.

What you can expect from us

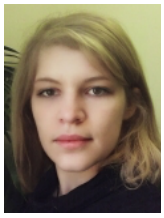
We will provide you extensive introductions to the relevant fields of research, e.g. biology, medicine, genomics, and with hands-on experiences, e.g. in in-memory database technology. For that, you will have access to latest server hardware. You will obtain insights in specific software development processes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

Setting

The project team will work on latest server hardware, in-memory, and multi-core technology provided by the “Enterprise Application Architecture Laboratory” at our group and HPI’s “Future SOC Lab”. The laboratory builds the foundation for HPI’s in-memory technology activities. Due to our cooperations 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.

Contact

Please feel free to contact us, if you have any questions. You can reach us either in person at Campus II, August-Bebel-Str. 88 or via e-mail.



Milena Kraus
milena.kraus@hpi.de
Room V-0.01

Dr. Matthieu-P. Schapranow
schapranow@hpi.de
Room V-1.01