Advanced large-scale data management infrastructure for science

Proposal for PhD Thesis at KIT

Huge quantities of information are produced by scientific experiments worldwide. Data formats, underlying storage engines, and sampling rates are varying significantly. The components of large international experiments are often located in multiple, sometimes hardly accessible places, and are developed by only loosely connected research groups. To handle the distributed and heterogeneous nature of modern experiments, a conceptually new design of data management system is required.

The goal of the project is to propose a new model for storage of large and growing archives of multi-dimensional time series. The model has to be implemented as a flexible and customizable data management framework applicable to a wide range of experimental conditions. It should provide the building blocks required to organize the data flow in scientific experiments and include components simplifying design of easy-to-use web interfaces for data analysis and visualization. The concept should be used for the mobile meteorological experiment KITcube that integrates about 30 complementary devices for atmospheric studies. Application to further experiments is foreseen.

The task requires knowledge of broad areas of computer technology from high-speed parallel computing to database optimization, and web-based visualization. Emerging technologies should be studied to bring faster and more convenient interface to the users. Particularly, WebCL and WebGL are of extreme importance for web-based data analysis and visualization. A sophisticated technique of data preprocessing and storage should be developed to quickly generate previews summarizing large quantities of data. The whole range of client platforms ranging from high-end visualization stations to handheld multi touch smartphones has to be supported.

**Qualification:**
Master in Computer Science, Electrical Engineering, Physics, or Mathematics

**Required Skills:**
- Web technologies
- Database design
- Parallel programming
- 3D rendering
- Mathematical statistics

**Contact:**
Suren Chilingaryan <suren.chilingaryan@kit.edu> Phone: +49 721 / 608 26579
Andreas Kopmann <andreas.kopmann@kit.edu> Phone: +49 721 / 608 24910