Master Thesis *A scalable storage platform for large archives of multi-dimensional meteorological data*

To advance the understanding of turbulence and convection in the atmosphere, KIT designed a mobile setup, KITCube, consisting of multiple meteorological instruments surveying an atmospheric volume of about 10 km side length.

These instruments continuously produce large amounts of multi-dimensional time series data, capturing different aspects of weather situations in a variety of formats. To understand this phenomena, it is crucial to correlate measurements from different instruments with a high spatial and temporal accuracy. Hence, an advance data management and preservation system is required. Traditional relational databases have serious limitations in handling multi-dimensional data efficiently but the newly emerged family of NoSQL databases can provide a better approach by its flexible data model, low latency, simple data query and scalability.

You are expected to evaluate different database engines and find an optimal solution to handle the data produced by KITCube, be it with NoSQL database or as a hybrid system of relational and non-relational database. You should identify the standard data access patterns and benchmark the considered storage engines in each selected scenario. The reliability and scalability in the clustered environment have to be taken into account as well. The goal is to create a data storage system for KITCube to handle large archives of meteorology data. This work will be carried out in close collaboration of the IPE data management group with the IMK meteorologists.

Required Skills: Relational model, SQL, Python. Prior knowledge of NoSQL is a plus.

Experience Gained: Data management, NoSQL databases, Clustering, Meteorology

Data plot by WindTracer

3D plot of wind velocity with Lidar

Contact:
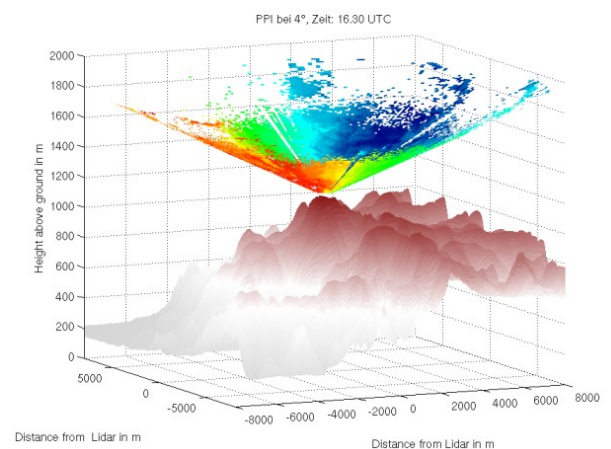Nicholas Tan Jerome <nicholas.jerome@kit.edu>
Phone: +49 721 / 608 25512

Suren Chilingaryan <suren.chilingaryan@kit.edu>
Phone: +49 721 / 608 26579