OpenStack based data management for URDME

Computational systems biology. URDME is a software that generates spatial stochastic realizations of biochemical reaction-diffusion systems on unstructured meshes. The software can be used to study diverse biochemical control systems in computational systems biology. As an example, it has been used to study variability in stem cell differentiation and as a fraemwork to develop many new multiscale simulation methods for stochastic chemical kinetics.

The goals of this project is to design, implement and evaluate a Cloud computing platform for systematic, flexible and efficient analysis of the large amounts of time series data generated by the URDME (www.urdme.org) software for typical modeling projects. The ambition is that such a platform can be used in an observational approach to model exploration of spatial stochastic models of biochemical reaction-diffusion networks.

Cloud computing and storage. The ever-growing use of computational resources by scientific applications such as URDME increases the potential to discover and understand complex phenomena by simulation. Such applications often generate data on a massive scale. It is essential to efficiently analyze that data and also to make it available for longer time in a way that facilitates collaborative exchange of data between researchers. Different data management tools and techniques have been developed to address such requirements, ranging from scalable distributed file systems to storage Clouds for long-term data availability. In this project our aim is to construct an analysis and data management platform for the URDME software based on OpenStack Cloud. The OpenStack Cloud suite is one of the fastest growing Cloud stacks both for computations and storage.

We will formulate the specifics of the research project together with the prospective student based on interests and background. Possible questions to address include:

- Can Openstack's Swift data management system be a good solution for addressing the data analysis requirements of applications such as URDME?
- How do we best design and develop an improved data analysis platform that is adjacent to the available data and provide answer to the queries rather than raw data?
- Can we implement techniques to identify interesting patterns in the data generated by the application?
- How can the proposed platform be made generic enough so that other applications besides URDME can benefit from this work?

You will:

Develop your ability to conduct independent research in an academic environment. Learn about spatial stochastic simulations in systems biology and about the challenges faced in large scale computing projects. Get practical experience with the OpenStack private Cloud.

Collaborations

We will collaborate with Dr. Brian Drawert at the University of California, Santa Barbara.

For more information, contact:

Andreas Hellander, andreas.hellander@it.uu.se, www.andreashellander.se
Salman Toor, salman.toor@it.uu.se