Contribution ID: **18** Type: **Full presentation: short (15 mins.)**

Managed Services for Accelerating Science in the 2020's - a SURF perspective

Tuesday, 3 November 2020 13:15 (15 minutes)

In this talk we will present Spider, a managed service by SURF. Spider is a versatile high-throughput data-processing platform aimed at processing large structured data sets. It runs on top of our in-house elastic Cloud. In combination with superb network and hierarchical storage this allows for processing on Spider to scale from many terabytes to petabytes.

In recent years Cloud technology has given IT users and providers near-limitless possibilities in terms of customizable and dedicated infrastructure for computing. However, in parallel researchers have faced increased publication pressure (publish or perish), explosive growth in data and a diversion of funding towards personal grants (rather than institutional/structural funding). This has decreased the effective time that researchers can afford to spend on designing IT solutions for their scientific problems and steadily widened the gap in IT knowledge between researchers/research-institutes and IT providers. Hence, although using Cloud technology to deploy tailored computing environments scales technically, this scalability also requires support, automation and fault tolerance which bring many new challenges that researchers cannot tackle and often are not interested in.

A new balance has to be found to more fully support researchers with IT intensive problems. Accepting the above paradigm shift, this means that any effective solution, for sustainable data processing, has to go beyond the virtualization layer. We believe that this solution can be found in providing researchers with managed, persistent and where possible shared services. Such a solution would not only accelerate science but will also reduce its carbon footprint.

In our model the researcher focuses on the the scientific algorithms and the IT provider is responsible for the infrastructure and the platforms built on top. These platforms are built on generic solutions and only where required are tailored to the needs of a particular user community. In this vision the infrastructure itself remains Cloud-native to preserve the proven strengths of this technology such as rapid deployment, robust adaptation and dynamic scaling.

Managed services we believe pave the road towards unburdening researchers and allow users and providers to focus on their respective strengths and achieve increased synergy. Furthermore, armed with modern technology (e.g., containers, virtual environments, shared & local filesystems, role-based access, collaborative spaces, private nodes/partitions and secure networks) managed services can flourish and fulfill the requirements of a broad and diverse set of research communities.

Spider combines these technologies in an effort to provide a low-threshold, managed data processing platform that appeals to a broad set of scientific disciplines. Here we discuss its technical setup, the possibilities for customization, its potential within a distributed computing federation and share some of the many current use-cases. The deployment and integration of managed services on the EGI infrastructure does not feature within the current EGI service model and through this talk we also aim to start a discussion on the need for including such services as part of this infrastructure and the European Open Science Cloud.

Primary author: OONK, Raymond (SURFsara BV)

Presenter: OONK, Raymond (SURFsara BV)

Session Classification: Cloud computing - Part 1