Introduction

Scientific Workflow Management Systems (SWFMS) provide a framework to compose, execute and monitor scientific applications. Provenance information need to be recorded to enable the reproducibility of scientific applications. System metrics allow the monitoring of the underlying infrastructure and the system availability. However, provenance and system metrics differ in scope of information and are provided by different sources, which makes the integrated analysis difficult and time-consuming.

Context-aware Information Integration and Exploration Framework

We propose a context-aware information integration and exploration framework to effectively integrate provenance with the system logs. It enables users to:

- Analyze the execution time of service-based scientific workflows with the help of provenance information
- Detect bottlenecks that cause workflow performance degradation
- Visualize results

Use case

We constructed a simple workflow using Taverna made of six tasks spread over three VMs:
- VM A hosts tasks LW1, CPU2 and Mem2
- VM B hosts tasks CPU1 and Mem1
- VM C hosts CPU3

Summary

- Our solution allows scientists or service developers to analyze the execution of service-based scientific workflows and to visualize possible bottlenecks related with the infrastructure by getting a detailed view of the resource usage of each workflow task.
- Infrastructure administrators, service developers or application controllers may configure the provisioned virtualized infrastructure.
- We aim to use statistical and AI algorithms to detect and predict possible workflow execution failures, and also support more complex workflows.

This work is partially supported by EU H2020 ENVRIPlus, ENVRI-FAIR and ARTICONF projects and performed at the University of Amsterdam.

Contact: dr. Zhiming Zhao (z.zhao@uva.nl)