Linking provenance with system logs: a context aware information integration and exploration framework for analyzing workflow execution

Elias el Khalidi Ahanach, Spiros Koulouzis, Zhiming Zhao,
Informatics Institute,
University of Amsterdam,
Amsterdam, The Netherlands,
elias.el.khaldi@gmail.com, {S.Koulouzis,Z.Zhao}@uva.nl

ABSTRACT

When executing scientific workflows in a distributed environment, anomalies of the workflow behavior are often caused by a mixture of different issues, e.g., careless design of the workflow logic, buggy workflow components, unexpected performance bottlenecks or resource failure at the underlying infrastructure. The provenance information only defines data evolution at the workflow level, which does not have an explicit connection with the system logs provided by the underlying infrastructure. Analyzing provenance information and apposite system metrics requires expertise and a considerable amount of manual effort. Moreover, it is often time-consuming to aggregate this information and correlate events occurring at different levels in the infrastructure. In this paper, we propose an architecture to automate the integration among the workflow provenance information with the performance information collected from infrastructure nodes running workflow tasks. Our architecture enables workflow developers or domain scientists to effectively browse workflow execution information together with the system metrics, and analyze contextual information for possible anomalies.

REFERENCES