Abstract

Scientific workflow management systems are providing the ability to manage and query the provenance of data products. Understanding the workflow lifecycle is essential because scientific workflow often deals with proprietary modules as well as private or confidential data, such as medical or health information. Comparing the workflow runs and understanding the difference between them is thus important. This paper discusses the i) Workflow lifecycle and challenges ii) Research issues in provenance for the scientific workflows. It is aimed to provide overview of scientific and business workflows. iii) Privacy issues in a scientific workflow—provenance privacy, data privacy and module privacy. In short the data provenance is an overloaded term that has been defined differently by different people. Data provenance can be examined from different point of perspective such as semantics. By the execution of two workflows with same specification leads to problem of differencing the provenance of two data products. At the end the paper discusses about the challenging task related to increasing number of workflows in the cloud environment is that managing various workflows, VMs & workflow execution on VM instances. Resolving workflow execution is very important because
workflow and cloud became most popular in cyber infrastructure projects. Workflow has the capability to build flexible applications and cloud provides scalable and economic services. Hence by clearly defining dependent services the proposed workflow architecture Workflow Flow as a Service (WFaaS) proves to manage large number of workflows and VMs. The workflow architecture proposed in the cloud has been concluded by discussing on future work.

References

29. Workflow as a Service in the Cloud: Architecture and Scheduling Algorithms Jianwu Wang1, Prakashan Korambath2, Ilkay Altintas1, Jim Davis3, Daniel Crawl1

Index Terms

Computer Science
Distributed Systems

Keywords

Data Provenance, Workflow, Scientific Workflow Management, Workflow as a Service (WFaaS), Cloud, Metadata.