

# A Study of Workflow Management Systems in the Cloud Environment

Prerana K A

Dept. of Computer Science and Engg.  
Acharya Institute of Technology  
Bangalore, India 560 107  
Email: prerana.mtc.15@acharya.ac.in

Naidila Sadashiv

Dept. of Computer Science and Engg.  
Acharya Institute of Technology  
Bangalore, India 560 107  
Email: sadashiv@acharya.ac.in

**Abstract**—Cloud computing ensures service flexibility, scalable storage of data, and high performance to users whose numbers are increasing by the day. It has enhanced its role in distributed computing systems by offering superior internet services that augment and complete functionalities of grid computing, web-enabled distributed computing, and peer-to-peer networks. Workflow for analysing data along with pre-processing tools, distributed datasets, knowledge models, and data mining algorithms are used for eliciting information of value from data, which is often a complicated process. Since cloud systems offer storage and processing services that are scalable, along with a software platform for creating a data analysis environment, they could be efficiently utilised to manage workflows for data analysis. In this paper, the various management systems of visual workflow, which are suitable for a cloud environment, are discussed.

**Index Terms**—Cloud Computing; Grid Computing; Workflows

## I. INTRODUCTION

Digital repositories contain large amounts of data that are massive and distributed, and computing platforms, suitable data analysis techniques, scalable storage are necessary for its analysis. Parallel and distributed data analysis methods or techniques are used to handle big data repositories. Data analysis environments are used by scientists and professionals to run complex simulations, endorse models, study, and share results with co-workers across the globe. Grids and clouds come with high performance computing abilities that allow scientists and professionals to run increasingly more complicated applications to connect with and work through huge data repositories, and experiment scientifically on disseminated computing stages. Workflow enacting systems are a popular technology to flexibly define and enact complex data processing tasks in business and e-science [21], [25], [26]. The entire or partial business process automation in which information, documents, or tasks are moved between participants, conforming to a collection of procedural rules, is called a workflow. Workflows for data analysis integrate pre-processing tools, distributed datasets, knowledge models, and data mining algorithms for eliciting information of value from data [10]. They are designed to bring about some data transformation. Many such systems that utilise a dataflow model for presenting computations could possess a built-in

control programme that schedules and sequences each step from a central location [1]. Workflows are first defined and then saved and recalled either for modification or re-execution. Such an important benefit permits consumers to establish typical forms, while re-utilising them in varied scenarios. The rest of this paper is organized as follows.

A study of existing works is presented in Section II that discusses about visual workflow management system. Table I highlights the state-of-the-art. Section III presents the concluding remarks and scope for future work.

## II. EXISTING VISUAL WORKFLOW MANAGEMENT SYSTEM

### A. Galaxy

Galaxy is a general tool to carry out analysis of data and computations for a wide array of life science applications. It provides a browser-based user interface and runs on Linux/Unix based servers. With an updated web browser, the end user can access Galaxy from any computer operating system, e.g. Google Chrome, Internet Explorer, Mozilla Firefox, etc. It has been formulated mainly to permit the implementation of the software tools on the local systems while providing simplified data access to biological resources located remotely [2], [4]. A vast repository of tools and pre-prepared data processing scripts are required for its installation. Workflows, using JSON format, are represented in its own language, while it efficiently executes on local infrastructures.

### B. Swift/T

To enhance the Galaxy's effectiveness for interfacing with computing systems on a large scale and, in running, parallel workflows, it has been integrated with Swift/T. The scripting language Swift is a parallel language, and was developed to run several occurrences of simple sequences simultaneously on various distributed and parallel resources. It is a powerful software designing model that is used to express and coordinate the type of workflows created by Galaxy users. Rapid creation of complex application flows and speedy execution using a variety of remote computation systems is Swift's speciality [14], [17]. The runtime consists of a group of services that achieve parallel processing of Swift/T scripts,