International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 4– No.9, December 2012 – www.ijais.org

Workflow Management in Cloud Computing

Monika Bharti
M.E. student
Computer Science and
Engineering Department
Thapar University, Patiala

Anju Bala
Assistant Professor
Computer Science and
Engineering Department
Thapar University, Patiala

ABSTRACT

Cloud computing is a paradigm that provides demand service resources like software, hardware, platform, and infrastructure. Under cloud environment, workflow is an emerging technique for future scalable applications. This paper discusses the various tools for generating workflow and these tools have been compared on the basis of operating system, databases, architecture and so on. The application on workflow is generated with Pegasus tool which can be further deployed on its compatible cloud platforms like Eucalyptus, Amazon EC2, Open Stack etc.

General Terms

Cloud Computing, Workflows.

Keywords

Cloud Computing,Pegasus,Workflows.

1. INTRODUCTION

Cloud computing has recently emerged as a new paradigm for hosting and delivering services over the Internet. Cloud computing is attractive to business owners as it eliminates the requirement for users to plan ahead for provisioning, and allows enterprises to start from the small and increase resources only when there is a rise in service demand [1]. It provides web-based software, middleware and computing resources on demand. The research issues for cloud computing are security, load balancing, resource provisioning, energy efficiency, workflows and so on.

This paper focus on workflows, which works behind cloud to manage resources, various clients, cost constraints. The concept of workflow is proposed by fixed work procedures with conformist activities. The tasks are divided into subtasks, roles, rules and processes to execute and observe the workflow, workflow system boost the level of production of organization and work efficiency. Various types of workflows are business workflow, abstract workflow, concrete workflow, scientific workflow and so on. Business workflow allows controlled flow of execution and simplifies workflow management. It provides support for security, reliability, transactions, and performance. Its performance can be increased by use of faster server. Its workflow lifecycle is design, deployment, execution, monitoring and finally refinement. Scientific workflow supports for large data flows and need to do parameterized execution of large number of jobs. It is also to monitor and control workflow execution including ad-hoc changes. The input given to workflow is written in languages like Java, Perl, Python and the output generated is the workflow. These workflows are managed and coordinated by workflow management system, which provides the end user with the required data and the appropriate application program for their tasks. It allocate tasks to end-user based only on the performance of constraints like control flow, data flow, transition conditions or pre- and post-conditions. The issues that arise with workflow and its management are workflow scheduling, fault tolerance, energy efficiency and so on. Workflow scheduling maps and manages the execution of inter-dependent tasks on the distributed resources.Fault tolerant is when a system's service failure can be avoided when faults are present in the system.

In this paper, workflow tools and implementation of workflows using Pegasus tool is discussed. Section 2 represents challenges for workflows in cloud computing. Section 3 shows detailed comparison between various workflow tools whereas Section 4 represents experimental results of the generated workflow using Pegasus tool.

2. RESEARCH CHALLENGES FOR WORKFLOWS IN CLOUD COMPUTING

2.1Security

While running software and keeping data on virtual machine appears daunting to many. Well-known security issues such as data loss, phishing pose serious threats to organization's data and software [2].

2.2DataLock-IN

The customer cannot easily extort their data and programs from one site to run on other site. The solution is to standardize the API's, so that the SaaS developer could deploy their services and data across multiple providers [3].

2.3Reliability and Performance

Performance and availability of the applications are important criteria defining the success of an enterprise's business. However, the fact that organizations lose control over IT environment and important success metrics like performance and reliability. This are dependent on factors outside the control of the IT organizations makes it dangerous for some mission critical applications [4].

2.4Storage

Existing workflow systems often rely on parallel and distributed file systems to ensure that tasks landing on any node can access the outputs of previous tasks that may have executed on another node. It is highly inefficient, and time-consuming. In addition, it may be costly in a commercial cloud that charges by the number of bytes transferred. The solution is to deploy a temporary shared file system in the cloud as part of a virtual cluster, but it is complex, potentially costly, and needs to make sure that desired outputs are transferred to permanent storage. A better solution would be a permanent, scalable, parallel file system similar to what existing clusters and grids use[5].

3. COMPARISON BETWEEN VARIOUS WORKFLOW TOOLS

The various tools for the workflows being compared as shown in Table 1 , are YAWL, OOZIE, UGENE, Open Bonita Solution and so on. UGENE is free open-source cross-platform, integrates number of biological tools and algorithms, provides both graphical user and command line interfaces [6]. The business processes can be graphically modified using Bonita Studio. The processes can also be connected to other pieces of the Information System to generate an autonomous business application accessible as a web form[7].

OrangeScape is browser based development environment and provides 4 design perspectives: Model design, Form design, Process design and Action design and is meant for building process oriented business applications example building SaaS applications that runs on OrangeScape Cloud [8]. Google App Engine, GAE is a platform as a service (PaaS) cloud computing platform used for developing and hosting web applications in Google-managed data centers [9]. A new workflow language called YAWL (Yet Another Workflow Language) offers comprehensive support for the

S.No	Name of tool	OS	Language	Year	Founder	Description	Architecture	Database	Companies
1	**UGENE**	Cross platform	C++, QtScript	Dec 2011	Unipro	Integrates tools and algorithms	Client-server	NCBI, PDB, UniprotKB /Swiss-Prot	IBM
2	**Bonita Open Solution**	Cross platform	JAVA	Jan, 2011	French National Institute For Research in CS	Creates high-tech workflows and spreadsheets	Client-server	ERP, ECM	VENTECH
3	**Orange Scape**	JEE Application server	JAVA	2003	Google	shared-Processing multitenan-cy models	Client-server	Oracle, MYSQL,IBM DB2	WIPRO
4	**Google App Engine**	Windows	Python, APIs, URL fetch	2008	Google	Allows user to run web application	Client-server	Python, java	IBM, MICROSOFT
5	**YAWL**	Cross platform	XML Schema, XPath and XQuery.	2002	Wil van der Aalst, Arthur ter Hofstede	based on the one hand on Petri nets and on the other hand on the well-known Workflow Patterns.	Client-server	MY SQL	GECKO
6	**OOZIE**	Cross platform	hpdl	2006	Team of Cape Town, South Africa	Java Web-Application that runs in a Java servlet-container	Client-server	Apache Tomcat	Microsoft

7	**KAAVO**	Cross platform	Java, PHP	2007	JAMAL MAZHER	Application centric approach to the management of cloud	Application-centric	Amazon, Rackspace	IBM
8	**PEGASUS**	LINUX , WINDOWS	JAVA, PERL, PYTHON	2003	Ewa Deelman	translate complex computational tasks into workflows	Client-server	XML, JAVA	AMAZON EC2

Table 1: Comparison of Various Workflows Tools in Cloud Computing

control-flow patterns and has a proper formal foundation. It also supports unique support for dynamic workflow through the Worklets approach. Workflows can thus develop over time to meet new and changing requirements [10]. Oozie is a Java Web-Application that runs in a Java servlet-container. Oozie workflow is a collection of actions arranged in a control dependency DAG (Direct Acyclic Graph), specifying a sequence of actions execution, specified in hPDL (a XML Process Definition Language) [11].Kaavo provides a framework to mechanize the deployment and run-time management of applications and workloads on multiple clouds. It takes a top-down application-centric approach for deploying and managing applications in the cloud. [12]. Pegasus Pegasus consists of a set of components like The Pegasus Mapper, Execution Engine and Task Manager that run and manage workflow-based applications in different environment, including desktops, clusters, grids, and now clouds. It bridges the scientific domain and the execution environment by automatically mapping high-level workflow descriptions onto distributed resources [13].

4. EXPERIMENTAL RESULTS

After this comparison, work is done on Pegasus and the workflow is generated. Pegasus is a configurable system for mapping and executing abstract application workflows over a wide range of execution environment including a laptop, a campus cluster, a Grid, or a commercial or academic cloud. Today, Pegasus runs workflows on Amazon EC2, Nimbus, Open Science Grid, the TeraGrid, and many campus clusters. One workflow can run on a single system or across a heterogeneous set of resources [15].`

Case 1: Data Flow Diagram for working of Pegasus
The user input the workflow code to the Pegasus WMS where processing is done by the system using its own database and finally the html link is generated which can be copied to web browser and then graphs and plots can be viewed(Fig 1).

Case 2: Use-Case Diagram for Pegasus
The workflow code given by user is submitted to Pegasus WMS. At the same time, user can choose the environment like cloud, grid, desktop and campus cluster, where he can deploy its workflow depending upon Pegasus tool compatibility with the environment (Fig 2).

Case 3: Design of Workflow
The design of the workflow is generated using Pegasus Workflow Management system. The design shows the various sections of Thapar University. It is broadly divided into school, academics and centres. Academics further tell about various departments of engineering like computer science, chemical, electronics communication, mechanical and biotechnology environmental and science (Fig 3).

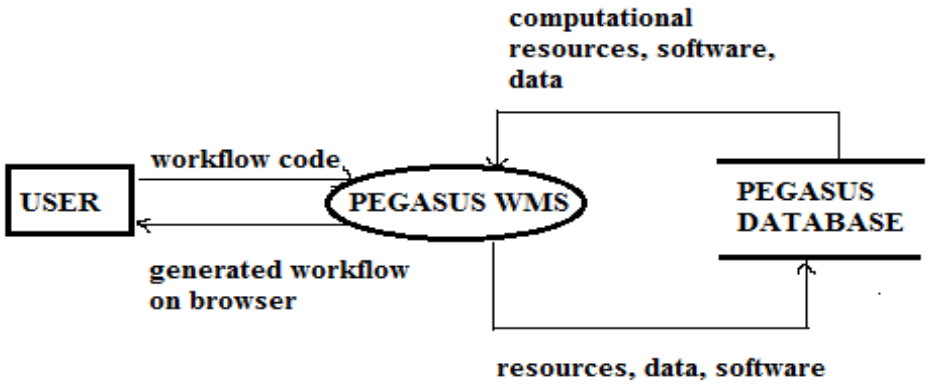

Fig 1:Dataflow Diagram for Pegasus WMS

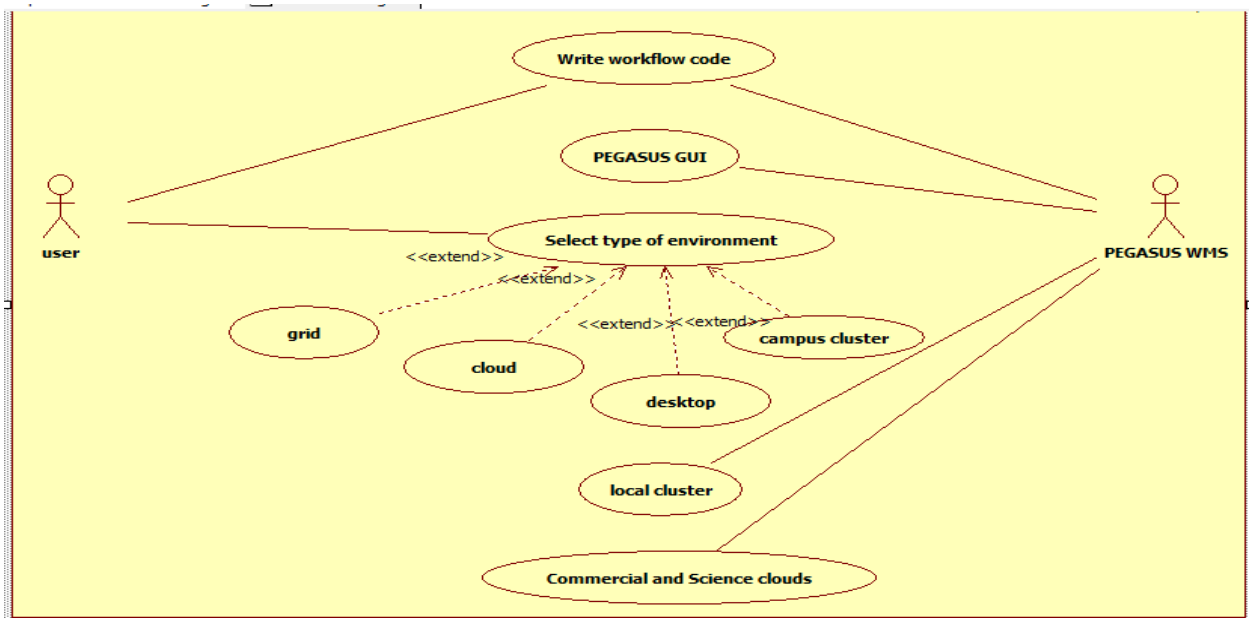

Fig 2: Usecase Diagram for Pegasus WMS

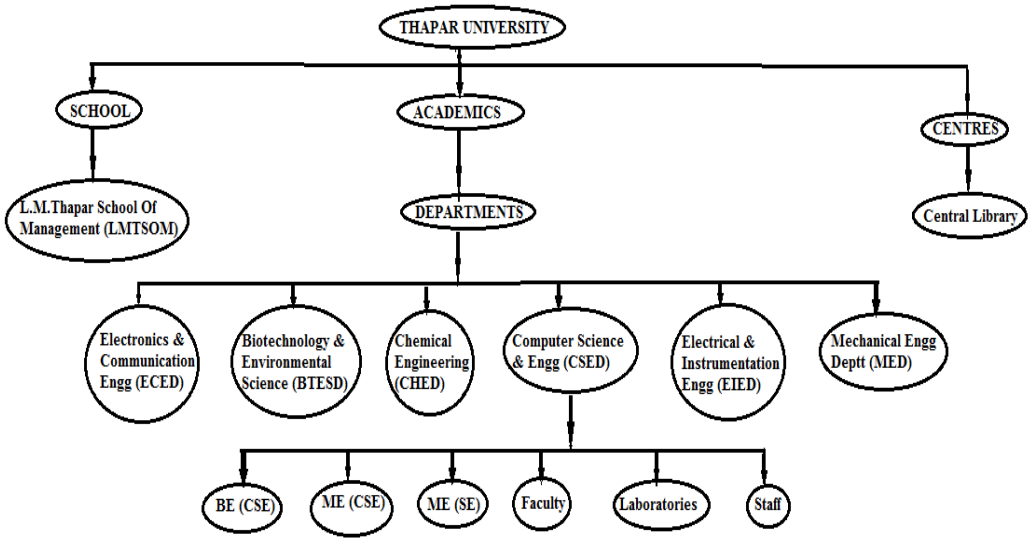

Fig 3: Design of the solution

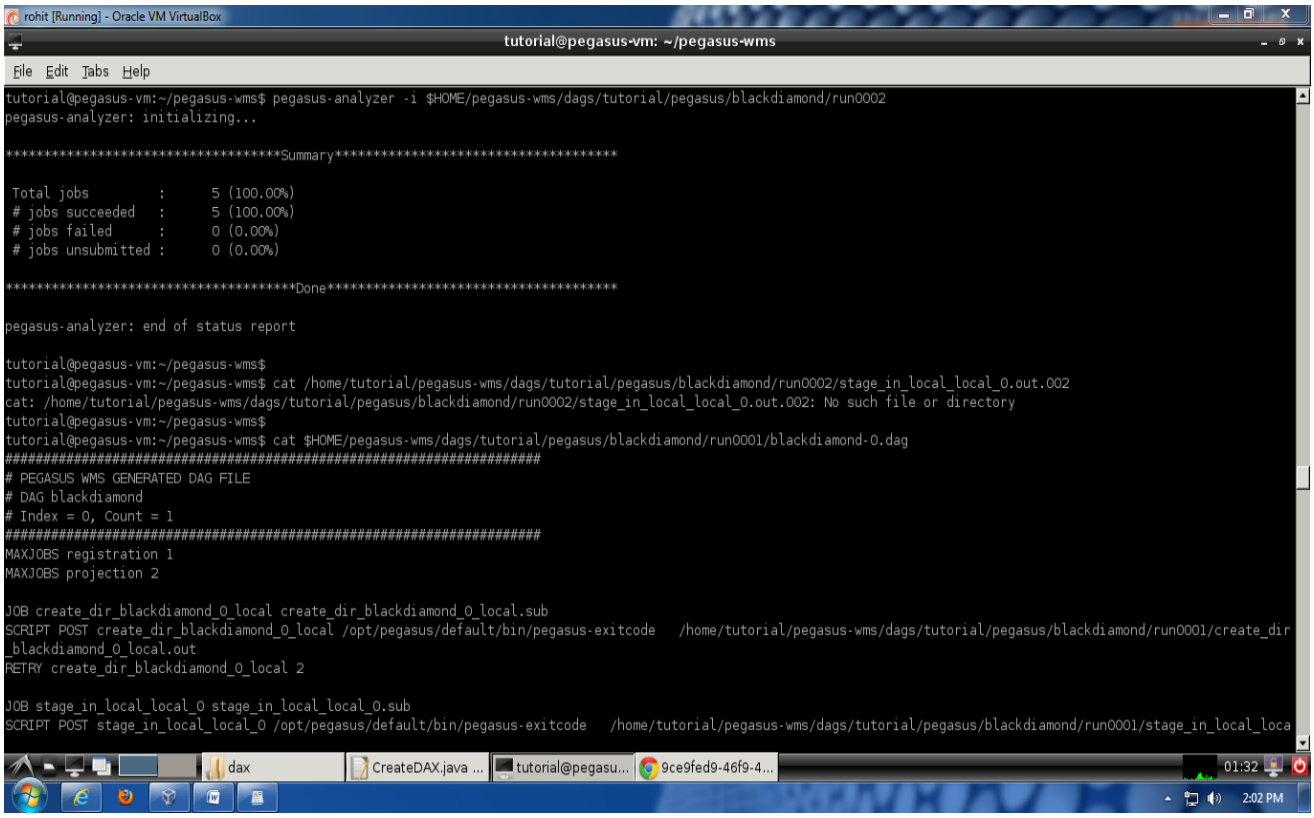

Fig 4: Analysis of jobs

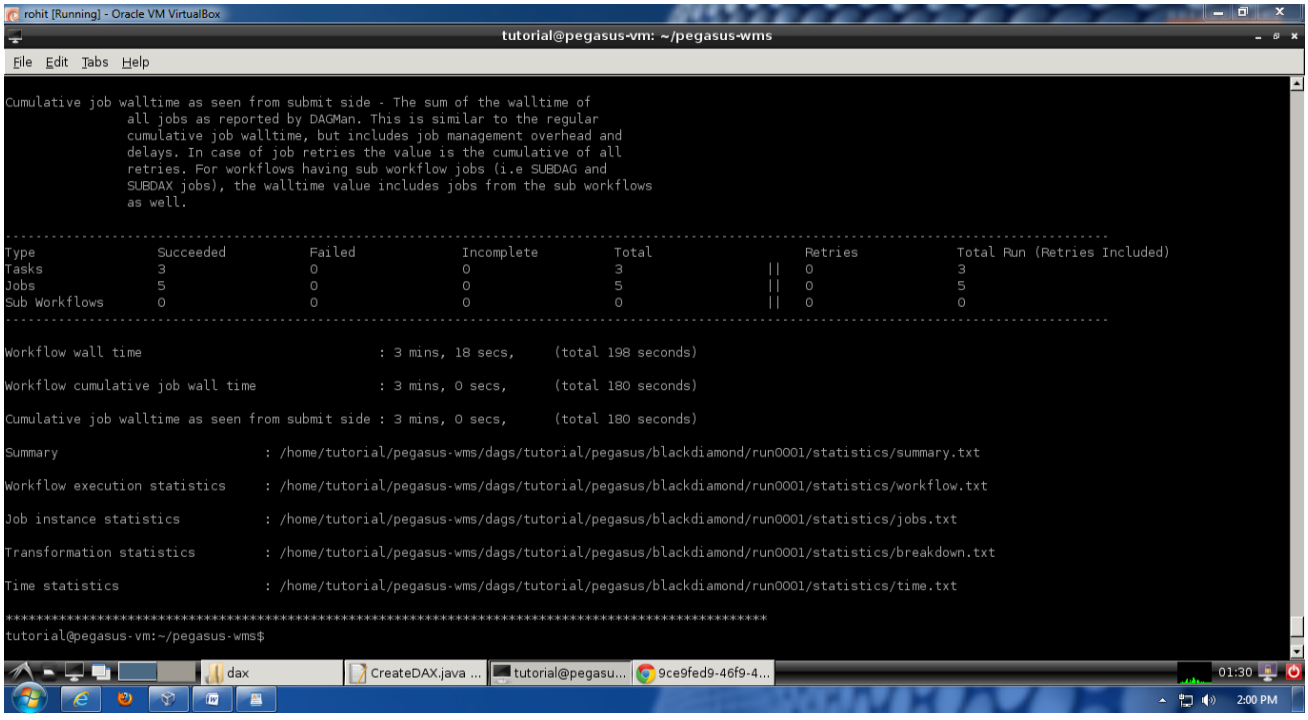

Fig 5: Scheduling and Monitoring of jobs

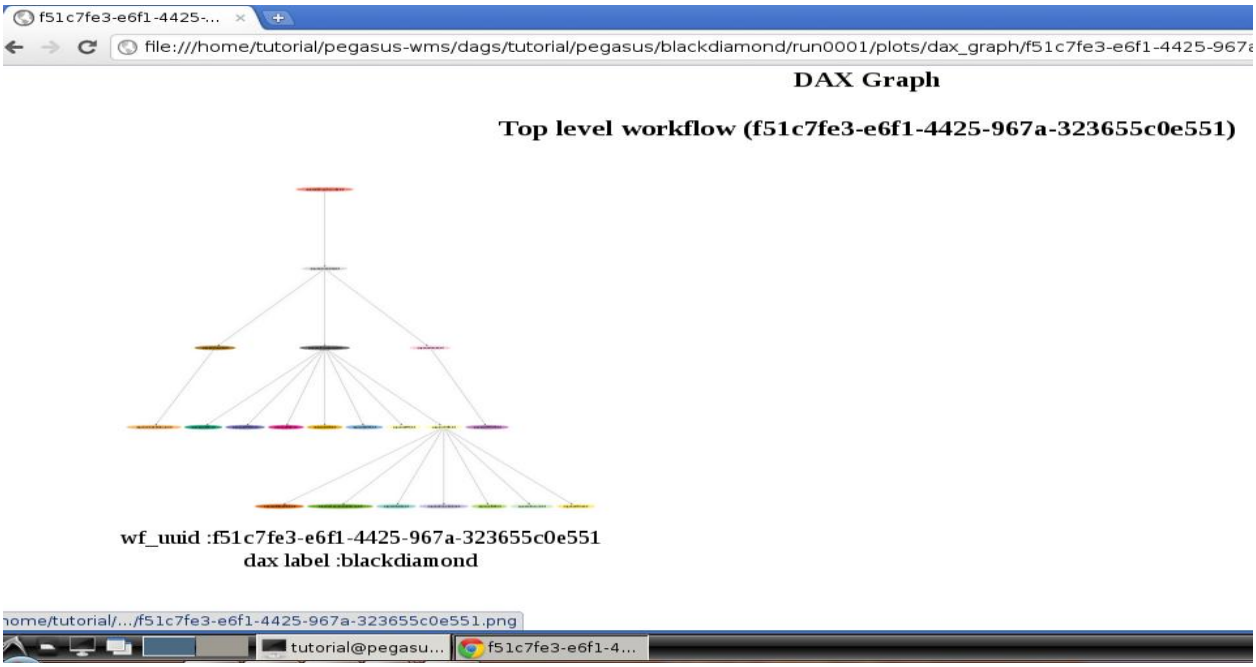

Fig 6: DAX Graph

Case 4: Analysis of jobs
It shows that a couple of jobs are running under the main Dagman process as shown in Fig2. It keeps track if jobs are succeeded. Failed or unsubmitted. The detail for each job is shown in the Fig 4.

Case5: Scheduling and Monitoring of jobs
It contains information about the workflow run like total execution time, failure of job. It gives complete details about workflow runtime, cumulative workflow runtime, total jobs, jobs succeeded, jobs failed, and jobs uncommitted as shown in Fig 5.

Case6: DAX Graph

The output in the form of DAX graph can be generated using command shown on the terminal. It results into Dag graph, DAX graph, Workflow Environment, Workflow execution Ghaint chart (per workflow), Host over time chart(per workflow), Invocation Breakdown chart (per workflow and across workflow). The output of DAX Graph is shown in Fig 6.

5. CONCLUSION

Workflows in Cloud computing plays important role in managing and coordinating tasks and workflow management system works under workflow to overcome constraints like control flow, data flow, transition conditions or pre- and post-conditions. In this paper, various tools have been discussed and workflow is generated using Pegasus workflow tool. Workflow generated will be deployed on compatible cloud environment like Eucalyptus, Amazon EC2. Once workflow is deployed, workflow scheduling will be next step using condor.

6. REFERENCES

[1] J. Deng, Scott C.-H. Huang, S. Han Yunghsiang, and J.H. Deng," Fault-Tolerant and Reliable Computation in Cloud Computing, Intelligent Automation", Inc., Rockville, MD, USA, 2010.

[2] Y. Chen, V. Paxson, and R. Katz, "What's New About Cloud Computing Security?," 2010.

[3] D. DeWitt and M. Stonebraker, "MapReduce: A major step backwards". DatabaseColumn Blog. http://www.databasecolumn.com/2008/01/mapreduce-a-major-step-back.html.

[4] J. Gideon and E. Deelman, "ScientificWorkflows in the Cloud", Gideon Juve University of Southern California, Marina del Rey, CA e-mail: juve@usc.edu and Ewa Deelman University of Southern California, Marina del Rey, CA e-mail:

[5] Unipro UGENE User Manual Version 1.10,December 29, 2011

[6] http://en.wikipedia.org/wiki/Bonita_Open_Solution

[7] http://en.wikipedia.org/wiki/OrangeScape

[8] http://en.wikipedia.org/wiki/Google_App_Engine

[9] YAWL-USER MANUAL , 2009 The YAWL Foundation

[10] http://en.wikipedia.org/wiki/Oozie, http://archive.cloudera.com/cdh/3/oozie/DG_Overview.html

[11] http://en.wikipedia.org/wiki/Kaavo

[12] Pegasus-user-guide.pdf, http://pegasus.isi.edu/wms/docs/3.1/pegasus-user-guide.pdf

[13] http://pegasus.isi.edu/wms/docs/3.1/pegasus-user-guide.pdf