

The Systematic Study of Scientific Workflows in Distributed and Cloud Environments: A Review of Privacy Issues in Workflow Provenance, Challenges and Opportunities

Shridevi Erayya Hombal
Data Analytics, IIIT Bangalore
Bangalore, 560100, India

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.

General Terms

Privacy, Confidential, flexible applications

Keywords

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

1. INTRODUCTION

The challenging aspect in distributed and heterogeneous environment is to “design and manage” the scientific workflow. Scientific workflow acts as a main concept for enabling science in terms of scientific analysis, the number of scientists required to collaborate to resolve the scientific problem. Workflow provides a simple representation of complex analyses composed of various models. Workflow also provides the facility to trace back to get to know about the data from where it is derived. Workflow can be shared with other experts or students or researchers. Consider an example of vacation to illustrate the workflow system how it works: Once the employee entered preferred dates, the department

manager automatically receive an approval form through a web browser. If the request is approved then control flow goes to the HR system. The next step is to inform the employee about where the workflow instance can be completed.

Data Provenance: It is the term which generates the obvious question of “why” and “where”. Both questions deal with a solution of discovering or tracing the path to get to know about the source from which the data came. Basically provenance is valuable and it is hard to collect. Other way of looking the provenance as metadata, contains experimental workflows, annotations and also workflow results. The best initiation of provenance is from data flow workflows. Querying of provenance data is needed to reveal the private information.

The provenance is needed for workflow, distributed, grid computation and also in bioinformatics and other curated scientific databases. Queries over applications, which includes the databases, storage systems or the provenance data set typically answers for the question like what are the input data or steps that led to the creation or modification of data item. Provenance is incomplete and it cannot be shared across applications. This challenge initiates the research of exploring the separate semantics for the provenance that can be agreed upon by people from different domains. A research on discovering an ontology which clarifies the semantics of data provenance is needed. For example in case of differential privacy [1], which involves the aggregate queries, these facilitate to get the private information about the individual from the data set by adding random noise. How to use this differential privacy on provenance is not clear because queries on provenance data is not a type of aggregate queries and also adding the random noise on the provenance data is of little use. Making Availability of complete provenance data leads to privacy concerns. Because the data might also contains the sensitive information. There is always a trade-off existing between the amount of information that can be revealed and the privacy guarantees of the modules of the workflow. A careful study of different activities and the connections between these activities needed to provide the privacy and security for the provenance data. Basically the provenance data maintains two type of information those are metadata and the processes which transform the raw data into new useful information.

2. LITERATURE SURVEY

Workflow technologies are meant for integrating the distributed services. Because of availability of set of tools and frameworks choosing the right tool is a challenging aspect. Researchers are not aware of the technologies; hence this paper presents the survey of existing workflow technologies in the business and scientific domain.

Service oriented architecture is basically used to build the applications by creating the independent services; these services provide a greater degree of flexibility and scalability. There are 2 types of services:

1) Loosely Coupled Services: The client invokes the remote service but they might not be aware of inner functionality.

2) Traditional Coupled Services: Different components are bound to one another, sharing the libraries, semantics and often state; making it difficult to evolve the applications.

For a given problem, SOA provides a solution, whereas for a specific task or goal, it needs co-ordination of services. This co-ordination of services is done by Workflow architecture.

2.1 Business Workflow Technology

Workflow technology was first adopted by the business community. Workflow implementation involves 2 approaches:

1) Service Orchestration: It refers to an executable business process which involves both the internal and external services. The standard business process language is Business Process Execution Language (BPEL)[2]. One more language for analysing the workflow patterns is Yet Another Workflow Language(YAWL)[3].

2) Service Choreography: It describes the collaboration between the services to achieve the common goal. All the processes are treated equally, in a peer-to-peer fashion. Choreography works on message exchange, each services involved in the business process knew about their partners and have the responsibility to perform the operation.

The difference between the above two approaches is that the service orchestration basically defines the process flow between services involving one participant, whereas service choreography defines the sequence of messages involving multiple parties.

2.2 Scientific Workflow Technology

The Workflow technology can be applied to e-Science. Scientists need advanced tools, to prove the scientific hypothesis, because business workflows are more like traditional programming language. Scientific experiment involves set of steps these steps are captured by designing the workflow in order to reach the goal. In general the scientific workflow [4] system provides an environment for analysis, visualisation and simulation. Scientific workflow need to support user-driven, prototypical and incremental approach to workflow composition. The scientific workflow is data flow oriented whereas business workflow emphasis on control flow patterns. The workflows become a valued commodity because of iterative and lengthy design process. The output of workflows can be used as basis for future research. The advantage of designing the workflow is reusability and reproducibility. In order to reproduce the workflow, the provenance information must be recorded which includes from where the data originated, what parameter settings were used and which components were used. The scientific workflow systems contain set of frameworks and languages to allow the scientists to work on workflow systems. Below are listed the survey of workflow systems.

- **Kepler:** It is dataflow-oriented workflow architecture. Kepler [7] is an actor oriented design. It provides a platform for designing, executing, reusing, sharing, evolving, and archiving the scientific workflows. The scientific workflows produced by kepler are directed graphs. In kepler the nodes are called "Actors" and the edges are called "channels". Actors are web

services or data base calls etc. The actor also contains the data wrapper used to access the data structure. Actors are reusable components. In order to achieve some specific goal the kepler targets the use of a workflow metaphor to organize the computational tasks.

- **Taverna:** It provides loosely coupled, transparent platform for the scientists to perform data intensive in-silico [5] experiments. In Taverna, Provenance [6] is having its importance such as it provides the information such as what services were used and who conducted the experiment. Basically the tavern is service oriented architecture.

- **Triana:** Triana [8] is inherently flow based. In Dataflow data arriving at component triggers execution whereas in control flow control commands trigger execution. Data or control messages sent from sender to receiver causes receiver to execute. Triana is not only providing the problem solving environment but also acts as a test application for GridLab project [9]. The purpose of designing the Traiana is to define, analyse, manage, execute, process and monitor the scientific workflows. The triana provides toolkit to compose the workflow by dragging the tools on to a workspace. Triana also distribute the part of workflows to remote machines with the help of peer to peer network. Triana supports different languages by using different workflow readers/writers including: Directed Acyclic Graph(DAG), Web Services Flow Language(WSFL), Business Process Execution Language(BPEL).

- **Pegasus:** It can be used to map the scientific workflows to distributed resources such as Grid. It allows the users to represent the workflows at an abstract level without needing the information of execution system. Pegasus [10] uses the proprietary language, DAX at the abstract level which is an XML representation of DAG. The abstract workflows are independent of any resources. This feature allows the user to design the workflow without considering which resources to use. For improving the application performance Pegasus attempts to find a mapping of the tasks to the available resources.

- **GridNexus:** A graphical system for creating and executing the workflows in a grid environment. GridNexus [11] allows the user to analyse the complex processes by building directed acyclic graph. It solves the complex tasks by enabling the composition of larger software modules from the smaller software modules and it separates the graphical user interface from the execution environment of the workflow to increase the flexibility. The workflow which is designed using the GridNexus need to converted to XML based language. Once the workflow creation is done then it can be executed locally or remotely because the GridNexus separates the GUI from the execution environment of the workflow.

- **DiscoveryNet:** DiscoveryNet [12] is an EPSRC funded project. Building the high performance applications requires protocols and grid technologies.

3. WORKFLOW LIFE CYCLE

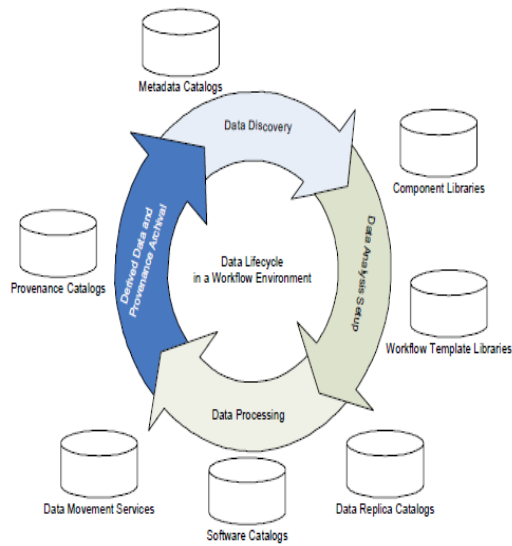


Figure 1: Data Lifecycle in a Workflow.

The description of the components mentioned in the above figure 1 workflow life cycle is given below:

- **Workflow Template:** It describes the steps of analysis to be taken with respect to the identifying component. Designing the workflow template involves searching over the workflows and component libraries. Workflow template also contains more than one component. Both the user and expert have the authority to validate the correctness of templates. The workflow templates can be re-used.
- **Workflow Instance:** Workflow template gets the functionality by pushing the data by creating the workflow instance. It is an essential component for reproducibility of workflow results.
- **Executable Workflow:** Execution consumes set of resources such as CPU, memory. Once the workflow instance is ready then executable workflow maps the instance to the available resources. The mapping involves adding nodes to the workflow to perform the desired task. Workflow engine executes the executable workflow.

4. CHALLENGES IN WORKFLOW LIFE CYCLE

Collaborative scientific workflows play an important role to collaborate among the research groups, datasets and computational tasks. Hence to execute the workflows, the collaborators need to be interconnected through the internet. The challenging aspect in the area of maintaining the collaborative workflows lies in storing, transferring and reusing the intermediate data among the collaborators. The challenges are listed below:

- 1) **Data Management:** Maintaining the data, computational resources, I/O speeds seriously needed to increase the scalability of application. Handling Data staging to and from the set of resources is a challenging aspect. Choosing the right resource for the computation is also a challenge. Scalability and performance can be improved by provenance data access patterns such as intermediate data, data locations and set of methodology to generate the data product.
- 2) **Service Management:** Data staging from one service to another service is critical to consider. It is hard to manage large

number of services in terms of service invocation, management, and destruction. Service Oriented Architecture provides the interoperability among the services.

3) **Data Provenance:** In addition to managing the metadata, the provenance data management is essential for scientific and engineering reproducibility [22]. Different workflow systems produces different amount of information. Managing the provenance data produced by different workflow system is a challenge. Because the data provenance technologies are still being developed [13]. Discovery of data where it is originated from several systems to facilitate the data provenance representation is underway [14, 15].

4) **Community Standards:** Scientists and researchers spend effort to accept the metadata ontology this is enough to describe the meaning of data sets. Hence it is a challenging aspect for communities to agree on standards for specifying the metadata and provenance. In the domain of bioinformatics the effort have been made successful, where scientists or researchers not only defining the meta data standards but also sharing the myExperiment [16].

5) **Workflow:** The workflow creation allows retracing the workflow. Vistrails [17] provides the graphical interface for workflow creation and it allows the systematic growth. Workflow restructuring process needs to recorded [18] to map between executable workflow and what the user specified [19].

6) **Data Dependencies:** During the workflow executions the execution manager keeps track of all the tasks.

7) **Metadata Catalogs:** Relational data bases or RDF triple stores [20] are useful to query the metadata attributes. The community defines the metadata ontology these are the database schemas. The Metadata Catalog Service [21] provides a set of generic metadata attributes.

8) **Metadata Management:** Extracting and saving the metadata information in the metadata catalog is important to consider. Thus additional workflow components can be provided. As there are no standard way for software components to generate the metadata, and in some cases the software components does not provide any metadata for the results. In general, it is very In general, it is very formats [22] which include the metadata. Some enhanced work need to be accomplished in the area of defining the standards and formats.

9) **Integrations:** The integration challenges involves while executing the workflow. During the execution of workflow on cloud resources it is mandatory to check whether the jobs are service based or task based. Because the jobs in workflow may be service or task on one hand but on the other hand the services for the task can be software tools, software tools, storage resources or services. After identifying it is better to assign on the available resources. These steps requires some of the challenging task such as integration of sub workflows to interconnect intermediate data or to transfer the data

10) **User Experiences:** Users don't have a fixed analysis rather they design the workflow through the exploration, by developing different versions of same workflow or modifying the existing workflow. Users need to difficult for workflow systems to know about the information systematically catalog the about the workflow. The metadata associated with the derived data. In some scientific scenarios, there are standard challenge here is to supply the information at the right level which can be easily understandable by the user.

11) Optimization: Optimizing the workflow need to be taken care for efficient overall workflow performance [23]. The solution to increase the performance is to schedule the individual jobs to the available resources. The challenging aspect in case of data intensive workflow is to schedule the computation close to the data.

12) Workflow Execution: The challenge in the workflow execution is the ability of handling and documenting failures. Failures can happen because networks go down, bugs in the software system, or might be in the application components, unavailability of data or resources.

13) Distributed Data Environment: Major challenge in the real time applications is managing the data in the distributed environment. Most of the data is stored in the archives and it can be staged to the computational sites on demand. Even though processing power is available but getting the data to the available resource is time consuming. There is a distinguish between shared storage and storage local to computation node within the computational site. In Major challenge in the real time applications is managing the data in the distributed environment. Most of the data is stored in the archives and it can be staged to the computational sites on demand. Even though processing power is available but getting the data to the available resource is time consuming. There is a distinguish between shared storage and storage local to computation node within the computational site. In distributed environment identifying the location of data sets is a challenging task.

Based on the knowledge of present state of resources which may be provided by information services [24], the workflow planner selects the data replicas. In particular the planner selects the copies of data which are close to the resources where the work flow is executing its tasks. Figure shows the distributed nature of the scientific data.

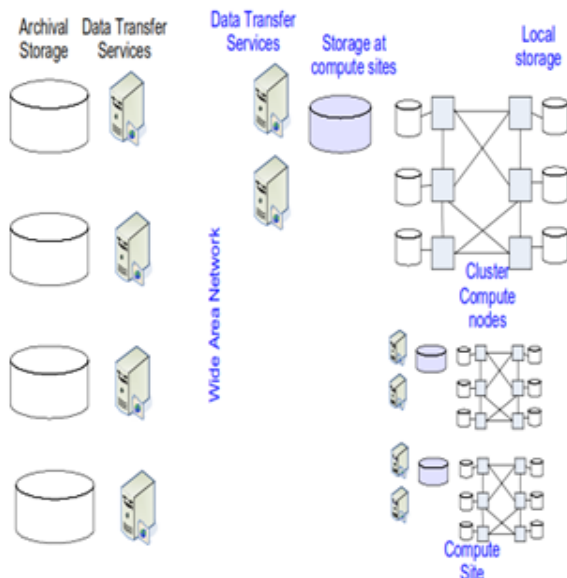


Figure 2: View of the Distributed Data Environment.

Another challenge in the workflow execution is to collect the necessary information about the failures so they can be avoided in the future. This can be solved by placing the monitoring components at all levels of workflow execution.

5. WORKFLOW MODEL

Developing an abstract model which captures the workflows is a major challenge to consider. Workflow is a set of activities that are necessary to complete the task. It comprises of set of modules where each module takes data as input and produces the final data as output. Consider an example of workflow as shown in below figure 3.

A workflow consists of modules where each module receives the initial data as input or data might be generated by other modules on its input ports and sends the output data or generates the final data on its output port. Data items sent by a module to another module are called intermediate data. In the above figure proc module is holding intermediate data.

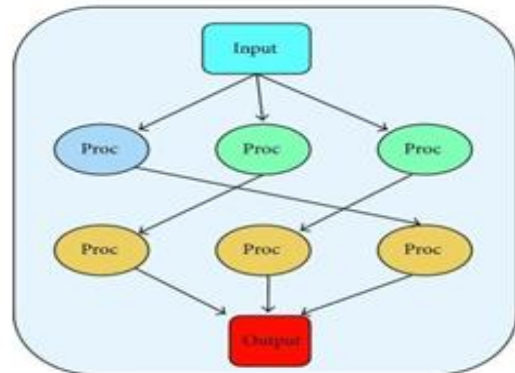


Figure 3: Workflow Model

6. EXAMPLES MOTIVATES PRIVACY ISSUES

Example 1

In a medical diagnosis the final goal of tracing the workflow is to predict whether the patient has a particular disease. The module in a workflow predict the presence of disease by taking the input as set of patient attributes and produces the output whether patient has particular disease. The module corresponds to private algorithm. The owner concerned to some approximation of algorithm could be re-engineered.

Example 2:

Annotation workflows are being developed to provide the annotations of good quality. The work on discovering the better annotation need to enhance because of the exponential growing number of sequenced genomes that are becoming available, functional annotation is essential for determining the function or role of each protein in a genome. This annotation workflow takes the protein sequence as an input and produces an annotation which is expressed using Gene Ontology (GO) terms. Motivated by the above examples, below three points describes the privacy issues in scientific workflows:

- **Module Privacy:** The user should not be able to guess the output for any input, because module contains private information. The module contains the algorithm; the algorithm employs the private information
- **Data Privacy:** The modules connect to the data base to get the private information. The data flow between modules may be private information. This type of private information is intermediate data.
- **Provenance Privacy:** The process of how the data product has generated by the execution of workflow is also private.

Among all these privacy, preserving the module privacy is better idea. The mechanism of carefully hiding only chosen subset of intermediate data is important for module privacy. This mechanism limits the amount of visibility of provenance data.

Another mechanism for preserving the module privacy is to add noise to the answer of a query in order to meet the estimated privacy. This mechanism is reasonable for the scenarios like statistical databases, because the estimated solution is to get the answer within some precision.

The technique of adding noise to the output information in the workflow modules does not give guarantee on the integrity of data. Therefore among these two above mentioned techniques, hiding the data is the best compared to adding noise.

Preserving the module privacy involves challenges such as

1. Standalone Module Privacy: Preserving the privacy for the modules which are not part of workflow systems.

2. In-network Module Privacy: Preserving the privacy for the modules which are in the network of the workflow system.

Among these two module privacy standalone module privacy is better in terms of complexity which arises because of interactions between the modules in case of in- network module privacy.

7. ADVANTAGES OF USING PROVENANCE FOR WORKFLOWS

- **Data Exploration:** The provenance is useful for exploratory process. In addition to this provenance contain useful knowledge, which can be used further for the construction of the workflow
- **Education:** The workflow adopt provenance will support data exploration component such as visualization and mining. The provenance simplifies the learning experience. For each of student queries the instructors keeps detailed record of all steps by using provenance enabled tool. At any time have the authority to see all the steps followed to solve the problem by tracking provenance data.
- **Reproducibility:** Maintaining Documentation of the steps to be taken to reach a goal gives some idea to others to reproduce and validate these outcomes. This is the benefit of maintaining provenance of the result. The usage of this reproducibility feature exits where authors need their publication reproducible.

8. DISADVANTAGES OF USING PROVENANCE FOR WORKFLOWS

- Work flow and provenance information play an important role to solve the problem hence there is need of developing efficient techniques to manage these data. One important challenge in information management is usability [25,26],because of the system growth especially in the scientific domain usability plays an important role
- **Information Overload:** This is one of the challenges which arise because of growth of provenance data.
- **Visualization:** Scientist can reproduce; debug their task in a workflow to obtain the better results. This can possible through visualization of provenance data. The challenge is mining and extracting the knowledge from provenance data.
- **Database Connection:** Understanding the provenance of result is important in order to connect

provenance information across workflow and databases. This lead to challenge of developing a frame work where in which provenance information can be combined and all workflow modules are treated uniformly.

- **Interoperability:** Designing the complex products involves the long processing chains and requires multiple tools and better technologies and resources. It is necessary to integrate the provenance derived from the different systems in order to provide the detailed provenance data for data products. The integration of provenance data is the second provenance challenge [27]. Although it is possible to integrate the results but it requires more principled approach to solve this problem. One solutions is the creation of standard for representing provenance[28]

9. WORKFLOW AS A SERVICE IN THE CLOUD (WFaaS)

To provide the scalable and future enhancement, this paper provides the WFaaS architecture with independent services. The major advantage of this architecture is to manage continuous workflow requests and scheduling the executions in the cloud.

Implementation of dynamic real time monitoring, analysis and modeling is essential for the development of a sophisticated smart manufacturing intelligence. Some real time projects considers workflow as an essential technique in the implementation of automation and dynamic decision making process through the analysis of real time data, whereas the cloud platform provides the scalability and reduces the cost of deploying the services by providing data storage resources to complete jobs in real time.

Among all the services provided by the cloud computing, WFaaS is good model to support workflow publishing, query execution [29]. WFaaS provides a generalised model to handle set of services based on certain logic. WFaaS is flexible to integrate set of application via workflows by facilitating not only providing the

service but also facilitating management environment. WFaaS also utilizes other cloud computing services such as IaaS, DaaS to get proper data, software packages and VM for its execution.

9.1. Architecture for WFaaS in the Cloud.

Workflow run on VMI in the cloud environment one VM provides a set of VMI. Workflow brings challenges on VM management. It is not a better solution to run set of tasks one VMI because it consumes lot of resources. This causes VM management difficulties in the cloud. The VM and Workflow tightly coupled in the approach because VM creation depends on Workflow system.

The WFaaS architecture in the cloud is shown in the below Figure 4 where it shows the role separation for workflow management and virtual machine.

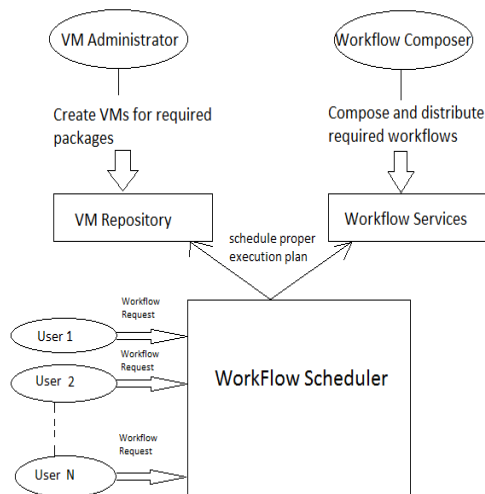


Figure 4: WFaaS in the Cloud

10. CONCLUSION AND FUTURE WORK

This paper discusses about the scientific workflow management and privacy issues in workflows, and briefly discusses about the architecture of workflow as a service in the cloud by clearly defining the independent services. The paper gives the information about several challenges associated with creating and managing large scale workflows in distributed environment. Based on the WFaaS architecture, the future work is to present the algorithm for efficient workflow execution in the cloud.

11. ACKNOWLEDGMENT

I wish to thank my parents for their support and encouragement throughout my study.

12. REFERENCES

- [1] C. Dwork. Differential privacy: A survey of results. In: TAMC, pages 1-19, 2008.
- [2] The OASIS Committee: Web Services Business Process Execution Language (WSBPEL) Version 2.0 (2007)
- [3] van der Aalst, W., ter Hofstede, A.: Yet another workflow language. *Information Systems* 30(4) (2005) 245–275
- [4] Deelman, E., Gil, Y.: Workshop on the Challenges of Scientific Workflows. Technical report, Information Sciences Institute, University of Southern California (2006)
- [5] Oinn, T., et al: Taverna: a tool for the composition and enactment of bioinformatics workflows. *Bioinformatics* 20(17) (2004) 3045–3054
- [6] Zhao, J., et al: Annotating, linking and browsing provenance logs for e-Science. In: 1st Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data, Sanibel Island, Florida, USA. (2003)
- [7] Ludascher, B., et al.: Scientific workflow management and the kepler system. *Concurrency and Computation: Practice and Experience* 18(10) (2005) 1039–1065
- [8] Taylor, I.J., et al.: Distributed P2P Computing within Triana: A Galaxy Visualization Test Case. In: 17th International Parallel and Distributed Processing Symposium (IPDPS 2003), IEEE Computer Society (2003) 16–27

- [9] Allen, G., et. al: Enabling Applications on the Grid: A GridLab Overview. *International Journal of High Performance Computing Applications: Special Issue on Grid Computing: Infrastructure and Applications* 17(4) (2003) 449–466
- [10] Deelman, E., et al.: Pegasus: A framework for Mapping Complex Scientific Workflows onto Distributed Systems. *Scientific Programming Journal* 13(3) (2005) 219– 237
- [11] Brown, J.L., et al.: GridNexus: A Grid Services Scientific Workflow System. *International Journal of Computer Information Science (IJCIS)* 6(2) (2005) 72–82
- [12] Rowe, A., et al.: The Discovery Net System for High Throughput Bioinformatics. *Bioinformatics* 19(1) (2003) 225–231
- [13] L. Moreau, et al., "Concurrency and Computation: Practice and Experience, Special Issue on the First Provenance Challenge," 2007.
- [14] L. Moreau, et al., "The Open Provenance Model," University of Southampton 2007. <http://eprints.ecs.soton.ac.uk/14979/1/opm.pdf>
- [15] L. Moreau, et al., "The First Provenance Challenge," *Concurrency and Computation: Practice and Experience*, 2007.
- [16] C. A. Goble, et al., "myExperiment: social networking for workflow-using e-scientists," *WORKS*, pp. 1-2, 2007.
- [17] J. Freire, et al., "Managing Rapidly-Evolving Scientific Workflows.," *IPAW*, vol. 4145, pp. 10-18, 2006.
- [18] S. Miles, et al., "Connecting Scientific Data to Scientific Experiments with Provenance," in *e- Science*, 2007
- [19] E. Deelman, et al., "Pegasus: a Framework for Mapping Complex Scientific Workflows onto Distributed Systems," *Scientific Programming .1*, vol. 13, 2005.
- [20] O. Lassila, et al., "Resource Description Framework (RDF) Model and Syntax Specification," 1999.
- [21] E. Deelman, et al., "Grid-Based Metadata Services," in *16th International Conference on Scientific and Statistical Database Management*, 2004.
- [22] "Flexible Image Transport System." <http://fits.gsfc.nasa.gov/>
- [23] Workflow optimization of performance and quality of service for bioinformatics application in high performance computing Rashid Al-Ali, Nagarajan Kathiresan, Mohammed El Anbari, Eric R. Schendel, Tariq Abu Zaid
- [24] K. Czajkowski, et al., "Grid Information Services for Distributed Resource Sharing," in *HPDC 2001*
- [25] L. Haas. Information for people. [http://www.almaden.ibm.com/cs/people/laura/Information For People keynote.pdf](http://www.almaden.ibm.com/cs/people/laura/Information%20For%20People%20keynote.pdf), 2007. Keynote talk at ICDE.
- [26] H. V. Jagadish. Making database systems usable <http://www.eecs.umich.edu/db/usable/usability-sigmod.ppt>, 2007. Keynote talk at SIGMOD.
- [27] Second provenance challenge. <http://twiki.ipaw.info/bin/view/Challenge/SecondProvenanceChallenge>, 2007. J. Freire, S. Miles, and L. Moreau (organizers).

[28] L. Moreau, J. Freire, J. Futrelle, R. McGrath, J. Myers, and P. Paulson. The open provenance model, December 2007. <http://eprints.ecs.soton.ac.uk/14979>.

[29] Workflow as a Service in the Cloud: Architecture and

Scheduling Algorithms Jianwu Wang¹ , Prakashan Korambath² , Ilkay Altintas¹ , Jim Davis³ , Daniel Crawl¹