A New Approach for Improvement of Workflow Execution Performance on Grid Environment

Mohammad Nikfahm, Majid Rezaei-pour, Mohsen Ramezani, Arman Eisapour, Milad Nemati

1Jaber-Ebne-Hayyan Higher Education Institute - Rasht - Iran.
2Sardar Jagad Higher Education Institute - Rasht – Iran.
3University of Guilan – Iran.
4Islamic Azad University Of Qazvin – Iran.
5Islamic Azad University Of Fouman – Iran.

ABSTRACT

Background: Large-scale applications are being built by scientific collaborations in physics, astronomy, biology, earthquake science etc. These applications are often structured as workflows that express an application by specifying a set of interdependent task. Different commercial projects have been done in field of building, directing and executing workflows in Grid environment; most of them are entered manually into the system containing services not available physically. Such services are assigned as physical during execution; this phenomenon has several advantages; for instance, they are created dynamically in the services’ distributed environments and disappear. So if they are selected during execution, the work quality is promoted and also the work is felt on the architectures taking into consideration the services’ quality into the composition. This article is to present new mechanisms; the techniques used in it are artificial intelligence and web services.

INTRODUCTION

An application workflow is asset of tasks with data dependencies between them. It can be represented as a directed acyclic graph (DAG) where the vertices are the compute tasks and the edges are the data dependencies between the tasks. The input data required by a task should be available before the task begins execution and the output data produced can be transferred to its child tasks only when the task has completed execution. This is in contrast to the data pipeline model where data is streamed between the tasks. In this paper, we focus on large-scale workflows containing thousands of tasks.

Web Services provide a new way of distributed computing that achieve the interoperability between heterogeneous applications through platform and language independent interfaces. As the number of available web services in a system increases, it becomes more difficult to find the specific service that can perform the task at hand. There may also be on single service capable of performing that task, but a combination of existing services may provide the capability. One of the main focus of web services is the ability to easily combine existing components to create compositions that provide novel functionality that was not directly available from the existing services [7].

Much work has focused on developing heuristics for mapping the application tasks to appropriate resources based on performance models in order to minimize the application makes pan. The makes pan of the workflow is defined as the time interval between the start time of the first task and the completion time of the last task in the workflow.

Automatic composition of workflows from web and Grid services is considered an important challenge. Allowing business processes to vary between Grid deployments implies the need to abstract them. In most of the projects developed in this field the abstract workflow is either built by the user or is predefined and stored in a repository[8].

Corresponding Author: Mohammad Nikfahm, Jaber-Ebne-Hayyan Higher Education Institute - Rasht - Iran
First, in this paper we explain the primary concepts of workflow systems and Grid environment. Second section is to examine the basic notions concerning web service and their composition techniques. In third section we examine our proposed framework and its details. And finally in fourth section we present the conclusion and analyze our following activities.

In the frame proposed by us the QoS parameters description is done by a QoS semantic language. The frame uses abstract service of 'next Grid' project in order to provide the possibility to create abstract workflow. The abstract workflows creations have been taken into consideration dynamically in a few Grid projects and no semantic language is used for this purpose. In this field most Grid projects give abstract workflows to the system by the user and convert it to real workflow during execution. Another feature of the frame is to present a new solution to select the services in form of a 'csp'.

**Primary concepts:**

One of the main advantages of workflows management system is its application in distributed environments where the processes controlling the activities possible in the system are on the systems already distributed [1].

The discussions related to workflows have been emphasized and examined widely since 2002. The workflows have been created to build processes which follow the operations and a defined goal. This continuation of the operations is categorized in different classes based on the execution order. Some of the operations have priority to others because they are more necessary; because of the same priority of some of them they are executed in parallel [2].

One of the most applied distributed environments is Grid [5]. Generally Grid is a set of linked computers which interacted with each other and function as a figurative machine to meet the users' needs in a way that the user cannot distinguish he(she) is in relation with some computers (Not only with one). In addition, Grid is a distributed environment with three special advantages [6].

a – Grid should provide QoS for the service to meet the users' needs; QoS describes the quality necessary for the users of the services presented by Grid.

b – No central source controls Grid.

c – It uses standard and multipurpose protocols to secure, discover and access the sources.

Web service is a software system known by a URL. The Interface of web service is defined and described by XML. The agent relating the web services is the messages described by XML.

The web service architecture is fourfold whose the highest has some facilities to discover and find other web services and the lowest one is related to describe the web service; the WSDL [10] standard has been introduced for this purpose founded on XML.

The next layer is recall service to send the messages between the client and server. Also SOAP[9] standard has been introduced for this layer. The messages to demand data towards server and returned from it are exchanged based on SOAP standard namely SOAP defines the format of the messages. This protocol is the most famous standard presented for this purpose and implemented on XML and finally the last layer is Transport containing some standards to transfer the messages between client and server. The protocol usually selected for this layer is HTTP.

**Proposed frame:**

Our proposed frame follows abstract and real workflows model in which we are in relation with abstract and real services. The frame uses the language: OWL-WS [3]. Abstract services concept has been used in the frame. It is necessary to have an abstract system similar to actual one; it is implemented easily. A similar service is searched among the abstract ones for each service registered in related place; if no service is found, an abstract one is built from the service. It should be noted that it is not necessary to compose the services based on QoS because the services are abstract and QoS is related to the real services.

Figure 1 shows general schema of the frame including three phases: the abstract workflows are composed automatically in the first phase. Next phase finds some real services for the abstract ones to be adapted with them. In the last phase the abstract services are chosen from the real ones to meet the user's QoS needs.

**First phase, Composing automatically the abstract workflows:**

By virtue of the techniques related to the range describing abstract services and primary and favorable situation description OWL-WS to PDDXML convertor was used to convert the services as a technique in the phase. Then PDDXML is converted to PDDL and the files are sent to Planner to be executed. The result of Planner execution is an abstract workflow in which none of actual services are assigned to special service and implementation.

**Second phase, Discovering adapted services:**

These services should be converted to real services. OWL-S Matchmaker [4] was used to convert these services to the real ones. Generally Matchmaking algorithms are to examine the possibility of creating
adapts between the two services. The UDDI is the standard to resolve the problem concerning web services recovery.

UDDI uses XML to describe the web services model, but XML may not describe the meanings. Our proposed registries use OWL-S. When services are formed based on OWL-S it is possible to use the search mechanisms. Such mechanisms are known as ‘Capability Search Mechanism’. Having used these mechanisms and matchmaking algorithms it is possible to find similar services. Generally by virtue of web services OWL-S descriptions these services are registered where the files of OWL-S type are accepted to enable the programs to find and retrieve them.

The programs should describe the profile of the requested OWL-S services in order to discover the registry and adaptations. The Matchmaker is to elect the services from the advertising services registered in OWL-S UDDI before. An advertising service is adapted to the requested service, if the descriptive operations of both resemble enough to each other.

**Fig. 1**: The architecture of the presented framework.

**Third phase, Selecting among real services:**

Considering there are many services distributed in the same environment and very likely they are the same specially in an environment such as Grid the services are selected according to QoS.

In this phase QoS descriptions were extracted related to the real services adapted to each abstract one and described in a similar XML file as follows:

The QoS descriptions of real services related to their abstract ones are saved in a tag of the service. The selection among the real service is entitled ‘Service Selection’. Such operation usually done by two groups of algorithm. First group: The algorithms define the operations as multidimensional rucksack. The second group includes the algorithms presenting Graph and resolving and presenting the operations by selecting optimal path.

In this article a new method is presented to resolve the problem; The selection operations model the services in one of the Constraint Satisfaction Problems (csp) states.

Csp is one of the artificial intelligence issues with a special formulation; we have some variables in it and each one has a special or common range while there are some limitations in the field, too. It is to make the
variables decimal in a way that the limitations would be removed. Following definitions are presented to formulate the services selection:

*The variables and abstract services are the same in this method:*

The variables range is m*n matrix where 'm' is the number of QoS and 'n' is the number of real matrices discovered for each abstract service (Variable) by the adaptor. Figure 2 shows the services model and related matrix. S1 and S2 are abstract services adapted by Matchmaker for two and three real services, respectively. Figure 3 shows the matrix created for S1 service.

![Fig. 2: The services model and related matrix](image)

![Fig. 3: The matrix created for S1 abstract service](image)

A matrix with the features mentioned for each abstract service is created from the QoS descriptions extracted from the real services in a XML file in order to achieve the goal. The limitations concerning this problem are defined as follows: total selected values from each variable range should be equal to or less than the values requested by the user; in other words, total selected entries of the m*1 matrices should be equal to or less than their similar entries in m*1 matrix presenting the user's needs. The services selection terminates all operations; such selection includes a real workflow by taking into consideration the quality problem of the services.

**Comparison and Conclusion:**

In table 1 you see a comparison by the frame with important and famous projects proposed and implemented in the field before.

In the frame proposed by us the QoS parameters description is done by a QoS semantic language. The frame uses abstract service of 'next Grid' project in order to provide the possibility to create abstract workflow. The abstract workflows creation have been taken into consideration dynamically in a few Grid projects and no semantic language is used for this purpose. In this field most Grid projects give abstract workflows to the system by the user and convert it to real workflow during execution. Another feature of the frame is to present a new solution to select the services in form of a 'csp'.

<table>
<thead>
<tr>
<th>Project</th>
<th>QoS - Awareness</th>
<th>Composition</th>
<th>User Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHOP2</td>
<td>N/A</td>
<td>Automatic</td>
<td>Init State &amp; Domain Description</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HTN Planning</td>
<td>A Set of tasks</td>
</tr>
<tr>
<td>next Grid</td>
<td>N/A</td>
<td>N/A</td>
<td>Abstract Workflow</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Supports Composite and Simple Processes</td>
<td></td>
</tr>
<tr>
<td>Proposed Architecture</td>
<td>(\lor)</td>
<td>(\lor)</td>
<td>Can be Solved by any numeric expression AI Planner</td>
</tr>
<tr>
<td></td>
<td>Exploits a 2 Layer Ontology</td>
<td>Init State, Domain Description &amp; Goal State</td>
<td></td>
</tr>
</tbody>
</table>
It is possible to increase the efficiency of the proposed frame following operations:
- Taking into consideration the simple and compound processes.
- Comparing the proposed model to select the services with similar proposed one.
- Using ‘preplanning’ techniques to improve the efficiency.

REFERENCES