ASKALON

Grid Application Development and Computing Environment

K-Wf Grid

Online Monitoring and Performance Analysis of Grid Scientific Workflows

Hong-Linh Truong, Thomas Fahringer

Institute for Computer Science, University of Innsbruck, Technikerstrasse 21A, A-6020 Innsbruck, Austria.

E-mail: {truong,tf}@dps.uibk.ac.at, http://dps.uibk.ac.at

ABSTRACT

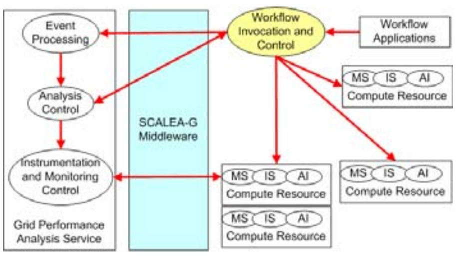
While existing work concentrates on developing QoS models of business workflows and Web services, few tools have been developed to support the monitoring and performance analysis of scientific workflows in Grids. This paper describes a Grid service for performance monitoring and analysis of Grid scientific workflows. The service utilizes workflow graphs and various types of performance data including monitoring data of resources, execution status of activities, and performance measurement obtained from the dynamic instrumentation, to support the monitoring and performance analysis in a dynamic and flexible way. We store workflows and their relevant information, devise techniques to compare constructs of different workflows, and support multi-workflow analysis.

MOTIVATION

As the Grid is diverse, dynamic and inter-organizational, even with a particular scientific experiment, there is a need of having a set of different workflows because (i) one workflow mostly is suitable for a particular configuration of underlying Grid systems, and (ii) available resources allocated for a scientific experiment and their configuration are changed in each run on the Grid. The client of performance tools wants to compare the performance of different WF constructs with respect to the resources allocated in order to determine which WF construct should be mapped onto which topology of the underlying Grid. Therefore, multi-workflow analysis, the analysis and comparison of the performance of different WF constructs, ranging from the whole WF to a specific construct (e.g. a fork-join construct), is an important feature. Even though numerous tools have been developed for constructing and executing scientific workflows on the Grid, Grid workflow programmers suffer from the lack of supportive monitoring and performance tools. To understand the performance of WFs on the Grid, we need to collect and analyze a variety of types of data relevant to the execution of the WFs from many sources.

ASKALON GRID TOOLSET

PERFORMANCE ANALYSIS SERVICE



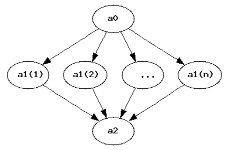
MS: Monitoring Service, IS: Instrumentation Service, AI: Activity Instance

Figure 1. Model of monitoring and performance analysis of workflow-based application

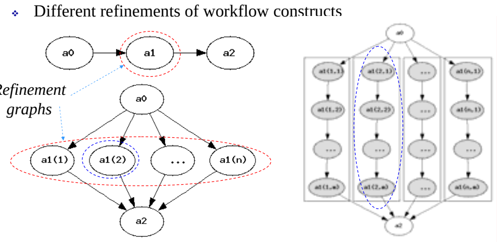
Dynamic approach in performance monitoring and analysis of Grid scientifc workflows

- Monitoring execution status of activities on the fly
- Dynamic instrumentation of invoked applications of activities
- Online profiling analysis of invoked applications and performance analysis of workflows
- Unifying applications and resource monitoring in a single system

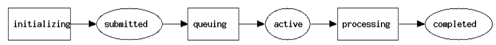
SUPPORTED WORKFLOW PARADIGMS



- **DAG-based workflows**
- **Fork-join model**
- **Multi-workflow model**
  - Different workflows for the same workflow application
  - Different refinements of workflow constructs



- **Activity execution status graph modeled by a discrete process**



PERFORMANCE ANALYSIS

- **Activity level**
  - Capture performance metrics of activities and instrumented code regions.
  - Incrementally provide performance metrics of code regions during the execution of the WF.
  - Focus on analyzing response-time of activities.
- **Workflow level**
  - Monitor and analyze performance metrics that characterize the interaction and performance impact among activities.
  - Utilize WF graphs, execution status information and performance data for performance analysis
- **Multi-workflow analysis**
  - Collect and store different DAGs of WFs, performance data and machine information into an experiment repository
  - Associate refinement graphs with their performance metrics.
  - Represent the refinement relationship between workflows or workflow constructs.

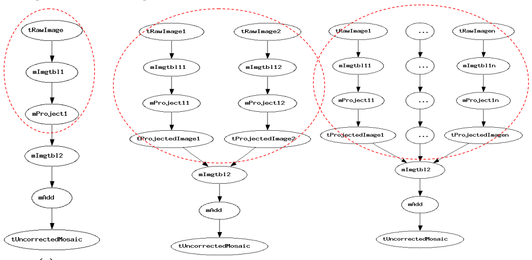
IMPLEMENTATION

Workflow Invocation Control in our experiment is currently implemented based on JavaCog. JGraph and JFreeChart are used to visualize workflow DAGs and performance results, respectively.
The performance analysis service has been integrated into SCALEA-G and will be part of the EU funded K-Wf Grid environment.

EXPERIMENTS

- **Experimental application**

Montage (montage.ipac.caltech.edu) is a software for generating astronomical image mosaics with background modeling and rectification capabilities. Based on the Montage tutorial, we develop a set of WFs, each generates a mosaic from 10 images without applying any background matching.



(a) (b) (c)

Figure 2. Experimental workflows of the Montage pplication: (a) workflow executed on single resource, (b) workflow executed on two resources, and (c) workflow executed on n resources.

Experiments are conducted on sites named GUP (University of Linz), UIBK (University of Innsbruck), AURORA6 (University of Vienna) and VCPC (University of Vienna) of the Austrian Grid.

- **Online monitoring execution status**

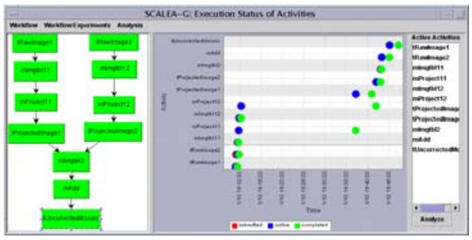


Figure 3. Monitoring execution status of a Montage workflow executed on 2 resources.

- **Dynamic instrumentation of invoked applications of workflows**

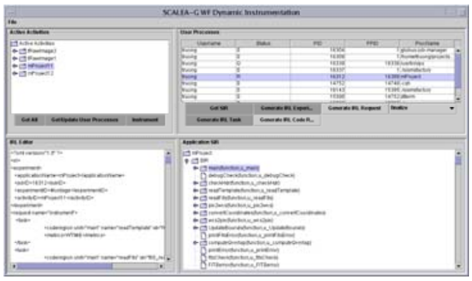


Figure 4. GUI used to control the dynamic instrumentation of invoked applications of activity instances.

- **Performance analysis of workflows**

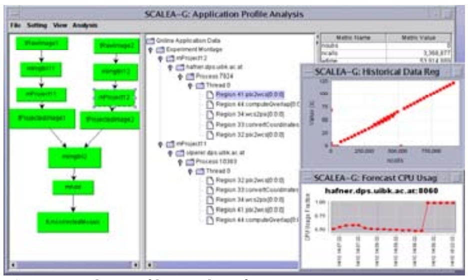


Figure 5. Online profiling analysis for WF activities..

(a) (b)

Figure 6. Analysis of Montage executed on 5 machines
(a) response time and synchronization delay of mImgtbl
(b) load imbalance of mProject.

- **Multi-workflow analysis**

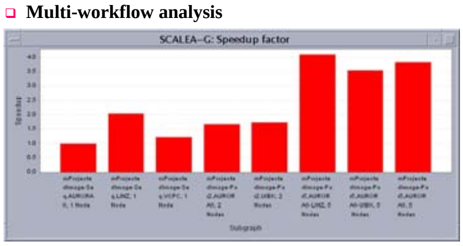


Figure 7. Speedup factor for refinement graphs ProjectedImage (tRawImage → mImgtbl1→ mProject1→ tProjectedImage) of Montage WFs.

FUTURE WORK

We will the extend WF specification language with directives specifying monitoring conditions. These directives will be translated into code used to publish the execution status to the monitoring middleware. The performance analysis for workflows will be extended to cover AGWL (Abstract Grid Workflow Language) workflows which covers also loops.

supported by the European Commission under contract IST-511385