K-WfGrid Distributed Monitoring and Performance Analysis Services for Workflows in the Grid

Hong-Linh Truong¹, Peter Brunner¹, Thomas Fahringer¹, Vlad Nae¹, Francesco Nerieri¹, Robert Samborski¹, Bartosz Balis², Marian Bubak², Kuba Rozkwitalski²

Email: truong@dps.uibk.ac.at
http://www.kwfgrid.eu

¹Distributed and Parallel Systems Group
University of Innsbruck, Austria
http://www.dps.uibk.ac.at

²Academic Computer Centre CYFRONET
AGH, Poland
http://www.cyf-kr.edu.pl
Outline

- Motivation
- Architecture of K-WfGrid Performance Monitoring and Analysis Services
  - Performance Service Interfaces and Data Representation
  - Workflow monitoring and instrumentation
  - Performance analysis of Grid workflows
- Implementation
- A short demonstration movie
Motivation

- The lack of performance monitoring and analysis tools supporting Grid workflows composed from Web/WSRF services.

- The challenge of understanding performance of Grid workflows at multiple levels of abstraction.

- The need to simplify the interoperability and integration among performance services and their clients, and to provide performance knowledge for semi-automatically composition and execution of workflows.

→ online SOA-based Grid workflow performance services for end-users, developers and middleware.
K-WfGrid Performance Monitoring and Analysis Services

- Performance overhead analysis and search for performance problems
- PDQS, WIRL, SIRWF, XML-based performance data
- Grid Monitoring and Instrumentation Service
- XML-based monitoring data

DIPAS Portal

WP3 Performance Monitoring and Analysis Core Services

PDQS, WARL

Other K-WfGrid Services

Grid resources and applications

PDQS, WIRL, SIRWF, XML-based performance data
Performance data representation and service interfaces

- All types of monitoring data are described in XML
  - Information about monitoring data is described in OWL
- PDQS (Performance Data Query and Subscription)
  - For subscription and query any kinds of monitoring data
- WARL (Workflow Analysis Request Language)
  - Performance analysis requests and search for performance problems.
- For controlling the instrumentation and measurement
  - SIRWF (Standardized Intermediate Representation for Workflows) and WIRL (Workflow Instrumentation Request Language)
Easy to search performance data provider using ontology

```
<?xml version="1.0"?>
<pdqs xmlns="http://net.kwfgrid/dr/pdqs">
  <dataTypeID>Wfa.event</dataTypeID>
  <resourceID>truong_810cf130-eb24-11da-8ebd-a46bfd55290</resourceID>
  <subscriptionTime>
    <from>0</from>
    <to>0</to>
  </subscriptionTime>
</pdqs>
```
Workflow Monitoring and Instrumentation

- **Workflow level**
  - Providing data for analyzing workflow, workflow region, and activity
  - Statically instrument GWES and collect workflow and activity events.

- **Invoked application level**
  - Providing data for analyzing invoked application and code region
  - Support dynamically-enabled instrumentation for C code and static/byte-code/dynamic instrumentation for Java

- **Integrated with existing infrastructure monitoring (Ganglia, Iperf)**

- **Data correlation using workflow/activity id**
  - Passing id using SOAP header

- **Support both data query (pull) and subscription (push)**
Workflow Performance Analysis

- **Workflow execution tracing**
  - Tracing all execution phases of all activity instances

- **Workflow overhead analysis**
  - Support a novel of performance overhead classification for Grid workflows
  - Provide application and/or middleware overheads

- **Search for performance problems**
  - Based on performance conditions
  - Conditions established based on performance metrics, overheads, user preferences
  - Conditions can be specified during runtime as well as before the workflow/activity is running

- **A unified system for performance analysis of Grid infrastructure and workflows**
Example of request for analyzing performance overheads and search for performance problems.

<?xml version="1.0" encoding="UTF-8"?>
<warl>
  <constraint>
    <startTime>0</startTime><endTime>0</endTime>
    <workflowID>truong_3d6c4330-eb2a-11da-8ebd-a46bfd55290e</workflowID>
    <concepts>
      <concept name="truong_3d6c4330-eb2a-11da-8ebd-a46bfd55290e" type="Workflow"/>
      <concept name="computeStartZonePolyg" type="Activity"/>
      <concept name="computeEndZonePolyg" type="Activity"/>
      <concept name="computeStartNodes" type="Activity"/>
    </concepts>
  </constraint>
  <analyze>
    <metric>LoadIm</metric><metric>TotalOverhead</metric><metric>QueuingTime</metric>
  </analyze>
  <perfProblemSpecs>
    <perfProblemSpec><metric>ElapsedTime</metric><operator>GE</operator><value>30</value>
    </perfProblemSpec>
    <perfProblemSpec><metric>QueuingTime</metric><operator>GE</operator><value>5</value>
    </perfProblemSpec>
  </perfProblemSpecs>
</warl>
WSRF-based performance services
- Using GT 4.0
- Monitoring data supports subscription based on ICE (Internet Communications Engine)

Performance visualizations
- Based on JGraph and JFreeChart

Portal based on Gridsphere

Not finished yet
- Monitoring and analysis of invoked applications has not been fully integrated
Allow monitoring and analysis of Grid workflows and infrastructure at the same time.

Monitoring and Analysis Portal

Monitoring Service Plugin: scaling
Registry Info: DPS-Innsbruck

Data Query and Subscription

ResourceID | Availability
-- | --
tcp:goedis.dps.uibk.ac.at:zeus72.cfyr.edu.pl | 100%
http://goedis.dps.uibk.ac.at:zeus72.cfyr.edu.pl:60000 | 100%
http://goedis.dps.uibk.ac.at:zeus72.cfyr.edu.pl:8080 | 100%
tcp:goedis.dps.uibk.ac.at:zeus72.cfyr.edu.pl | 100%

Error Messages

H.-L. Truong
- CTM (Coordinated Traffic Management) Workflows
- Web services deployed in Berlin, Bratislava, Cracow, Genoa and Innsbruck

Short movie
Summary and future work

Summary

- Support performance monitoring and analysis of Grid workflows at multiple levels of abstraction
- Target to end-users, developers, and Grid middleware
- Performance services can be used/adapted in/to other projects

Future work

- Integrating monitoring and analysis of invoked applications and code regions
- Storing performance results in the workflow performance ontology (WfPerfOnto) into knowledge storage

http://www.dps.uibk.ac.at/projects/kwfgrid
http://www.kwfgrid.eu