



# K-Wf Grid



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

Hong-Linh Truong<sup>1</sup>, Peter Brunner<sup>1</sup>, Thomas Fahringer<sup>1</sup>,  
Vlad Nae<sup>1</sup>, Francesco Nerieri<sup>1</sup>, Robert Samborski<sup>1</sup>,  
Bartosz Balis<sup>2</sup>, Marian Bubak<sup>2</sup>, Kuba Rozkwitalski<sup>2</sup>

Email: [truong@dps.uibk.ac.at](mailto:truong@dps.uibk.ac.at)  
<http://www.kwfgrid.eu>

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

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

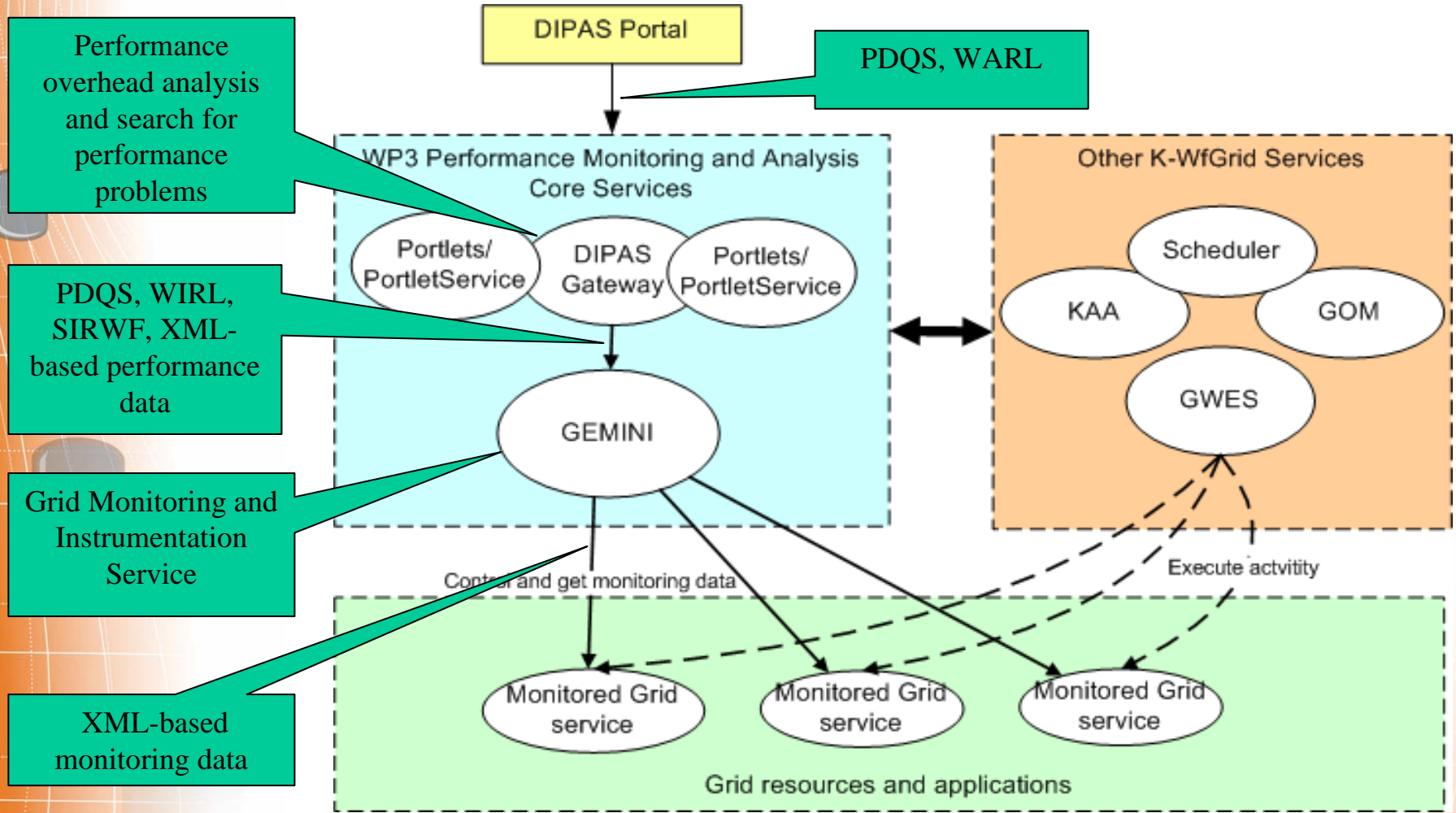


Singular Logic



- ❑ **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**

- ❑ **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**



- ❑ **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)**



OWL

```
<dg:DataObject rdf:ID="MD1148476305524_DO">
  <dg:contains>
    <dg:MonitoringData rdf:ID="MD1148476305524">
      <dg:hasDataType rdf:datatype="...">wfa.event</dg:hasDataType>
      <dg:ofResource rdf:datatype="..."
        >truong_810cf130-eb24-11da-8ebd-a46bfd55290e
      </dg:ofResource>
      <dg:validFrom rdf:datatype="...">1148476305524</dg:validFrom>
      <dg:validTo rdf:datatype="...">0</dg:validTo>
    </dg:MonitoringData>
  </dg:contains>
  <dg:isStoredIn rdf:resource="http://gom.kwfgrid.net/gom/ontology/
    ServiceRegistry/CMN#MSa6240bba-3c48-4cc6-ad31-648e9b60124b"/>
</dg:DataObject>
```

PDQS

```
<?xml version="1.0"?>
<pdqs xmlns="http://net.kwfgrid/dr/pdqs">
  <dataTypeID>wfa.event</dataTypeID>
  <resourceID>truong_810cf130-eb24-11da-8ebd-a46bfd55290e</resourceID>
  <subscriptionTime>
    <from>0</from>
    <to>0</to>
  </subscriptionTime>
  H.- </pdqs>
```

- ❑ **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 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**



```
<?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-a46bf
    <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>
```



Target to  
Grid  
middleware

- ❑ **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

Willkommen | **Monitoring and Analysis**

Infrastructure PMA | Workflow PMA

Workflow performance monitoring and analysis

File Monitoring View Analysis

Suspend Workflow | Res | Workflow info  
Execution Phases  
Load Imbalance | 96 (INITIATED)  
Mean Execution Time  
Instances Distribution  
Workflow Overhead  
Performance Severity

ZoomStd | ZoomIn | ZoomOut | Update | + 144% | - | Time

Workflow ID : csac8265\_80dda320-5538-11db-9cab-8c0cab51a496  
Current Status : completed @ Fri Oct 06 12:49:09 GMT 2006

Active Activities

Workflow: Execution Time (s)

0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150

---

**MonitoringService Configuration**

Monitoring Service Plugin: scaleag

Registry Info: DPS-Innsbruck

Set configuration

---

**Data Query and Subscription**

DataTypeID=>ResourceID:

service.available=>tcp:petzeck.dps.uibk.ac.at->zeus72.cyf-kr.edu.pl

user.ps=>petzeck.dps.uibk.ac.at

user.ps=>goedis.dps.uibk.ac.at

service.available=>icmp:goedis.dps.uibk.ac.at->clown.first.fraunhofer.de

service.available=>http:petzeck.dps.uibk.ac.at->pc6163-c703.uibk.ac.at:40065/wsrf/services/monitor/service

service.available=>icmp:goedis.dps.uibk.ac.at->pc6163-c703.uibk.ac.at

service.available=>http:goedis.dps.uibk.ac.at->fhrg.first.fraunhofer.de:8080/gwes/services/GWES?wsdl

service.available=>tcp:goedis.dps.uibk.ac.at->zeus72.cyf-kr.edu.pl

service.available=>http:petzeck.dps.uibk.ac.at->pc6163-c703.uibk.ac.at:40065/wsrf/services/kwfgrip/DIPASFa

service.available=>http:goedis.dps.uibk.ac.at->pc6163-c703.uibk.ac.at:40350/gwes/services/GWES?wsdl

service.available=>http:goedis.dps.uibk.ac.at->zeus72.cyf-kr.edu.pl:8080/wsrf/services/gom/service/GOMServ

service.available=>http:goedis.dps.uibk.ac.at->pc6163-c703.uibk.ac.at:40065/wsrf/services/kwfgrip/DIPASFa

service.available=>http:petzeck.dps.uibk.ac.at->kwfgrip.dps.uibk.ac.at/WSDL/net-kwfgrip-ctm-NetFileParser-k

service.available=>tcp:petzeck.dps.uibk.ac.at->zeus72.cyf-kr.edu.pl

service.available=>http:petzeck.dps.uibk.ac.at->pc6163-c703.uibk.ac.at:40350/gwes/services/GWES?wsdl

service.available=>http:goedis.dps.uibk.ac.at->kwfgrip.dps.uibk.ac.at/WSDL/net-kwfgrip-ctm-NetFileParser-kw

service.available=>tcp:goedis.dps.uibk.ac.at->portal.ui.sav.sk

service.available=>icmp:petzeck.dps.uibk.ac.at->clown.first.fraunhofer.de

service.available=>icmp:petzeck.dps.uibk.ac.at->pc6163-c703.uibk.ac.at

service.available=>tcp:petzeck.dps.uibk.ac.at->portal.ui.sav.sk

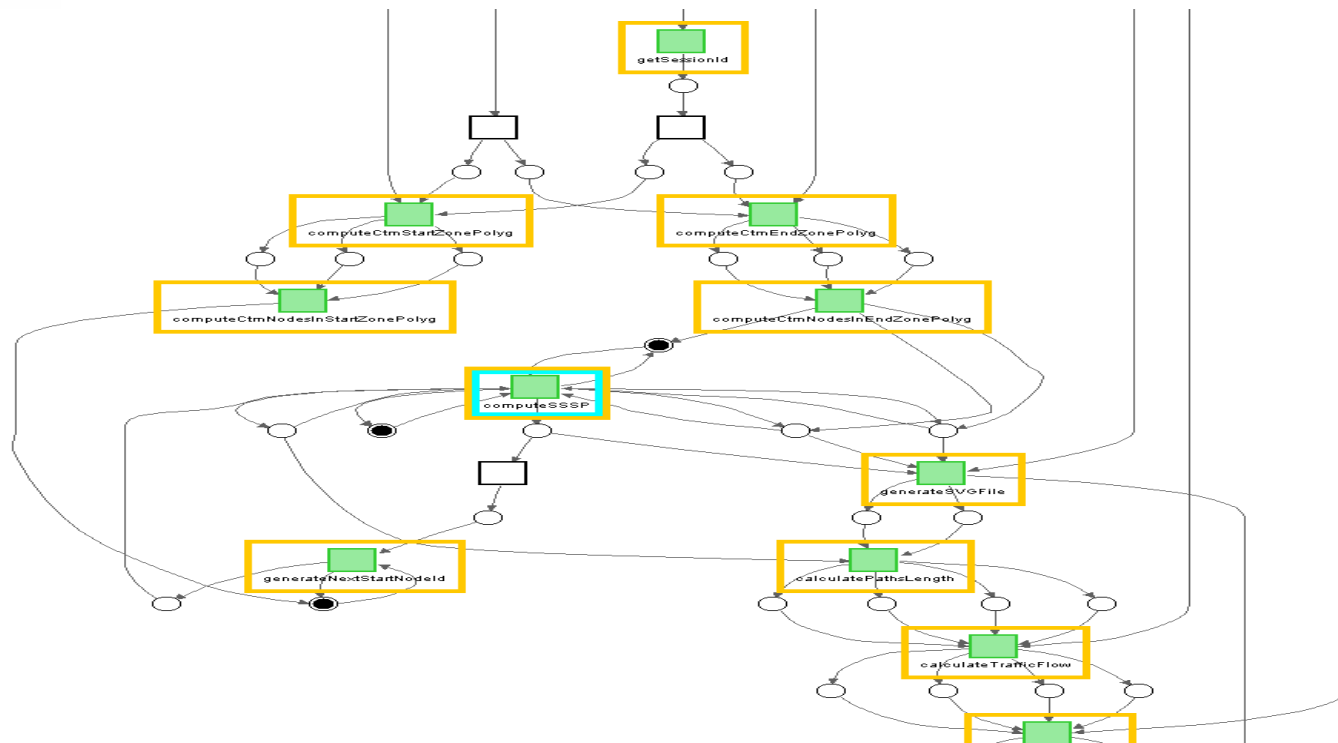
service.available=>http:goedis.dps.uibk.ac.at->grid02.softeco.it/WSDL/net-kwfgrip-ctm-NetFileParser-grid02.w

---

**Error Messages**

ResourceID	Availability
tcp:goedis.dps.uibk.ac.at->zeus72.cyf-kr.edu.pl	<b>Availability = 100 %</b> Start of monitoring: Thu Oct 05 17:19:02 GMT 2006 End of monitoring: Thu Oct 05 19:51:03 GMT 2006 Total number of measurements: 20 Average measurement interval: 480.06 sec. First UP: Thu Oct 05 17:19:02 GMT 2006 Last UP: Thu Oct 05 19:51:03 GMT 2006
http:petzeck.dps.uibk.ac.at->zeus72.cyf-kr.edu.pl:8080/wsrf/services/gom/service/GOMService?wsdl	<b>Availability = 100 %</b> Start of monitoring: Thu Oct 05 17:20:34 GMT 2006 End of monitoring: Thu Oct 05 19:52:31 GMT 2006 Total number of measurements: 20 Average measurement interval: 479.849 sec. First UP: Thu Oct 05 17:20:34 GMT 2006 Last UP: Thu Oct 05 19:52:31 GMT 2006
http:goedis.dps.uibk.ac.at->zeus72.cyf-kr.edu.pl:8080/wsrf/services/gom/service/GOMService?wsdl	<b>Availability = 100 %</b> Start of monitoring: Thu Oct 05 17:19:02 GMT 2006 End of monitoring: Thu Oct 05 19:51:03 GMT 2006 Total number of measurements: 20 Average measurement interval: 480.057 sec. First UP: Thu Oct 05 17:19:02 GMT 2006 Last UP: Thu Oct 05 19:51:03 GMT 2006
tcp:petzeck.dps.uibk.ac.at->zeus72.cyf-kr.edu.pl	<b>Availability = 100 %</b> Start of monitoring: Thu Oct 05 17:20:34 GMT 2006 End of monitoring: Thu Oct 05 19:52:31 GMT 2006 Total number of measurements: 20 Average measurement interval: 479.849 sec. First UP: Thu Oct 05 17:20:34 GMT 2006 Last UP: Thu Oct 05 19:52:31 GMT 2006

- ❑ **CTM (Coordinated Traffic Management) Workflows**
- ❑ **Web services deployed in Berlin, Bratislava, Cracow, Genoa and Innsbruck**



- ❑ [Short movie](#)

## □ 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>