

# **Ganglia Information Provider**

DRAFT

# Ganglia Information Provider

## Abstract

The Ganglia Information Provider collects information from two sources: the scheduler and the cluster information system (for example Ganglia or Hawkeye) using the XML mapping of the GLUE schema and reports it to a GRAM4 service, which publishes it as resource properties. The information from the scheduler and the cluster will be merged to form a single output resource property in the GLUE schema. The official Ganglia site states the following:

Ganglia is a scalable distributed monitoring system for high-performance computing systems such as clusters and Grids. It is based on a hierarchical design targeted at federations of clusters. It leverages widely used technologies such as XML for data representation, XDR for compact, portable data transport, and RRDtool for data storage and visualization

This provider publishes information such as the following:

- basic host data (name, ID)
- memory size
- OS name and version
- file system data
- processor load data
- other basic cluster data

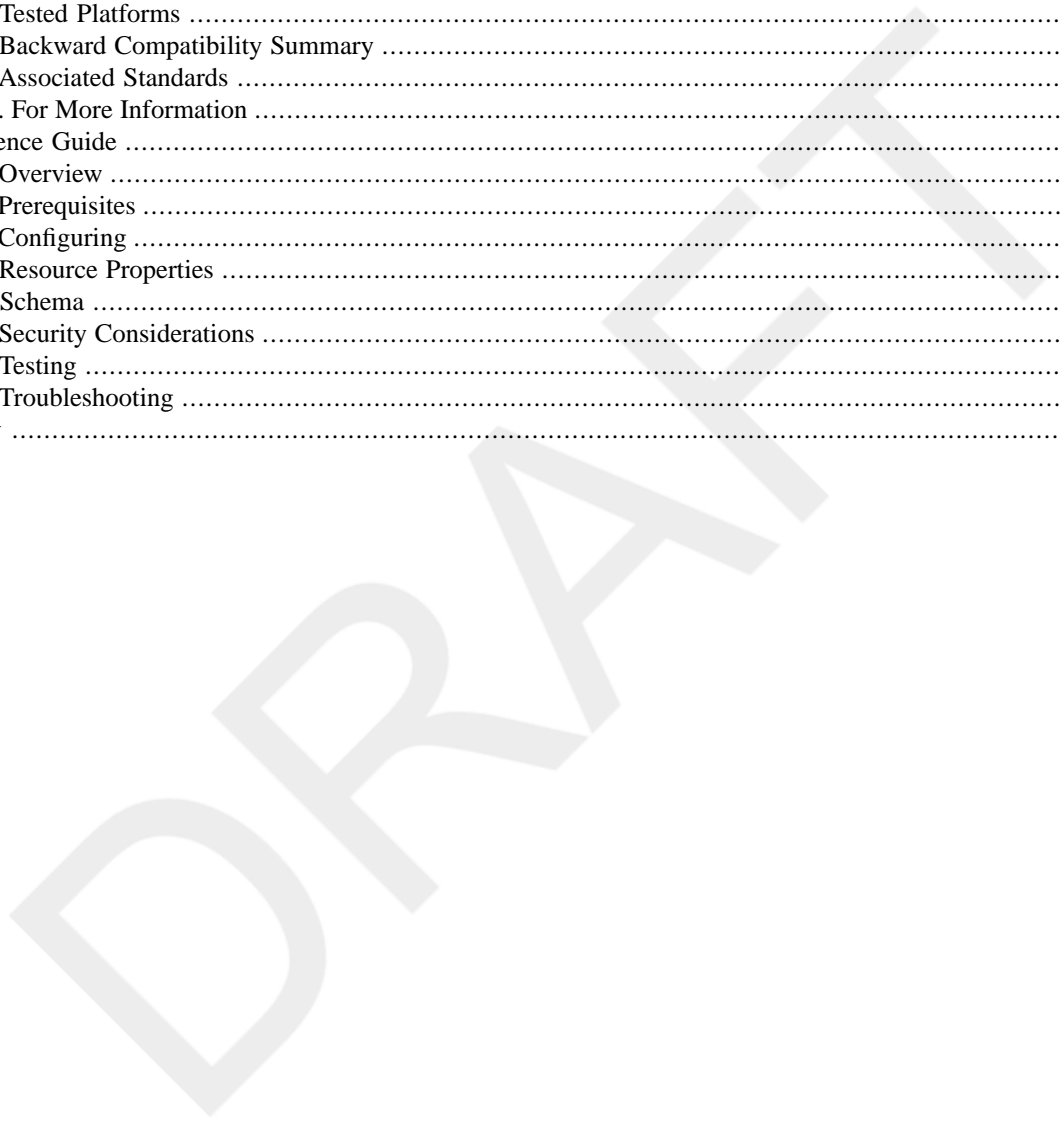
You can download a PDF version of Ganglia Information Provider information [here](#)<sup>1</sup>.

---

<sup>1</sup> ganglia.pdf

# Table of Contents

- 1. Release Notes ..... 1
  - 1. Component Overview ..... 1
  - 2. Feature Summary ..... 1
  - 3. Changes Summary ..... 1
  - 4. Bug Fixes ..... 1
  - 5. Known Problems ..... 1
  - 6. Technology Dependencies ..... 2
  - 7. Tested Platforms ..... 2
  - 8. Backward Compatibility Summary ..... 2
  - 9. Associated Standards ..... 2
  - 10. For More Information ..... 2
- 2. Reference Guide ..... 3
  - 1. Overview ..... 3
  - 2. Prerequisites ..... 3
  - 3. Configuring ..... 3
  - 4. Resource Properties ..... 5
  - 5. Schema ..... 6
  - 6. Security Considerations ..... 6
  - 7. Testing ..... 6
  - 8. Troubleshooting ..... 6
- Glossary ..... 7



# Chapter 1. GT 4.2.0 Release Notes: Ganglia Info Information Provider

## 1. Component Overview

The Ganglia Information Provider collects information from two sources: the scheduler and the cluster information system (for example Ganglia or Hawkeye) using the XML mapping of the GLUE schema and reports it to a GRAM4 service, which publishes it as resource properties. The information from the scheduler and the cluster will be merged to form a single output resource property in the GLUE schema. The official Ganglia site states the following:

Ganglia is a scalable distributed monitoring system for high-performance computing systems such as clusters and Grids. It is based on a hierarchical design targeted at federations of clusters. It leverages widely used technologies such as XML for data representation, XDR for compact, portable data transport, and RRDtool for data storage and visualization

This provider publishes information such as the following:

- basic host data (name, ID)
- memory size
- OS name and version
- file system data
- processor load data
- other basic cluster data

## 2. Feature Summary

Features new in release 4.2.0:

- This provider publishes information such as the following: basic host data (name, ID), memory size, OS name and version, file system data, processor load data, other basic cluster data.

## 3. Changes Summary

This is an existing information provider available from GT4 to GT 4.2.0.

## 4. Bug Fixes

There are ongoing fixed bugs for this information provider (see Bugzilla).

## 5. Known Problems

- There may currently be bugs for this information provider (see Bugzilla).

## 6. Technology Dependencies

This information provider depends on the following GT components:

- [Java WS Core](#)
- [GRAM4](#)

This information provider depends on the following 3rd party software:

- A working Perl installation
- A working Ganglia (gmond) installation

## 7. Tested Platforms

Tested Platforms for this information provider:

- N/A

Tested containers for this information provider

- Java WS Core container

## 8. Backward Compatibility Summary

This information provider works with all GT4 and WS MDS releases.

## 9. Associated Standards

Associated standards for this Information Provider:

- N/A

## 10. For More Information

See [Chapter 2, GT 4.2.0: Ganglia Information Provider Reference](#) for more information about this information provider.

# Chapter 2. GT 4.2.0: Ganglia Information Provider Reference

## 1. Overview

The Ganglia Information Provider collects information from two sources: the scheduler and the cluster information system (for example Ganglia or Hawkeye) using the XML mapping of the GLUE schema and reports it to a GRAM4 service, which publishes it as resource properties. The information from the scheduler and the cluster will be merged to form a single output resource property in the GLUE schema. The official Ganglia site states the following:

Ganglia is a scalable distributed monitoring system for high-performance computing systems such as clusters and Grids. It is based on a hierarchical design targeted at federations of clusters. It leverages widely used technologies such as XML for data representation, XDR for compact, portable data transport, and RRDtool for data storage and visualization

This provider publishes information such as the following:

- basic host data (name, ID)
- memory size
- OS name and version
- file system data
- processor load data
- other basic cluster data

This information provider is included in the GT4 toolkit and is used for reporting GLUE Computing Element (CE) information. The GLUE resource property (as used by GRAM) collects information from two sources: the scheduler and the cluster information system (for example Ganglia or Hawkeye). These will be merged to form a single output resource property in the GLUE schema.

## 2. Prerequisites

1. A working Perl installation
2. A working Ganglia (gmond) installation

## 3. Configuring

The following configuration is required to use this information provider

1. **Make sure usefulRP is enabled (enabled by default in 4.2.0+):** Find the service's or resource's `server-config.wsdd` file, in this case `$GLOBUS_LOCATION/etc/globus_wsrp_mds_index/server-config.wsdd`, and make sure there is a `DefaultIndexService Handler` section similar to the following (if it is not there, you can copy and paste the following):

```

<service name="DefaultIndexService" provider="Handler" use="literal" style="document"
  <parameter name="providers"
    value="org.globus.mds.usefulrp.rpprovider.ResourcePropertyProviderCollection
    org.globus.wsrfl.impl.servicegroup.ServiceGroupRegistrationProvider
    GetRPPProvider
    GetMRPPProvider
    QueryRPPProvider
    DestroyProvider
    SetTerminationTimeProvider
    SubscribeProvider
    GetCurrentMessageProvider"/>
  <parameter name="rpProviderConfigFile"
    value="/etc/globus_wsrfl_mds_usefulrp/gluece-rpprovider-sample-config.xml"
  <parameter name="scope" value="Application"/>
  <parameter name="allowedMethods" value="*/>
  <parameter name="handlerClass" value="org.globus.axis.providers.RPCProvider"/>
  <parameter name="className" value="org.globus.mds.index.impl.DefaultIndexService"/>
  <wsdlFile>share/schema/mds/index/index_service.wsdl</wsdlFile>
</service>

```

Verify two things:

- In the *provider* section, one of the values should be `org.globus.mds.usefulrp.rpprovider.ResourcePropertyProviderCollection`
  - There should be an *rpProviderConfigFile* parameter with the OS-native file path to an RPPProvider config file. There should be a sample included by default. You can add to this config file or make your own config file and just make sure the path is updated here. This file contains all required information for generating one or more Resource Properties for the hosting service or resource.
2. **Enable the provider:** Edit the RPPProvider config file that was specified in the previous step to include the config block for Ganglia. The following configures the GLUE Resource Property provider with element producers using Ganglia to provide cluster information and PBS for scheduler information. This sample configures the provider to generate cluster information using Ganglia on the localhost with the default Ganglia port, and configures PBS as the scheduler information provider. The period of execution is set to 300 seconds for each element producer, but may be configured separately if desired. This configuration mirrors a common information provider setup in the GRAM4 ManagedJobExecutable service. You can also use a command line client to generate a new configuration file. For more information, see FIXME. Using the RPPProvider Framework, it is possible to generate this information in other services as well:

```

<ns1:ResourcePropertyProviderConfigArray
  xsi:type="ns1:ResourcePropertyProviderConfigArray"
  xmlns:ns1="http://mds.globus.org/rpprovider/2005/08"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <ns1:resourcePropertyProviderConfiguration xsi:type="ns1:resourcePropertyProvider
  <ns1:resourcePropertyName xsi:type="xsd:QName" xmlns:mds="http://mds.globus.org/g
  <ns1:resourcePropertyImpl xsi:type="xsd:string">org.globus.mds.usefulrp.rpprovide
  <ns1:resourcePropertyElementProducers xsi:type="ns1:resourcePropertyElementProduc
  <ns1:className xsi:type="xsd:string">org.globus.mds.usefulrp.glue.GangliaElemen
  <ns1:arguments xsi:type="xsd:string">localhost</ns1:arguments>
  <ns1:arguments xsi:type="xsd:string">8649</ns1:arguments>

```

```

    <ns1:period xsi:type="xsd:int">300</ns1:period>
    <ns1:transformClass xsi:type="xsd:string">org.globus.mds.usefulrp.rpprovider.tr
  </ns1:resourcePropertyElementProducers>
  <ns1:resourcePropertyElementProducers xsi:type="ns1:resourcePropertyElementProdu
    <ns1:className xsi:type="xsd:string">org.globus.mds.usefulrp.rpprovider.produce
    <ns1:arguments xsi:type="xsd:string">libexec/globus-scheduler-provider-pbs</ns1
    <ns1:transformClass xsi:type="xsd:string">org.globus.mds.usefulrp.rpprovider.tr
    <ns1:period xsi:type="xsd:int">300</ns1:period>
  </ns1:resourcePropertyElementProducers>
</ns1:resourcePropertyProviderConfiguration>
</ns1:ResourcePropertyProviderConfigArray>

```

3. *Restart the container:* The information should now appear when querying the Index Service using [wsrf-query].

The configuration file format for the ResourcePropertyProviderCollection operation provider is simply the XML-serialized form of the ResourcePropertyProviderConfig stub object, as defined in the schema file rpprovider.xsd.

## 3.1. Generating an RPPProvider config file for GLUE

On resources running Ganglia:

1. Change working dir to \$GL/etc/gram-service-PBS (or -LSF, or -Condor, depending what you installed)
2. Run

```
mds-gluerp-configure pbs ganglia
```

If you're not using PBS, look for \$GLOBUS\_LOCATION/globus-scheduler-provider-\*. Use the value that appears there. You should see the following output:

```
Successfully wrote configuration output file to:
gluerp-config.xml
```

3. If Ganglia is running on the same (local) host as the container, and on the default port, then you can stop here. Otherwise, you will need to edit the gluerp-config.xml file to change the host and/or port. Open the file and look for the following lines:

```

<ns1:className xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xsi:type="xsd:string">org.globus.mds.usefulrp.glue.GangliaElementProducer
  </ns1:className>
<ns1:arguments xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xsi:type="xsd:string">localhost</ns1:arguments>
<ns1:arguments xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xsi:type="xsd:string">8649</ns1:arguments>

```

The last two *ns1:arguments* lines are the host and port parameters, respectively. Change them to match your Ganglia installation host and port.

## 4. Resource Properties

- The data gathered is published as part of the GLUECE RP

## 4.1. Namespace URI

- The GLUE namespace is: `http://mds.globus.org/glue/ce/1.1`
- The CE namespace is: `http://mds.globus.org/glue/ce/1.1`
- The XML namespace is: `http://www.w3.org/2001/XMLSchema`

## 5. Schema

- `$GLOBUS_LOCATION/share/schema/mds/usefulrp/ce.xsd`

## 6. Security Considerations

General security considerations associated with the container and all MDS services apply. See: [Aggregator Framework](#).

### 6.1. WS MDS Aggregator Services (Index Service and Trigger Service) Security Considerations

By default, the *aggregator sources* do not use authentication credentials -- they retrieve information using anonymous SSL authentication or no authentication at all, and thus retrieve only publicly-available information. If a user or administrator changes that configuration so that a service's aggregator source uses credentials to acquire non-privileged data, then that user or administrator must configure the service's aggregator sink to limit access to authorized users.

## 7. Testing

Once `gmond` is running and your configuration is complete, restart the container and the information should appear when querying the Index Service using `wsrf-query`.

## 8. Troubleshooting

Be sure that the `server-config.wsdd` and the `RPPProvider` configuration files are configured properly exactly as shown above. Simply having too much whitespace between the tags can cause errors.

# Glossary

## A

aggregator source

A Java class that implements an interface (defined as part of the Aggregator Framework) to collect XML-formatted data. WS MDS contains three aggregator sources: the query aggregator source, the subscription aggregator source, and the execution aggregator source.

DRAFT