caArray 1.6 *MAGE-OM* Grid Service API

Installation Guide

National Cancer Institute
Center for Bioinformatics

Table of Contents

Introduction

Microarray Gene Expression Object Model (MAGE-OM) is a data exchange model for microarray experiments which has been modeled using the Unified Modeling Language (UML). The *MAGE-OM API Installation Instructions* is intended to help system administrators to install the MAGE-OM Application Programming Interface (API). This guide describes how to install the three scenarios described in the *Overview of caArray MAGE-OM Grid Service API Installation* on page 2.

Overview of caArray

The caArray (http://caarray.nci.nih.gov/) software has been developed by the NCI Center for Bioinformatics (NCICB) to create an information-sharing network modeled on the World Wide Web. caArray consists of a microarray database and microarray data visualization and analysis tools. caArray is an open source project, and the source code and APIs are available in the download site at the NCICB web site, http://ncicb.nci.nih.gov/download/index.jsp. caArray is designed to make microarray data publicly available, and to develop and bring together open source tools to analyze these data.

caArray MAGE-OM API

The caArray MAGE-OM API is a set of Java objects that adhere to the object model defined by OMG's Gene Expression v1.1. The caArray MAGE-OM API objects provide access to data in the caArray database via Remote Method Invocation (RMI) call issued to a dedicated MAGE server at NCI or any other site with an accessible MAGE-OM server installation.

There are two primary types of objects defined in the API:

1. MAGE-OM-compliant interfaces
2. Custom MAGE-OM Impl (implementation) objects

The MAGE-compliant objects are defined as Java interfaces, which the custom MAGE-OM Impl Java classes implement. This ensures the custom MAGE-OM Impl provide a MAGE-OM compliant API.

The MGED Society website is an excellent source for supplemental material on the MAGE object model - http://www.mged.org/.

caArray MAGE-OM Grid Service API

The caArray MAGE-OM Grid Service API is a caGrid Data Service that exposes the functionality of MAGE-OM API over the grid. The Grid Service provides access to data in the caArray database via a web service call issued to a dedicated Grid Service server at NCI or any other site with an accessible caArray MAGE-OM Grid Service installation.

For more information about caGrid, see http://www.cagrid.org

The caGrid 1.0 User and Programmer Guides are excellent sources for additional material, see http://gforge.nci.nih.gov/frs/?group_id=25

NOTE:	Existing caArray development documentation can be found on the caArray page of the NCICB web site: http://caarray.nci.nih.gov/documentation. Existing caArray installation, technical and end-user documentation can be found on the download page of the NCICB website: http://ncicb.nci.nih.gov/download/downloadcaarray.jsp.

Overview of caArray MAGE-OM Grid Service API Installation	This installation guide outlines the steps necessary for installing MAGE-OM Grid Service API. This includes the steps for configuring, building and running a local instance of caArray Mage-OM Grid Service server. It includes the steps needed to build Mage-OM Grid Service server for deployment and access data from a local deployment of caArray at a non-NCICB center. Before installing the MAGE-OM Grid Service API, make sure your system has the minimum requirements as specified in the next section, *Minimal System Requirements*,

Minimal System Requirements

Minimal System Requirements

The hardware environment that has been tested and verified by NCICB includes the following:

- **Processor**: 1+GHZ Dual (SPARC) or 2+Ghz (x86 Intel)
- **Memory**: 1+ GB
- **Hard Drives**: 2+GB (Mirrored)

The amount of memory required depends on the amount of data stored in your database (see *Error! Reference source not found.* on page **Error! Bookmark not defined.**) and your systems performance will vary depending on hardware processing power.

MAGE-OM Grid Service Software and Technology Requirements

Required Software—Not Included in caArray

You must download and install the required software that is not included with the caArray MAGE-OM Grid Service (listed in Table 1 .) The software name, version, description, and URL hyperlinks (for download) are indicated in the table.

Where appropriate, apply patches to the installed applications.

Software Name	*Version*	*Description*	*URL*
Java 2 Platform Standard Edition 5.0 Update 6 (J2SE 5.0)	J2SE 5.0 Update 6 or higher	The J2SE Development Kit (JDK) supports creating J2SE applications.	http://java.sun.com/products/archive/j2se/5.0_06/index.html (select JDK 5.0 Update 6)
caGrid Infrastructure	1.0	The caGrid Infrastructure includes several dependent software packages including Globus Toolkit, Apache Tomcat and Apache Ant.	http://gforge.nci.nih.gov/frs/download.php/1450/caGrid-1.0_Installer.jar http://gforge.nci.nih.gov/frs/download.php/1453/caGrid-1.0_Installer.jar.md5 (signed)
caArray MAGE-OM API	1.5	The caArray MAGE-OM API for a local installation that refers to a non-NCICB installation of the caArray Portal	http://ncicb.nci.nih.gov/download/downloadcaarray.jsp
caArray Portal	1.5	The non-NCICB caArray Portal installation that will contain the data to be exposed over the grid.	http://ncicb.nci.nih.gov/download/downloadcaarray.jsp

Table 1 Required software and technology for the SDK

Configuring the System Environment

Complete the following steps to configure your system environment for all scenarios:

Step	Action
1	Download and install Java 2 SDK version 1.5.0_06+ (http://java.sun.com/javase/downloads/index.jsp) and set the JAVA_HOME environment variable to point to the SDKs installation directory. You should also put {JAVA_HOME}\bin directory on the front of your PATH variable.
2	Follow the instructions to install the caGrid 1.0 Infrastructure located at http://gforge.nci.nih.gov/frs/download.php/1454/caGrid-1.0_Installer_Instructions.pdf. The automated installer will install the dependent software packages, Apache Tomcat and Apache Ant.

Step	Action
3	Complete the installation of a local caArray Portal instance if one does not already exist. For more information, see the caArray 1.5 Data Portal Local Installation Guide, available at http://ncicb.nci.nih.gov/download/downloadcaarray.jsp.
4	Complete the installation of a local caArray MAGE-OM Server instance if one does not already exist. For more information, see the caArray 1.5 MAGE-OM_1.1 API Installation Guide, available at http://ncicb.nci.nih.gov/download/downloadcaarray.jsp.
5	Complete the installation.

Configuring Tomcat

Complete the following steps to configure the installed Tomcat container:

Step	Action
| 1 | Create a file called .java.security.policy in the <tomcatuser_home> directory. The contents of this file should be:<br>```
grant    {
         permission java.security.AllPermission;
         };
``` |
| 2 | Edit the $CATALINA_HOME/bin/catalina.sh file and add the following property at the beginning of the file:<br>```
JAVA_OPTS=-Djava.security.policy=<tomcatuser_home>/.java.security.policy
``` |

MAGE-OM Grid Service API Server Installation

This section describes the steps necessary to download and install the MAGE-OM Grid Service from the source distribution.

BEFORE YOU BEGIN	There must be a running instance of caArray MAGE-OM server available in order to deploy and run MAGE-OM Grid Service. Therefore, a running instance of caArray Portal must be available in order to deploy and run MAGE-OM server. The easiest configuration is to have caArray Portal and MAGE-OM server on the same machine. Follow the respective installation guides for assistance prior to completing this installation. For more information, see the caArray Download website, http://ncicb.nci.nih.gov/download/downloadcaarray.jsp.

MAGE-OM Grid Service API Server Dependency

MAGE-OM Grid Service (server-side) relies on an instance of the MAGE-OM server to be running. In other words, the Grid Service is a client of the MAGE-OM server and therefore, it depends on the MAGE-OM client binaries.

Downloading and Unzipping MAGE-OM Grid Service Files

Complete the following steps to download the appropriate MAGE-OM Grid Service files:

Step	Action
1	Go to the caCORE download web site: http://ncicb.nci.nih.gov/download/index.jsp.
2	Provide your email, name, and institution. Click **Enter the Download Center**.
3	Select **caArray**, agree to the caArray software license by selecting **Checking this box indicates that you agree to the above terms**, and click **Download.**
4	Select the appropriate MAGE-OM Grid Service download files and save them to your computer. **File Description** / **Example Filename** MAGE-OM Grid Service API Source Code — `caarray-mageom-gridservice-src.{version}.zip` MAGE-OM Grid Service API Java Documents — `caarray-mageom-gridservice-javadoc.{version}.zip`
5	Create a MAGE-OM Grid Service root directory (indicated by `{GRIDSVC_ROOT}`)
6	Copy `caarray-mageom-gridservice-src.{version}.zip` to `{GRIDSVC_ROOT}`.
7	Unzip `caarray-mageom-gridservice-src.{version}.zip` to `{GRIDSVC_ROOT}`. Assuming jar is in your path, type: ***jar -xvf caarray-mageom-gridservice-src.{version}.zip*** **OR** ***unzip caarray-mageom-gridservice-src.{version}.zip***

After unzipping MAGE-OM Grid Service API, the directory structure should resemble Figure 1.
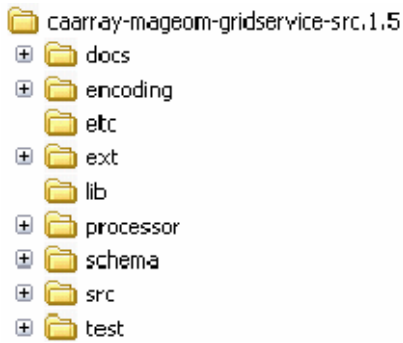
Figure 1 MAGE-OM Grid Service directory structure

Deploying MAGE-OM Grid Service

Complete the following steps to deploy MAGE-OM Grid Service:

Step	Action
1	Edit the {GRIDSVC_ROOT}/build.properties file, specifying the properties for your environment's configuration. See Table 2 for details.
2	Edit the {GRIDSVC_ROOT}/etc/serviceMetadata.xml. Update the <ns1: hostingResearchCenter> … </ns1: hostingResearchCenter> element block for your installation. The following is an example of what is used at NCICB: ``` <ns1:hostingResearchCenter> <ns69:ResearchCenter displayName="National Cancer Institute Center for Bioinformatics" shortName="NCICB" xmlns:ns69="gme://caGrid.caBIG/1.0/gov.nih.nci.cagrid.metadata.common"> <ns69:Address country="US" postalCode="20852" stateProvince="MD" street1="6116 Executive Blvd, Suite 403"/> <ns69:pointOfContactCollection> <ns69:PointOfContact affiliation="NCICB Application Support" email="ncicbmb@mail.nih.gov" firstName="" lastName="" role="maintainer"/> </ns69:pointOfContactCollection> </ns69:ResearchCenter> </ns1:hostingResearchCenter> ```
3	From {GRIDSVC_ROOT} type the following to configure your local installation: ***ant update-deploy-props update-service-props update-mageom-libs***
4	Stop the Tomcat instance.

Step	Action
5	From `{GRIDSVC_ROOT}` type the following to build the server: ***ant -Denv.CATALINA_HOME=work tar4Tomcat***
6	Untar the `caArrayGridSvc4Tomcat.tar` into `$CATALINA_HOME` directory*. Set owner of files that were untar'd to the \<user\> that runs the Tomcat instance (using chown)**. *If the build is not done on the destination server, the `caArrayGridSvc4Tomcat.tar` needs to be copied to the destination server. **You may need a system administrator (or super user access).
7	Because caArray uses a patched version of Castor 1.0.4, you must remove any other versions of Castor, namely the version provided by the caGrid installation, `$CATALINA_HOME/webapps/wsrf/WEB-INF/lib/castor-0.9.9.jar`.
8	Start the Tomcat instance.

Parameter	*Description*
`mageom-client.zip.file`	The file path to the corresponding mageom-client.zip that was built and deployed as a part of the MAGE-OM Server (non-NCICB) local installation. *Example:* `{MAGE-OM_ROOT}/target/local/mageom-client.zip`
`cqlQueryProcessorConfig_rMIServerURL`	The RMI URL for the SearchCriteriaHandler of your MAGE-OM Server (non-NCICB) local installation. *Example:* `//<publichostname>:<port>/SearchCriteriaHandler`
`cqlQueryProcessorConfig_secureSessionManagerURL`	The RMI URL for the SecureSessionManager of your MAGE-OM Server (non-NCICB) local installation. *Example:* `//<publichostname>:<port>/SecureSessionManager`
`index.service.url`	The appropriate index service for your local data service with which to register. *Example:* http://cagrid-index.nci.nih.gov:8080/wsrf/services/DefaultIndexService

Table 2 build.properties parameters

Testing the MAGE-OM Grid Service

Once the server is running, you should run some tests to verify that everything is working properly. The basic JUnit tests are defined in:
`{GRIDSVC_ROOT}/test/src/gov/nih/nci/cagrid/caArray/QueryTestCase.java.`

Complete the following steps to test MAGE-OM Grid Service installation:

Step	Action
1	Copy `{GRIDSVC_ROOT}\ext\test\lib\junit.jar` to `{ANT_HOME}\lib` to run the JUnit tests; the MAGE-OM API tests use JUnit (http://junit.sourceforge.net).
2	Make sure that the Jars in the library are in the classpath.
3	To compile the tests, type the following from the `{GRIDSVC_ROOT}` directory: *ant compileTests*
4	To run the tests, type the following from the `{GRIDSVC_ROOT}` directory: *ant test -Dtest.serviceUrl=<YourLocalGridServiceUrl>* *Example:* *ant test –* *Dtest.serviceUrl=http://localhost:8080/wsrf/services/caGrid/ CaArraySvc*

Troubleshooting

Should you have any problems while following the procedures in this installation guide, contact NCICB Application Support.

Contacting Application Support

NCICB   http://ncicb.nci.nih.gov/NCICB/support
Application   Telephone: 301-451-4384
Support   Toll free: 888-478-4423