

GridCom, Grid Commander: графический интерфейс для работы с задачами и данными в Гриде

В.В. Галактионов

Лаборатория информационных технологий ОИЯИ, Дубна

Программный пакет GridCom [1], разработан для обеспечения автоматизации доступа к средствам (задачам и данным) WLCG (Worldwide LHC Grid Computing) [2] – программно-вычислительного комплекса распределенной архитектуры, предназначенного для сбора, хранения и обработки больших объемов данных, поступающим с экспериментов на LHC (Large Hadron Collider). Для понимания и применения этого пакета ниже будут приведены основные сведения и основы функционирования и архитектуры WLCG.

Все операции с задачами и данными выполняются пользователем в командном режиме **CLI** (Command Language Interface). Громоздкие, неудобные и многословные конструкции языка управления задачами и данными (типа 40-символьных hash-кодов в идентификаторах задач и наименованиях файлов) представляют собой совершенно недружественный интерфейс пользователя к WLCG. Поэтому, одной из **главных целей** данной работы было обеспечить более удобные средства для пользователя в форматах уже известных современных графических средств оперирования данными и задачами, в частности с использованием средств WEB-доступа к программному обеспечению WLCG.

1. Job Management. Управление задачами

Работа с задачами в WLCG выполняется в несколько этапов.

Подготовка задачи заключается в ее описании на языке **JDL** (Job Definition Language) в текстовом файле, как правило, с расширением имени **.jdl**. Это описание содержит основные параметры задачи:

- тип задачи,
- название исполнительного модуля,
- перечень файлов с программами, которые составят передаваемый в Грид пакет,
- перечень файлов с данными, которые будут передаваться в Грид в пакете (StdInput, InputSandbox),
- перечень файлов для результатов счета задачи (StdOutput, StdError, OutputSandbox),

- перечень Грид-файлов, необходимых задаче для обработки данных и размещения результатов обработки (InputData и OutputData),
- другие требования задачи для выбора подходящего для счета **CE** (Computing Element).

Запуск задачи и контроль ее состояния выполняется пользователем в режиме CLI на UI (User interface)-машине:

Запуск задачи:

```
glite-wms-job-submit -a <jdl-файл>
```

Задача передается в **WMS** (Workload Management System) и после успешного синтаксического и содержательного анализа jdl-файла пользователь получает сообщение типа:

```
Connecting to the service
https://lcgwms01.jinr.ru:7443/
glite_wms_wmproxy_server
```

с именем WMS-сервера, принявшего задачу, а также идентификатор задачи **jobID** типа:

```
https://lcglb11.jinr.ru:9000
/HdG_Pv8WwtVoVo0HyBK0pA
```

Запрос **состояния** переданной задачи выполняется командой

```
glite-wms-job-status <jobID>.
```

При появлении статуса задачи типа **Done** (Done (Success) или Done (Failed)) можно затребовать на UI-машину результаты счета, описанные JDL-параметрами StdError и StdOut:

```
glite-wms-job-output <jobID> или
glite-wms-job-output -dir <dir_name>
<jobID>
```

Результаты счета записываются на UI-машине в установленной UI-конфигуратором директории (первый вариант) либо в указанной пользователем команде директории (второй вариант). После этого задача принимает состояние **Cleared**.

2. Data Management. Управление данными

В WLCG используются несколько типов операций с файлами:

- *Массовая пересылка* данных применяется специальными уполномоченными администраторами для перемещения больших объемов данных между основными хранилищами данных, например между Tier0 и Tier2. Эта задача выполняется сервисом **FTS** (File Transfer Service).

- *Пользовательский уровень* работы с файлами. Эти операции выполняются двумя группами команд: **lfc-*** - для работы с **LFC** (LCG File Catalogue)-каталогом и **LCG Data Management tools (lcm_util)** для непосредственных операций с файлами (заведение, копирование, удаление и др.).
- *Специальный* (нижний уровень) работы с файлами: **GSIFTP** (edg-gridftp-* и globus-url-copy).

2.1. Операции с файлами пользовательского уровня

Upload, копирование файла из локальной UI-машины на SE в WLCG и включение его имени в LFN (Logical File Name)-формате в LFC-каталог. Обязательные параметры команды – название виртуальной организации (для правильного выбора LFC-каталога) и имя локального файла в UI-машине. Остальные атрибуты Грид-файла (**SE** (Storage Element) и LFN-имя файла) устанавливаются автоматически (по умолчанию) либо могут быть определены пользователем. В любых операциях с LFC-каталогом должно быть установлено значение переменной **LFC_HOST**.

Download, копирование файла из хранилища SE в локальную на UI-машине файловую систему.

3. Архитектура GridCom

Программный пакет GridCom состоит из двух частей: *клиентской* части **GridCom** и *исполнительной* - **Executor**.

Клиент GridCom не требует установки на конечных пользовательских компьютерах каких-либо специальных программ, выполняется в операционных системах Windows или UNIX стандартными сетевыми приложениями-браузерами, обеспечивает интерактивное взаимодействие с пользователем в графическом режиме с применением Java-технологий:

- Апплет-структура пакета обеспечивает WEB-интерфейс пользователя к исполнительной части. Этим обеспечивается популярная в Сети парадигма клиента: минимум нагрузки и инсталляций для машин потребителя.
- Две части клиентской программы **Job Management** (Управление задачами) и **Data Management** (Управление данными) выполняются в разных окнах и работают совершенно независимо. Программа формирует запросы пользователя в виде записей в базу данных (*семафоры* и данные), запоминает в очереди и периодически в режиме polling сканирует состояние запроса.

- Существует вариант GridCom, выполненный в виде Java-приложения.

Серверный компонент Executor выполняется только в специальных вычислительных системах с программным обеспечением доступа к WLCG – UI-машинах. Таким, например, является LINUX-кластер ЦИВК ЛИТ ОИЯИ. Executor получает запросы пользователя и формирует из них наборы команд в формате CLI для работы с локальной файловой системой и с задачами и данными WLCG. Для эффективного функционирования серверная программа выполняет запросы, используя *поточковый* механизм распараллеливания процессов, группируя их по типам (работа с задачами, данными и др.).

База данных. В GridCom используется база данных общего пользования MySQL из стандартного программного обеспечения UNIX-кластера ЛИТ ОИЯИ. Для MySQL JDBC-драйверы содержатся в свободно распространяемом пакете **mysql-connector-java-3.1.12-bin.jar**.

4. Управление задачами в GridCom

Как уже упоминалось выше управление задачами в GridCom выполняется загружаемым через Интернет Java-апплетом (WEB-интерфейс) графическими изобразительными средствами и интерактивными запросами пользователя (рис.1). Все операции сводятся к выбору объекта (операция **Select**) в окне (мышью), минимальные действия с ним (редактирование) и выбора действия с ним в роруп-меню. Выбранный пользователем тип действия над задачей трансформируется в команды **gLite** [4], которые, используя механизм **Runtime**, выполняются серверным компонентом **Executor** в операционной системе UI-машины. Основные операции с задачами:

- подготовка задачи,
- передача задачи в WMS,
- получение результатов счета задач.

5. Управление данными в GridCom

Управление данными в GridCom выполняется в отдельном окне “**Data management**”(Рис. 2) независимо от окна управления задачами. Основные типы операций с данными в WLCG описаны выше. Как и в случае окна “Job Management” в клиентской части GridCom графические средства используются для выбора объекта, типа операции над ним и формирования запроса в исполнительную часть – Executor, в котором эти запросы формируются в систему gLite-команд и выполняются на UI-машине.

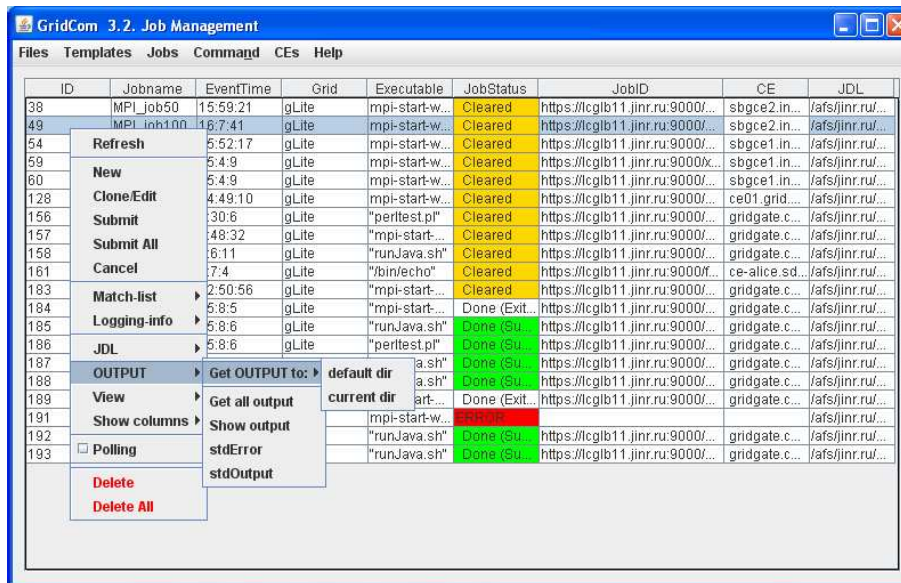


Рис. 1: Окно управления задачами и роруп-меню

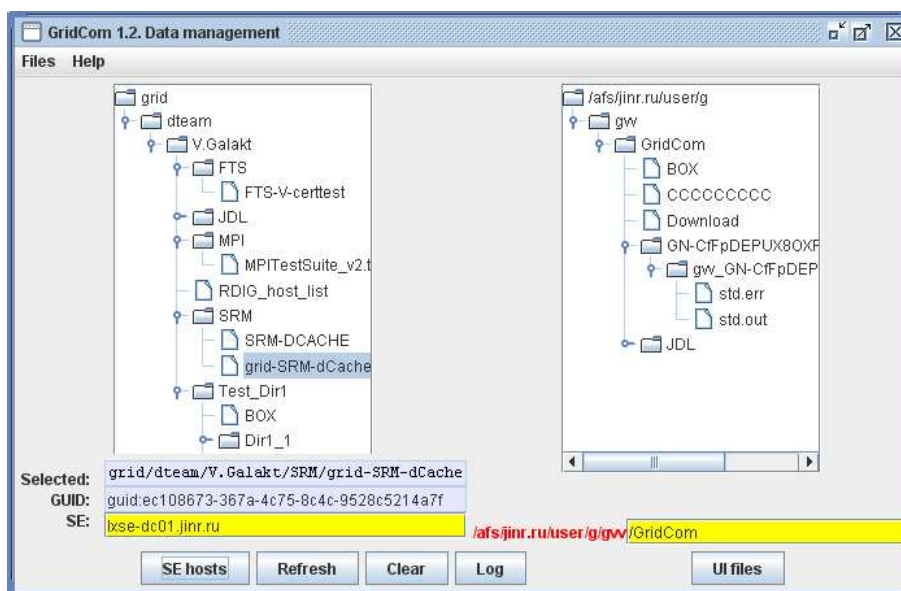


Рис. 2: Окно управления данными

References

[1] В.В. Галактионов. GridCom, Grid Commander: графический интерфейс для работы с задачами и данными в Гриде. Сообщение ОИЯИ, P10-2011-55, Дубна, 2011.

[2] Worldwide LHC Computing Grid.
<http://lcg.web.cern.ch/lcg/>

[3] gLite 3.1 User Guide.
<https://edms.cern.ch/file/722398/1.3/gLite-3-UserGuide.html>.