

## РАСШИРЕНИЕ ВОЗМОЖНОСТЕЙ КЛАСТЕРНЫХ СИСТЕМ УПРАВЛЕНИЯ ДЛЯ ИНФОРМАЦИОННОГО ОБСЛУЖИВАНИЯ ГРИД-ДИСПЕТЧЕРА\*

**Коваленко Е.И., Шорин О.Н.**

Институт прикладной математики им. М.В.Келдыша РАН;  
Россия, 125047, Москва, Миусская пл. 4; тел. (095)250-79-82,  
kei@keldysh.ru, shorin@keldysh.ru

### **Введение.**

В Институте прикладной математики им. М.В. Келдыша разработан программный комплекс Грид-диспетчер [1]. Основной особенностью планирования в нем является использование локальных расписаний расхода кластерных ресурсов на будущее время. Поставку таких расписаний должны обеспечивать все кластеры, находящиеся под управлением Грид-диспетчера. Существующие кластерные системы управления такую функциональность не обеспечивают, и в этом контексте возникла задача расширения их возможностей функцией построения расписания.

Эта функция должна выполняться синхронно с работой локального планировщика системы управления пакетной обработкой кластера, осуществляя обновление локальных расписаний и пересылку их Грид-диспетчеру в соответствии с событиями, приводящими к их изменению.

Для решения этой задачи была разработана и реализована компонента локальной кластерной системы – Агент Грид-диспетчера.

### **Задачи Агента Грид-диспетчера.**

Задача планирования в среде Грид во многом аналогична той, которая решается на локальном уровне системами управления пакетной обработкой (СПО). Только теперь задания должны распределяться из глобальной очереди по исполнительным узлам, которые сами являются сложными кластерными системами.

Планировщику Грид-диспетчера (Метапланировщику) необходимо оценивать возможности запуска глобальных заданий в каждом исполнительном кластере. Между Грид-диспетчером и СПО заключается договоренность о правилах предоставления локальных ресурсов глобальным заданиям. В нашем проекте условием предоставления ресурса является его цена, вычисляющаяся динамически в зависимости от того, какова приоритетность локального задания, претендующего на данный ресурс. На основании таких соглашений и дополнительной информации о состоянии кластерных ресурсов и их загрузки Метапланировщик принимает решение о возможности запуска глобального задания в конкретном кластере на конкретной машине.

---

\* Работа выполнена при поддержке Российского фонда фундаментальных исследований (проекты 02-01-00282 и 04-07-90299).

Информация, поступающая от кластеров, должна включать набор статических параметров кластерных узлов (их ОС, архитектуру, число процессоров, объем памяти и др.), позволяющих Метапланировщику выбрать подмножество кластеров, подходящих для работы данного глобального задания. Для выбора же единственного, окончательного исполнителя этого задания нужна еще информация о загрузке ресурсов в этих кластерах. Поскольку в Грид масштабы и условия сетевого взаимодействия отличаются от локальных, Метапланировщику необходимо иметь информацию о состоянии ресурсов заранее. Этого можно достичь, строя прогноз использования кластерных ресурсов в будущем и получая тем самым **локальные расписания** использования ресурсов. Локальное расписание содержит адреса и объемы отводимых заданиям кластера ресурсов, а также время их занятия и освобождения. Взаимодействуя со всеми кластерными системами и получая от них локальные расписания, Метапланировщик строит общее, глобальное расписание использования ресурсов на будущее. Создание локальных расписаний и их доставка Грид-диспетчеру – основная задача Агента.

Для сбора информации и построения расписаний Агенту необходимо взаимодействовать с локальным планировщиком кластерной системы управления, поскольку именно локальный планировщик принимает решение о запуске всех заданий в кластере и, следовательно, именно его данные о состоянии кластерных ресурсов существенны для работы Метапланировщика. Следует отметить, что такая информация (данные о конфигурируемых ресурсах) не всегда совпадает с теми данными, которые можно получить непосредственно от ОС машины или от других объектов СПО.

Информация, собранная Агентом, должна передаваться Грид-диспетчеру в те моменты, когда она обновляется. Поэтому Агенту необходимо отслеживать кластерные события, приводящие к ее изменению, и инициировать передачу новой информации Грид-диспетчеру.

Таким образом, Агент, установленный в каждой исполнительной кластерной системе, должен решать следующие задачи:

- обеспечивать доступ к нужной Грид-диспетчеру информации, хранящейся у локального планировщика;
- фиксировать события, изменяющие эту информацию;
- по событиям получать или обновлять информацию;
- формировать локальное расписание использования ресурсов в будущем времени;
- организовывать доставку обновленной информации Грид-диспетчеру.

### **Реализация агента Грид-диспетчера**

Наш Агент Грид-диспетчера создан на базе системы управления пакетной обработкой OpenPBS [1] с внешним планировщиком MAUI [2]. Выбор системы OpenPBS связан, главным образом, с ее большей популярностью среди СПО, *хотя* Агент будет работать с любыми системами, которые совместимы с планировщиком MAUI. Замена штатного планировщика OpenPBS на MAUI для нас существенна, так как большой набор его возможностей способствует реализации основных функций Агента.

Среди многих достоинств MAUI для нас важно то, что это открытый продукт и можно получить тексты его программных модулей. Кроме того, наличие специального режима MAUI (SIMULATION) позволяет моделировать процесс размещения заданий в кластере в будущем времени, обеспечивая механизм генерации локальных расписаний. И, наконец, в MAUI есть механизм предварительного резервирования ресурсов, а это позволяет обеспечить гарантированный запуск глобальных заданий.

Для работы Агенту необходимо иметь доступ к данным, размещенным в информационных структурах MAUI. К сожалению, в настоящее время нет готового API для доступа к такой информации, поэтому Агент реализован путем модификации MAUI307p6, то есть нужные программные блоки внесены непосредственно в его программные модули (на рис.1. это блоки A-1 и A-2). В результате мы получили модифицированный планировщик, который кроме своих действий выполняет и ряд функций по предоставлению информации Агенту.

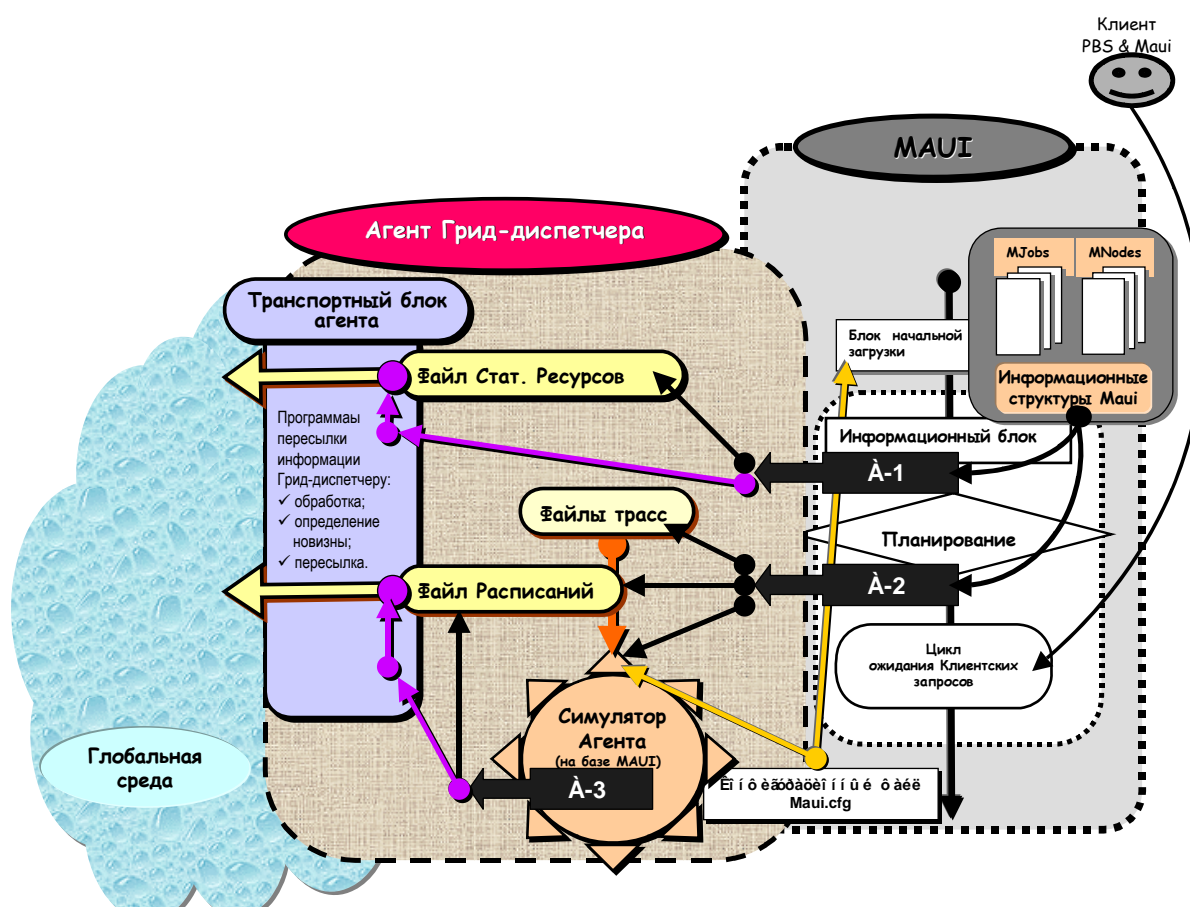


Рис.1. Схема взаимодействий между компонентами MAUI и Агентом.

В каждом цикле работы модифицированного MAUI выполняется проверка наличия событий, обновляющих информацию. Если такое событие происходит, новая информация о статических и динамических параметрах кластерных объектов

выводится в файлы, которые далее используются для построения локального расписания. Агент формирует два выходных файла: файл статических ресурсов и файл расписания. Статические ресурсы обновляются при появлении нового узла в кластере, изменении состояния уже существующих узлов (рабочее–нерабочее) или при изменении конфигурационного файла MAUI. Файл расписания обновляется при каждом поступлении нового задания в кластер или при новом старте задания из очереди.

Основная задача Агента – построение локальных расписаний. Полное локальное расписание использования кластерных ресурсов должно определять времена стартов и окончаний всех заданий кластерной очереди, вместе с адресами и объемами отводимых им ресурсов. Поскольку рассматривается планирование для однопроцессорных заданий и компьютеров, то Грид-диспетчеру достаточно ограниченной информации. В настоящее время Агент строит расписание только для двух последовательных заданий на каждом кластерном узле (уже запущенного и следующего за ним – прогнозируемого).

Если информация об уже работающих заданиях присутствует в информационных структурах MAUI, то для заданий, на данный момент только ожидающих запуска, информации о том, где и когда они будут запущены, нет. Она может быть получена в результате моделирования процесса обработки кластерной очереди. Для этого мы используем возможности MAUI работать в режиме моделирования (SIMULATION). При необходимости построения расписания запускается новый процесс – Симулятор Агента (работает модифицированная программа MAUI307p6 в режиме SIMULATION). В результате его работы формируется прогноз о будущих запусках.

Для работы Симулятора создаются входные файлы (файлы трасс), в которых содержится информация о параметрах всех ресурсов и заданий кластера в данный момент времени. Формат файлов трасс отличается от обычных файлов трасс MAUI. Это объясняется тем, что для Симулятора Агента необходимы конкретные параметры для каждого прогнозируемого задания (время старта и окончания, приоритет запуска и т.д.), а не среднестатистические оценки, используемые в стандартном симуляторе MAUI. В связи с этим Агент использует модифицированную программу генерации файлов трасс.

Построение расписания происходит по событиям, приводящим к его изменению. При этом обновляются файлы трасс и инициируется работа Симулятора. Симулятор для своей работы использует значения параметров из конфигурационного файла планировщика MAUI. В качестве начальной информации Симулятор загружает данные о кластерных ресурсах и заданиях из файлов трасс и моделирует работу MAUI, то есть генерирует прогноз, как будут размещены задания по свободным ресурсам. После получения прогноза для всех кластерных узлов процедура моделирования завершается. В результирующий файл расписаний Агент помещает информацию об уже стартовавших заданиях (их ресурсных запросах, запрашиваемом времени выполнения, приоритете и др.) и заданиях, которые по прогнозу будут запущены следующими на соответствующие ресурсы.

При обновлении выходных файлов Агента происходит запуск его транспортных служб. Программы этого блока проверяют содержимое файлов статических ресурсов и

расписаний и готовят информацию для транспортировки. При наличии новых данных посредством Грид-служб[3] происходит их пересылка Грид-диспетчеру.

### ЛИТЕРАТУРА

[1]. Коваленко В.Н., Коваленко Е.И., Корягин Д.А., Любимский Э.З., Хухлаев Е.В., Шорин О.Н. Грид-диспетчер: реализация службы диспетчеризации заданий в Грид. Сборник докладов международной конференции «Распределенные вычисления и Грид-технологии в науке и образовании», Дубна, 2004

[1]. <http://www.OpenPBS.org>

[2]. <http://www.supercluster.org/maui>

[3]. Tuecke S., Czajkowski K., Foster I., Frey J., Graham S., Kesselman C. Grid Services Specification, p. 1-47, 2002.

<http://www.globus.org/research/papers/gsspec.pdf>