

Веб-инструментарий для научных исследований: текущее состояние и перспективы развития

А.П.Крюков, А.П.Демичев

НИИЯФ МГУ

ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014 - 2020 годы»
Соглашение № 14.604.21.0146

Грид → веб-платформы

- рост производительности отдельных ресурсов ⇒ концепция грида стала терять значительную часть своей привлекательности
 - масштабные грид-инфраструктуры ⇒ высокие накладные расходы
 - необходимость объединяющей организационной структуры (WLCG — CERN)
 - результаты проекта EGEE
 - центр внимания при разработке промежуточного ПО нового поколения смещается к построению удобных и эффективных средств доступа к отдельным крупным ресурсам
 - **веб-платформы**

Удаленный доступ к ресурсам (1/2)

- прямой доступ по протоколу ssh
 - гибкое использование возможностей ресурсов
 - высокие требования к уровню компьютерных знаний у пользователя
- часто требуется запускать большое количество **ОДНОТИПНЫХ** вычислительных заданий
 - модель "ПО как услуга" (Software as a Service, SaaS)
 - **Веб-платформа (ВП)**: совокупность специальных веб-сервисов + прикладные веб-интерфейсы для удаленного доступа к вычислительным ресурсам
 - *предметно-ориентированный веб-портал, виртуальная лаборатория, веб-ориентированный производственно-исследовательский центр и т.п.*

Удаленный доступ к ресурсам (2/2)

- заранее подготовленное задание (*запуск ППП, обращение к хранилищу данных и т.п.*) = "инструмент"
- набор ППП = композитные задания (поточные задания, workflow)
- предварительная настройка технических деталей (интернет-адреса ресурсов и т.п.) ⇒ оставляет пользователю формулировку существа конкретного запроса на естественном языке
- ВП могут предоставлять сервисы, для удаленного разворачивания авторизованными пользователями новых ППП (инструментов)

Типы веб-платформ

1. ВП запуска заданий: удаленный запуск, контроль выполнения и получение результатов заданий;
2. ВП запуска заданий и установки ПО;
3. веб-хабы: функции пп. 1, 2 + проф. соц. сети
- обмен опытом использования инструментов платформы, формирование рейтинга ППП;
4. ВП рынка прикладного ПО: (функции пп. 1, 2, 3) + взаимодействие поставщиков и потребителей ППП (аналог AppStore, Google Play и т.п.)

Основные служебные функциональные требования к ВП (1/2)

- Управление мандатами пользователей, дающими право на использование ресурсов
 - аутентификация, авторизация, делегирование
- Удаленное администрирование ВП посредством веб-браузера
 - учетные записи пользователей, включая предоставление прав пользования группами инструментов; учетные записи поставщиков услуг; управление выполнением заданием, включая отмену выполнения; аудит/журналирование заданий

Основные служебные функциональные требования к ВП (2/2)

- Управление выполнением заданий
 - композиция заданий из имеющихся инструментов; трансляция команд в формат ресурса; запуск заданий; мониторинг; получения результатов: возможность удаленного просмотра
- Управление передачей файлов данных
 - передача входных/выходных данных с/на локальный компьютер пользователя/ сервер хранения
- Управлением инструментами (услугами)
 - регистрация новых; создание шаблонов заданий; управление мандатами доступа на ресурсы

Администрирование ВП

- наиболее гибкий подход основан на понятии виртуальной организации (ВО)
 - динамические сообщества людей и/или учреждений, которые совместно используют вычислительные ресурсы в соответствии с согласованными между ними правилами
- в задачу менеджера ВО входит создание и пополнение репозитория инструментов для данной ВО

Особенности ВП с функцией удаленной установки прикладного ПО

- различные типы ресурсов, **если** прикладное ПО разворачивается **администраторами** ресурсов
 - инструментарий для облегчения взаимодействия разработчиков ПО и администраторов ресурсов: средства преобразования прикладного ПО в SaaS
 - NanoHUB/HUBZero
- самостоятельная установке прикладного ПО
 - ресурсы - облачные системы, виртуальные машины
 - при установке ПО используется модель предоставления ресурсов "платформа как сервис" (Platform as a service; PaaS).

Примеры конкретных разработок (1/4)

- Учебно-методического программного комплекса "Многомасштабное моделирование в нанотехнологиях" (ФГБУН Центр фотохимии РАН; <http://www.nanomodel.ru>);
- система «Персональный виртуальный компьютер» (Южно-Уральский государственный университет; <http://supercomputer.susu.ac.ru/pvc>);
- технологическая платформа программы «Университетский кластер» Unihub (Институт системного программирования РАН; <https://unihub.ru>);

Примеры конкретных разработок (2/4)

- вычислительная облачная платформа УрО РАН (Институт математики и механики УрО РАН);
- веб-портал системы управления суперкомпьютером (Институт кибернетики НАН Украины; <http://melkon.com.ua/ru/cms>);
- многофункциональная инструментально-технологическая платформа поддержки облачных вычислений CLAVIRE (ИТМО; <http://clavire.ru>);
- веб-платформа Everest [15] (ИППИ РАН; <http://everest.distcomp.org>)

Примеры конкретных разработок (3/4)

- веб-хаб в области нанотехнологий nanoHUB (консорциум университетов США; <http://www.nanohub.org>);
- веб-платформа для удаленного запуска заданий eQUEUE (Advanced Clustering Technologies, Inc.; <http://www.advancedclustering.com/products/software/equeue>);
- научный веб-портал Nucleonica (Institute for Transuranium Elements; <http://www.nucleonica.net>);

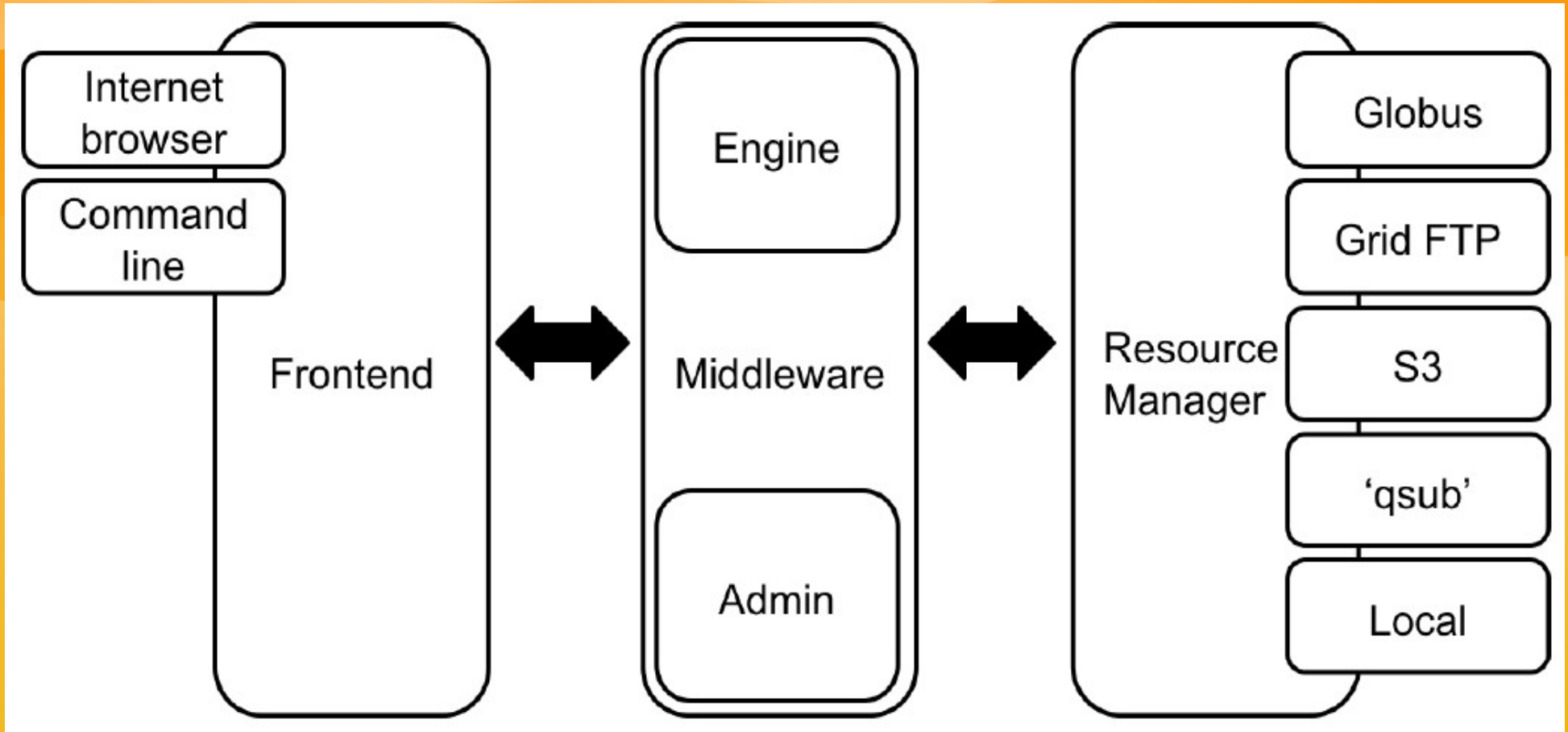
Примеры конкретных разработок (4/4)

- веб-платформа WebMO (Hope College, Holland, USA; <http://www.webmo.net>);
- веб-платформа e-Science Central [18] (Newcastle University, UK; <http://www.esciencecentral.co.uk>).
- веб-платформа Yabi [17] (Centre for Comparative Genomics, Murdoch, Australia; <https://ccg.murdoch.edu.au/yabi>)

Проект Yabi

- выполнение композитных заданий, состоящих из **последовательного** ряда задач
- развитый и удобный административный интерфейс для
 - конфигурирования "инструментов"
 - управления правами доступа пользователей
- ППО состоит из двух *основных* компонентов:
 - модуля администрирования
 - модуля выполнения композитных заданий (движок Yabi)
- GNU GPL v3 (некоммерческое использование)

Проект Yabi: архитектура



Проект Yabі: ресурсы

- менеджер ресурсов Yabі предоставляет два типа сервисов:
 - вычислительные ресурсы
 - ресурсы хранения данных
- оба типа имеют расширяемую (plug-in) архитектуру
 - позволяет стыковать их различными типами вычислительных ресурсов и хранилищ данных
- для взаимодействия с внешними сервисами используются мандаты, зашифрованные от имени пользователя (с помощью пары логин-пароль)

Проект Yabi: интерфейс

- Графический веб-интерфейс (браузер) + CLI
- Веб-интерфейс — три вкладки
 - Создание/редактирование WF
 - Созданные ранее WF для повторного использования
 - Автоматические фильтры возможных зависимостей
 - Задания: доступ к ранее созданным заданиям
 - Поиск по имени, дате и т.п.
 - Файлы: доступ к файлам на внешних хранилищах
 - Обеспечивает типичные операции — копирование, уничтожение, переименование

Проект Yabi: веб-интерфейс

The screenshot displays the Yabi web interface with the following components:

- Header:** Yabi logo and navigation tabs for 'jobs', 'design' (active), and 'files'. Copyright information for 2006 and 2011 CCG, Murdoch University, and a 'log out ebiuser6' button are also present.
- Left Panel:** A sidebar with a search bar (Find tool: in:*) and a 'show all' button. Below it, a 'Use selection to auto-filter?' toggle is set to 'on'. A list of tools is provided, including 'unix', 'select data', and 'EBI' tools like 'eprimer3', 'fetchdb_by_id', 'clustalw2', 'split fasta', 'fetchdb', 'fasta', 'ebi_ncbiblast', 'emboss_pepinfo', 'emboss_tmap', 'emboss_sirna', 'emboss_getorf', and 'ebi_wublast', each with an 'add' button.
- Main Panel:** A workflow editor titled 'ebi webservice blast' with a 'run' button. It shows a sequence of steps: 'start', '1 - select file', and '2 - ebi_ncbiblast'. The '2 - ebi_ncbiblast' step is highlighted in blue and shows configuration options for 'accepts' (fn, fastq, frn, fa, fna, fasta, faa) and 'outputs' (*).
- Right Panel:** A configuration panel for 'Options for 2 - ebi_ncbiblast' with a 'show all options' button. It includes dropdown menus for '--format' (out), '--program' (blastn), and '--database' (EMBL Mammal). It also features a '--stype' dropdown (DNA/RNA) and a '--sequence' dropdown (1 - select file) with associated file format buttons (fn, fastq, frn, fa, fna, fasta, faa).

Проект Yabi: администрирование (1/3)

- Администрирование посредством веб-интерфейса (браузер)
 - Создание новых инструментов
 - Веб-форма описания
 - Контроль доступа пользователей к инструментам и файлам
 - Контроль использования инструментов и ресурсов

Проект Yabi: администрирование (2/3)

Создание инструментов



The screenshot displays the 'Change tool' configuration page in the Yabi administration interface. The page is titled 'Change tool' and is located under the path 'Home > Yabi > Tools > Imp_input'. The interface includes a navigation bar with links for 'jobs', 'design', 'files', 'account', and 'admin'. The configuration fields are as follows:

- Name:** Imp_input (Unique toolname for internal use.)
- Display name:** Imp_input (Tool name visible to users.)
- Path:** /home/yabiapp/yabi/Imp (The path to the binary for this file. Will normally just be binary name.)
- Description:** Running LAMMPS with an Input file as a parameter and output to STDOUT (staging out). (The description that will be sent to the frontend for the user.)
- Enabled** (Enable tool in frontend.)
- Backend:** Example Execution Server - ssh://tb31.ngrid.ru:22/ (The execution backend for this tool.)
- Fs backend:** Example File Server - scp://tb31.ngrid.ru:22/home/ (The filesystem backend for this tool.)
- Accepts input** (If checked, this tool will accept inputs from prior tools rather than presenting file select widgets.)

Проект Yabi: администрирование (3/3)

Информация о
выполненных
заданиях

Yabi Administration						Welcome, andrew . Change password / Log out
Home > Backend image_gen (2)						
Workflow Summary						
Name	User	Start	End	Log File Path	Status	
image_gen (2)	andrew	2011-06-15 15:54:56	2011-06-15 15:59:23	/	complete Edit	
Order	Status	Start	End	Command		
0	complete	2011-06-15 15:34:57.506788	None	<SelectFile> ({})	Edit	
	complete	Task:			Edit Syslog JSON	
				Src: gridftp://amacgregor@xe-gt4.ivec.org/scratch/bi01/amacgregor/0demo_files/MYD01.A2008271.0430.005.2008271161919.hdf Dst: scp://andrew@yabi.localdomain:22/data/yabi/sge/b66daa71-3c56-47db-bc02-d3459a20cc3f/input/MYD01.A2008271.0430.005.2008271161919.hdf	Edit	
1	complete	2011-06-15 15:34:57.565990	None	<Command:modis_L1A_to_GEO.csh> ([<BatchSwitch:inputFile= Value:None>])	Edit	
	complete	Task: modis_L1A_to_GEO.csh "/scratch/bi01/amacgregor/4b992e9c-6720-4198-8aef-413838539cc7/input/MYD01.A2008271.0430.005.2008271161919.hdf"			Edit Syslog JSON	
				Src: scp://andrew@sooty.localdomain:22/export/home/tech/macgregor/yabi/image_gen (2)/1 - select file/MYD01.A2008271.0430.005.2008271161919.hdf Dst: gridftp://amacgregor@xe-gt4.ivec.org/scratch/bi01/amacgregor/4b992e9c-6720-4198-8aef-413838539cc7/input/MYD01.A2008271.0430.005.2008271161919.hdf	Edit	
2	complete	2011-06-15 15:34:57.612145	None	<Command:modis_L1A_to_L1B.csh> ([<BatchSwitch:inputFileHDF= Value:None>, <BatchSwitch:inputFileGEO= Value:None>])	Edit	
	complete	Task: modis_L1A_to_L1B.csh "/scratch/bi01/amacgregor/14170cb7-0941-45d6-89d0-bc56b4d1b9e8/input/MYD01.A2008271.0430.005.2008271161919.hdf" "/scratch/bi01/amacgregor/14170cb7-0941-45d6-89d0-bc56b4d1b9e8/input/A2008271043000.GEO"			Edit Syslog JSON	
				Src: scp://andrew@sooty.localdomain:22/export/home/tech/macgregor/yabi/image_gen (2)/1 - modis L1A to GEO (6.1)/A2008271043000.GEO Dst: gridftp://amacgregor@xe-gt4.ivec.org/scratch/bi01/amacgregor/14170cb7-0941-45d6-89d0-bc56b4d1b9e8/input/A2008271043000.GEO	Edit	
				Src: scp://andrew@sooty.localdomain:22/export/home/tech/macgregor/yabi/image_gen (2)/1 - select file/MYD01.A2008271.0430.005.2008271161919.hdf Dst: gridftp://amacgregor@xe-gt4.ivec.org/scratch/bi01/amacgregor/14170cb7-0941-45d6-89d0-bc56b4d1b9e8/input/MYD01.A2008271.0430.005.2008271161919.hdf	Edit	

Проект Yabi: пример - genomic analysis

The screenshot displays the Yabi web interface, which is used for managing genomic analysis jobs. The interface is divided into several sections:

- Jobs List:** A table showing the status of various jobs. Two jobs are highlighted: 'large_seq_htp' and 'gprotein-htp', both completed on 2011-06-18.
- Job Details (large_seq_htp):** Shows a list of steps: 're-use', 'start', '1 - select file', '2 - repeatmasker', '3 - gscan', and '4 - gscan2gff'. Step 4 is currently selected.
- Job Details (gprotein-htp):** Shows a list of steps: 're-use', 'start', '1 - select file', '2 - getorf', and '3 - gpcrmm'. Step 3 is currently selected.
- File Outputs:** Two panels show the output of the selected jobs. The first panel shows the output of 'output.genscan.gff' (21.1 kB) with a table of contigs and genes. The second panel shows the output of 'STDOUT.txt' (0.40 MB) with a table of sequence identifiers and their coordinates.

File Outputs (output.genscan.gff):

contig1	assembly	contig	0	497001	.	+	.	.
contig1	GeneScan	gene	1922	12584	.	+	.	Gene
contig1	GeneScan	mRNA	1922	12584	.	+	.	mRNA

File Outputs (STDOUT.txt):

Sequence identifier	global	local	pred
TC15164_1	190.54	-99.04	No
TC15164_2	-97.22	-	No
TC15164_3	-150.32	-	No
TC15165_1	-131.14	-	No
TC15166_1	-186.33	-	No
TC15166_2	-220.80	-	No
TC15166_3	-158.70	-	No
TC15167_1	-223.49	-	No
TC15167_2	-120.63	-	No
TC15167_3	-256.96	-	No
TC15167_4	-172.27	-	No
TC15168_1	61.64	60.31	GPCR
TC15168_2	-206.28	-	No
TC15168_3	-278.22	-	No
TC15168_4	-333.85	-	No
TC15169_1	-155.19	-	No
TC15170_1	-160.98	-	No
TC15170_2	-108.06	-	No
TC15171_1	75.09	83.12	GPCR
TC15172_1	-239.50	-	No
TC15172_2	-100.73	-	No
TC15174_1	-163.33	-	No
TC15175_1	-139.00	-	No
TC15180_1	-140.16	-	No
TC15181_1	-145.28	-	No
TC15186_1	-131.66	-	No
TC15188_1	-86.74	-	No
TC15190_1	-78.84	-	No
TC15191_1	-164.02	-	No
TC15193_1	-117.75	-	No

(a)

(b)

Заключение (1/3)

- дальнейшее развитие этого направления может быть связано не только с количественным увеличением числа ВП и расширением научно-инженерных и производственных областей, в которых они применяются
- но и с совершенствованием технологии удаленного развертывания нового прикладного ПО на ресурсах
- это позволит преодолеть важную проблему связанную с использованием концепции SaaS в научно-технических областях: **ограниченный набор** программ и пакетов, предлагаемый провайдерами услуг SaaS

Заключение (2/3)

- В настоящее время вопросы предоставления услуг для поставщиков прикладного ПО в контексте научно-ориентированных веб-платформ проработаны **недостаточно** глубоко
 - в некоторых разработках (eScience Central) присутствуют сервисы для удаленной установки прикладного ПО, они пока **недостаточны** для того, чтобы обеспечить создание веб-площадки, способной выполнить весь спектр задач, который характерен для **свободного открытого рынка**

Заключение (3/3)

- технология создания таких веб-платформ рынка прикладного ПО может быть основана как на оригинальных решениях, так и на синтезе и адаптации решений, используемых при создании научных хабов (например, nanoHUB), облачных и грид-систем, а также on-line магазинов приложений
- будущие веб-платформы будут предоставлять **единую** точку входа как для **поставщиков** веб-услуг, так и для их **потребителей**