Development of Web platform for provision of application software as a Web service (SaaS) in open market approach

A. Kryukov, A. Demichev

SINP MSU

Supported by RFBR grant No.15-07-09309

Application software and distributed computing

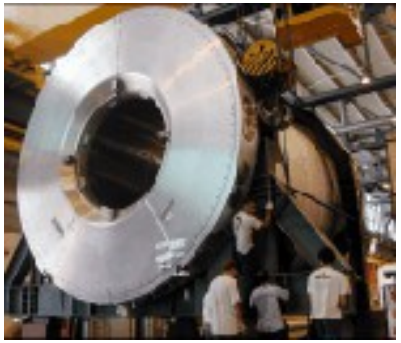
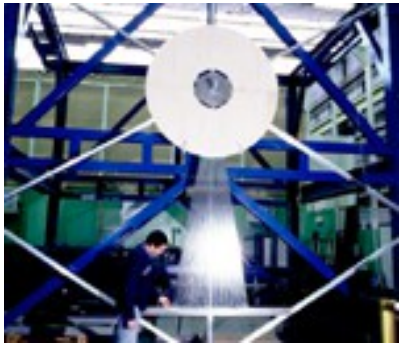
- Application software (especially simulation of complex systems) is a time expensive and resource consuming task
- simulation models require many simulation runs (sweep calculation)
  - ⇒ often local resources are insufficient
- a viable solution to speed up the process is to run the simulations on a distributed system
  - a good example: WLCG project

Success story of distributed computing (1/3)

- The Worldwide LHC Computing Grid (WLCG)
  - an international collaborative project
  - **grid**-based computer network infrastructure incorporating over 170 computing centers in 36 countries
  - It was designed by CERN to handle the prodigious volume of data produced by Large Hadron Collider (LHC) experiments
    - approximately 25 petabytes per year per experiment

CMS ***LHCb*** ***ATLAS*** ***ALICE***



Success story of distributed computing (2/3)

- Inspired by the WLCG success, grid computing became popular beyond the high-energy physics
  - bioinformatics, nanotechnology, geophysics, etc.
    - to share and combine the power of computers
    - and sophisticated, often unique, scientific instruments
- The most ambitious attempt to extend the grid technology to other scientific areas undertaken in the framework of a series of projects

  DataGrid → EGEE → EGI (2001 - 2015)

Grid → Web platforms

- With the growth of performance of individual resources (supercomputers, data storages, cloud systems, etc.) the grid conception began to lose a significant part of its appeal
  - a large-scale distributed computing grid infrastructure ⇒ high overheads
    ⇒ requires powerful unifying organizational structure, maintaining a cumbersome grid infrastructure (WLCG — CERN)
  - the focus in the development of a new generation of middleware shifts to building convenient and efficient means for accessing individual computing resources
    - **web platforms**

Remote access to HPC resources (1/2)

- direct access using the ssh protocol remains quite popular
  - provides the user with the flexibility in using the capabilities of the resources (+)
  - users have to learn many subtleties of the environment in which the applications are launched (-)
- in many cases researchers need to run a large number of **similar** computing tasks
  - The model Software as a Service, **SaaS**
  - **Web platform** (WP) for remote access to computing resources: a set of specialized web services + web application interfaces

Remote access to HPC resources (2/2)

- A pre-arranged task (*a launch of an application package, an access to a data storage, etc*) = "a tool"
- A series of tasks depending on each other in a hierarchical way are called **workflows**
- Having in their disposal a set of pre-configured "tools", users only needs to formulate the essence of a specific request in a natural language
- Web platforms can offer specific services for deploment by authorized user of new software packages (tools)

Types of Web Platforms (WP)

1. WPs for job submission: remote submission, monitoring, and obtaining the job results;
2. WPs for job submission and software installation: item 1 + remote installation and configuring of application packages (tools);
3. Web hubs: items 1, 2 + providing the features of professional social networks (sharing experiences, rating of tools, etc);
4. WPs of application software market: items 1, 2, 3 + services for interaction between the providers and consumers of «tools» on market principles.

Basic Functional Requirements to WPs (1/2)

- management of user credentials granting the right to use the available resources
  - Authentication, authorization, delegation
- remote administration of the web platform via a web browser
  - Accounts, granting the rights to use of sets of tools; accounts of tool providers

Basic Functional Requirements to WPs (2/2)

- job execution management
  - creation of composite jobs (workflow); representation of the job description in the format of resource; job submission; monitoring; recive results and visualization of it
- data files transfer management
  - Data transfere from/to local computer or storage service
- tools (services) management
  - Register of new tools, building templates, handling of access grants to resources

Administration of the WPs (1/2)

- The most flexible approach to distributed resources managing is based on the concept of virtual organization (VO)
  - a dynamic set of individuals or institutions defined around a set of resource-sharing rules and conditions
- VO may be responsible for
  - the development/installation of the tools,
  - day-by-day interaction with the resource owners/VO members: within the VO both consumers and owners of the resources cooperate.
- administration of resources gets easier because their owners only deal with VO managers, not with each user individually.

Administration of the WPs (2/2)

- Users can just:
  - compose the tools available for their VO into a workflow
  - provide files with input data and execution parameters for the tool of the workflow.
- VO administrators' objective should include the creation and completion of a repository of VO's tools + installation on the widest possible range of resources.
- Repository of tools can be filled up both by open source software suitable for use in the area of interest of the VO and by the software developed by the members of the VO.

Examples of WPs Implementations (1/4)

- web platform of educational-methodical software package «Multiscale modeling in nanotechnology» (Photochemistry Center of the Russian Academy of Sciences (RAS); http://www.nanomodel.ru);
- «Personal virtual computer» system (South Ural State University; http://supercomputer.susu.ac.ru/pvc);
- UniHUB, the technological platform of the National «University Cluster» program [13] (Institute for System Programming of RAS; https://unihub.ru);
- computing cloud platform of the Ural Branch of the RAS (Instit. of Math. and Mechanics, UrB RAS);

Examples of WPs Implementations (2/4)

- web portal of the supercomputer management system (Glushkov Institute of Cybernetics of NAS of Ukraine; http://melkon.com.ua/ru/cms);
- Everest web platform (Institute for Information Transmission Problems of the RAS; http://everest.distcomp.org)
- multifunctional instrumental and technological platform for cloud computing support CLAVIRE (SPbg State University of Information Technologies, Mechanics and Optics; http://clavire.ru);
- nanoHUB, web hub in nanotechnology (consortium of the US universities; http://www.nanohub.org);

Examples of WPs Implementations (3/4)

- eQUEUE, web platform for the remote job submission (Advanced Clustering Technologies, Inc.; http://www.advancedclustering.com/products/software/equeue);
- Nucleonica, scientific web portal (Institute for Transuranium Elements; http://www.nucleonica.net);
- WebMO web platform (Hope College, Holland, USA; http://www.webmo.net);
- веб-платформа e-Science Central (Newcastle University, UK; http://www.esciencecentral.co.uk).
- Yabi web platform (Centre for Comparative Genomics, Murdoch, Australia; https://ccg.murdoch.edu.au/yabi)

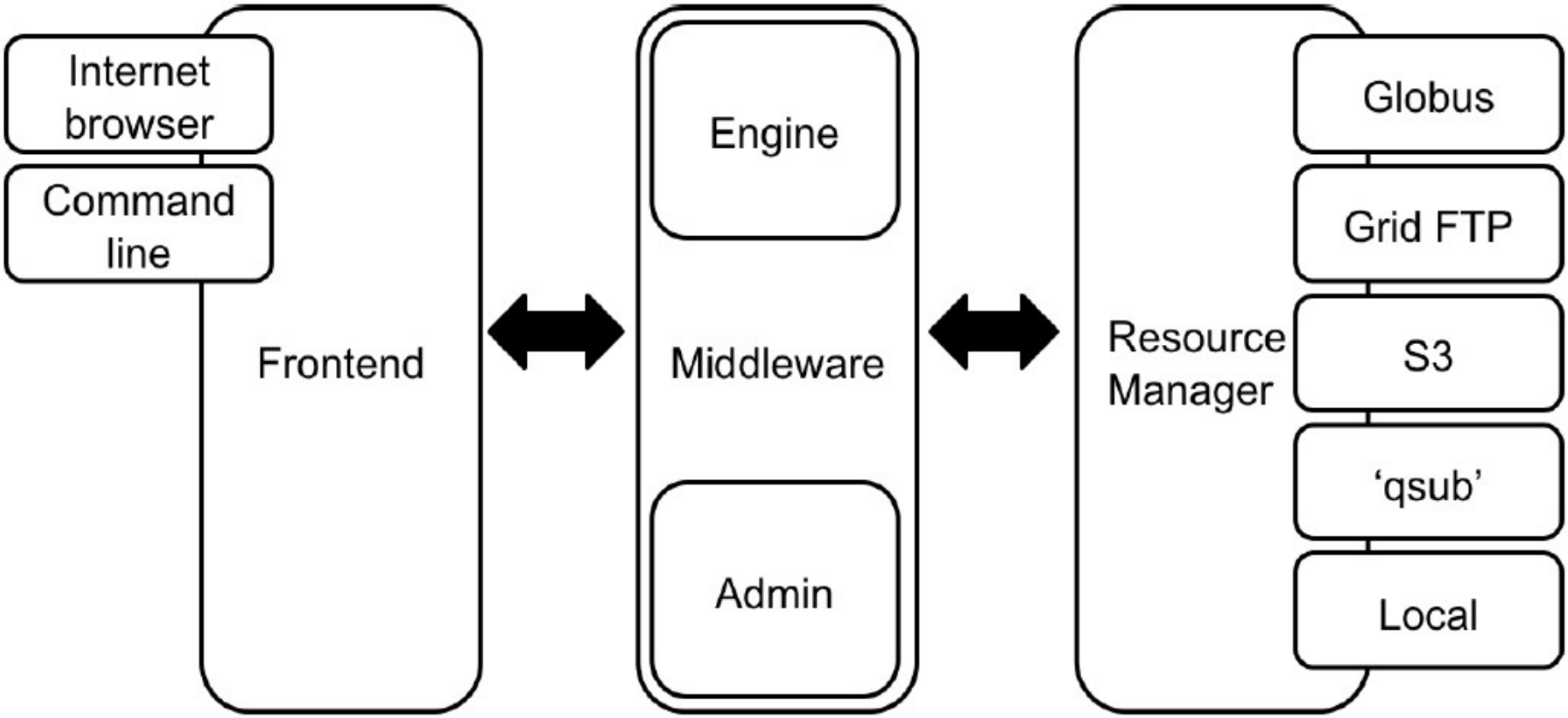
Examples of WPs Implementations (4/4)

- almost all of these developments ensure the provision of end-users in advance preset simulation tools
- they are still insufficient to ensure the creation of a web platform capable of performing the whole range of tasks characteristic for a free open market

Typical example: Yabi Project

- provides the execution of workflows consisting of successive tasks
  - accesses to databases,
  - use of the results as input for computing tasks etc.
- has well developed and convenient administrative interface for
  - configuring the "tools"
  - controlling the user access to these tools
- GNU GPL v3 (noncommercial use)

Yabi Project: architecture



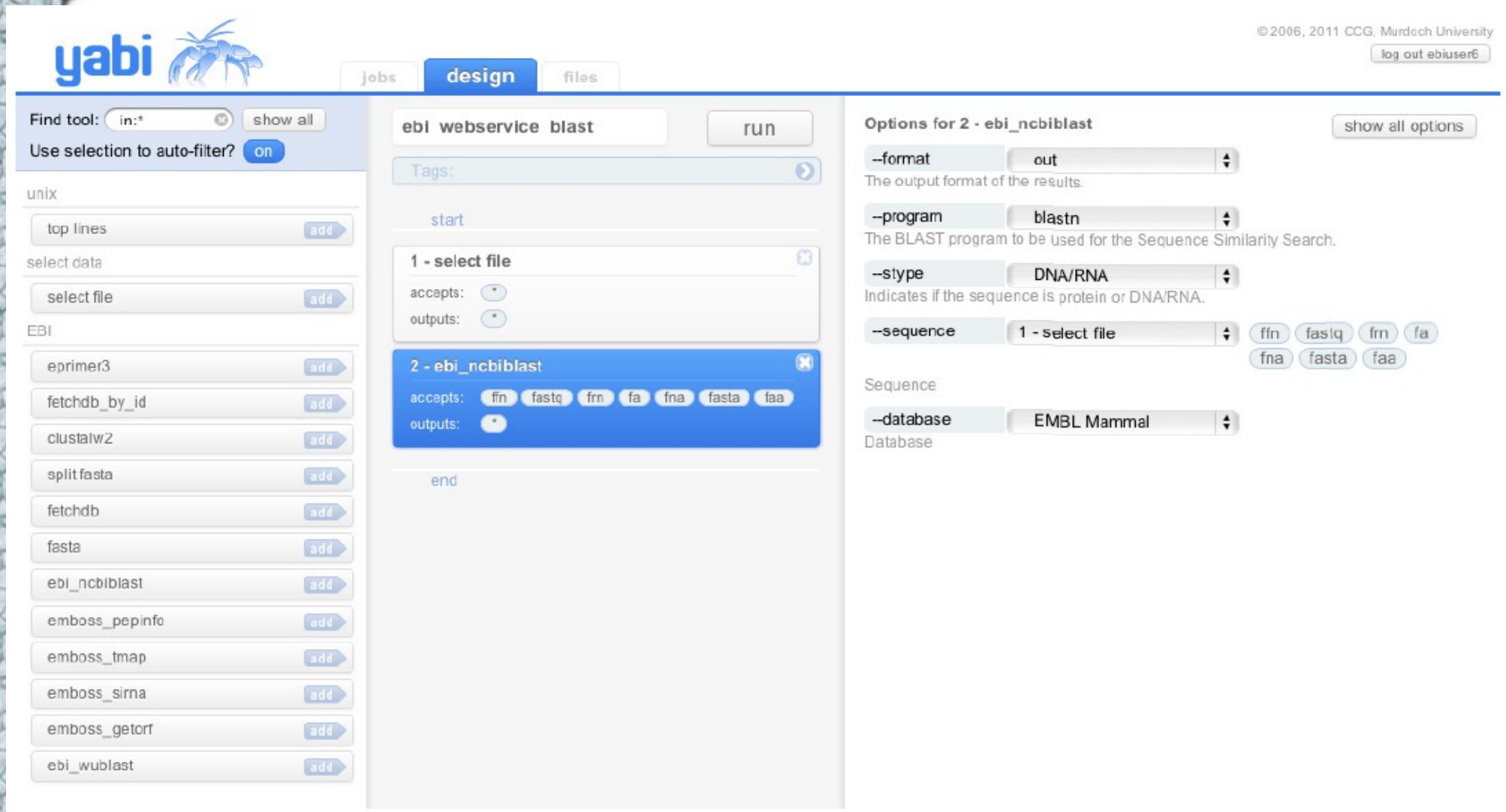
Yabi Project: Resources

- Yabi resource manager provides the two types of services:
  - compute services
  - data services,
- both types have plug-in architecture
  - allows them to communicate with various compute resources and file storage systems
- for the interaction with the external services
  - credentials encrypted on behalf of the users (by using the login/password pair)

Yabi Project: Web Interface

- Web GUI (browser) + CLI
- Web GUI — three tabs:
  - the design view which allows construction of new workflows,
  - the jobs view which shows previously submitted workflows along with their parameters and input files
  - the files view that represents the data storages accessible by the user

Yabi Project: Web GUI



Yabi Project: administration (1/3)

- administration module provides web interface that allows an authorized user (Yabi administrator) to manage all aspects of the web platform operation including
  - the creation of new tools from application software preinstalled on the resources
  - controlling user access to various tools

Yabi Project: administration (2/3)

Creation of tools

Yabi Project: administration (3/3)

Information on completed jobs

Yabi Administration

Welcome, **andrew** . Change password / Log out

Home › Backend image_gen (2)

Workflow Summary

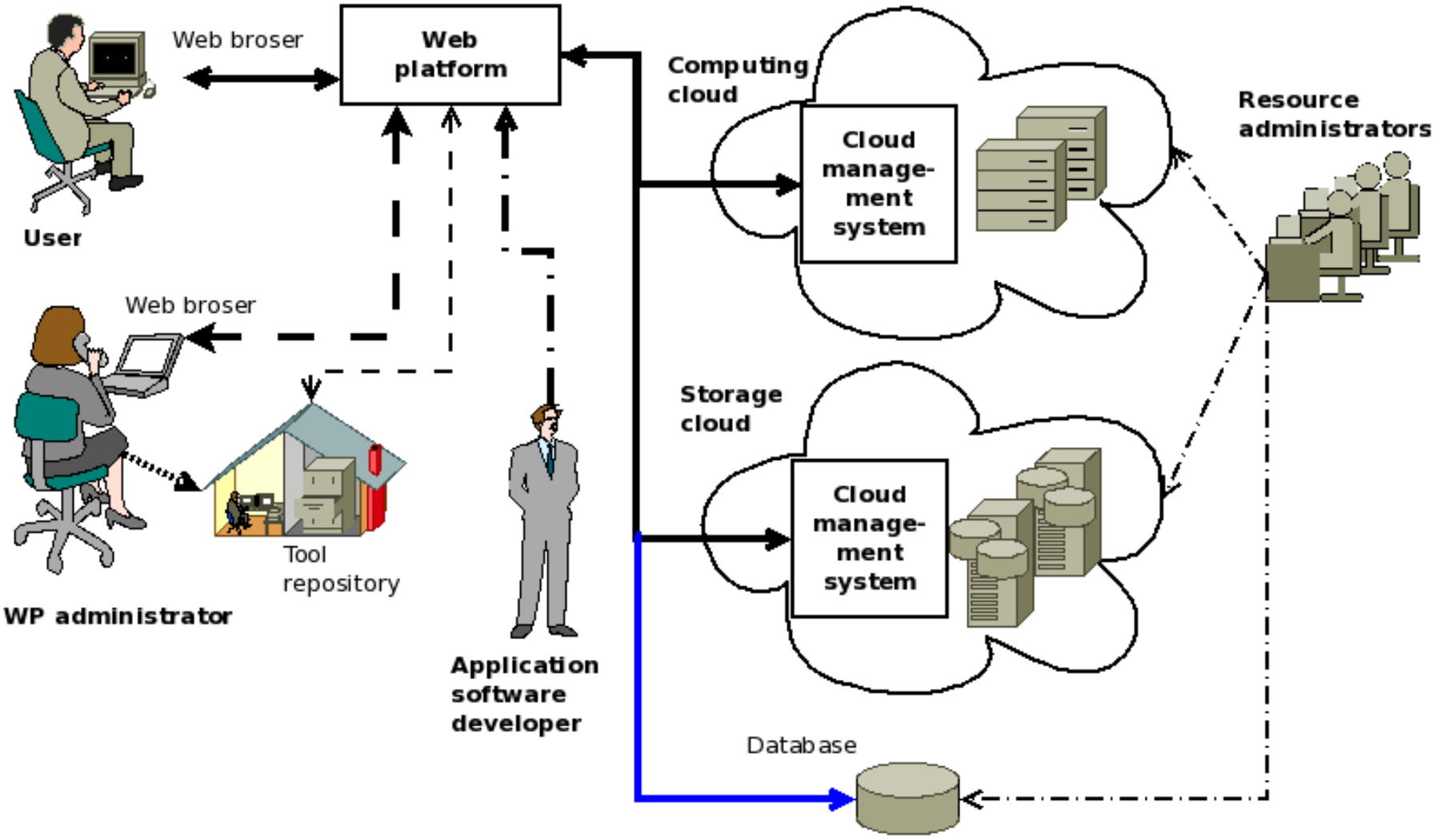
Name	User	Start	End	Log File Path	Status	
image_gen (2)	andrew	2011-06-15 15:54:56	2011-06-15 15:59:23	/	complete	Edit

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)/2 - 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

Peculiarities of WPs with services of remote deployment of application software

- various types of resources provided that application software is deployed by the resources administrators
  - WP may contain tools to facilitate interaction between software developers and resources administrators
    - these tools can include means for converting application software into SaaS
  - NanoHUB/HUBZero
- remote software installation by users themselves
  - resources is practically limited to the cloud systems (virtual machines)
  - model Platform as a service (PaaS).

AppStore Paradigm (1/3)



AppStore Paradigm (2/3)

- Software developers interact with the cloud management system via the WP
  - PaaS model
  - software deployment + convertation into the web tool
- WP contains modules needed for market-like interrelations between providers and users of tools
  - monitoring
  - logging and billing

AppStore Paradigm (3/3)

- A prototype of such a platform is currently being developed at SINP MSU
  - Supported by RFBR grant No.15-07-09309
- From a technological point of view, the WP prototype be implemented as a set of web services
  - with extensive use of architectural style REST
  - data exchange between the services - in the JSON format

Summary(1/3)

- Further line of development of the web toolkit may be related not only with the quantitative increase in the number of web-based platforms for remote access and the expansion of scientific, engineering, and manufacturing areas in which they are used,
- but also with the improvement of the technology of remote deployment of new application software on resources interacting with the web platforms
- This approach will help to overcome an important problem associated with the use of the SaaS model in scientific areas, namely, limited set of application packages offered by SaaS providers

Summary (2/3)

- Currently, the provision of services for providers of application software in the context of scientific-oriented web platforms is not developed enough.
  - Although some implementations (for example, e-Science Central) have services for remote application software deployment, they are still insufficient to ensure the creation of a web platform capable of performing the whole range of tasks characteristic for a free open market.

Summary (3/3)

- The technology of creating such web platforms market of application software can be based both on the original solutions and on the synthesis and adaptation of the solutions used in research hubs (e.g., nanoHUB; nanohub.org), cloud and grid systems, supercomputers, as well as in on-line app stores.
- However, it seems that unlike the on-line app stores, the platform should not only provide information services for searching the tools needed by users, but also provide the feasibility of direct using of the necessary tools.
- Thus, the future web platforms will provide a single entry point both for web service providers and for their customers.

Conclusions

In the SINP MSU was started a new project for research and development of Web platform for scientific software application called eSciMarket which has next primary targets:

- Provides the users of software (SaaS) and provider of these software by the common Web platforms where they will be interact with each others.
- The basic principles of such interaction should be principles of free open market.
- The platform should be a tools itself to development of new SaaS, for example in the form “Workflow as a Service”

Thank you for attention! Questions?