NEW METHODS OF SOLVING PROBLEMS OF COMPUTATIONAL CHEMISTRY IN GRID ENVIRONMENTS

Pivushkov Alexander, Ph.D.
Volokhov V.M., Varlamov D.A., Volokhov A.V.

Institute of Problem of Chemical Physics, Russia, Chernogolovka

Dubna-2012, Distributed Computing and Grid-technologies in Science and Education

Participation of IPCP in grid projects (in Russian)

1. Программа № 13 фундаментальных исследований Президиума РАН на 2009-2012 годы «**Проблемы создания национальной научной распределенной информационно-вычислительной среды на основе развития GRID технологий и современных телекоммуникационных сетей**», проект «**Исследование методов виртуализации вычислительных сред и приложений в области вычислительной химии. Динамическое формирование параллельных программных сред на распределенных ресурсах**»;
2. Государственный контракт № 07.514.11.4019 от 23 сентября 2011 г. с ИПХФ РАН в рамках ФЦП «**Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2013 годы**» по теме «**Разработка технологии проведения высокопроизводительных расчетов в области вычислительной химии в различных распределенных средах с применением методов виртуализации приложений и ресурсов**»
3. Грант РФФИ № 11-07-00686-а «**Разработка методов динамического формирования параллельных вычислительных сред на различных типах грид ресурсов (на примере приложений квантовой химии)**
4. Федеральная целевая программа (2011 г.) по теме «**Создание комплекса проблемно-ориентированных сервисов доступа к прикладным программным ресурсам грид-инфраструктуры Национальной нанотехнологической сети (ГридННС) на основе веб-сервисных технологий и интеллектуальных интерфейсов пользователя**»

Used distributed grid polygons

- National nanotechnology Network (GridNNN, VO «*NanoChem*», «*GAMESS*», «*Gaussian*»), middleware – modified Globus Toolkit 4, up to 8000 CPU
- Pilot zone of Russian Grid System for high-efficience calculations (under management of RF Ministry of Communications) – middleware Globus Toolkit 5

Also earlier developed technologies were tested in grid environments under different middleware on the following polygons:

- Unicore 6.2 (SKIF-Polygon);
- gLite (polygon EGEE-RDIG, subsequently – EGI-RU-NGI)

The main problems of computing in grid environments

1. Problems compatibility of the applied and system software of grid applications with already installed on operating grid resources (the conflicts of programs, system libraries, presence of middleware(s) of various distributed environments);
2. Problems of safety of resources and information isolation;
3. Problems efficiency of application of various parallel methods of calculations

The technologies offered to development

1. Technology of creation and support of virtual machines (VM) for placement of managing and network services of resource sites of grid polygons;
2. "Virtualization" of grid applications and use of dynamically formed computing images (operation with “virtual containers”);
3. Creation and management of "pools" of grid jobs for work with big tasks on uniform "grids" of data or parameters (with use of multiple parameter tasks or applied packages of the GAMESS type). They can be presented in the form of association of a large number in parallel carried-out tasks independent from each other.
4. Use of graphic high-efficiency video adapters (GPU) for substantial increase of efficiency of parallel calculations in areas of quantum chemistry and molecular dynamics on grid resource sites

Physical units (computers) - 1-2, with KVM hypervisor

The Virtual Machines isolated from each other (up to 4-5 per 1 physical unit)

Network and Grid Services

Resource site

Grid polygon 1

Resource sites

Grid polygons 2, 3, …

Grid-oriented cluster

OS Scientific Linux 5.4
task manager
PBS/Torque
(at the same time
carried out 3 copies)

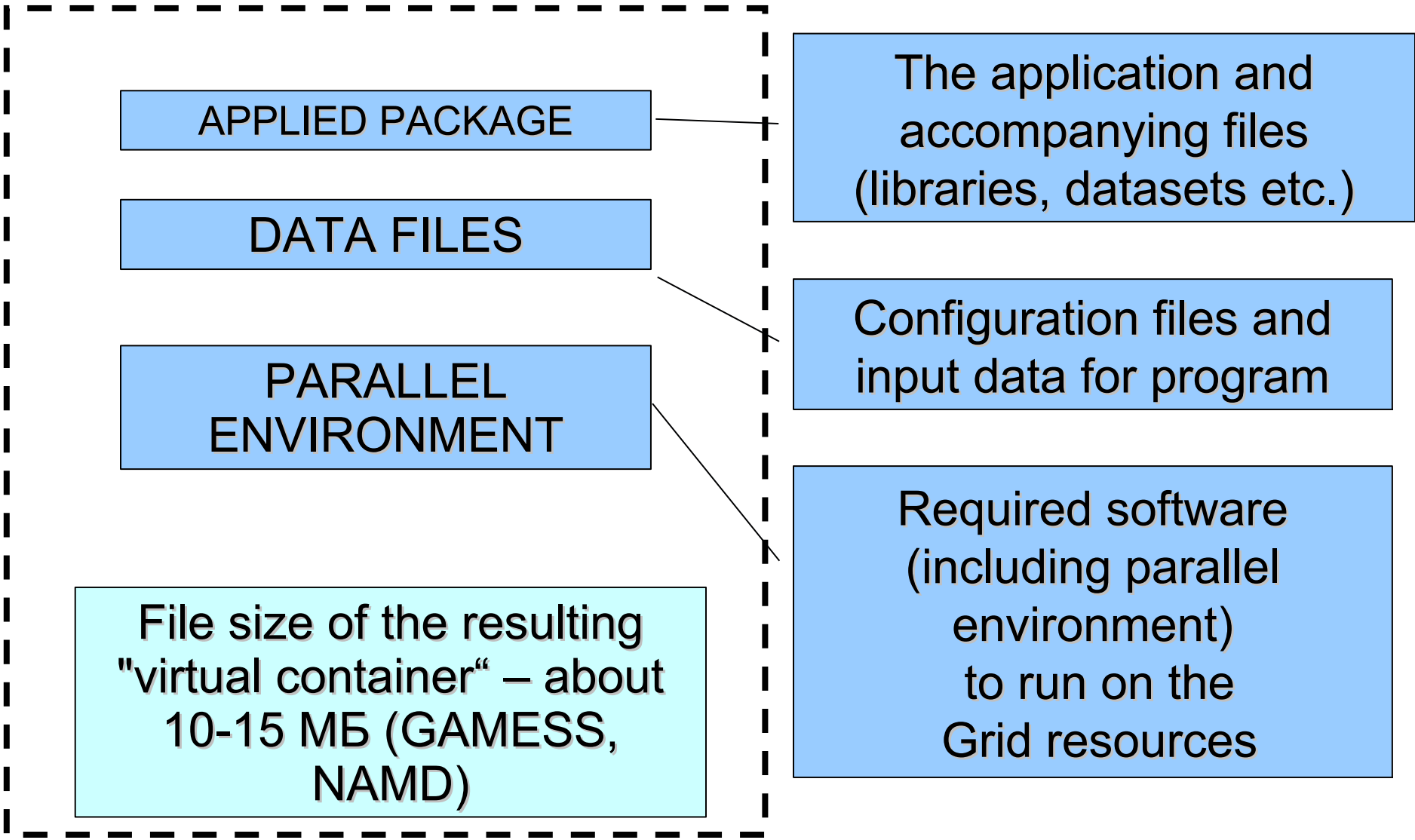
Applied software

- **GAMESS-US**
- **Gaussian-03**
- **NAMD-2**
- **PWscf**
- **CPMD**
- **Dalton-2011**

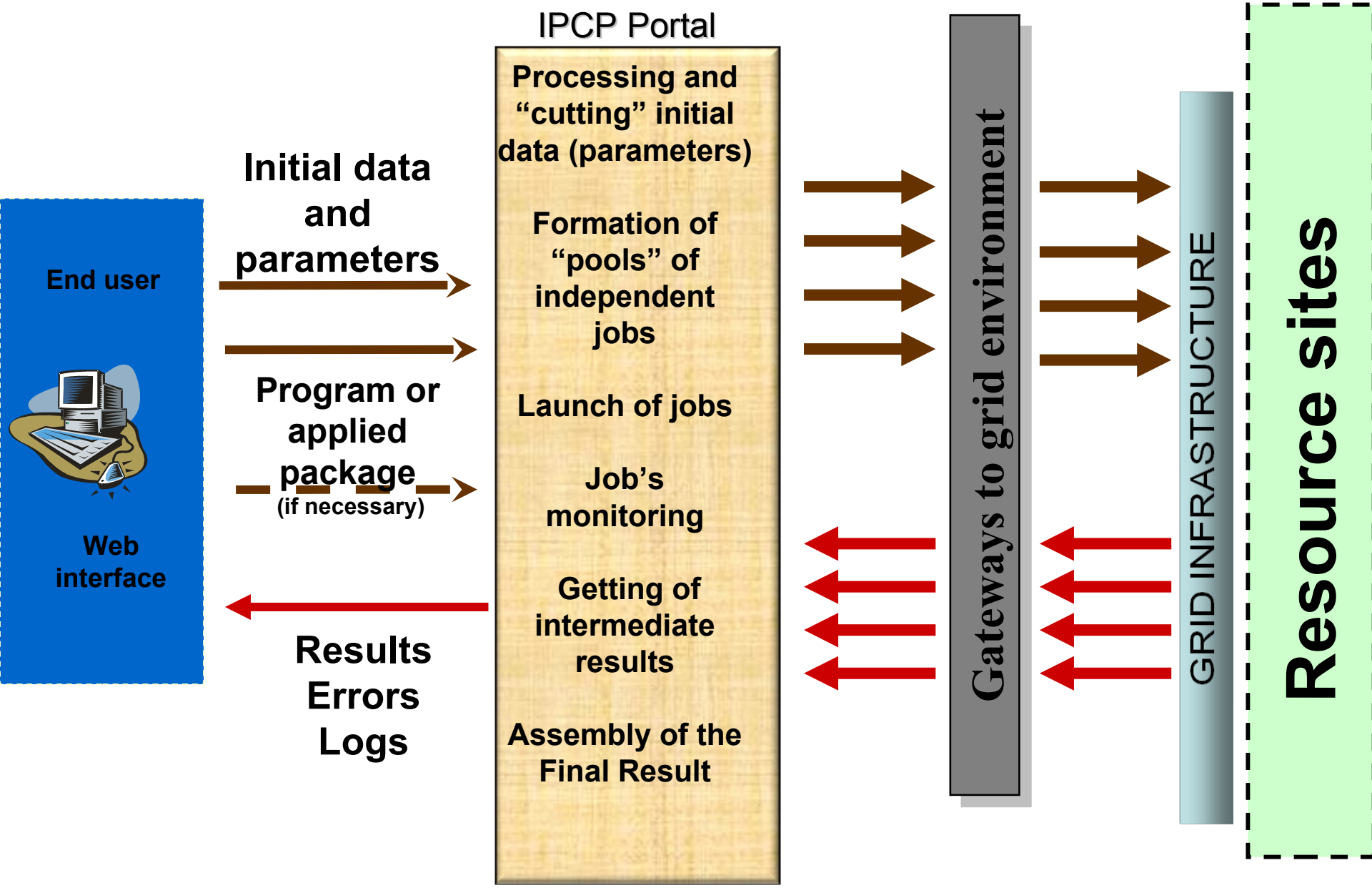
The technology of "virtualization" of grid applications

- For the selected application dynamic "container" is created. It contains the applied package, program files and libraries necessary for its work, control scripts for installation and configuration on a grid-resource, initial data, configuration files.
- Packed (TAR gzipped) "container" through grid infrastructure is delivered to a remote grid resource (cluster) and is started as a usual grid job
- The resource is temporarily adjusted under needs of an applied package: own parallel environment (Mpich-2, Charm++) can be started (if necessary), the package is installed on available worknodes and (after necessary tests) is started on calculation;
- After completion of calculations results through grid infrastructure are transferred back to the user, then the resource "is cleared" from temporarily created computing environment.

Technology of start of parallel tasks in the distributed environments



Creation and management of "pools" of grid jobs for work with big tasks on uniform "grids" of data or parameters



The use of graphics adapters (GPU) in applications of quantum chemistry

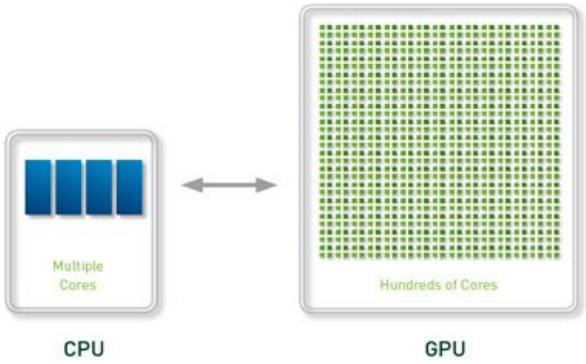
GPU computing are sharing the CPU and GPU resources in a heterogeneous computing model. Standard part of the application running on the CPU, and more demanding of computing parallel computing – use multiconveyors on the GPU.

From a user perspective the application is much faster because it uses parallelization and performance GPU.

Example: NAMD applied package accelerated at one node at almost 12 times due to CUDA™, speedup on a cluster with GPU-accelerated nodes - up to 330 times.

Already use: GROMACS, LAMMPS, NAMD, VASP, TeraChem
Tested: GAMESS, AbInit, Gaussian, NWChem

4-8 CPU cores



128-2056 cores
up to 0,5 Тф

Acceleration assessment for a number of programs of quantum chemistry and MD

(c) www.nvidia.ru

Applied package	Supported Features:	Acceleration, times
Abalone	Molecular simulations	**4-29**
AMBER	Method of Evald summary, PMEMD: Explicit and Implicit Solvent	**8**
CHARMM	Molecular simulations	**4-29**
GAMESS-US	Libqc with Rys Quadrature Algorithm, Integral evaluation Closed shell Fock matrix construction.	**2-12(under development)**
GROMACS	Method of Evals summary: Implicit (5x), Explicit(2x) Solvent	**2-5**
NWChem	Triples part of Reg-CCSD(T), CCSD & EOMCCSD	**3-8**
LAMMPS	Methods of Lennard-Jones, Gay-Berne	**6**
NAMD	Non-Bond Force calculation	**2-7**
Schrodinger Core Hopping	• Prime Minimization • DESMOND • Glide Grid generation • PyMOL visualization system • MD trajectory analysis	**up to 5000**
TeraChem	Full GPU-based solution	**44-650**
VASP	Davidson iteration scheme	**3-6**
VMD	High-quality rendering, Large structures (100 million atoms)	**64-128**

Conclusions

At IPCP RAS a number of the computing grid-technologies directed on increase in efficiency of carrying out calculations in the field of computing chemistry in grid-environments is developed. Introduction of similar technologies:

- Essentially accelerates carrying out quantum and chemical calculations and their scalability;
- increases efficiency of use of grid-resources;
- simplifies administration of resources and services;
- considerably increases reliability of operating sites;
- lowers expenses on grid-infrastructure support.

Developed technologies are protected as the main components of object of intellectual property «A specialized problem-oriented package of computing services of various level for the solution of problems of computing and quantum chemistry in the distributed GRID networks», certificate № **2011611232**