

Concept of a cloud service for data preparation and computational control on custom HPC systems in application to molecular dynamics

S. Markizov, D. Puzyrkov, V. Podryga, S. Polyakov

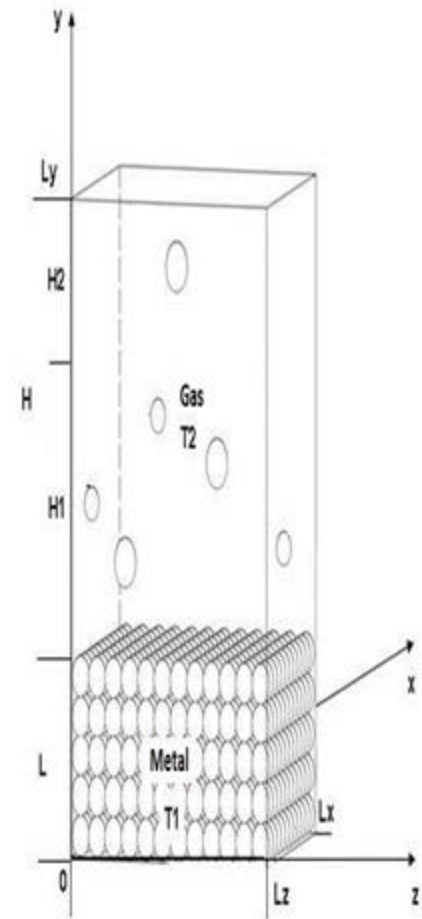
arh11msn@yandex.ru, dpuzyrkov@gmail.com,
polyakov@imamod.ru, pvictoria@list.ru

KIAM RAS, Moscow

MMCP2017, Dubna

Introduction

- Molecular Dynamics method operates with particles whose dynamics are described using the laws of classical mechanics.
- Large-scale MD simulations requires a lot of computational resources.
- MD initial data preparation is a complex procedure.
- If the scientist has access to multiple HPC systems, he usually needs to run simulations manually on every system.



Problem statement

Obviously, it is quite useful to

- Unite accessible HPC systems using a single management platform.
- Combine this platform with a CAD system for construction of the computational domain.
- Provide user a tool for re-using previous simulations results in a simple, cataloged way.
- Make the user interactions with this system independent from simulator program he choose.

Problem statement

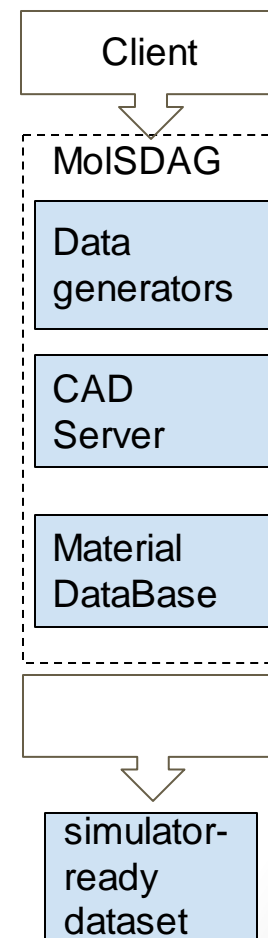
In fact, all of this points goes expands in 4 tools, that can be developed as a web platform:

- Computational domain geometry editor;
- Database of molecular calculations, geometries and metadata;
- Applications for converting the internal-formated datasets into a format known for the computational application;
- Deployment, monitoring and management system.

KIAM MoISDAG web service

The purpose of this service is to create a cloud-based multi-user environment for

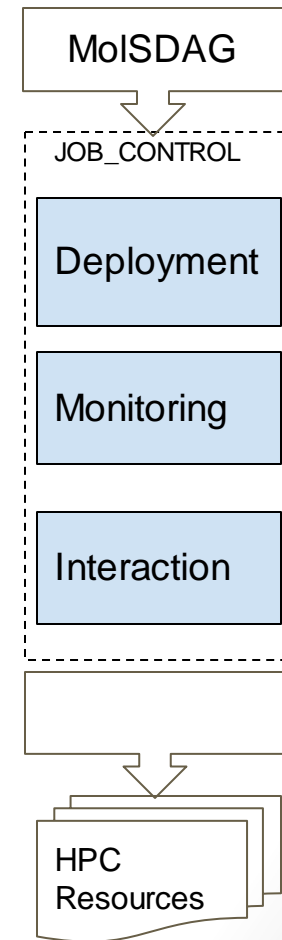
- designing of microstructures and metadata for subsequent MD calculations;
- accounting and storage of already studied microstructures and their properties;
- parallel generation of new microstructures and raw data for various MD simulators;



KIAM MoISDAG web service

Combining this with the KIAM JOB_CONTROL task management system will also provide tools for:

- deployment and monitoring of user calculations
- relocation of the running calculations from one resource to another if quota has expired.
- communication with the simulation

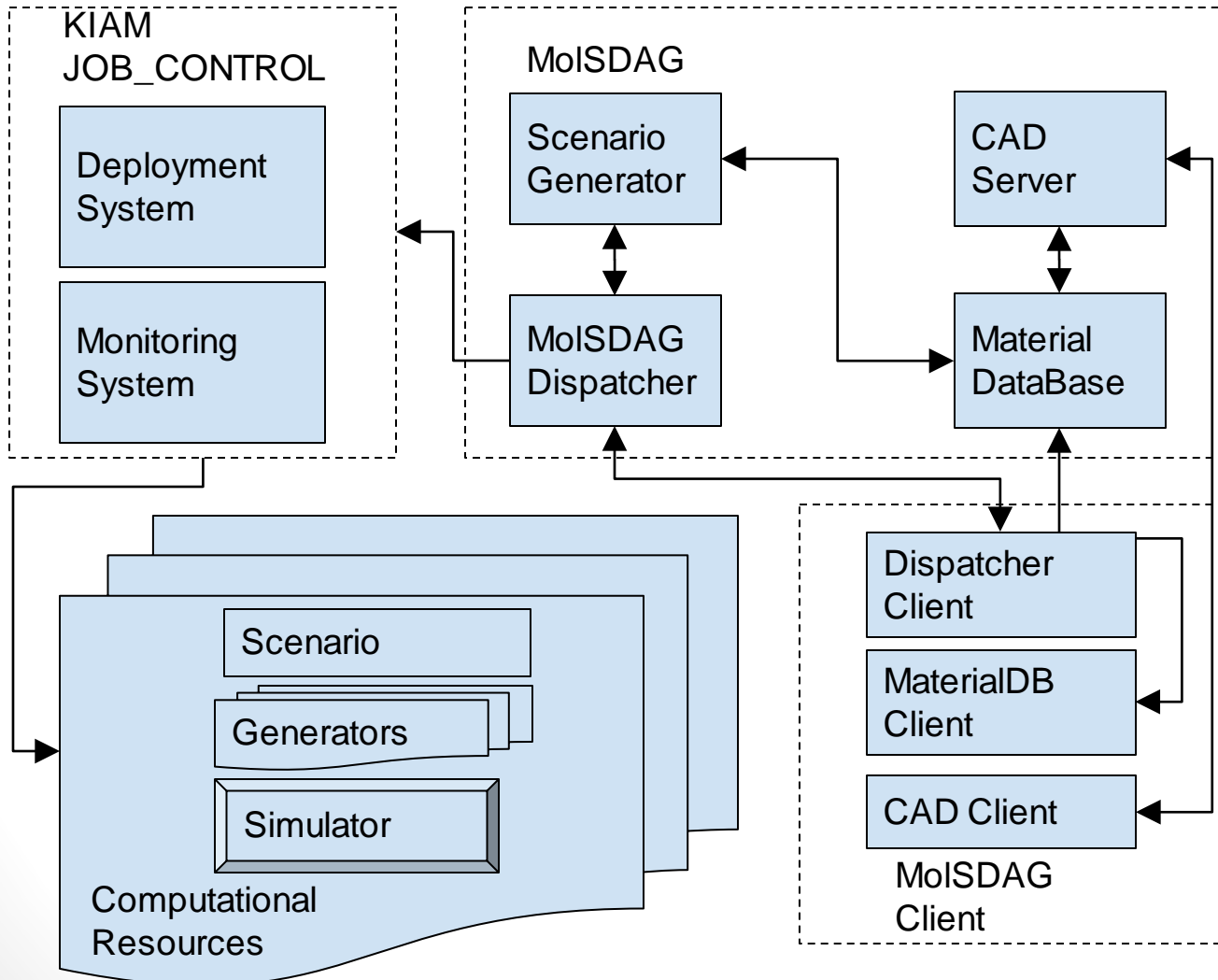


KIAM MoISDAG structure

Basic components of this prototype are

- Editor of atomistic structure of computational domain:
 - client part (GUI)
 - server part (DBMC Interactions and helpers)
- DBMC (DataBase for MolecularCalculations)
 - Results of calculations
 - Metadata for structure generator
- Target system scenario generator
 - Building applications on target system
 - Initial data construction on target system
- Dispatcher
 - JOB_CONTROL description files
 - JOB_CONTROL integration

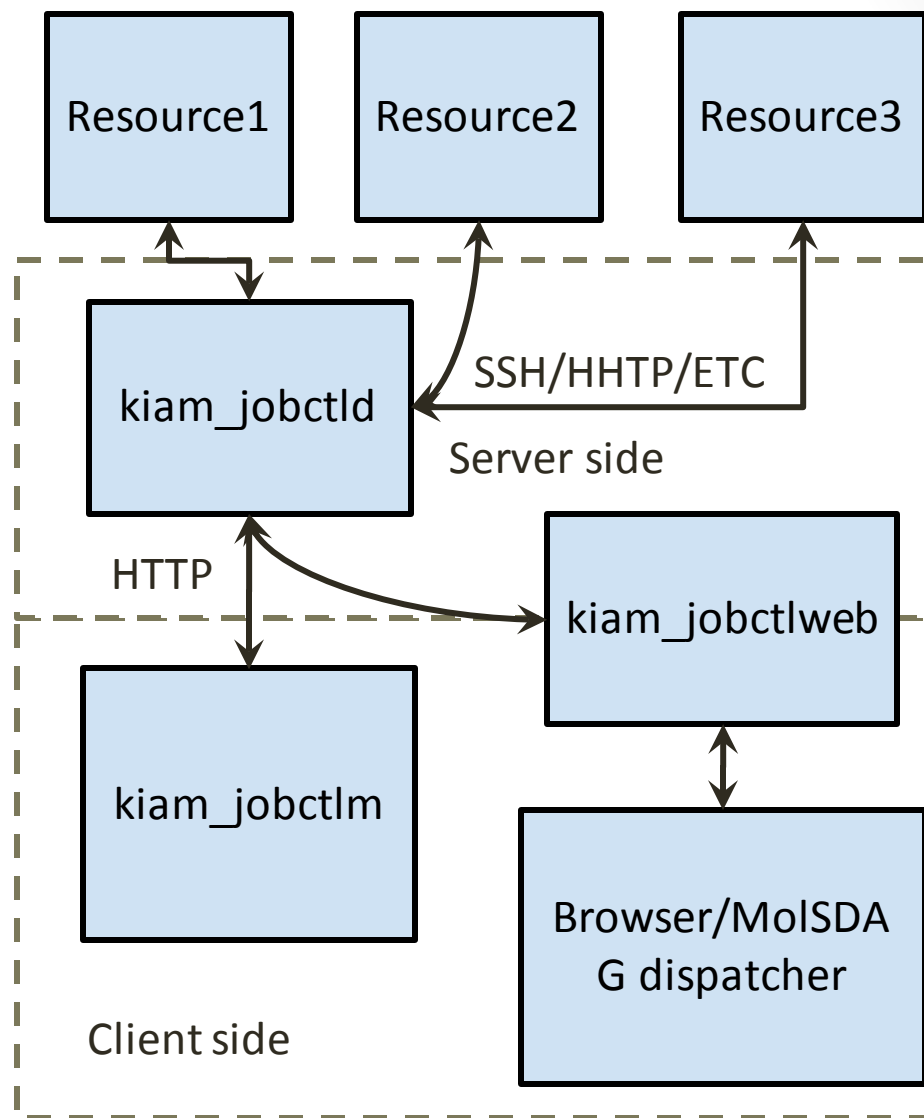
KIAM MoISDAG structure



KIAM JOB_CONTROL

Structure

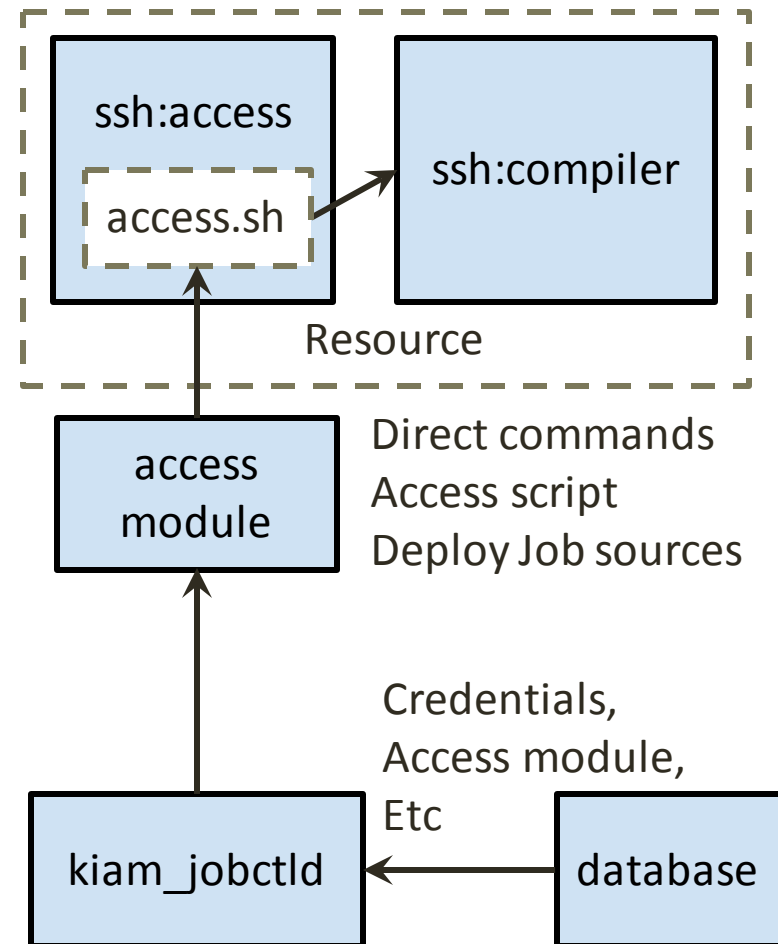
- **kiam_jobctld** - a server part, provides HTTP API for interaction, and does all of the management tasks.
- **kiam_jobctlm** – user command line interface, provides access to the kiam_jobctld.
- **kiam_jobctlweb** – web UI application.



KIAM JOB_CONTROL daemon

Basic components of this prototype are

- Server application, provides
 - HTTP api
 - Job deployment system
 - Job monitoring system
 - Job relocation system
- Different access modules for different access types to HPC systems
 - Special access.sh scripts for inter-resource commands.



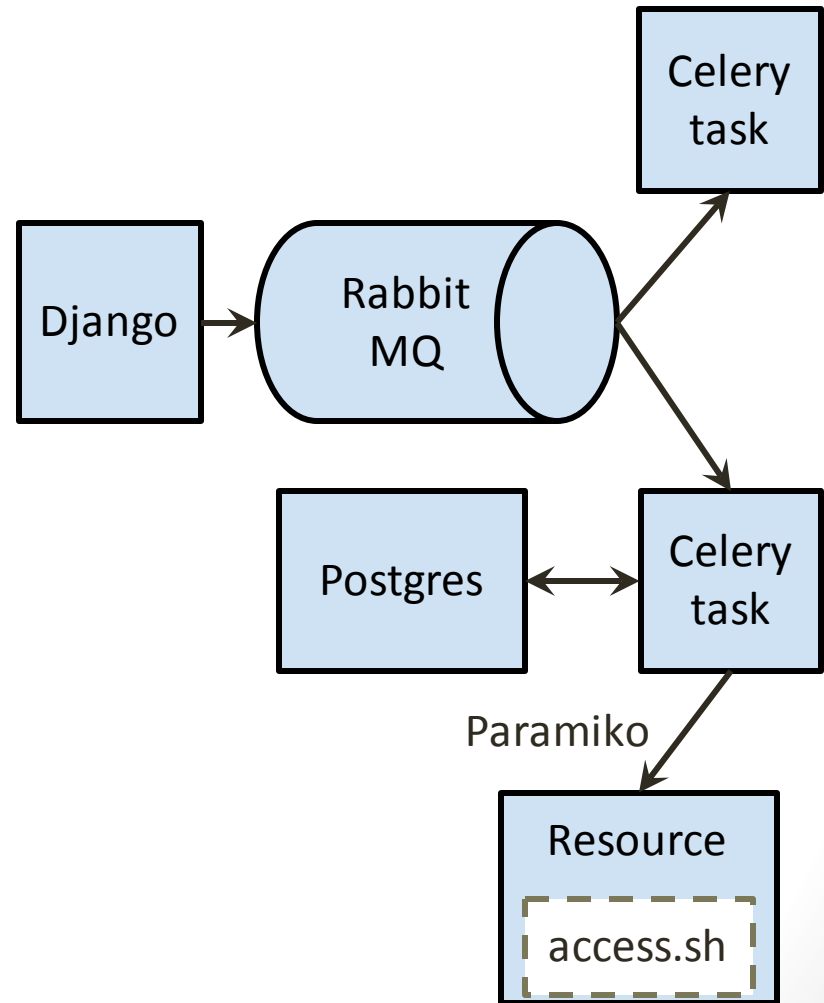
Technologies

Python:

- - Django for web pages and api
- - Celery+RabbitMQ for asynchronous task queue
- - Paramiko for ssh access
- - Postgres as a database

Sh scripts :

- for inter-resource communication



Prototype application

- Large-scale MD simulation:
 - - 5 variants of simulation
 - - 3 different resources:
 - MVS10-P (MSC RAS)
 - K1 (NICEVT)
 - IMM6 (KIAM RAS)

Tasks:

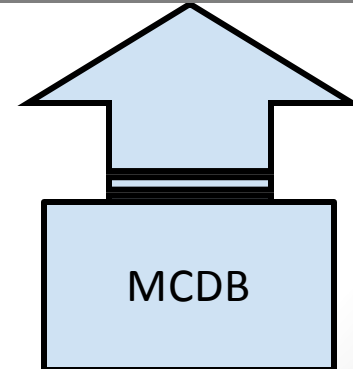
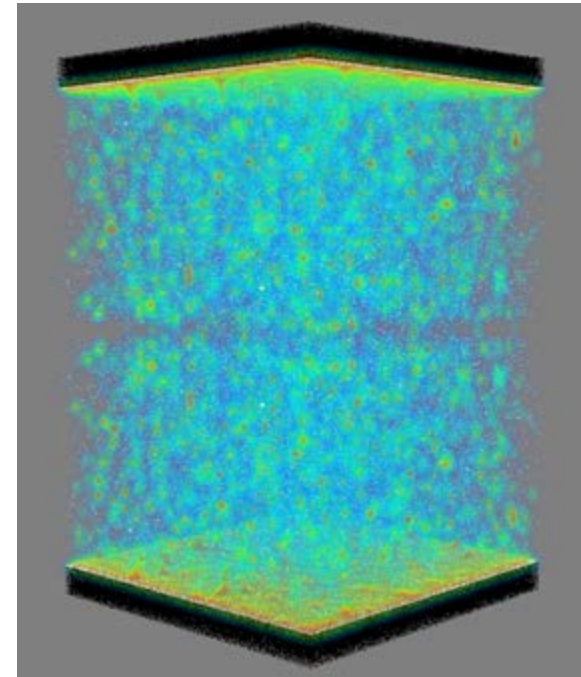
- computational domain design
- reuse of previous computed datasets
- automatic data relocation

Resources specs

- MVS10-P (MSC RAS)
 - 49 TFlops
 - 8-24 hours per week
 - 0.5 Tb disk quota per calculation
- K1 (NICEVT)
 - 4.2 TFlops
 - 48-94 hours per week
 - 10 Tb disk quota
- IMM6 (KIAM RAS)
 - 2 TFlops
 - 168 hours per week
 - 60 TB storage

Re-use of data and design of computational domain

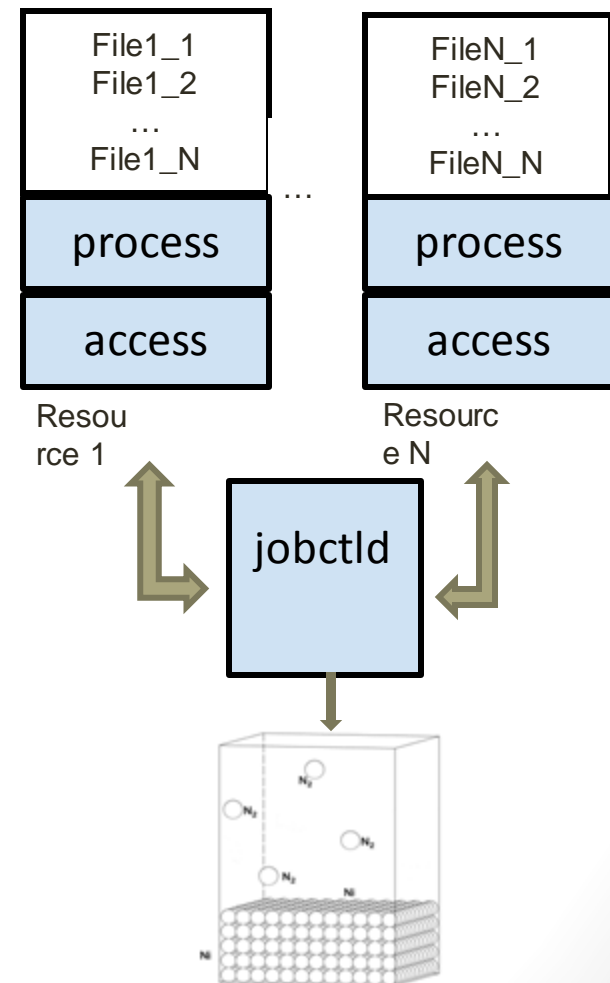
- Desired MD system: two nickel plates and nitrogen between them.
 - geometry design based on geometric primitives and rules of particles inserting in them.
- Re-use of data:
 - previously simulated to the thermodynamical equilibrium components are used.



Automatic data relocation

KIAM JOB_CONTROL makes it possible to

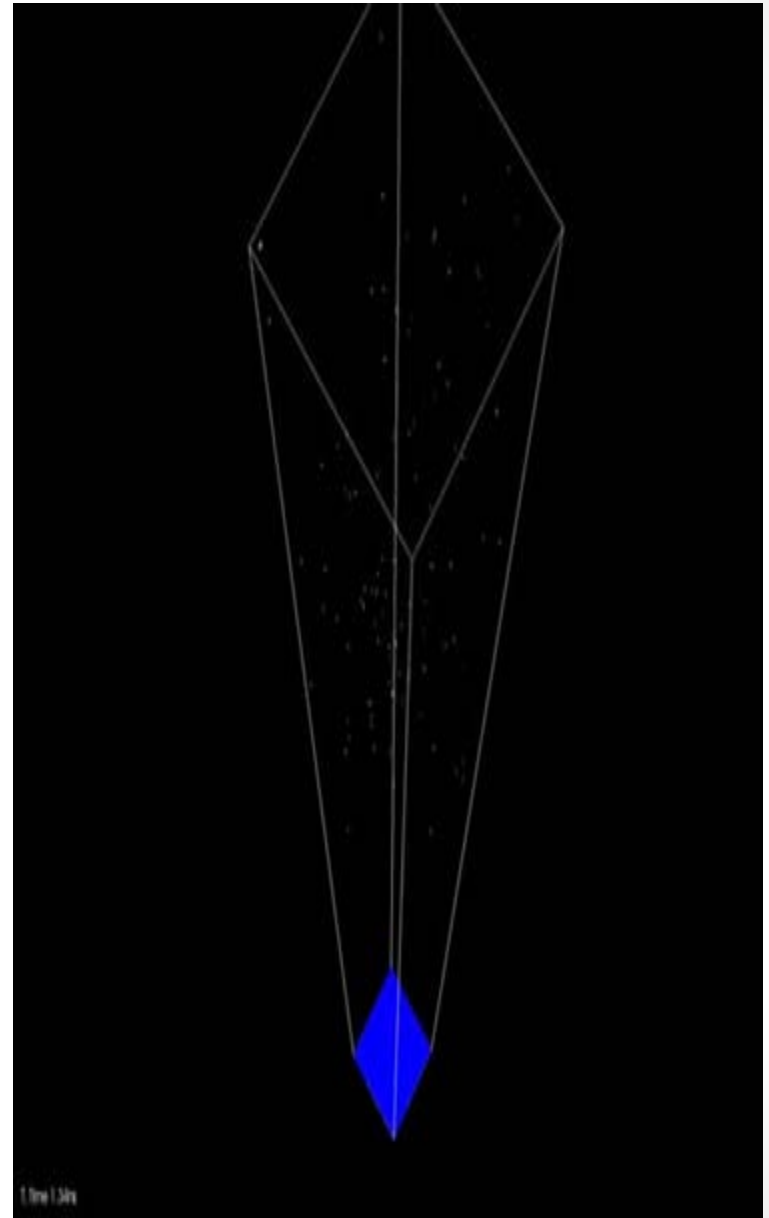
- Manage jobs execution process.
- Run jobs on available resources according simple timetable or a quota information.
- Relocate calculation data from target resource to IMM6 storage in automatic mode.
- Run post-processing jobs in parallel with calculations.



Results

- Implemented simple prototype of described cloud service
- Used it to prove a concept in real calculations

On the video - a visualization of nitrogen clusters appearance in the simulation.



Thank you!