

# Combined use of HPC resources and grid infrastructures with Everest cloud platform

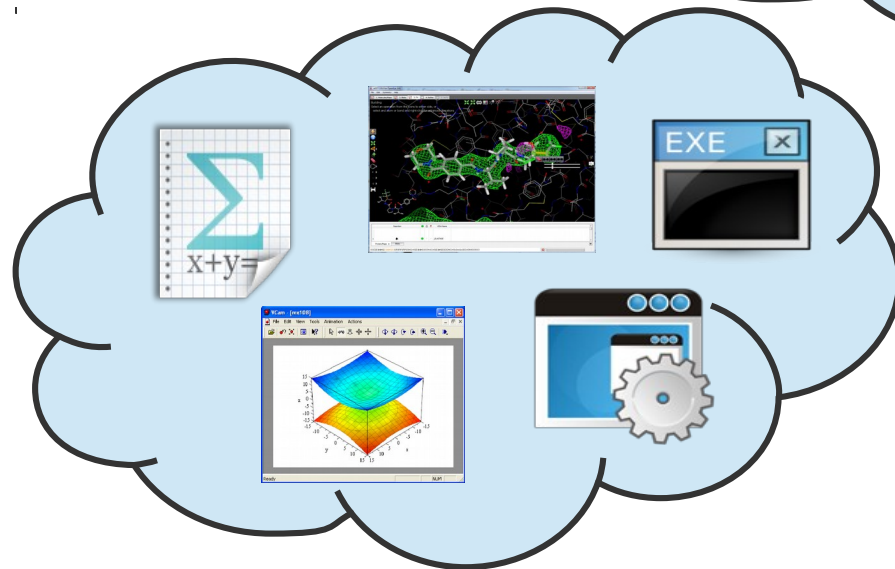
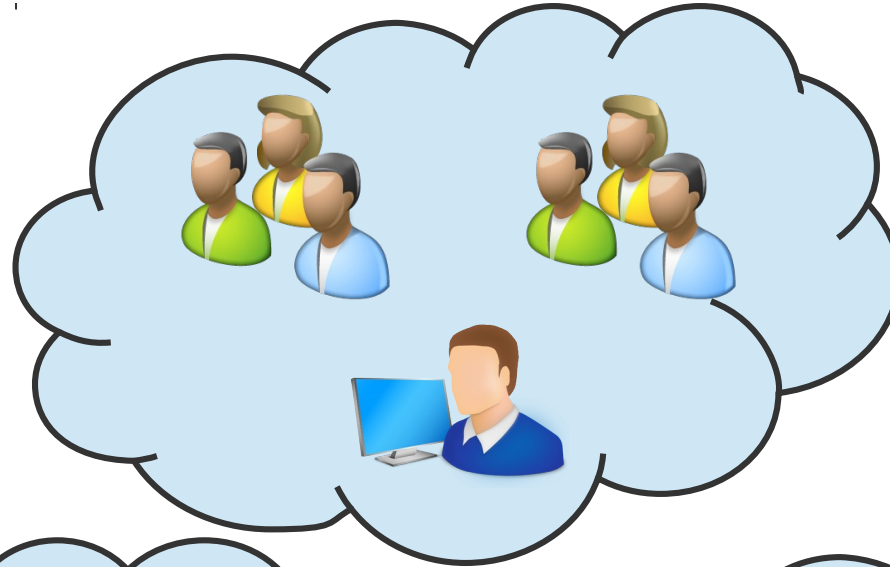
Oleg Sukhoroslov

Institute for Information Transmission Problems (Moscow, Russia)



# Motivation

Researchers



Applications

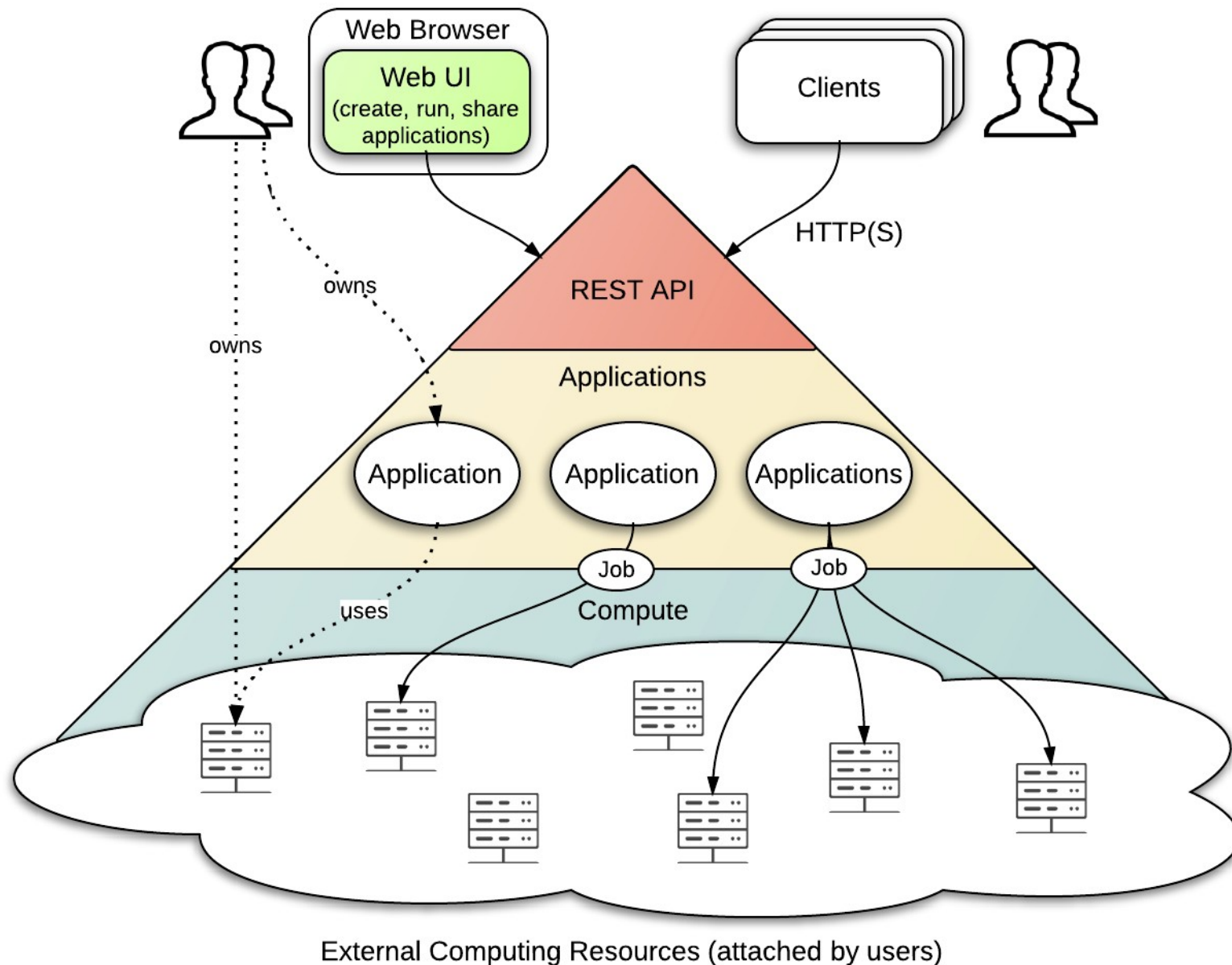


Computing Resources

# Everest

- Web-based platform supporting
  - Publication of computational applications as services
  - Binding applications to external computing resources
  - Running applications on arbitrary sets of resources
  - Sharing applications and resources with other users
- <http://everest.distcomp.org/>

# Everest Architecture



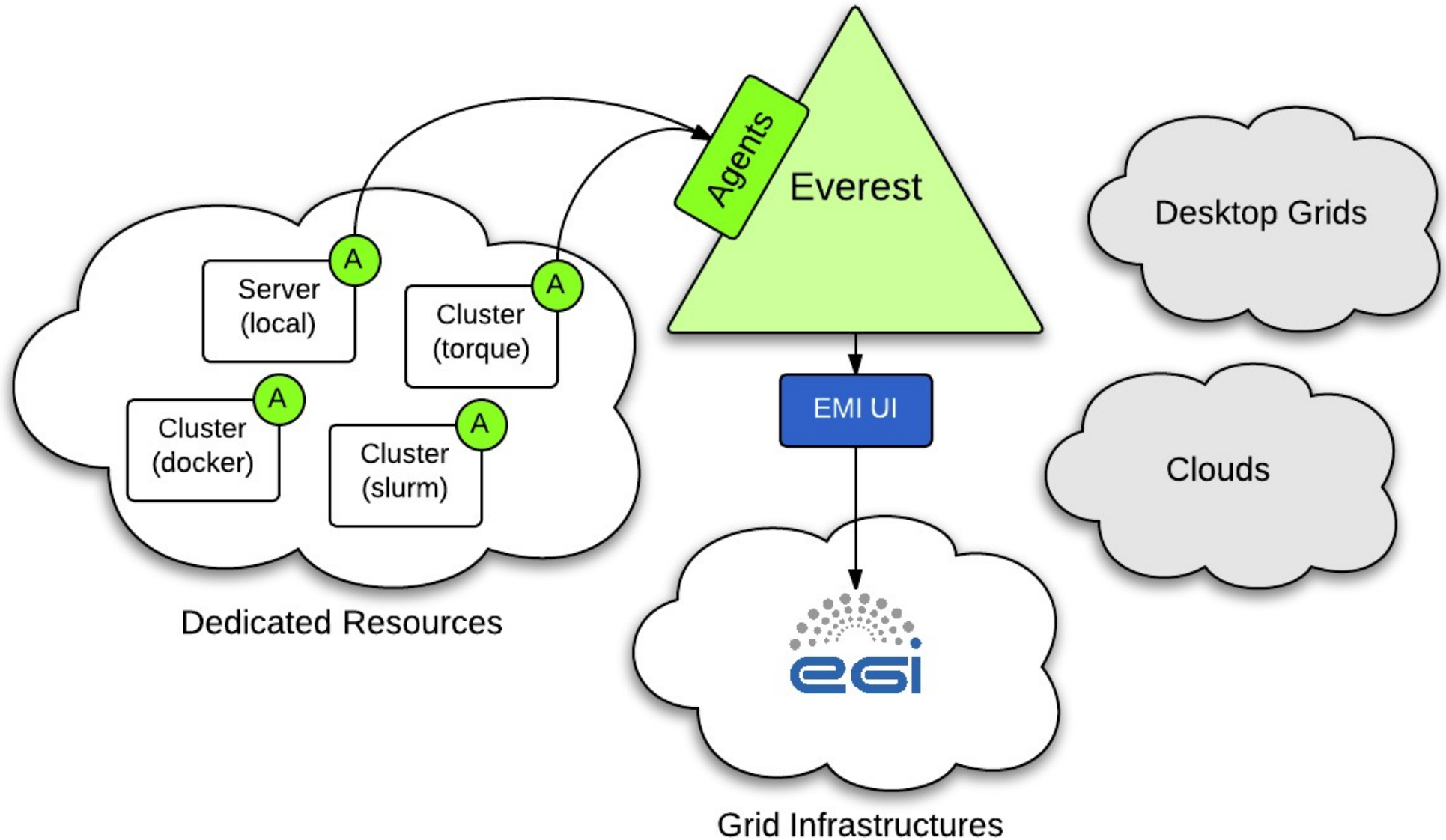
# Supported Applications

- Command
  - Generic skeleton to publish existing applications
  - Single task
- Parameter Sweep
  - Generic application to run parameter sweep experiments
  - Experiment is described using declarative syntax
  - Many independent tasks
- Workflow
  - Described using Python API
  - Many dependent jobs
  - Publication as a service (work in progress)

# Ad-hoc Computing Infrastructures

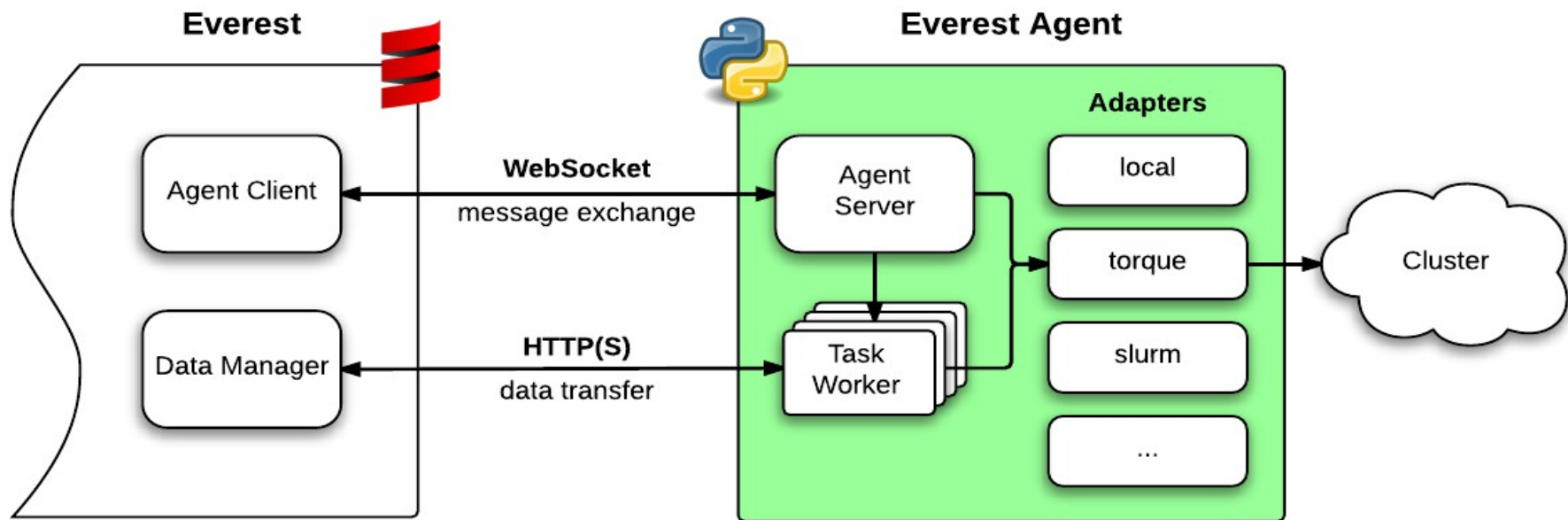
- Combination of available resources/infrastructures
  - Local servers and clusters, shared supercomputing centers, grid infrastructures, clouds, volunteer resources
- Suitable for
  - HTC and MTC applications (parameter sweep, workflows)
  - Load balancing
- Personal (user-level) or shared (project-level) ad-hoc infrastructures
  - Should be easy to setup and manage
  - Should support different resource types
  - Should not require admin privileges or complex middleware installation
- Everest
  - Provide necessary middleware as a service
  - Users attach their resources and combine them for running applications

# Integration with Computing Resources



# Everest Agent

- A mediator between the resource and the platform
- Supporting servers, clusters and resources behind a firewall
- Security mechanisms: white list, execution of tasks in Docker containers
- Open Source: <https://gitlab.com/everest/agent/>

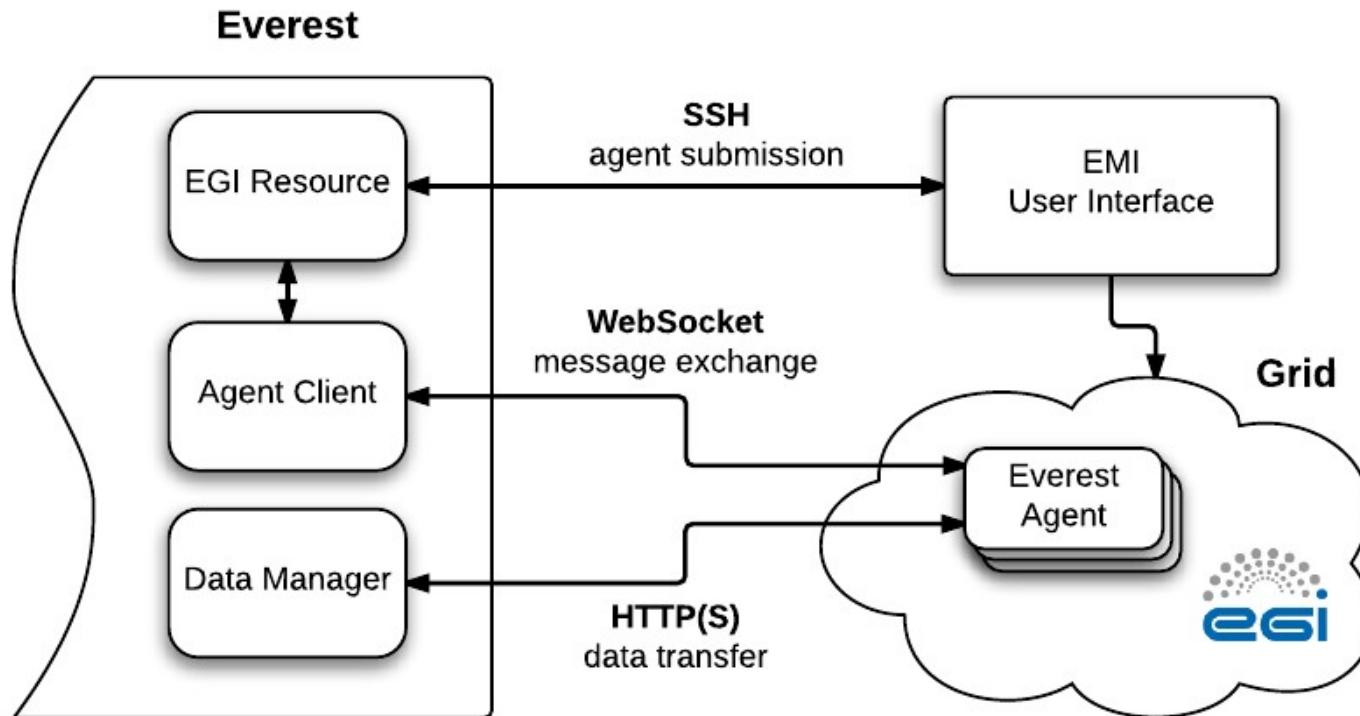




# Integration with EGI

- European Grid Infrastructure (EGI)
  - ~300 resource centers, ~500K CPU cores
- Challenges
  - Some users don't have access to grid UI
    - Using hosted docker image with configured EMI UI via SSH
  - Job submission requires a grid user certificate and VO
    - Enable user to pass proxy certificate to Everest
    - Create a single Everest resource per VO
  - Unpredictable delays while scheduling jobs in grid
    - Use “pilot jobs” strategy to allocate grid resources before scheduling
    - Reuse agents as pilots (a single resource is backed by many agents)

# Integration with EGI



Name	State	Type	Total Slots	Free Slots	Max Tasks	Total Tasks	Running Tasks
BIOMED	ONLINE	EGI	396743	25461	42	42	42
ENMR.EU	ONLINE	EGI	552187	31103	0	0	0
ESR	ONLINE	EGI	136611	5008	0	0	0

# Combined use of HPC cluster and EGI

- Parameter sweep (geophysics) with 670 tasks

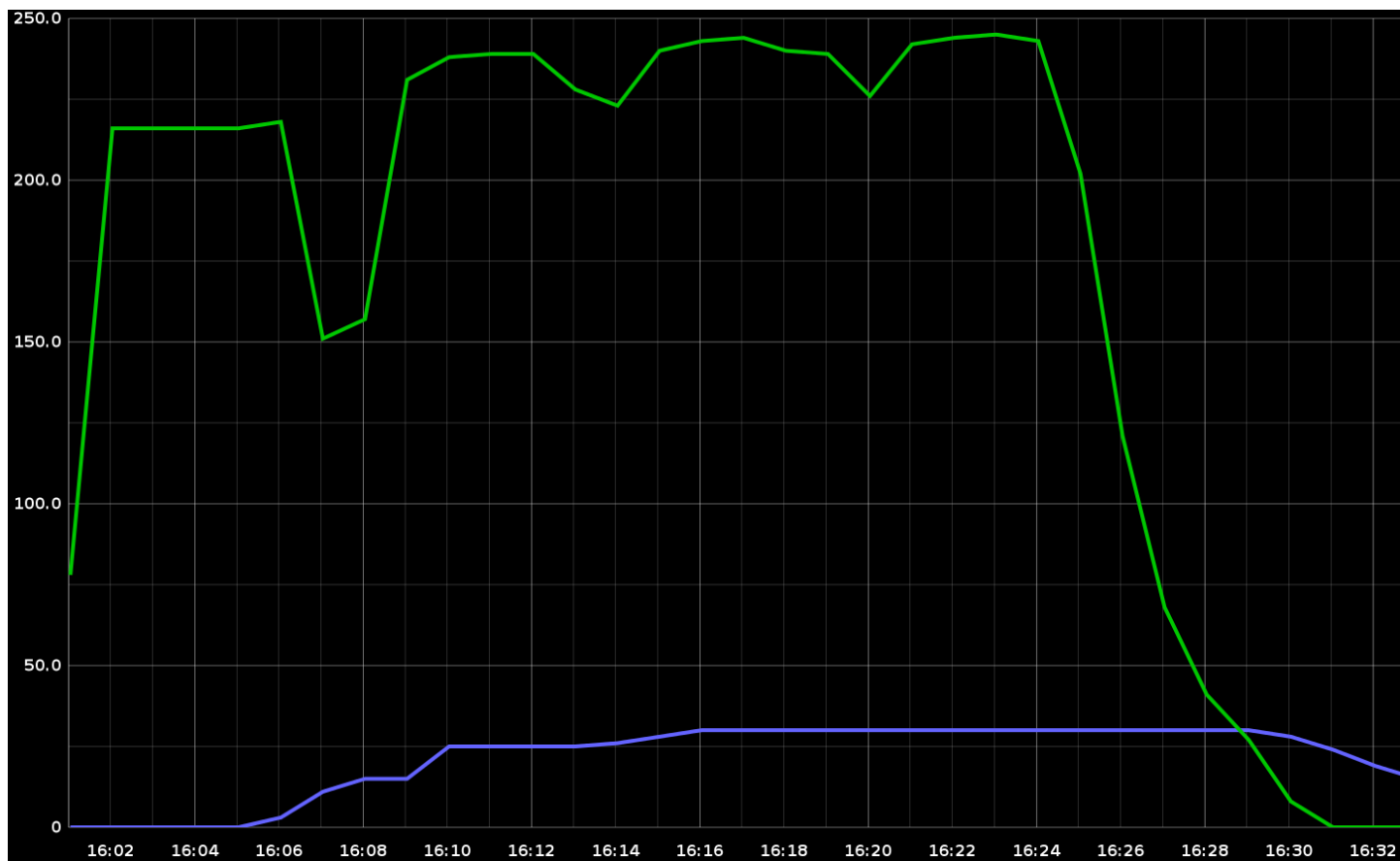
Name	State	Type	Total Slots	Free Slots	Max Tasks	Total Tasks	Running Tasks
<a href="#">ESR</a>	ONLINE	EGI	136579	4897	51	51	51
<a href="#">My Cluster</a>	ONLINE	torque	216	0	216	216	216



# Combined use of HPC cluster and EGI

- Parameter sweep (molecular docking) with 1000 tasks

Name	State	Type	Total Slots	Free Slots	Max Tasks	Total Tasks	Running Tasks
BIOMED	ONLINE	EGI	396115	26334	30	30	30
My Cluster	ONLINE	torque	216	0	216	216	216



# Future Work

- EGI integration improvements
  - Prefer “responsive” computing elements
  - Allocate multiple cores per agent, prestage task data
  - Generate and update proxy certificates
- Integration with other types of resources
  - Cloud computing services, volunteer computing
- Efficient scheduling across multiple resources
  - Collect and analyze resource metrics, predict task run time
- Optimization of data transfer
  - Cache input files on agents, support other data transfer protocols

# Conclusion

- Everest enables users to publish and run applications on ad-hoc computing infrastructures
  - Platform as a Service, accessible via web browser
  - Based on user-level agents deployed on resources
  - Combining HPC resources and grid infrastructures
- More information: <http://everest.distcomp.org/>