HTCaaS: Efficient and Simplified Large-Scale Scientific Computing over Supercomputers, Grids and Cloud

Soonwook Hwang, Seoyoung Kim, Jik-Soo Kim, Seungwoo Rho, Sangwan Kim, Seokkyoo Kim
National Institute of Supercomputing and Networking (NISN) at KISTI, Daejeon, Republic of Korea

Problem Statement

Exploiting Computing Resources for High-Throughput Computing
- A large number of independent, sequential jobs that can be scheduled on many different computing infrastructures
- Leveraging the complexity of integrating various computing resources
- Effective management and exploitation of all available computing resources

The Number of Jobs are increasing
- Growing in the number of jobs drives HTC area into *Many-Task Computing*
- Needs efficient job submission, processing and load balancing

HTC-as-a-Service

- Hiding heterogeneity and complexity of integrating various computing infrastructures from users throughout *pluggable resource interfaces*
- Allowing users to efficiently submit a large number of jobs *at once* (based on the *Meta-job*) and monitor them
- Exploiting *intelligent scheduling algorithms* to support multiple users submitting large numbers of tasks to distributed computing resources
- Successfully integrated with *PLSI Supercomputers in Korea, International Computational Grids, and Amazon EC2

*PLSI : Partnership and Leadership for the Nationwide Super- computing Infrastructure


System Architecture

- Agent-based multi-level scheduling & streamlined job dispatching
- User-level job scheduling & accounting which can dynamically adapt to the current system loads
- Jobs & In/Outputs are managed by Job manager and User data manager
- Agents are dispatched from Agent manager and process jobs in Supercomputers, Grids and Clouds
- Loosely-coupled Service-oriented architecture with WS-Interface
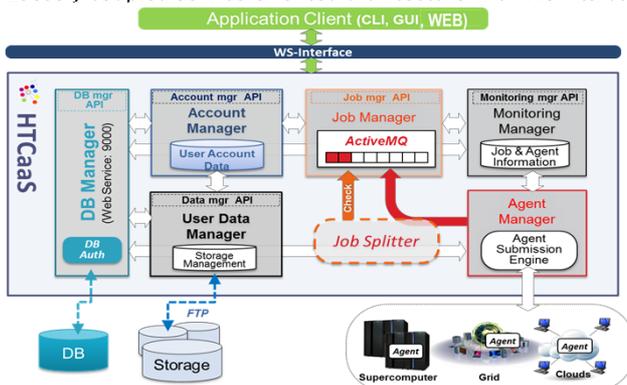
Interfaces & Client Tools

- Powerful *Meta-job* descriptions which allow users to easily specify a large amount of computations (e.g., parameter sweeps)
- Supporting *many client interfaces* including a native WS-interface, Java API, and easy-to-use client tools (CLI/GUI/Web portal)

Web portal

CLI

GUI
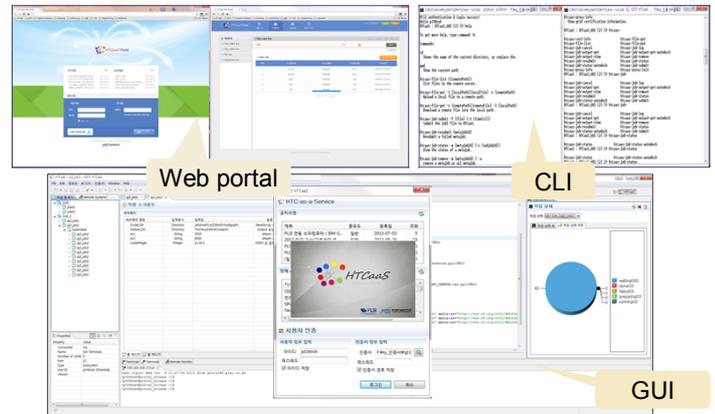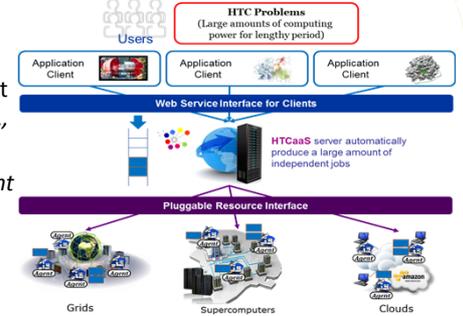
Application Support

END-USER SUPPORTS

Autodock; Pharmaceutical & New drug Discovery
- New drug discovery by simulating protein docking
- A target protein of 3CL-pro of SARS with 1.1 Million chemical compounds for protein docking
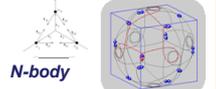
Madgraph5; High Energy Physics(HEP)
- Compute the collider processes of high-energy particles
- 5-million particle collisions resulting in tens of millions of interactions for QCD simulations
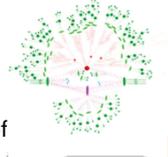
N-body Calculations; Nuclear Physics
- Calculations for simulation of a dynamical system of particles, usually under the influence of physical forces such as gravity

3rd PARTY DEVELOPER SUPPORT(API)

ezBioNet; Bio-technology
- A modeling & simulation system for analyzing biological reaction networks
- Support connection to computations on supercomputing

Current status & Future work

Efforts to improve HTC-as-a-Service
- Applying job profiling based allocation and provisioning model on Clouds
- Carrying out *performance studies* using micro-benchmark and stress test on HTCaaS

Acknowledgement