

**ISGC 2014**

**Abstracts book**

## Table of contents

Large-Scale Scientific Simulations throughout HTCaaS: Technologies, Practice and Applications . . . . .	3
---------------------------------------------------------------------------------------------------------	---

Abstract ID : 23

# Large-Scale Scientific Simulations throughout HTCaaS: Technologies, Practice and Applications

## Content :

We have designed and implemented HTCaaS(High-Throughput Computing as a Service) to provide researchers with ease of exploring large-scale and complex scientific problems by effectively leveraging national supercomputing infrastructures. HTCaaS has been successfully integrated with PLSI (Partnership & Leadership for the nationwide Supercomputing Infrastructure) supercomputers in Korea and is currently running as a pilot service on top of it. PLSI provides researchers with an integrated view of geographically distributed supercomputing clusters to solve complex and demanding scientific problems which is similar to the XSEDE in U.S. and the PRACE in Europe. PLSI also provides a common software stack for accounting, monitoring, global scheduling (based on LoadLeveler) and a global shared storage system (based on GPFS).

HTCaaS exploits a synthesis of well known techniques and its own intelligent scheduling algorithm to effectively support multiple users independently submitting large numbers of tasks to a collection of computing resources. HTCaaS employs Meta-Job (consisting of many tasks described as OGF JSDL) based automatic job split & submission to support parameter sweeps or N-body calculations, Agent(Pilot-Job)-based multi-level scheduling mechanism (exploiting LoadLeveler as a first-level scheduler) to efficiently support a bag of tasks. To ensure fairness among multiple users in the system, HTCaaS provides user-level scheduling and dynamic fairness by maintaining separate job queues and agents per user. By employing data staging mechanism, HTCaaS can effectively leverage local disks of each compute resource to circumvent the performance bottleneck of the shared storage file system in PLSI. Finally, HTCaaS provides many client interfaces including native WS-interface/Java API for developers and easy-to-use client tools (CLI/GUI/Web portal).

HTCaaS has been applied to various scientific applications including pharmaceutical domain (new drug discovery and drug repositioning by simulating protein docking based on Autodock3/Autodock-Vina), Monte-Carlo simulations in high energy particle physics (MadGraph5-PYTHIA) and N-Body calculations from nuclear physics. All of these applications have a large number of tasks to be processed and require massive computing resources. For example, virtual screening process based on Autodock3 targets 5 different proteins (diseases) with 500,000 ligands (potential new drugs) which required 2 weeks of processing by utilizing more than 1000 cores. In high energy particle physics, Monte-Carlo methods are useful in the analysis of almost all experimental data to simulate the event feature of signal processes and their backgrounds. Specifically, we have utilized MadGraph5-PYTHIA MC tools in the high energy physics to simulate millions of particle collisions resulting in tens of millions of interactions which required 10 days of processing with more than 500 cores. By effectively integrating multiple geographically distributed PLSI computing clusters based on multi-level scheduling mechanism, HTCaaS has successfully supported these various scientific applications with easy-to-use end-user client tools.

HTCaaS is still actively being developed and improved to support more challenging scientific applications. Our future work can include improving the scalability of our HTCaaS system to accommodate massive amount of computing resources (i.e., agents),

supporting more complex scientific applications consisting of HPC, HTC or scientific workflows, and stabilizing current version of pilot system into production-level service on top of national supercomputing resources in Korea.

**Primary authors** : Dr. KIM, Jik-soo (KISTI)

**Co-authors** : Mr. RHO, Seungwoo (KISTI) ; Ms. KIM, Seoyoung (KISTI) ; Mr. KIM, Sangwan (KISTI) ; Dr. KIM, Seok-kyoo (KISTI) ; Dr. HWANG, Soonwook (KISTI)

**Presenter** : Dr. KIM, Jik-soo (KISTI)

**Track classification** : High Performance & Technical Computing (HPTC)

**Contribution type** : --not specified--

**Submitted by** : Dr. KIM, Jik-soo

**Submitted on** Thursday 14 November 2013

**Last modified on** : Thursday 14 November 2013

**Comments** :