

# SAGA

A Simple API for Grid Applications

## A Brief Introduction to SAGA



omi-uk  
[www.omii.ac.uk](http://www.omii.ac.uk)



# SAGA

A Simple API for Grid Applications

All material from this tutorial can be found at:

<http://saga.cct.lsu.edu/software/cpp/documentation/tutorials/loni-training-2010>

And at:

[https://svn.cct.lsu.edu/repos/saga-projects/tutorial/general\\_tutorial](https://svn.cct.lsu.edu/repos/saga-projects/tutorial/general_tutorial)

## General Information and Documentation

- ▣ General information
  - ▣ <http://saga.cct.lsu.edu/>
- ▣ Documentation:
  - ▣ <http://saga.cct.lsu.edu/software/cpp/documentation>
- ▣ API documentation
  - ▣ Python
    - ▣ <http://static.saga.cct.lsu.edu/apidoc/python/latest/>
  - ▣ C++
    - ▣ <http://static.saga.cct.lsu.edu/apidoc/cpp/latest/>
- ▣ Programmers Guide:
  - ▣ [https://svn.cct.lsu.edu/repos/saga/core/trunk/docs/manuals/programming\\_guide/tex/saga-programming-guide.pdf](https://svn.cct.lsu.edu/repos/saga/core/trunk/docs/manuals/programming_guide/tex/saga-programming-guide.pdf)

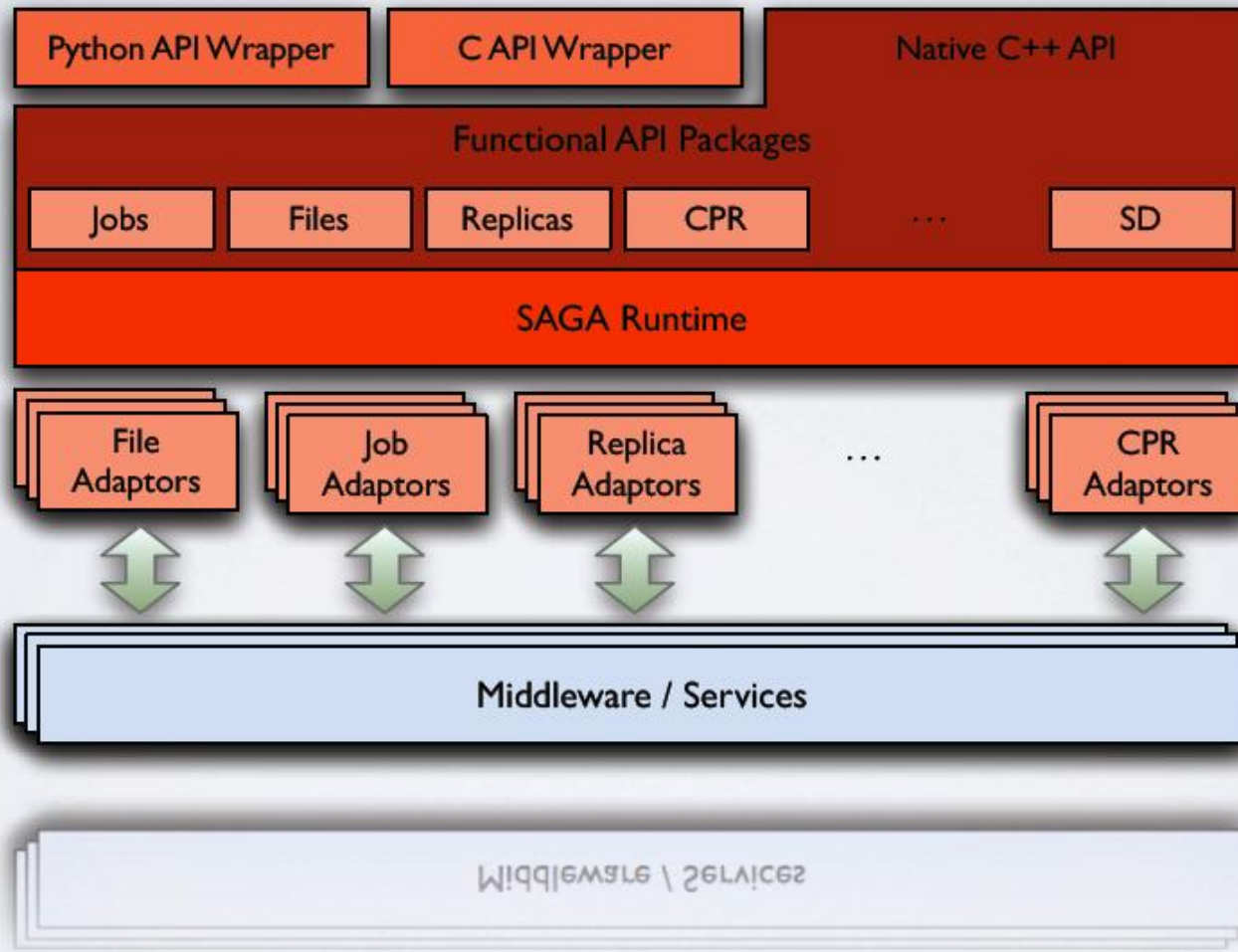
## Distributed Applications Development Challenges

- ▣ Developing Distributed Applications is fundamentally hard:
  - ▣ Intrinsic:
    - ▣ Control/Coordination & execution over Heterogeneous sites
    - ▣ Complex Design point/Models of Distributed Applications,
      - ▣ Reasons for using distributed CI -- more than (peak) performance result
  - ▣ Extrinsic:
    - ▣ (Complex) Underlying infrastructure & its provisioning
    - ▣ Large number Programming systems, tools and environments
      - ▣ Lack of well-defined interfaces & abstractions
      - ▣ Interoperability and extensibility become difficult
- ▣ Number of “effective” distributed applications that utilize resources sequentially, concurrently or asynchronously is low
  - ▣ Distributed CI: Is the whole > than the sum of the parts?
- ▣ See: DPA Survey Paper:
  - ▣ [http://www.cct.lsu.edu/~sjha/dpa\\_publications/dpa\\_surveypaper.pdf](http://www.cct.lsu.edu/~sjha/dpa_publications/dpa_surveypaper.pdf)

## SAGA: In a nutshell

- There exists a lack of Programmatic approaches that:
  - Provide general-purpose, basic & common grid functionality for applications and thus hide underlying complexity, varying semantics..
  - The building blocks upon which to construct “consistent” higher-levels of functionality and abstractions
  - Meets the need for a Broad Spectrum of Application:
    - Simple scripts, Gateways, Smart Applications and Production Grade Tooling, Workflow...
  
- Simple, integrated, stable, uniform and high-level interface
  - Simple and Stable: 80:20 restricted scope and **Standard**
  - Integrated: Similar semantics & style across
  - Uniform: Same interface for different distributed systems
  
- SAGA: Provides Application\* developers with units required to compose high-level functionality across (distinct) distributed systems  
(\* ) One Person's Application is another Person's Tool

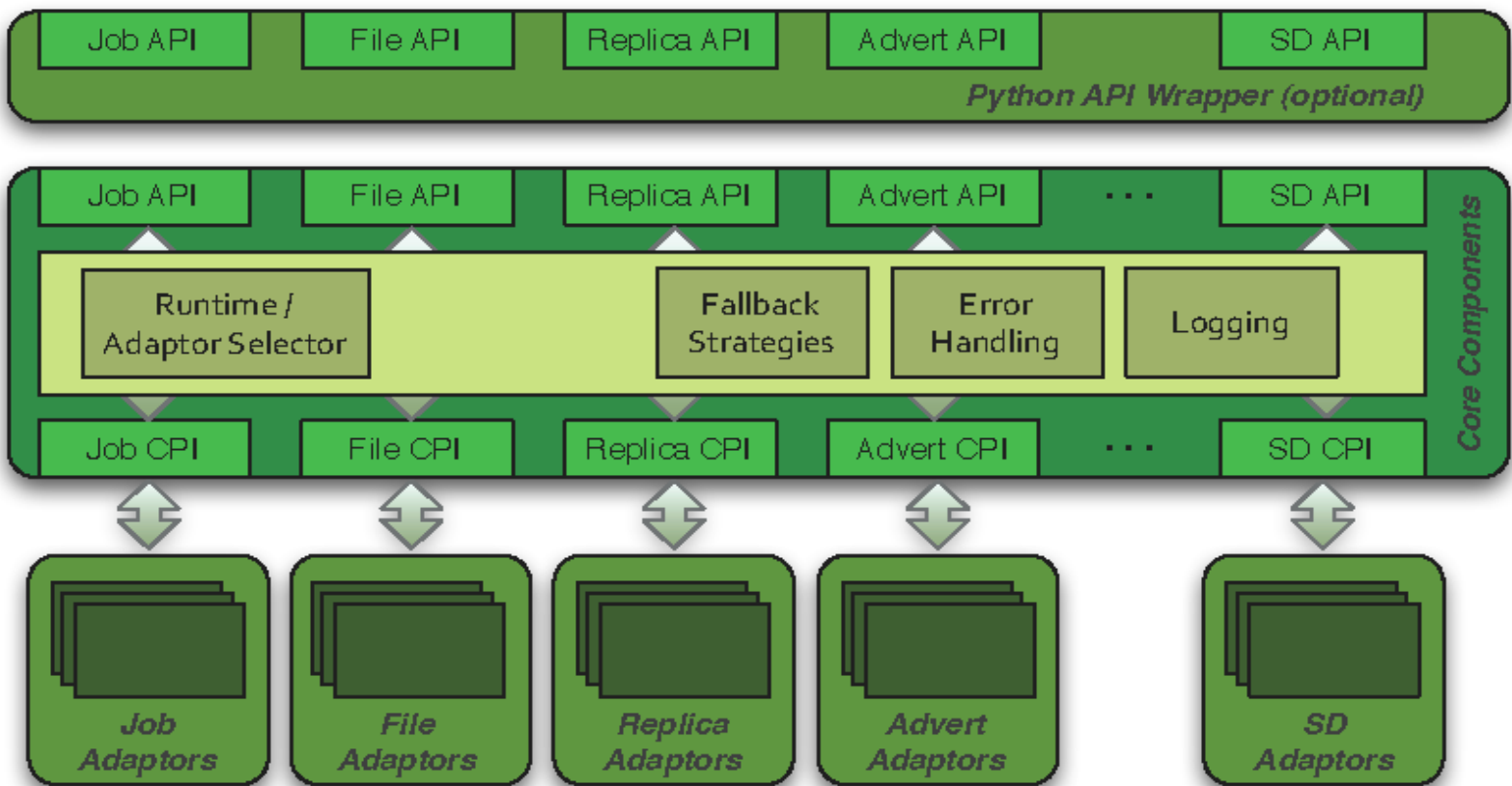
# SAGA: In a thousand words



# SAGA

A Simple API for Grid Applications

## SAGA: Architecture



## How is SAGA Used?

- SAGA is used to develop applications that are distributed by definition:
  - Simple extensions of “localized applications” (eg scripting)
    - MW applications, workers submitted to >8 back-ends
  - Novel Distributed Programming Models (eg Rep-Exch)
- SAGA: Build tools and implement abstractions that enable the execution of applications over distributed resources, *without modifying the applications*
  - Eg. Infrastructure Independent Pilot-Jobs
- SAGA: To provide uniform access layers to heterogeneous CI
  - Uniform access to EGI (ARC, gLite, Globus and Unicore/BES)
  - Simplify the building of tools and Gateways



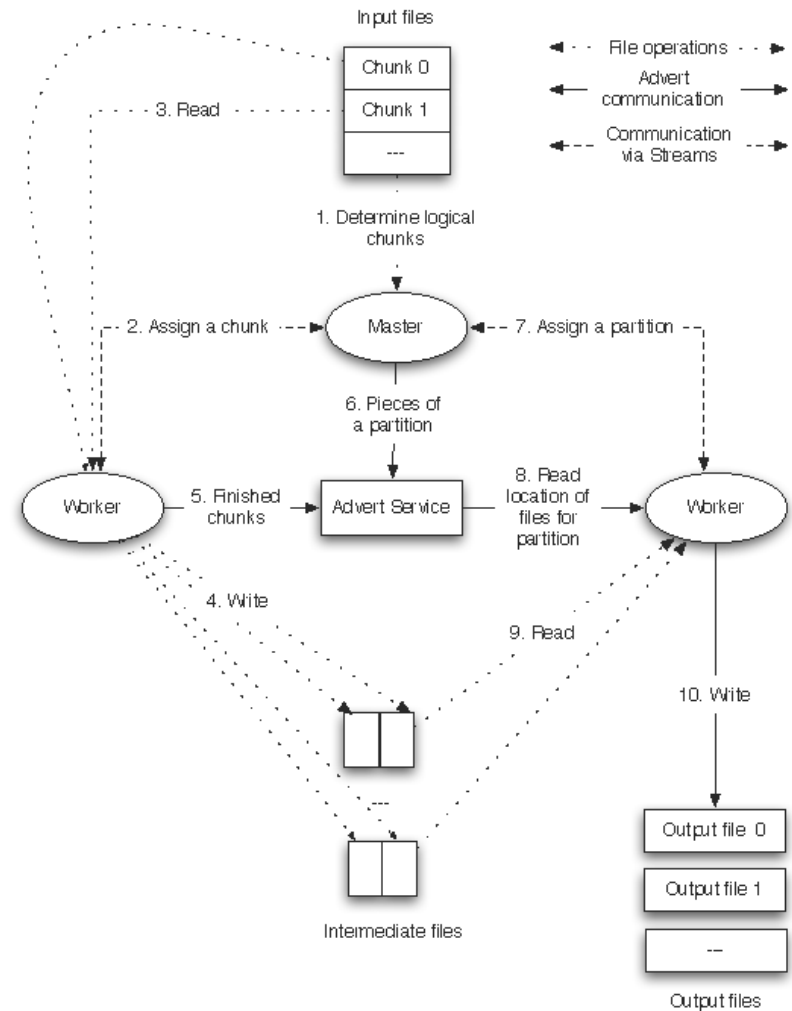


# 1. Develop applications that are distributed by definition

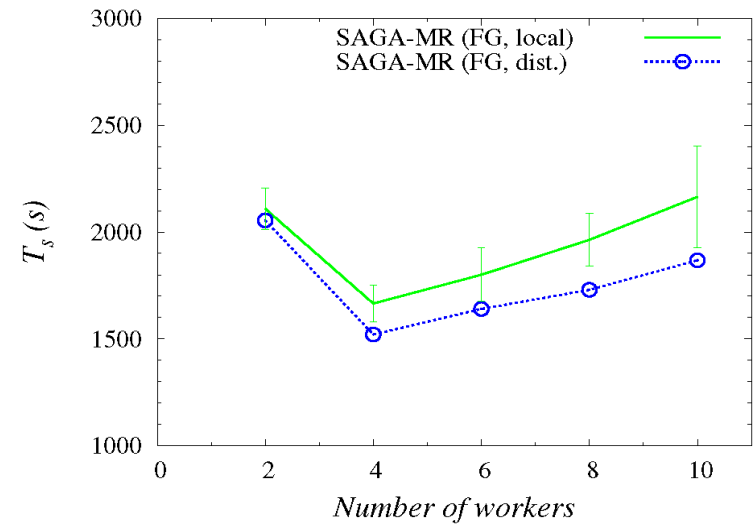
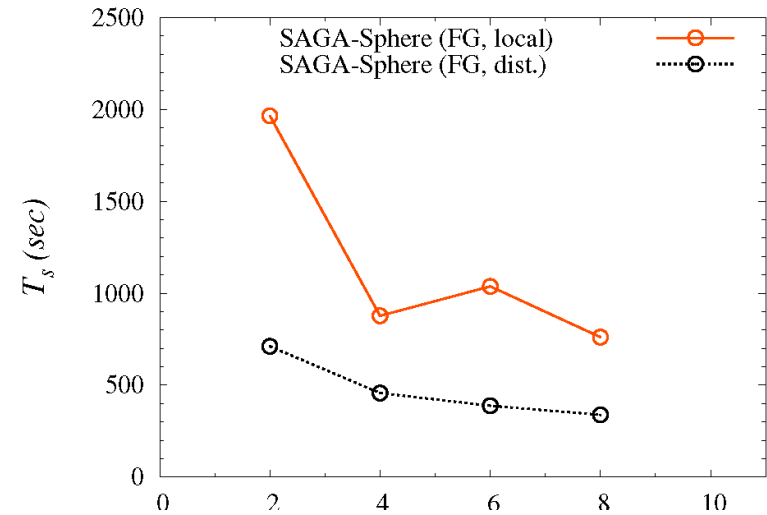
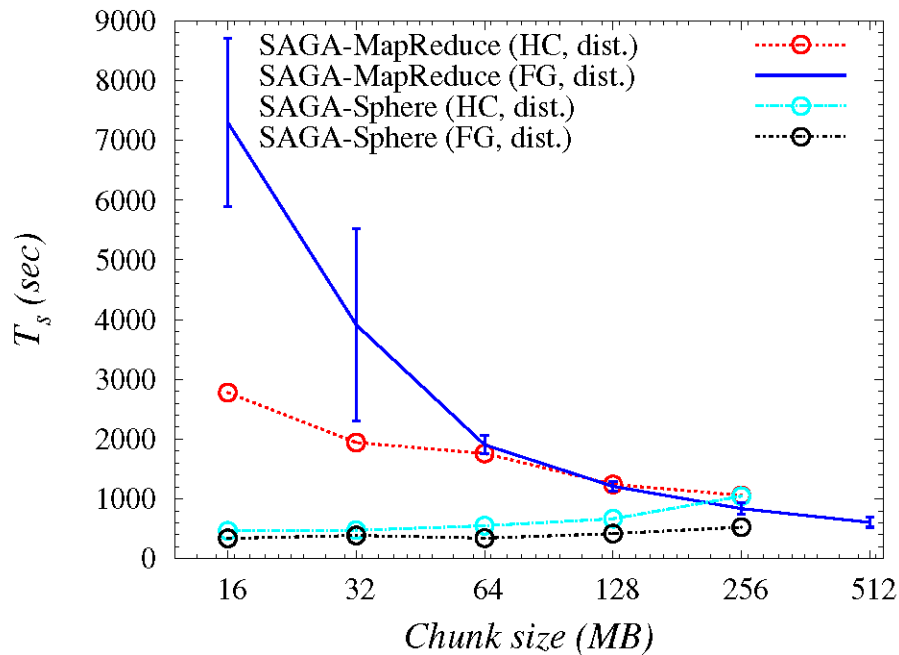
```
0 ([fork://localhost]-[2516])
1 ([ssh://cyder.cct.lsu.edu]-[[fork://localhost]-[2..
2 ([https://grass1.man.poznan.pl:19021/]-[<Activity..
3 ([https://grass1.man.poznan.pl:19022/]-[<Activity..
4 (https://qb1.loni.org:50897/23012/1288218915/)
5 ([https://localhost:10001/arex-ut]-[<ActivityIden..
6 ([epr://localhost/Users/merzky/.saga/fg.india.sho..
7 ([https://localhost:10003/DEMO-SITE/services/BESF..
8 ([ssh://ec2-50-16-45-213.compute-1.amazonaws.com]..
9 ([fork://localhost]-[2524])
10 ([fork://localhost]-[2516])
11 ([ssh://cyder.cct.lsu.edu]-[[fork://localhost]-[..
12 ([https://grass1.man.poznan.pl:19021/]-[<Activit..
13 ([https://grass1.man.poznan.pl:19022/]-[<Activit..
14 (https://qb1.loni.org:50897/23012/1288218915/)
15 ([https://localhost:10001/arex-ut]-[<ActivityIde..
16 ([epr://localhost/Users/merzky/.saga/fg.india.sh..
17 ([https://localhost:10003/DEMO-SITE/services/BES..
18 ([ssh://ec2-50-16-45-213.compute-1.amazonaws.com..
19 ([fork://localhost]-[2524])
```

# SAGA: Develop applications that are distributed by definition

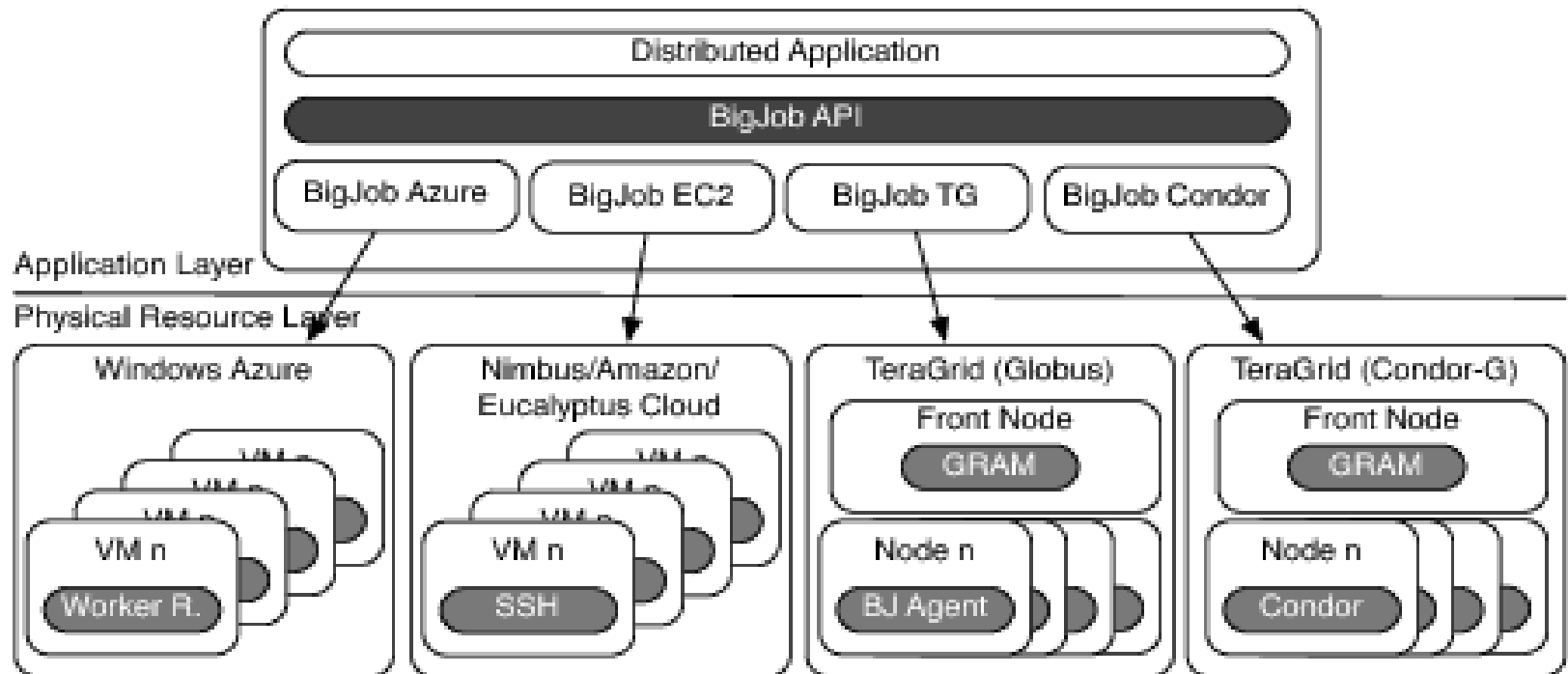
- How to develop a simple MR that is interoperable across infrastructure concurrently?
- Same application, same programming model:
  - Very different performance dependence
- Same application, different programming models
  - Very different performance dependence



## Understanding Distributed Programming Models

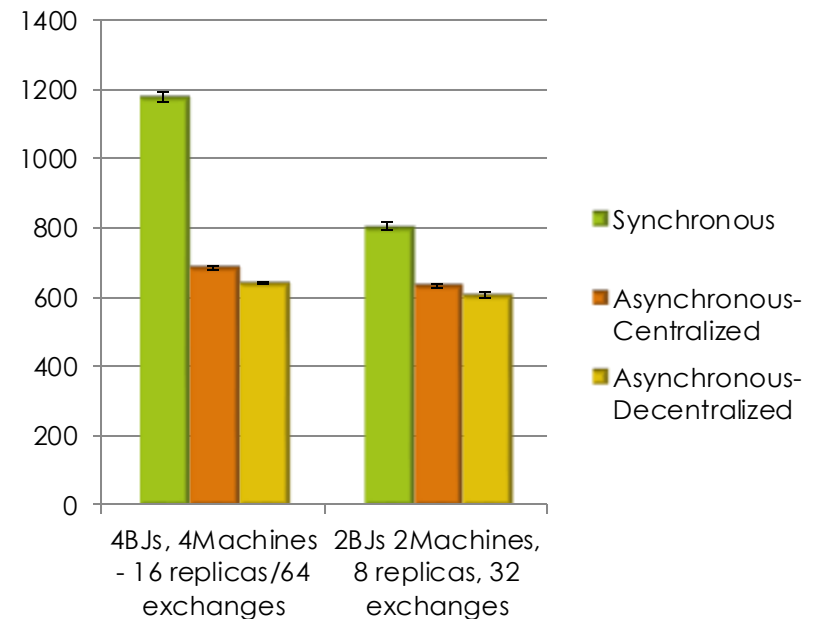
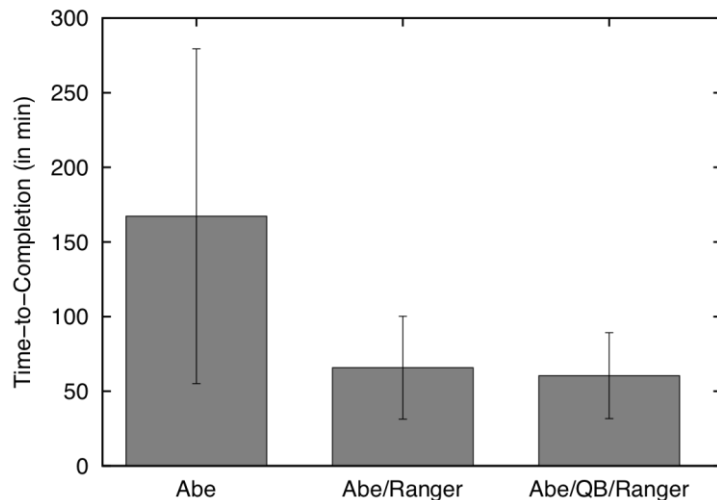


## 2. Tools for Effective Distributed Execution



## Distributed Adaptive Replica Exchange (DARE) Multiple Pilot-Jobs on the “Distributed” TeraGrid

- Ability to dynamically add HPC resources. On TG:
- Innovations in Distributed Algorithms:
  - Variants of RE: Sync (local) vs async (distr.)



# 3. Provides uniform access layers to heterogeneous CI



**TeraGrid™**  
EMPOWERING DISCOVERY



**NAREGI**

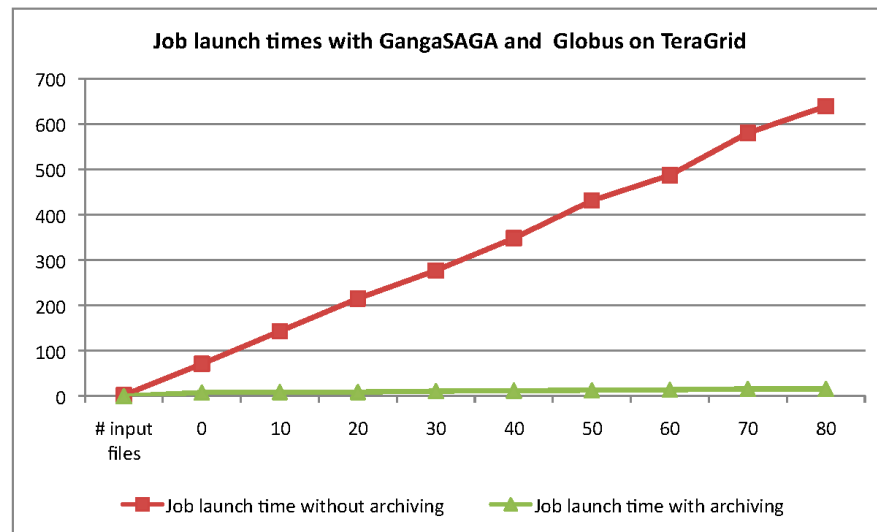
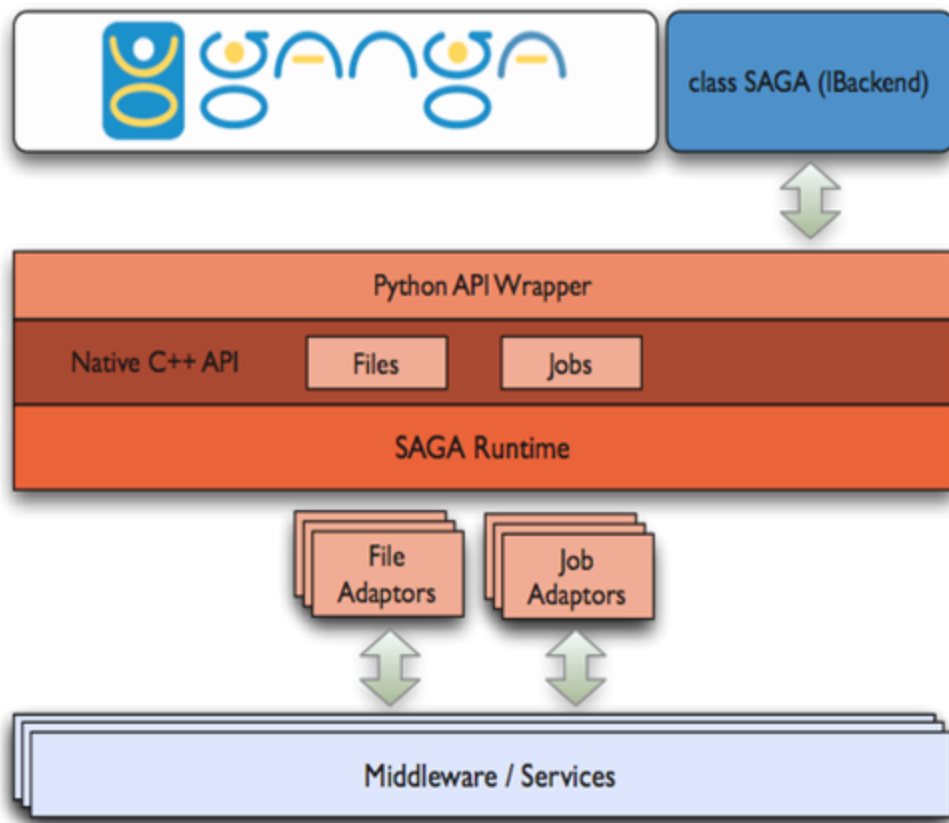
**NAREGI**

**eGEE**  
Enabling Grids  
for E-science

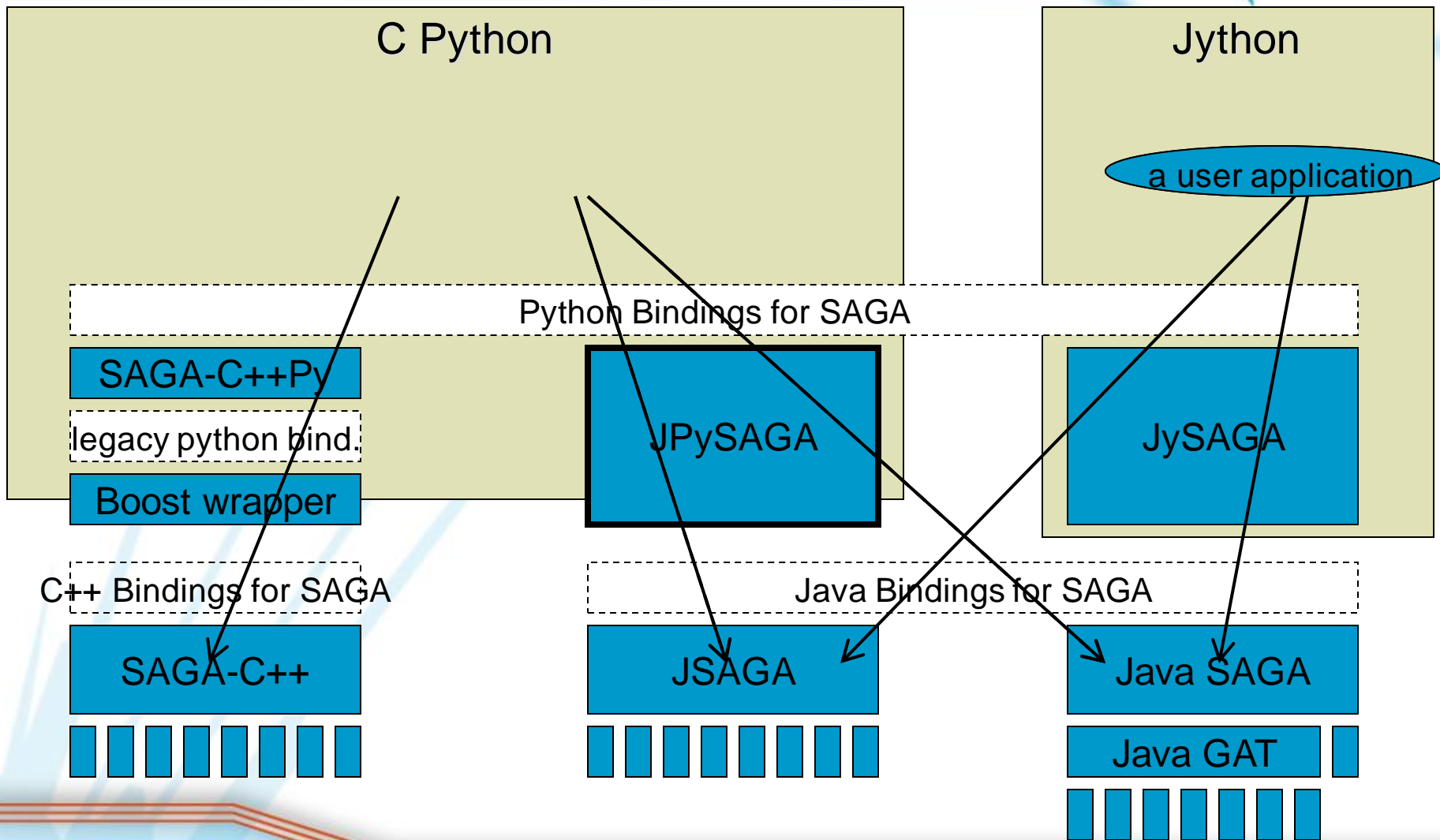
**eGEE**  
Enabling Grids  
for E-science



# SAGA-GANGA Integration



# Java-based Python SAGA wrapper



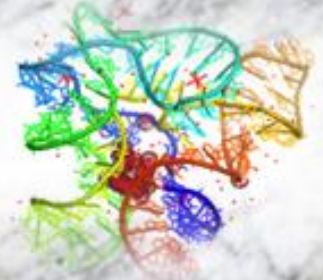




# DARE – Gateway for RNA-folding (Joohyun Kim, CyD)

## DARE-Rfold

RNA structure prediction utilizing multi-physics principles



HOME ABOUT DARE\_RFOLD JOB SUBMISSION▼ DOWNLOAD CONTACT

- DARE-Gateway:
  - Integrated, Extensible
  - Balanced: Scale-Up and Scale-out to
- DARE-RFOLD, DOCK, Bioscope (NG Sequence Data), STMD (Molecular Dynamics)

