Enabling Multi-task computation on Galaxy based Gateways using Swift

Ketan Maheshwari, Alex Rodriguez, David Kelly, Ravi Madduri, Justin Wozniak, Michael Wilde, Ian Foster

Argonne National Laboratory & University of Chicago

Overview

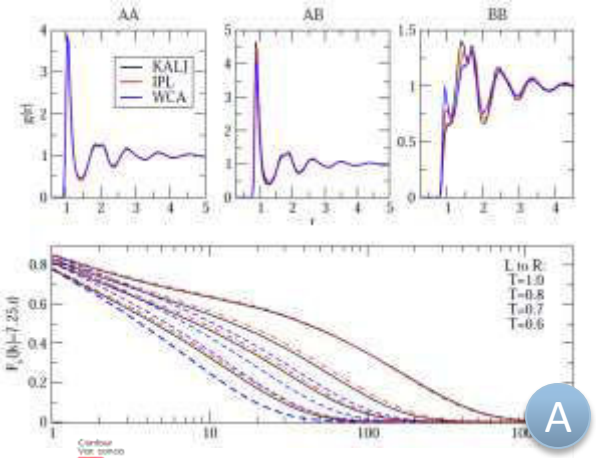
- Couple the **Swift** and **Galaxy** gateway frameworks
- Combine the features offered by Galaxy and Swift into an integrated platform
- Benefits to user communities from both systems
- Ease in the uptake of new resources
- Different integration *schemes* based on user requirements, and application characteristics

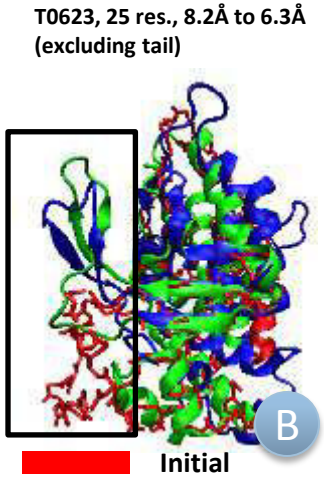
Motivation

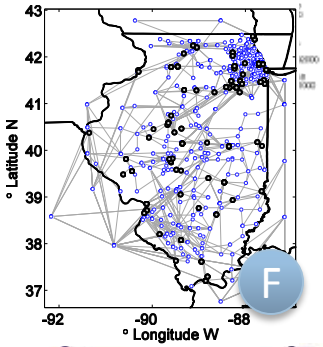
- Swift and Galaxy offer complementary functionalities to scientific community
- Galaxy (galaxyproject.org) offers a simple, user-friendly web-based interface for composing, execution, monitoring workflows
- Galaxy workflow results are sharable, reproducible and reusable
- Swift on the other hand, provides a sophisticated parallel and distributed computing platform
- Swift scripts are structured expressions of complex application flows which are readily executable on multiple, diverse and independent remote resources

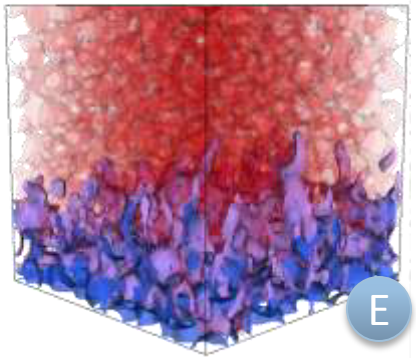
Swift: Enabling many-task applications

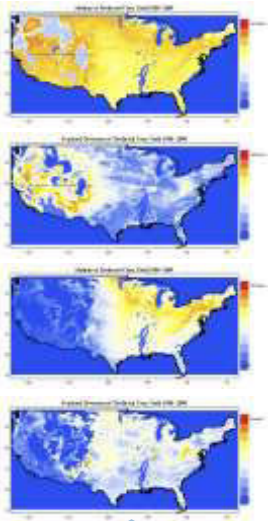
- A Simulation of super-cooled glass materials
- B Protein folding using homology-free approaches
- C Climate model analysis and decision making in energy policy
- D Simulation of RNA-protein interaction
- E Multiscale subsurface flow modeling
- F Modeling of power grid applications
- > All have published science results obtained using Swift

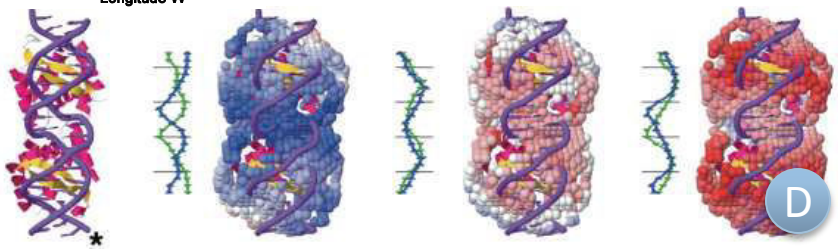




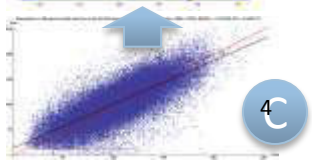




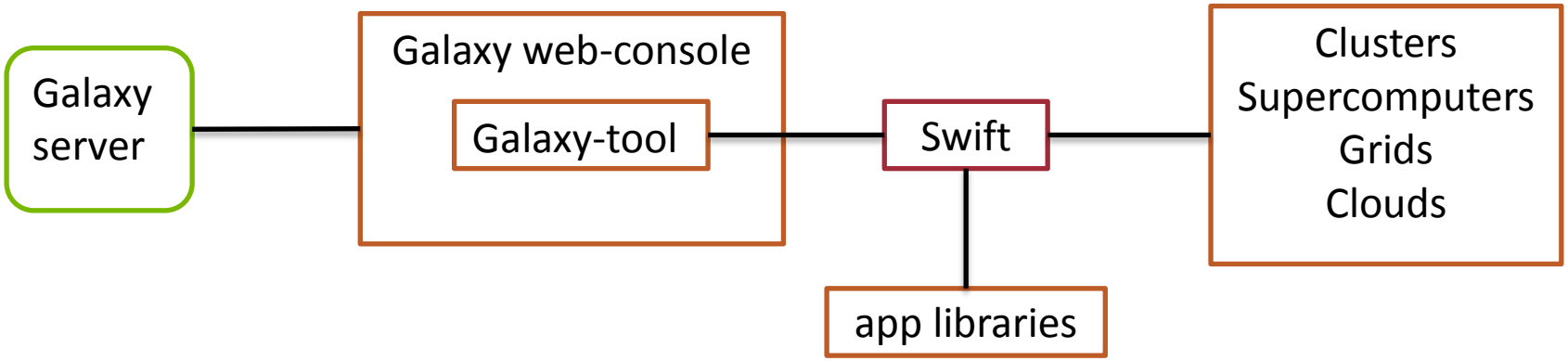




Protein loop modeling. Courtesy A. Adhikari

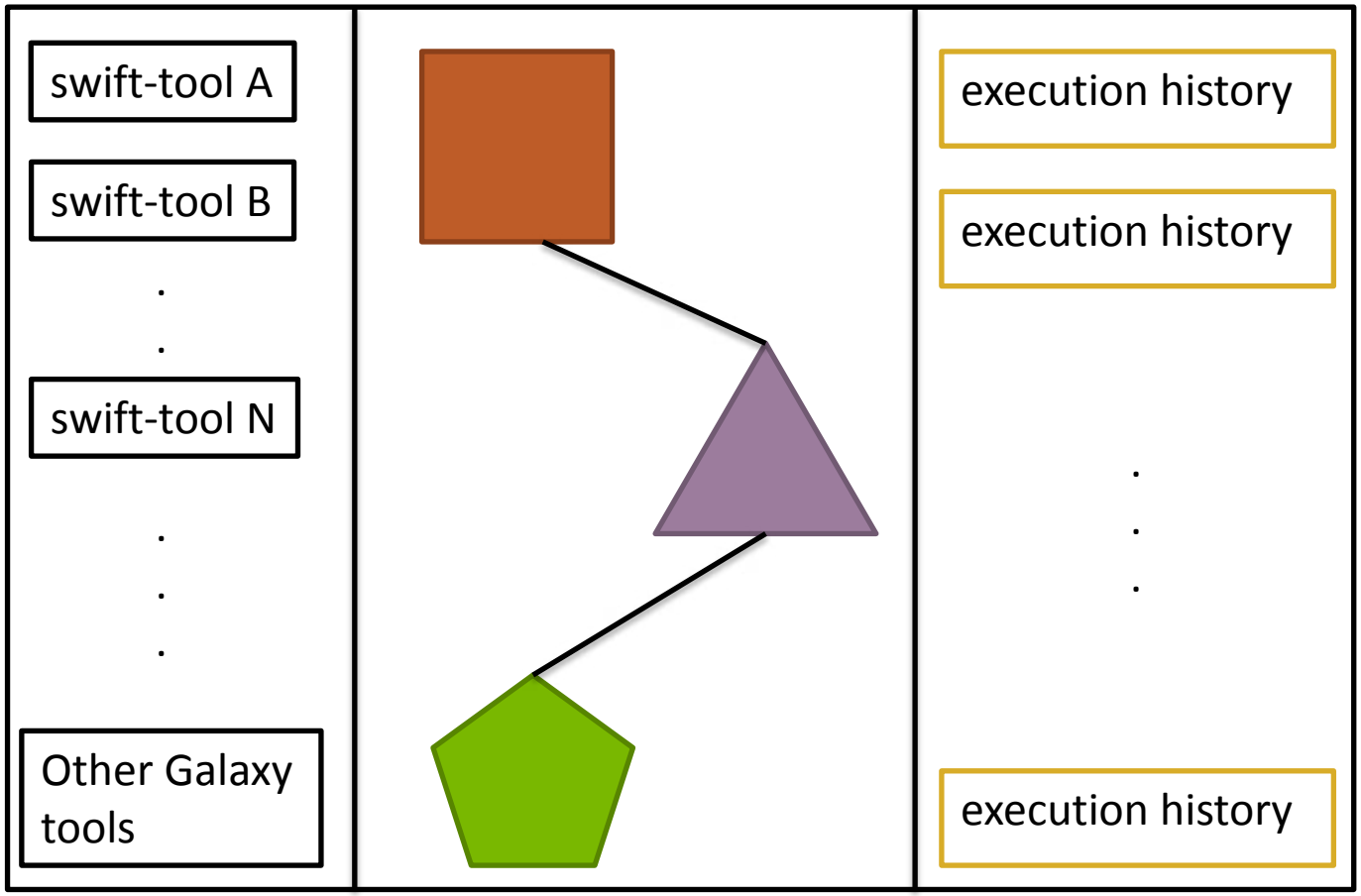


Swift-Galaxy Integration Schemes



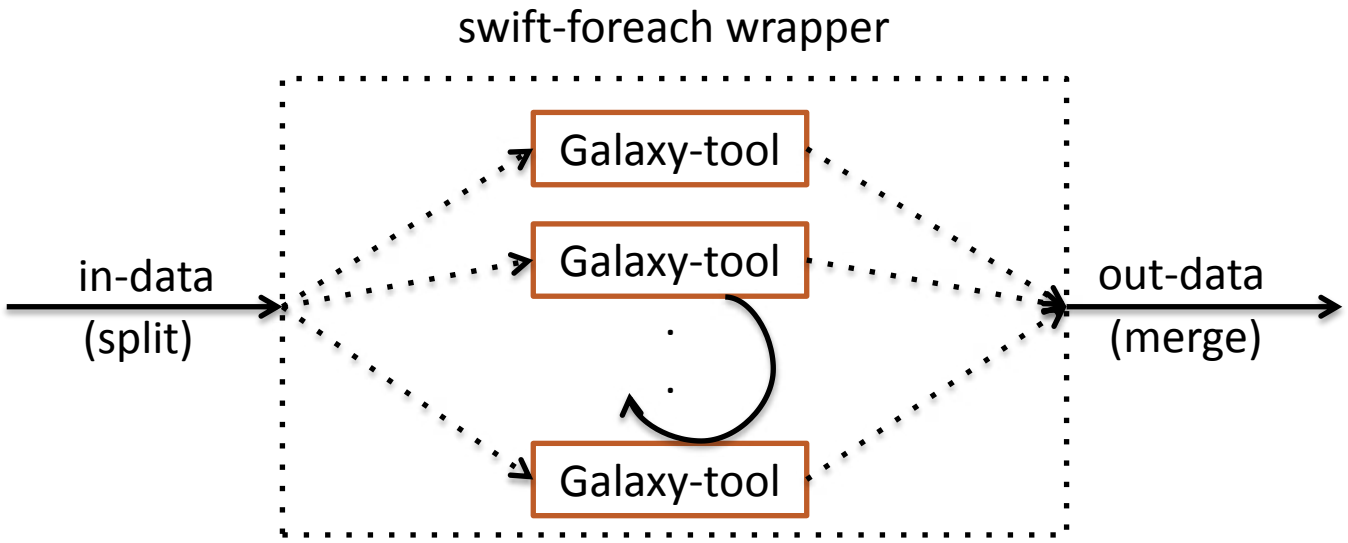
- Approaches enabling integration in different ways:
  - At tool level
  - At Workflow level
  - At language/expression level

Scheme1: Wrap Swift around Galaxy Tools



Scheme 2:
Interoperability between Swift and Galaxy expressions

Scheme 3: Harness Data Parallelism using foreach



Data Management

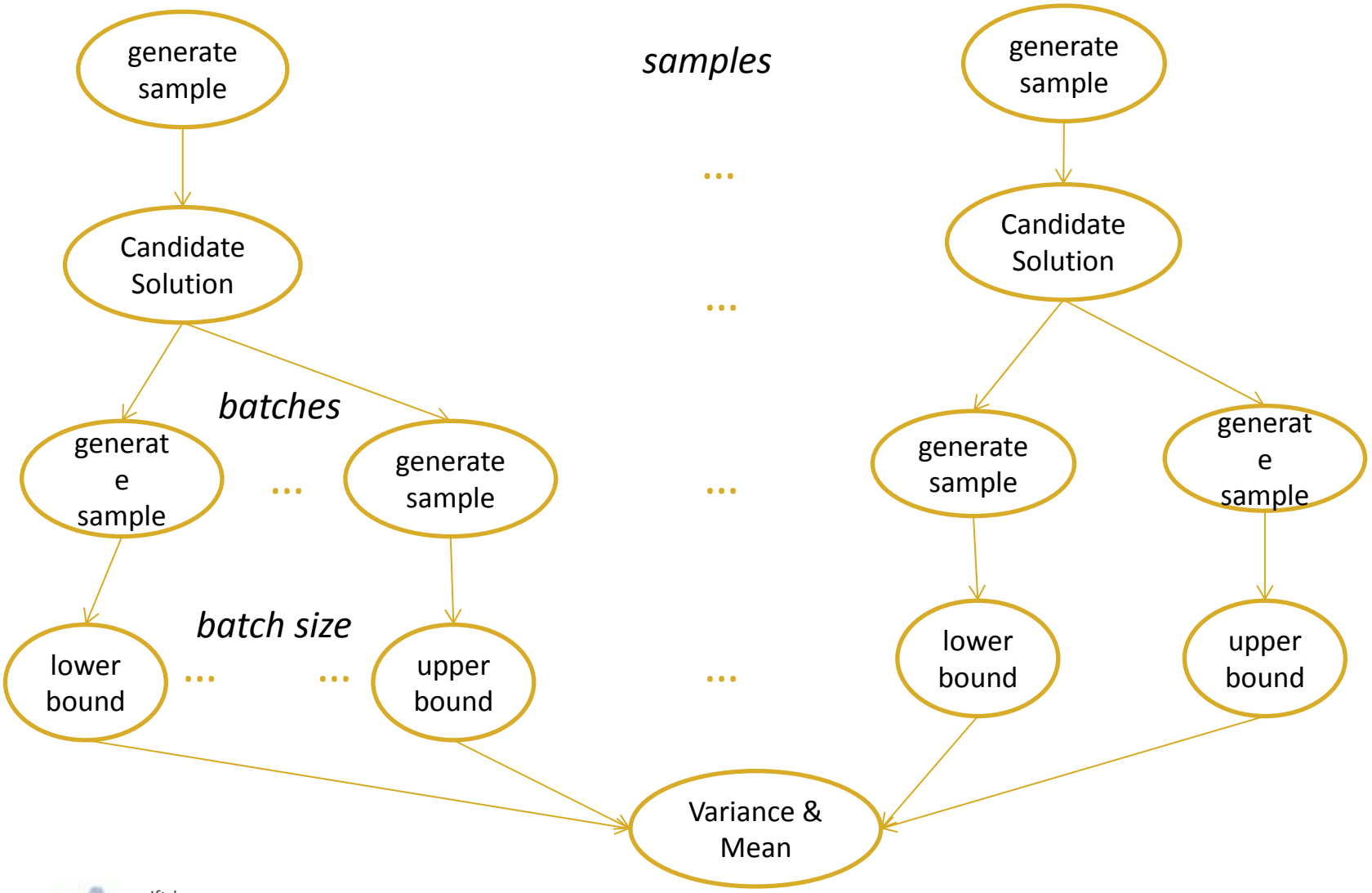
- Both Galaxy and Swift offer various data management capabilities
- Galaxy offers remote data uploading and viewing capabilities
- Swift allows disc resident data to be operated upon as program variables
- Swift's data-providers are interfaced with various data management protocols and can manage data motions at runtime

Computational Infrastructure

- Galaxy offers a limited support for Resource Managers
  - Needs additional configuration
  - Constrained, e.g. needs shared file system*
- Swift is robustly interfaced to a wider types of Resource Managers with finer control over job submission parameters:
  - Supports: PBS/Torque, SGE, SLURM, Condor
  - Supports bag-of-workstations: clouds, workstation clusters
  - Supports distributed file system, multiple execution sites simultaneously

* To the best of our knowledge

Evaluation: Inference analysis for power prices



Swift Script for Inference Analysis

```
import "mappings";
import "apps”;
type file;

int nS[] = [10, 100, 1000, 10000, 100000];
foreach S, idxs in nS {

  sample0 = gensample(S, wind_data);
  obj[idxs] = ampl(sample0);
  foreach B, idxb in [10:40:10] {
    foreach k in [0:B]{

      sample1 = gensample(S, wind_data);
      obj_l[idxs][idxb][k] = ampl_L(sample1);

      sample2 = gensample(S, wind_data);
      obj_u[idxs][idxb][k] = ampl_U(sample2, obj[idxs]);
}}}
```

Summary

- Swift-Galaxy integration improves science gateways:
  - User control
  - Structured distributed computing
  - Simple
  - Interactive
- Commonalities in basic execution model of Galaxy and Swift leads to many avenues of integration schemes
- Broadly, Swift acts as a backend manager while Galaxy being the frontend for operations
- Example of combining command-line and GUI based frameworks

Future Work

- A generic approach for each of the integration schemes
- Wider application adaptation
- Finer as well as broader exposure to configuration options to users
- Interactive run monitoring features
- Authentication features, Globus based identity management

Acknowledgements

- This work was supported by the U.S. Department of Energy, Office of Science, under Contract DE-AC02-06CH11357
- Colleagues at Swift and Globus groups

Evaluation: Demonstration via a Screencast Video

Thank you!

swift-lang.org