

Tigres: Template Interfaces for Agile Parallel Data-Intensive Science

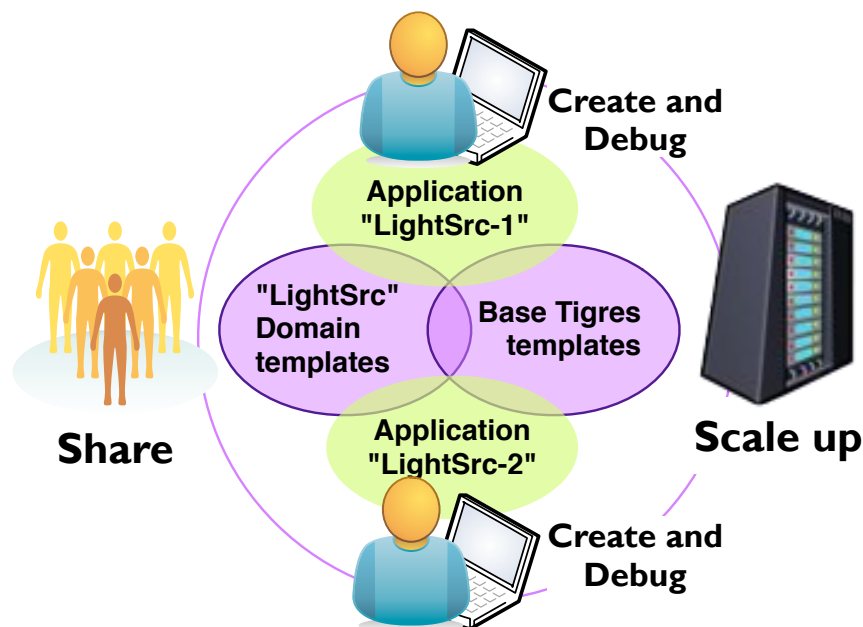
**Lavanya Ramakrishnan
Deb Agarwal**

Lawrence Berkeley National Lab

Tigres Team

- **Core Team**
 - **Deb Agarwal (PI), Lavanya Ramakrishnan, Dan Gunter**
 - **Valerie Hendrix, Gilberto Pastorello, Sarah Poon**
 - **Ryan Rodriguez, James Fox**
- **CS Research groups**
 - **John Shalf, Shane Canon, Nicholas Wright**
- **Science research groups**
 - **Cosmology - Alex Kim, Rollin Thomas, Stephen Bailey**
 - **Gamma Ray - Dan Chivers**
 - **Advanced Light Source - Dula Parkinson**
 - **HEP - Paolo Calafiura**
 - **Materials – Kristin Persson**

Tigres: Design *templates* for common scientific workflow patterns



Workflow Library: Implement templates as a library in an existing language

Basic Templates: Sequence, Parallel, Split, Merge

Early python release is now available!

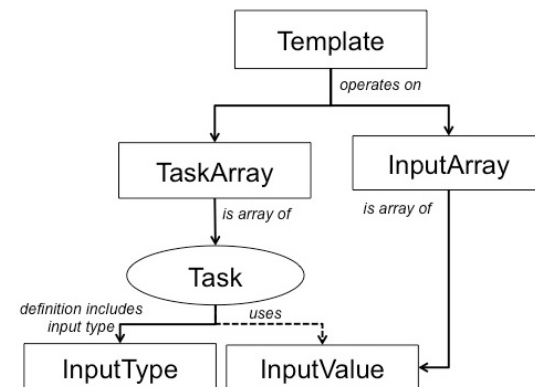
<http://tigres.lbl.gov>

Key Aspects of Tigres

- Targeted for large-scale data-intensive workflows
 - Motivated by “MapReduce” model
- Library model embedded in existing languages such as Python and C
 - “Extend current scripting/programming tools”
 - API-based, embedded in code
- Light-weight execution framework
 - “As easy to run as an MPI program on an HPC resource”
 - No persistent services
- **Scientist-Centered Design Process**
 - Get feedback from user continuously

Tigres: Current Status (in Release)

- **Iterative workflow development**
 - Simple data model
 - **Python API to compose and execute**
 - **Use programming language constructs for complex logic flows**
- **Execution**
 - Existing application binaries, functions
 - **Seamlessly run on Desktops, Clusters and HPC**
- **Monitoring, Provenance**
 - **Visual representation of graph that ran**
 - **Extensive monitoring from workflow execution**
 - **Support for adding user-level provenance**
- **Extensive documentation, examples and tutorials**



Tigres data model

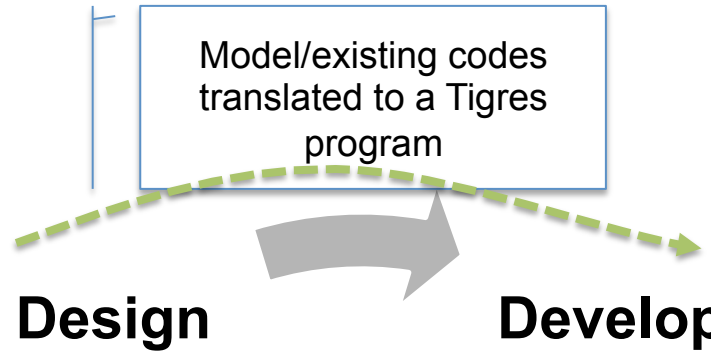
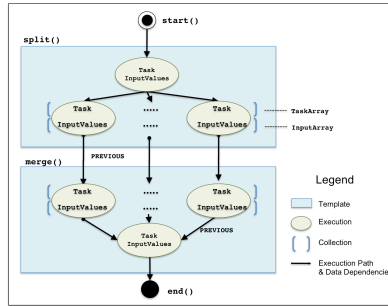
Tigres: Current Status

- **Recover failed workflows from logs (Testing)**
- **C API in development (90% done)**
- **Active Code Generation (Prototype)**
- **Fault tolerance and failure recovery API (Design)**

Scientist-Centered Design Process

- Usability studies provides semi-structured feedback from end-users
 - *Not the same as requirements gathering*
 - Limited literature on doing usability for APIs
- Round 1: Paper API & Google Docs Coding Session
 - Goal: Nomenclature and desired features
 - Priorities: Nomenclature, Monitoring, Dependency syntax, ..
- Round 2: Initial Prototype with documentation
 - Goal: Effectiveness of using API for specific problems
 - Understanding experience relative to programming work styles – Opportunistic, Pragmatic, Systematic
 - Questionnaire and interview and a 3/6 month follow-up

Iterative Scientific Workflow Process

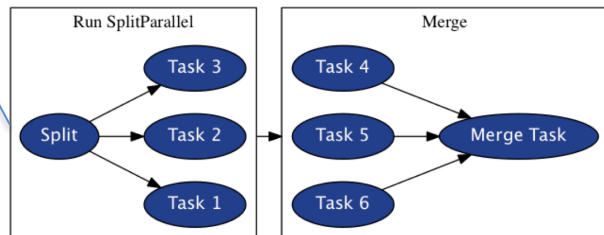


```
start(name="MyWorkflow")  
...  
split(name="Split"...)  
merge(name="Merge"...)  
...  
end()
```



Program state
available during and
after runs

Feedback



Run

Desktop

HPC



May be a partial
recovery run

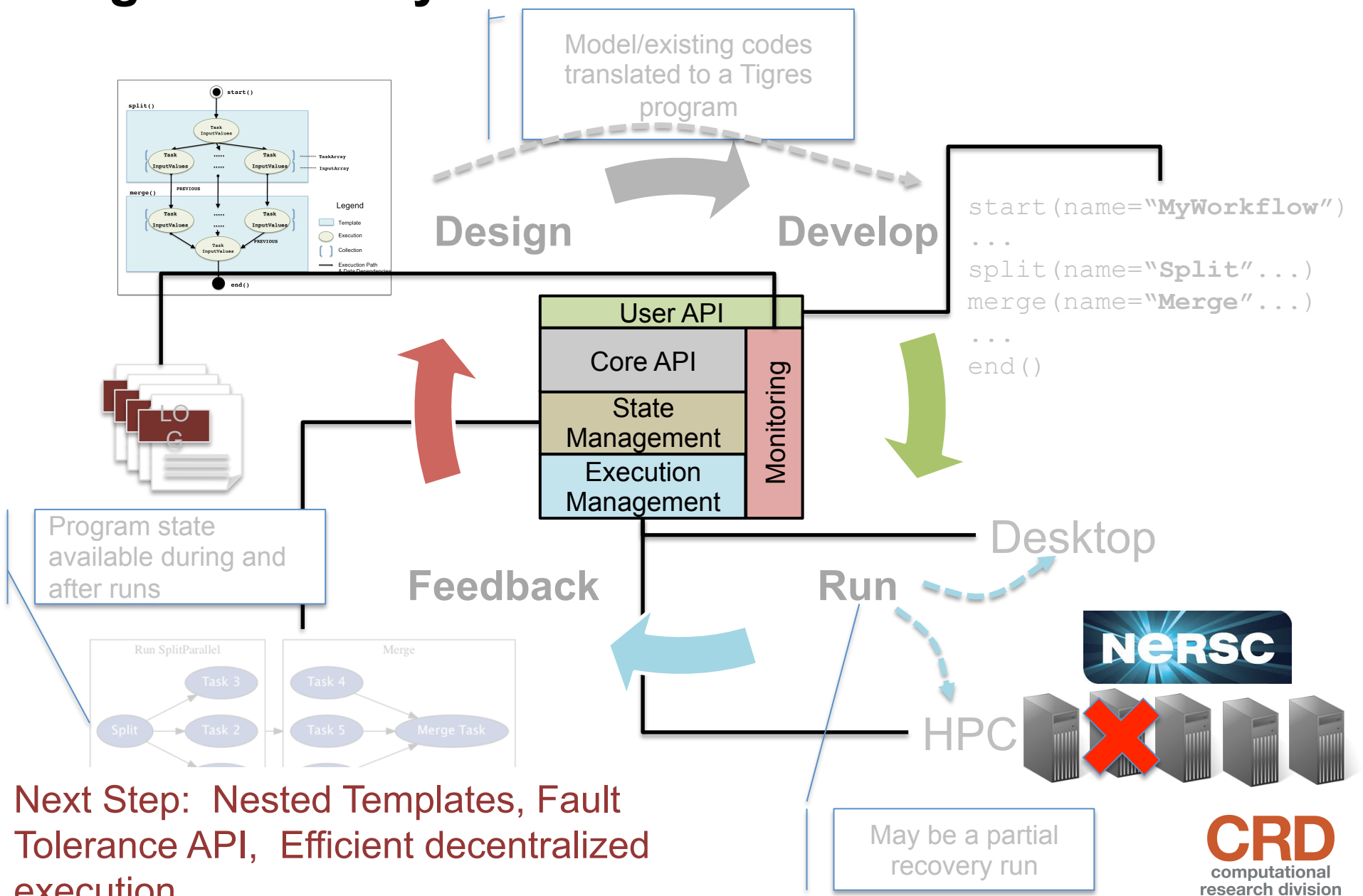


U.S. DEPARTMENT OF
ENERGY

Office of
Science

CRD
computational
research division

Tigres “Library” Model



Other Collaborations

- **DALHIS: INRIA Associated Team**
 - Building a data analysis environment using shared space execution and cloud models
 - *Paper: Combining Workflow Templates with a Shared Space-based Execution Model, WORKS 2014*
- **NERSC**
 - Identifying next-generation workflows and supporting services needs at HPC centers
- **ARES**
 - Use of Tigres for managing shared data-analysis workflows
- **Additional communities**
 - Climate CASCADE SFA, Berkeley Institute for Data Science (BIDS), ..

Open Research Topics

- **How does a “computational/data” workflow tie with the larger scientific process and scientist’ development environment?**
- **How do we balance the dynamic, interactive and iterative needs with performing global optimizations needed for exascale?**
- **How do we provide a framework that allows for data fusion from multiple diverse sources that can be used to derive knowledge?**

Questions?

- Website: <http://tigres.lbl.gov>
- LRamakrishnan@lbl.gov