Embracing Big Data using Scalable Workflows

Contact: Jianwu Wang, Ph.D.
jianwu@umbc.edu
http://userpages.umbc.edu/~jianwu/

Research Interests: Big Data Application/Workflow, Provenance, Cloud Computing, Service-Oriented Computing

Research goals
- Improve usability at front end
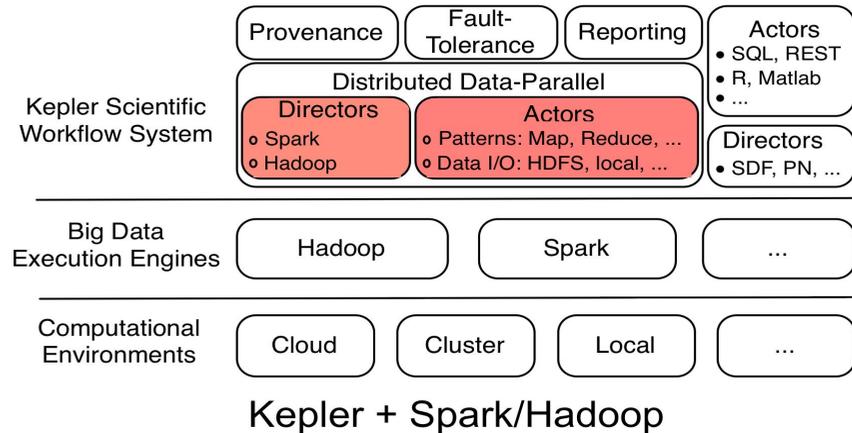- Optimize execution at back end

Applications

Big Batch Data
- Bayesian network ensemble learning
- Bioinformatics workflows
- Health IT
- ...

Big Stream Data
- Traffic sensor data processing
- Travel companion recommendation
- Smart farm
- Intrusion detection
- ...

System

Kepler + Spark/Hadoop

Extended Storm System

Contributions

- Easy Big Data application construction through visual programing
- Adaptive Big Data application execution on different engines
- Optimal execution for legacy tool parallelization and Big Data pattern selection

- Finer bolt operations for richer semantics and easier reusability
- Better throughput via: 1) Hash B-tree for caching, 2) Historical data partitioning and 3) Better parallelism among finer bolt operations