**STAMPEDE: A Framework for Monitoring and Troubleshooting of Large-Scale Applications on National Cyberinfrastructure**

Fabio Silva¹, Christopher Brooks⁵, Ewa Deelman³, Monte Goode¹, Dan Gunter¹, Gideon Juve³, Gaurang Mehta³, Priscilla Moraes⁴, Taghrid Samak¹, Martin Swany⁴, Prasanth Thomas³ and Karan Vahi³

¹Lawrence Berkeley National Laboratory, ²University of Southern California, ³University of Southern California Information Sciences Institute, ⁴University of Delaware, Newark, ⁵University of San Francisco

---

**Log Collection Tools**
- *monitord* parses data in real-time
- Collects workflow execution logs

**NetLogger Tools**
- *nl-loader* stores data in a database, such as SQLite, MySQL, etc.
- Data goes to a broker, where it can then be sent to many subscribers

**STAMPEDE database schema**
- Represents both the abstract workflow plan and running workflow, including the associations between the two
- Handles parent and child workflows
- Stores data at high granularity

---

**Large-Scale Workflows**
- Composed of thousands to millions of coordinated tasks
- Executed in complex distributed environments
- Difficult to track failures, search through thousands of files

**Failure Analysis Algorithms**
- Machine learning algorithms predict workflow failures based on behavior patterns
- Hard failures can be easily determined using database queries
- Soft failures are often stochastic, should be detected early for quick error recovery

**End-user Tools**
- **Stampede Python Analysis API**: Simple, uniform access to back-end database
- **Stampede-analyzer**: Quickly debug a workflow after execution is completed
- **Stampede-statistics**: Generates statistics about a running or finished workflow
  - Number of tasks/jobs/sub workflows ran/succeeded/failed/retried, ...
  - Job execution site, scheduler queuing time, execution delay, ...
- **Stampede-plots**: Interactive graphs and charts for workflow visualization

---

**Plan and Execute**

**Transform and Archive**

**Real-time Analysis**

**Scientific Workflow**

**Stampede DB**

**PeriScope**
- Provides end-to-end system performance view to users
- Presents data from processes, hosts, and network elements using a scalable analysis and presentation framework

---

STAMPEDE is funded by the National Science Foundation under the OCI program grant #0943705