ModSim Challenges for Experimental/ Observation Data Workflows

Ewa Deelman, Ph.D

Science Automation Technologies Group

USC Information Sciences Institute

Funding from DOE, NSF, and NIH

LIGO Experiment: Searching for Gravitational Waves

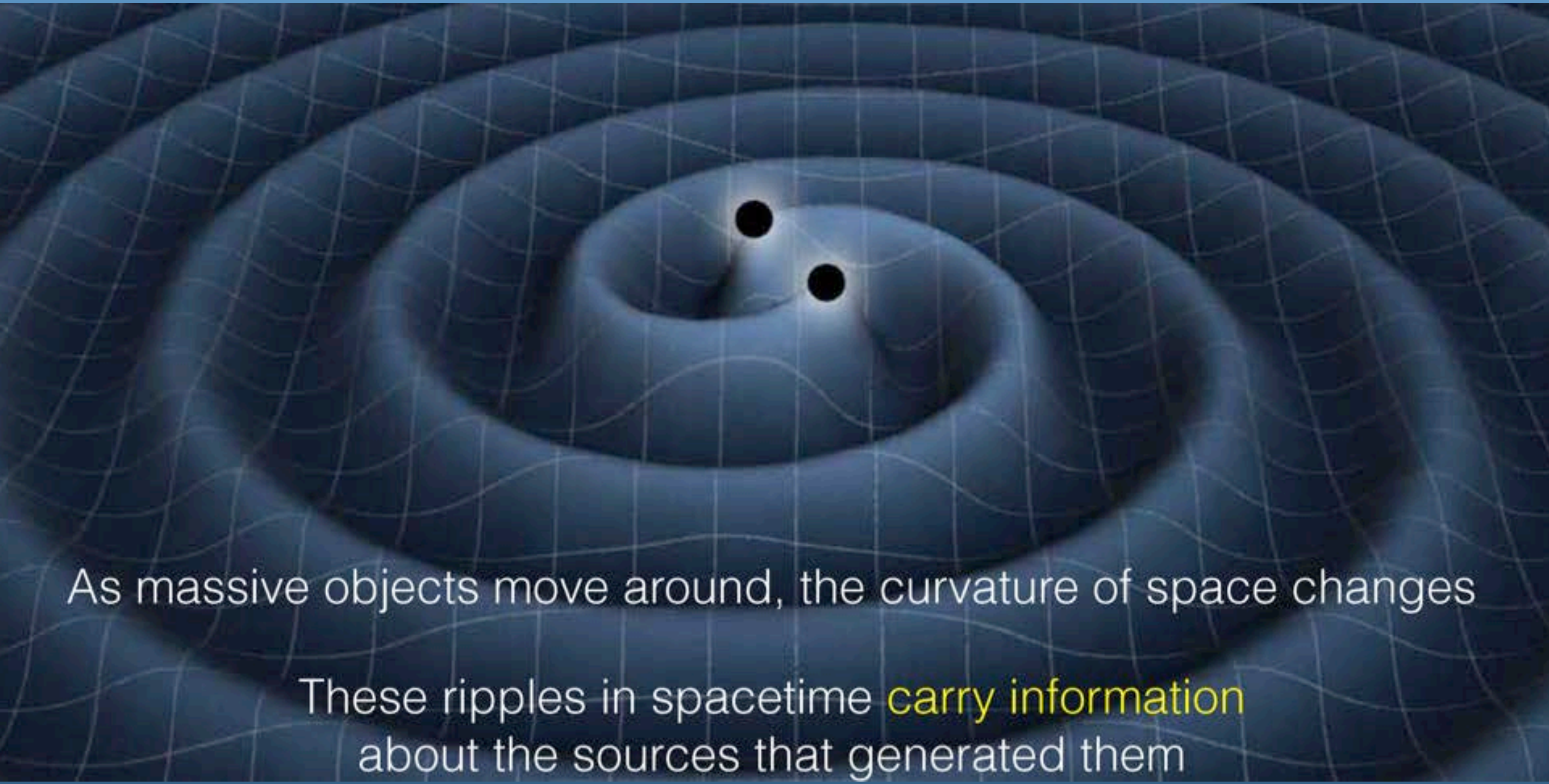


Image courtesy of LSC

LIGO (Laser Interferometer Gravitational-Wave Observatory)

- **LSC (LIGO Scientific Collaboration)**
  - Collaboration involved in research of the data coming out of the detectors.
  - 1000 scientists from universities in US and 14 other countries
  - 250 students
  - Responsible for developing analysis methodologies and detector technology.

- **Background**
  - Largest ever NSF funded project
  - Two 4km long detectors in the US (Hanford, Washington, and Livingston, Louisiana)
  - **Phase I (Initial LIGO 2002 – 2010)**
    - No gravitational waves detected.
    - But a lot of analysis pipelines and computing infrastructure
    - **Late 2010 - Passed Blind Injection Test**
  - **Upgrade of the detectors** (Designed to be 10 times more sensitive than Phase I)
  - **Phase II (Advanced LIGO September 2015 onwards**
    - Currently operating at 4 times the Initial LIGO sensitivity

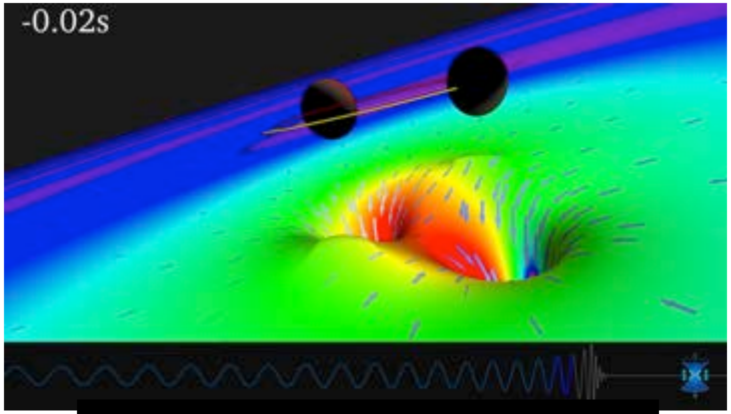
Aerial View of the LIGO Livingston Laboratory
Image Credit: Caltech/MIT/LIGO Lab

LIGO Engineering from www.ligo.caltech.edu/page/facts

- “ ***Most sensitive:*** LIGO is designed to detect a change in distance between its mirrors*1/10,000th the width of a proton!* Equivalent to measuring the distance to the nearest star to an accuracy smaller than the width of a human hair!
- ***World's second-largest vacuum chambers:*** Encapsulating 10,000 m3 (350,000 ft3), each vacuum chamber encloses as much volume as 11 Boeing 747-400 commercial airliners. The air removed from *each* of LIGO’s vacuum chambers could inflate two-and-a-half MILLION footballs, or 1.8 million soccer balls! LIGO's vacuum volume is surpassed only by the LHC
- ***Ultra-high vacuum:*** The pressure inside LIGO's vacuum tubes is *one-trillionth* of an 'atmosphere' (in scientific terms, that’s 10-9 torr). It took 40 days (1100 hours) to remove all 10,000 m3 (353,000 ft3) of air and other residual gases from each of LIGO’s vacuum tubes to reach an air pressure one-trillionth that at sea level.
- ***Curvature of the Earth***: LIGO’s arms are so long that the curvature of the Earth is a measurable 1 meter (vertical) over the 4 km length of each arm. The most precise concrete pouring and leveling imaginable was required to counteract this curvature and ensure that LIGO’s vacuum chambers were "flat" and level. Without this work, LIGO's lasers would hit the end of each arm 1 m above the mirrors it is supposed to bounce off of!”

LIGO's Gravitational Wave Detection

- **LIGO announced first ever detection of gravitational waves (Feb 2016)**
  - Created as a result of coalescence of a pair of dense, massive black holes.
  - Confirms major prediction of Einstein Theory of Relativity



- **Detection Event**
  - Detected by both of the operational Advanced LIGO detectors (4km long L shaped interferometers)
  - Event occurred at September 14, 2015 at 5:51 a.m. Eastern Daylight Time

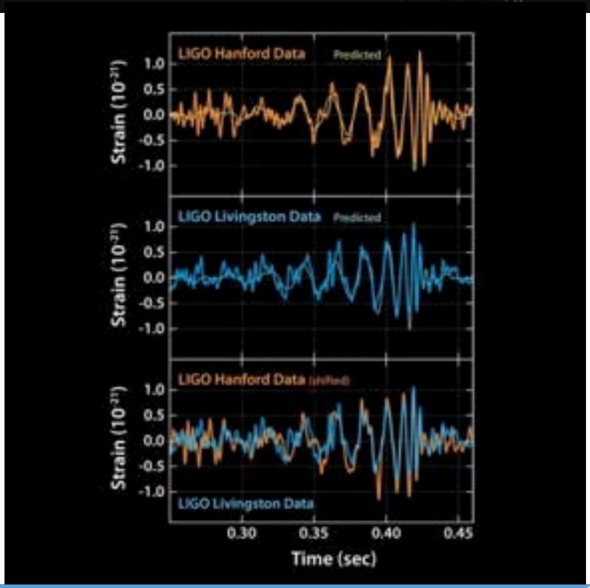


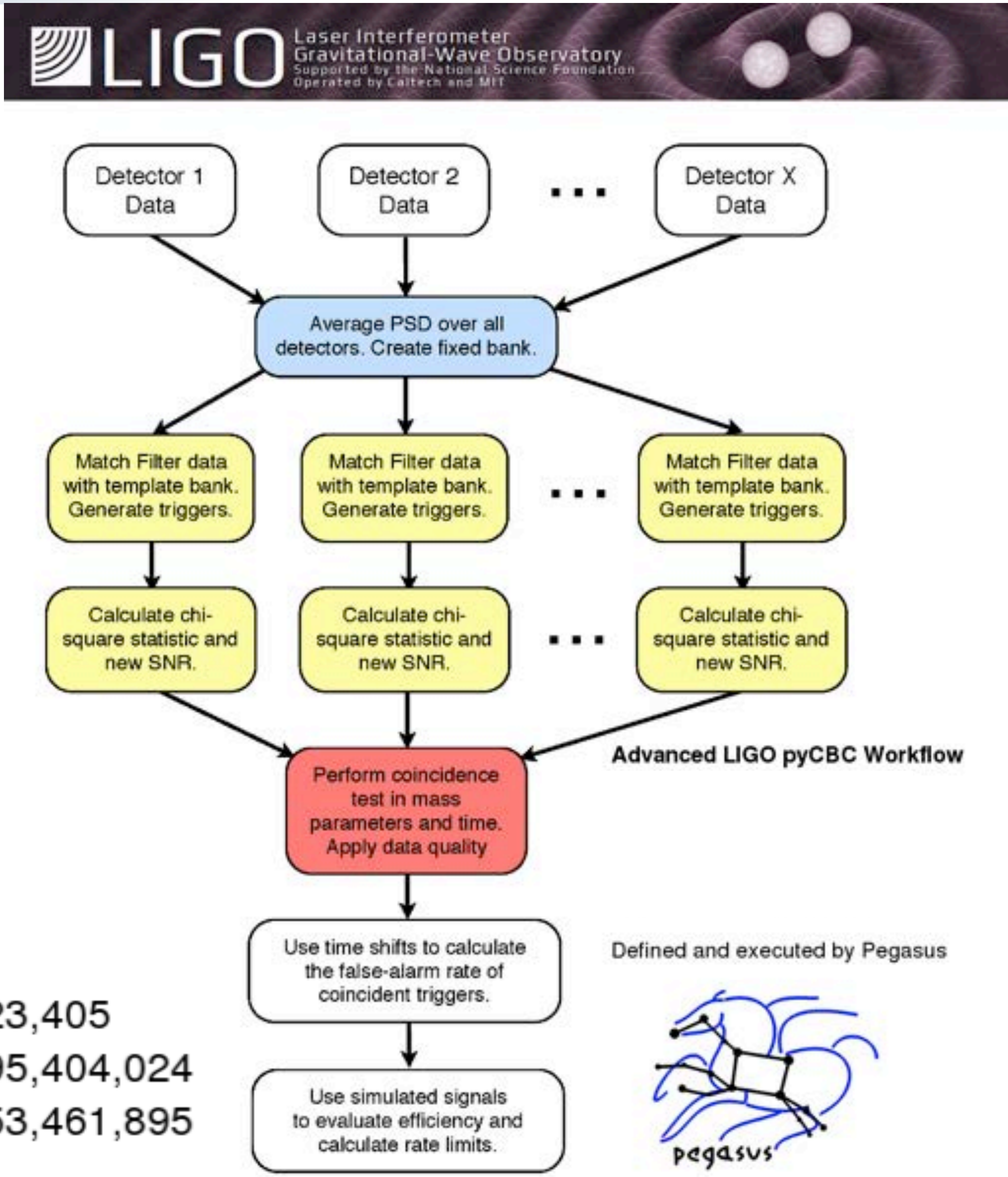
Image Credits:
0.2 Second before the black holes collide: SXS/LIGO
Signals of Gravitational Waves Detected: Caltech/MIT/LIGO Lab

LIGO Detection – Behind the Scenes

- **A variety of complex analysis pipelines were executed.**
- **Some were low latency that initially alerted people to look at a specific piece of data containing the signal.**
- **However, to verify that signal is a valid candidate,**
  - A large amount of data needed to be analyzed.
  - Statistical significance of the detection should be at 5-sigma level
- **Pipelines executed on LSC Data Grid, OSG, and XSEDE**
  - Consists of approximately 11 large clusters at various LIGO institutions and affiliates
  - Data is replicated at sites in the US and Europe
  - Each LIGO cluster has Grid middleware and HTCondor installed.
  - GridFTP used for data transfers.
- **Pipelines are modeled as scientific workflows**

Advanced LIGO PyCBC Workflow

- One of the main pipelines to measure the statistical significance of data needed for discovery.
- Contains 100's of thousands of jobs and accesses on order of terabytes of data.
- Uses data from multiple detectors.
- For the detection, the pipeline was executed on Syracuse and Albert Einstein Institute Hannover
- Use our Pegasus software to automate the execution of tasks and data access



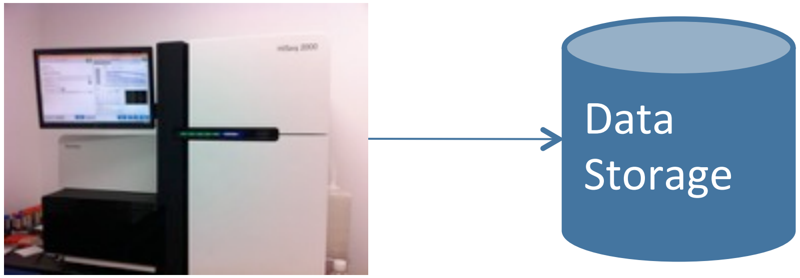
Workflows:	23,405
Tasks:	95,404,024
Jobs:	53,461,895

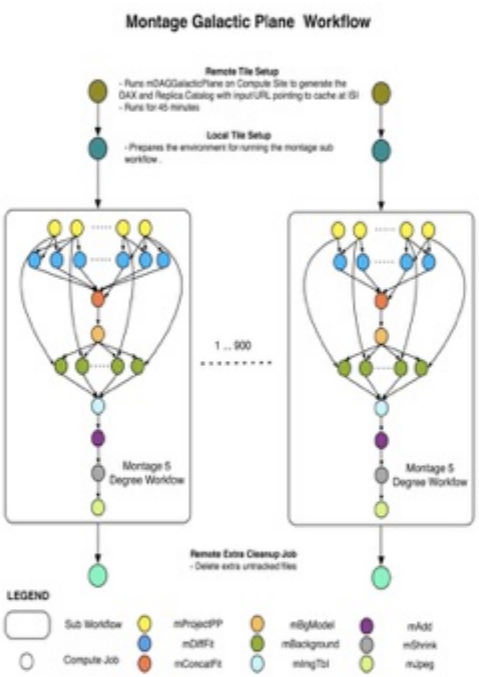
***Image Credit**: Samantha Usman, Duncan Brown et al*

Outline

- Example Pegasus Workflows
- Pegasus Workflow Management System
- ModSim Challenges
- Research directions

Sometimes the environment is complex





Work definition

Local Resource

Data Repositories

Campus Cluster

XSEDE

NERSC

ALCF

OLCF

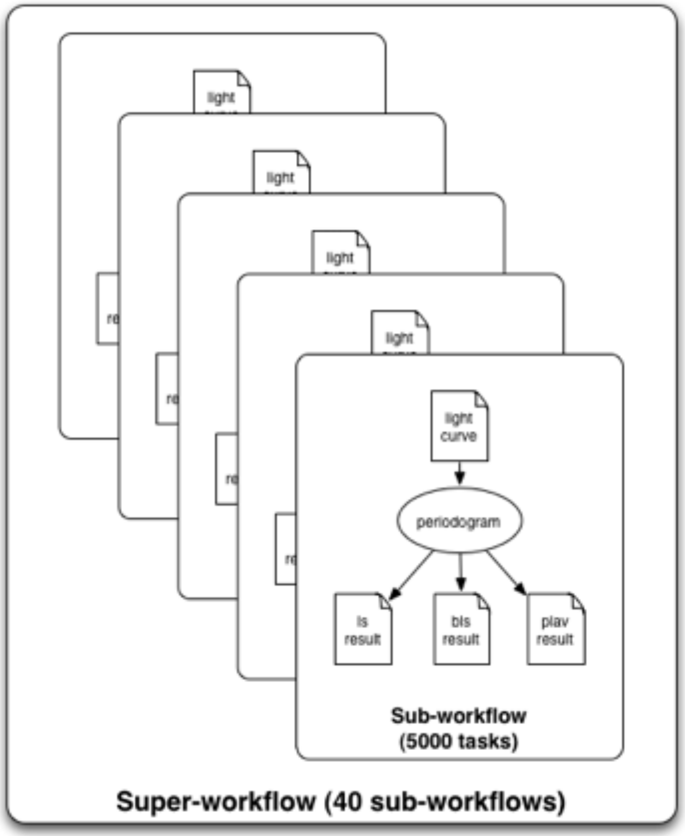
Open Science Grid

Chameleon

Amazon Cloud

Sometimes the environment is just not exactly right

Single core workload



Cray XK7 System Environment / Designed for MPI codes

Sometime you want to change or combine resources

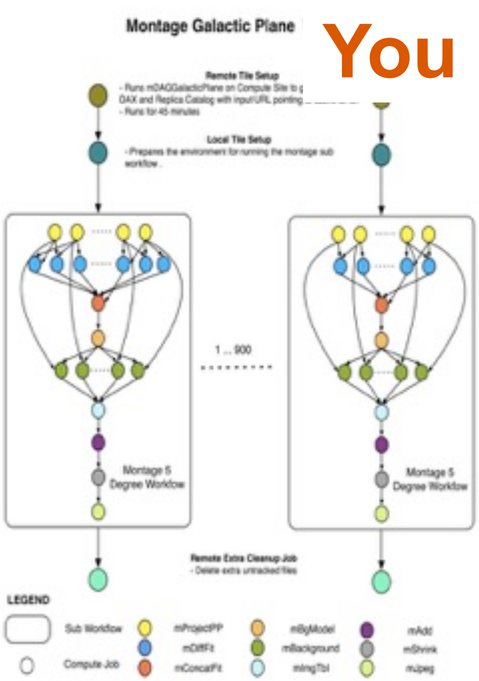
Data Storage

data

Data Repositories

Campus Cluster

XSEDE



You don't want to recode your workflow

ALCF

OLCF

work

Open Science Grid

FutureGrid

Local Resource

Amazon Cloud

Our Approach: Submit locally, Compute globally

Data Storage

data

Campus Cluster

XSEDE

NERSC

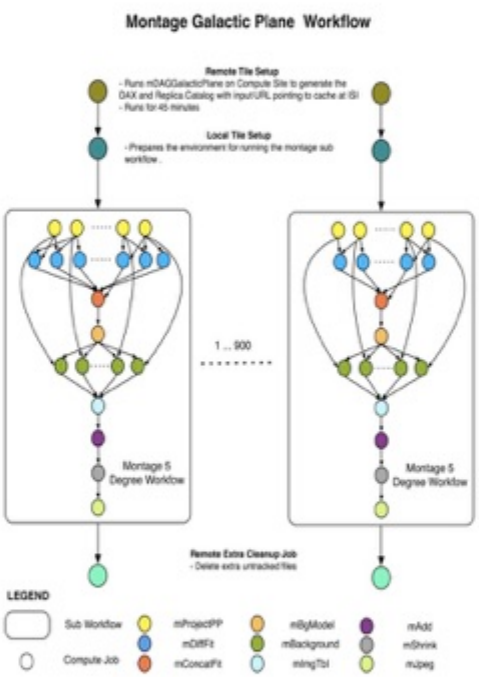
ALCF

OLCF

Open Science Grid

Chameleon

Amazon Cloud



Work definition

Workflow Management System

work

Local Resource

Workflow Management System (WMS) Functions

- Discover what resources (computation, data, software) are available
- Select the appropriate resources based on a architecture, availability of software, performance, reliability, availability of cycles, storage,..
- Devise a plan:
  - What resources to use
  - How to best adapt the workflow to the resources
  - What protocols to use to access the data, to schedule jobs
  - What data to save
- Execute the plan
  - In a reliable way
  - Keep track of what data was accessed, generated and how
- Outside of the WMS functions
  - Resource provisioning

Pegasus Workflow Management System (est. 2001)

Collaboration with HTCondor, UW Madison

- A workflow “compiler”/planner
  - Input: abstract workflow description, resource-independent
  - Auxiliary Info (catalogs): available resources, data, codes
  - Output: executable workflow with concrete resources
  - Automatically locates physical locations for both workflow tasks and data
  - Transforms the workflow for performance and reliability
- A workflow engine (DAGMan)
  - Executes the workflow on local or distributed resources (HPC, clouds)
  - Task executables are wrapped with *pegasus-kickstart* and managed by Condor *schedd*
- Provenance and execution traces are collected and stored
- Traces and DB can be mined for performance and overhead information

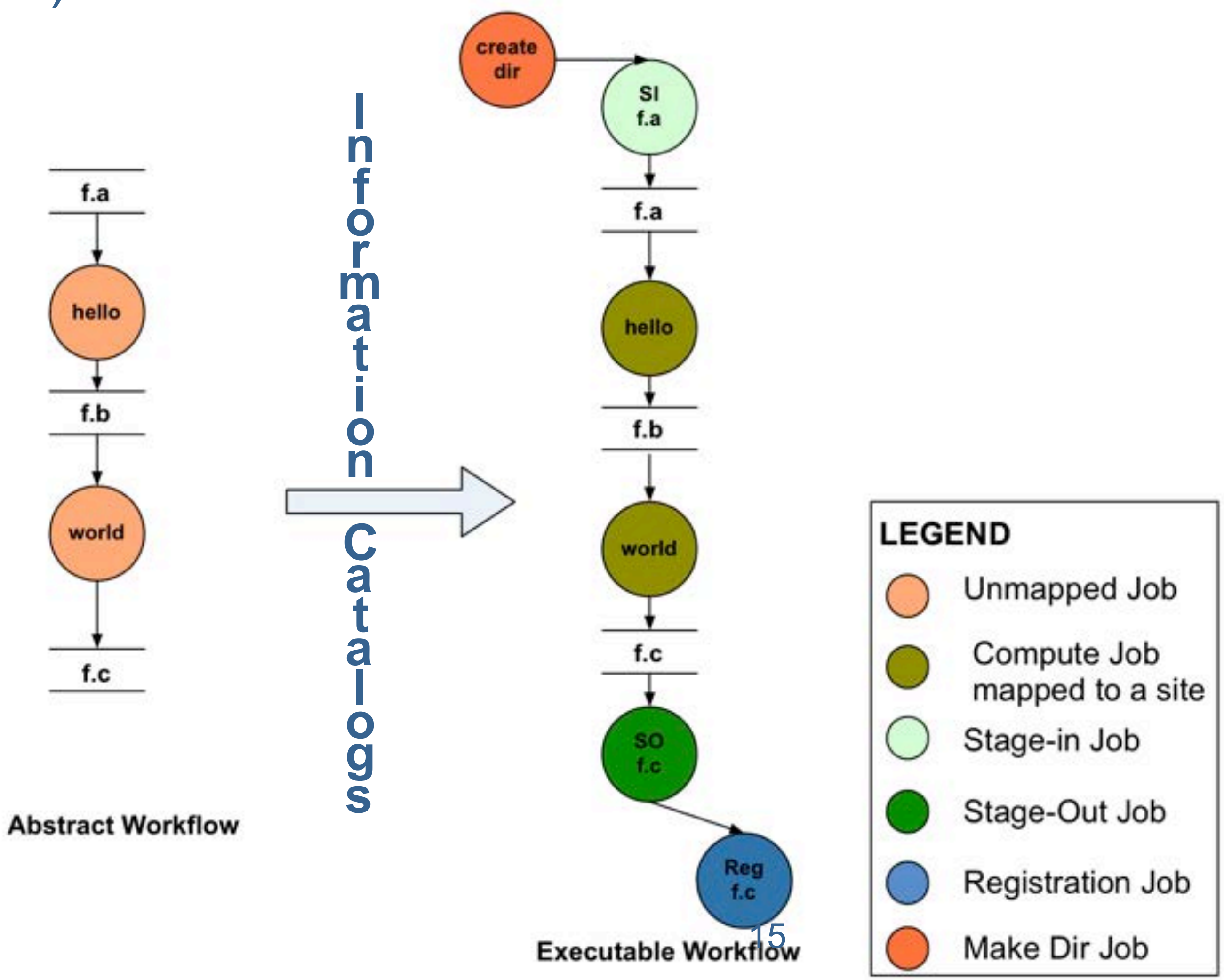
Generating executable workflows

(DAX)

APIs for workflow specification (DAX--- DAG in XML)

Java, Perl, Python

R - prototype



Pegasus Dashboard

Workflow Wall Time	5 days 1 hour
Workflow Cumulative Job Wall Time	2065 days 10 hours
Cumulative Job Walltime as seen from Submit Side	2068 days 23 hours
Workflow Cumulative Badput Time	58 mins 32 secs
Cumulative Job Badput Walltime as seen from Submit Side	1 hour 32 secs
Workflow Retries	5

Workflow Statistics

Job Breakdown Statistics

Workflow Statistics

Show 10 entries   Search:

Transformation	Count	Succeeded	Failed	Min	Max	Mean	Total
dagman::post	15301	14819	482	5	554	9.607	146993
inspiral-FULL_DATA-H1_ID9	7621	7620	1	901.357	20055.080	12507.034	95316108.773
inspiral-FULL_DATA-L1_ID10	6641	6640	1	1589.662	19588.902	12504.586	83042955.049
pegasus::transfer	108	108	0	0	205.304	9.664	1043.731
coinc-FULL_DATA_FULL-H1L1_ID14	20	20	0	263.256	348.672	297.379	5947.588
pegasus::dirmanager	7	7	0	0	5	2.857	20
condor::dagman	6	6	0	607	833	700.167	4201
dagman::pre	6	6	0	11	75	27.167	163
single_template-P1_0-H1_ID5	5	5	0	360.602	383.866	373.194	1865.968
single_template_plot-P1_0-H1_ID6	5	5	0	4.013	8.382	5.008	25.041

Showing 1 to 10 of 314 entries   First Previous 1 2 3 4 5 … 32 Next Last

Job Statistics

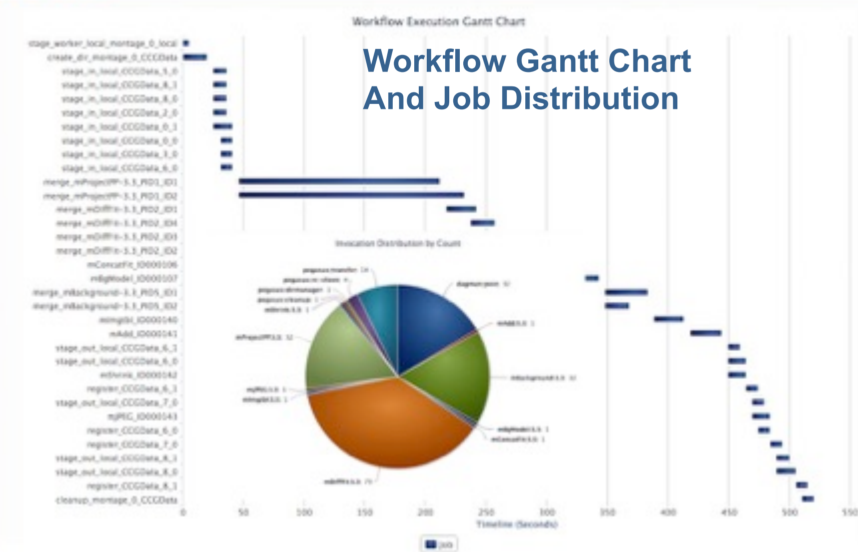
Charts

Job Distribution

Time Chart

Gantt Chart

Workflow Gantt Chart And Job Distribution



Workflow Listing Page Shows Successful, Failed and Running Workflows

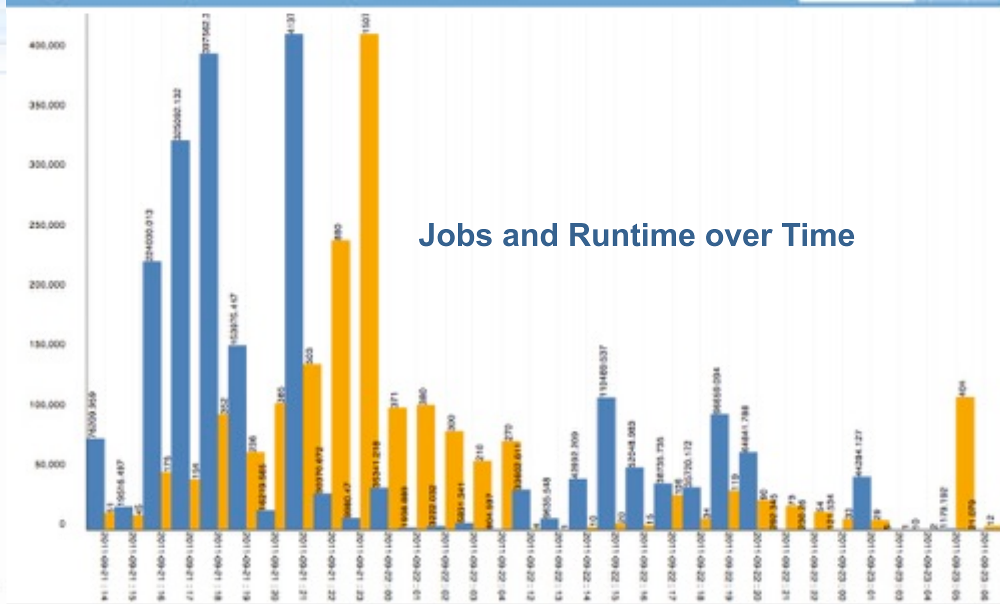
how results for all

Show 10 entries   Search: injections

Workflow Label	Submit Host	Submit Directory	State	Submitted On
analysis2-C01-injections	sugar-dev2.phy.syr.edu	/usr1/amber.lenon/pycbc-tmp.4a07mk2LXe/work	Successful	Sat, 06 Feb 2016 13:27:15
analysis8-C01-injections	sugar-dev2.phy.syr.edu	/usr1/amber.lenon/pycbc-tmp.cqBQirspEI/work	Successful	Mon, 08 Feb 2016 15:25:05
analysis7-C01-injections	sugar-dev2.phy.syr.edu	/usr1/amber.lenon/pycbc-tmp.tpHketeH7X/work	Successful	Mon, 08 Feb 2016 11:45:22
analysis3-C01-injections	sugar-dev2.phy.syr.edu	/usr1/amber.lenon/pycbc-tmp.LU6iToRyVA/work	Running	Tue, 23 Feb 2016 16:27:30
analysis4-C01-injections	sugar-dev2.phy.syr.edu	/usr1/amber.lenon/pycbc-tmp.gArV7UNU9C/work	Running	Tue, 23 Feb 2016 16:27:44
analysis5-C01-injections	sugar-dev2.phy.syr.edu	/usr1/amber.lenon/pycbc-tmp.d95VFhwkKu/work	Running	Wed, 24 Feb 2016 11:49:17
analysis6-C01-injections	sugar-dev2.phy.syr.edu	/usr1/amber.lenon/pycbc-tmp.O1anUn5hGE/work	Running	Wed, 24 Feb 2016 11:55:55
analysis9-C01-injections	sugar-dev2.phy.syr.edu	/usr1/amber.lenon/pycbc-tmp.Lf2hB7UTuG/work	Running	Wed, 24 Feb 2016 12:07:11

Showing 11 to 18 of 18 entries (filtered from 33 total entries)   First Previous 1 2 Next Last

Jobs and Runtime over Time



Tools to calculate job statistics

Workflow makespan, Cumulative time

Task Type	Count	Runtime(s)	IO Read (MB)	IO Write (MB)	Memory Peak(MB)	CPU Utilization(%)
mProjectPP	2102	1.73	2.05	8.09	11.81	86.96
mDiffFit	6172	0.66	16.56	0.64	5.76	28.39
mConcatFit	1	143.26	1.95	1.22	8.13	53.17
mBgModel	1	384.49	1.56	0.10	13.64	99.89
mBackground	2102	1.72	8.36	8.09	16.19	8.46
mImgtbl	17	2.78	1.55	0.12	8.06	3.48
mAdd	17	282.37	1102	775.45	16.04	8.48
mShrink	16	66.10	412	0.49	4.62	2.30
mJPEG	1	0.64	25.33	0.39	3.96	77.14

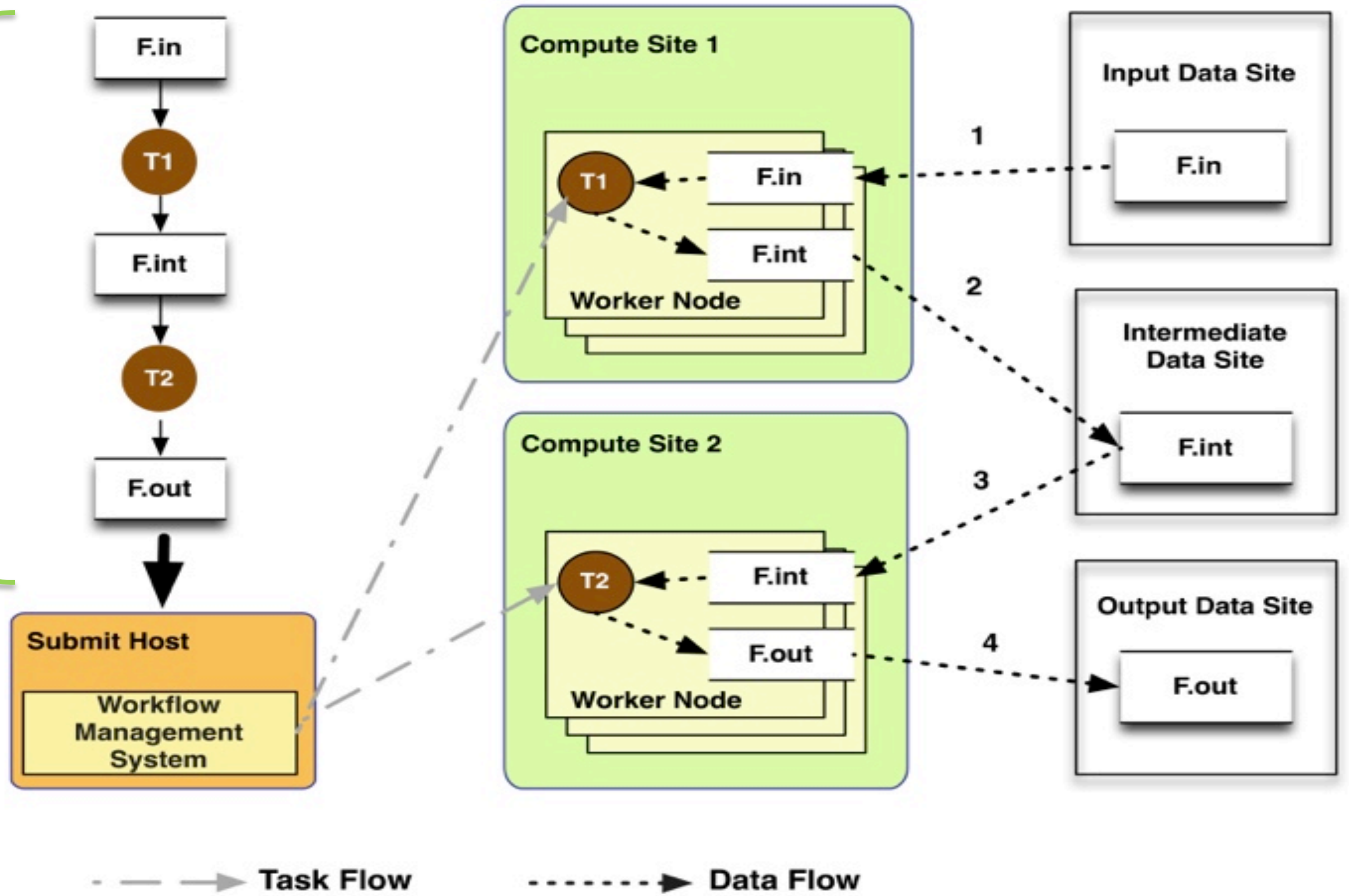
Execution profile of the Montage workflow, averages calculated

Outline

- Example Pegasus Workflows
- Pegasus Workflow Management System
- ModSim Challenges
- Research directions

Variety of file system deployments: need to model different storage systems and networks

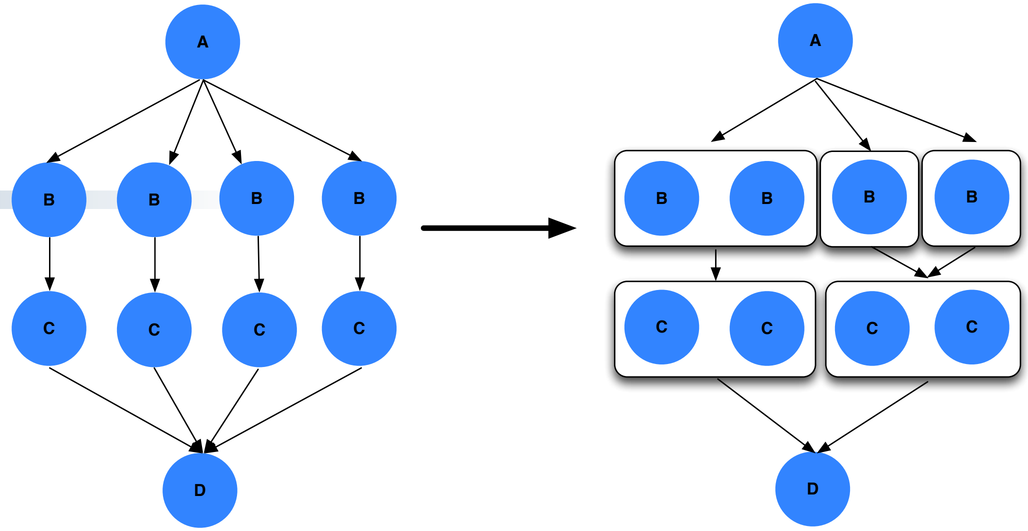
User workflow

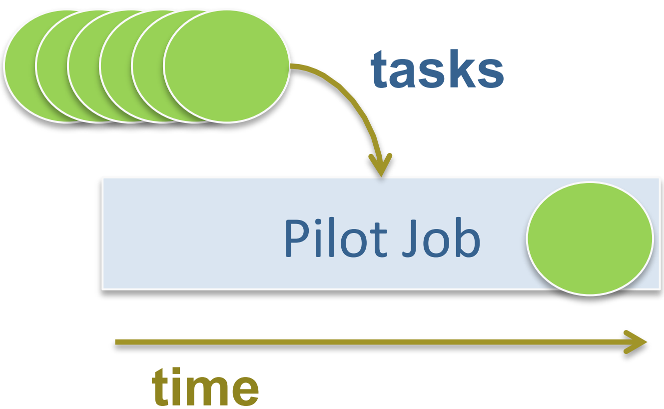


Allows LIGO to run easily on Open Science Grid, Clouds

Workload is overlaid on top the infrastructure

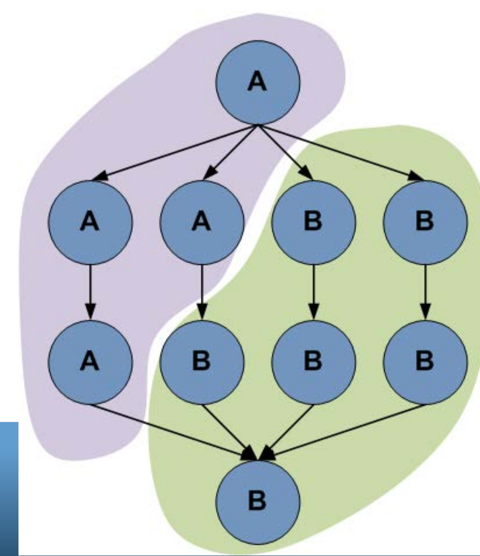
Cluster tasks





Use "pilot" jobs to dynamically provision a number of resources at a time

Partition the workflow into subworkflows and send them for execution to the target system



ModSim Workflow Challenges

- Modeling of wide area networks
- Quantifying scheduling delays at the resources
- Finding out and modeling resource usage policies
  - how many jobs you can have in a queue
- Failures at all levels of the system
- Very heterogeneous applications within workflows

ModSim Challenges

Panorama Project

- How to develop models that can predict the behavior of complex, data-intensive, scientific workflows executing on large-scale infrastructures?
- What monitoring information and analysis are needed for performance prediction and anomaly detection in scientific workflow execution?
- How to adapt the workflow execution and the infrastructure to achieve the potential performance predicted by the models?
- How to automate the modeling, monitoring, and adaption processes?

Ewa Deelman, Gideon Juve, Dariusz Krol, Rafael Ferreira da Silva (USC/ISI)
Anirban Mandal, Paul Ruth, Ilya Baldin (RENCI)
Jeffrey Vetter, Vickie Lynch, Ben Mayer, Jeremy Meredith, (ORNL)
Chris Carothers, Mark Blanco, Noah Wolfe (RPI), Brian Tierney (LBL)

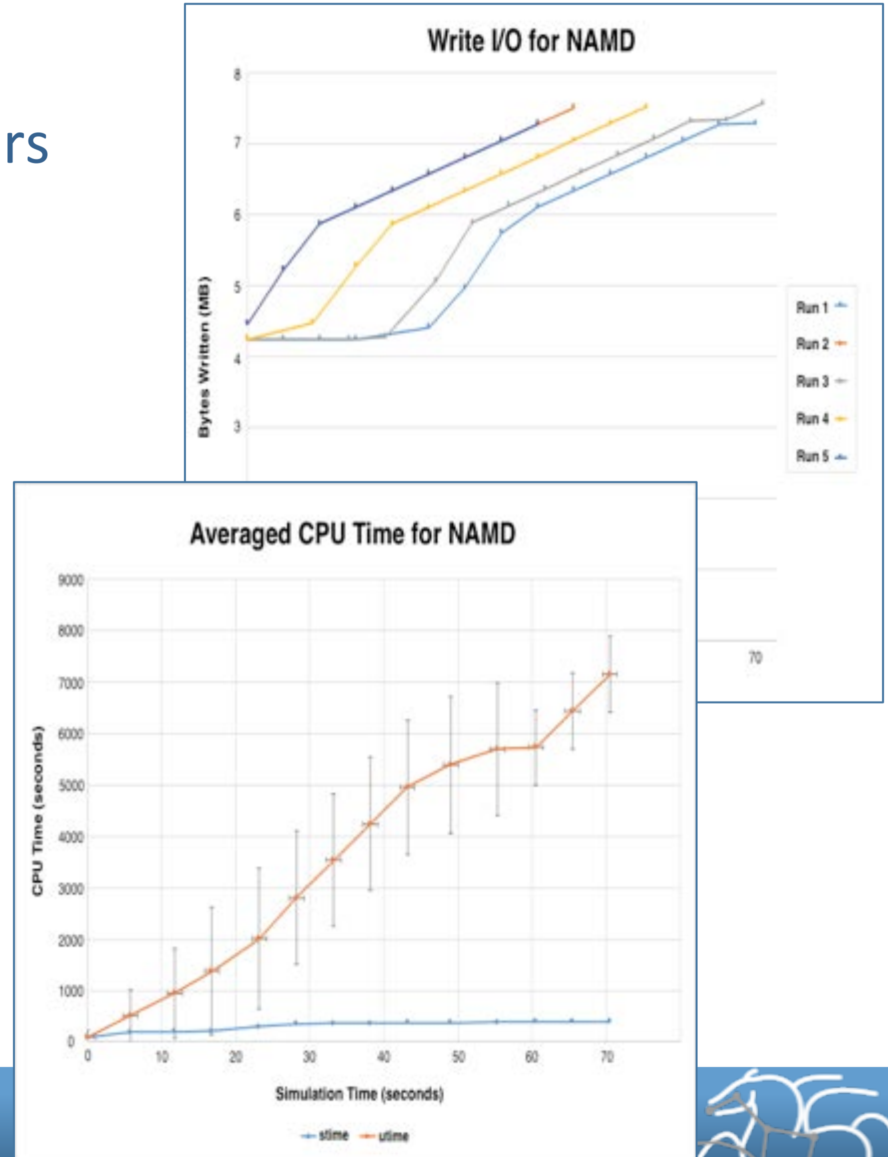
Application and Infrastructure Monitoring

Application Monitoring

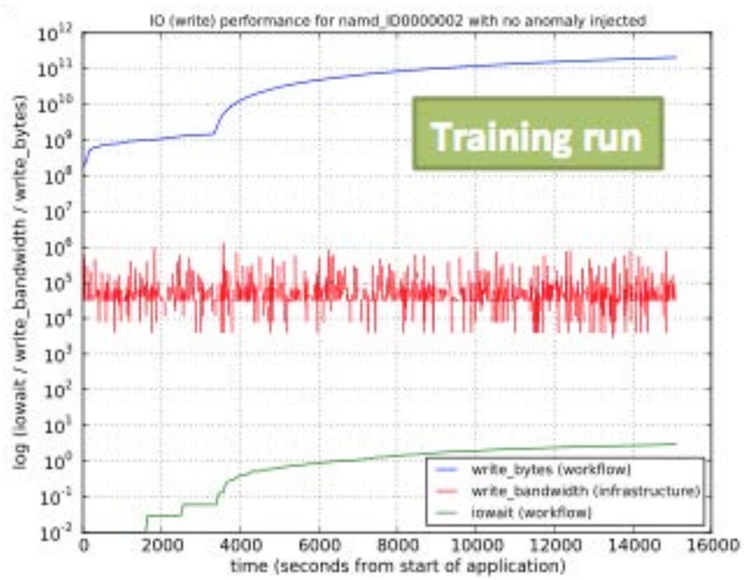
- CPU, I/O, memory, perf counters
- Function interposition
- MPI and serial jobs
- Real-time reporting

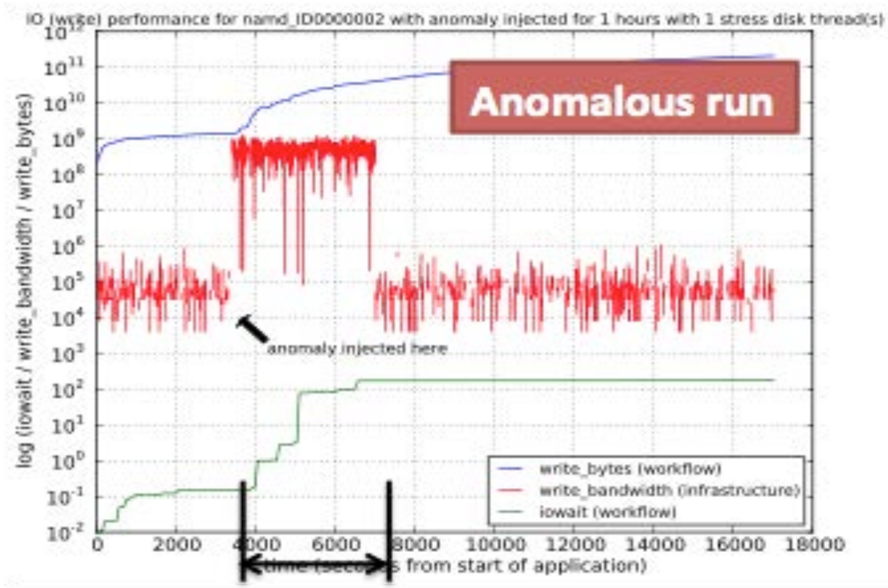
Infrastructure Monitoring

- Load, disk I/O, network, etc.
- Standard tools
- Data stored in time series DB

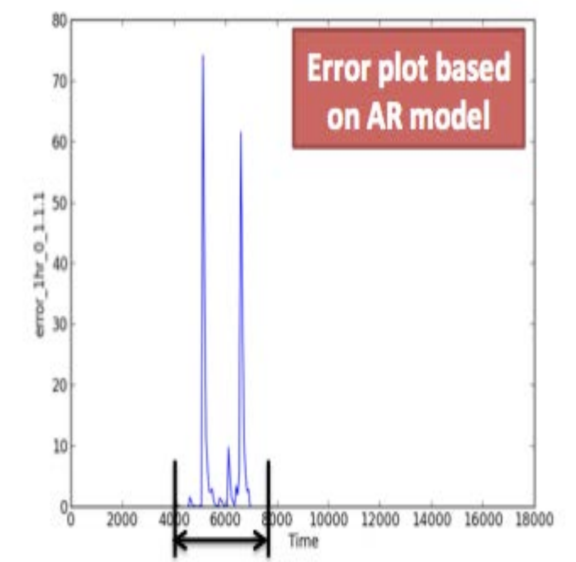


Anomaly Detection using AR(N) - iowait with NAMD





- AR coefficients calculated from training run - AR model
- Used AR model to predict for anomaly case
- Interval corresponding to maximum error (predicted vs. actual) overlapped with anomaly interval



Application trends

- **More data management**
  - Collecting and archiving data from sensors and instruments
- **More "live" processing**
  - Analysis of instrumental data on the fly
  - Coupling simulation and visualization/analytics
- **Applications that are composed of multiple workflows (ensembles)**
  - UQ applications

- ***We need realistic, representative workflows***

Additional Challenges for ModSim

- Workflows in HPC environments
  - Support for in-situ processing
  - Different checkpoint mechanisms
  - Use of novel architectural components (NVM)

  - We need models for energy consumption, so we can make data management decisions within HPC
- Workflows in Virtual Environments
  - Clouds
  - Software defined infrastructures (SDX– Anirban Mandal's poster)
- Sometimes resource provisioning is part of the workflow
- In general we need models and simulations at different scales

A different role for ModSim

"Research is required to develop the science of workflows to fully understand how workflows behave. **Did the workflow behave as expected? Did the infrastructure** (computer, instrument, network, storage) **behave as expected**? **Can the data or metadata be trusted**? Is the experiment repeatable?"

http://science.energy.gov/~/media/ascr/pdf/programdocuments/docs/workflows_final_report.pdf