USC Viterbi School of Engineering
Information Sciences Institute

Science Impact of Sustained Cyberinfrastructure: The Pegasus Example

Ewa Deelman, Ph.D.

University of Southern California,

Information Sciences Institute

NSF Office of Advanced CyberInfrastructure Webinar

May 17, 2018

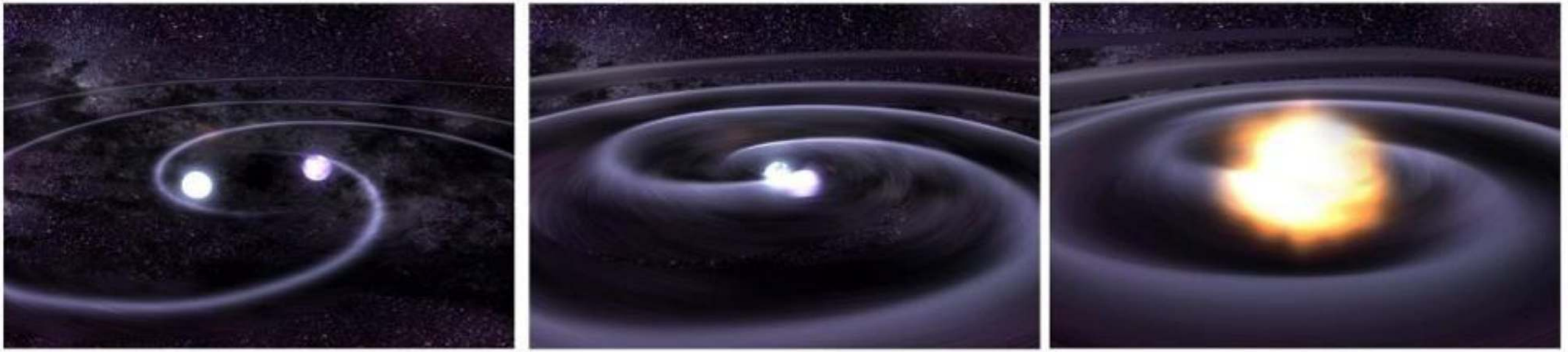
Pegasus represents a long standing collaboration with Miron Livny, University of Wisconsin, Madison

HTCondor
High Throughput Computing

October 16th 2017: " LIGO and Virgo make first detection of gravitational waves produced by colliding neutron stars"

USC Viterbi School of Engineering Information Sciences Institute

And kick off a new era of multi-messenger astronomy

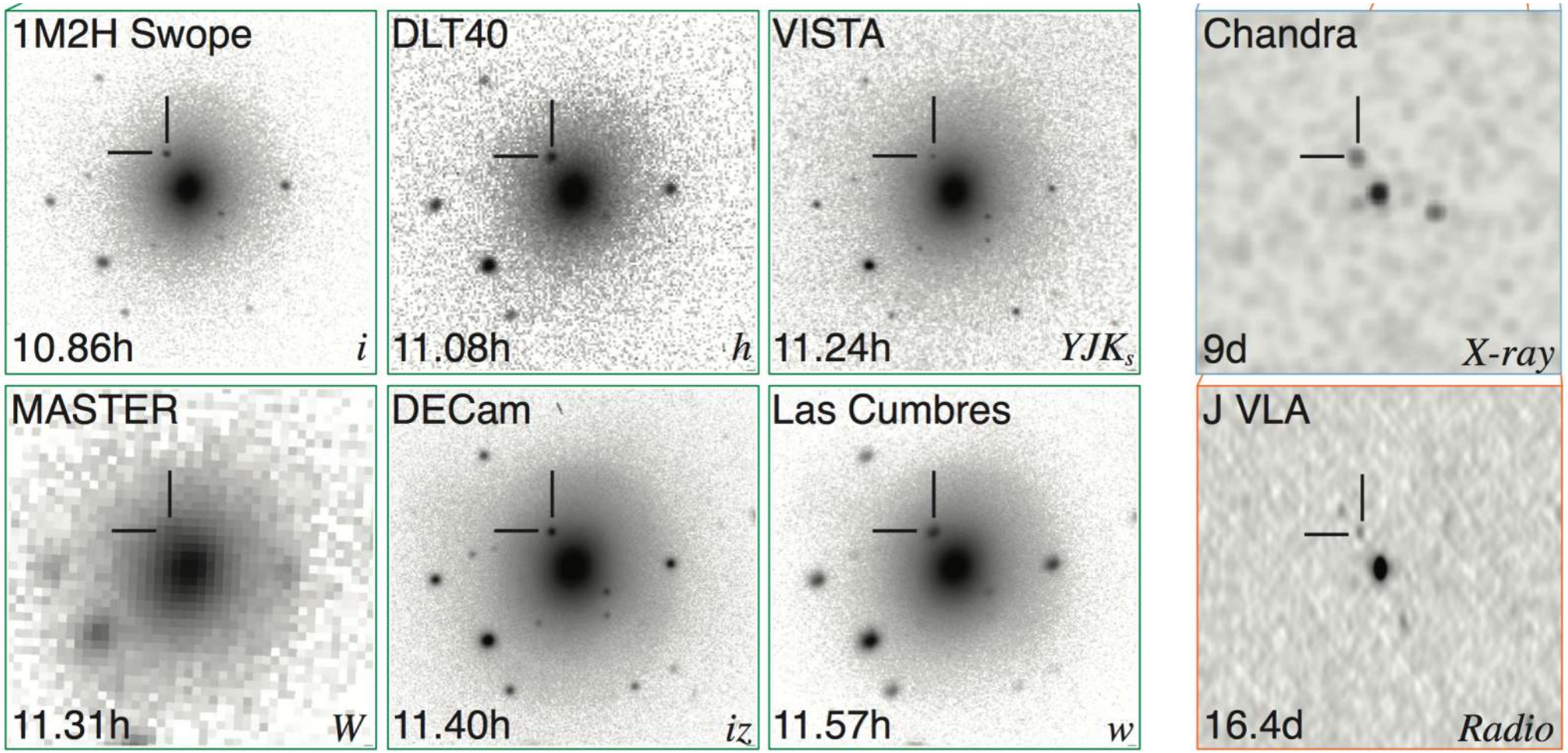


"The inspiral and merger of two neutron stars, as illustrated here, should produce a very specific gravitational wave signal, but the moment of the merger should also produce electromagnetic radiation that's unique and identifiable as such.", credit LIGO

NASA's Fermi space telescope had detected a burst of gamma rays at about the same time

Images credit: LIGO Scientific Collaboration

Targeting Telescopes on the Neutron Star Merger



"aftermath of the BNS merger.. On the left are six optical images taken between 10 and 12 hours after the merger by different telescopes. On the right are images constructed from x-ray and radio observations. The x-ray image was taken 9 days after the merger by NASA's Chandra X-ray Observatory. 16 days after the merger NRAO's Jansky Very Large Array (VLA) captured the radio image" from LIGO.org

Images credit: LIGO Scientific Collaboration

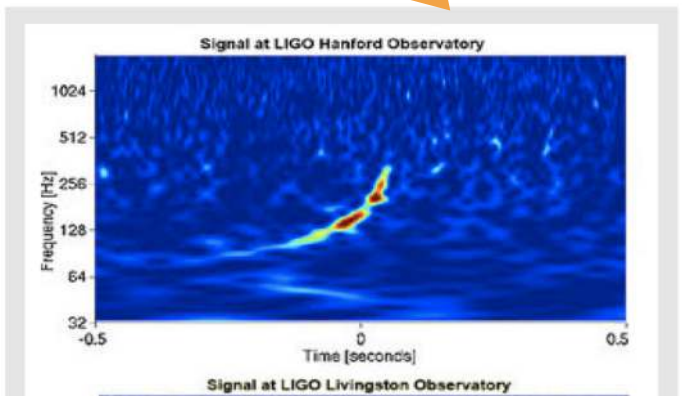
Dependable Cyberinfrastructure

Takes time, the Pegasus and LIGO Partnership

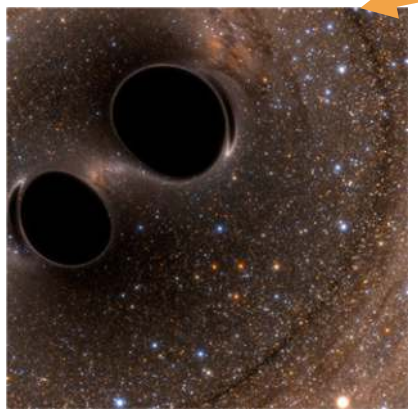
Nobel Prize

2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017

First Pegasus prototype



Blind injection detection



First detection of a GW from black hole collision

Multi-messenger neutron star merger observation

Image credit: LIGO Scientific Collaboration

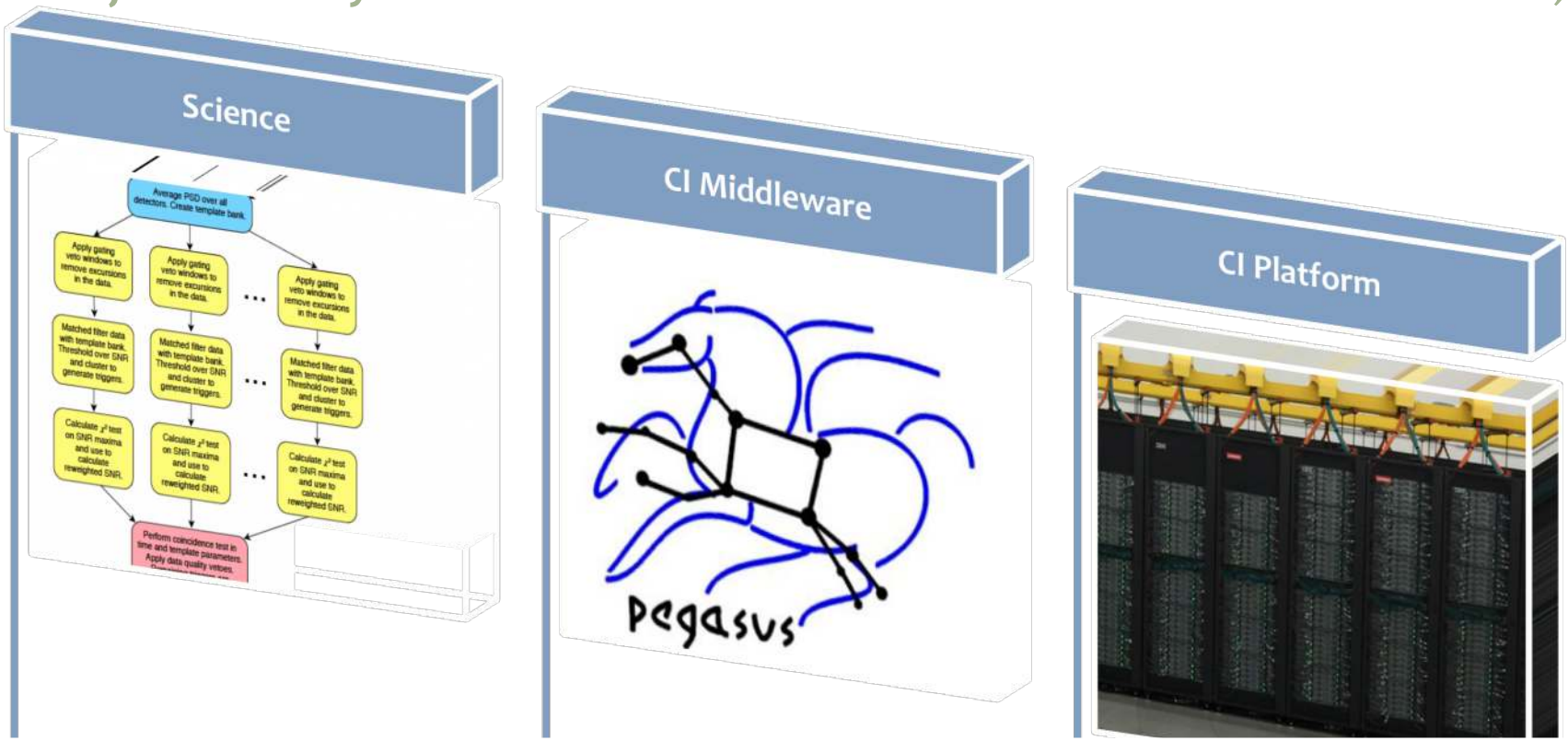
First GW detection: Pegasus automated ~ 21K workflows with ~ 107M tasks

USC Viterbi School of Engineering Information Sciences Institute

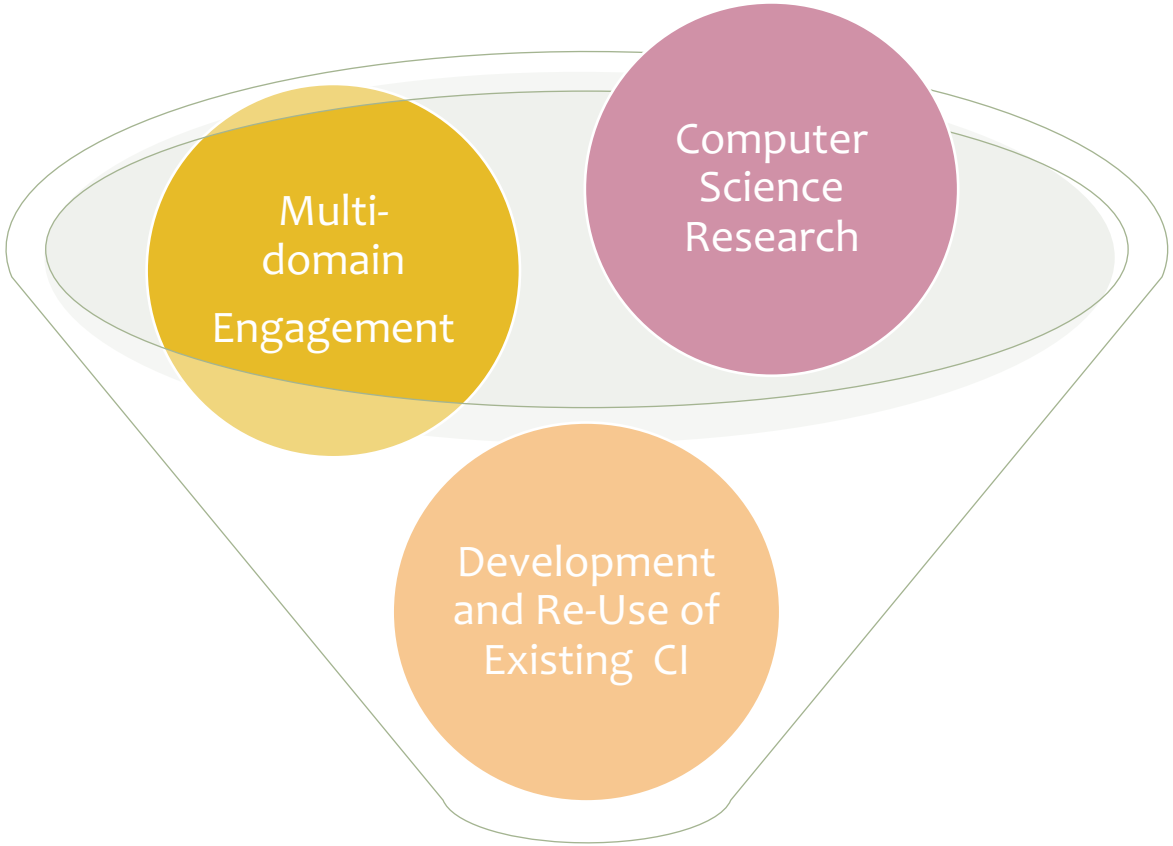
Science workflow:
measure the statistical significance of data needed for discovery

Automated by Pegasus
execution of tasks and data access

Distributed Power
LIGO, Open Science Grid, XSEDE, Blue Waters



What does it take to build and sustain Cyberinfrastructure?

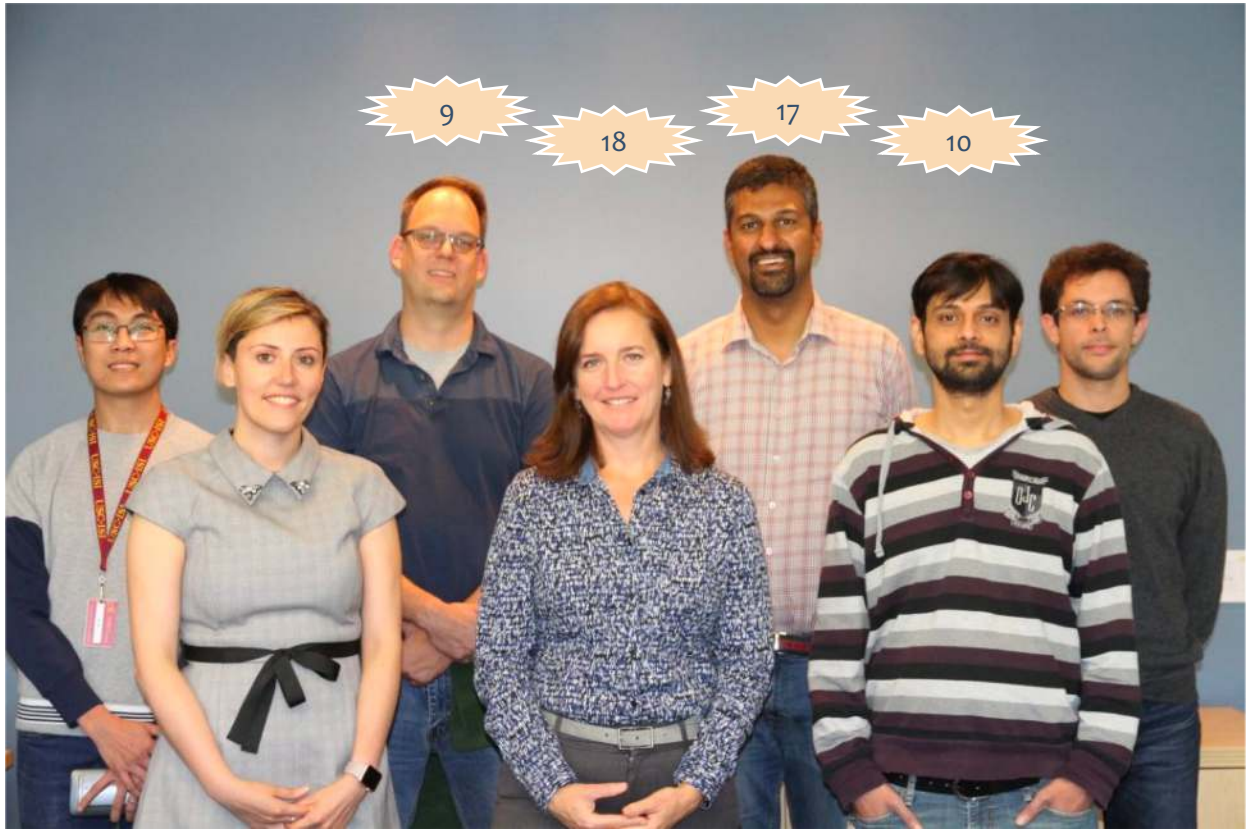


The Pegasus lesson- One needs a holistic approach to build dependable CI!

Dependable CI

Takes time to build a team and expertise

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018



5

Back Row: Tu Mai Anh Do, Mats Rynge, Karan Vahi, George Papadimitriou
Front Row: Rosa Filgueira, Ewa Deelman, Rajiv Mayani
Missing: Rafael Ferreira da Silva, Ashwin Venkatesha

Takes Contributions from Many People

Developers

GRAs

PostDocs

Master Students and Visitors

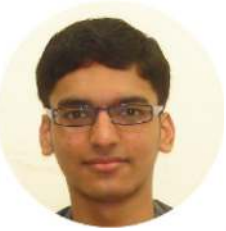
















Currently at Amazon, Google, NetApps, SpaceX, Samsung, startups

And others

Takes Collaboration with Many CS and Domain Scientists

CS research

How did Pegasus Start?

Extend the concept of view materialization in DBs to distributed environments

The Virtual Data Grid (VDG) Model

- Data suppliers publish data to the Grid
- Users request raw or derived data from Grid, without needing to know
  - Where data is located
  - Whether data is stored or computed

NSF ITR: GriPhyN Project: Ian Foster (PI), Paul Avery, Carl Kesselman, Miron Livny, (co-PIs)

How do you translate the Computer Science idea to the needs of science?

Circa. 2001

Virtual Data Scenario

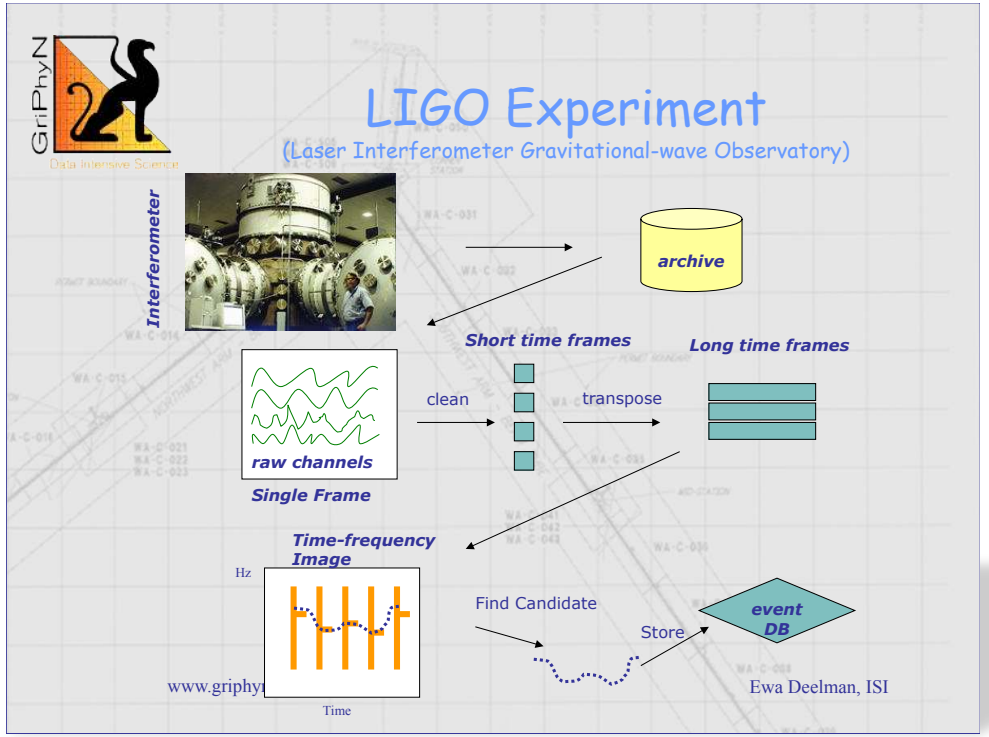
- (LIGO) "Conduct a pulsar search on the data collected from Oct 16 2000 to Jan 1 2001"
- For each requested data value, need to
  - Understand the request
  - Determine if it is instantiated; if so, where; if not, how to compute it
  - Plan data movements and computations required to obtain all results
  - Execute this plan

CS research

USC Viterbi School of Engineering Information Sciences Institute

Challenge: How Translate a Science Request to an Actionable Plan?

Explore AI planning techniques



Lost in translation: high-level abstraction for this science domain

Found: new research direction: management of workflows in distributed environments

CS research

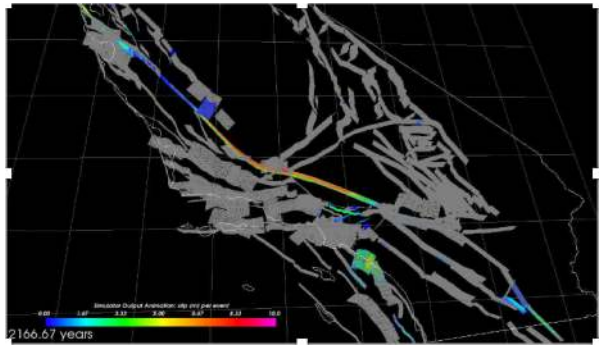
USC Viterbi School of Engineering Information Sciences Institute

Challenges of Workflow Management

- Working with LIGO and other applications (astronomy, earthquake science), found common challenges:
  - Need to describe complex workflows in a simple way
  - Need to access distributed, heterogeneous data and resources
  - Need to deal with resources/software that change over time

Sky mosaic, IPAC, Caltech

- Our focus:
  - Separation between workflow description and workflow execution
  - Workflow planning and scheduling (scalability, performance)
  - Task execution (monitoring, fault tolerance, debugging)



Earthquake simulation, SCEC, USC

Multi-domain

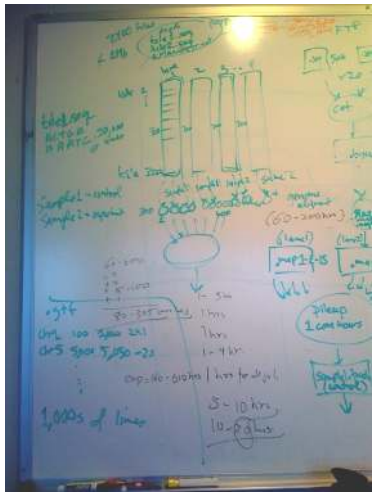
Benefits of Scientific Workflows (from the point of view of an application scientist)

- Conducts a series of computational tasks
- Chaining (outputs become inputs) replaces manual hand-offs
- Ease of use: gives non-developers access to sophisticated codes
- Provides framework to host or assemble community set of applications, can be multi-disciplinary
- Framework to define common formats or standards when useful

Slide credit: David Okaya, SCEC, USC

Multi-domain

Typical local computational environment



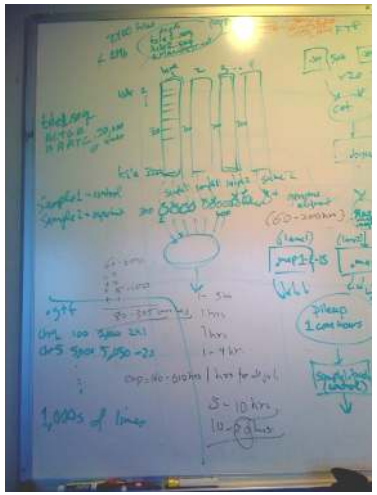
Work Definition

Local Resource

Local Data Storage

Multi-domain

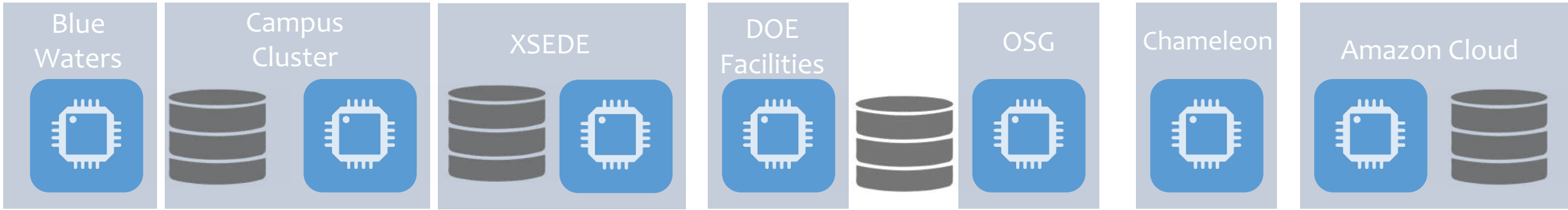
Typical local computational environment

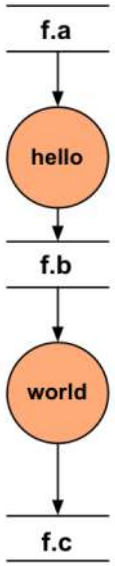


Work Definition

Local Resource

Local Data Storage





To run Hello World on TACC's Wrangler

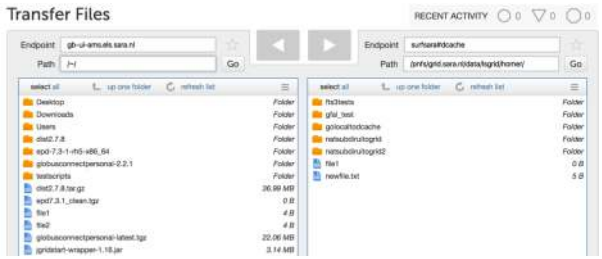
1. Login to TACC

```
localhost$ ssh -l deelman wrangler.tacc.utexas.edu
login1.wrangler$ emacs myjob.sub
```

2. Write submit script

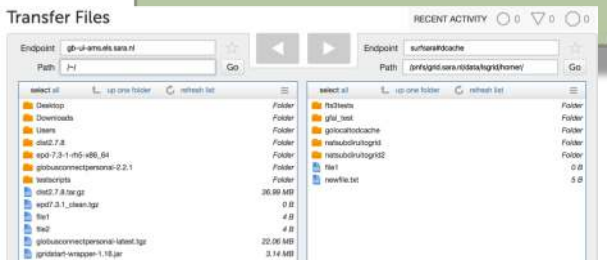
```
#/bin/bash
#SBATCH -J myjob
#SBATCH -o myjob.o%j
#SBATCH -e myjob.e%j
#SBATCH -p normal
#SBATCH -N 1
#SBATCH -n 1
#SBATCH -t 01:30:00
#SBATCH --mail-
user=deelman@gmail.com
#SBATCH --mail-type=all
#SBATCH -A myproject

mkdir $WORK/helloworld
cd $WORK/helloworld
cp $WORK/data/inputs/* .
~/hello
~/world
cp * $WORK/data/outputs/
~/my_output_files/
```

3. Find and bring in your input data



4. Submit script for execution

```
login1.wrangler$ squeue myjob.sub
```

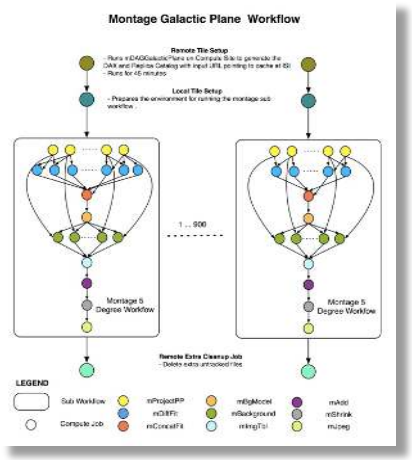


5. Stage out data for further analysis

What if Wrangler goes down/gets decommissioned? What if the job crashed? What about running on multiple platforms?

Multi-domain

Our Approach: Submit locally, Compute globally



Work Definition

Local Resource

Workflow Management System

work

Local Data Storage

data

HTCondor High Throughput Computing

Blue Waters

Campus Cluster

XSEDE

DOE Facilities

OSG

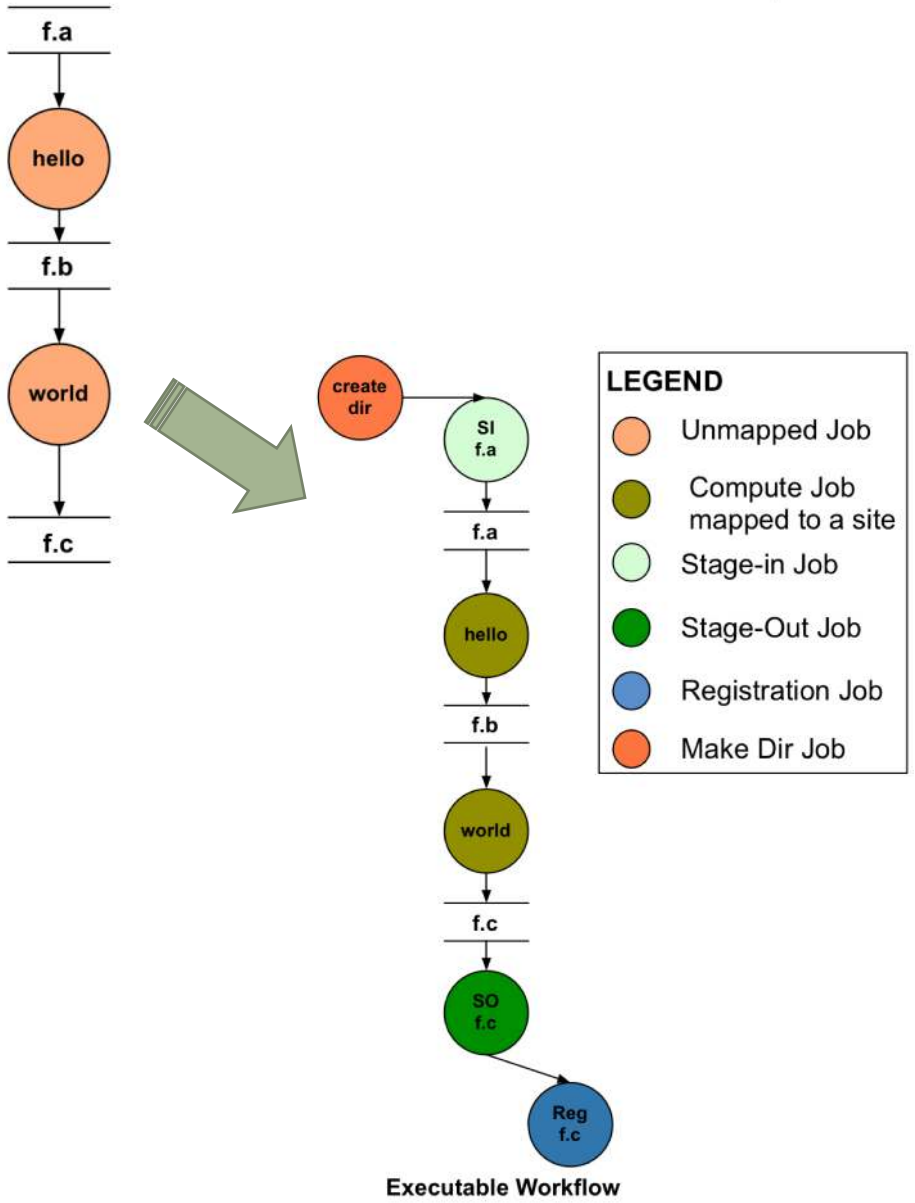
Chameleon

Amazon Cloud

Multi-domain

Pegasus Today

- Scientists describe their computational processes (workflows) at a logical level, without including details of the underlying CI
  - Operates at the level of files and individual codes
- Pegasus maps the abstract workflow to the available resources and infers the needed data transfers
- Pegasus generates an executional workflow, writes out submit files, and executes the workflow
- Underpins other user facing portals: NanoHub (Purdue)
- Provides workflow management for workflow composition tools: Wings (USC)

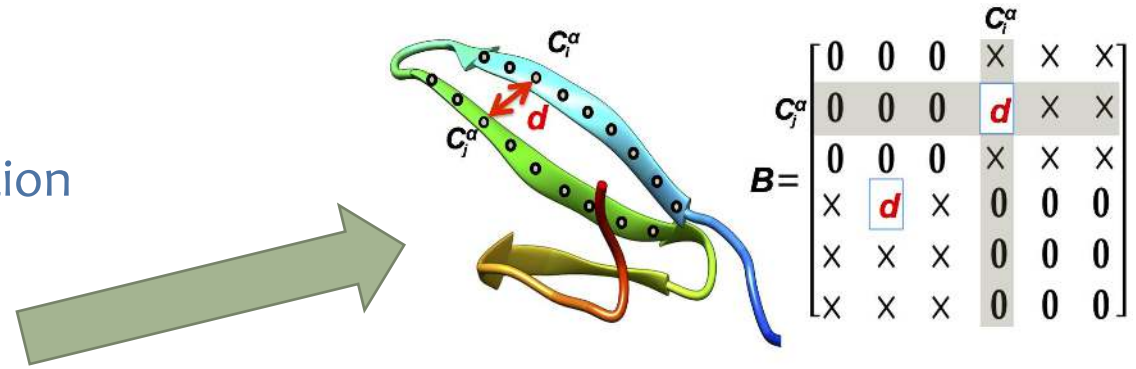


CS research

USC Viterbi School of Engineering Information Sciences Institute

CS Principles Help in Cyberinfrastructure Development

- Structure workflows as **directed acyclic graphs (DAGs)**
  - Re-use of graph traversal algorithms, node clustering, pruning, other complex graph transformation
- Use hierarchical structures in DAGs
  - To achieve scalability, recursion, dynamic behavior
- Develop new algorithms:
  - Task clustering
  - Data placement
  - Data re-use
  - Resource usage estimation
  - Resource provisioning
  - ***In situ* workflows**



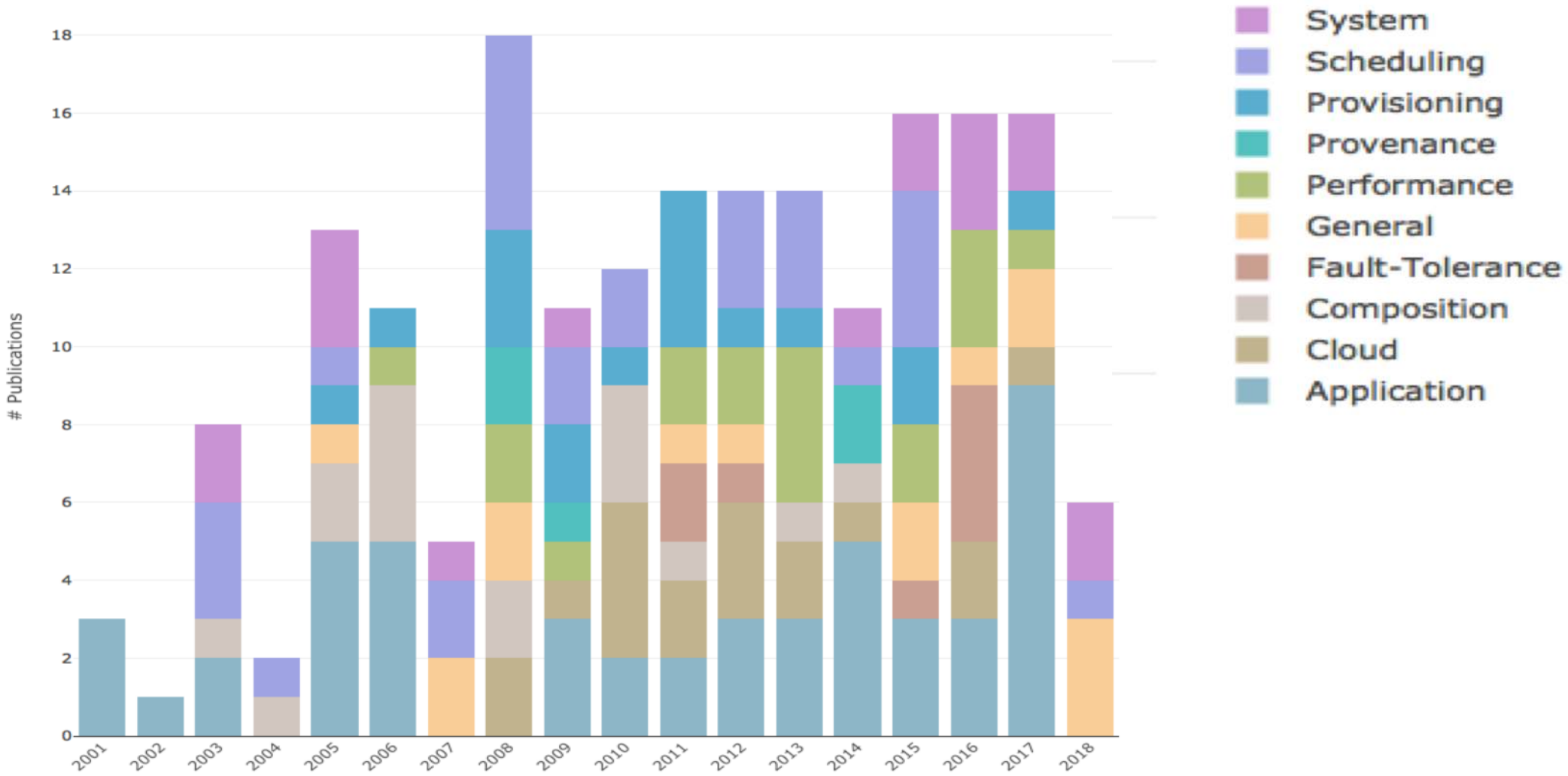
New Direction: In-memory coupling of simulation and analytics Collaboration with U Delaware, Cornell, UTEP

Image credit: Michela Taufer, U. of Delaware

CS research

USC Viterbi School of Engineering Information Sciences Institute

Publications are important for dissemination, education, workforce development, career path, and funding



Publications over the years

Re-use

Leveraging Proven Solutions Key to Innovation

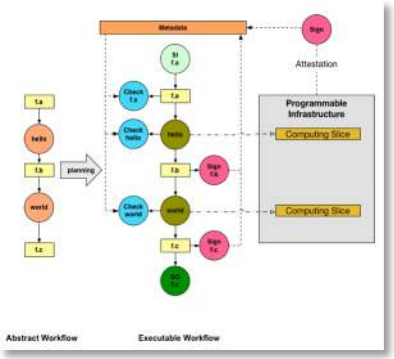
USC Viterbi School of Engineering Information Sciences Institute

- **Leveraged HTCondor's**
  - Job submission to heterogeneous, distributed resources
  - Managing job dependencies expressed as DAGs
  - Job retries and error recovery

HTCondor High Throughput Computing

- **Allowed us to focus on other aspects of automation:**
  - Workflow planning, and re-planning in case of failures
  - Automated data management
  - Specialized workflow execution engines for HPC systems
  - APIs for workflow composition: Python, R, Java, Perl, Jupyter Notebook
  - User-friendly monitoring and debugging tools
  - Provenance tracking
  - **Data integrity**

SWIP Scientific Workflow Integrity with Pegasus



Indiana University
RENCI

CS research

CI Dev.

Using Real Applications Provides Realistic Testing and Evaluation

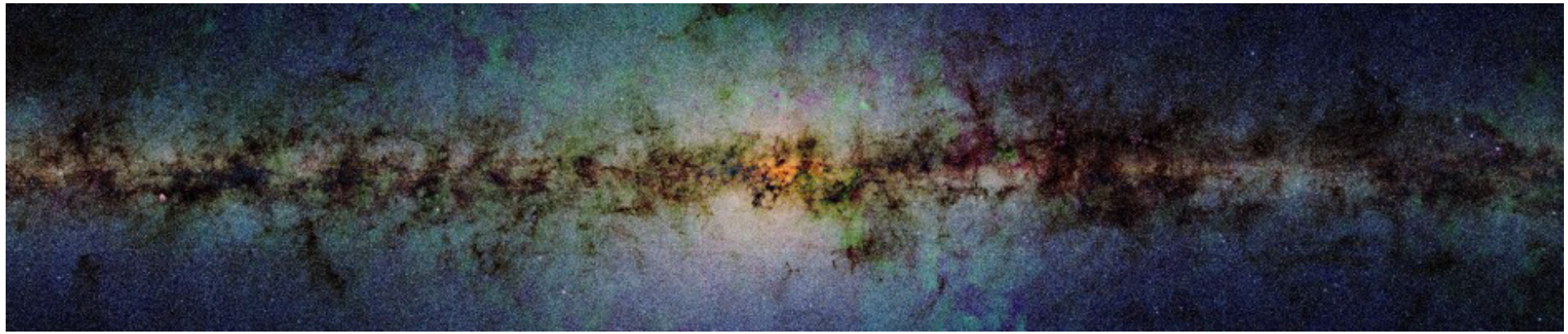
USC Viterbi School of Engineering Information Sciences Institute

Montage: Important application for CS and CI

Open source, open data, scalable, robust

Helps advance CS and test CI: workflow scheduling, resource provisioning, provenance tracking

One of the workflows used in Pegasus' nightly build and test



Montage, an important astronomy application, collaboration with Caltech since 2002

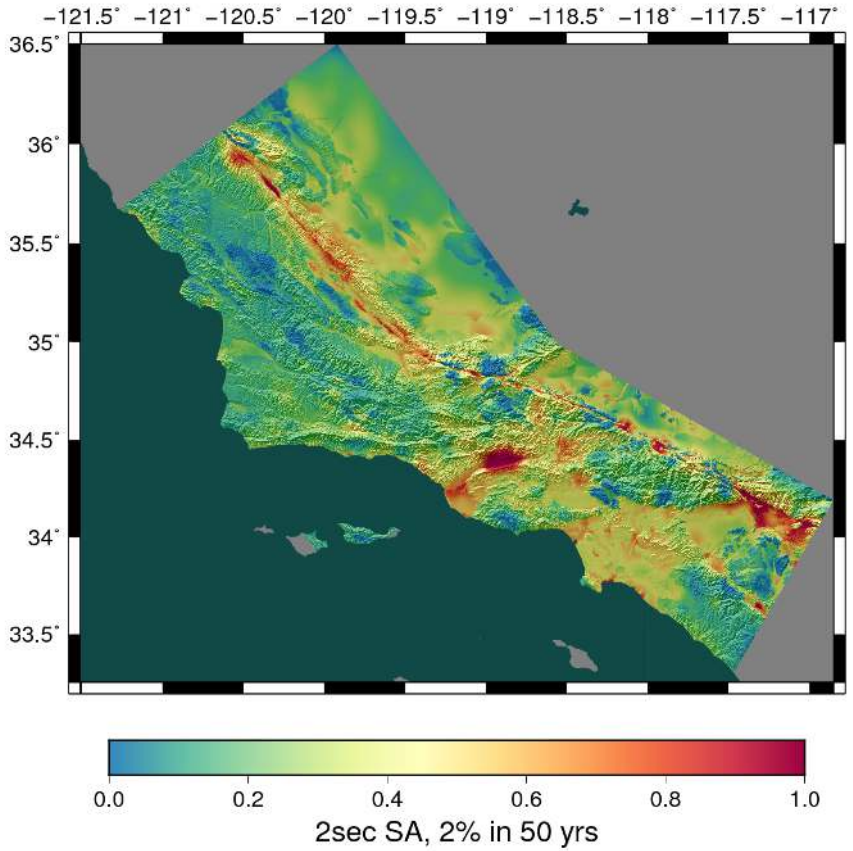
Slide credit: B.Berriman, J.Good, Caltech

Multi-domain

Need applications that push the boundaries of what you can do

USC Viterbi School of Engineering Information Sciences Institute

SCEC's CyberShake: What will peak earthquake shaking be over the next 50 years?



Useful information for:

- Building engineers
- Disaster planners
- Insurance agencies

2017: 21.6 million core hours, 777TB of data

On ORNL's Titan and NCSA's Blue Waters

Slide credit: Southern California Earthquake Center

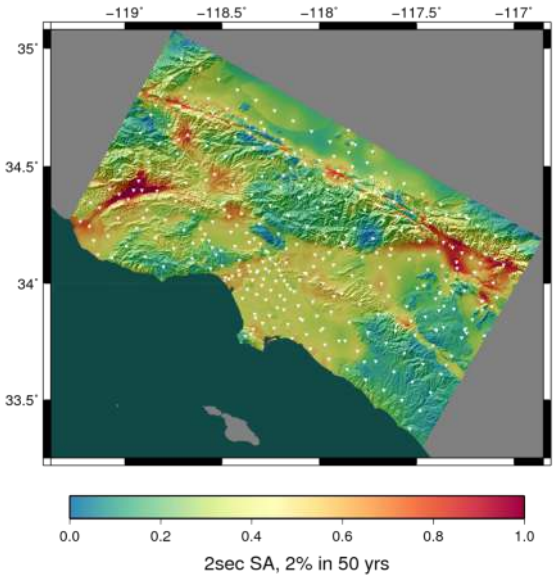
Multi-domain

CS research

Since 2007: CyberShake ran on 9 different HPC systems 100 million core-hours (11,416 years)

Pegasus Optimizations:

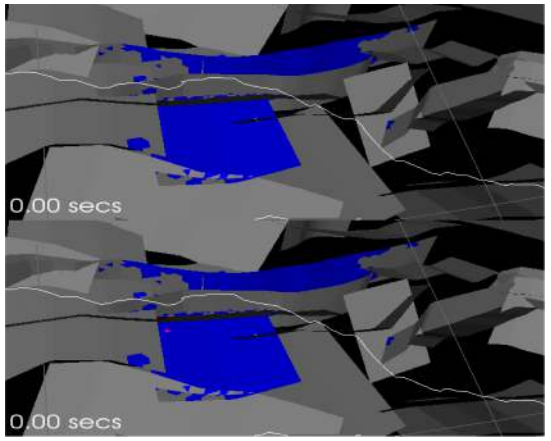
- Task clustering
- MPI-based workflow engine



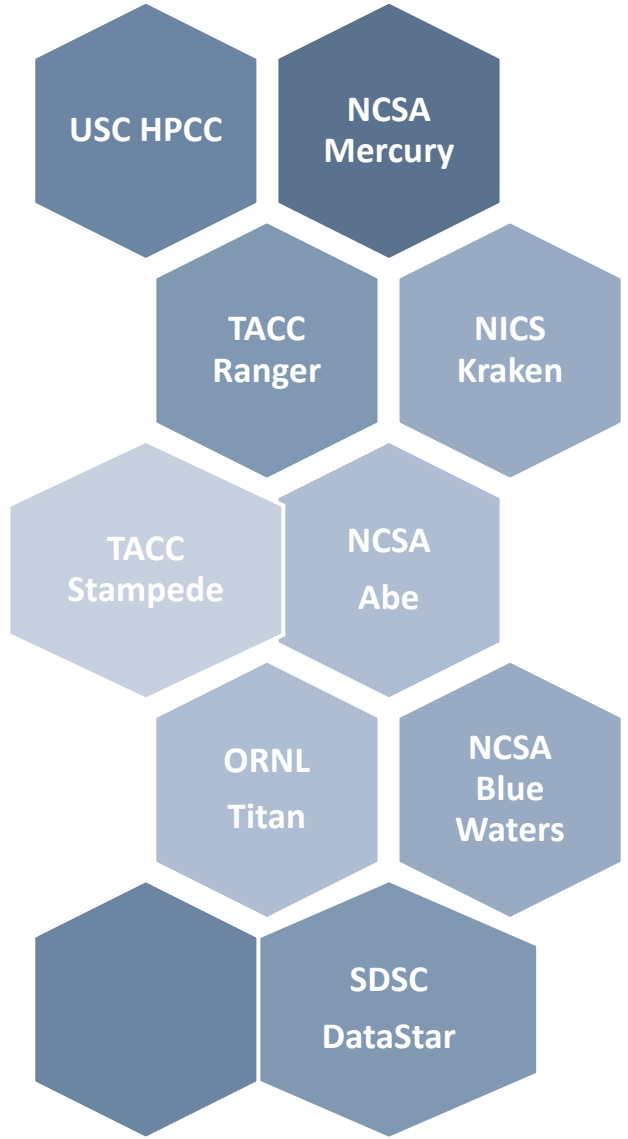
2010: World's first physics-based probabilistic seismic hazard map,

Application Optimizations:

- Workflow restructuring
- MPI/code tuning
- Porting to GPUs



2018: Incorporating earthquake simulator with a 1 million-year catalog of California seismicity

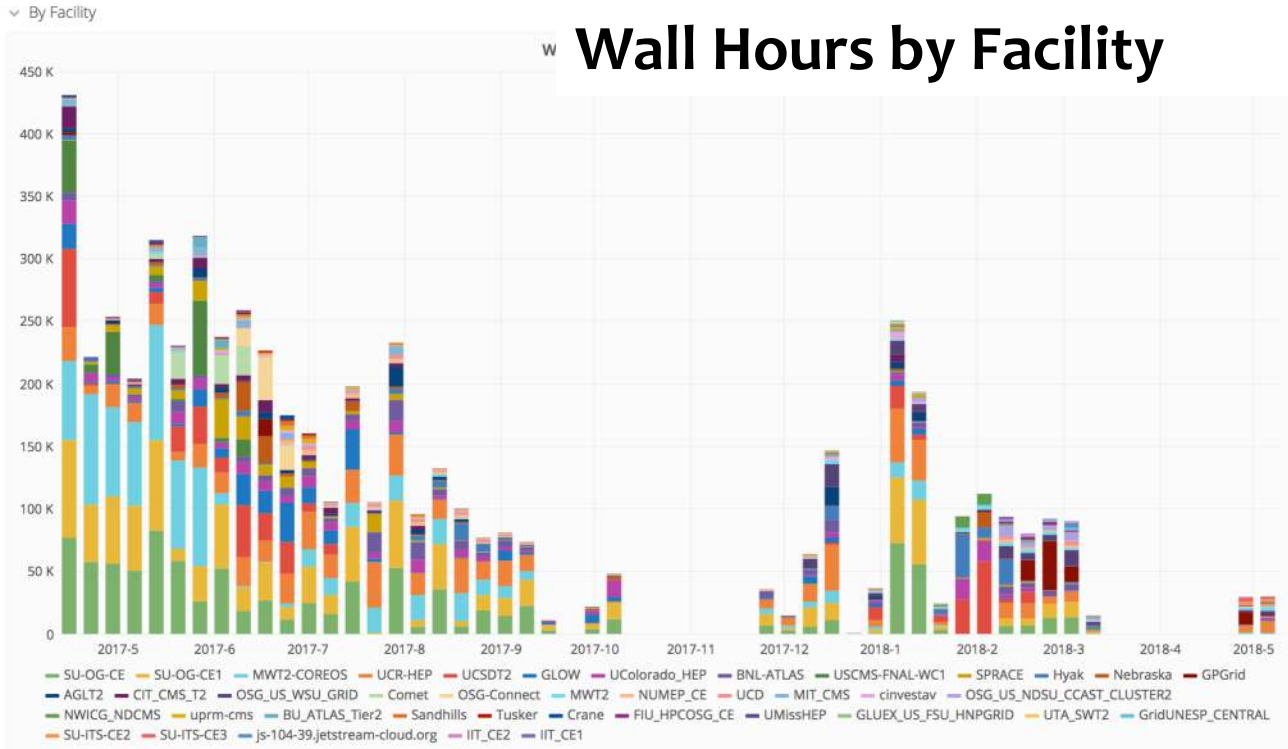


Slide credit: Southern California Earthquake Center

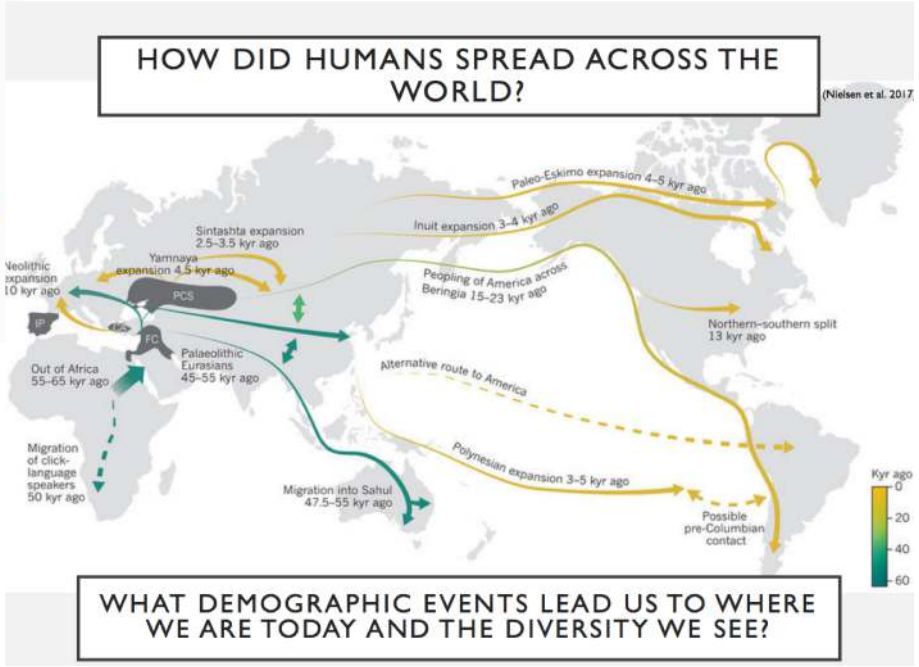
Multi-domain

Arming Individual Scientists with Pegasus on OSG

USC Viterbi School of Engineering Information Sciences Institute



342 workflows
12 million jobs
40 execution sites
~ 7.3 Million Wall Hours



Ariella Gladstein, Ph.D. Student
University of Arizona

Graph credit: Open Science Grid, Image credit: Gladstein

CI Dev. ulti- main

Cross-pollination between domains is highly beneficial

USC Viterbi School of Engineering Information Sciences Institute

LIGO Driven

Development started

Support for Replica Catalog 7/03

Data Cleanup 12/04

New data transfer and monitoring tools, 11/10

Online monitoring dashboard 1/13

1.1 1.2 1.3 1.4 2.0 2.1 2.2 2.3 2.4 3.0 3.1 4.0 4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8 4.9

2001 2003 2005 2007 2009 2011 2013 2015 2017 2019

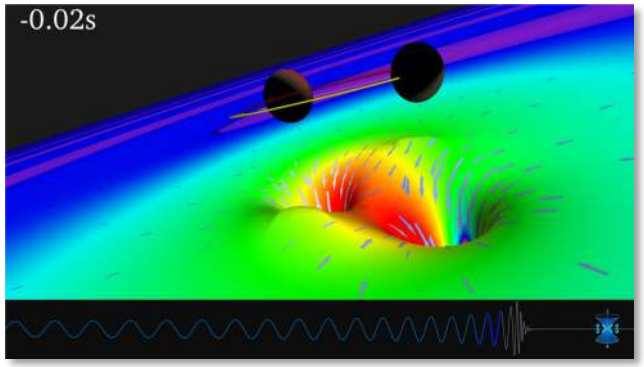
SCEC Driven

10/05 Task clustering

10/06 New partitioning and clustering

8/12 Pegasus MPI-cluster

LIGO's first GW detection



0.2 Second before a pair of massive black holes collide

Benefits the applications

Benefits the software

But, can make the software more complex

Image credit: LIGO Scientific Collaboration

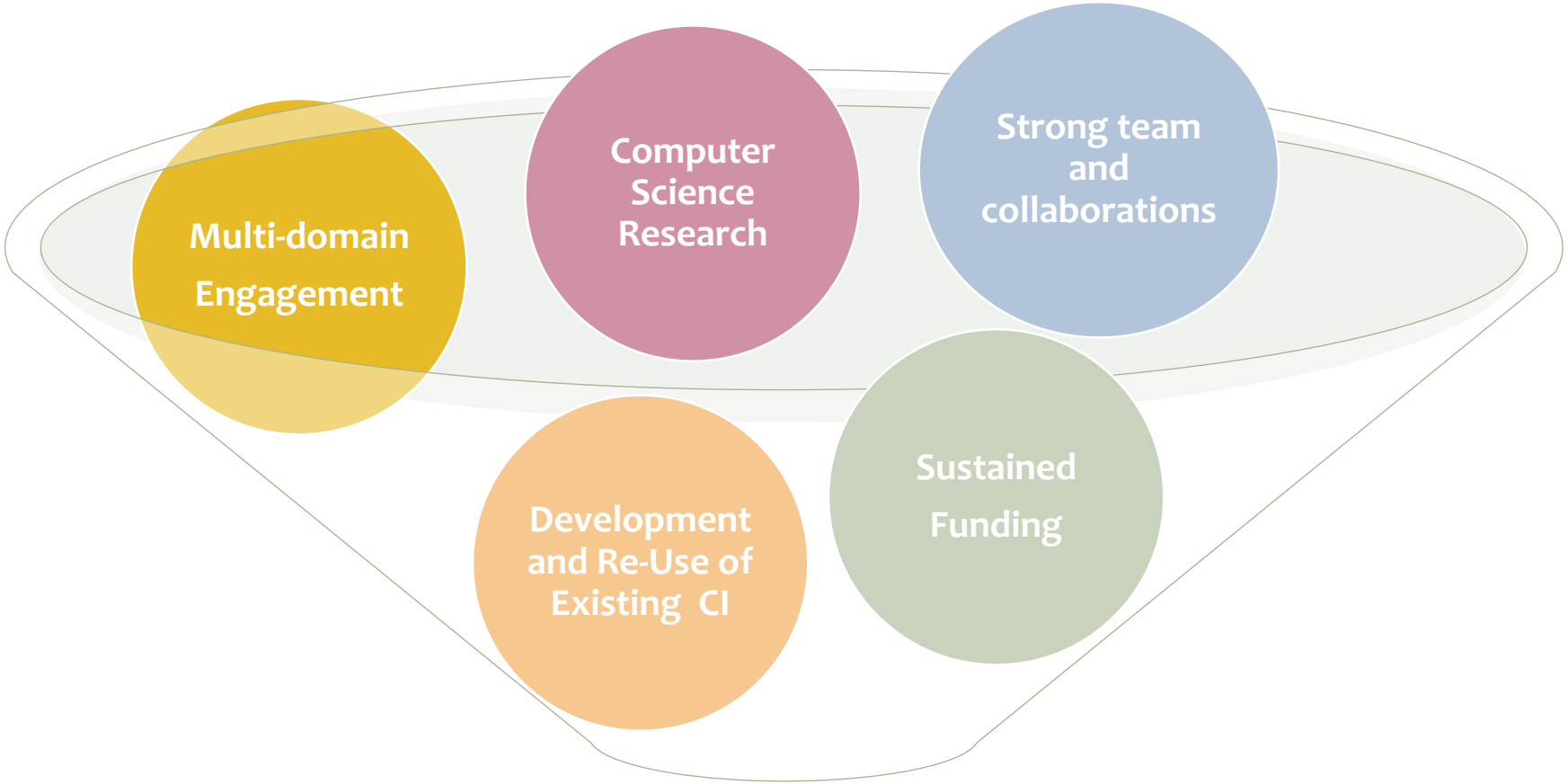
To sustain software, need many different funding sources and need to interleave research, software development, and user support

Pegasus-related funding

2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Pegasus: NSF SDCI program															
Workflow Design: NSF															
LIGO subcontract															
	SCEC subcontract														
	NASA subcontract														
		Monitoring effort: NSF													
		Resource Provisioning effort: NSF													
		SCEC subcontract													
		Data Placement Research: NSF													
				SCEC subcontract											
					Pegasus: NSF SSI Program										
					Open Science Grid subcontract										
					Resource Management: DOE										
						XSEDE subcontract									
						Elastic workflows: NSF									
							LIGO/Pegasus grant: NSF								
							Performance modelling: DOE								
									New directions: Circuit design:DARPA						
									XSEDE subcontract						
									Data integrity: NSF						
										NHERI SimCenter:NSF					
										Pegasus : NSF SSI Program					
										Insitu workflows:NSF					
		Application to bioinformatics: NIH													

Core funding
Application funding
CS Research
Infrastructure support

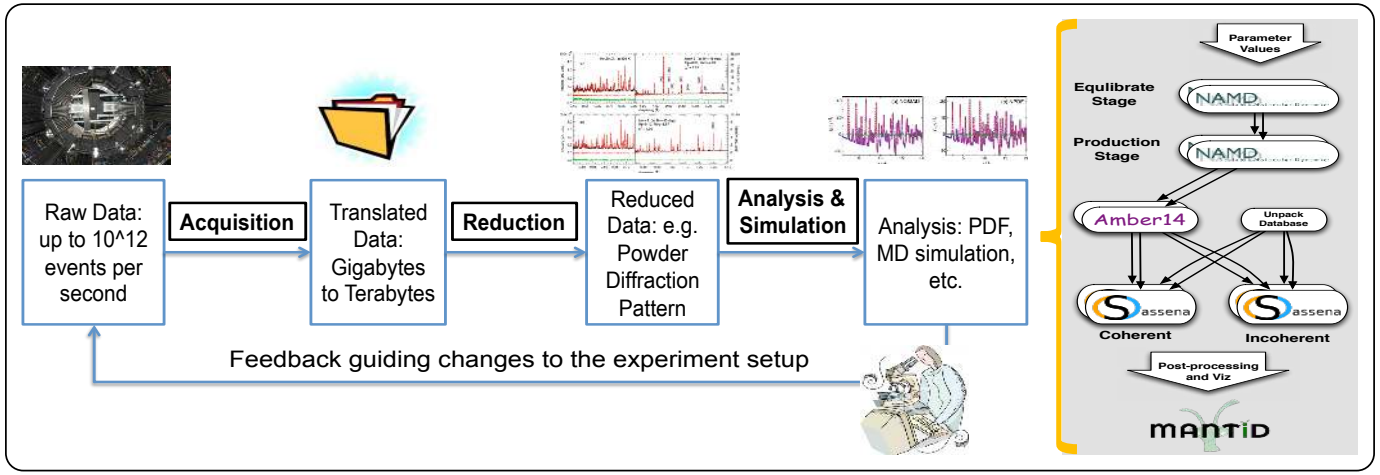
Summary of Observations



Dependable CI

Looking ahead: Application Trends

Spallation Neutron Source



Planned CyberShake for Northern California:

- 869 geographic sites
- 16,000 workflow jobs
- 70 million core-hours on Blue Waters and Titan
- 800 TB of data

- More complex
- Faster time to solution: instrument steering
- More individual researchers in need of significant CI
  - Need intuitive workflow composition, better monitoring, error handling, assisted debugging

Outreach: How do you reach scientists that don't know you are out-there?
Many scientists are going through the same pain
Leverage/enhance existing engagement: NSF's Campus efforts? OSG/XSEDE outreach?
Education and outreach at instruments and experimental facilities?

Image credit: Vickie Lynch, ORNL

Looking ahead: Growing Demand for Automation

HPC Systems
- Complex
- Heterogeneous
- Specialized data storage
- Increasingly faulty

Distributed Systems
- Software Defined capabilities
- Specialized data storage

Clouds
- New platform for science
- Very heterogenous
- Can be costly

Resource Management is Key

Constraints: time to solution, budget

Faulty environment: failure detection and attribution

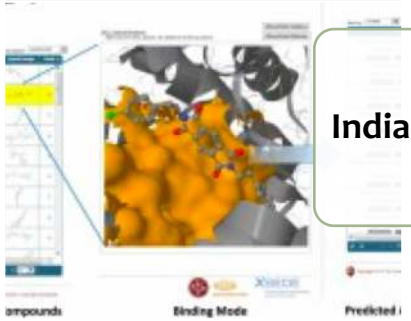
Heterogenous storage: memory, burst buffers, file systems, data xfer nodes

Need to keep track of big data technologies and machine learning solutions that are being developed at a rapid pace by industry

Cancer genetics

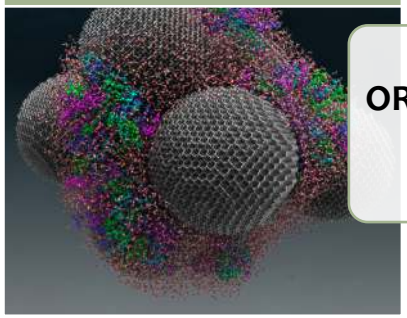
Clemson

Structural Protein-Ligand Interactome



Indiana U.

Molecular Dynamics

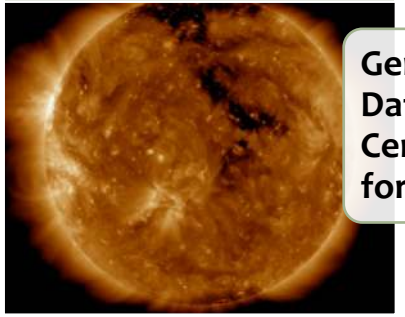


ORNL

Thank you to the team, collaborators and funders

Example Pegasus-enabled Applications

Helioseismology

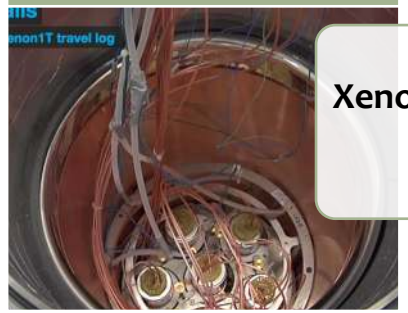


German Data Center for SDO

Next Generation Sequencing

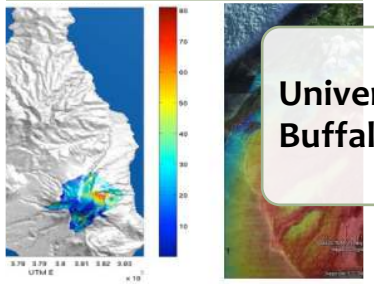
Children's Hospital of Philadelphia

Dark Matter Detection



Xenon1T

Volcano Mass Flow

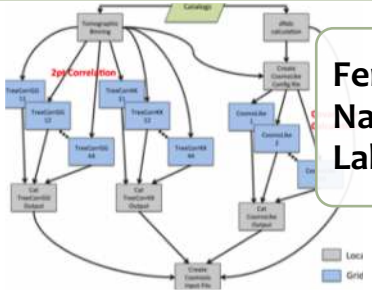


University of Buffalo

Soybean Studies

University of Missouri

Weak Gravitational Lensing



Fermi National Lab

We look forward to future collaborations