The Role of Cyberinfrastructure in Science: Challenges and Opportunities

Ewa Deelman, Ph.D.
University of Southern California,
Information Sciences Institute

SC’19 Invited presentation
Denver, CO, November 20, 2019
Modern Science is Done Across Scales

40 years from hypothesis to discovery

$5.5$ billion instrument

The ATLAS and CMS teams: 3000 members each

> 100 nations participate in LHC

Ariella Gladstein
University of Arizona

40 computational resources
12 million jobs
Over 1.5 Million Wall Hours
Open Science Grid

HOW DID HUMANS SPREAD ACROSS THE WORLD?

WHAT DEMOGRAPHIC EVENTS LEAD US TO WHERE WE ARE TODAY AND THE DIVERSITY WE SEE?
Cyberinfrastructure (CI)

“consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible.”

Enablers of Modern Science: Connecting Scientists, CI Practitioners and CI Facilitators

Individual Platforms
- Individual Large Facilities
- DOE National Labs
- Projects
- Campuses
- Software Institutes

Science Gateways
Community Institute

2019 OLCF User meeting

2018 OSG Summer School

Gateways 2019
Enablers of Modern Science: Connecting Data

October 28, 2019 #GettyFire

UCSD WiFire Platform
Connected Data Informs Emergency Services

October 28, 2019
#GettyFire

UCSD WiFire Platform helped LAFD set up evacuation zones

Ilkay Altintas
Invited talk, Thursday am
Enabler of Modern Science: Connecting Instruments and Compute Resources

Images courtesy of Benedikt Reidel, IceCube

- 5000 light-sensitive detectors on 86 cables
- Create tiny blue flashes of light when neutrinos react with ice
- Direction of neutrinos can be found
- Reconstructs which regions of space neutrinos are coming from

- Raw data (~3 TB/day) - written to tape/disk at the South Pole and shipped to UW Madison once a year
- Filtered data – filtered “online” at the South Pole (i.e., Level 1)
- Alerts of interesting events can be created (triggers)
October 16th 2017: “LIGO and Virgo make first detection of gravitational waves produced by colliding neutron stars”

And kicked off a new era of multi-messenger astrophysics

“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
Connecting Domain and Computer Scientists

Connections take time: the Pegasus and LIGO Example

- **2001**: First Pegasus prototype
- **2004**: Blind injection detection
- **2007**: First detection of black hole collision
- **2011**: Multi-messenger neutron star merger observation
- **2015**: Nobel Prize

Image credits: LIGO Scientific Collaboration
Workflow Management: Connecting Applications and CI

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

Analysis measures the statistical significance of collected data

Efficient, scalable, and robust execution of tasks and data access

LIGO, Open Science Grid, XSEDE, Blue Waters

Science Workflow

Automation

Distributed Power

2015/16
Challenges of Workflow Management

**Challenges across domains**
- Need to describe complex workflows in a simple way
- Need to access distributed, heterogeneous data and resources (heterogeneous interfaces)
- Need to deal with resources/software that change over time

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

```
#SBATCH --job myjob
#SBATCH --o myjob.o@j
#SBATCH --e myjob.e@j
#SBATCH --partition normal
#SBATCH --nodes 1
#SBATCH --ntasks 1
#SBATCH --time 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
-1/world
cp * $WORK/data/outputs/
-1/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

```
```
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

```
#SBATCH --job-name=helloworld
#SBATCH --nodes=1
#SBATCH --tasks-per-node=1
#SBATCH --time=01:30:00
#SBATCH --mail-user=deelman@gmail.com
#SBATCH --mail-type=all
#SBATCH --output=/my_output_files/
```

3. Find and bring in your input data

4. Submit script for execution

5. Stage out data for further analysis

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

Local Data Storage

Work Definition

Local Resource
Typical Local Computational Environment

- Campus Clusters
- HPC
- HTC
- Clouds

Local Data Storage
Local Resource
Work Definition
Connecting Local and Global Environments

Local Data Storage

Workflow Management System

Local Resource

Work Definition

Campus Clusters

HPC

Data

Work

Data

Work

HTC

Clouds
Pegasus Workflow Management System

- Operates at the level of files and individual applications
- Allows scientists to describe their computational processes (workflows) at a logical level
- Without including details of target heterogeneous CI (portability)
- Scalable to $O(10^6)$ tasks, TBs of data
- Captures provenance and supports reproducibility
- Includes monitoring and debugging tools

Composition in Python, R, Java, Perl, Jupyter Notebook
Portable Description

Users do not worry about low level execution details

**Abstract Workflow**

**Executable Workflow**

- **Logical filename (LFN)**
  - Platform independent (abstraction)

- **Transformation**
  - Executables (or programs)
  - Platform independent

- **Stage-in job**
  - Transfers the workflow input data

- **Cleanup job**
  - Removes unused data

- **Stage-out job**
  - Transfers the workflow output data

- **Registration job**
  - Registers the workflow output data
Connecting to Heterogeneous Resources

**Single resource**
- *stage-in job*
- *cleanup job*
- *stage-out job*
- *registration job*

**Two resources**

**HPC resource**
Connecting to Heterogeneous Storage

- Workflows execute across Open Science Grid and European Grid Infrastructure
- Rucio used for data management
- MongoDB used for tracking science runs and data products

Pegasus provides interfaces for a variety of data discovery and movement

XENONnT - Dark Matter Search

HTTP, SCP, GridFTP, Globus Service, iRods, Amazon S3, Google Storage, SRM, FDT, stashcp, cp, ln -s

Image credit: XENONnT project
Connecting Scientists to Resources at Scale

SCEC’s CyberShake: What will the peak earthquake motion be over the next 50 years?

Useful information for:

- Building engineers
- Disaster planners
- Insurance agencies

Slide credit: Southern California Earthquake Center
Connecting Scientists to Resources at Scale

- 120 million core-hours
- 39,285 jobs
- 1.2 PB of data managed
- 157 TB of data automatically transferred
- 14.4 TB of output data archived

- NCSA Blue Waters
- OLCF Titan

Total map: 170 million core hours > 19,407 core years

2018-2019 Mapping Northern California

Slide credit: Southern California Earthquake Center
Connecting Mix Workloads to Heterogeneous/Changing CI

Since 2007: 215 million core-hours (24,543 years)
9 different supercomputers

Pegasus Optimizations:
• Task clustering
• MPI-based workflow engine

Application Optimizations:
• Workflow restructuring
• MPI/code tuning
• Porting to GPUs

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

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

Slide credit: Southern California Earthquake Center
Connecting Workflow Management and Resource Provisioning Systems

CASA: Collaborative Adaptive Sensing of the Atmosphere

Tracking of rare events requires additional resources and dynamic resource provisioning capabilities

- 7M people,
- >100K businesses, >1500 Corporate HQs

Maximum wind velocity Sends alerts to users
Nowcasts: predict the immediate weather
Connecting Workflow Management and Resource Provisioning Systems

• Compute and storage resources on both ExoGENI and Chameleon clouds
• Dynamic resource provisioning on ExoGENI and Chameleon clouds
• High speed data movement via ExoGENI’s dedicated layer-2 overlay networks
• Pegasus interacts with the Dynamo resource provisioners to acquire resources as needed
Connecting CS Research and CI Development

• Structure workflows as **directed acyclic graphs (DAGs)**
  • Re-use of graph traversal algorithms, node clustering, pruning, other complex graph transformation

• Use hierarchical structures and recursion in DAGs
  • To achieve scalability and dynamic behavior

• Develop new algorithms:
  • Task clustering
  • Data placement
  • Data re-use
  • Resource usage estimation
  • Resource provisioning
  • **Insitu workflows**

New Direction: In-memory coupling of simulation and analytics. Collaboration with U Tennessee Knoxville, Cornell, U. of New Mexico

Image credit: Michela Taufer, U. of Tennessee, Knoxville
Connecting Robust CI Components

Since 2001 leveraged HTCondor’s capabilities:
• Job submission to heterogeneous, distributed resources
• Managing job dependencies expressed as DAGs
• Job retries and error recovery

Submit host/local/community resource

Remote Execution Environment
Connecting Robust CI Components

• Allowed us to focus on other aspects of automation:
  • Workflow planning, and re-planning in case of failures
  • Automated data management
  • User-friendly monitoring and debugging tools
  • Specialized workflow execution engines for HPC systems
  • Provenance tracking
  • Data integrity
Using Real Applications Provides Realistic Testing and Evaluation

Montage, an important Astronomy Application, collaboration with Caltech since 2002

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

Connecting CS researchers and Domain Applications

Slide credit: B.Berriman, J.Good, Caltech
Connecting Application Requirements: Cross-pollination between domains is highly beneficial

LIGO Driven

Support for Replica Catalog

Data Cleanup

New data transfer tools, pegasus-statistics, pegasus-plots

Online monitoring dashboard

Development started


SCEC Driven

Task clustering

New partitioning and clustering

Pegasus MPI-cluster

Benefits the applications

Benefits the software

But, can make the software more complex

Image credit: LIGO Scientific Collaboration
Connecting to Testing Platforms

- Need testbeds for evaluation
- Need realistic workloads, benchmarks and traces for evaluation and comparative studies

National-scale Computer Science research testbed infrastructure
- Leverages existing NSF investments: PAWR (Wireless), Cloud platforms (CloudLab, Chameleon, Cloud Access), national supercomputing facilities and testbeds
- BYOE (Bring your own Equipment) Research platform for Cybersecurity, ML/AI, IoT, network protocols, distributed systems and applications
- Sandbox for scientific workflow experimentation

Ilya Baldin, RENCI, PI
Develop a model and a plan for a Cyberinfrastructure Center of Excellence

• Dedicated to the enhancement of CI for science
• Platform for knowledge sharing and community building
• Forum for discussions about CI sustainability and workforce development and training
• Currently fostering working group discussions around the science data life cycle and identity management
• Key partner for the establishment and improvement of large-scale projects with advanced CI architecture designs
• Partnering with other community efforts (TrustedCI, OSG,..) to support science

http://cicep-pilot.org/

10/2018– 9/2020
Increased Connectivity May Increase CI Complexity

Increased need for
- automation
- autonomy

Role of ML

Current challenges increase

Trust: How do you know that what we observe is real?
Inspect
Understand
Reproduce
Automation Changes the Workforce Landscape

How will the scientist of the future look like?
How will the human machine interfaces look like?

BIG Thanks to the Pegasus Team and amazing collaborators!
Connecting CI, Connecting Science, Connecting People

Instruments
CI Practitioners
CI Resources
CI Facilitators
CI Software
Computer Scientists
Domain Scientists

http://cicoe-pilot.org  http://pegasus.isi.edu