SCIENCE IMPACT OF SUSTAINED CYBERINFRASTRUCTURE:

The Pegasus Workflow Management System: Evolution and Impact

Ewa Deelman, Ph.D.

University of Southern California,
Information Sciences Institute
Exploring a Scientific Question

**Scientific Problem**
Earth Science, Astronomy, Neuroinformatics, Bioinformatics, etc.

**Computational Scripts**
Shell scripts, Python, Matlab, etc.

**Analytical Formulation**

\[
U \frac{d^2}{dx^2} U^{-1} g = \left( \frac{d}{dx} U^{-1} \right) \times \left( \frac{d}{dx} U^{-1} \right) g
= \frac{d}{dx} \left[ \psi + \frac{1}{2} \frac{\psi'}{\psi} \right] \cdot \psi + \frac{1}{2} \left[ \psi' + \frac{1}{2} \frac{\psi''}{\psi} \right] \cdot \psi' \psi''
= \psi'' + 2 \psi' + \frac{1}{2} \frac{\psi''}{\psi} - \frac{1}{2} \frac{\psi'''}{\psi''}
\]

**Automation**

**Distributed Computing**
Clusters, HPC, Clouds, etc.

**Scientific Result**
Models, Quality Control, Image Analysis, etc.

**Monitoring and Debug**
Fault-tolerance, Provenance, etc.
Exploring a Scientific Question

Scientific Problem
Earth Science, Astronomy, Neuroinformatics, Bioinformatics, etc.

Computational Scripts
Shell scripts, Python, Matlab, etc.

Analytical Formulation

$U \frac{d^2 U}{dx^2} U^{-1} g = \left( \frac{d}{dx} U^{-1} \right) \times \left( \frac{d}{dx} U \right) g = \frac{d}{dx} \left[ \frac{d}{dx} - \frac{1}{2} \frac{d^2}{dx^2} - \frac{1}{2} \frac{d^2}{dx^2} \right] \cdot g + \frac{1}{2} \frac{d^2}{dx^2} \cdot \psi + \frac{1}{2} \frac{d^2}{dx^2} \cdot \psi^2
= g \psi^2 + 2g \psi + \frac{1}{2} g \times \frac{\psi^2}{\psi^2 + \psi^2}$

Distributed Computing
Clusters, HPC, Cloud, etc.

Automation

Monitoring and Debug
Fault-tolerance, Provenance, etc.

Scientific Result
Models, Quality Control, Image Analysis, etc.
Working with LIGO (Laser-Interferometer Gravitational Wave Observatory)

- **2001**: First Pegasus prototype
- **2007**: Blind injection detection
- **2014**: First detection of black hole collision
- **2017**: Multi-messenger neutron star merger observation

Nobel Prize

Image credit: LIGO Scientific Collaboration
First GW detection: ~21K workflows with ~107M tasks

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 Pegasus?

Cyberinfrastructure (CI) =
computing systems
+ data storage systems
+ advanced instruments
+ data repositories
+ visualization environments
+ people

Connected by software
and high-performance networks
for research and breakthroughs
not otherwise possible

Dependable Cyberinfrastructure
Takes time to build a team and expertise

Front Row: Tu Mai Anh Do, Rajiv Mayani, Ryan Mitchell, Ragini Church, Ewa Deelman, Mukund Murrali
Back Row: Karan Vahi, Mats Rynge, George Papadimitriou, Rafael Ferreira da Silva

USC Viterbi School of Engineering
How did Pegasus Start?

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

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

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

Circa. 2001
Challenge: How Translate a Science Request to an Actionable Plan?

**Explore AI planning techniques**

**LIGO Experiment**
(Laser Interferometer Gravitational-wave Observatory)

**Lost in translation: high-level abstraction for this science domain**
Found: new research direction: management of workflows in distributed environments

**Work with Yolanda Gil and Jim Blythe**
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

• Our focus:
  • Separation between workflow description and workflow execution
  • Workflow planning and scheduling (scalability, performance)
  • Task execution (monitoring, fault tolerance, debugging)
Typical local computational environment

Multi-domain

Work Definition

Local Resource

Local Data Storage
Typical local computational environment

Work Definition

Local Resource

Local Data Storage

Blue Waters
Campus Cluster
XSEDE
DOE Facilities
OSG
Chameleon
Amazon Cloud

Information Sciences Institute

USC Viterbi
School of Engineering
To run Hello World on USC’s HPC System

1. Login to System

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

2. Write submit script

```
#!/bin/bash
#SBATCH --job-name=myjob
#SBATCH --output=myjob.o
#SBATCH --error=myjob.e
#SBATCH --partition=normal
#SBATCH --nodes=1
#SBATCH --time=01:30:00
#SBATCH --user deelman@gmail.com
#SBATCH --mail-type=all
#SBATCH --mail-user deelman@gmail.com

mkdir $WORK/helloworld
cd $WORK/helloworld
cp $WORK/data/inputs/* .
rm -rf hello
rm -rf world
cp * $WORK/data/outputs/
```

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 the system goes down/gets decommissioned? What if the job crashed? What about running on multiple platforms?
Our Approach: Submit locally, Compute globally

Local Resource

Workflow Management System

Work Definition

Local Data Storage

Blue Waters
Campus Cluster
XSEDE
DOE Facilities
OSG
Chameleon
Amazon Cloud

Multi-domain
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 WINGS, Python, R, Java, Perl, Jupyter Notebook
Abstract Workflow

Portable Description
Users do not worry about low level execution details

Executable Workflow

**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

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

*Transformation*
Executables (or programs) platform independent

Http://cicoe-pilot.org
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
  – **Insitu workflows**

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

Image credit: Michela Taufer, U. of Tennessee
Leveraging HTCondor’s
  – Job submission to heterogeneous, distributed resources
  – Managing job dependencies expressed as DAGs
  – Job retries and error recovery

Allowed us to focus on other aspects of automation:
  – Workflow planning, and re-planning in case of failures
  – Automated data management
  – APIs for workflow composition in Python, R, Java, Perl, Jupyter Notebook
  – 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: 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
Supporting Large-Scale Applications

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

Useful information for:
- Building engineers
- Disaster planners
- Insurance agencies

- 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
Cross-pollination between domains is highly beneficial.

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

Online monitoring dashboard

LIGO’s first GW detection

0.2 Second before a pair of massive black holes collide

LIGO Driven

Support for Replica Catalog 7/03
Data Cleanup 12/04

Development started

SCEC Driven

Task clustering 10/05
New partitioning and clustering 10/06
Pegasus MPI-cluster 8/12

Benefits the applications
Benefits the software

But, can make the software more complex

Image credit: LIGO Scientific Collaboration
Arming Individual Scientists with Pegasus on OSG

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

Ariella Gladstein, Ph.D. Student
University of Arizona
Summary of Observations

- Multi-domain Engagement
- Computer Science Research
- Development and Re-Use of Existing CI
- Strong team and collaborations
- Sustained Funding

Dependable and Impactful Software

Information Sciences Institute
Looking ahead: Growing Demand for Automation

High Performance Computing 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, budget, resource capabilities

Faulty environment: detection and attribution

Heterogenous storage: memory, BB, FS, WAN

IoT devices

Programmable networks

Significant and applicable innovation in industry:
Need to keep track of big data technologies and machine learning solutions
Increased use of automation and ML presents a new set of challenges

- Trust: How do you know that what we observe is real?
- Transparency
- Understanding
- Reproducibility
Automation Changes the Workforce Landscape

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

http://pegasus.isi.edu

BIG Thanks to the Pegasus Team and amazing collaborators!

Information Sciences Inst.