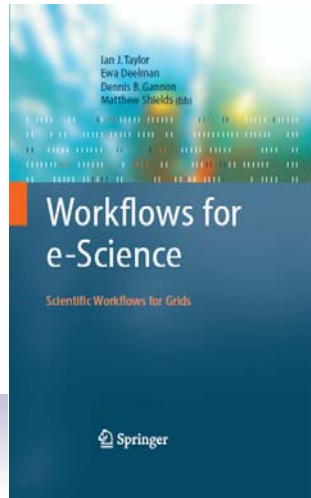
Managing Scientific Workflows on OSG with Pegasus

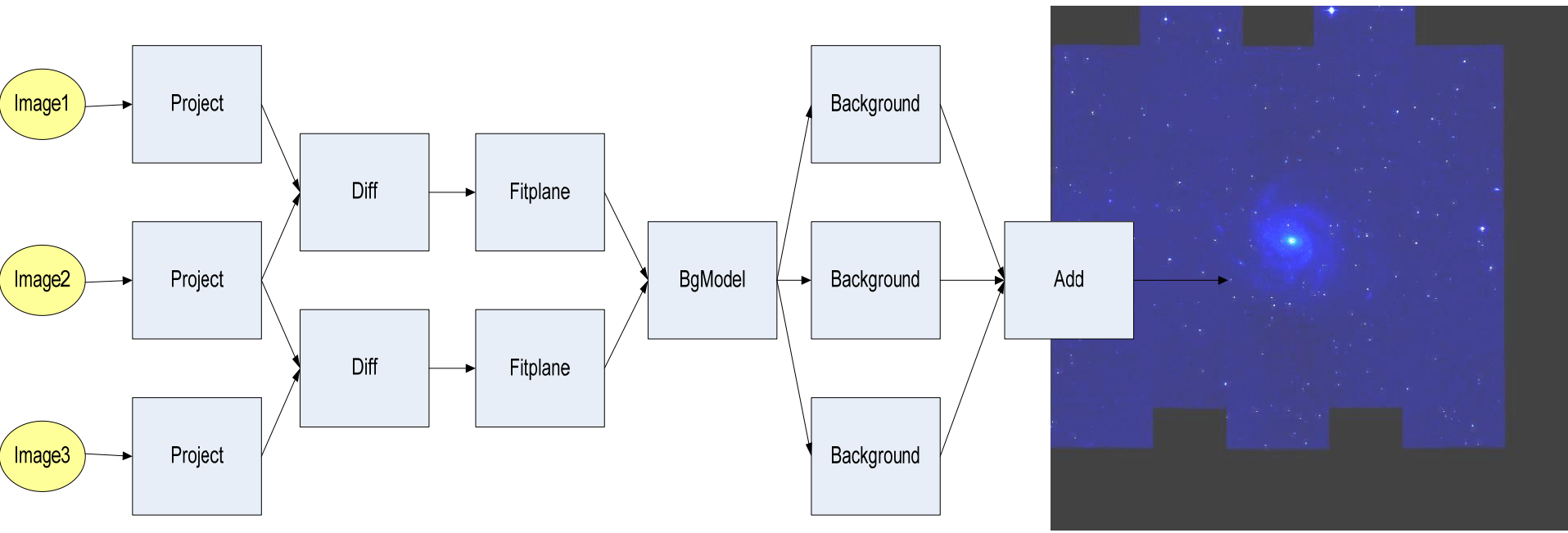
Ewa Deelman

USC Information Sciences Institute



Scientific (Computational) Workflows

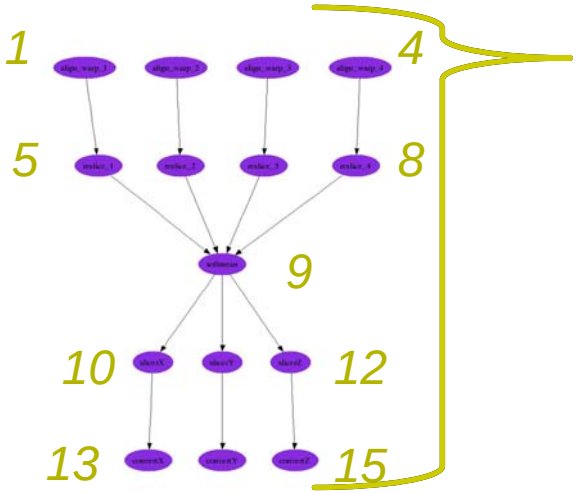
- Enable the assembly of community codes into large-scale analysis
- Montage example: Generating science-grade mosaics of the sky (Bruce Berriman, Caltech)



Pegasus: Planning for Execution in Grids, Est. 2001

- Automatically maps high-level resource-independent workflow descriptions onto distributed resources such as the TeraGrid and the Open Science Grid
- Enables the construction of workflows in abstract terms without worrying about the details of the underlying CyberInfrastructure
- Pegasus is used to map complex, large-scale scientific workflows with thousands of tasks processing TeraBytes of data
- Pegasus applications include LIGO's Binary Inspiral Analysis, NVO's Montage, SCEC's CyberShake simulations, Neuroscience, Artificial Intelligence, Genomics (GADU), others
- Pegasus improves the performance of applications through:
  - Data reuse to avoid duplicate computations and provide reliability
  - Workflow restructuring to improve resource allocation
  - Automated task and data transfer scheduling to improve overall runtime
- Pegasus provides reliability through dynamic workflow remapping and DAGMan workflow execution
- Pegasus uses Condor and Globus to provide the middleware for distributed environments

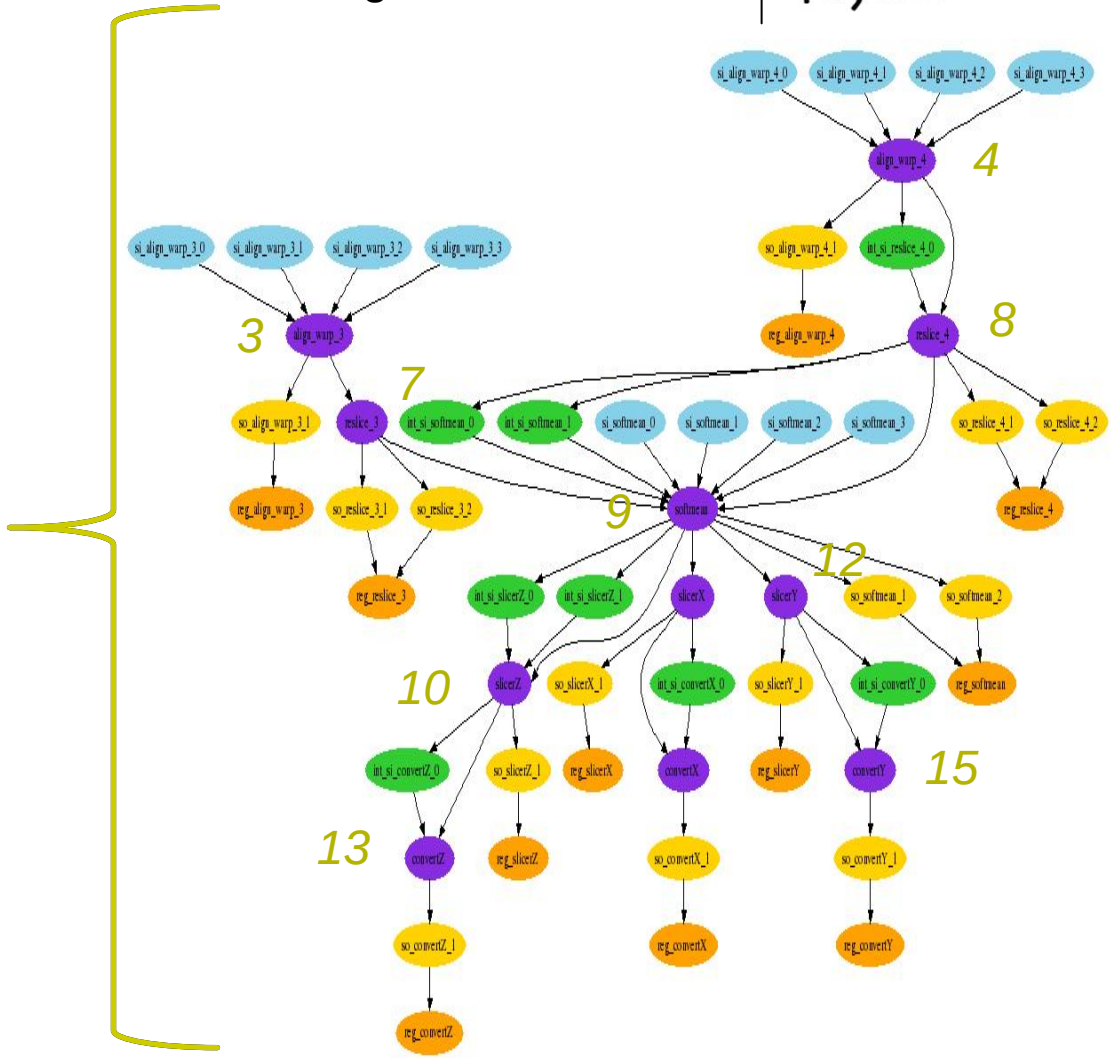
Pegasus Workflow Mapping



Original workflow: 15 compute nodes devoid of resource assignment

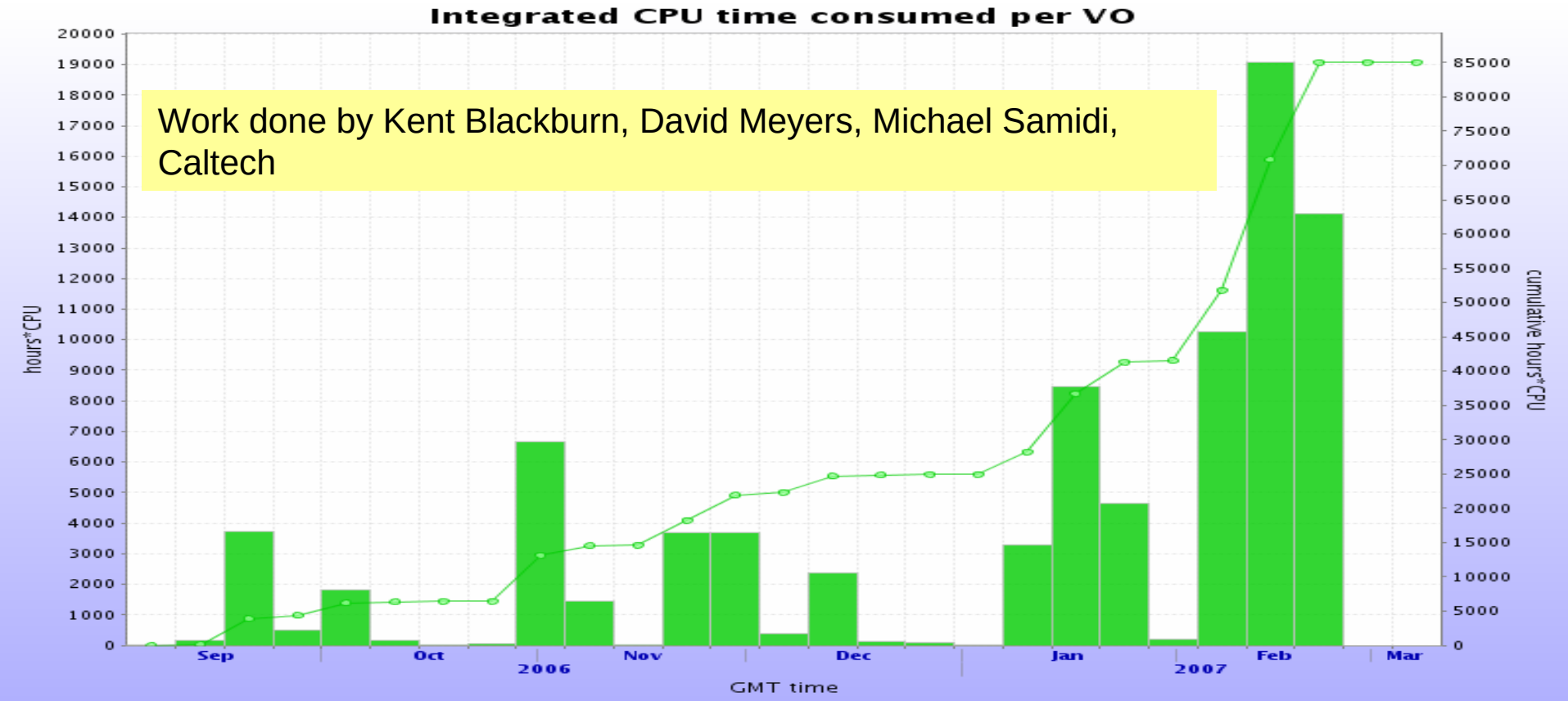
Resulting workflow mapped onto 3 Grid sites:

- 11 compute nodes (4 reduced based on available intermediate data)
- 13 data stage-in nodes
- 8 inter-site data transfers
- 14 data stage-out nodes to long-term storage
- 14 data registration nodes (data cataloging)

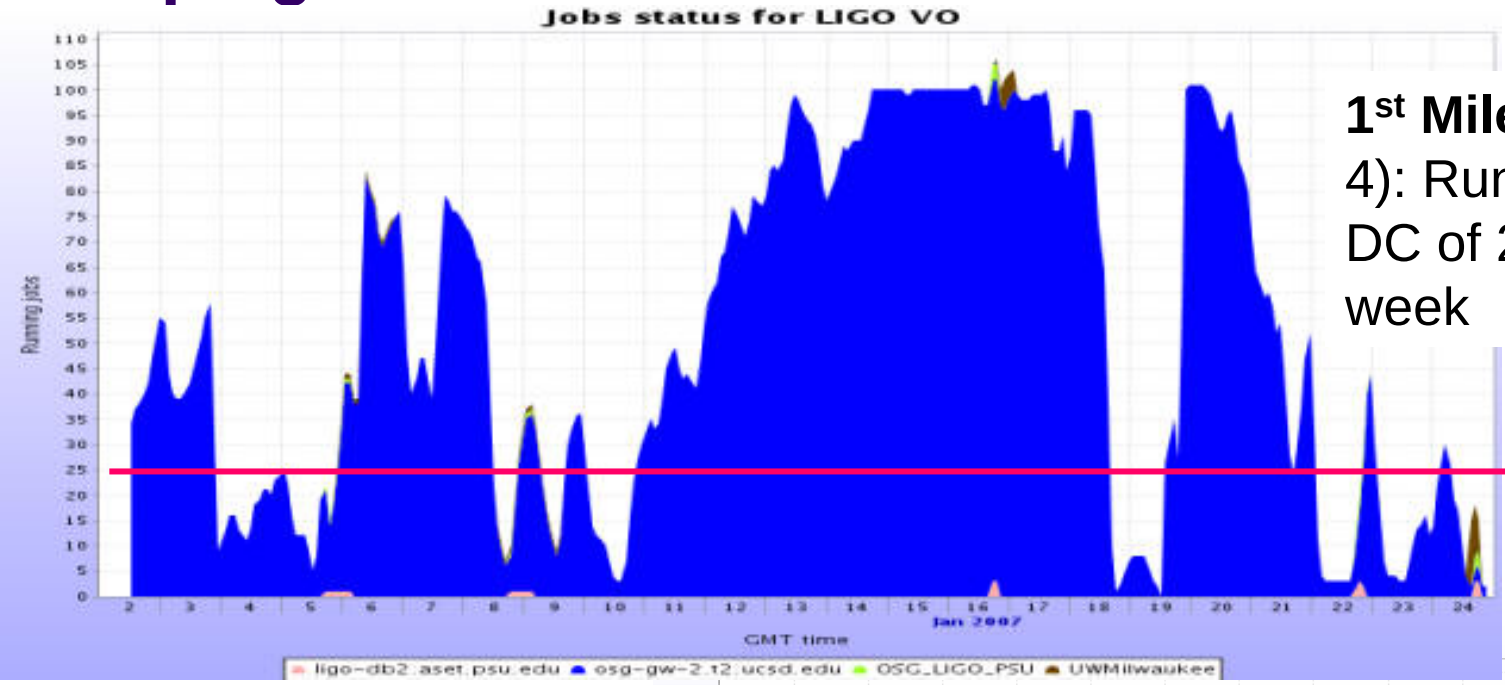


Supporting OSG Applications

- LIGO—Laser Interferometer Gravitational-Wave Observatory
  - Aims to find gravitational waves emitted by objects such as binary inpirals 9.7 Years of CPU time over 6 months

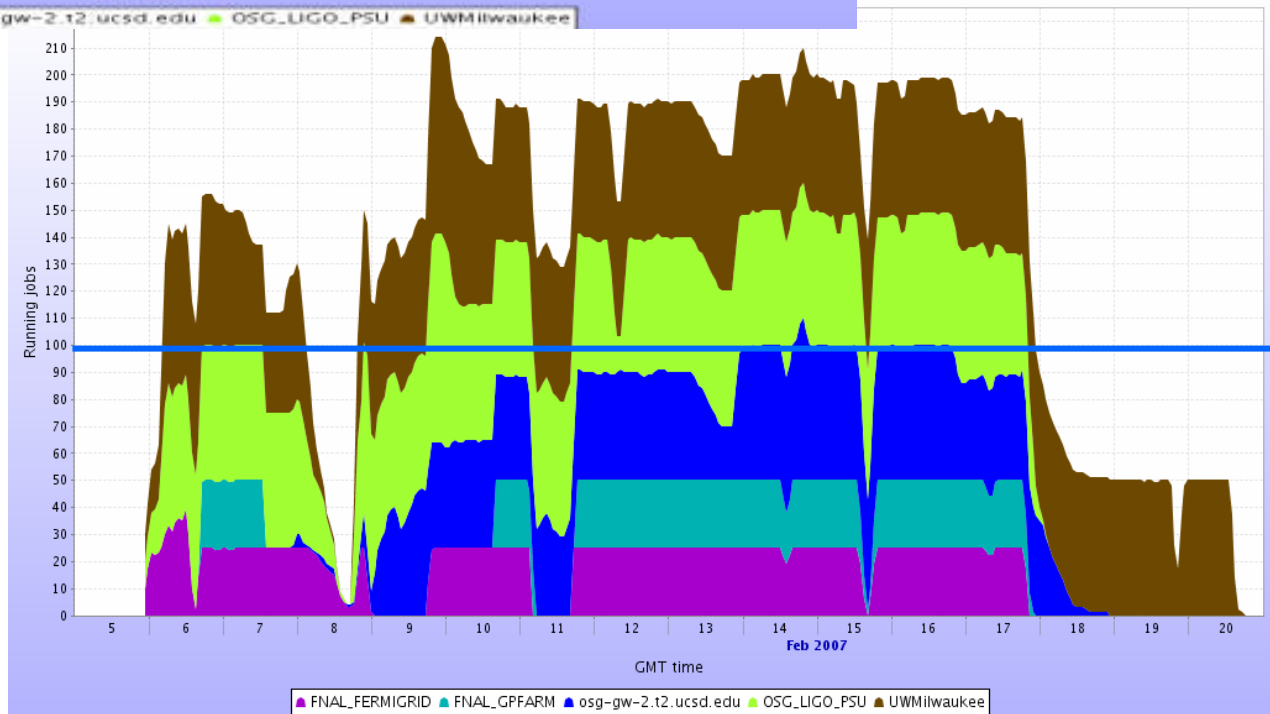


Helping to meet LIGO/OSG milestones



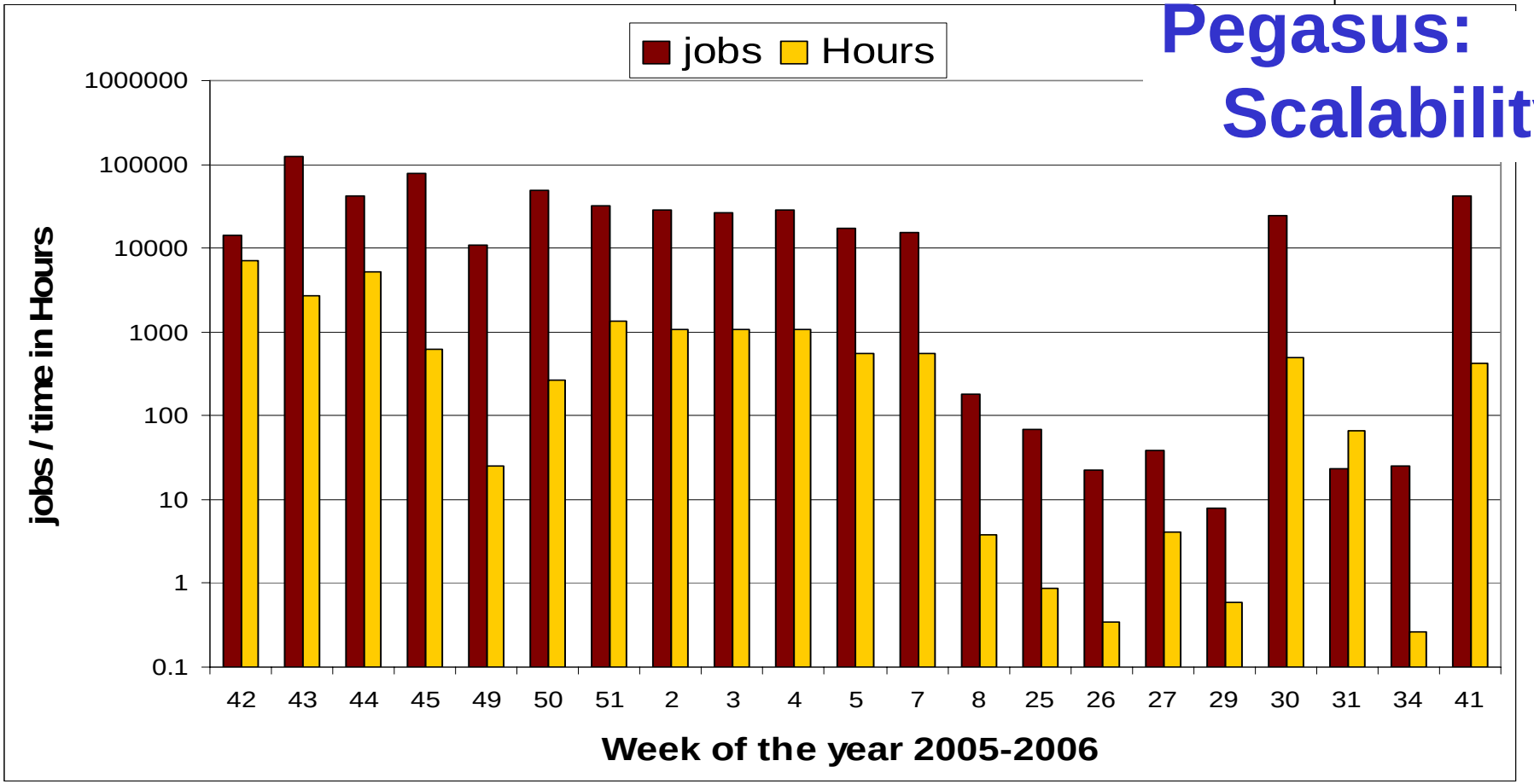
1st Milestone (month 4): Run at UCSD with DC of 25 slots for one week

2nd Milestone (month 8): Run on OSG with DC of 100 slots for one week



Work done by Kent Blackburn, David Meyers, Michael Samidi, Caltech

Pegasus: Scalability

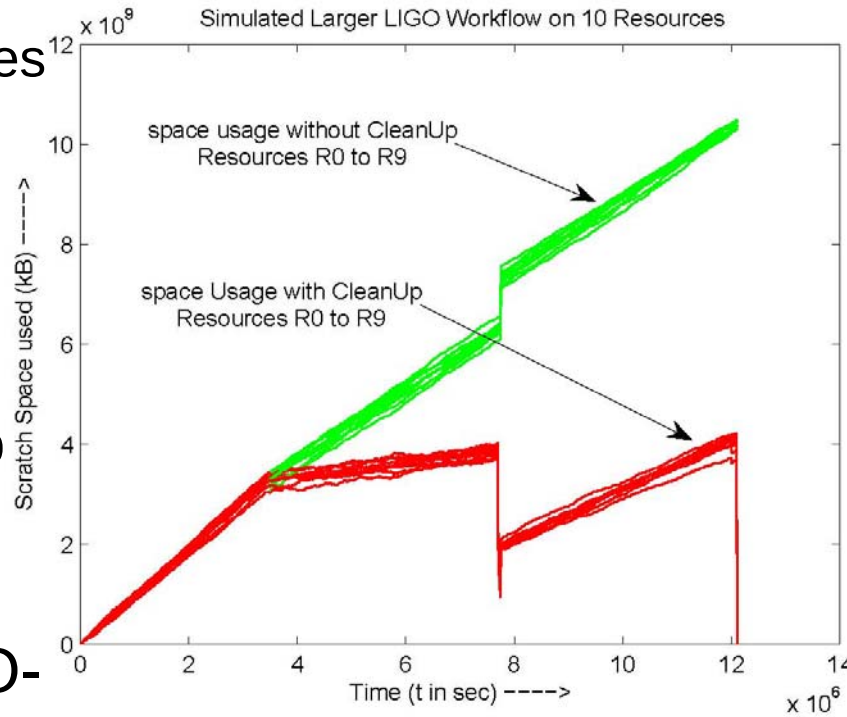


SCEC workflows run each week using Pegasus and DAGMan on the TeraGrid and USC resources. Cumulatively, the workflows consisted of over half a million tasks and used over 2.5 CPU Years.

Managing Large-Scale Workflow Execution from Resource Provisioning to Provenance tracking: The CyberShake Example, Ewa Deelman, Scott Callaghan, Edward Field, Hunter Francoeur, Robert Graves, Nitin Gupta, Vipin Gupta, Thomas H. Jordan, Carl Kesselman, Philip Maechling, John Mehringer, Gaurang Mehta, David Okaya, Karan Vahi, Li Zhao, e-Science 2006, Amsterdam, December 4-6, 2006, **best paper award**

Pegasus: Efficient data handling

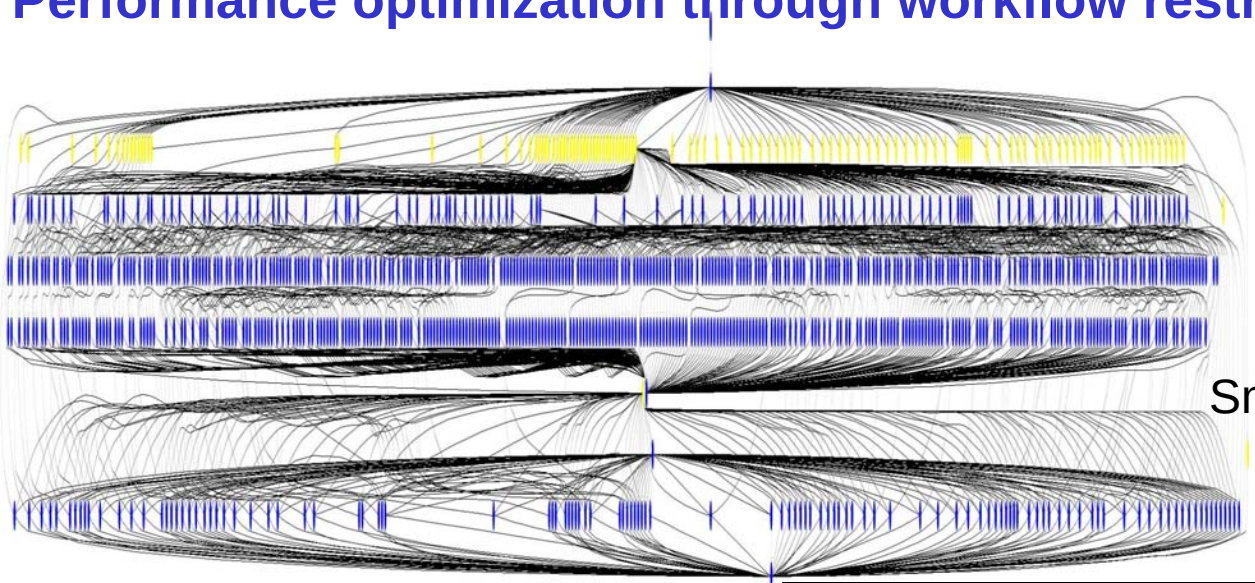
- Input data is staged dynamically, new data products are generated during execution
- For large workflows 10,000+ input files
  - Similar order of intermediate/output files
  - Not enough space-failures occur
- Solution:
  - Determine which data are no longer needed and when
  - Add nodes to the workflow do cleanup data along the way
- Benefits: simulations show up to 57% space improvements for LIGO-like workflows



- Next steps: experiment with LIGO inspiral workflow on OSG

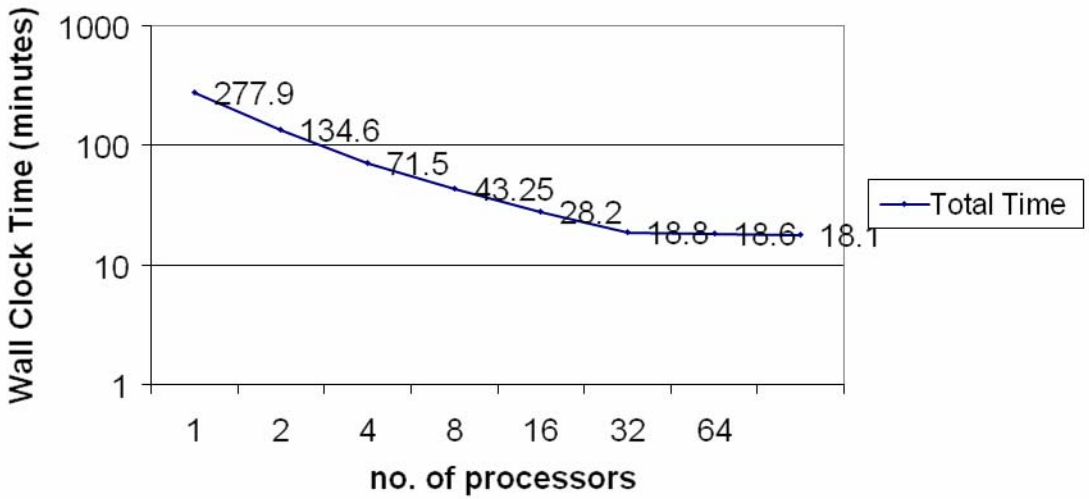
"Scheduling Data-Intensive Workflows onto Storage-Constrained Distributed Resources", A. Ramakrishnan, G. Singh, H. Zhao, E. Deelman, R. Sakellariou, K. Vahi, K. Blackburn, D. Meyers, and M. Samidi, accepted to CCGrid 2007

Performance optimization through workflow restructuring



Small 1,200 Montage Workflow

Montage application
~7,000 compute jobs in instance
~10,000 nodes in the executable workflow
same number of clusters as processors
speedup of ~15 on 32 processors

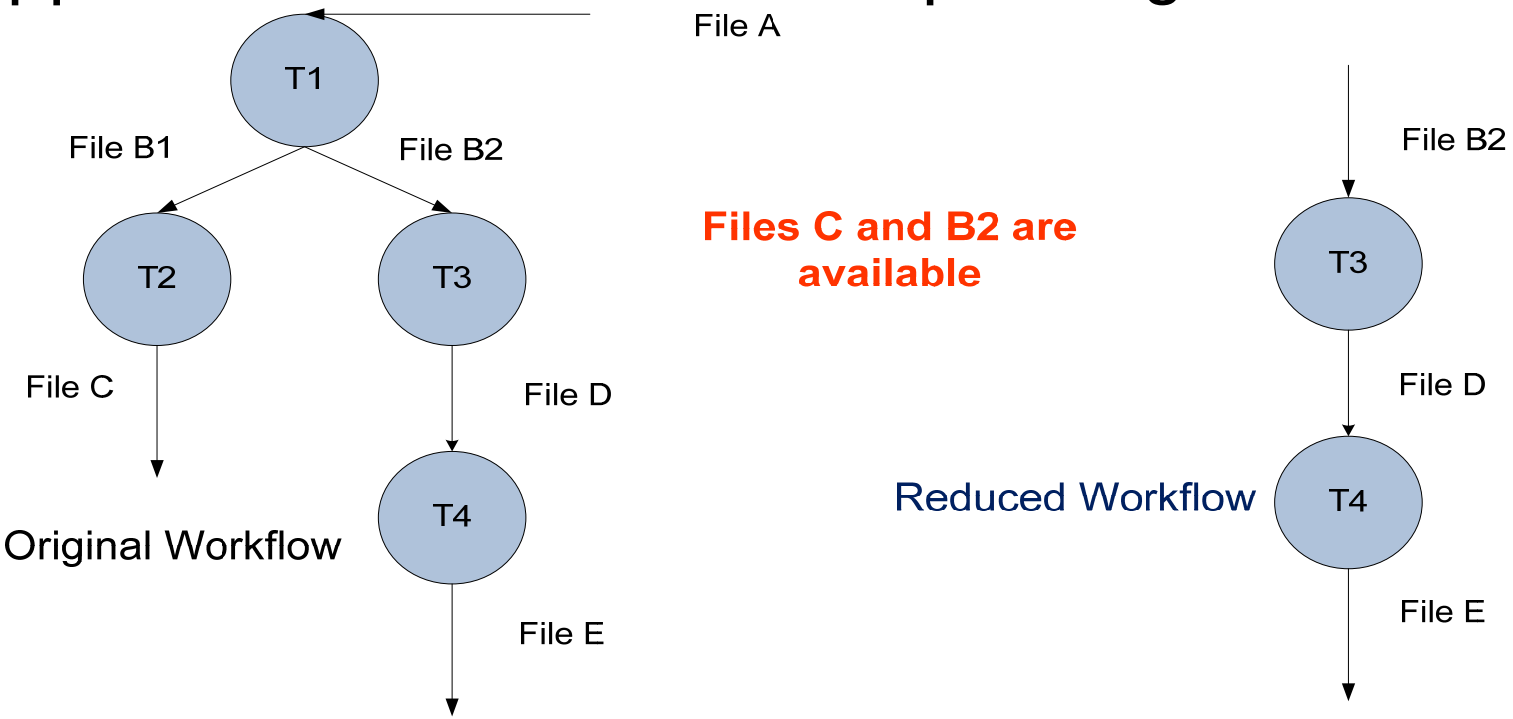


Pegasus: a Framework for Mapping Complex Scientific Workflows onto Distributed Systems, Ewa Deelman, Gurmeet Singh, Mei-Hui Su, James Blythe, Yolanda Gil, Carl Kesselman, Gaurang Mehta, Karan Vahi, G. Bruce Berriman, John Good, Anastasia Laity, Joseph C. Jacob, Daniel S. Katz, Scientific Programming Journal, Volume 13, Number 3, 2005

Data Reuse

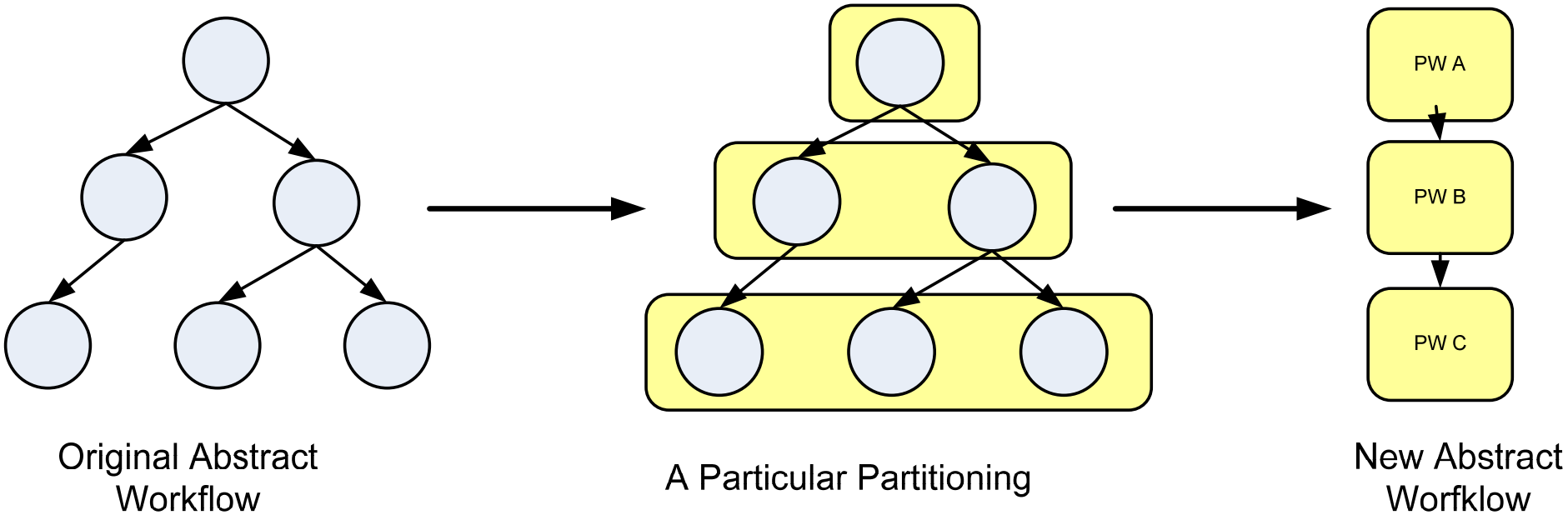
Sometimes it is cheaper to access the data than to regenerate it

Keeping track of data as it is generated supports workflow-level checkpointing



Mapping Complex Workflows Onto Grid Environments, E. Deelman, J. Blythe, Y. Gil, C. Kesselman, G. Mehta, K. Vahi, K. Backburn, A. Lazzarini, A. Arbee, R. Cavanaugh, S. Koranda, *Journal of Grid Computing, Vol.1, No. 1, 2003., pp25-39.*

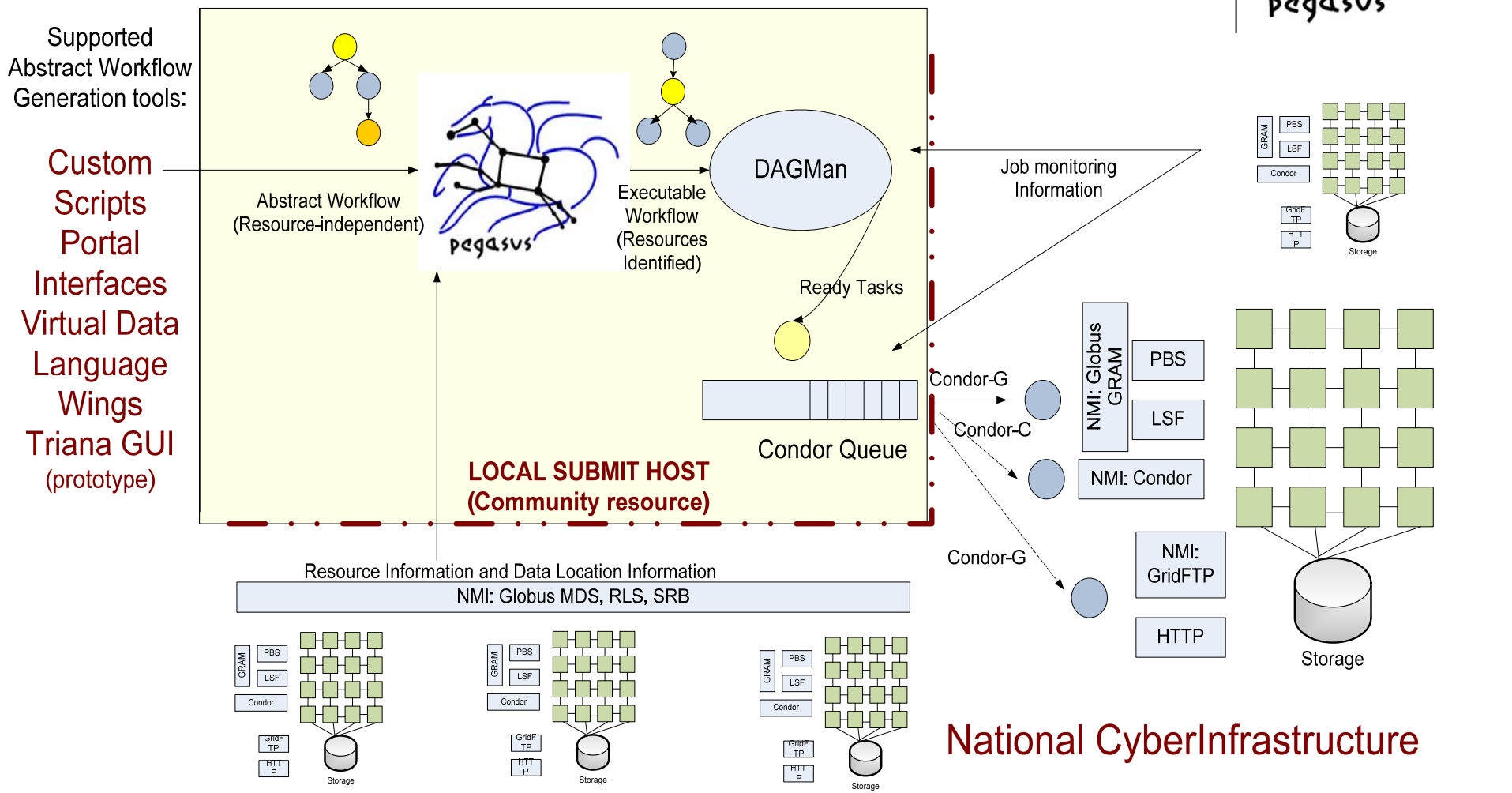
Managing execution environment changes through partitioning



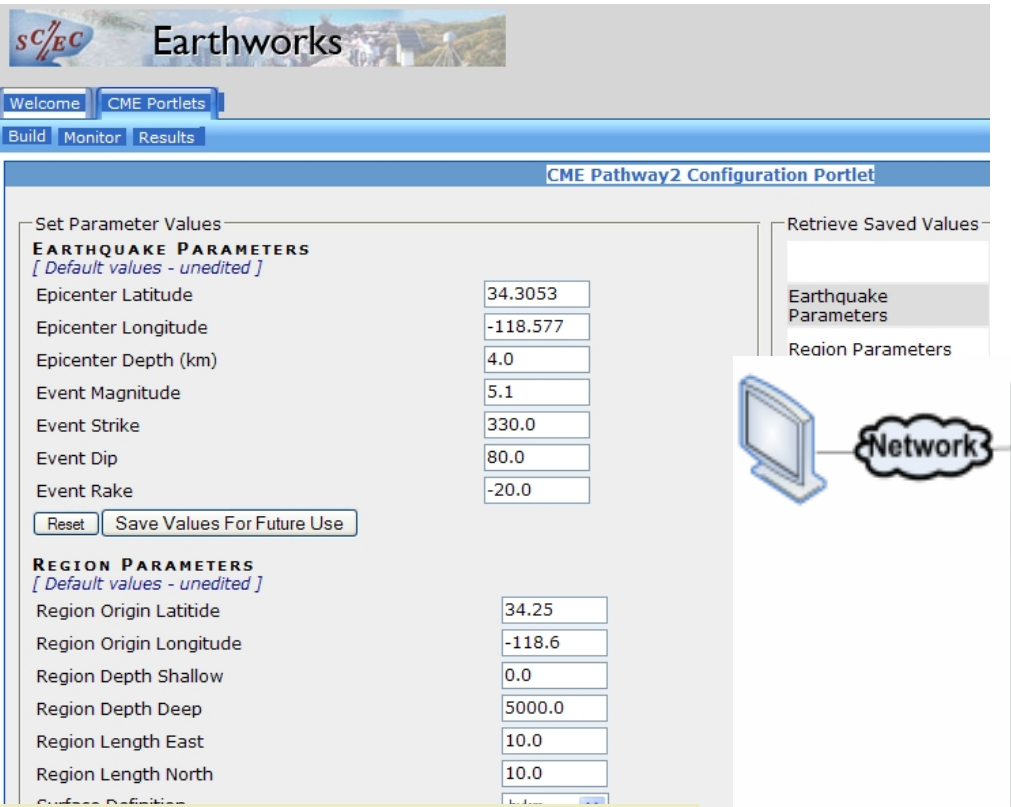
Provides reliability—can replan at partition-level
Provides scalability—can handle portions of the workflow at a time

E. Deelman, et al. Pegasus: a Framework for Mapping Complex Scientific Workflows onto Distributed Systems, Scientific Programming Journal, Volume 13, Number 3, 2005

Pegasus Deployment

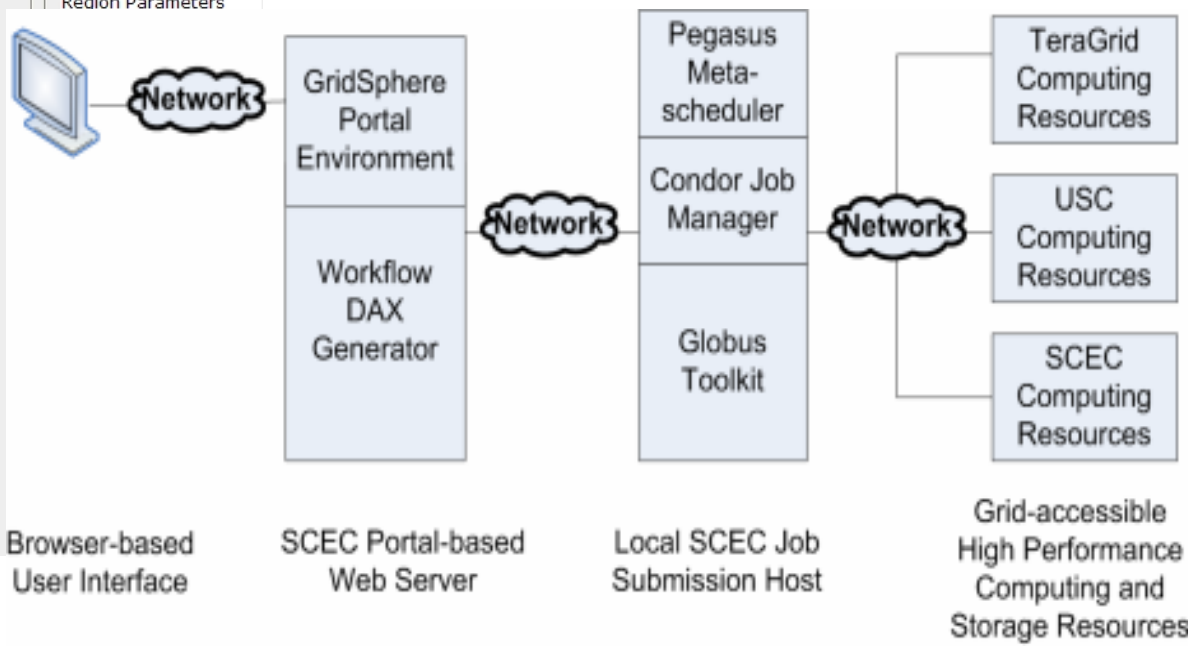


Portals, Providing high-level Interfaces



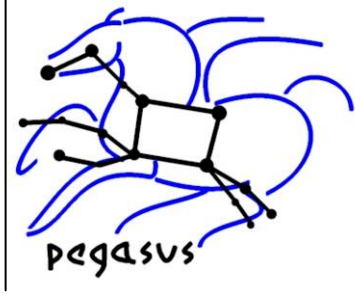
EarthWorks Project (SCEC), lead by with J. Muench P. Maechling, H. Francoeur, and others

TG Science Gateway, Washington University



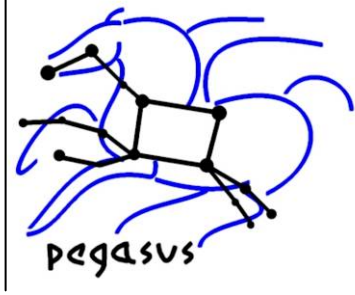
SCEC Earthworks: Community Access to Wave Propagation Simulations, J. Muench, H. Francoeur, D. Okaya, Y. Cui, P. Maechling, E. Deelman, G. Mehta, T. Jordan
TG 2006

Future OSG-focused Developments



- Working towards 3rd Milestone: (month 12): reach 1000 slots peak in MonaLisa on OSG
  - *Dynamic clean-up of completed data dependencies*—put into production
  - *SRM/dcache support for Pegasus workflows*
  - *Limit use of shared file systems / job manager fork / total number of jobs*.
    - Use of node clustering techniques
    - Support for the placement of the data and jobs in temporary OSG directories.
    - Support for local disk I/O and vertical clustering of jobs

New project targeting OSG: Windward--Yolanda Gil, ISI (PI)

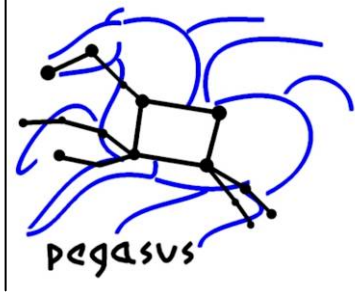


- Investigates the use of workflow technologies for Artificial Intelligence applications
  - data analysis and knowledge discovery tasks
- Explores automatic workflow component selection
- Explores the use of learning techniques to improve the performance of the workflow
- Investigates mechanisms to support autonomous and robust execution of concurrent workflows over continuously changing data

Windward: Supporting multiple workflows with real time requirements

- Workflow performance estimation ahead of the execution
- Workflows must meet deadlines (24hrs for example)
- Automated resource provisioning prior to workflow mapping
- Optimization across workflows
  - Reusing data across workflows
- Prioritization of workflows
  - Fast tracking of critical workflows

What does Pegasus do for an application?



- Provides a level of abstraction above gridftp, condor-submit, globus-job-run, etc commands
- Provides automated mapping and execution (via DAGMan) of workflow applications onto distributed resources
- Manages data files, can store and catalog intermediate and final data products
- Improves successful application execution
- Improves application performance
- Provides provenance tracking capabilities
- Supplies client-side tools
- Provides an OSG-aware workflow management tool

Acknowledgments

- LIGO: Kent Blackburn, David Meyers, Michael Samidi
- Pegasus: Gaurang Mehta, Mei-Hui Su, Karan Vahi (developers), Nandita Mandal, Arun Ramakrishnan, Tsai-Ming Tseng (students)
- Other Collaborators: Yolanda Gil, Miron Livny and the Condor team, Jihie Kim, Paul Cohen, Varun Ratnakar
- Thanks to Montage collaborators: Bruce Berriman, John Good, Dan Katz, and Joe Jacobs
- Thanks to SCEC collaborators: Tom Jordan, Robert Graves, Phil Maechling, David Okaya, Li Zhao

Relevant Links

- Pegasus: pegasus.isi.edu
  - Currently released as part of VDS and VDT
  - Standalone pegasus distribution v 2.0 coming out in April 2007, will remain part of VDT
- NSF Workshop on Challenges of Scientific Workflows : www.isi.edu/nsf-workflows06, E. Deelman and Y. Gil (chairs)
- Workflows for e-Science, Taylor, I.J.; Deelman, E.; Gannon, D.B.; Shields, M. (Eds.), Dec. 2006
- Open Science Grid: www.opensciencegrid.org
- LIGO: www.ligo.caltech.edu/
- SCEC: www.scec.org
- Montage: montage.ipac.caltech.edu/
- Condor: www.cs.wisc.edu/condor/
- Globus: www.globus.org
- TeraGrid: www.teragrid.org