June 18, 2013

# Introduction to Scientific Workflows

Kim Dillman
Purdue University

XSEDE
Extreme Science and Engineering Discovery Environment

NSF

# Introduction

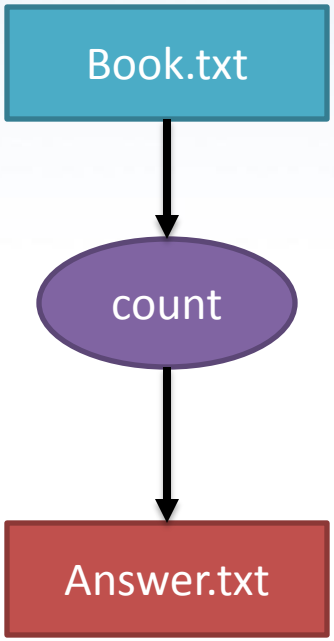
- What are Scientific Workflows?
- What does a Scientific Workflow look like?
- When to use a Scientific Workflow?
- Why would I want to use a program to manage Scientific Workflows?
- Fields of Science using Workflows
- Scientific Workflow Management Programs
- Other Info…

# What are Scientific Workflows?

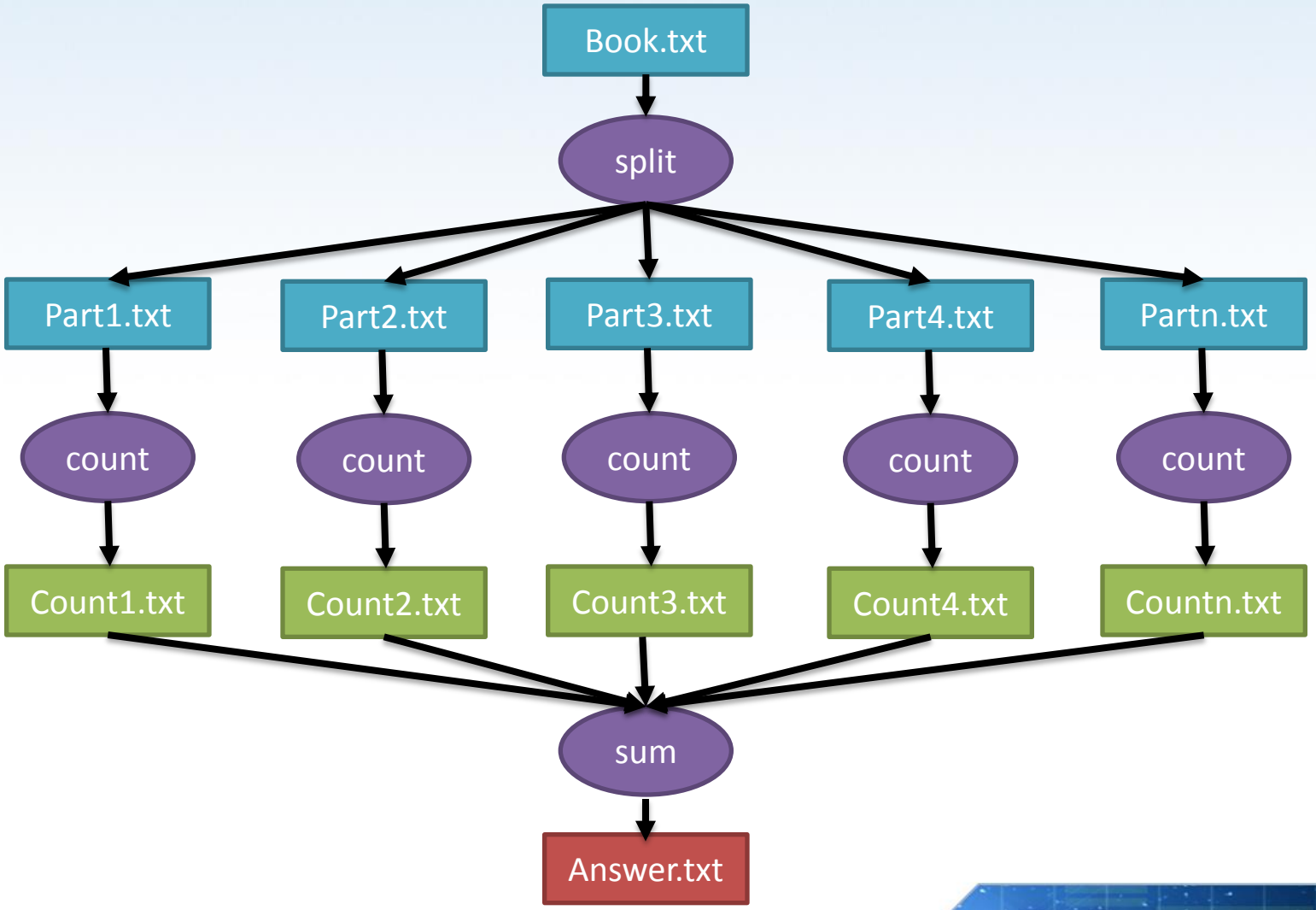
Scientific workflows allow users to easily express multi-step computational tasks, for example retrieve data from an instrument or a database, reformat the data, and run an analysis. A scientific workflow describes the dependencies between the tasks and in most cases the workflow is described as a directed acyclic graph (DAG), where the nodes are tasks and the edges denote the task dependencies. A defining property for a scientific workflow is that it manages data flow. The tasks in a scientific workflow can be everything  from short serial tasks to very large parallel tasks (MPI for example) surrounded by a large number of small, serial tasks used for pre- and post-processing.

***Information provided by [http://pegasus.isi.edu/](http://pegasus.isi.edu/)***

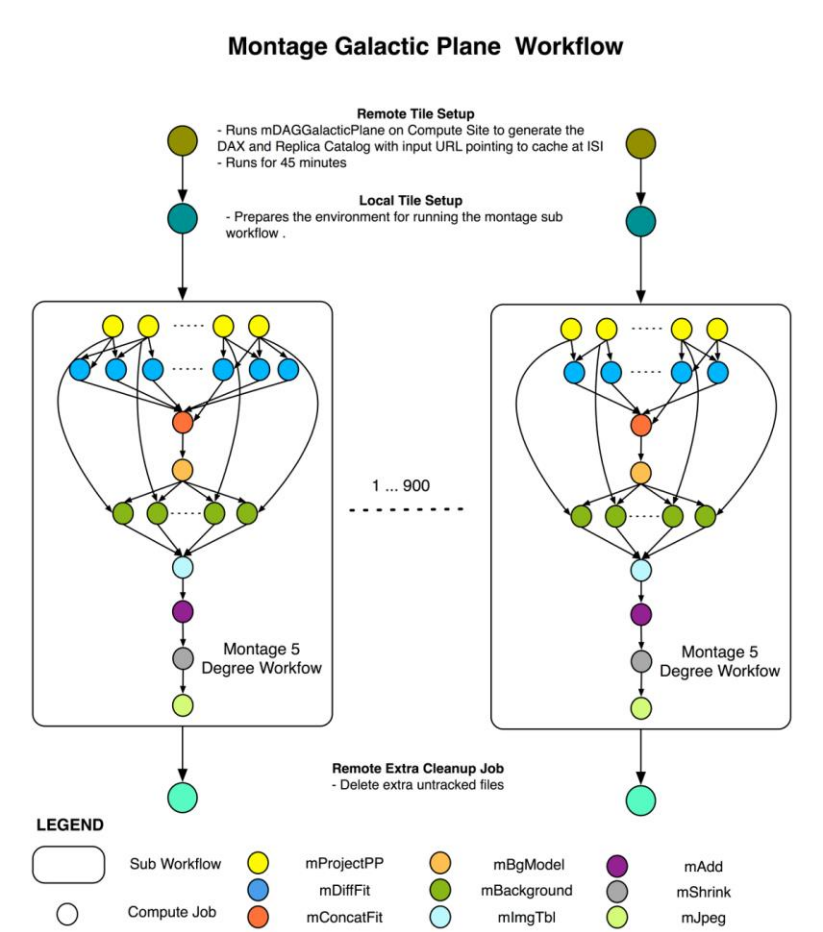
# What does a Scientific Workflow look like?

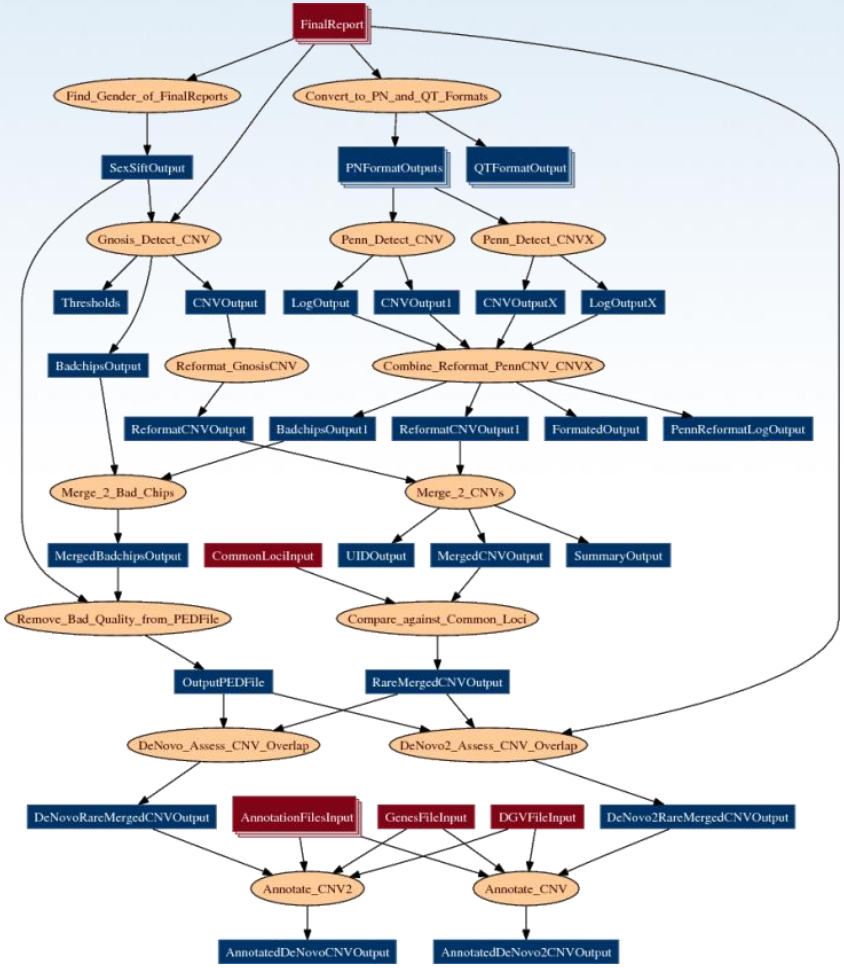


# What does a Scientific Workflow look like?



# What does a Scientific Workflow look like?




http://pegasus.isi.edu/applications/galactic-plane

http://pegasus.isi.edu/applications/combined%20cnv_detection

# When to use a Scientific Workflow?

- Lots of independent tasks that can be run in parallel
- Tasks that require pre and/or post processing steps
- Tasks that require monitoring for failures and/or cleanup work
- Processing of large data using many independent tasks
- Complicated series of parallel processing steps
- And more…

# Why would I want to use a program to manage Scientific Workflows?

- Monitor a large number of tasks for successful completion (automatic restart)
- Run a complex scientific workflow in different environments
- The ability to “scale up” the workflow
- Performance optimization of workflows
- Obtain statistics and debugging information
- Data management for workflow
- Error recovery
- And more….

# Fields of Science using Workflows
## (some examples)

- Astronomy
- Bioinformatics
- Botany
- Chemistry
- Climate Modeling
- Computer Science
- Earthquake Science
- Genome Analysis
- Helioseismology
  - "the study of the propagation of wave oscillations, particularly acoustic pressure waves, in the Sun"
- Limnology
  - also called freshwater science, is the study of inland waters
- Neuroscience
- Ocean Science
- Physics

# Scientific Workflow Management Programs
## (some)

- DAGMan (Condor)
  - "Directed Acyclic Graph MANager"
  - http://research.cs.wisc.edu/htcondor/dagman/dagman.html
- Pegasus
  - http://pegasus.isi.edu/
- Kepler
  - https://kepler-project.org/
- Taverna
  - http://www.taverna.org.uk/
- BigJob (SAGA)
  - "Simple Api for Grid Applications"
  - https://github.com/saga-project/BigJob/wiki
  - BigJob on XSEDE:
    - http://kb.iu.edu/data/bcnm.html
- Makeflow
  - "Make" + "Workflow"
  - http://www3.nd.edu/~ccl/software/makeflow/
- Airavata (Apache)
  - http://airavata.apache.org/

# Upcoming Tutorials…

- XSEDE13
  - Tutorial - “An Introduction to BigJob and BigData – Using Pilots to Manage Computation and Data on XSEDE”
  - Monday, July 22, 8 AM – noon
  - http://xsede13.sched.org/event/497b4f190c3e586f7708eac3633eeb97?iframe=no&w=900&sidebar=yes&bg=no#
- Campus Champions Tech Talks
  - Pegasus on XSEDE and OSG
  - TBD
- ?

# More “online” tutorials…

- Pegasus
  - Tutorial: http://pegasus.isi.edu/wms/docs/latest/tutorial.php
  - Tutorial VM download: http://pegasus.isi.edu/pegasus/download?filename=4.2/PegasusTutorialVM-4.2.2.zip
  - OSG/XSEDE: http://pegasus.isi.edu/news/osg-xsede
- BigJob
  - Tutorial: https://github.com/saga-project/BigJob/wiki/BigJob-Tutorial
  - Info: “BigJob on XSEDE” https://www.xsede.org/bigjob
- Condor/DAGMan
  - Tutorials : http://research.cs.wisc.edu/htcondor/tutorials/

# Links of interest…

- Examples of scientific workflows
  - https://pegasus.isi.edu/applications
- Examples of scientific workflow management programs on Wikipedia
  - http://en.wikipedia.org/wiki/Scientific_workflow_system#Examples
- Scientific Workflows Survey Page (doesn't seem recent…)
  - http://www.extreme.indiana.edu/swf-survey/


Questions?
Our reach will forever
exceed our grasp, but,
in stretching our horizon,
we forever improve our world.
XSEDE
Extreme Science and Engineering
Discovery Environment