Why Pegasus?

Automates Complex, Multi-stage Processing Pipelines

Enables Parallel, Distributed Computations

Automatically Executes Data Transfers

Reusable, Aids Reproducibility

Records How Data was Produced (Provenance)

Handles Failures with to Provide Reliability

Keeps Track of Data and Files

Ensures Data Integrity during workflow execution
How to build workflows with Pegasus?

Try our **self-guided tutorial** available in the Pegasus website!
What information does Pegasus need? from the abstraction to execution

**Site Catalog**
Describes the execution sites where the workflow jobs are to be executed
**automatically created for default local and condorpool sites**

**Transformation Catalog**
Describes the executables (called "transformations") used by the workflow

**Replica Catalog**
Describes all of the input data stored on external servers
Portable Description

Users do not worry about low level execution details

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
Stage-out generated output data

Registration Job
Registers the workflow output data
Could you talk more about execution?

**COMPUTE**

- Desktop/Laptop
  - HTCondor, PBS, Slurm, LSF, SGE

- Local/Campus Cluster
  - XSEDE, TACC, ORNL, ANL, NERSC, etc.

- HPC Systems
  - Amazon AWS, Google Cloud, Chameleon Cloud, etc.

- Clouds
  - Open Science Grid

**STORAGE**

- Transfer Protocols
  - HTTP, SCP, GridFTP, Globus Online, iRods, Amazon S3, Google Storage, SRM, FDT, Stashcp, Rucio, cp, ln -s

- File Systems
  - Shared and non-shared file systems, and HTCondor I/O

- Parallel Transfers
  - Automated Retries

- Task Clustering
  - Reduces execution overhead

- Data Reuse
  - Avoids re-computations

- Fault-tolerance
  - Checkpoints, Retries, Rescue DAGs

- Large-scale Workflows
  - Hierarchical execution
What to do next?

Grab us during the **break**

Come to **office hours** @12:30pm PST / 3:30pm EST

Do a **self-guided tutorial**

https://pegasus.isi.edu/documentation/user-guide/tutorial.html

We are happy to learn about your application and are here to help