CI/CS WORKSHOP THE COMMUNITY TOGETHER

End to End Workflow Monitoring and Execution

Ryan Tanaka Programmer Analyst USC Information Sciences Institute

CI/CS WORKSHOP ·I

| **%℃ICoE**≣

The Pegasus Workflow Management System

- Bridges the scientific domain and execution environment by mapping high level workflow descriptions onto distributed resources
- Enables scientists to:
 - **Automate** their work, as portable workflows
 - **Recover** from failures at runtime
 - **Debug** failures in their computations
- Built on top of HTCondor, a proven DHTC workhorse









Outline

Introducing the Pegasus WMS

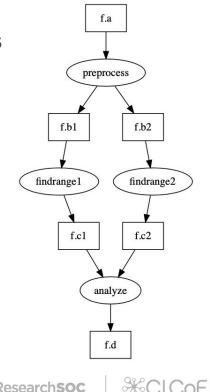
- Concepts
- Features
- **Production Use**





Workflows as DAGs

- Workflows are multi-step computational tasks organized as directed acyclic graphs (DAG)
- Define abstract workflow using one of our Python, Java, or R APIs
 - Abstract in the sense that users need not map jobs to resources or create file transfer jobs for input and output files
 - Pegasus will plan the abstract workflow into an executable workflow
- Example..



Defining Workflow Inputs



Defining Workflow Executables



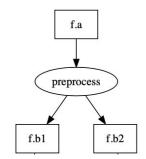
wf = Workflow("blackdiamond")



·IJĮI• Research**soc** | ⅔CICOE ₪

wf = Workflow("blackdiamond")

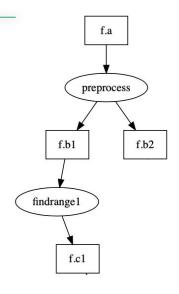
```
fb1 = File("f.b1")
fb2 = File("f.b2")
job_preprocess = Job(preprocess)\
                          .add_args("-a", "preprocess", "-T", "3", "-i", fa, "-o", fb1, fb2)\
                         .add_inputs(fa)\
                         .add_outputs(fb1, fb2)
```





·IIII Research**soc** | ⅔℃| COE

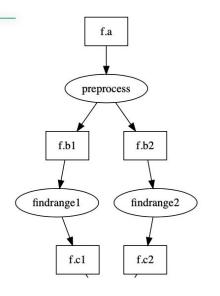
```
wf = Workflow("blackdiamond")
```





```
wf = Workflow("blackdiamond")
```

```
fb1 = File("f.b1")
fb2 = File("f.b2")
job preprocess = Job (preprocess) \
                      .add args("-a", "preprocess", "-T", "3", "-i", fa, "-o", fb1, fb2)
                      .add inputs(fa) \
                      .add outputs (fb1, fb2)
fc1 = File("f.c1")
job findrange 1 = Job(findrange) \setminus
                      .add args("-a", "findrange", "-T", "3", "-i", fb1, "-o", fc1)
                      .add inputs(fb1) \
                      .add outputs (fc1)
fc2 = File("f.c2")
job findrange 2 = Job(findrange) \setminus
                      .add args("-a", "findrange", "-T", "3", "-i", fb2, "-o", fc2)
                      .add inputs(fb2)\
                      .add outputs (fc2)
```



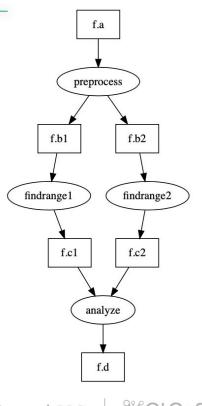


·illi Research**soc**



```
wf = Workflow("blackdiamond")
```

```
fb1 = File("f.b1")
fb2 = File("f.b2")
job preprocess = Job(preprocess) \
                      .add args("-a", "preprocess", "-T", "3", "-i", fa, "-o", fb1, fb2)
                      .add inputs(fa) \
                      .add outputs (fb1, fb2)
fc1 = File("f.c1")
job findrange 1 = Job(findrange) \setminus
                      .add args("-a", "findrange", "-T", "3", "-i", fb1, "-o", fc1)\
                      .add inputs(fb1) \
                      .add outputs (fc1)
fc2 = File("f.c2")
job findrange 2 = Job(findrange) \
                      .add args("-a", "findrange", "-T", "3", "-i", fb2, "-o", fc2)
                      .add inputs(fb2)\
                      .add outputs (fc2)
fd = File("f.d")
job analyze = Job(analyze) \
               .add args("-a", "analyze", "-T", "3", "-i", fc1, fc2, "-o", fd)
               .add inputs(fc1, fc2) \
```

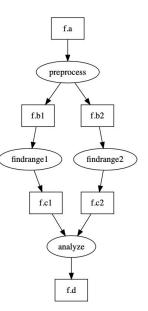


CI/CS WORKSHOP

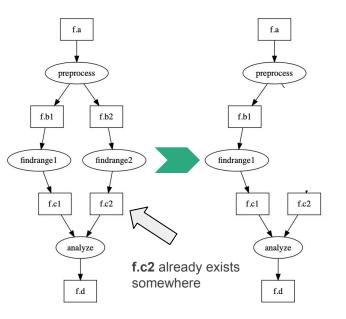
.add outputs (fd)

III Research**SOC**

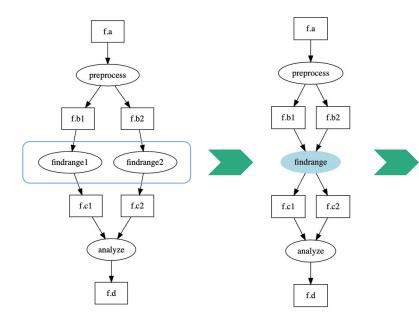
- Workflow planning process
 - Data reuse module will optionally prune jobs for which output files already exist
 - Task clustering optimizations may be performed for small independent jobs
 - Mapping jobs onto physical compute resources
 - Add auxiliary jobs for data staging, cleanup, file registration, etc.
- Generated executable workflow submitted to through HTCondor to be run
 CI/CS WORKSHOP



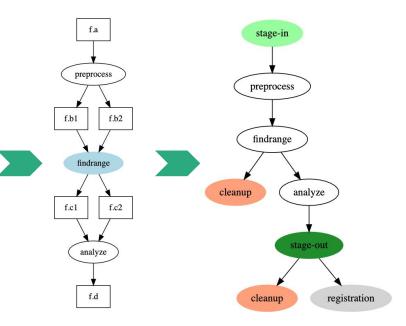
- Workflow planning process
 - Data reuse module will optionally prune jobs for which output files already exist
 - Task clustering optimizations may be performed for small independent jobs
 - Mapping jobs onto physical compute resources
 - Add auxiliary jobs for data staging, cleanup, file registration, etc.
- Generated executable workflow submitted to through HTCondor to be run
 CI/CS WORKSHOP



- Workflow planning process
 - Data reuse module will optionally prune jobs for which output files already exist
 - Task clustering optimizations may be performed for small independent jobs
 - Mapping jobs onto physical compute resources
 - Add auxiliary jobs for data staging, cleanup, file registration, etc.
- Generated executable workflow submitted to through HTCondor to be run
 CI/CS WORKSHOP



- Workflow planning process
 - Data reuse module will optionally prune jobs for which output files already exist
 - Task clustering optimizations may be performed for small independent jobs
 - Mapping jobs onto physical compute resources
 - Add auxiliary jobs for data staging, cleanup, file registration, etc.
- Generated executable workflow submitted to through HTCondor to be run
 CI/CS WORKSHOP





Outline

Introducing the Pegasus WMS



- **Features**
- **Production Use**





Data Staging Configurations

- HTCondor I/O (HTCondor pools, OSG, ...)
 - Worker nodes do not share a file system
 - Data is pulled from/pushed to the submit host via HTCondor file transfers
 - Submit host used as staging site
- Non-shared File System (Clouds, OSG, ...)
 - Worker nodes do not share a file system
 - Data is pulled / pushed from a staging site, possibly not co-located with the computation

I. Research**SOC**

- Shared File System (HPC sites, XSEDE, campus clusters, ...)
 - I/O directly against the shared file system

pegasus-transfer internal file transfer utility

- Directory creation, file removal
- Two stage transfers between incompatible protocols
 - E.g., GridFTP to S3 is executed as: GridFTP to local file, local file to S3
- Parallel transfers
- Automatic retries
- Credential management

CI/CS WORKSHOP

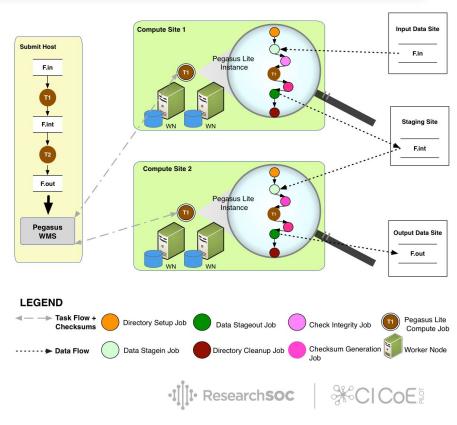
HTTP SCP GridFTP Globus Online iRods Amazon S3 Google Storage SRM FDT Stashcp Rucio ср ln -s

IIII Research**SOC**



Automatic Integrity Checking

- Pegasus automatically performs integrity checksums on input files right before jobs begin
 - Checksums can be specified for inputs coming from external sources
 - All intermediate and output files have checksums which are generated and tracked within the system
- Checksum validation failure results in job failure



Monitoring and Debugging Tools CLI

• pegasus-status

- View current status of running workflow
- View summary of jobs and sub workflows
- pegasus-analyzer
 - View errors from any failed jobs
- pegasus-statistics
 - View summary of workflow statistics
 - Succeeded jobs, failed jobs, retries, workflow walltime, etc.

\$ pega	asus-s	tatus	-l /Woi	rkflow/	'dags/d	irecto	ory		
STAT	IN ST	ATE J	OB						
Run	07	:01 1	evel-3-	-0					
Run	06	:25	-sleep	p IDOOC	005				
Run	06	:20	\ subda	ax leve	el-2 ID	000004	1		
Run	05	:44	-s]	leep II	0000003				
Run	05	:39	\ sı	ıbdax l	evel-1	ID000	002		
Run	05	:03	_ \	_sleep	_ID000	001			
Summai	су: б	Condor	jobs t	total ((R:6)				
UNRDY	READY	PRE	IN_Q	POST	DONE	FAIL	%DONE	STATE	DAGNAME
0	0	0	1	0	1	0	50.0	Running	level-2_I
0	0	0	2	0	1	0	33.3	Running	level-2_I
0	0	0	3	0	1	0	25.0	Running	*level-3-0
0	0	0	6	0	3	0	33.3		TOTALS (9
Summ	ary: 3	B DAGs	total	(Runni	ng:3)				

CI/CS WORKSHOP

III Research**soc**

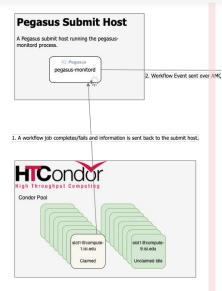


Monitoring and Debugging Tools dashboard

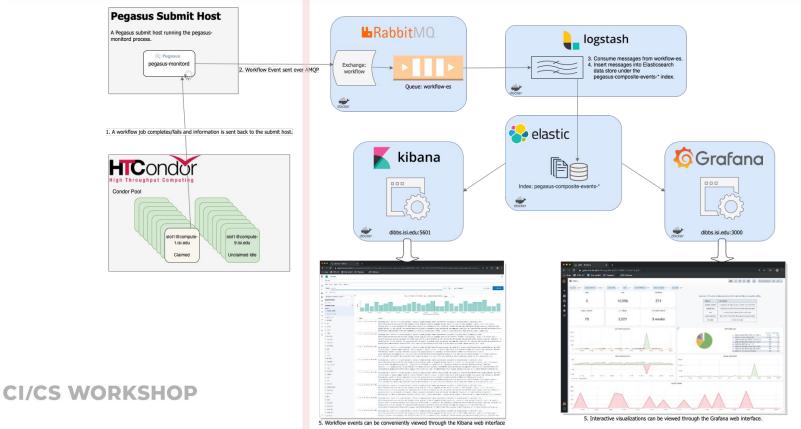
兪				
Workflow	Listing			
		Successful 8 Running 0 faled. 2 Successful 8 Running 0 Successful		
Show results fo	or all +		Search	h:
Show 10 0 (Workflow		Submit Directory o	Search	n: Submitted On
Show 10 0 e	entries	Submit Directory o		
Show 10 0 (Workflow Label	entries Submit Host ≎		State 0	Submitted On Fri, 23 Oct 2015 16:04:0
Show 10 0 C Workflow 0 Label 0	Submit Host o workflow.isi.edu	/nfs/ccg3/ccg/home/examples/split/split/run0006	State ©	Submitted On Fri, 23 Oct 2015 16:04:1 Fri, 23 Oct 2015 15:56:1
Show 10 0 C Workflow 0 split split	Submit Host o workflow.isi.edu	/nfs/ccg3/ccg/home/examples/split/split/run0006 /nfs/ccg3/ccg/home/examples/split/split/run0004	State O Running Failed	Submitted On Fri, 23 Oct 2015 16:04:1 Fri, 23 Oct 2015 15:56:1 Fri, 23 Oct 2015 15:50:
Show 10 0 C	Submit Host o workflow.isi.edu workflow.isi.edu workflow.isi.edu	/nfs/ccg3/ccg/home/examples/split/split/run0008 /nfs/ccg3/ccg/home/examples/split/split/run0004 /nfs/ccg3/ccg/home/examples/diamond/diamond/run0002	State Running Failed Successful	Submitted On Fri, 23 Oct 2015 16:04:1 Fri, 23 Oct 2015 15:56:1 Fri, 23 Oct 2015 15:50: Fri, 23 Oct 2015 15:41:
Show 10 0 4	Submit Host workflow.isi.edu workflow.isi.edu workflow.isi.edu workflow.isi.edu	/mfs/ccg3/ccg/homs/examples/split/split/un0006 /mfs/ccg3/ccg/homs/examples/split/split/un0004 /mfs/ccg3/ccg/homs/examples/diamond/diamond/un0002 /mfs/ccg3/ccg/homs/examples/diamond/liamond/un0003	State O Running Failed Successful Failed	Submitted On Fri, 23 Oct 2015 16:04:0 Fri, 23 Oct 2015 15:56:0 Fri, 23 Oct 2015 15:50:1 Fri, 23 Oct 2015 15:04:2 Fri, 23 Oct 2015 15:04:2
Show 10 0 4	Submit Host o Submit Host o workflow.isi.edu workflow.isi.edu workflow.isi.edu workflow.isi.edu	/nfs/ccg3/ccg/homa/examples/split/split/vn0006 /nfs/ccg3/ccg/homa/examples/split/split/vn0004 /nfs/ccg3/ccg/homa/examples/daimond/ulamond/vn0002 /nfs/ccg3/ccg/homa/examples/split/split/vn0003 /nfs/ccg3/ccg/homa/examples/split/split/vn0002	State C Running Failed Successful Failed Successful	Submitted On Fri, 23 Oct 2015 16:04:4 Fri, 23 Oct 2015 16:56:4 Fri, 23 Oct 2015 16:56:0 Fri, 23 Oct 2015 16:41: Fri, 23 Oct 2015 16:04:2 Fri, 23 Oct 2015 16:00:2
Show 10 0 0 Workflow 0 split split split split process	submit Heat o workflow.isi.edu workflow.isi.edu workflow.isi.edu workflow.isi.edu workflow.isi.edu workflow.isi.edu	/nfs/ccg3/ccg/homs/examples/split/split/vni0006 /nfs/ccg3/ccg/homs/examples/split/split/vni0004 /nfs/ccg3/ccg/homs/examples/daimond/un0002 /nfs/ccg3/ccg/homs/examples/aplit/split/vni0003 /nfs/ccg3/ccg/homs/examples/aplit/split/vni0002 /nfs/ccg3/ccg/homs/examples/aplit/split/vni0002	State C Running Failed Successful Successful Successful	Submitted On Fri, 23 Oct 2015 16:04:4 Fri, 23 Oct 2015 15:56:4 Fri, 23 Oct 2015 15:56: Fri, 23 Oct 2015 15:54: Fri, 23 Oct 2015 15:04: Fri, 23 Oct 2015 15:00: Fri, 23 Oct 2015 15:00:
Show 10 0 0 Workflow 0 split	Submit Host O workflow.isi.edu workflow.isi.edu workflow.isi.edu workflow.isi.edu workflow.isi.edu workflow.isi.edu workflow.isi.edu	Infs/ccg3/ccg/home/examples/split/split/vni0006 /nfs/ccg3/ccg/home/examples/split/split/vni0004 /nfs/ccg3/ccg/home/examples/split/split/vni0003 /nfs/ccg3/ccg/home/examples/split/split/vni0003 /nfs/ccg3/ccg/home/examples/split/split/vni0002 /nfs/ccg3/ccg/home/examples/split/split/vni0001	State C Running Failed Successful Successful Successful	Submitted On

pegasus-dashboard / Workflow Workflow Details 1145e2d5-ad2f-45d6-a3ce-4bd68499d8af Summary Files 6 Metadata 2 Label diamond root-wf Type Successfu Progress Submit Host cartman User bamboo Submit Directory C // /ifs1/software/bamboo/data/xml-data/build-dir/PEGASUS-WT-T39A/test/core/039-bl. **DAGMan Out File** & diamond-0.dag.dagman.out Wall Time 5 mins 9 secs **Cumulative Wall Time** 5 mins 52 secs Job Status (Per Workflow) Job Status (Entire Workflow) Unsubmitted: 0 Jobs: 0 Workflows: 0 Enilad: 0 Jobs: 0 Workflows: 0 Total: 0 Jobs: 26 Workflows: 0 Successful: 26 Total: 26 Unsubmitted Eailed Successful Running Failed Successful Charts Statistics Sub Workflows Failed Running Successful Failing Show 10 entries Search: Job Name Time Taken analyze_ID0000004 1 min an up local level 3.0 5 secs I. Research**SOC**

Monitoring and Debugging Tools AMQP endpoint



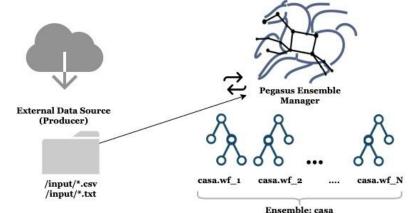
Monitoring and Debugging Tools AMQP endpoint



CICOE

Ensemble Manager workflow management & dynamic triggering

- Service for managing collections of workflows called ensembles
- Allows for throttling of concurrent planning and running workflows
- Support for triggering of new workflow runs based on arrival of new input files which match one or more given patterns





Outline

-Introducing the Pegasus WMS

Concepts

Features

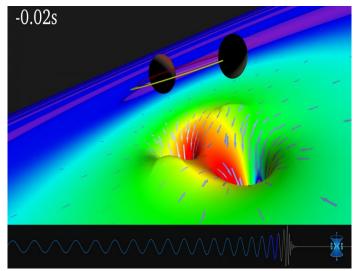
Production Use

CI/CS WORKSHOP

LIGO PyCBC Workflows for Gravitational Wave Science

- Laser Interferometer Gravitational Wave Observatory
 - Facility for gravitational wave research
 - Methods:
 - PyCBC software package
 - Pegasus WMS workflows
 - Compute using OSG, XSEDE, etc.

What do these workflows look like..

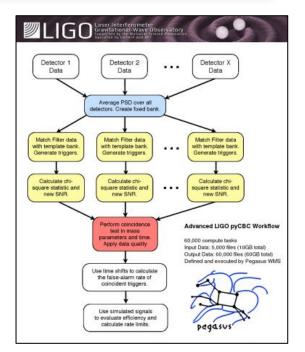


0.2 Second before the black holes collide. Image credit: SXS/LIGO



LIGO PyCBC Workflows for Gravitational Wave Science

- Advanced PyCBC Workflows
 - 40,000 compute tasks
 - 1,100 input files
 - 63 output files
 - 725 GB processed data
 - Compute: LIGO Data Grid, OSG, EGI, XSEDE



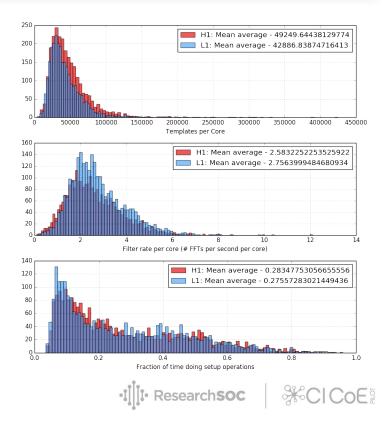
Advanced LIGO pyCBC Workflow. Image Credit: Samantha Usman, Duncan Brown et al

III: Research**soc**



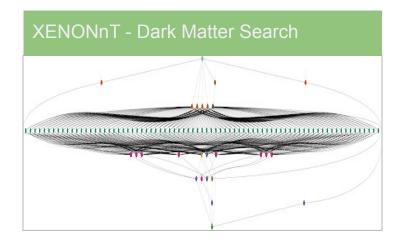
LIGO PyCBC Workflows for Gravitational Wave Science

- Plots typically generated as part of post run analysis
- Using the AMQP data collection setup, these charts are able to be updated live as jobs complete, affording LIGO researchers better monitoring capabilities of the PyCBC workflow runs

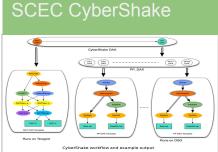


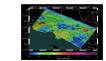
Other Production Use

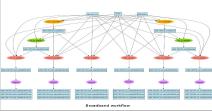
In the last 12 months, Pegasus users ran 240K workflows, 145M jobs



CI/CS WORKSHOP

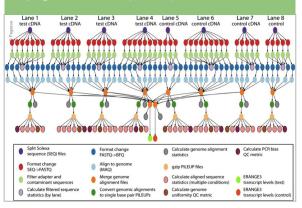






Epigenomics (USC)

I. Research**SOC**





Pegasus est. 2001

Automate, recover, and debug scientific computations

Get Started

Pegasus Online Office Hours

https://pegasus.isi.edu/blog/online-pegasus-office-hours

Bi-monthly basis on the second Friday of the month, where we address user questions and also apprise the community of new developments. Pegasus Website https://pegasus.isi.edu/
 Users Mailing Liet

Users Mailing List pegasus-users@isi.edu

Pegasus Website

pegasus-support@isi.edu

USC Viterbi Information Sciences Institute

Questions?

CI/CS WORKSHOP

•Ijjl • Research**soc**

%℃ICoE§

Thank you!

tanaka@isi.edu

CI/CS WORKSHOP

•J]]I• Research**soc**

%CICoE₽