Using Condor for On-line Data Analysis within the LIGO Scientific Collaboration

Kipp Cannon, University of Wisconsin Milwaukee

March 15, 2005

Condor-based LIGO Data Analysis

- Stand-alone processing tasks are submitted to compute clusters using Condor.
- Complex analysis pipelines are sequenced using DAGMan.
- Has traditionally been performed "off-line".
- Latencies of several months.

However

- Instrument operators require rapid feed-back on the quality of the data.
- Soon instruments will be running continuously.

We Need "On-line" Data Analysis

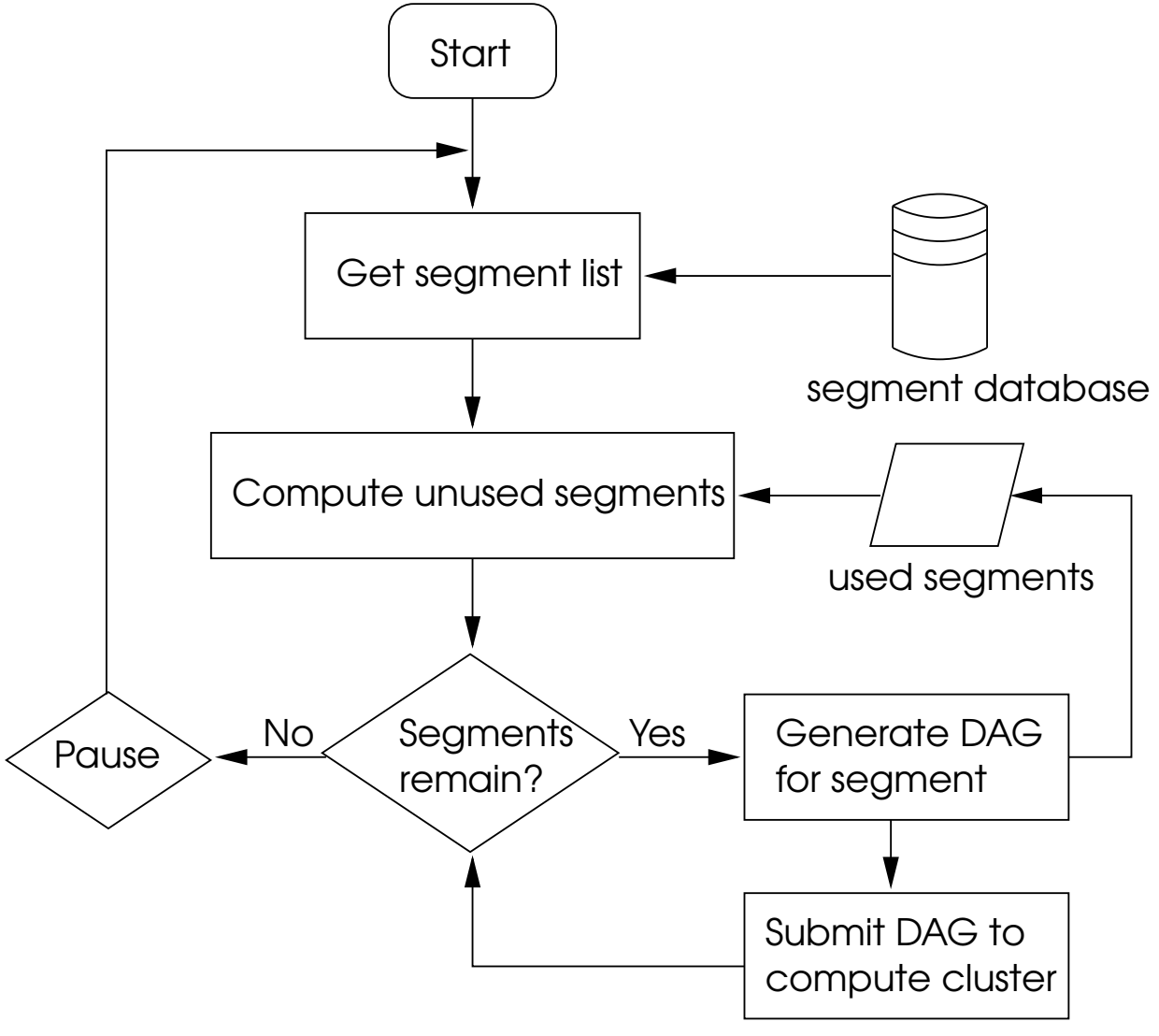
- Low(er) latency, and continuous, analysis.
- Sequencing the same complex work flows seen in off-line analyses.

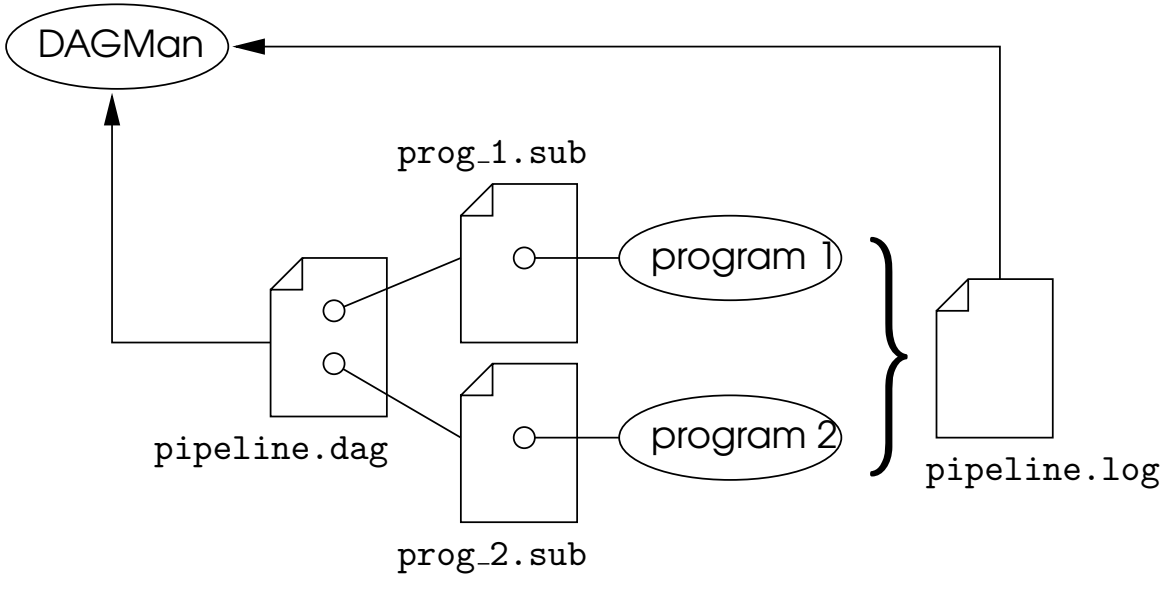
On-Line Software Goals

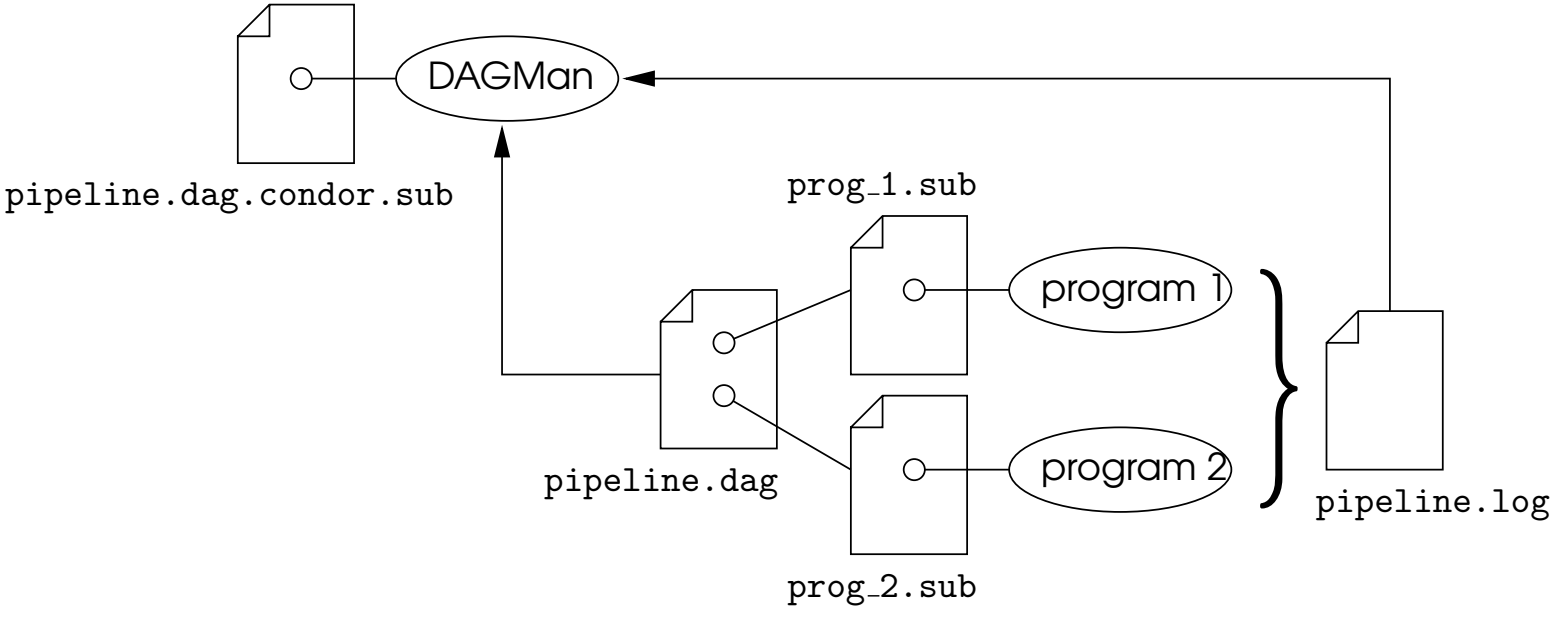
- Latencies of hours.
- Minimize the developer cost in converting an off-line analysis to an on-line analysis.
- Be robust against software and hardware failures.
- Provide status reports and error notification.

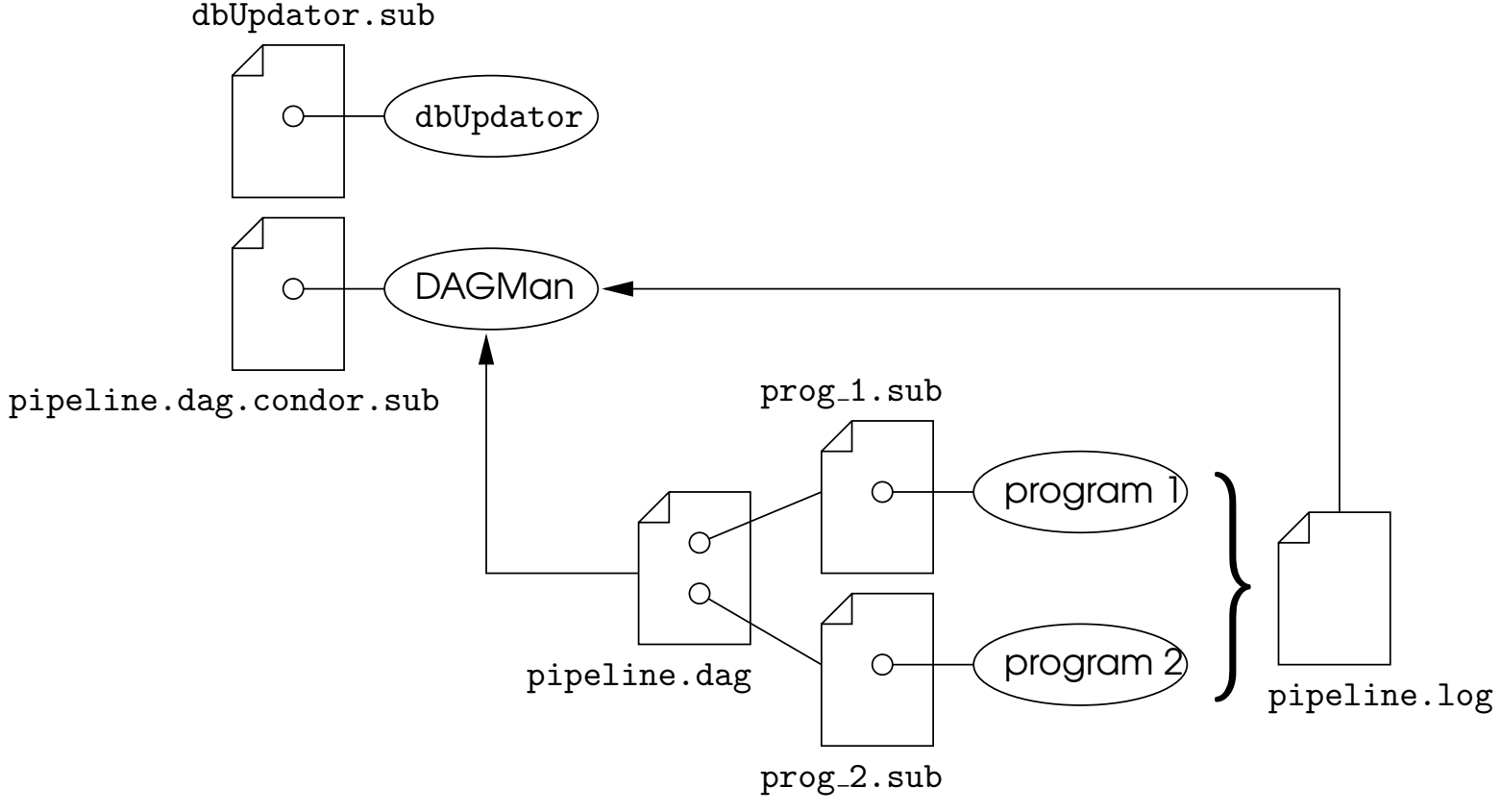
How we do it

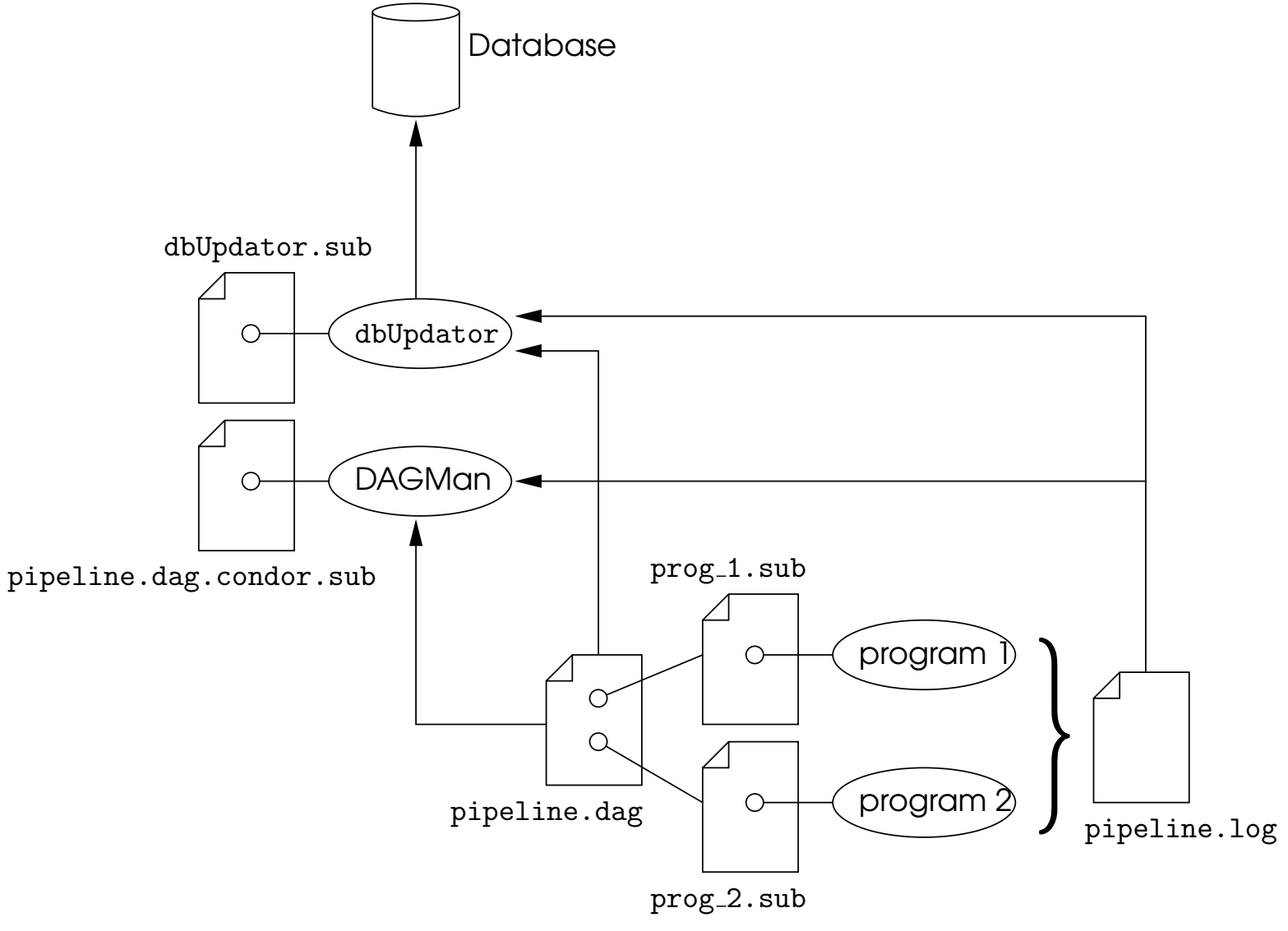
- Retain DAGMan as the work flow sequencer.
- User provides a script to generate a DAG that will analyze an interval of data given the start and end time of that data.
- A daemon watches for data, and calls the user's script as each new data segment becomes available.
- A job monitoring process runs alongside each DAG, periodically updating a MySQL database with the DAG's state.
- Scripts query the database, and provide a live view of the status of all analyses via a web interface.

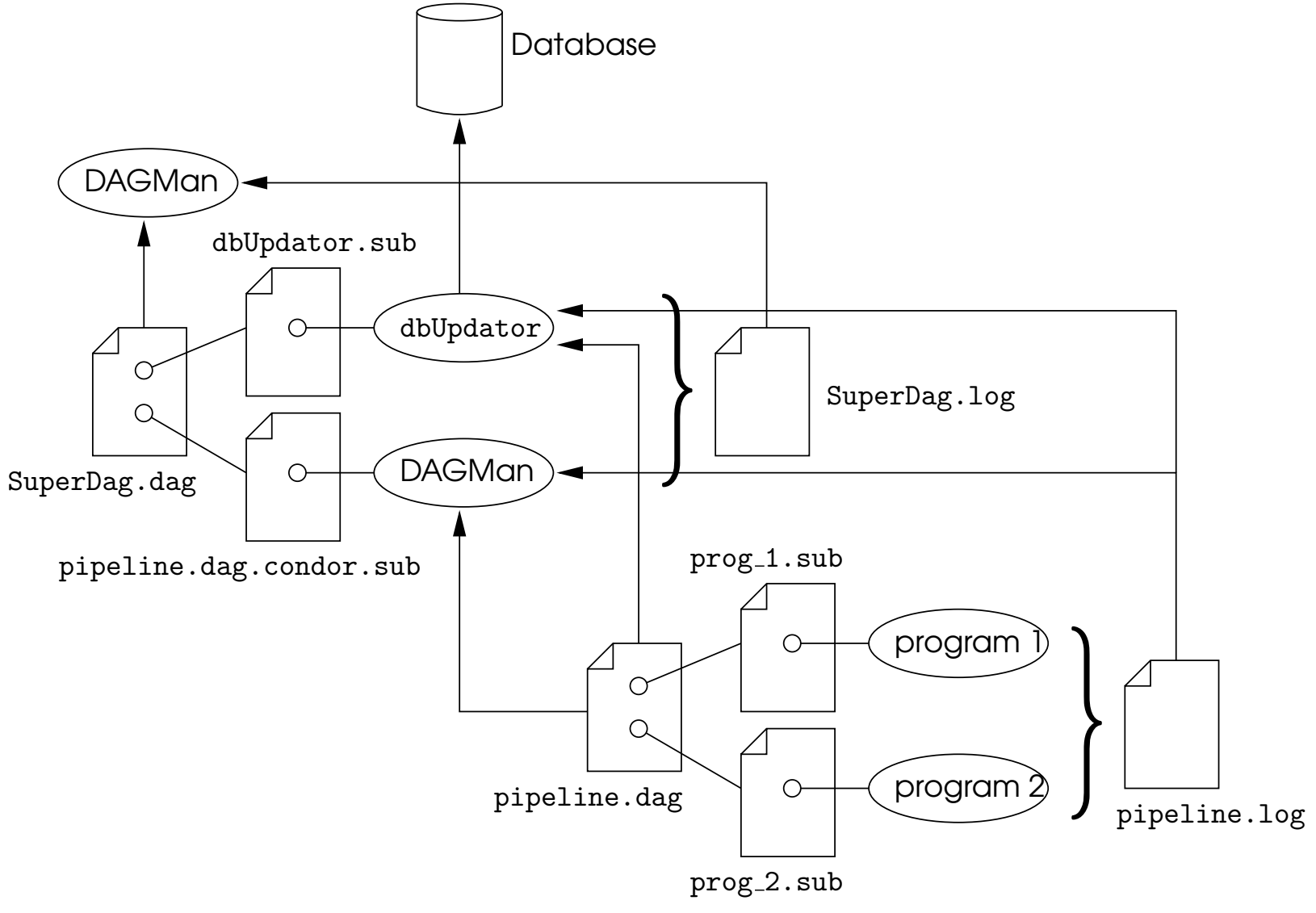

Start
Get segment list
segment database
Compute unused segments
used segments
Segments remain?
No
Pause
Yes
Generate DAG for segment
Submit DAG to compute cluster


DAGMan
prog_1.sub
program 1
pipeline.dag
program 2
pipeline.log
prog_2.sub


DAGMan
pipeline.dag.condor.sub
prog_1.sub
program 1
pipeline.dag
program 2
prog_2.sub
pipeline.log


dbUpdator.sub
dbUpdator
DAGMan
pipeline.dag.condor.sub
prog_1.sub
program 1
pipeline.dag
program 2
prog_2.sub
pipeline.log


Database
dbUpdator.sub
dbUpdator
DAGMan
pipeline.dag.condor.sub
prog_1.sub
program 1
program 2
pipeline.dag
pipeline.log
prog_2.sub

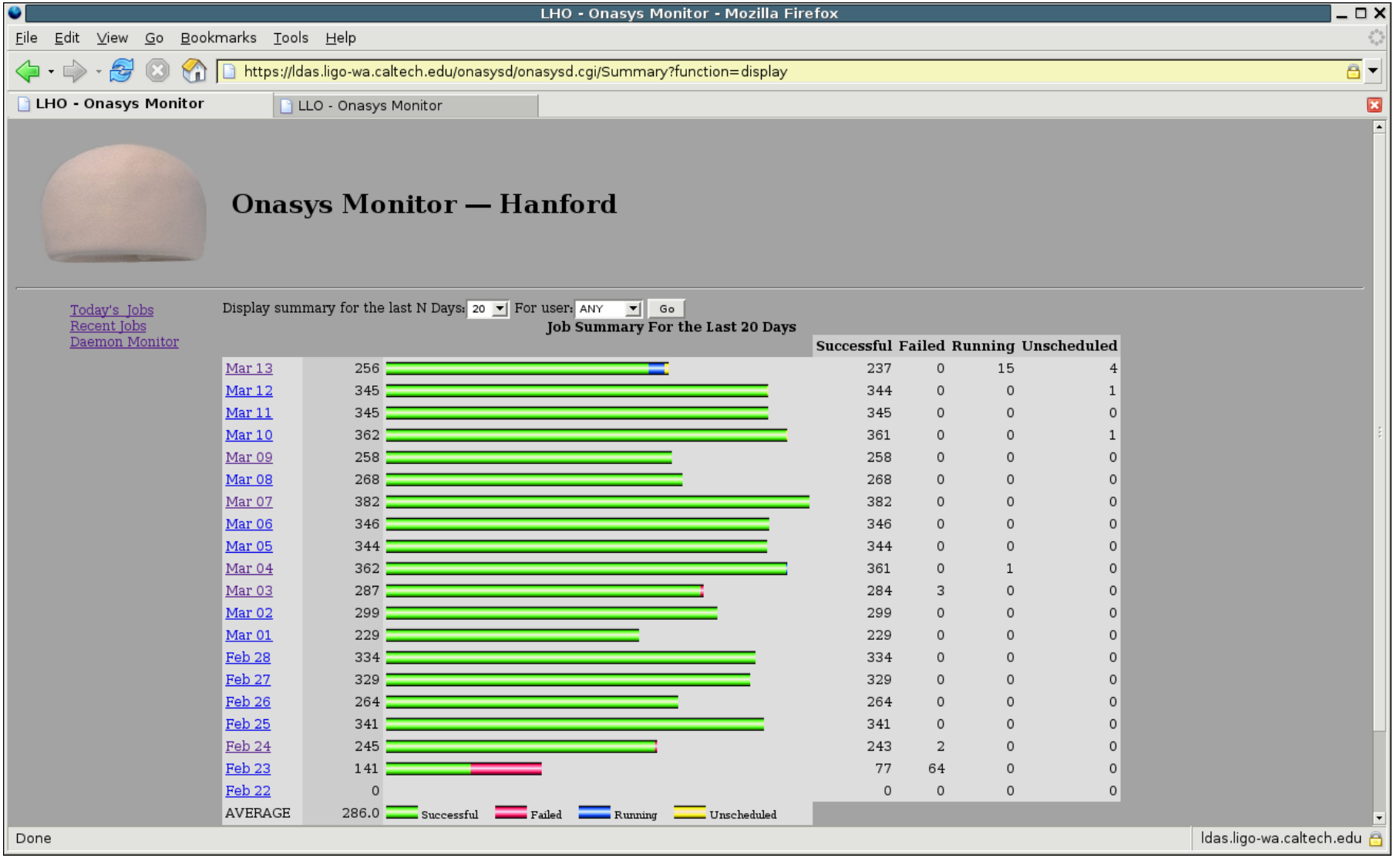

Database
DAGMan
dbUpdator.sub
dbUpdator
SuperDag.log
SuperDag.dag
DAGMan
pipeline.dag.condor.sub
prog_1.sub
program 1
pipeline.dag
program 2
pipeline.log
prog_2.sub

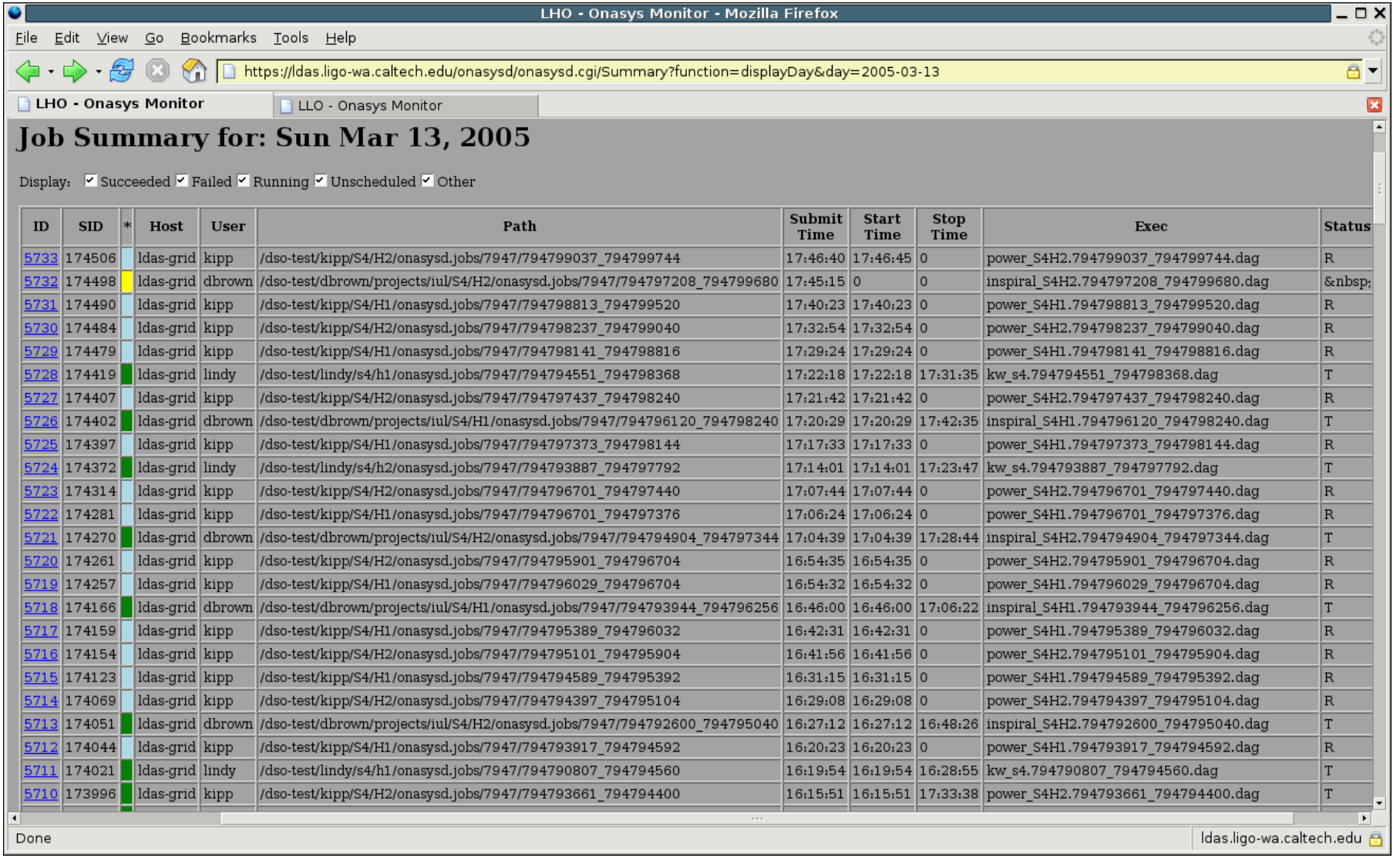
Advantages

- No direct input from the on-line tools to the data analysis pipeline, only monitoring.
- Little interaction between on-line tools and data analysis pipelines means very few restrictions — user can run nearly any off-line DAG virtually as-is.
- Uses Condor and DAGMan to provide execution reliability for the various components.

Potential Problems

- "A DAG every ten minutes" might bury us under log files — reliable machine-assisted job monitoring and diagnostics are a must.
- Long-term scalability. Existing setup will survive current science run without problems. The next, six month long, science run can almost certainly be accommodated as well. Will need to develop automated house-keeping tool(s) for continuous running.

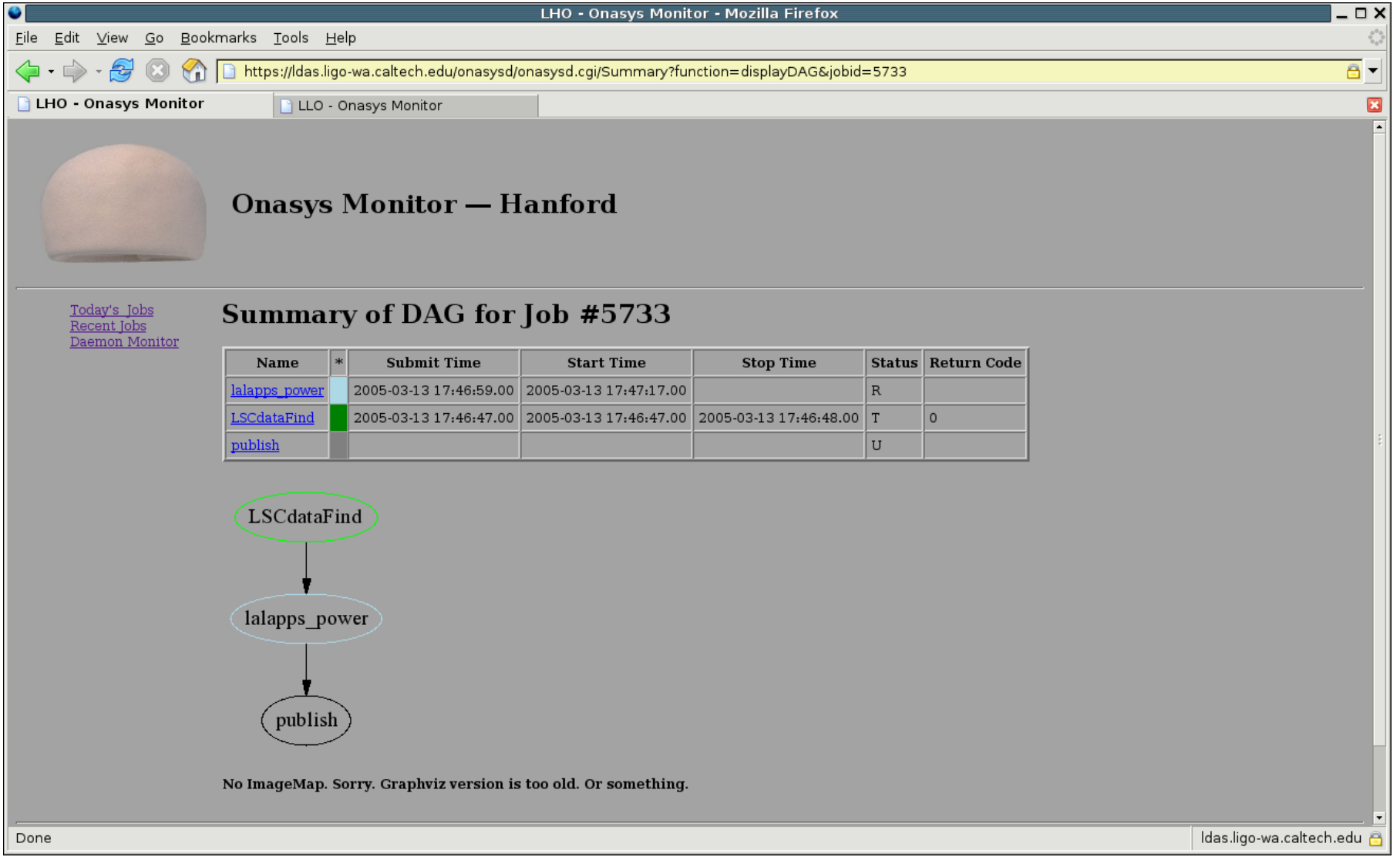

LHO - Onasys Monitor - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
https://ldas.ligo-wa.caltech.edu/onasysd/onasysd.cgi/Summary?function=display
LHO - Onasys Monitor
LLO - Onasys Monitor
Onasys Monitor — Hanford
Today's Jobs
Recent Jobs
Daemon Monitor
Display summary for the last N Days: 20 For user: ANY Go
Job Summary For the Last 20 Days
Successful Failed Running Unscheduled
Mar 13 256 237 0 15 4
Mar 12 345 344 0 0 1
Mar 11 345 345 0 0 0
Mar 10 362 361 0 0 1
Mar 09 258 258 0 0 0
Mar 08 268 268 0 0 0
Mar 07 382 382 0 0 0
Mar 06 346 346 0 0 0
Mar 05 344 344 0 0 0
Mar 04 362 361 0 1 0
Mar 03 287 284 3 0 0
Mar 02 299 299 0 0 0
Mar 01 229 229 0 0 0
Feb 28 334 334 0 0 0
Feb 27 329 329 0 0 0
Feb 26 264 264 0 0 0
Feb 25 341 341 0 0 0
Feb 24 245 243 2 0 0
Feb 23 141 77 64 0 0
Feb 22 0 0 0 0 0
AVERAGE 286.0 Successful Failed Running Unscheduled
Done
ldas.ligo-wa.caltech.edu

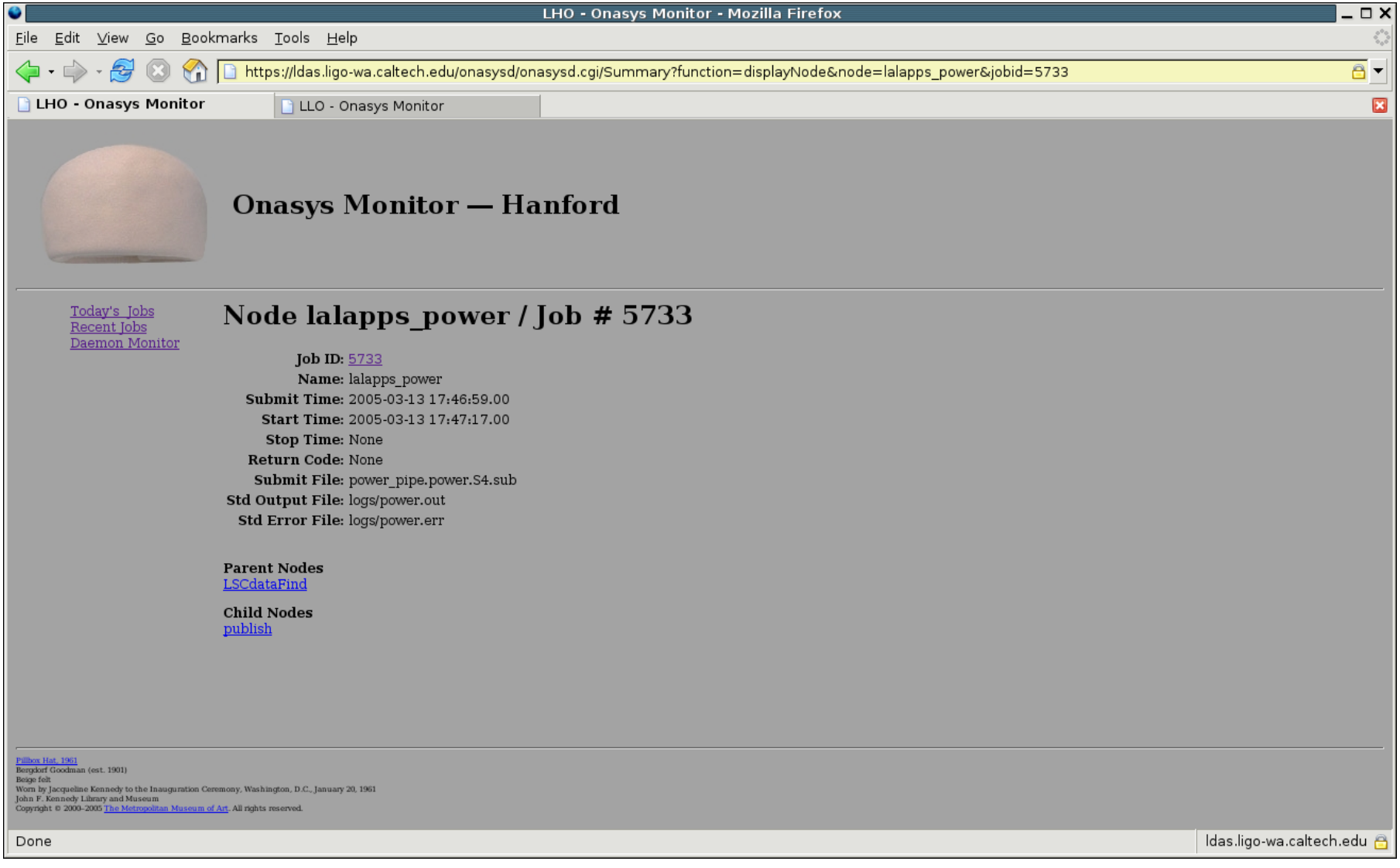


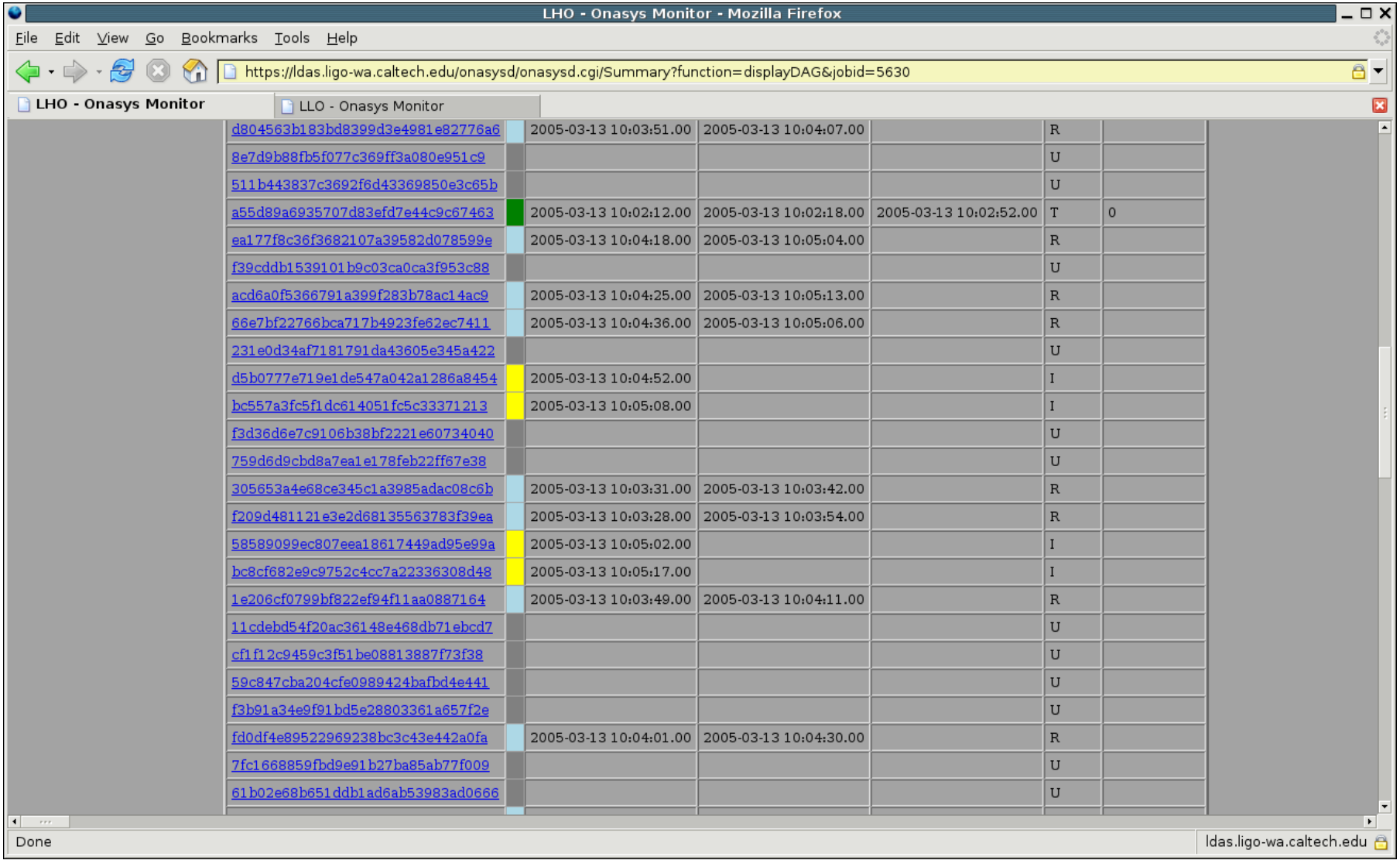
Job Summary for: Sun Mar 13, 2005

Display: ☑ Succeeded ☑ Failed ☑ Running ☑ Unscheduled ☑ Other

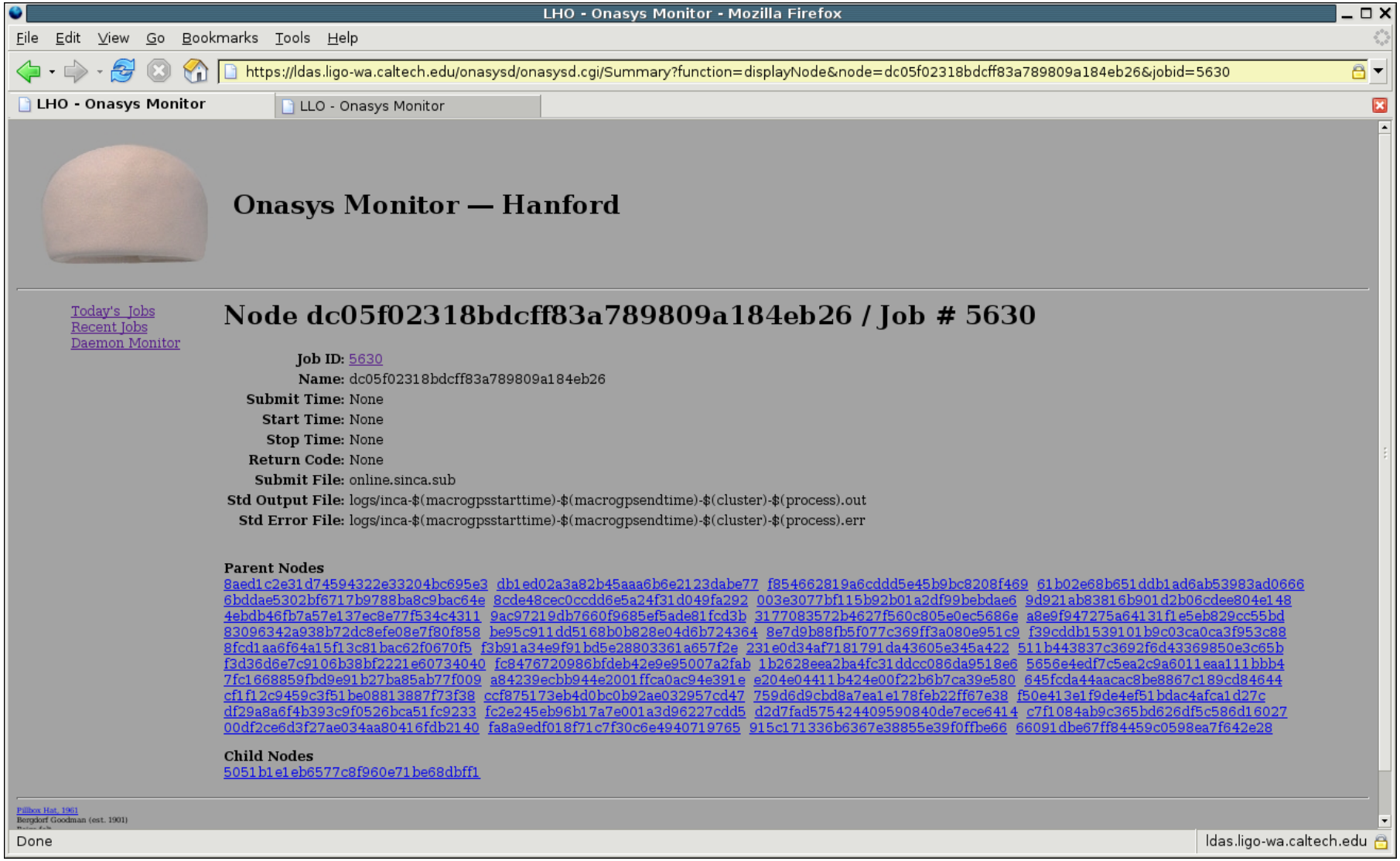
ID	SID	*	Host	User	Path	Submit Time	Start Time	Stop Time	Exec	Status
5733	174506		ldas-grid	kipp	/dso-test/kipp/S4/H2/onasysd.jobs/7947/794799037_794799744	17:46:40	17:46:45	0	power_S4H2.794799037_794799744.dag	R
5732	174498		ldas-grid	dbrown	/dso-test/dbrown/projects/iul/S4/H2/onasysd.jobs/7947/794797208_794799680	17:45:15	0	0	inspiral_S4H2.794797208_794799680.dag	
5731	174490		ldas-grid	kipp	/dso-test/kipp/S4/H1/onasysd.jobs/7947/794798813_794799520	17:40:23	17:40:23	0	power_S4H1.794798813_794799520.dag	R
5730	174484		ldas-grid	kipp	/dso-test/kipp/S4/H2/onasysd.jobs/7947/794798237_794799040	17:32:54	17:32:54	0	power_S4H2.794798237_794799040.dag	R
5729	174479		ldas-grid	kipp	/dso-test/kipp/S4/H1/onasysd.jobs/7947/794798141_794798816	17:29:24	17:29:24	0	power_S4H1.794798141_794798816.dag	R
5728	174419		ldas-grid	lindy	/dso-test/lindy/s4/h1/onasysd.jobs/7947/794794551_794798368	17:22:18	17:22:18	17:31:35	kw_s4.794794551_794798368.dag	T
5727	174407		ldas-grid	kipp	/dso-test/kipp/S4/H2/onasysd.jobs/7947/794797437_794798240	17:21:42	17:21:42	0	power_S4H2.794797437_794798240.dag	R
5726	174402		ldas-grid	dbrown	/dso-test/dbrown/projects/iul/S4/H1/onasysd.jobs/7947/794796120_794798240	17:20:29	17:20:29	17:42:35	inspiral_S4H1.794796120_794798240.dag	T
5725	174397		ldas-grid	kipp	/dso-test/kipp/S4/H1/onasysd.jobs/7947/794797373_794798144	17:17:33	17:17:33	0	power_S4H1.794797373_794798144.dag	R
5724	174372		ldas-grid	lindy	/dso-test/lindy/s4/h2/onasysd.jobs/7947/794793887_794797792	17:14:01	17:14:01	17:23:47	kw_s4.794793887_794797792.dag	T
5723	174314		ldas-grid	kipp	/dso-test/kipp/S4/H2/onasysd.jobs/7947/794796701_794797440	17:07:44	17:07:44	0	power_S4H2.794796701_794797440.dag	R
5722	174281		ldas-grid	kipp	/dso-test/kipp/S4/H1/onasysd.jobs/7947/794796701_794797376	17:06:24	17:06:24	0	power_S4H1.794796701_794797376.dag	R
5721	174270		ldas-grid	dbrown	/dso-test/dbrown/projects/iul/S4/H2/onasysd.jobs/7947/794794904_794797344	17:04:39	17:04:39	17:28:44	inspiral_S4H2.794794904_794797344.dag	T
5720	174261		ldas-grid	kipp	/dso-test/kipp/S4/H2/onasysd.jobs/7947/794795901_794796704	16:54:35	16:54:35	0	power_S4H2.794795901_794796704.dag	R
5719	174257		ldas-grid	kipp	/dso-test/kipp/S4/H1/onasysd.jobs/7947/794796029_794796704	16:54:32	16:54:32	0	power_S4H1.794796029_794796704.dag	R
5718	174166		ldas-grid	dbrown	/dso-test/dbrown/projects/iul/S4/H1/onasysd.jobs/7947/794793944_794796256	16:46:00	16:46:00	17:06:22	inspiral_S4H1.794793944_794796256.dag	T
5717	174159		ldas-grid	kipp	/dso-test/kipp/S4/H1/onasysd.jobs/7947/794795389_794796032	16:42:31	16:42:31	0	power_S4H1.794795389_794796032.dag	R
5716	174154		ldas-grid	kipp	/dso-test/kipp/S4/H2/onasysd.jobs/7947/794795101_794795904	16:41:56	16:41:56	0	power_S4H2.794795101_794795904.dag	R
5715	174123		ldas-grid	kipp	/dso-test/kipp/S4/H1/onasysd.jobs/7947/794794589_794795392	16:31:15	16:31:15	0	power_S4H1.794794589_794795392.dag	R
5714	174069		ldas-grid	kipp	/dso-test/kipp/S4/H2/onasysd.jobs/7947/794794397_794795104	16:29:08	16:29:08	0	power_S4H2.794794397_794795104.dag	R
5713	174051		ldas-grid	dbrown	/dso-test/dbrown/projects/iul/S4/H2/onasysd.jobs/7947/794792600_794795040	16:27:12	16:27:12	16:48:26	inspiral_S4H2.794792600_794795040.dag	T
5712	174044		ldas-grid	kipp	/dso-test/kipp/S4/H1/onasysd.jobs/7947/794793917_794794592	16:20:23	16:20:23	0	power_S4H1.794793917_794794592.dag	R
5711	174021		ldas-grid	lindy	/dso-test/lindy/s4/h1/onasysd.jobs/7947/794790807_794794560	16:19:54	16:19:54	16:28:55	kw_s4.794790807_794794560.dag	T
5710	173996		ldas-grid	kipp	/dso-test/kipp/S4/H2/onasysd.jobs/7947/794793661_794794400	16:15:51	16:15:51	17:33:38	power_S4H2.794793661_794794400.dag	T


LHO - Onasys Monitor - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
https://ldas.ligo-wa.caltech.edu/onasysd/onasysd.cgi/Summary?function=displayDAG&jobid=5733
LHO - Onasys Monitor
LLO - Onasys Monitor
Onasys Monitor — Hanford
Today's Jobs
Recent Jobs
Daemon Monitor
Summary of DAG for Job #5733
Name | * | Submit Time | Start Time | Stop Time | Status | Return Code
lalapps_power | 2005-03-13 17:46:59.00 | 2005-03-13 17:47:17.00 | | R |
LSCdataFind | 2005-03-13 17:46:47.00 | 2005-03-13 17:46:47.00 | 2005-03-13 17:46:48.00 | T | 0
publish | | | | U |
LSCdataFind
lalapps_power
publish
No ImageMap. Sorry. Graphviz version is too old. Or something.
Done
ldas.ligo-wa.caltech.edu


LHO - Onasys Monitor - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
https://ldas.ligo-wa.caltech.edu/onasysd/onasysd.cgi/Summary?function=displayNode&node=lalapps_power&jobid=5733
LHO - Onasys Monitor
LLO - Onasys Monitor
Onasys Monitor — Hanford
Today's Jobs
Recent Jobs
Daemon Monitor
Node lalapps_power / Job # 5733
Job ID: 5733
Name: lalapps_power
Submit Time: 2005-03-13 17:46:59.00
Start Time: 2005-03-13 17:47:17.00
Stop Time: None
Return Code: None
Submit File: power_pipe.power.S4.sub
Std Output File: logs/power.out
Std Error File: logs/power.err
Parent Nodes
LSCdataFind
Child Nodes
publish
Pillbox Hat, 1961
Bergdorf Goodman (est. 1901)
Beige felt
Worn by Jacqueline Kennedy to the Inauguration Ceremony, Washington, D.C., January 20, 1961
John F. Kennedy Library and Museum
Copyright © 2000-2005 The Metropolitan Museum of Art. All rights reserved.
Done
ldas.ligo-wa.caltech.edu



d804563b183bd8399d3e4981e82776a6		2005-03-13 10:03:51.00	2005-03-13 10:04:07.00		R	
8e7d9b88fb5f077c369ff3a080e951c9					U	
511b443837c3692f6d43369850e3c65b					U	
a55d89a6935707d83efd7e44c9c67463		2005-03-13 10:02:12.00	2005-03-13 10:02:18.00	2005-03-13 10:02:52.00	T	0
ea177f8c36f3682107a39582d078599e		2005-03-13 10:04:18.00	2005-03-13 10:05:04.00		R	
f39cddb1539101b9c03ca0ca3f953c88					U	
acd6a0f5366791a399f283b78ac14ac9		2005-03-13 10:04:25.00	2005-03-13 10:05:13.00		R	
66e7bf22766bca717b4923fe62ec7411		2005-03-13 10:04:36.00	2005-03-13 10:05:06.00		R	
231e0d34af7181791da43605e345a422					U	
d5b0777e719e1de547a042a1286a8454		2005-03-13 10:04:52.00			I	
bc557a3fc5f1dc614051fc5c33371213		2005-03-13 10:05:08.00			I	
f3d36d6e7c9106b38bf2221e60734040					U	
759d6d9cbd8a7ea1e178feb22ff67e38					U	
305653a4e68ce345c1a3985adac08c6b		2005-03-13 10:03:31.00	2005-03-13 10:03:42.00		R	
f209d481121e3e2d68135563783f39ea		2005-03-13 10:03:28.00	2005-03-13 10:03:54.00		R	
58589099ec807eea18617449ad95e99a		2005-03-13 10:05:02.00			I	
bc8cf682e9c9752c4cc7a22336308d48		2005-03-13 10:05:17.00			I	
1e206cf0799bf822ef94f11aa0887164		2005-03-13 10:03:49.00	2005-03-13 10:04:11.00		R	
11cdebd54f20ac36148e468db71ebcd7					U	
cf1f12c9459c3f51be08813887f73f38					U	
59c847cba204cfe0989424bafbd4e441					U	
f3b91a34e9f91bd5e28803361a657f2e					U	
fd0df4e89522969238bc3c43e442a0fa		2005-03-13 10:04:01.00	2005-03-13 10:04:30.00		R	
7fc1668859fbd9e91b27ba85ab77f009					U	
61b02e68b651ddb1ad6ab53983ad0666					U	


LHO - Onasys Monitor - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
https://ldas.ligo-wa.caltech.edu/onasysd/onasysd.cgi/Summary?function=displayNode&node=dc05f02318bdcff83a789809a184eb26&jobid=5630
LHO - Onasys Monitor
LLO - Onasys Monitor
Onasys Monitor — Hanford
Today's Jobs
Recent Jobs
Daemon Monitor
Node dc05f02318bdcff83a789809a184eb26 / Job # 5630
Job ID: 5630
Name: dc05f02318bdcff83a789809a184eb26
Submit Time: None
Start Time: None
Stop Time: None
Return Code: None
Submit File: online.sinca.sub
Std Output File: logs/inca-$(macrogpsstarttime)-$(macrogpsendtime)-$(cluster)-$(process).out
Std Error File: logs/inca-$(macrogpsstarttime)-$(macrogpsendtime)-$(cluster)-$(process).err
Parent Nodes
8aed1c2e31d74594322e33204bc695e3 db1ed02a3a82b45aaa6b6e2123dabe77 f854662819a6cddd5e45b9bc8208f469 61b02e68b651ddb1ad6ab53983ad0666
6bddae5302bf6717b9788ba8c9bac64e 8cde48cec0ccdd6e5a24f31d049fa292 003e3077bf115b92b01a2df99bebdae6 9d921ab83816b901d2b06cdee804e148
4ebdb46fb7a57e137ec8e77f534c4311 9ac97219db7660f9685ef5ade81fcd3b 3177083572b4627f560c805e0ec5686e a8e9f947275a64131f1e5eb829cc55bd
83096342a938b72dc8efe08e7f80f858 be95c911dd5168b0b828e04d6b724364 8e7d9b88fb5f077c369ff3a080e951c9 f39cddb1539101b9c03ca0ca3f953c88
8fcd1aa6f64a15f13c81bac62f0670f5 f3b91a34e9f91bd5e28803361a657f2e 231e0d34af7181791da43605e345a422 511b443837c3692f6d43369850e3c65b
f3d36d6e7c9106b38bf2221e60734040 fc8476720986bfdeb42e9e95007a2fab 1b2628eea2ba4fc31ddcc086da9518e6 5656e4edf7c5ea2c9a6011eaa111bbb4
7fc1668859fbd9e91b27ba85ab77f009 a84239ecbb944e2001ffca0ac94e391e e204e04411b424e00f22b6b7ca39e580 645fcda44aacac8be8867c189cd84644
cf1f12c9459c3f51be08813887f73f38 ccf875173eb4d0bc0b92ae032957cd47 759d6d9cbd8a7ea1e178feb22ff67e38 f50e413e1f9de4ef51bdac4afca1d27c
df29a8a6f4b393c9f0526bca51fc9233 fc2e245eb96b17a7e001a3d96227cdd5 d2d7fad575424409590840de7ece6414 c7f1084ab9c365bd626df5c586d16027
00df2ce6d3f27ae034aa80416fdb2140 fa8a9edf018f71c7f30c6e4940719765 915c171336b6367e38855e39f0ffbe66 66091dbe67ff84459c0598ea7f642e28
Child Nodes
5051b1e1eb6577c8f960e71be68dbff1
Pillbox Hat, 1961
Bergdorf Goodman (est. 1901)
Done
ldas.ligo-wa.caltech.edu

Summary

- As far as we can tell, any DAG at all can be run using these tools, and monitored via the web interface.
- For three weeks, the system has delivered latencies of 10s of minutes for all three searches being run using it.
- System has proven itself to be fault tolerant, surviving a power failure at one observatory with minimal user intervention.
- The machine-assisted monitoring and fault analysis has proven itself to be highly effective at quickly identifying failures, and directing the user to the specific problem that has occurred.

Acknowledgements

- Stuart Anderson, Duncan Brown, Ben Johnson, Scott Koranda, Greg Mendell, Brian Moe