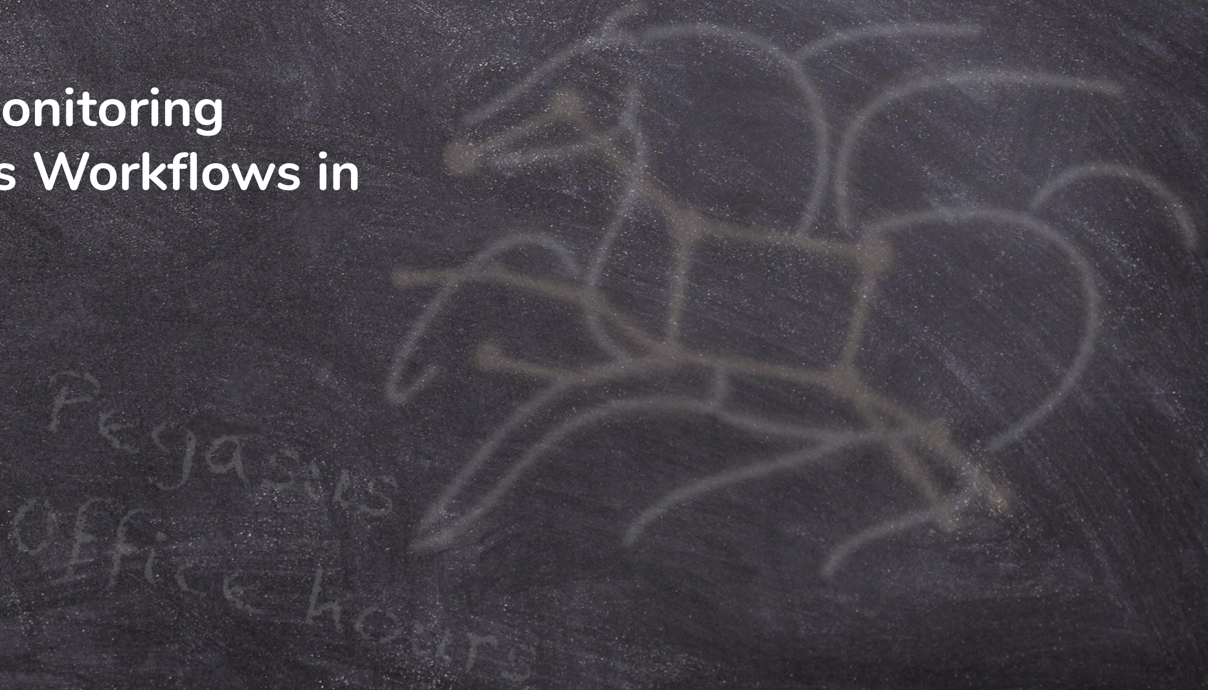
Data Collection and Monitoring Across Heterogeneous Workflows in Pegasus

Ryan Tanaka
tanaka@isi.edu



USC Viterbi
Information Sciences Institute

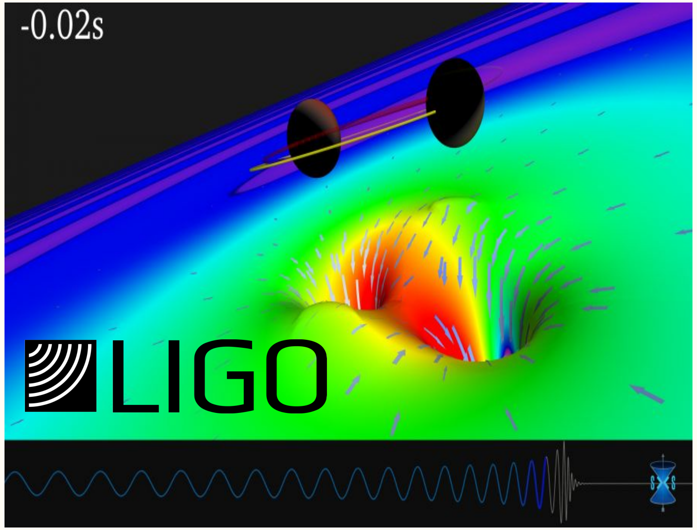
https://pegasus.isi.edu/

Background: Gravitational Waves, Workflows, LIGO

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

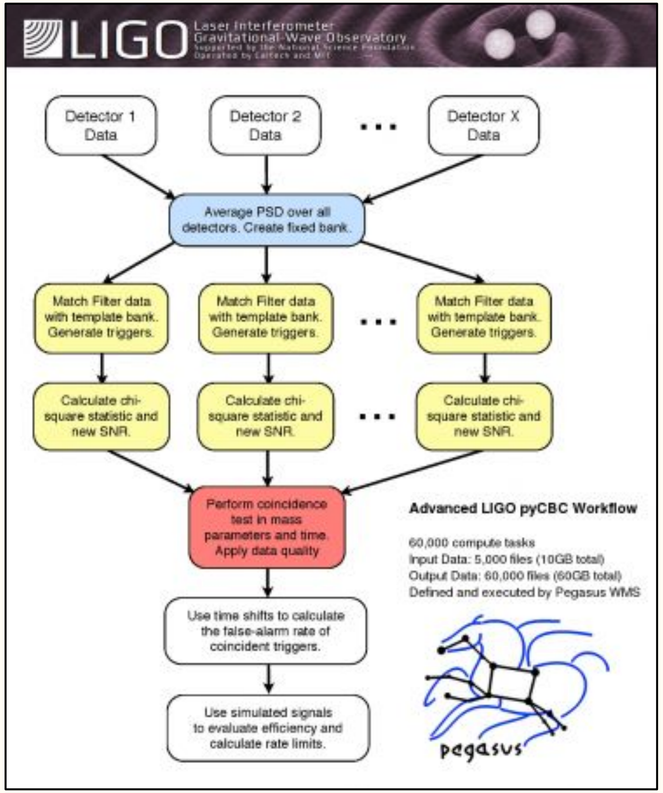
Background: Gravitational Waves, Workflows, LIGO

Advanced PyCBC Workflows:

- 60,000 compute tasks
- 5,000 input files (10GB total)
- 60,000 output files (60GB total)
- Post run analysis

Challenges:

- Error analysis on workflows of this scale
- Monitoring across multiple runs, users, and submit machines


LIGO
Laser Interferometer Gravitational-Wave Observatory
Supported by the National Science Foundation
Operated by Caltech and MIT
Detector 1 Data
Detector 2 Data
Detector X Data
Average PSD over all detectors. Create fixed bank.
Match Filter data with template bank. Generate triggers.
Calculate chi-square statistic and new SNR.
Perform coincidence test in mass parameters and time. Apply data quality
Use time shifts to calculate the false-alarm rate of coincident triggers.
Use simulated signals to evaluate efficiency and calculate rate limits.
Advanced LIGO pyCBC Workflow
60,000 compute tasks
Input Data: 5,000 files (10GB total)
Output Data: 60,000 files (60GB total)
Defined and executed by Pegasus WMS
pegasus

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

USC Viterbi
Information Sciences Institute

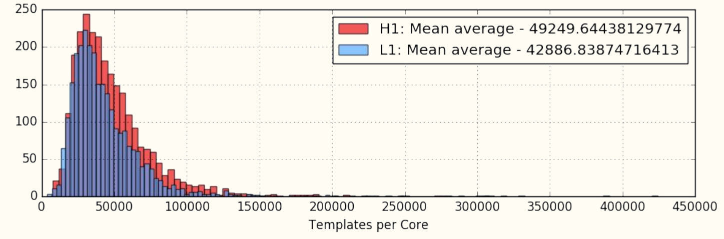
Background: Gravitational Waves, Workflows, LIGO

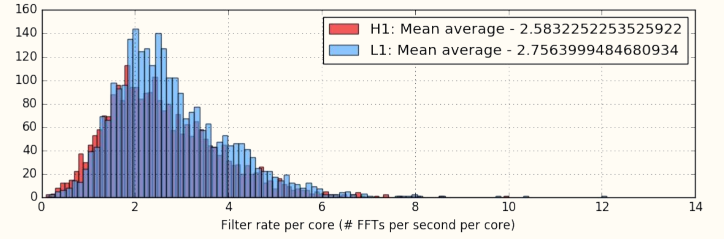
Advanced PyCBC Workflows:

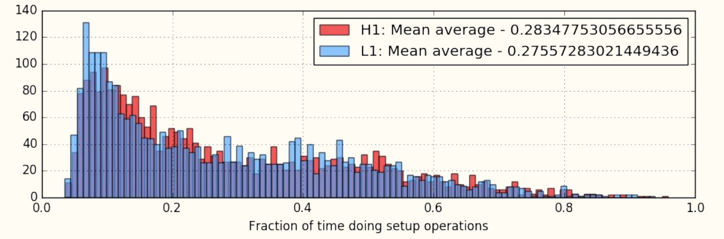
- 60,000 compute tasks
- 5,000 input files (10GB total)
- 60,000 output files (60GB total)
- Post run analysis

Challenges:

- Error analysis on workflows of this scale
- Monitoring across multiple runs, users, and submit machines







USC Viterbi
Information Sciences Institute

Background: Empowering LIGO Researchers

CIF21 DIBBs: Domain-Aware Management of Heterogeneous Workflows

Active Data Management for Gravitational-Wave Science

PI: Duncan Brown[1], Co-PIs: Peter Couvares[2] Ewa Deelman[3], Jian Qin[1] NSF Award ACI-1443047

1 Syracuse University. 2 LIGO Caltech, 3 USC Information Sciences Institute.

Goals:

- Develop new data management techniques in Pegasus
- Improve data access for LIGO researchers
- Enhance Pegasus workflow monitoring capabilities
- Enable LIGO researchers to conduct analysis across multiple PyCBC pipeline runs

Background: Empowering LIGO Researchers

Outcomes:

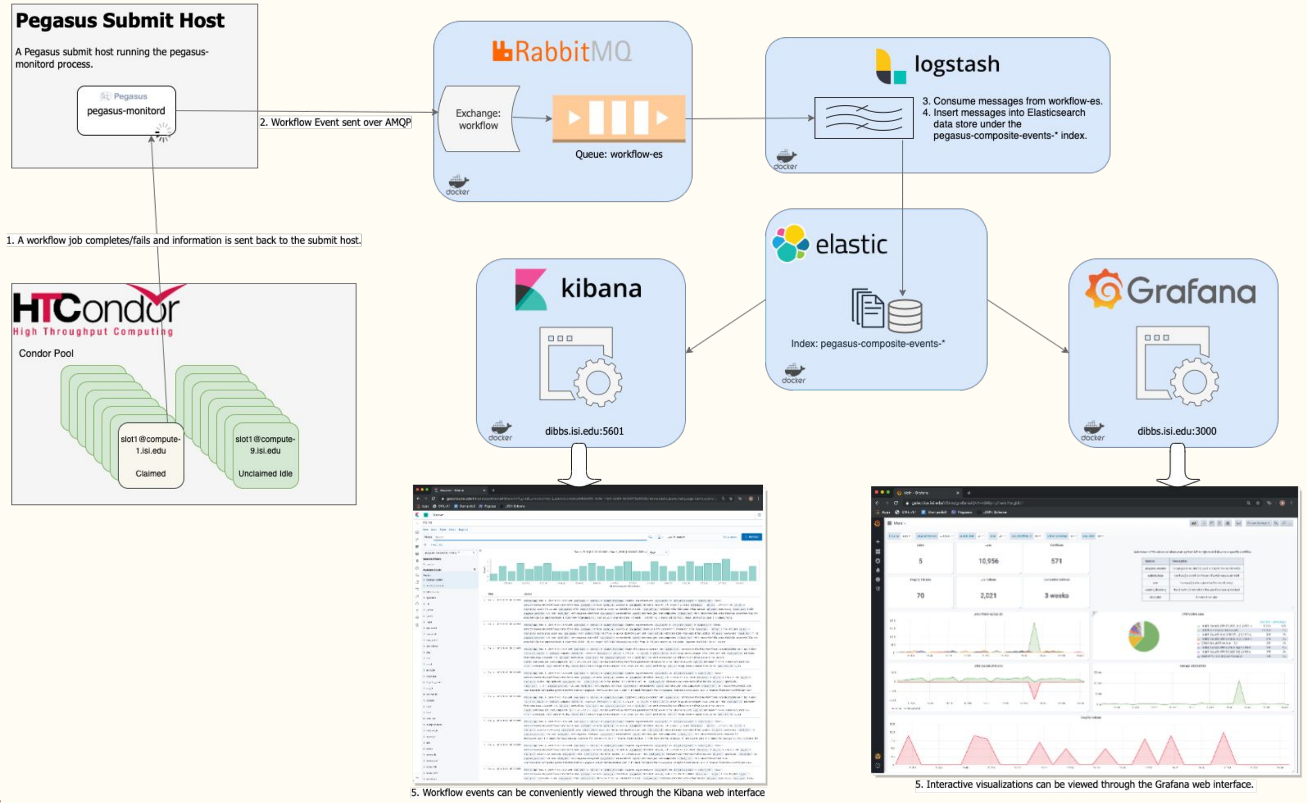
- Developed Pegasus extensions to capture runtime provenance metadata
- Data storage Solutions
  - relational datastore linked to a single PyCBC run
  - Elasticsearch persisting data across multiple PyCBC runs
- Browser based monitoring/analytics solutions
  - Kibana: query/explore Elasticsearch data via a browser
  - Grafana: dashboard for viewing workflow runs at varying levels of granularity

Outline

- ~~Background~~
- **Data collection pipeline**
- Getting started
- Demo

USC Viterbi
Information Sciences Institute

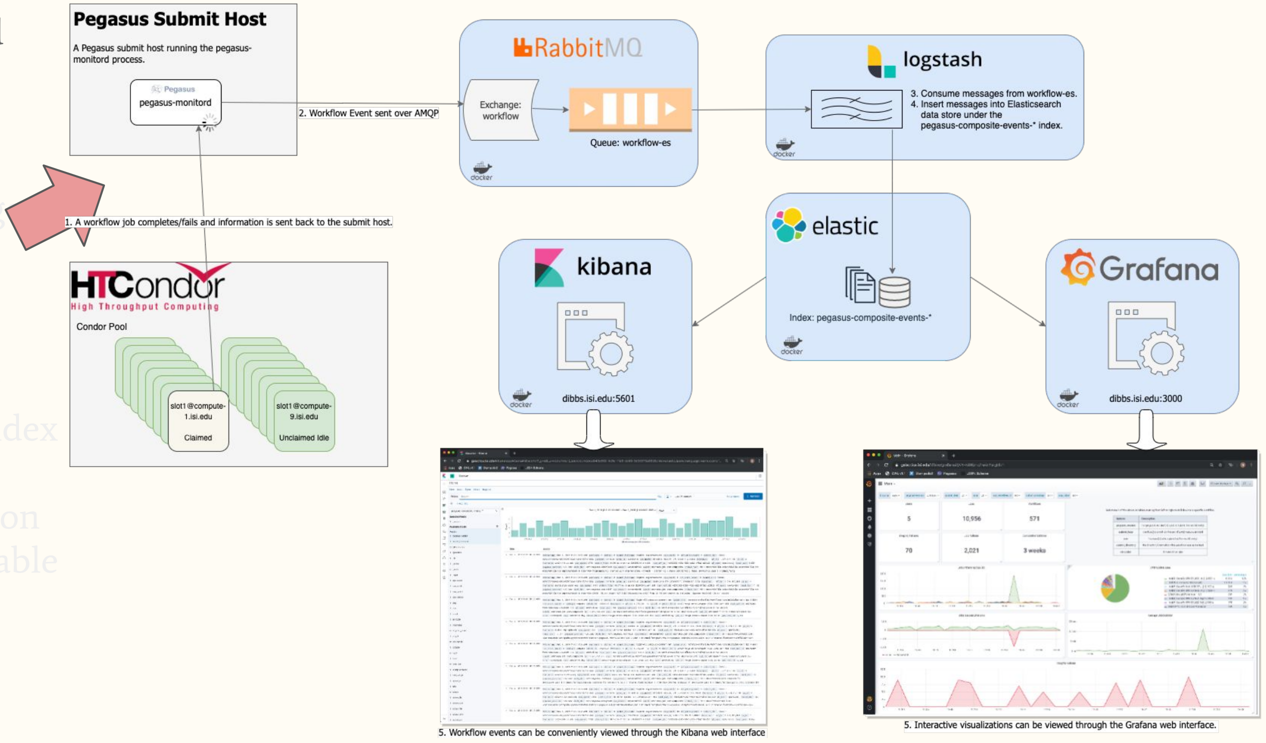
Data Collection Pipeline: Overview


Pegasus Submit Host
A Pegasus submit host running the pegasus-monitord process.
Pegasus
pegasus-monitord
2. Workflow Event sent over AMQP
RabbitMQ
Exchange: workflow
Queue: workflow-es
docker
logstash
3. Consume messages from workflow-es.
4. Insert messages into Elasticsearch data store under the pegasus-composite-events-* index.
1. A workflow job completes/fails and information is sent back to the submit host.
HTCondor
High Throughput Computing
Condor Pool
slot1@compute-1.isi.edu
Claimed
slot1@compute-9.isi.edu
Unclaimed Idle
elastic
Index: pegasus-composite-events-*
kibana
dibbs.isi.edu:5601
Grafana
dibbs.isi.edu:3000
5. Workflow events can be conveniently viewed through the Kibana web interface
5. Interactive visualizations can be viewed through the Grafana web interface.

Data Collection Pipeline: Data Flow

Sequence of Events:

1. Workflow job completes and information sent to pegasus-monitord
2. Workflow event sent over AMQP to RabbitMQ message queue
3. Logstash consumes message from queue
4. Logstash inserts event under pegasus-composite-even-* index in Elasticsearch data store
5. Data Exploration/Visualization
    a. Workflow events viewable through Kibana
    b. Dashboard viewable through Grafana



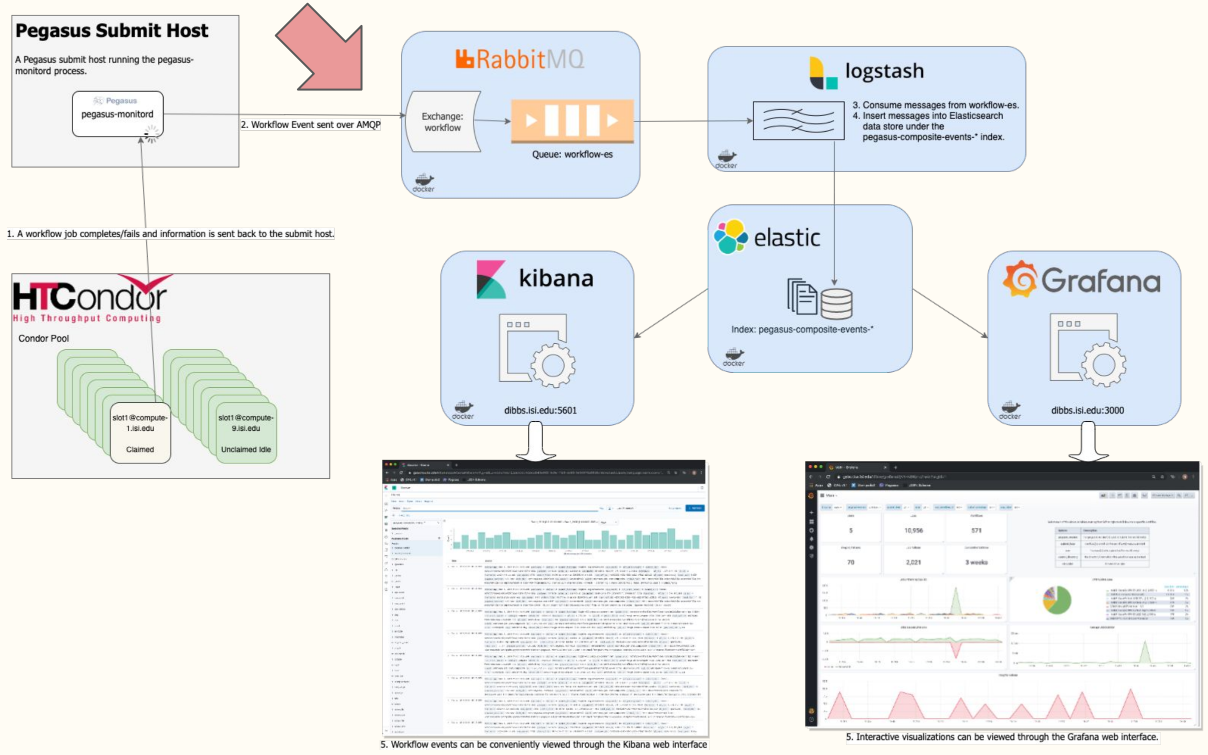
Pegasus Submit Host
A Pegasus submit host running the pegasus-monitord process.
Pegasus
pegasus-monitord
2. Workflow Event sent over AMQP
1. A workflow job completes/fails and information is sent back to the submit host.
HTCondor
High Throughput Computing
Condor Pool
slot1@compute-1.isi.edu
Claimed
slot1@compute-9.isi.edu
Unclaimed Idle
RabbitMQ
Exchange: workflow
Queue: workflow-es
docker
logstash
3. Consume messages from workflow-es.
4. Insert messages into Elasticsearch data store under the pegasus-composite-events-* index.
elastic
Index: pegasus-composite-events-*
kibana
dibbs.isi.edu:5601
Grafana
dibbs.isi.edu:3000
5. Workflow events can be conveniently viewed through the Kibana web interface
5. Interactive visualizations can be viewed through the Grafana web interface.

USC Viterbi
Information Sciences Institute

Data Collection Pipeline: Data Flow

Sequence of Events:

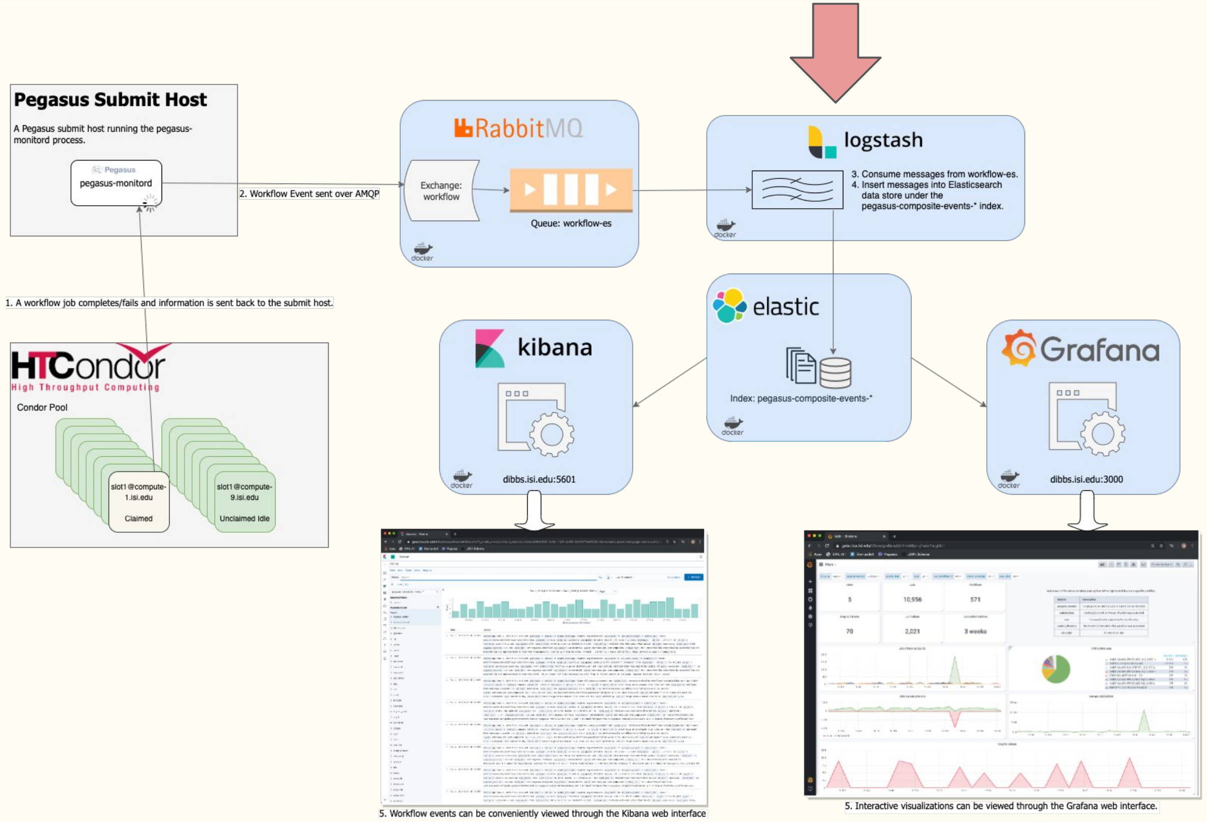
1. Workflow job completes and information sent to pegasus-monitord
2. Workflow event sent over AMQP to RabbitMQ message queue
3. Logstash consumes message from queue
4. Logstash inserts event under pegasus-composite-even-* index in Elasticsearch data store
5. Data Exploration/Visualization
    a. Workflow events viewable through Kibana
    b. Dashboard viewable through Grafana


Pegasus Submit Host
A Pegasus submit host running the pegasus-monitord process.
Pegasus
pegasus-monitord
2. Workflow Event sent over AMQP
RabbitMQ
Exchange: workflow
Queue: workflow-es
docker
logstash
3. Consume messages from workflow-es.
4. Insert messages into Elasticsearch data store under the pegasus-composite-events-* index.
1. A workflow job completes/fails and information is sent back to the submit host.
HTCondor
High Throughput Computing
Condor Pool
slot1@compute-1.isi.edu
Claimed
slot1@compute-9.isi.edu
Unclaimed Idle
elastic
Index: pegasus-composite-events-*
kibana
dibbs.isi.edu:5601
Grafana
dibbs.isi.edu:3000
5. Workflow events can be conveniently viewed through the Kibana web interface
5. Interactive visualizations can be viewed through the Grafana web interface.

Data Collection Pipeline: Data Flow

Sequence of Events:

1. Workflow job completes and information sent to pegasus-monitord
2. Workflow event sent over AMQP to RabbitMQ message queue
3. Logstash consumes message from queue
4. Logstash inserts event under pegasus-composite-even-* index in Elasticsearch data store
5. Data Exploration/Visualization
   a. Workflow events viewable through Kibana
   b. Dashboard viewable through Grafana

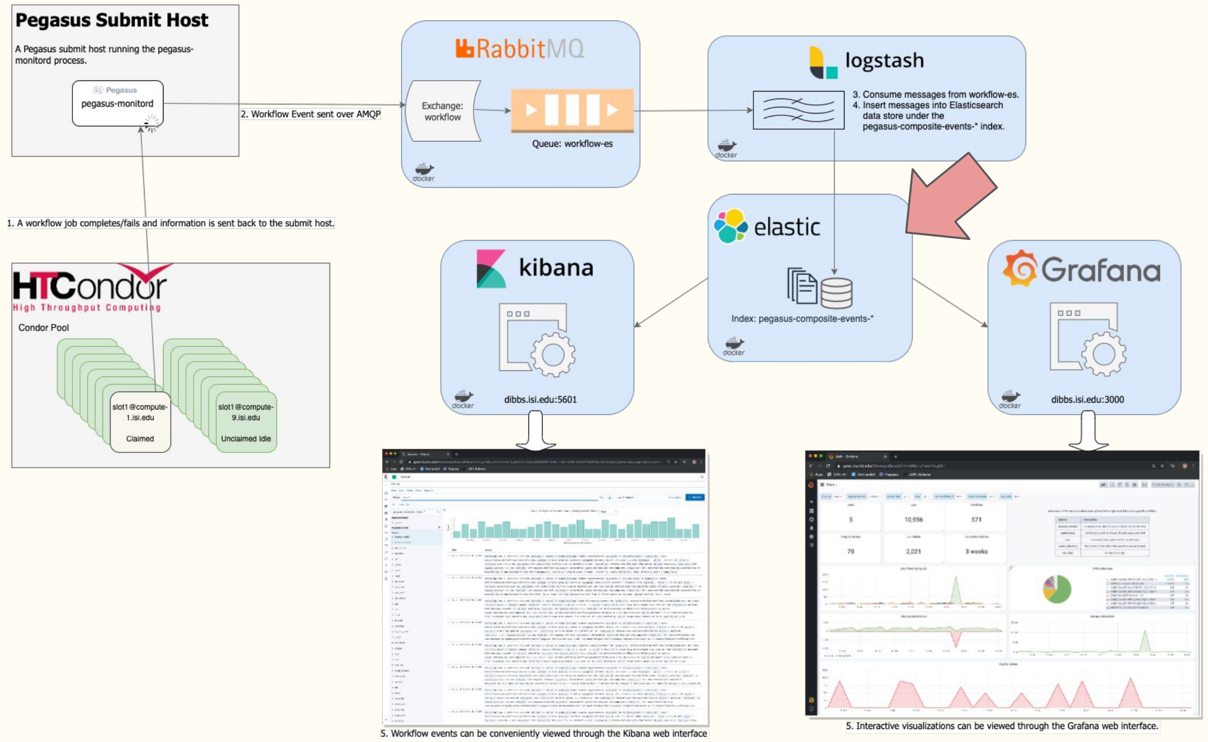

Pegasus Submit Host
A Pegasus submit host running the pegasus-monitord process.
Pegasus
pegasus-monitord
2. Workflow Event sent over AMQP
1. A workflow job completes/fails and information is sent back to the submit host.
HTCondor
High Throughput Computing
Condor Pool
slot1@compute-1.isi.edu
Claimed
slot1@compute-9.isi.edu
Unclaimed Idle
RabbitMQ
Exchange: workflow
Queue: workflow-es
docker
logstash
3. Consume messages from workflow-es.
4. Insert messages into Elasticsearch data store under the pegasus-composite-events-* index.
elastic
Index: pegasus-composite-events-*
kibana
dibbs.isi.edu:5601
Grafana
dibbs.isi.edu:3000
5. Workflow events can be conveniently viewed through the Kibana web interface
5. Interactive visualizations can be viewed through the Grafana web interface.

USC Viterbi
Information Sciences Institute

Data Collection Pipeline: Data Flow

Sequence of Events:

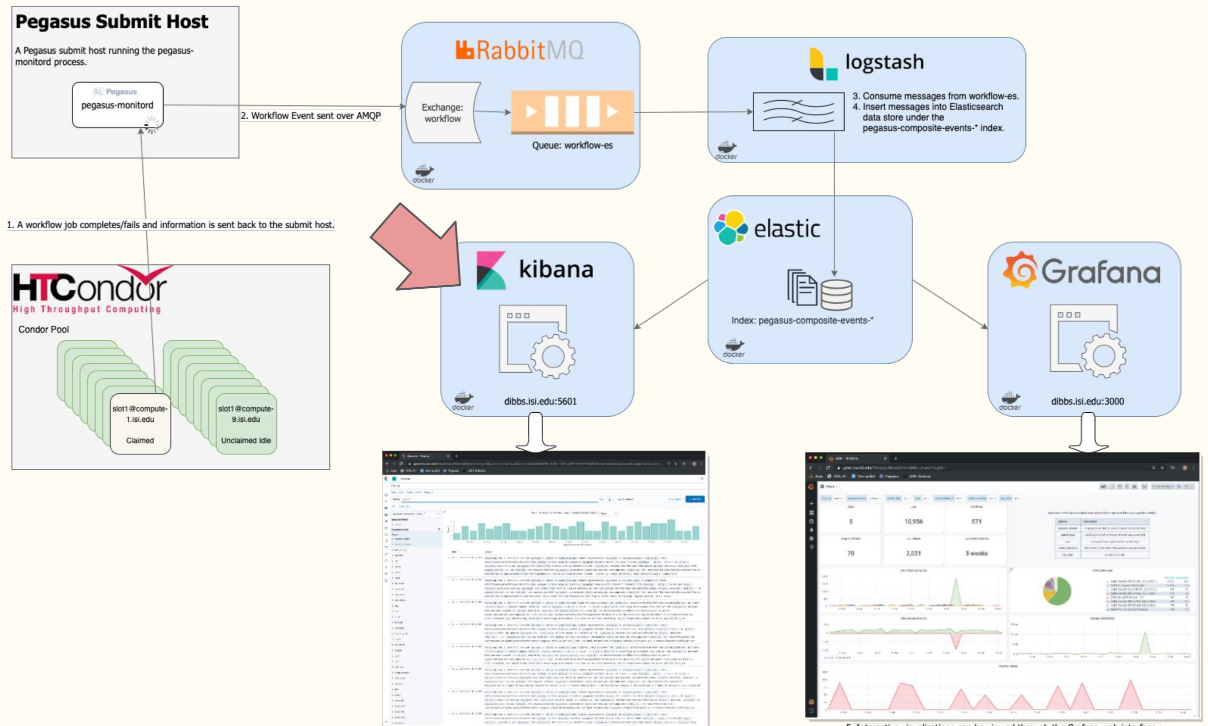
1. Workflow job completes and information sent to pegasus-monitord
2. Workflow event sent over AMQP to RabbitMQ message queue
3. Logstash consumes message from queue
4. Logstash inserts event under pegasus-composite-event-* index in Elasticsearch data store
5. Data Exploration/Visualization
   a. Workflow events viewable through Kibana
   b. Dashboard viewable through Grafana


Pegasus Submit Host
A Pegasus submit host running the pegasus-monitord process.
Pegasus
pegasus-monitord
2. Workflow Event sent over AMQP
RabbitMQ
Exchange: workflow
Queue: workflow-es
docker
logstash
3. Consume messages from workflow-es.
4. Insert messages into Elasticsearch data store under the pegasus-composite-events-* index.
docker
1. A workflow job completes/fails and information is sent back to the submit host.
HTCondor
High Throughput Computing
Condor Pool
slot1@compute-1.isi.edu
Claimed
slot1@compute-9.isi.edu
Unclaimed Idle
elastic
Index: pegasus-composite-events-*
docker
kibana
dibbs.isi.edu:5601
docker
Grafana
dibbs.isi.edu:3000
docker
5. Workflow events can be conveniently viewed through the Kibana web interface
5. Interactive visualizations can be viewed through the Grafana web interface.

Data Collection Pipeline: Data Flow

Sequence of Events:

1. Workflow job completes and information sent to pegasus-monitord
2. Workflow event sent over AMQP to RabbitMQ message queue
3. Logstash consumes message from queue
4. Logstash inserts event under pegasus-composite-event-* index in Elasticsearch data store
5. Data Exploration/Visualization
   a. Workflow events viewable through Kibana
   b. Dashboard viewable through Grafana


Pegasus Submit Host
A Pegasus submit host running the pegasus-monitord process.
Pegasus
pegasus-monitord
2. Workflow Event sent over AMQP
RabbitMQ
Exchange: workflow
Queue: workflow-es
docker
logstash
3. Consume messages from workflow-es.
4. Insert messages into Elasticsearch data store under the pegasus-composite-events-* index.
1. A workflow job completes/fails and information is sent back to the submit host.
HTCondor
High Throughput Computing
Condor Pool
slot1@compute-1.isi.edu
Claimed
slot1@compute-9.isi.edu
Unclaimed Idle
kibana
elastic
Index: pegasus-composite-events-*
Grafana
dibbs.isi.edu:5601
dibbs.isi.edu:3000

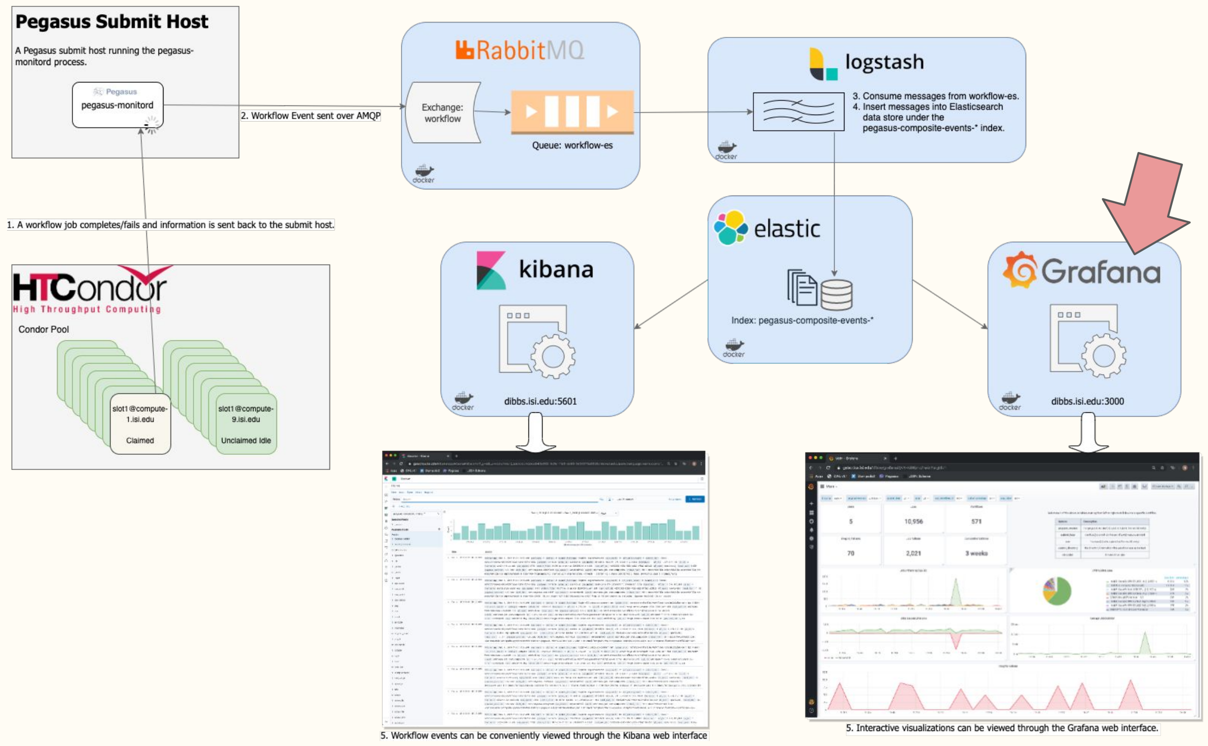
5. Workflow events can be conveniently viewed through the Kibana web interface

5. Interactive visualizations can be viewed through the Grafana web interface.

Data Collection Pipeline: Data Flow

Sequence of Events:

1. Workflow job completes and information sent to pegasus-monitord
2. Workflow event sent over AMQP to RabbitMQ message queue
3. Logstash consumes message from queue
4. Logstash inserts event under pegasus-composite-event-* index in Elasticsearch data store
5. Data Exploration/Visualization
   a. Workflow events viewable through Kibana
   b. Dashboard viewable through Grafana



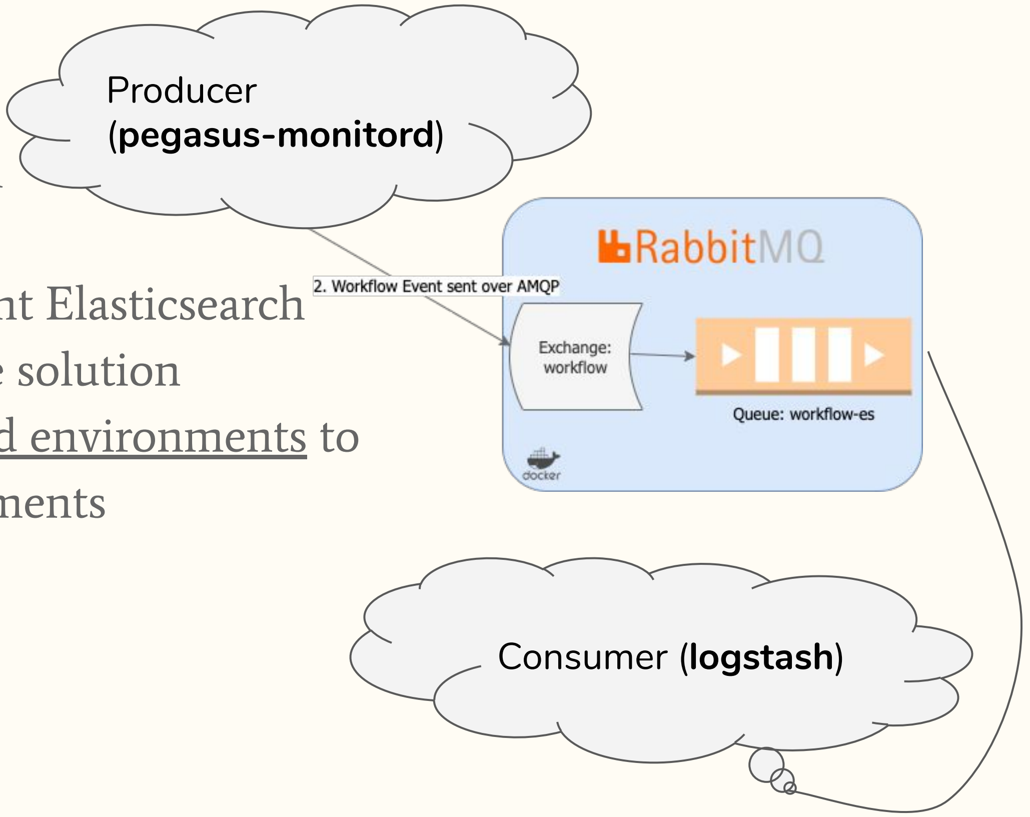
Pegasus Submit Host
A Pegasus submit host running the pegasus-monitord process.
Pegasus
pegasus-monitord
2. Workflow Event sent over AMQP
RabbitMQ
Exchange: workflow
Queue: workflow-es
logstash
3. Consume messages from workflow-es.
4. Insert messages into Elasticsearch data store under the pegasus-composite-events-* index.
1. A workflow job completes/fails and information is sent back to the submit host.
HTCondor
High Throughput Computing
Condor Pool
slot1@compute-1.isi.edu
Claimed
slot1@compute-9.isi.edu
Unclaimed Idle
elastic
Index: pegasus-composite-events-*
kibana
dibbs.isi.edu:5601
Grafana
dibbs.isi.edu:3000
docker
5. Workflow events can be conveniently viewed through the Kibana web interface
5. Interactive visualizations can be viewed through the Grafana web interface.

USC Viterbi
Information Sciences Institute

Data Collection Pipeline: RabbitMQ

- Lightweight message broker
- Relevant advantages
  - Promotes decoupling between producer/consumers
    - You can swap out/augment Elasticsearch with another data storage solution
  - Can be deployed in distributed environments to meet high availability requirements

    https://www.rabbitmq.com/



Data Collection Pipeline: ELK Stack

Elasticsearch:

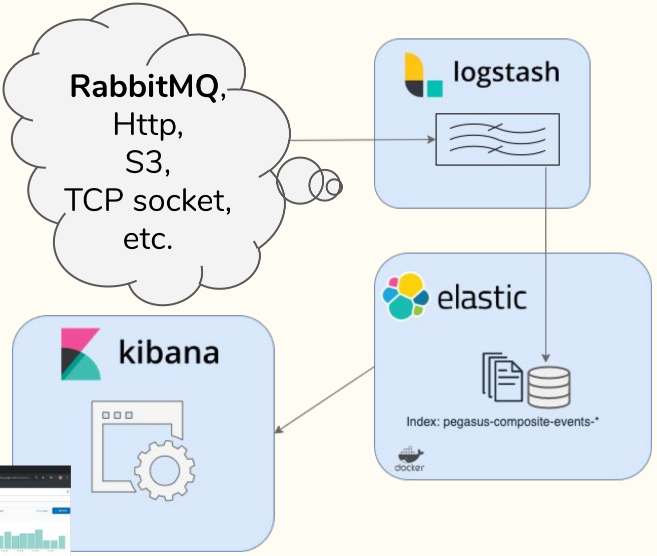
- distributed , RESTful search and analytics engine
- Log analytics, application performance monitoring, infrastructure metrics

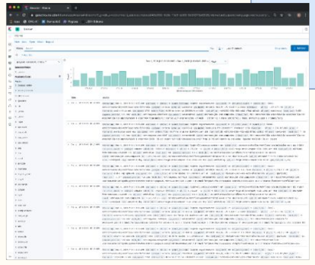
Logstash:

- Service that can aggregate and process data from various sources and insert it to elasticsearch

Kibana:

- Data visualization and exploration




5. Workflow events can be conveniently viewed through the Kibana web interface

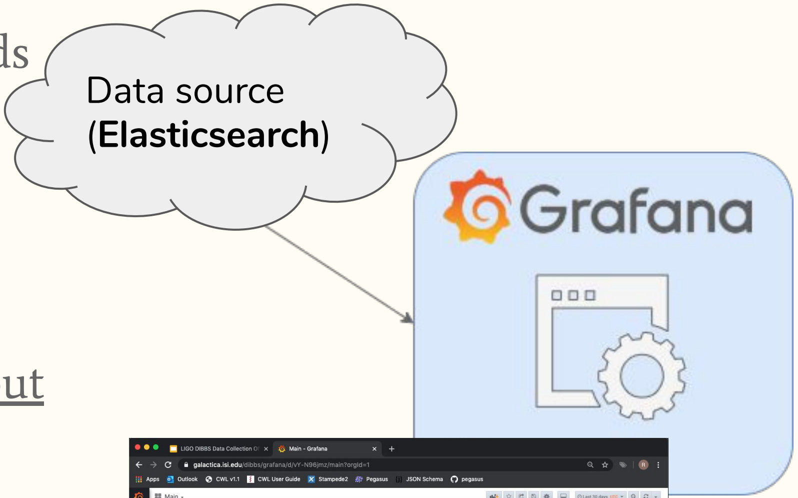
https://www.elastic.co/what-is/elk-stack

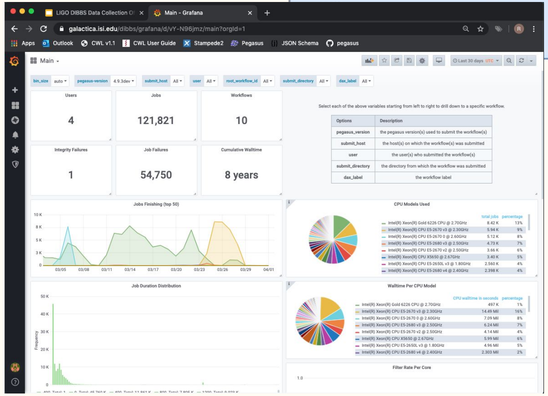
USC Viterbi
Information Sciences Institute

Data Collection Pipeline: Grafana

- Customizable browser based dashboards
- Support for multiple data sources
  - Elasticsearch
  - MySql
  - Etc.
- Enable data sharing across teams without exposing database

  https://grafana.com/





galactica.isi.edu/dibbs/grafana/d/vY-N96jmz/main?orgId=1

bin_size: auto · pegasus-version: 4.9.3dev · submit_host: All · user: All · root_workflow_id: All · submit_directory: All · dax_label: All

Users	Jobs	Workflows
4	121,821	10

Integrity Failures	Job Failures	Cumulative Walltime
1	54,750	8 years

Select each of the above variables starting from left to right to drill down to a specific workflow.

Options	Description
pegasus_version	the pegasus version(s) used to submit the workflow(s)
submit_host	the host(s) on which the workflow(s) was submitted
user	the user(s) who submitted the workflow(s)
submit_directory	the directory from which the workflow was submitted
dax_label	the workflow label

CPU Models Used

	total jobs	percentage
Intel(R) Xeon(R) Gold 6226 CPU @ 2.70GHz	8.42 K	13%
Intel(R) Xeon(R) CPU E5-2670 v3 @ 2.30GHz	5.94 K	9%
Intel(R) Xeon(R) CPU E5-2670 0 @ 2.60GHz	5.12 K	8%
Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50GHz	4.73 K	7%
Intel(R) Xeon(R) CPU E5-2670 v2 @ 2.50GHz	3.66 K	6%
Intel(R) Xeon(R) CPU X5650 @ 2.67GHz	3.40 K	5%
Intel(R) Xeon(R) CPU E5-2650L v3 @ 1.80GHz	2.560 K	4%
Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz	2.398 K	4%

Walltime Per CPU Model

	CPU walltime in seconds	percentage
Intel(R) Xeon(R) Gold 6226 CPU @ 2.70GHz	497 K	1%
Intel(R) Xeon(R) CPU E5-2670 v3 @ 2.30GHz	14.49 Mil	16%
Intel(R) Xeon(R) CPU E5-2670 0 @ 2.60GHz	7.09 Mil	8%
Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50GHz	6.24 Mil	7%
Intel(R) Xeon(R) CPU E5-2670 v2 @ 2.50GHz	4.14 Mil	4%
Intel(R) Xeon(R) CPU X5650 @ 2.67GHz	5.99 Mil	6%
Intel(R) Xeon(R) CPU E5-2650L v3 @ 1.80GHz	4.96 Mil	5%
Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz	2.303 Mil	2%

Data Collection Pipeline: How We Use It

Deployment:

- 3 node elasticsearch cluster on 3 VMs
- Single logstash instance in docker
- Single RabbitMQ instance

Usage:

- Also stores OSG data, system data
- Projects:
    - Panorama 360: Performance Data Capture and Analysis for End-to-end Scientific Workflows (https://panorama360.github.io/)
    - IRIS: Integrity Introspection For Scientific Workflows (http://nrig.renci.org/project/iris-integrity-introspection-for-scientific-workflows/)

Outline

- ~~Background~~
- ~~Data collection pipeline~~
- **Getting started**
- Demo

USC Viterbi
Information Sciences Institute

Getting Started: Prerequisites

Requirements:

- Host with a static ip (dibbs.isi.edu in this example)
  - Following ports unused:
    - 5673 <- RabbitMQ
    - 15672 <- RabbitMQ
    - 9200 <- Elasticsearch
    - 9600 <- Logstash
    - 5601 <- Kibana
    - 3000 <- Grafana
- Docker v17.02+
- Docker Compose v3.5

Getting Started: Up and Running

Clone Repository (https://github.com/pegasus-isi/dibbs-data-collection-setup):

```
git clone https://github.com/pegasus-isi/dibbs-data-collection-setup.git
```

Grant Read/Write Access Permissions:

- dibbs-data-collection-setup/**elasticsearch**/data
- dibbs-data-collection-setup/**grafana**/data
- dibbs-data-collection-setup/**kibana**/data
- dibbs-data-collection-setup/**rabbitmq**/data

Getting Started: Up and Running

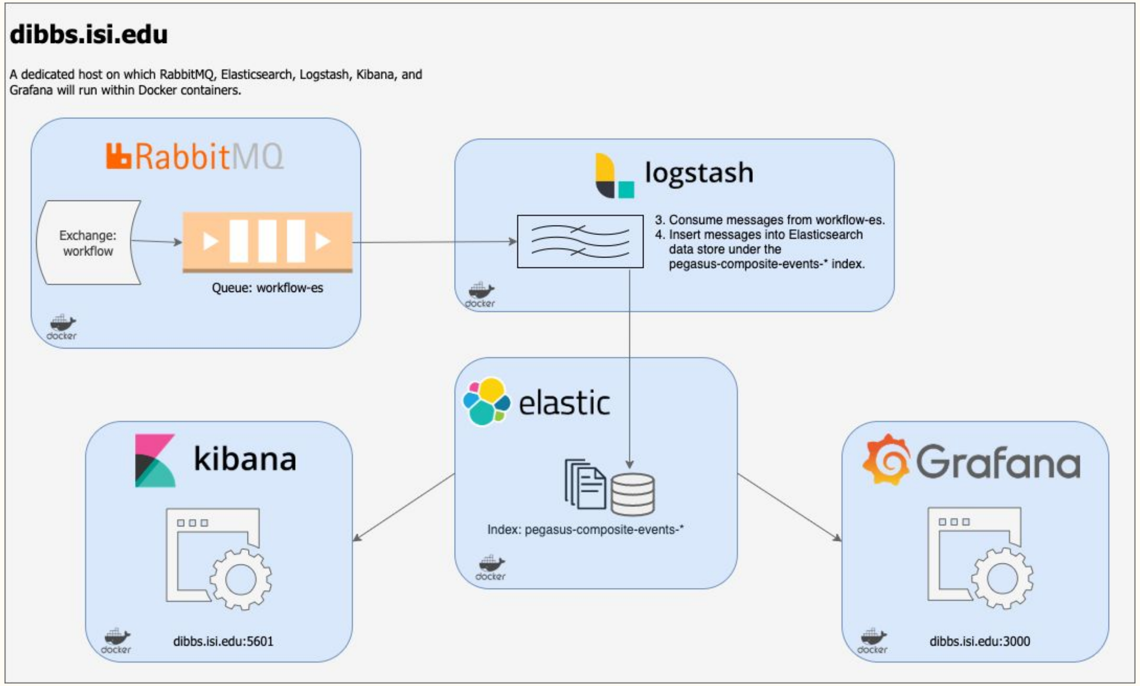
Starting the pipeline:

```
cd dibbs-data-collection-setup
```

```
docker-compose up
```

Access:

- Kibana: dibbs.isi.edu:5602
- Grafana: dibbs.isi.edu:3000


dibbs.isi.edu
A dedicated host on which RabbitMQ, Elasticsearch, Logstash, Kibana, and Grafana will run within Docker containers.
RabbitMQ
Exchange: workflow
Queue: workflow-es
logstash
3. Consume messages from workflow-es.
4. Insert messages into Elasticsearch data store under the pegasus-composite-events-* index.
elastic
Index: pegasus-composite-events-*
kibana
dibbs.isi.edu:5601
Grafana
dibbs.isi.edu:3000
docker

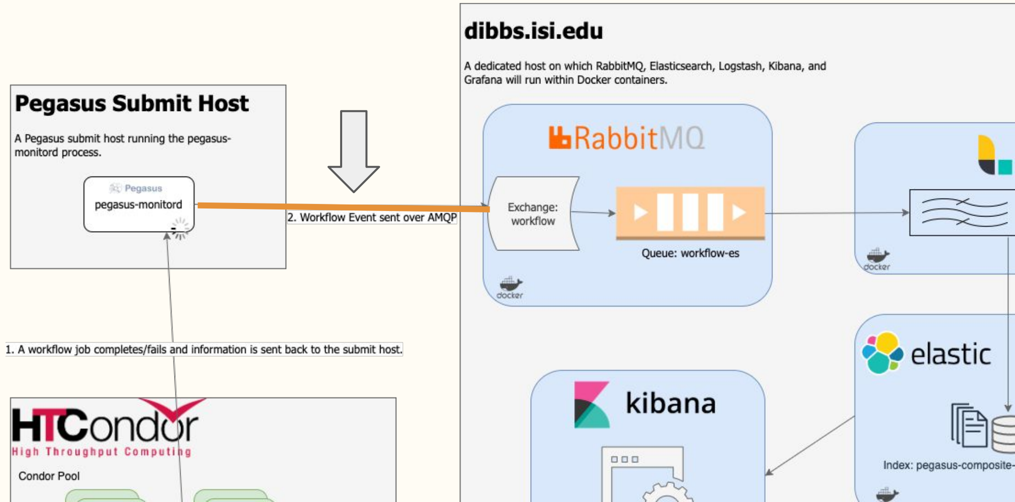
USC Viterbi Information Sciences Institute

Getting Started: Configuring Pegasus

Say that this setup is running on a host dibbs.isi.edu, then the following must be included in the pegasus configuration file used to run your workflows:

```
pegasus.monitord.encoding = json
pegasus.catalog.workflow.amqp.url = amqp://friend:donatedata@dibbs.isi.edu:5672/prod/workflows
```

**** requires Pegasus 4.9.2+***



dibbs.isi.edu
A dedicated host on which RabbitMQ, Elasticsearch, Logstash, Kibana, and Grafana will run within Docker containers.
Pegasus Submit Host
A Pegasus submit host running the pegasus-monitord process.
Pegasus
pegasus-monitord
2. Workflow Event sent over AMQP
RabbitMQ
Exchange: workflow
Queue: workflow-es
docker
1. A workflow job completes/fails and information is sent back to the submit host.
HTCondor
High Throughput Computing
Condor Pool
elastic
kibana
Index: pegasus-composite-

USC Viterbi
Information Sciences Institute

Getting Started: Sending Application Specific Metadata

**** requires Pegasus 4.9.3+***

```
@@@MONITORING_PAYLOAD - START@@@
```

```
{
   “ts”: <long>,
   “monitoring_event”: “metadata”,
   “payload”: [
      {
         “name”: <string>,
         “value”: <scalar|string>
      }
      …
   ]
}
```

```
@@@MONITORING_PAYLOAD - END@@@
```

```
# the start of the marker the monitord will look for in the stdout
echo '@@@MONITORING_PAYLOAD - START@@@'

# a json blurb describing the content following
# the actual content
cat <<EOF
{
    "ts": 1437688574,
    "monitoring_event": "metadata",
    "payload": [
        {
            "name": "num_template_banks",
            "value" : 3
        },
        {
            "name": "event_name",
            "value" : "binary start merger"
        }
    ]
}
EOF

# the end of the marker the monitord will look for in the stdout
echo '@@@MONITORING_PAYLOAD - END@@@'
```

```
"metadata_num_template_banks": 3,
"metadata_event_name": "binary start merger",
```

USC Viterbi
Information Sciences Institute

Outline

- ~~Background~~
- ~~Data collection pipeline~~
- ~~Getting started~~
- Demo

USC Viterbi
Information Sciences Institute

Pegasus est. 2001

Automate, recover, and debug scientific computations

Get Started

Pegasus Website
https://pegasus.isi.edu/

Users Mailing List
pegasus-users@isi.edu

Pegasus Website
pegasus-support@isi.edu

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
office hours

USC Viterbi
Information Sciences Institute