

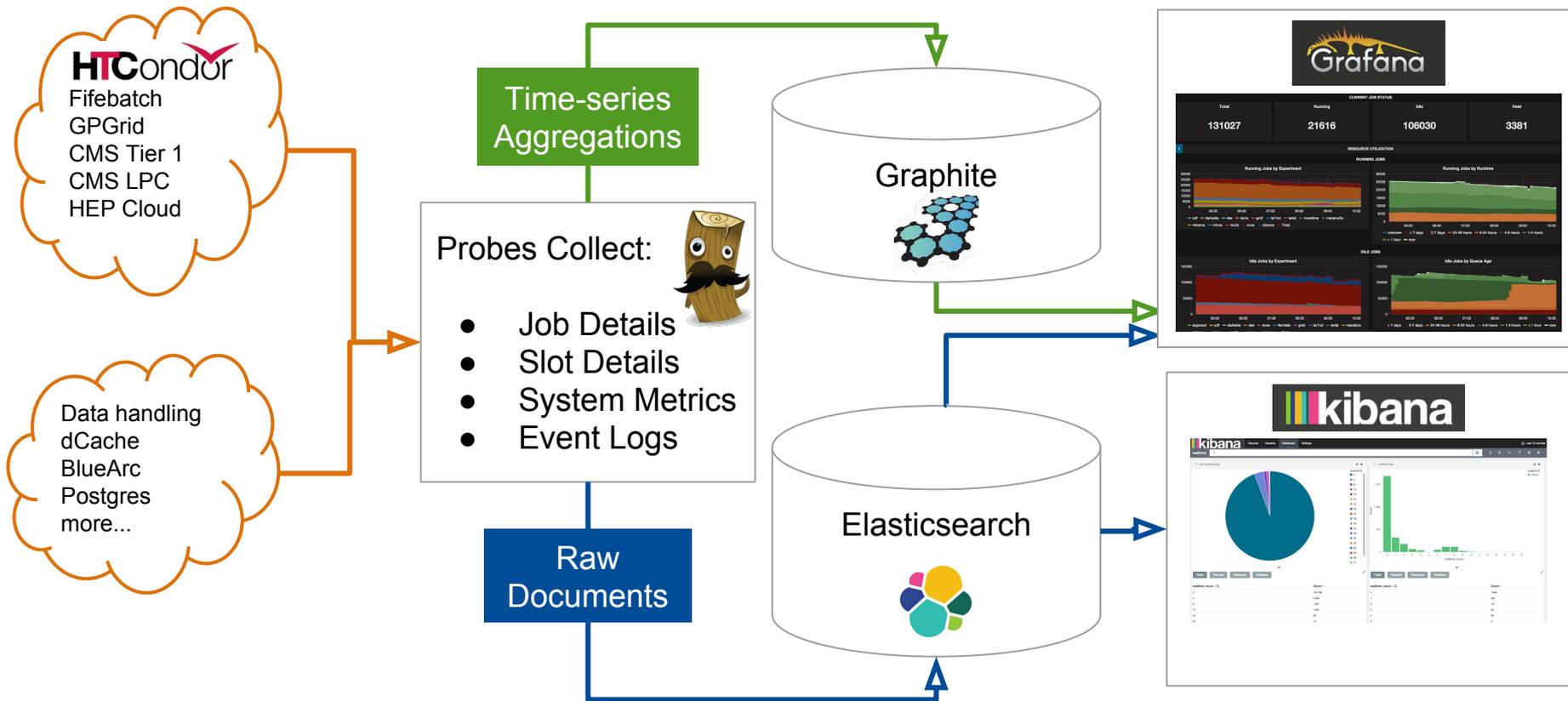
FIFE Batch System Analytics with Kibana

July 12, 2016

<https://fifemon.fnal.gov/kibana> (KCA)

<https://fifemon-pp.fnal.gov/kibana> (SSO)

FIFE Batch Monitoring Pipeline



FIFE Batch Analytics Pipeline

Probes: Collect data from batch system.



Elasticsearch: Stores fully indexed JSON documents, accessible via rich RESTful API.



Kibana: Graphical web frontend to Elasticsearch, perfect for ad-hoc queries and data exploration.

fifebatch1/2 **realtime**
Logstash collecting EventLog & JobSub logs

ifmongpvm02 **10 min**
Custom probes query fifebatch collector jobs (condor_q) and slots (condor_status)

fifemon-es
Five-node Elasticsearch cluster hosted at UNL HCC, primary purpose OSG accounting (GRACC).
<https://fifemon-es.fnal.gov>
Restricted API, certain read-only endpoints (e.g. _search) open, full access requires authorized certificate.

fifemon (fifemon1)
<https://fifemon.fnal.gov/kibana>
KCA, authorized DNs only

fifemon-pp (fifemon2)
<https://fifemon-pp.fnal.gov/kibana>
SSO, authorized users only

FIFE Batch Data in Elasticsearch (and Kibana)

Elasticsearch Concepts

- Data model and interchange format is JSON **documents** composed of multiple key:value **fields**.
- Data storage engine is Apache **Lucene**.
- Data is stored in **indices**, distributed between one or more **shards** - each shard is a Lucene database.
- Commonly data is stored in time-based **index patterns** for partitioning and scalability. Kibana will expand an index pattern (e.g. `fiFebatch-logs-YYYY.MM.dd`) based on time range being looked at. Also supports wildcards (e.g. `fiFebatch-logs-*`).

Elasticsearch Concepts

- An index can have multiple document **types**.
- A type has defined field **mappings** that are set at index time. The mapping defines the type of field (string, float, etc) and how it is **analyzed**. *Changing a field mapping requires re-indexing the data!*
- By default string fields are analyzed by lowercasing and splitting on word boundaries (space, dash, etc). Great for ordinary text search, not great for identifying strings (e.g. host names). Most strings in FIFE indices are set to `not_analyzed`.

Index: fifebatch-jobs

- Select job attributes (ClassAds) updated every 10 minutes from fifebatch HTCondor queue.
- `@timestamp` is set to the last time the job was updated.
- Once the job falls out of the queue (complete, cancel) it stops being updated, so **a job will remain forever in “running” state**. Query “Last 10m” for current jobs.
- Upcoming changes:
 - update job info from EventLog
 - capture final state
 - possibly separate indices for current and historical jobs

Index: fifebatch-jobs

Key Attributes

jobid "123456789.0@fifebatch2.fnal.gov"

cluster "123456789"

process "0"

schedd "fifebatch2.fnal.gov"

status

1: idle

2: running

~~3: cancelled~~

~~4: completed~~

5: held

Owner "novapro"

JobsubClientKerberosPrincipal

"amoren@FNAL.GOV"

JobSub_Group "nova"

Request Attributes

DESIRED_usage_model "DEDICATED"

JOB_EXPECTED_MAX_LIFETIME "28800" (s)

RequestCpus "1"

RequestMemory "2048" (MB)

memory_request "2GB" (scripted bytes)

RequestDisk "35000000" (KB)

Cmd "/fife/local/scratch/...2707850_0_1_wrap.sh"

Running Attributes

MATCH_GLIDEIN_Site "FNAL"

walltime "3065" (s)

cputime "2396" (s)

efficiency "0.782"

ResidentSetSize_RAW "1,343,928" (KB)

memory_usage "1.282GB" (scripted bytes)

DiskUsage_RAW "110" (KB)

Index: fifebatch-slots

- Select slot/machine attributes (ClassAds) updated every 10 minutes from fifebatch HTCCondor collector.
- `@timestamp` is set to the last time the slot was updated.
- Similar to jobs, once the slot disappears it stops being updated, so **a slot will remain forever in “Claimed” state**. Query “Last 10m” for current slots.

Index: fifebatch-slots

Key Attributes

Name "slot1_3@glidein_1045375_1086175008@fnpc4201.fnal.gov"

GLIDEIN_Site "FNAL"

GLIDEIN_ResourceName "GPGGrid"

SlotType

"Partitionable": pilot

"Dynamic": claimed

State "Claimed"

time_left "295596" (s)

time_left_hours "82"

time_left_days "3"

Resources

For "Partitionable" slot this is what's remaining unclaimed, for "Dynamic" this is what's claimed by the job.

Cpus "35GB"

Disk "35GB"

Memory "3GB" (raw in bytes)

Memory_mb "3072"

Memory_gb "3"

FifemonSlotWeight "1.5"

FifemonSlotWeight is fractional "standard" slots - 1 Core, 2GB RAM

Index: `fifebatch-logs-*`

- Daily indices (e.g. `fifebatch-logs-2016.07.12`)
- Currently keeping last 30 days
- Collecting logs from Fifebatch schedds:
 - HTCondor EventLog (“`type:condor_eventlog`”)
 - JobSub SubmitLog (“`type:jobsub_submitlog`”)

Type: fifebatch-logs-*/condor_eventlog

Key Attributes

jobid "123456789.0@fifebatch2.fnal.gov"
cluster "123456789"
process "0"
host "fifebatch2.fnal.gov"
Owner "novapro"
JobSub_Group "nova"
MachineAttrGLIDEIN_Site0 "FNAL"
MachineAttrGLIDEIN_ResourceName0 "GPGrid"

Event Attributes

TriggerEventTypeName "5"
TriggerEventName
 "ULOG_JOB_TERMINATED"
MyType "JobTerminatedEvent"

Key Events

1	SubmitEvent
2	ExecuteEvent
4	JobEvictedEvent
5	JobTerminatedEvent
6	JobImageSizeEvent
9	JobAbortedEvent
12	JobHeldEvent
13	JobReleasedEvent
22	JobDisconnectedEvent
23	JobReconnectedEvent

Type: fifebatch-logs-*/condor_eventlog

ImageSizeEvent (6)

ResidentSetSize "102.414MB" (raw in bytes)

resident_set_size_kb "104872"

MemoryUsage "103MB" (raw in bytes)

memory_usage_mb "103"

JobTerminatedEvent (5)

cpu_request "1"

cpu_allocated "1"

memory_request_mb "2000"

memory_allocated_mb "2000"

memory_usage_mb "1438"

disk_request_kb "35000000"

disk_allocated_kb "35705766"

disk_usage_kb "139305"

return_value "1"

JobAbortedEvent (9)

Reason "via condor_rm (by user dsmith)"

JobHeldEvent (12)

HoldReason

"SYSTEM_PERIODIC_HOLD Memory/limit
2578/2.0000000000000000E+03"

HoldReasonCode "26"

HoldReasonSubCode "1"

Type: fifebatch-logs-*/jobsub_submitlog

Key Attributes

jobid "123456789.0@fifebatch2.fnal.gov"

cluster "123456789"

process "0"

schedd "fifebatch2.fnal.gov"

user "novapro"

principal "satish@FNAL.GOV"

jobsub_version "1.2.1.5.1"

command

```
"jobsub_submit --group nova --role Production -N 1 --resource-provides=usage_model=DEDICATED
--disk=10000MB --memory=1900MB --expected-lifetime=300s -e SAM_PROJECT_NAME -e
SAM_STATION -e IFDH_BASE_URI -e IFDH_DEBUG -e EXPERIMENT -e GRID_USER satish-svd-
raw2root-test-20160711_1801-testjobs.sh --limit 1 --multifile --export 20160711_1801 --config
daq2rawdigitjob.fcl --source setup_nova.sh:-r:S16-02-02:-b:maxopt --source setup_local_test.sh -X
runNovaSAM.py -n 3 --jsonMetadata --copyOut --logs --zipLogs --outTier out1:artdaq"
```

Other Indices

`fife-events`: Outage notices, unusual event log. Used for annotations.

- e.g. <https://fifemon.fnal.gov/monitor/dashboard/db/scd-summary-fife>

`fifemon-logs-*`: Fifemon frontend logs

- Daily indices
- Grafana (“`type:grafana`”)
- Kibana (“`type:kibana`”)
- Supervisor (“`type:supervisor`”)

Kibana Overview

Overview



Discover:
Search and
Filter raw
documents

Dashboard:
Combine
visualizations

Visualize:
Graph
document
statistics

Settings:
Configure
Kibana

Timepicker:
change time
range of
interest and
refresh.

Toolbar: Save
and share
searches,
visualizations,
and
dashboards

Discover

Search

Index Pattern

Filter

Document Timeline

Documents First 500 (default).

Fields
Only fields that exist in documents that match the search terms, filters, and time range will be shown!

Discover Interface Details:

- Search bar: type:condor_eventlog
- Index Pattern: Jobsub_Group: "minerva"
- Selected Fields: @timestamp, Cluster, Jobsub_Group, MachineAttrGLIDEIN_ResourceName0, MachineAttrGLIDEIN_Site0, MyType, Owner, Return/Value, Jobid, @version, EventTime, EventTypeNumber, MachineAttrGLIDEIN_Entry_Name0, MemoryUsage, Proc, ReceivedBytes, ResidentSetSize, RunLocalUsage, RunRemoteUsage, SentBytes
- Available Fields: @timestamp, Cluster, Jobsub_Group, MachineAttrGLIDEIN_ResourceName0, MachineAttrGLIDEIN_Site0, MyType, Owner, Return/Value, Jobid, @version, EventTime, EventTypeNumber, MachineAttrGLIDEIN_Entry_Name0, MemoryUsage, Proc, ReceivedBytes, ResidentSetSize, RunLocalUsage, RunRemoteUsage, SentBytes
- Time Range: July 11th 2016, 14:17:05.218 - July 11th 2016, 18:17:05.219 — by 5 minutes
- Histogram: @timestamp per 5 minutes
- Document List (First 500):

Time	_source
July 11th 2016, 18:16:36.000	message: 028 (8885132.000.000) 07/11 18:16:36 Job ad information event triggered. Size = 1680968 Proc = 0 EventTime = "2016-07-11T18:16:36" TriggerEventTypeName = "ULOG_IMAGE_SIZE" Jobsub_Group = "minerva" MemoryUsage = 862 MachineAttrGLIDEIN_Site0 = "FNAL" TriggerEventTypeName = 6 MyType = "JobImageSizeEvent" MachineAttrGLIDEIN_ResourceName0 = "GPGrid" Owner = "sultana" Cluster = 8885132 Subproc = 0 ResidentSetSize = 881784 MachineAttrGLIDEIN_Entry_Name0 = "FNAL_GPGrid_ce02_mcore" EventTypeNumber = 28 type: condor_eventlog Jobsub_Group: minerva @timestamp: July 11th 2016, 18:16:36.000 @version: 1 tags: multiline path: /var/log/condor/EventLog
July 11th 2016, 18:16:35.000	message: 028 (8886354.000.000) 07/11 18:16:35 Job ad information event triggered. Size = 1664780 Proc = 0 EventTime = "2016-07-11T18:16:35" TriggerEventTypeName = "ULOG_IMAGE_SIZE" Jobsub_Group = "minerva" MemoryUsage = 745 MachineAttrGLIDEIN_Site0 = "FNAL" TriggerEventTypeName = 6 MyType = "JobImageSizeEvent" MachineAttrGLIDEIN_ResourceName0 = "GPGrid" Owner = "sultana" Cluster = 8886354 Subproc = 0 ResidentSetSize = 762128 MachineAttrGLIDEIN_Entry_Name0 = "FNAL_GPGrid_ce02_mcore" EventTypeNumber = 28 type: condor_eventlog Jobsub_Group: minerva @timestamp: July 11th 2016, 18:16:35.000 @version: 1 tags: multiline path: /var/log/condor/EventLog
July 11th 2016, 18:16:34.000	message: 028 (8885052.000.000) 07/11 18:16:34 Job ad information event triggered. Size = 1647988 Proc = 0 EventTime = "2016-07-11T18:16:34" TriggerEventTypeName = "ULOG_IMAGE_SIZE" Jobsub_Group = "minerva" MemoryUsage = 715 MachineAttrGLIDEIN_Site0 = "FNAL" TriggerEventTypeName = 6 MyType = "JobImageSizeEvent" MachineAttrGLIDEIN_ResourceName0 = "GPGrid" Owner = "sultana" Cluster = 8885052 Subproc = 0 ResidentSetSize = 731748 MachineAttrGLIDEIN_Entry_Name0 = "FNAL_GPGrid_ce02_mcore" EventTypeNumber = 28 type: condor_eventlog Jobsub_Group: minerva @timestamp: July 11th 2016, 18:16:34.000 @version: 1 tags: multiline path: /var/log/condor/EventLog

Search

- Kibana supports powerful full-text search queries using Elasticsearch [query_string_query](#), which are in turn based on (and turn into) Lucene expressions.
- Specify field to search against as “`field:term`”. If unspecified, term is matched against default field (which is typically `_all`, which contains all fields).
- Combine terms with boolean operations (OR, AND, NOT - caps are important!).
- Group terms with parens “`name:(foo OR bar)`”.
- Match wildcards “`name:foo*`” and regex “`name:/foo.*/*`”
- Numeric comparisons “`Memory_mb:>2000`”

Display & Filter Fields

GLIDEIN_ResourceName:GPGrid AND SlotType:Partitionable

July 12th 2016, 11:48:00.000 - July 12th 2016, 11:48:30.000

3 hits

July 12th 2016, 11:48:00.000 - July 12th 2016, 11:48:30.000 — [by second](#)

Visualize

Selected Fields

- # Memory

Quick Count (3 / 3 records)

15.625GB

100.0%

Visualize

- # Cpus
- # Disk

Available Fields

- @timestamp
- DaemonStartTime
- FifemomSlotWeight
- FractionCpus
- FractionDisk
- Memory_gb
- Memory_mb
- MyCurrentTime
- Name

Time

Cpus

Memory

Disk

Table [JSON](#)

@timestamp	July 12th 2016, 11:48:04.893
@version	1
# Cpus	8
DaemonStartTime	July 11th 2016, 22:50:47.000
# Disk	434.919GB
# FifemomSlotWeight	8
FractionCpus	1
# FractionDisk	1.002
# FractionMemory	1
GLIDECLIENT_group	GPGRID
GLIDEIN_ResourceName	GPGrid
GLIDEIN_Site	GPGrid

[Link to /fifebatch-slots/slot/slot1840qlidein_3061099_372593430840fnc3216.mat.gov](#)

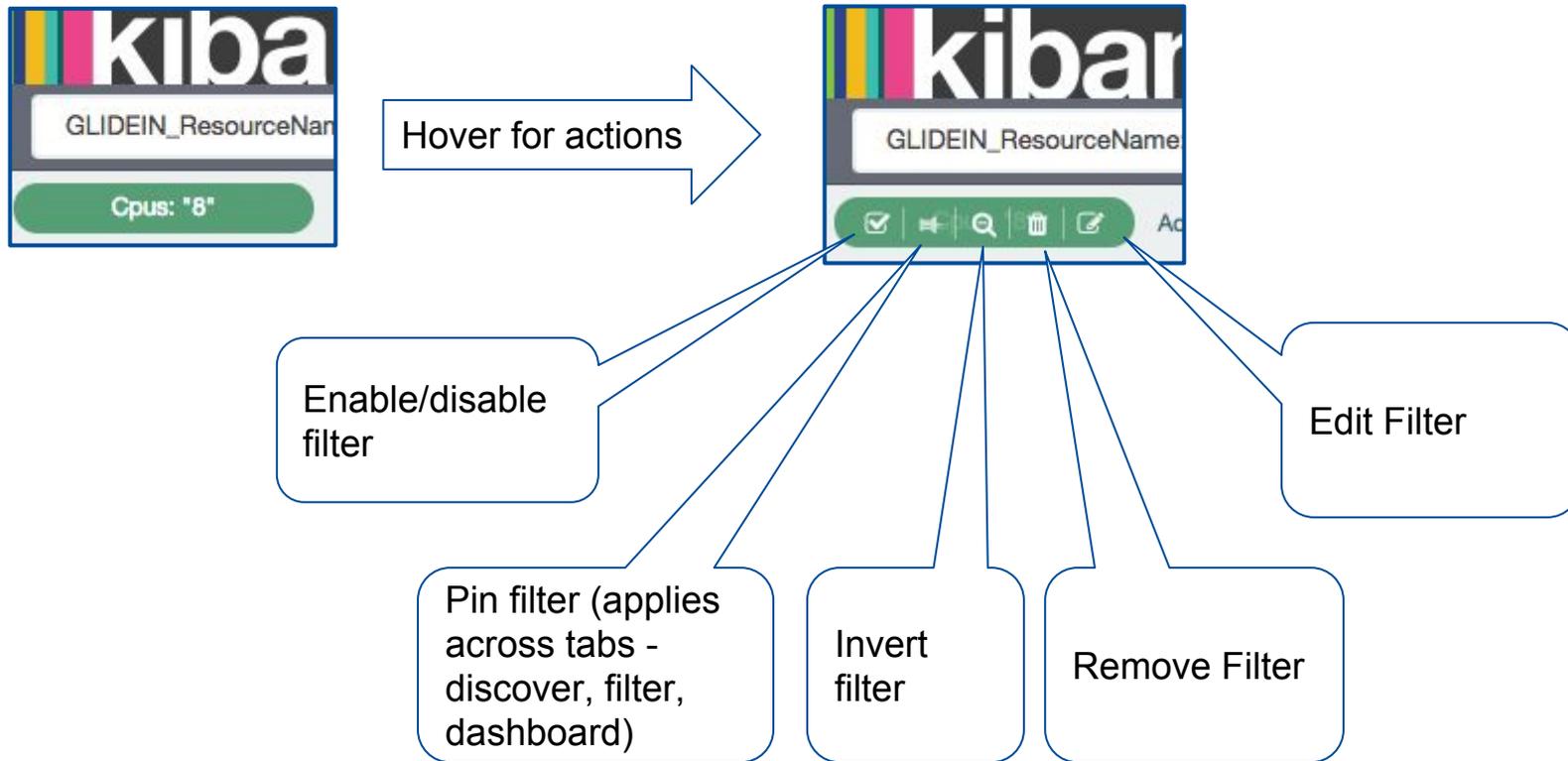
Show/hide all fields

Add/remove filter for term

Select field for display

Displayed fields. Hover for sort/move options

Filters



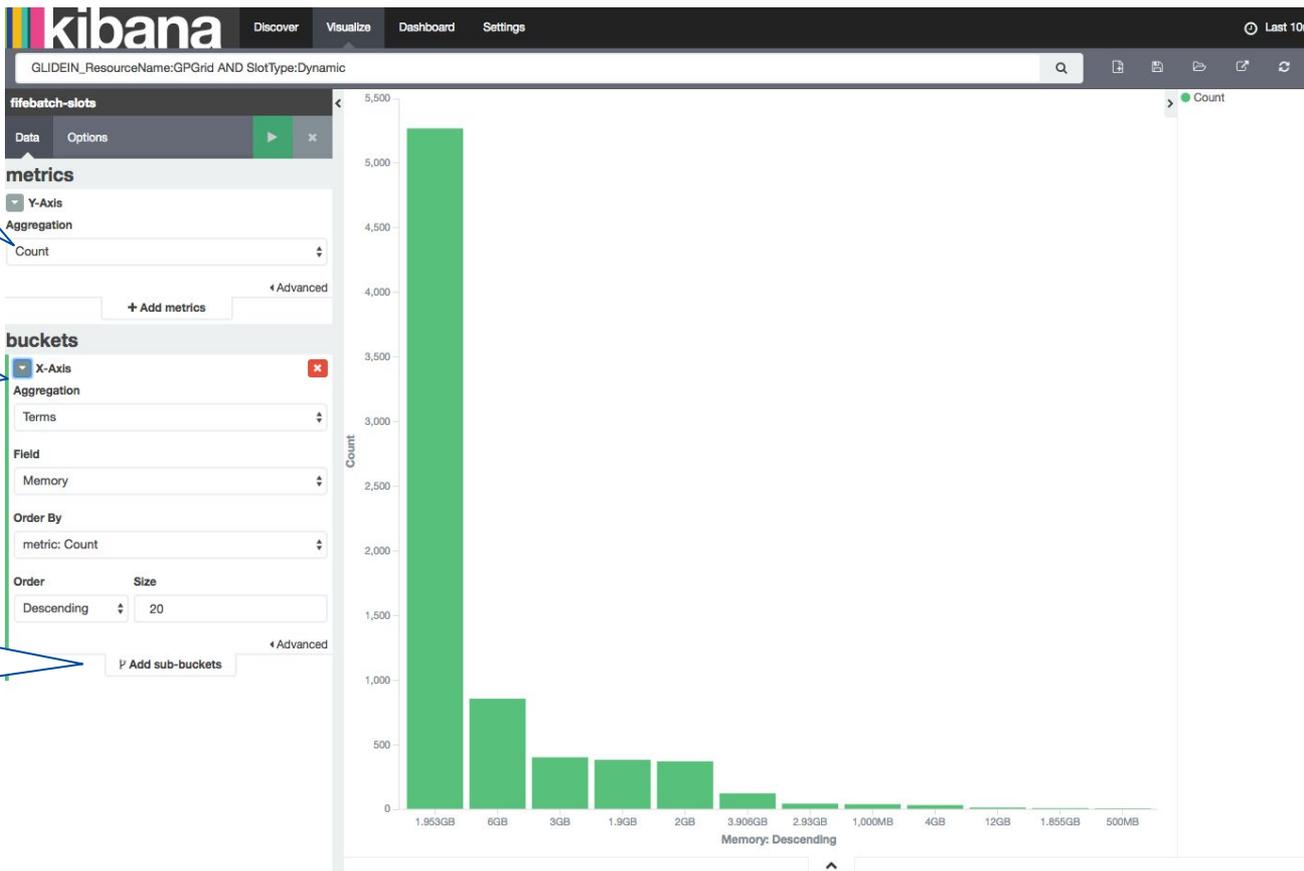
Quick Visualize

The screenshot shows the Kibana interface with the following elements:

- Search Bar:** Query: `GLIDEIN_ResourceName:GPGrid AND SlotType:D`. Results: 7,549 hits.
- Left Panel:**
 - Selected Fields: # Memory
 - Quick Count: 500 / 500 records
 - Memory usage bars: 1.953GB (73.0%), 6GB (9.4%), 3GB (5.2%), 1.9GB (4.6%), 2GB (4.0%)
 - Visualize button (highlighted with "Visualize!")
 - Available Fields: @timestamp, @version, AccountingGroup
- Chart:** Bar chart showing count over time. Title: July 12th 2016, 12:45:16.465 - July 12th 2016, 12:55:16.465 — by 10 seconds. X-axis: @timestamp per 10 seconds. Y-axis: Count (0 to 6,000).
- Table:**

Time	Cpus	Disk	Memory
July		34.699GB	1.953GB
July		34.699GB	3GB
July		34.699GB	1.953GB
July		33.431GB	1.953GB
July 12th 2016, 12:47:54.031	1	34.699GB	6GB
July 12th 2016, 12:47:54.031	1	34.533GB	1.953GB
July 12th 2016, 12:47:54.031	1	34.533GB	1.953GB

Visualize

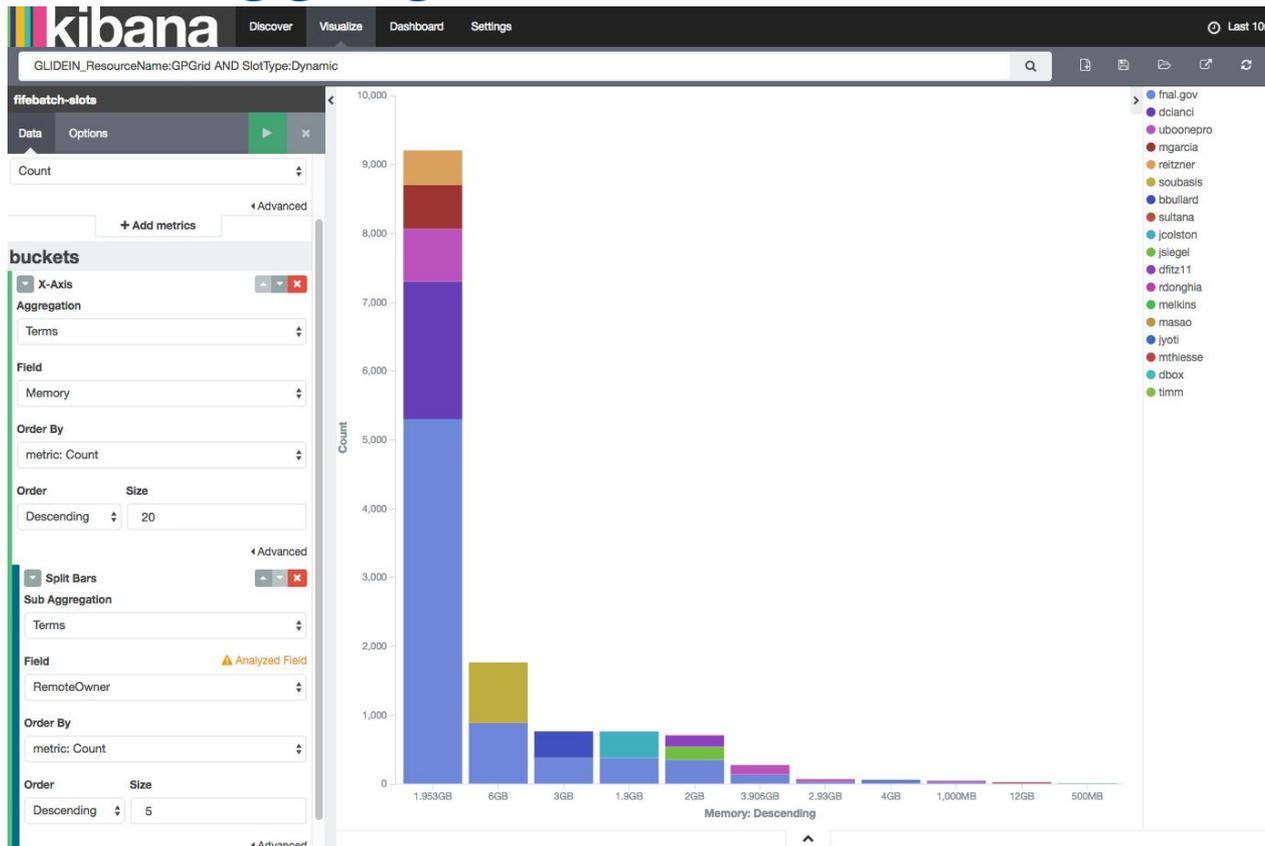


Quantity to display

How it should be grouped

Add sub-groups

Visualize: sub-aggregations



Analyzed field! I need to fix that...

Many Options...



Discover Visualize Dashboard Settings

Create a new visualization

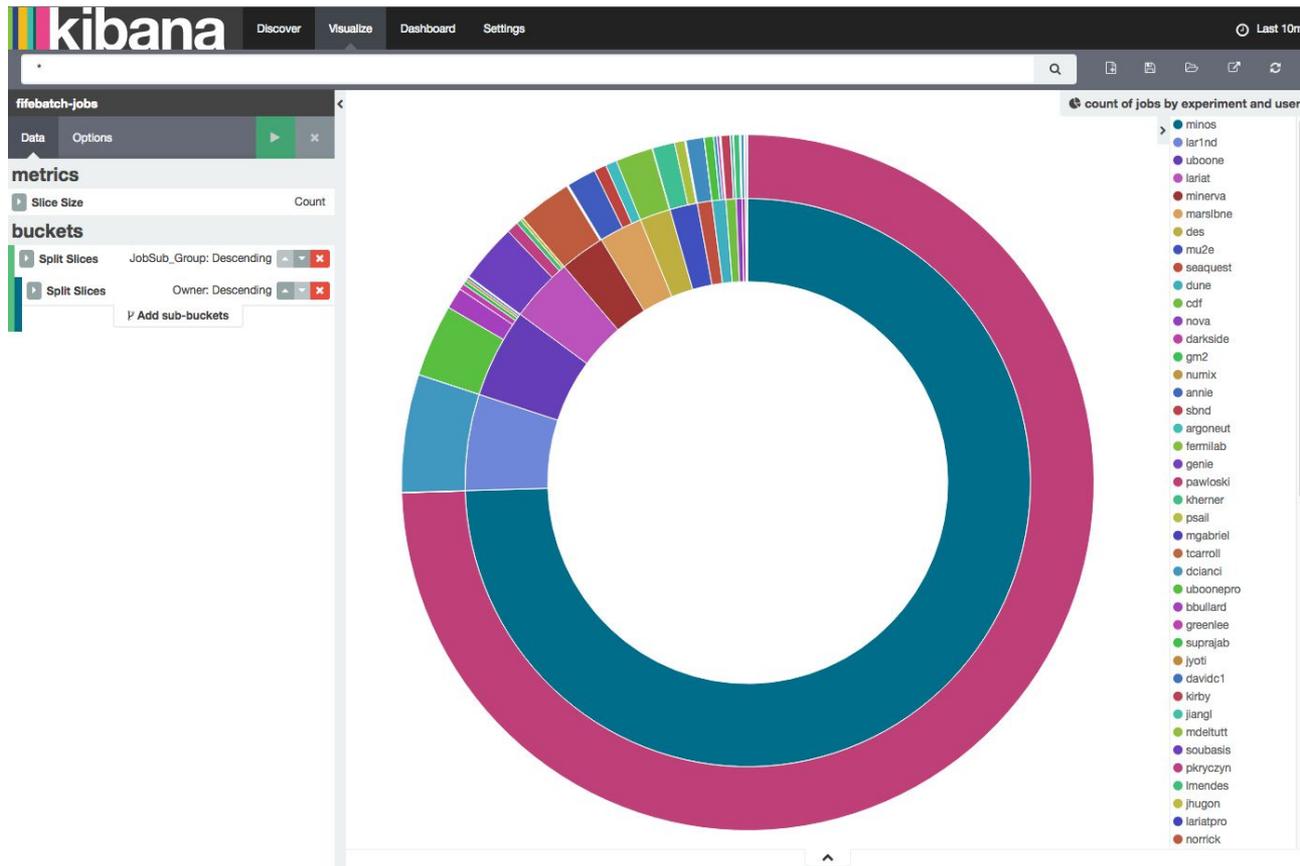
Step 1

 Area chart	Great for stacked timelines in which the total of all series is more important than comparing any two or more series. Less useful for assessing the relative change of unrelated data points as changes in a series lower down the stack will have a difficult to gauge effect on the series above it.
 Data table	The data table provides a detailed breakdown, in tabular format, of the results of a composed aggregation. Tip, a data table is available from many other charts by clicking grey bar at the bottom of the chart.
 Line chart	Often the best chart for high density time series. Great for comparing one series to another. Be careful with sparse sets as the connection between points can be misleading.
 Markdown widget	Useful for displaying explanations or instructions for dashboards.
 Metric	One big number for all of your one big number needs. Perfect for showing a count of hits, or the exact average a numeric field.
 Pie chart	Pie charts are ideal for displaying the parts of some whole. For example, sales percentages by department.Pro Tip: Pie charts are best used sparingly, and with no more than 7 slices per pie.
 Tile map	Your source for geographic maps. Requires an elasticsearch geo_point field. More specifically, a field that is mapped as type:geo_point with latitude and longitude coordinates.
 Vertical bar chart	The goto chart for oh-so-many needs. Great for time and non-time data. Stacked or grouped, exact numbers or percentages. If you are not sure which chart your need, you could do worse than to start here.

Or, open a saved visualization

Visualization Filter	manage visualizations
	30 visualizations
 AbortedJobsByVOAndUser	
 GridIFDHFailedTransferCountByHostByExperiment	

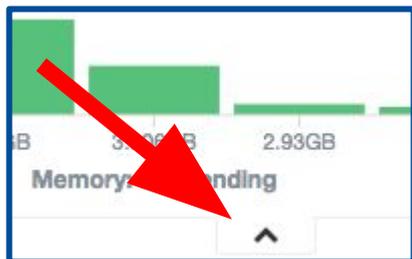
Pie Chart



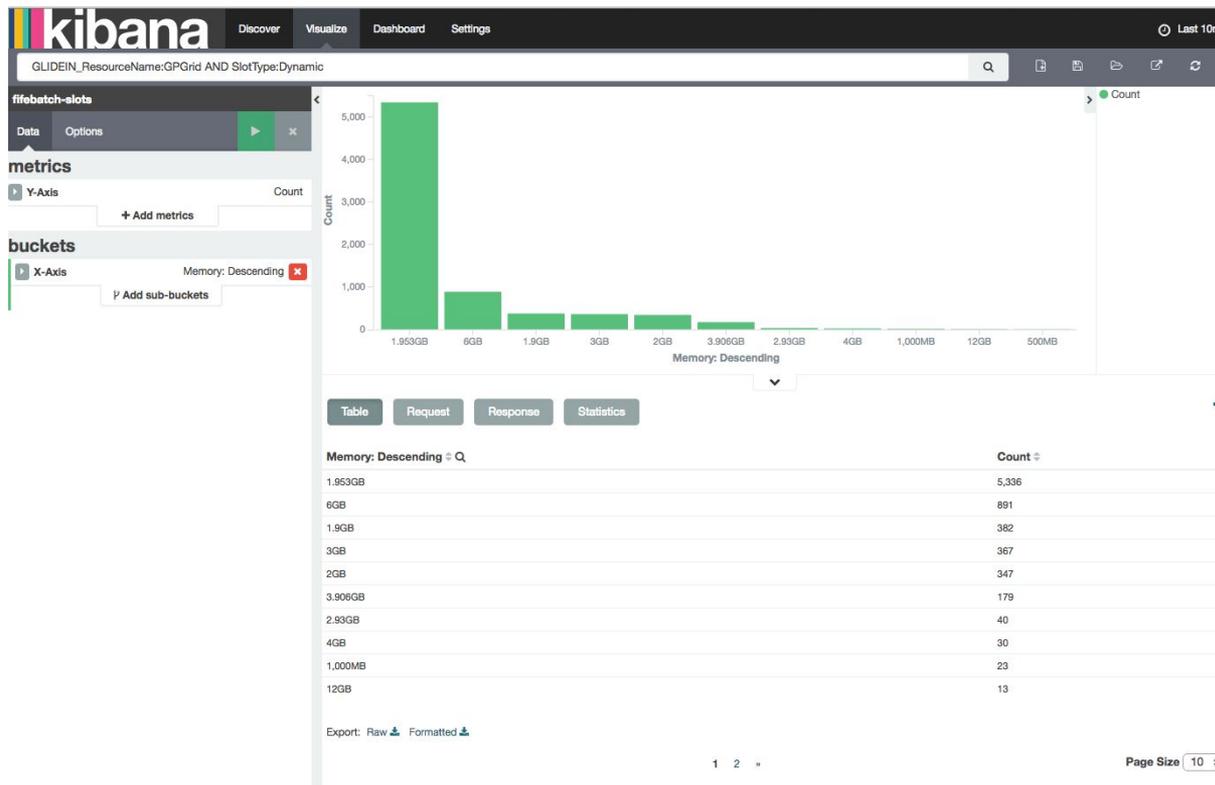
Table

You can always get a table for any visualization!

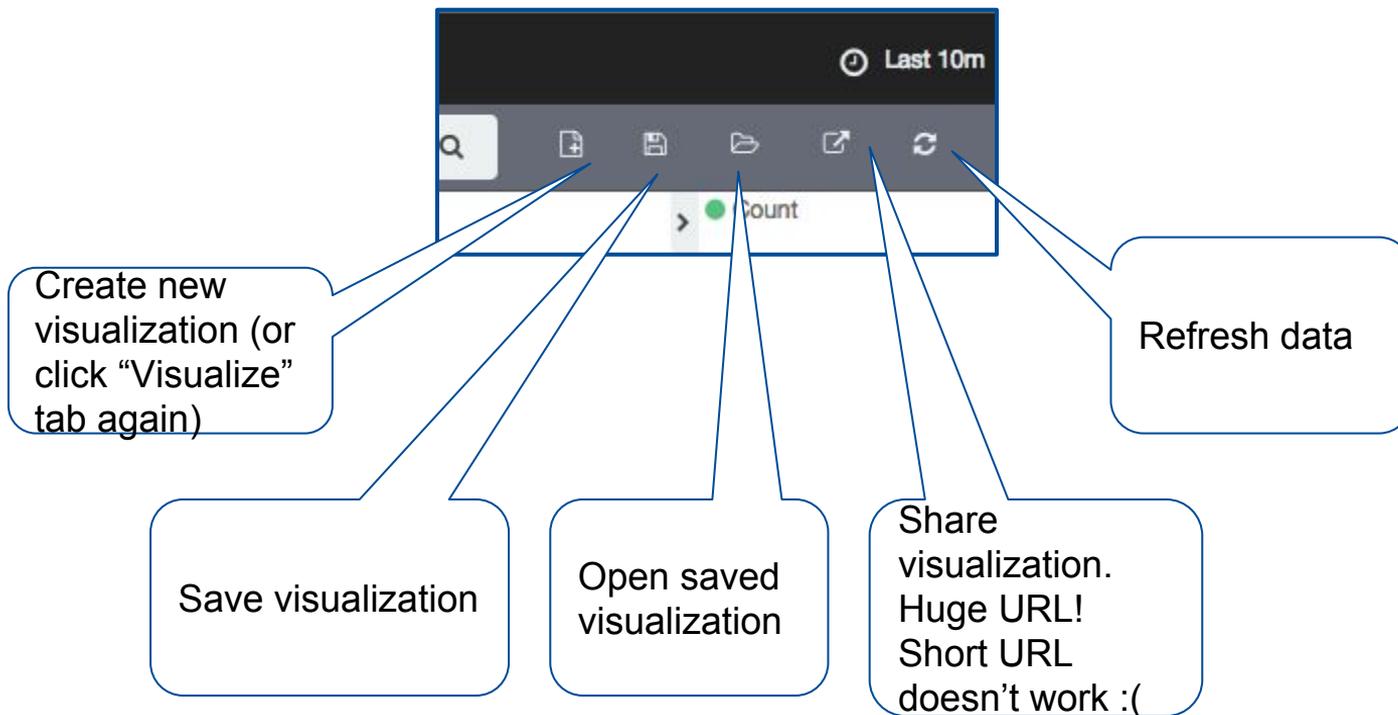
Click the arrow at the bottom to get the data table.



Can also see raw JSON search & response from Elasticsearch.

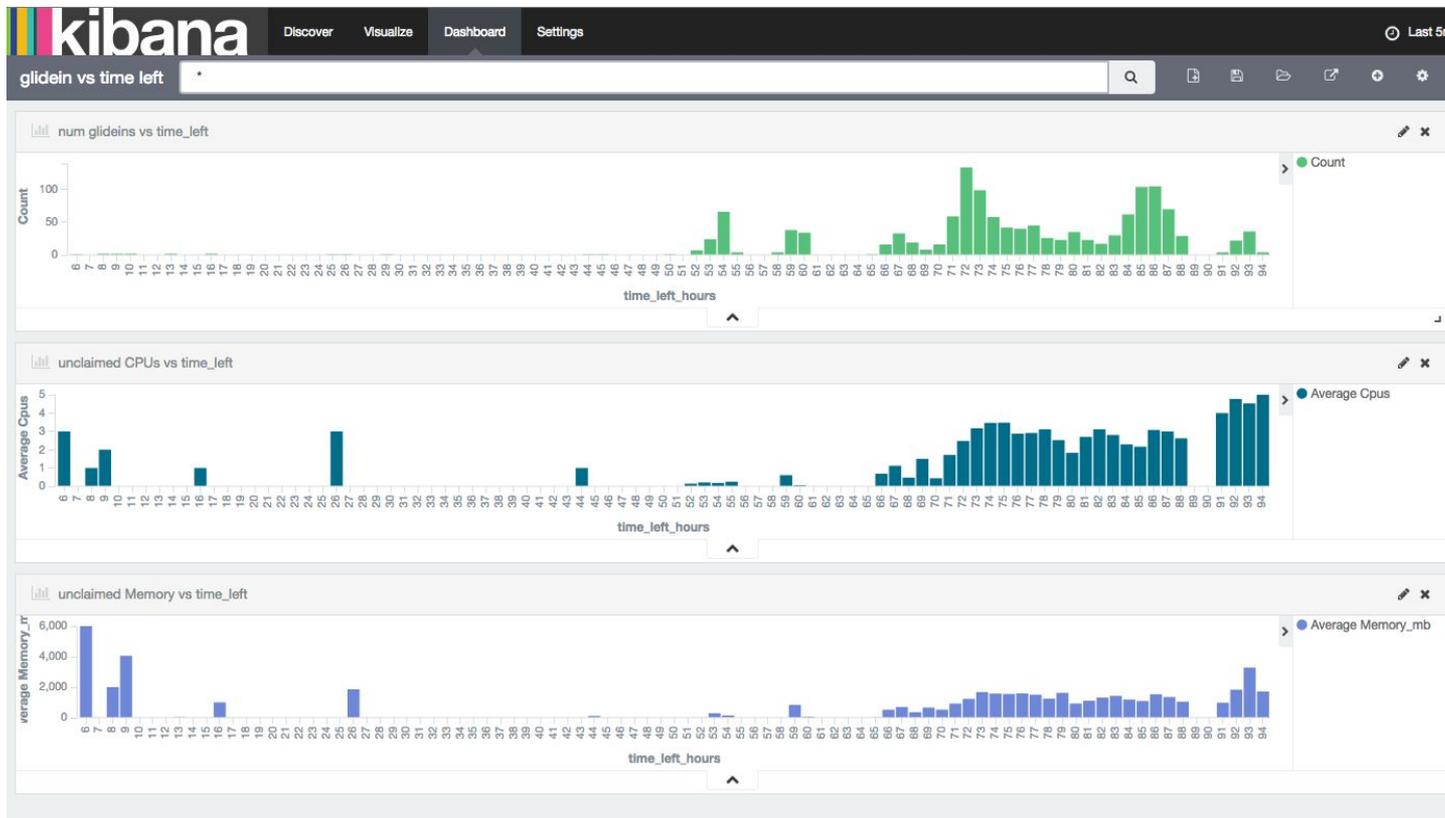


Visualization Toolbar



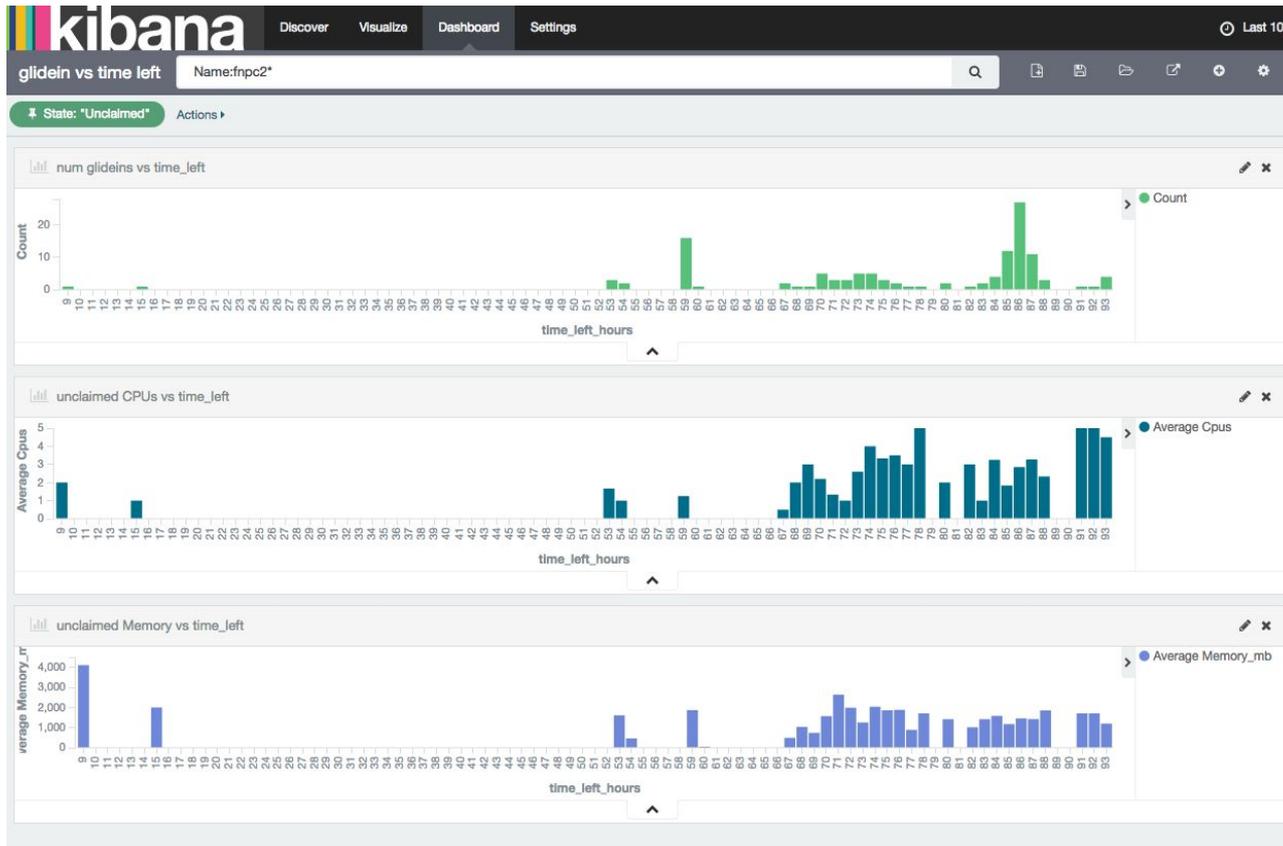
Dashboard

A dashboard is a combination of saved visualizations displayed on one page.



Dashboard

Can apply arbitrary search terms and filters.



Dashboard Toolbar

