

Data Retrieval

By Ravi Shende

Information Layout

There are 3 main data types for everything that is being retrieved:

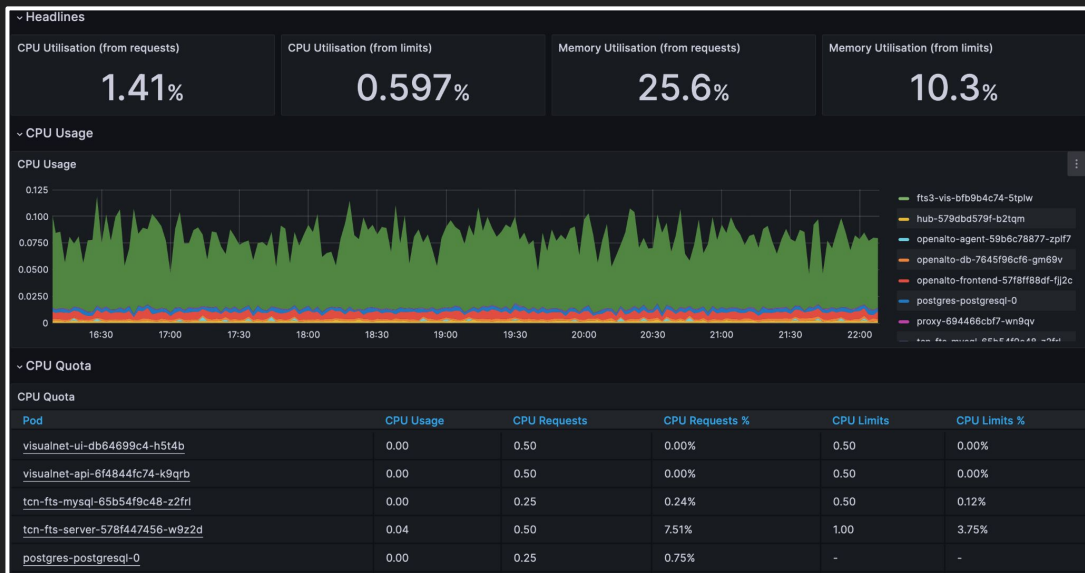
1. Data Points (Header vales - top 4 data points seen in [Grafana](#))
2. Tables
3. Graphs

Currently, all data is collected using PromQL, then represented in Pandas DataFrames, with each data frame containing values split up by Node and Pod.

What is Being Collected

All the data being collected currently is based on the main information categories displayed by [Grafana](#).

Any information collected by the site can also be collected by the program.



To the left, there is shown a portion of the information displayed on grafana.

The top 4 is the header information. Then there is a graph and table

Inputs

There are 3 main inputs that specify what information will be returned

1. Duration
 - For Tables (just Storage IO and Network Usage), specifies the time period to check for data. Then values are calculated by subtracting the two most recent points
 - For graphs, specifies the Δx for which data point values are calculated
2. Graph Step
 - How often data points are collected along an interval; resolution
3. Graph Time Offset
 - How far back from the end point to collect data from.
 - Default endpoint is current time but can be specified to any time when initializing the graphs class.

There are other inputs that can be specified (such as namespace), but these are the most commonly tweaked

Run Times and Efficiency

Currently the Header values and Tables are all collected quite quickly, within a matter of a few seconds. Graphs can sometimes take much longer to query data for, taking a few seconds to several minutes depending on the input parameters (mainly step and time offset).

In terms of code efficiency, the bottleneck is by far querying the api. Once that is done, all calculations, reformatting, and printing happen within hundredths of a second. There is not much that can be done about the time taken to query.

```
*****
*****      Graphs:      *****
8%|          |          | 0/8 [00:00<?, ?it/s]
time elapsed for querying: 32.73869367137146
total time elapsed: 32.74191379547119

12%|         |         | 1/8 [00:32<03:49, 32.74s/it]
time elapsed for querying: 93.67879388618469
total time elapsed: 93.6878908279341

25%|          |          | 2/8 [02:06<06:51, 68.59s/it]
time elapsed for querying: 68.11517818921533
total time elapsed: 68.12228989681135

38%|          |          | 3/8 [03:34<06:27, 77.51s/it]
time elapsed for querying: 49.3747879372406
total time elapsed: 49.38693889589277

50%|          |          | 4/8 [04:23<04:25, 66.41s/it]
time elapsed for querying: 11.882912949157715
total time elapsed: 11.885918169821086
```

The photo to the left shows the process of waiting while collecting graph data before it is all displayed.

Note: the green text showing time elapsed is only displayed if `show_graph_run_times = True` in `get_all_data()` in `main.py`

Data Being Collected

Per Data Type

Header Values

- CPU Utilization % (from requests)
- CPU Utilization % (from limits)
- Memory Utilization % (from requests)
- Memory Utilization % (from limits)

To the right, you can see the data frames with the information for the first 2 header categories by node and pod.

```
*****
Header:
*****
```

CPU Utilisation (from requests) %			
	Node	Pod	CPU Utilisation (from requests)
0	nrp-c12.nysernet.org	proxy-694466cbf7-w9qv	0.194183
1	nrp-c3.nysernet.org	fts3-vis-bfb9b4c74-5tplw	38.883338
2	gpu-11.nrp.mghpcc.org	hub-579dbd579f-b2tqm	1.615976
3	nrp-c16.nysernet.org	openalto-agent-59b6c78877-zplf7	0.000000
4	nrp-c16.nysernet.org	openalto-db-7645f96cf6-gm69v	2.562812
5	gpu-01.nrp.mghpcc.org	openalto-frontend-57f8ff88df-8gtkj	2.907101
6	nrp-c16.nysernet.org	postgres-postgresql-0	1.035963
7	nrp-c11.nysernet.org	tcn-fts-server-578f447456-w9z2d	11.813617
8	nrp-c12.nysernet.org	tcn-fts-mysql-65b54f9c48-z2frl	0.180633
9	rci-nrp-gpu-03.sdsu.edu	visualnet-api-6f4844fc74-7d8gr	0.000506
10	gpu-12.nrp.mghpcc.org	visualnet-ui-db64699c4-56kd8	0.000000

CPU Utilisation (from limits) %			
	Node	Pod	CPU Utilisation (from limits)
0	nrp-c12.nysernet.org	proxy-694466cbf7-w9qv	0.019418
1	nrp-c3.nysernet.org	fts3-vis-bfb9b4c74-5tplw	38.883338
2	gpu-11.nrp.mghpcc.org	hub-579dbd579f-b2tqm	0.080799
3	nrp-c16.nysernet.org	openalto-agent-59b6c78877-zplf7	0.000000
4	gpu-01.nrp.mghpcc.org	openalto-frontend-57f8ff88df-8gtkj	1.453550
5	nrp-c11.nysernet.org	tcn-fts-server-578f447456-w9z2d	5.906808
6	nrp-c12.nysernet.org	tcn-fts-mysql-65b54f9c48-z2frl	0.090317
7	rci-nrp-gpu-03.sdsu.edu	visualnet-api-6f4844fc74-7d8gr	0.000506
8	gpu-12.nrp.mghpcc.org	visualnet-ui-db64699c4-56kd8	0.000000

CPU Quota:

- ## Current Network Usage

- Current Receive Bandwidth
- Current Transmit Bandwidth
- Rate of Received Packets
- Rate of Transmitted Packets
- Rate of Received Packets Dropped
- Rate of Transmitted Packets Dropped

Memory Quota

- Memory Usage
- Memory Requests
- Memory Requests %
- Memory Limits
- Memory Limits %
- Memory Usage (RSS)
- Memory Usage (Cache)

Current Storage IO

- IOPS(Reads)
- IOPS(Writes)
- IOPS(Reads + Writes)
- Throughput(Read)
- Throughput(Write)
- Throughput(Read + Write)

=====

Tables:

=====

CPU Usage

	Pod	Node	CPU Usage	CPU Requests	CPU Requests %	CPU Limits	CPU Limits %
0	proxy-69446dcbf7-wm9qv	nnp-c12.myserver.net	0.000364436133333336857	0.2	0.178218	2	0.017822
1	ftsd-yis-bf9bc674-5td1u	nnp-c3.myserver.net	0.00432968367157658	0.5	0.086598	0.5	0.086598
2	hub-579dbd579f-b2tqm	gpu-11.nnp.mghppcc.org	0.0014735193333333324	0.1	1.423519	2	0.003247
3	openalto-agent-59b6c78877-zp1f7	nnp-c16.myserver.net	0	0.25	0.000000	0.5	0.000000
4	openalto-db-7645f9c6cf-gm9qv	nnp-c16.myserver.net	0.001858730133333334044	0.1	1.858730	0.5	0.371746
5	openalto-frontend-578f8f88d7-8gtjk	gpu-01.nnp.mghppcc.org	0.0083274881368000057	0.25	3.338433	1	0.032488
6	postgres-postgresql-0	nnp-c16.myserver.net	0.00271933066666661953	0.25	1.087734	0.5	0.543867
7	tcn-ft-server-578f447456-w9z2d	nnp-c11.myserver.net	0.02988122659966813	0.5	5.976245	0.5	5.976245
8	tcn-ft-mysql-4b554f9c48-z2fr1	nnp-c12.myserver.net	0.0004395282333333215	0.25	0.157000	0.5	0.037946
9	visualnet-agent-674844fc74-7d8gr	rci-nnp-gpu-03.sdsu.edu	0.000002857381484549764683	0.5	0.000571	NaN	NaN
10	visualnet-ui-d846497c4-56kdh	gpu-12.nnp.mghppcc.org	0	0.5	0.000000	NaN	NaN

Memory Usage

	Pod	Node	Memory Usage	Memory Requests	Memory Requests %	Memory Limits	Memory Limits %
0	proxy-69446dcbf7-wm9qv	nnp-c12.myserver.net	26587136	536878912	0.049522	1073741824	0.024761
1	ftsd-yis-bf9bc674-5td1u	nnp-c3.myserver.net	301119248	1073741824	0.280430	1073741824	0.280430
2	hub-579dbd579f-b2tqm	gpu-11.nnp.mghppcc.org	156410560	536878912	0.291710	1073741824	0.145855
3	openalto-agent-59b6c78877-zp1f7	nnp-c16.myserver.net	112711552	536878912	0.210001	1073741824	0.105845
4	openalto-db-7645f9c6cf-gm9qv	nnp-c16.myserver.net	736747552	524288000	0.140523	4294967296	0.017154
5	openalto-frontend-578f8f88d7-8gtjk	gpu-01.nnp.mghppcc.org	24078656	536878912	0.137222	1073741824	0.004641
6	postgres-postgresql-0	nnp-c16.myserver.net	174569232	268435456	0.056824	4294967296	0.004643
7	tcn-ft-server-578f447456-w9z2d	nnp-c11.myserver.net	551247872	1073741824	0.513390	2147483648	0.256695
8	tcn-ft-mysql-4b554f9c48-z2fr1	nnp-c12.myserver.net	200972480	536878912	0.309465	1073741824	0.194733
9	visualnet-agent-674844fc74-7d8gr	rci-nnp-gpu-03.sdsu.edu	132558848	1073741824	0.123455	1073741824	0.123455
10	visualnet-ui-d846497c4-56kdh	gpu-12.nnp.mghppcc.org	117450888	1073741824	0.109398	1073741824	0.109398

Current Network Usage

	Pod	Node	Current Receive Bandwidth	Current Transmit Bandwidth	Rate of Received Packets	Rate of Transmitted Packets
0	hub-579dbd579f-b2tqm	gpu-11.nnp.mghppcc.org	156	281.6	1.8	0
1	openalto-db-7645f9c6cf-gm9qv	nnp-c16.myserver.net	0	0	0	0
2	postgres-postgresql-0	nnp-c16.myserver.net	0	0	0	0
3	tcn-ft-mysql-4b554f9c48-z2fr1	nnp-c12.myserver.net	0	0	0	0
4	openalto-frontend-578f8f88d7-8gtjk	gpu-01.nnp.mghppcc.org	1078.4666666666667	18857.266666666666	565.72	15.233333333333333
5	tcn-ft-server-578f447456-w9z2d	nnp-c11.myserver.net	227.76666666666667	1020.3333333333334	266.66666666666666	44.83333333333333
6	proxy-69446dcbf7-wm9qv	nnp-c12.myserver.net	1467.6666666666667	1397.6	30.633333333333333	30.633333333333333
7	openalto-agent-59b6c78877-zp1f7	nnp-c16.myserver.net	0	0	0	0
8	ftsd-yis-bf9bc674-5td1u	nnp-c3.myserver.net	2347.6666666666667	266.66666666666666	266.66666666666666	266.66666666666666
9	visualnet-agent-674844fc74-7d8gr	rci-nnp-gpu-03.sdsu.edu	991.0666666666667	623	11.4	11.4
10	visualnet-ui-d846497c4-56kdh	gpu-12.nnp.mghppcc.org	1359.4	721.6666666666666	19.433333333333334	19.433333333333334

Current Storage IO

	Pod	Node	Current Read Bandwidth	Current Write Bandwidth	Rate of Read IOPS	Rate of Write IOPS
--	-----	------	------------------------	-------------------------	-------------------	--------------------

No Data

Graphs

- CPU Usage
- Memory Usage
- Receive Bandwidth
- Transmit Bandwidth
- Rate of Received Packets
- Rate of Transmitted Packets
- Rate of Received Packets Dropped
- Rate of Transmitted Packets Dropped
- IOPS(Read+Write)
- ThroughPut(Read+Write)

Rate of Transmitted Packets					
	Time	Node	Pod	Rate of Transmitted Packets	
0	2023-08-25 02:21:18.66199872	gpu-01.nrp.mghpcc.org	openalto-frontend-57f8ff88df-8gtkj	13.366667	
1	2023-08-25 02:26:18.661999872	gpu-01.nrp.mghpcc.org	openalto-frontend-57f8ff88df-8gtkj	8.066667	
2	2023-08-25 02:31:18.661999872	gpu-01.nrp.mghpcc.org	openalto-frontend-57f8ff88df-8gtkj	11.466667	
3	2023-08-25 03:06:18.661999872	gpu-01.nrp.mghpcc.org	openalto-frontend-57f8ff88df-8gtkj	9.133333	
4	2023-08-25 03:21:18.661999872	gpu-01.nrp.mghpcc.org	openalto-frontend-57f8ff88df-8gtkj	8.766667	
5	2023-08-25 02:21:18.661999872	gpu-11.nrp.mghpcc.org	hub-579dbd579f-b2tqm	3.333333	
6	2023-08-25 02:36:18.661999872	gpu-11.nrp.mghpcc.org	hub-579dbd579f-b2tqm	2.833333	
7	2023-08-25 02:51:18.661999872	gpu-11.nrp.mghpcc.org	hub-579dbd579f-b2tqm	2.100000	
8	2023-08-25 03:06:18.661999872	gpu-11.nrp.mghpcc.org	hub-579dbd579f-b2tqm	2.100000	
9	2023-08-25 03:21:18.661999872	gpu-12.nrp.mghpcc.org	hub-579dbd579f-b2tqm	3.033333	
10	2023-08-25 02:21:18.661999872	gpu-12.nrp.mghpcc.org	visualnet-ui-db64699c4-56kd8	10.833333	
11	2023-08-25 02:36:18.661999872	gpu-12.nrp.mghpcc.org	visualnet-ui-db64699c4-56kd8	8.466667	
12	2023-08-25 02:51:18.661999872	gpu-12.nrp.mghpcc.org	visualnet-ui-db64699c4-56kd8	8.466667	
13	2023-08-25 03:06:18.661999872	gpu-12.nrp.mghpcc.org	visualnet-ui-db64699c4-56kd8	7.766667	
14	2023-08-25 03:21:18.661999872	gpu-12.nrp.mghpcc.org	visualnet-ui-db64699c4-56kd8	8.566667	
15	2023-08-25 02:21:18.661999872	nrp-c11.nysernet.org	tcn-fts-server-578f447456-w9z2d	50.833333	
16	2023-08-25 02:36:18.661999872	nrp-c11.nysernet.org	tcn-fts-server-578f447456-w9z2d	54.933333	
17	2023-08-25 02:51:18.661999872	nrp-c11.nysernet.org	tcn-fts-server-578f447456-w9z2d	30.133333	
18	2023-08-25 03:06:18.661999872	nrp-c11.nysernet.org	tcn-fts-server-578f447456-w9z2d	40.600000	
19	2023-08-25 03:21:18.661999872	nrp-c11.nysernet.org	tcn-fts-server-578f447456-w9z2d	29.415688	
20	2023-08-25 02:21:18.661999872	nrp-c12.nysernet.org	proxy-694466cbf7-wm9qv	9.066667	
21	2023-08-25 02:36:18.661999872	nrp-c12.nysernet.org	proxy-694466cbf7-wm9qv	12.733333	
22	2023-08-25 02:51:18.661999872	nrp-c12.nysernet.org	proxy-694466cbf7-wm9qv	11.498677	
23	2023-08-25 03:06:18.661999872	nrp-c12.nysernet.org	proxy-694466cbf7-wm9qv	10.466667	
24	2023-08-25 03:21:18.661999872	nrp-c12.nysernet.org	proxy-694466cbf7-wm9qv	11.466667	
25	2023-08-25 02:21:18.661999872	nrp-c12.nysernet.org	tcn-fts-mysql-6b5b4f9c48-z2f1l	0.000000	
26	2023-08-25 02:36:18.661999872	nrp-c12.nysernet.org	tcn-fts-mysql-6b5b4f9c48-z2f1l	0.000000	
27	2023-08-25 02:51:18.661999872	nrp-c12.nysernet.org	tcn-fts-mysql-6b5b4f9c48-z2f1l	0.000000	
28	2023-08-25 03:06:18.661999872	nrp-c12.nysernet.org	tcn-fts-mysql-6b5b4f9c48-z2f1l	0.000000	
29	2023-08-25 03:21:18.661999872	nrp-c12.nysernet.org	tcn-fts-mysql-6b5b4f9c48-z2f1l	0.000000	
30	2023-08-25 02:21:18.661999872	nrp-c16.nysernet.org	openalto-agent-59b6c78877-zpl1f	0.000000	
31	2023-08-25 02:36:18.661999872	nrp-c16.nysernet.org	openalto-agent-59b6c78877-zpl1f	0.000000	
32	2023-08-25 02:51:18.661999872	nrp-c16.nysernet.org	openalto-agent-59b6c78877-zpl1f	0.000000	
33	2023-08-25 03:06:18.661999872	nrp-c16.nysernet.org	openalto-agent-59b6c78877-zpl1f	0.000000	
34	2023-08-25 03:21:18.661999872	nrp-c16.nysernet.org	openalto-agent-59b6c78877-zpl1f	0.000000	
35	2023-08-25 02:21:18.661999872	nrp-c16.nysernet.org	openalto-db-7645f96cf6-gm09v	0.000000	
36	2023-08-25 02:36:18.661999872	nrp-c16.nysernet.org	openalto-db-7645f96cf6-gm09v	0.000000	
37	2023-08-25 02:51:18.661999872	nrp-c16.nysernet.org	openalto-db-7645f96cf6-gm09v	0.000000	
38	2023-08-25 03:06:18.661999872	nrp-c16.nysernet.org	openalto-db-7645f96cf6-gm09v	0.000000	
39	2023-08-25 03:21:18.661999872	nrp-c16.nysernet.org	openalto-db-7645f96cf6-gm09v	0.000000	
40	2023-08-25 02:21:18.661999872	nrp-c16.nysernet.org	postgres-postgresql-0	0.000000	
41	2023-08-25 02:36:18.661999872	nrp-c16.nysernet.org	postgres-postgresql-0	0.000000	
42	2023-08-25 02:51:18.661999872	nrp-c16.nysernet.org	postgres-postgresql-0	0.000000	
43	2023-08-25 03:06:18.661999872	nrp-c16.nysernet.org	postgres-postgresql-0	0.000000	
44	2023-08-25 03:21:18.661999872	nrp-c16.nysernet.org	postgres-postgresql-0	0.000000	
45	2023-08-25 02:21:18.661999872	nrp-c3.nysernet.org	fts3-vis-bfb9b4c74-5tplw	192.300000	
46	2023-08-25 02:36:18.661999872	nrp-c3.nysernet.org	fts3-vis-bfb9b4c74-5tplw	8.266667	
47	2023-08-25 02:51:18.661999872	nrp-c3.nysernet.org	fts3-vis-bfb9b4c74-5tplw	7.166667	
48	2023-08-25 03:06:18.661999872	nrp-c3.nysernet.org	fts3-vis-bfb9b4c74-5tplw	175.866667	
49	2023-08-25 03:21:18.661999872	nrp-c3.nysernet.org	fts3-vis-bfb9b4c74-5tplw	18.233333	
50	2023-08-25 02:21:18.661999872	rci-nrp-gpu-03.sdsu.edu	visualnet-api-6f4844fc74-7d8gr	9.933333	
51	2023-08-25 02:36:18.661999872	rci-nrp-gpu-03.sdsu.edu	visualnet-api-6f4844fc74-7d8gr	6.599340	
52	2023-08-25 02:51:18.661999872	rci-nrp-gpu-03.sdsu.edu	visualnet-api-6f4844fc74-7d8gr	8.400000	
53	2023-08-25 03:06:18.661999872	rci-nrp-gpu-03.sdsu.edu	visualnet-api-6f4844fc74-7d8gr	8.933333	
54	2023-08-25 03:21:18.661999872	rci-nrp-gpu-03.sdsu.edu	visualnet-api-6f4844fc74-7d8gr	9.700000	

The above picture shows the data frame for the graph Rate of Transmitted Packets. Note for each node and pod, there are data points for different times. Each pod would represent a different colored line on a graph.

Additional Features

Sorted by Data Type

All Data Types

- Filter all data collected to only include pods that are bp3d-worker pods.
Display the pod name as the ensemble.
 - Ex: “bp3d-worker-f605f10” becomes “f605f10”
 - Set `only_include_worker_pods` to `True` in `main.py` in `get_all_data()`
- Store all data in one multi-level dictionary.
 - In `main.py`: `result_dict = get_all_data()`
 - Returns a dictionary of dictionaries.
 - `{‘header’: header_dict, ‘tables’: tables_dict, ‘graphs’: graphs_dict}`

Graphs

- Check for potential pods with data that was dropped and/or recovered (nonzero → 0 or 0 → nonzero). Requery those graphs for higher resolution at the drop time (optional)
 - To know if certain pods are untrustworthy (if they frequently drop data, they aren't trustworthy)
 - . Requerying is meant for checking if pods were actually dropped or if their values slowly went to 0.
 - in main.py calling check_graphs_losses().
 - Set print_info to True to print all information on potential pods dropped and recovered.
 - Set requery to True/False, or None if you want to prompt the user in the terminal.
- Display graphs as graphs instead of dataframes
 - Run graph_visualization.py
- Collect all graphs as one large dataframe instead of one dataframe per graph (useful for inputting graphs into a database)
 - Set get_graphs_as_one_df to True in main.py in get_all_data()
- Display the times in the time column as seconds since epoch (01/01/1970) or as a timestamp (%Y-%m-%d %H:%M:%S)
 - Set display_time_as_timestamp to False or True respectively in get_all_data() in main.py

Future Direction

With everything in pandas data frames, it shouldn't prove too challenging to integrate this data with any database we like.

If we choose to go with a time based database like InfluxDB, it is simple to include timestamps for header and tables, as they are already being collected by the queries, just not displayed.

- To Include them, edit `_generate_df` in `header.py` and `tables.py`
 - Uncomment the 6 lines containing the inline comment “ # for timestamp”, then delete the previous line if it's being rewritten by this new line
 - (Optional): add code to choose to display them as timestamps instead of seconds since epoch (look at `get_graphs_dict` in `graphs.py` for inspiration on how to do so)

For additional enhancements, there was the goal of collecting node temperature and power usage, however after further research, this information does not seem to be stored by nautilus, therefore it cannot be queried for. For more information on which node statistics can be queried, go to `extras/node_temp_power.pdf`