Data Retrieval

By Ravi Shende

Information Layout

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

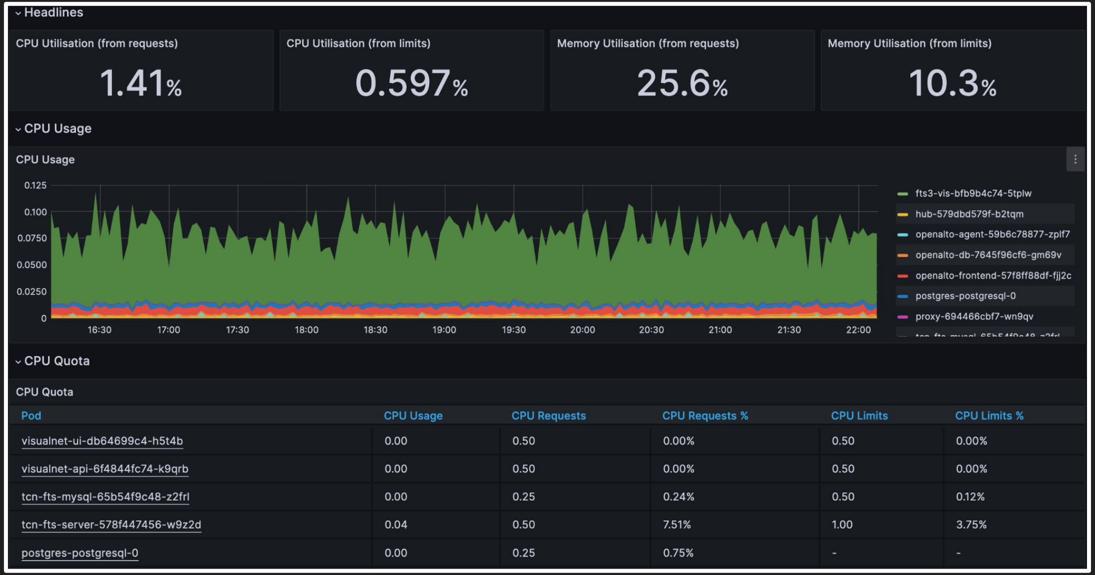
1. Data Points (Header vales - top 4 data points seen in Grafana)
2. Tables
3. Graphs

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.



Headlines

CPU Utilisation (from requests)	CPU Utilisation (from limits)	Memory Utilisation (from requests)	Memory Utilisation (from limits)
1.41%	0.597%	25.6%	10.3%

CPU Usage

CPU Usage

CPU Quota

CPU Quota

Pod	CPU Usage	CPU Requests	CPU Requests %	CPU Limits	CPU Limits %
visualnet-ui-db64699c4-h5t4b	0.00	0.50	0.00%	0.50	0.00%
visualnet-api-6f4844fc74-k9qrb	0.00	0.50	0.00%	0.50	0.00%
tcn-fts-mysql-65b54f9c48-z2frl	0.00	0.25	0.24%	0.50	0.12%
tcn-fts-server-578f447456-w9z2d	0.04	0.50	7.51%	1.00	3.75%
postgres-postgresql-0	0.00	0.25	0.75%	-	-

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 4 main inputs that specify what information will be returned

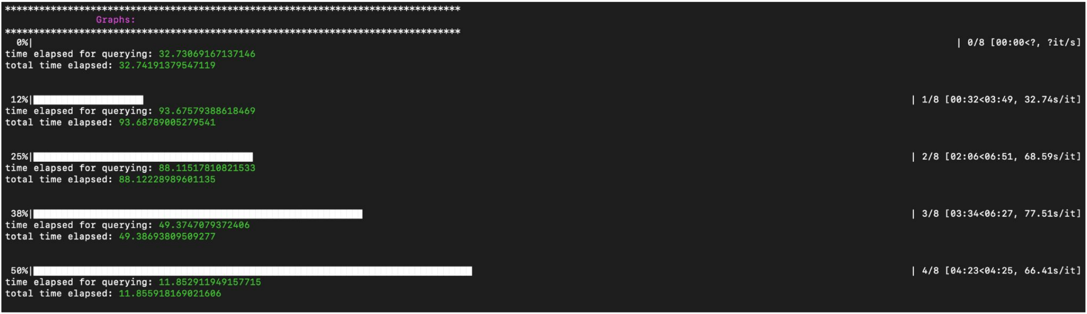
1. Duration
   - For Tables (just Storage IO and Network Usage), specifies the time period to check for data. Then values are calculated by subtracting the two most recent points
   - For graphs, specifies the Δx for which data point values are calculated
2. Graph Step
   - How often data points are 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.
4. Final Graph Time
   - End time for graph sampling: when the last datapoint on the graphs will be.

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:
********************************************************************************
  0%|                                                                                    | 0/8 [00:00<?, ?it/s]
time elapsed for querying: 32.73069167137146
total time elapsed: 32.74191379547119


 12%|███████████████████                                                     | 1/8 [00:32<03:49, 32.74s/it]
time elapsed for querying: 93.67579388618469
total time elapsed: 93.68789005279541


 25%|███████████████████████████████████████                                 | 2/8 [02:06<06:51, 68.59s/it]
time elapsed for querying: 88.11517810821533
total time elapsed: 88.12228989601135


 38%|██████████████████████████████████████████████████████████              | 3/8 [03:34<06:27, 77.51s/it]
time elapsed for querying: 49.3747079372406
total time elapsed: 49.38693809509277


 50%|█████████████████████████████████████████████████████████████████████████████ | 4/8 [04:23<04:25, 66.41s/it]
time elapsed for querying: 11.852911949157715
total time elapsed: 11.855918169021606
```

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

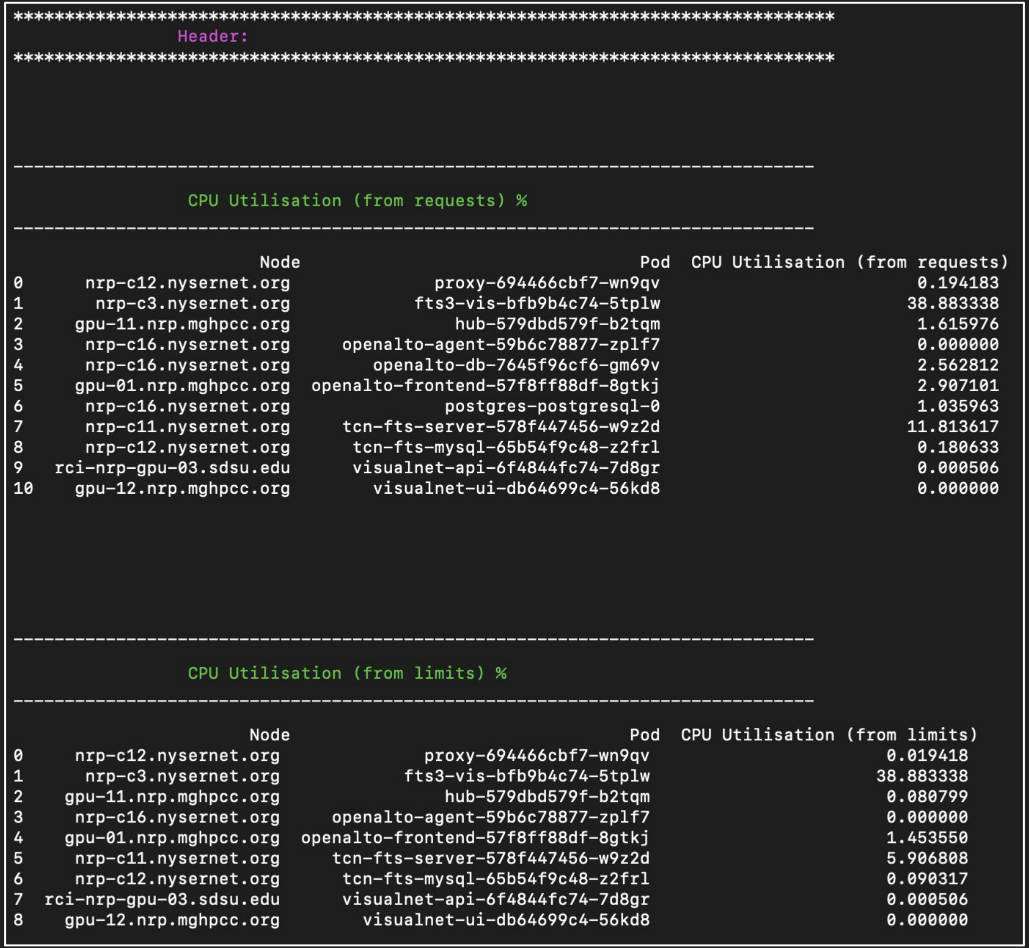
Data Being Collected

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-wn9qv                         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-wn9qv                       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
```

Tables

CPU Quota:

- CPU Usage
- CPU Requests
- CPU Requests %
- CPU Limits
- CPU Limits %

Memory Quota

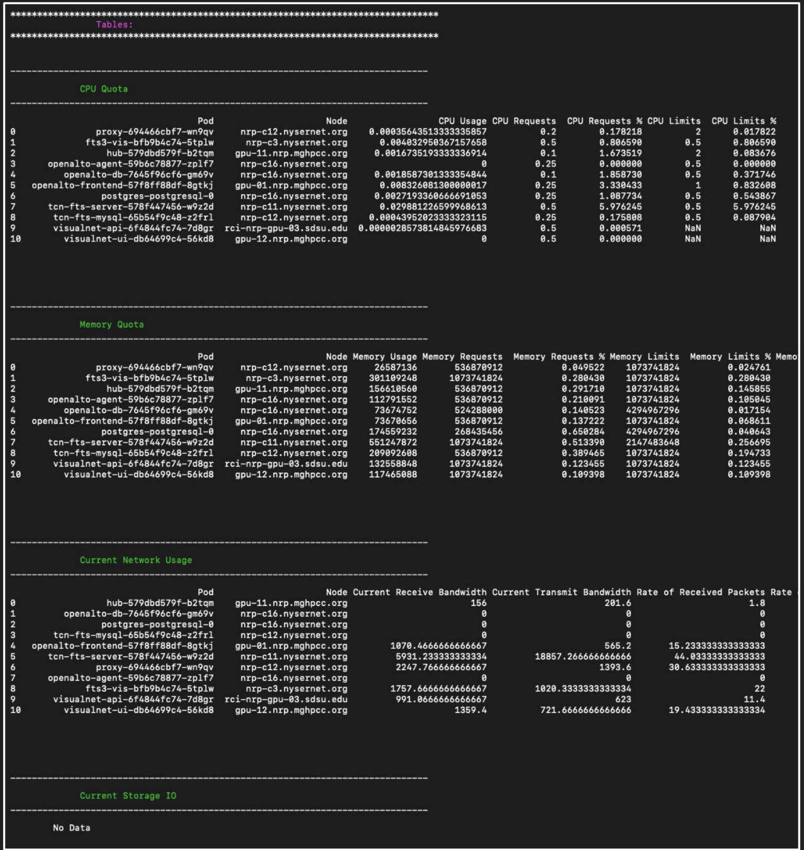
- Memory Usage
- Memory Requests
- Memory Requests %
- Memory Limits
- Memory Limits %
- Memory Usage (RSS)
- Memory Usage (Cache)

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

Current Storage IO

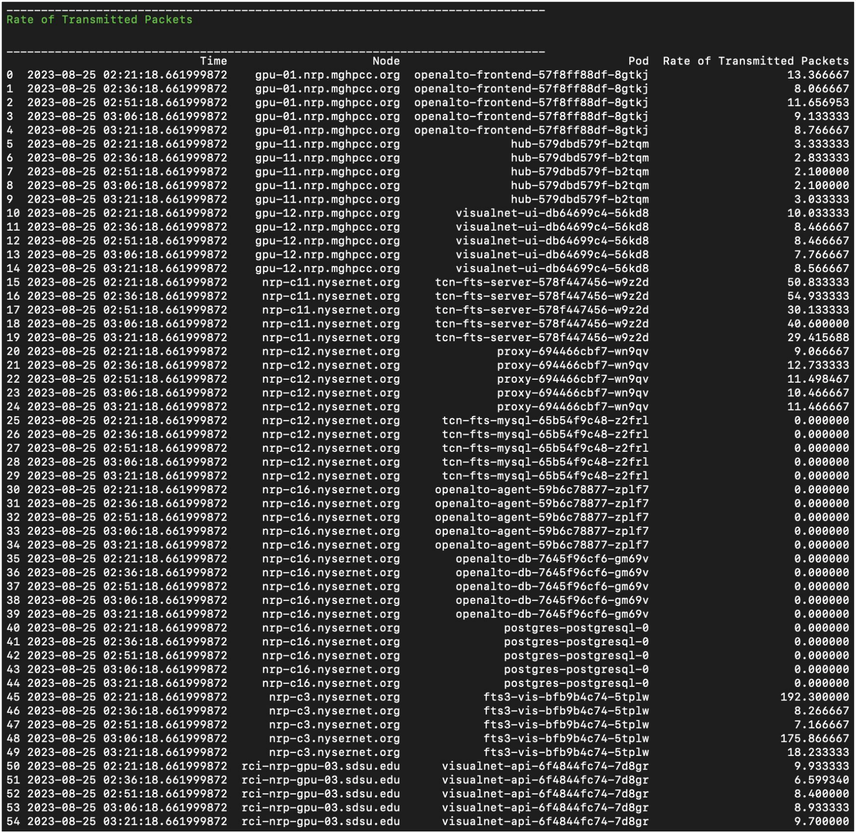
- IOPS(Reads)
- IOPS(Writes)
- IOPS(Reads + Writes)
- Throughput(Read)
- Throughput(Write)
- Throughput(Read + Write)



```
************************************************************************************
                Tables:
************************************************************************************


--------------------------------------------------------------------------------

             CPU Quota

--------------------------------------------------------------------------------

                                     Pod                        Node                   CPU Usage CPU Requests  CPU Requests % CPU Limits  CPU Limits %
0             proxy-694466cbf7-wn9qv       nrp-c12.nysernet.org    0.0003564351333335857          0.2        0.178218          2      0.017822
1           fts3-vis-bfb9b4c74-5tplw        nrp-c3.nysernet.org     0.004032950367157658          0.5        0.806590        0.5      0.806590
2              hub-579dbd579f-b2tqm      gpu-11.nrp.mghpcc.org     0.0016735193333336914          0.1        1.673519          2      0.083676
3     openalto-agent-59b6c78877-zplf7       nrp-c16.nysernet.org                        0         0.25        0.000000        0.5      0.000000
4        openalto-db-7645f96cf6-gm69v       nrp-c16.nysernet.org     0.0018587301333354844          0.1        1.858730        0.5      0.371746
5  openalto-frontend-57f8ff88df-8gtkj      gpu-01.nrp.mghpcc.org      0.00832608130000017         0.25        3.330433          1      0.832608
6             postgres-postgresql-0       nrp-c16.nysernet.org     0.0027193360666691053         0.25        1.087734        0.5      0.543867
7     tcn-fts-server-578f447456-w9z2d       nrp-c11.nysernet.org      0.02988122659968613          0.5        5.976245        0.5      5.976245
8      tcn-fts-mysql-65b54f9c48-z2frl       nrp-c12.nysernet.org     0.0004395202333323115         0.25        0.175808        0.5      0.087904
9     visualnet-api-6f4844fc74-7d8gr  rci-nrp-gpu-03.sdsu.edu  0.0000028573814845976683          0.5        0.000571        NaN           NaN
10      visualnet-ui-db64699c4-56kd8      gpu-12.nrp.mghpcc.org                        0          0.5        0.000000        NaN           NaN



--------------------------------------------------------------------------------

             Memory Quota

--------------------------------------------------------------------------------

                                     Pod                        Node Memory Usage Memory Requests  Memory Requests % Memory Limits  Memory Limits % Memo
0             proxy-694466cbf7-wn9qv       nrp-c12.nysernet.org     26587136       536870912           0.049522    1073741824         0.024761
1           fts3-vis-bfb9b4c74-5tplw        nrp-c3.nysernet.org    301109248      1073741824           0.280430    1073741824         0.280430
2              hub-579dbd579f-b2tqm      gpu-11.nrp.mghpcc.org    156610560       536870912           0.291710    1073741824         0.145855
3     openalto-agent-59b6c78877-zplf7       nrp-c16.nysernet.org    112791552       536870912           0.210091    1073741824         0.105045
4        openalto-db-7645f96cf6-gm69v       nrp-c16.nysernet.org     73674752       524288000           0.140523    4294967296         0.017154
5  openalto-frontend-57f8ff88df-8gtkj      gpu-01.nrp.mghpcc.org     73670656       536870912           0.137222    1073741824         0.068611
6             postgres-postgresql-0       nrp-c16.nysernet.org    174559232       268435456           0.650284    4294967296         0.040643
7     tcn-fts-server-578f447456-w9z2d       nrp-c11.nysernet.org    551247872      1073741824           0.513390    2147483648         0.256695
8      tcn-fts-mysql-65b54f9c48-z2frl       nrp-c12.nysernet.org    209092608       536870912           0.389465    1073741824         0.194733
9     visualnet-api-6f4844fc74-7d8gr  rci-nrp-gpu-03.sdsu.edu    132558848      1073741824           0.123455    1073741824         0.123455
10      visualnet-ui-db64699c4-56kd8      gpu-12.nrp.mghpcc.org    117465088      1073741824           0.109398    1073741824         0.109398



--------------------------------------------------------------------------------

             Current Network Usage

--------------------------------------------------------------------------------

                                     Pod                        Node Current Receive Bandwidth Current Transmit Bandwidth Rate of Received Packets Rate
0              hub-579dbd579f-b2tqm      gpu-11.nrp.mghpcc.org                       156                      201.6                      1.8
1        openalto-db-7645f96cf6-gm69v       nrp-c16.nysernet.org                         0                          0                        0
2             postgres-postgresql-0       nrp-c16.nysernet.org                         0                          0                        0
3      tcn-fts-mysql-65b54f9c48-z2frl       nrp-c12.nysernet.org                         0                          0                        0
4  openalto-frontend-57f8ff88df-8gtkj      gpu-01.nrp.mghpcc.org        1070.4666666666667                      565.2        15.233333333333333
5     tcn-fts-server-578f447456-w9z2d       nrp-c11.nysernet.org         5931.233333333334         18857.266666666666        44.03333333333333
6             proxy-694466cbf7-wn9qv       nrp-c12.nysernet.org        2247.7666666666667                     1393.6        30.633333333333333
7     openalto-agent-59b6c78877-zplf7       nrp-c16.nysernet.org                         0                          0                        0
8           fts3-vis-bfb9b4c74-5tplw        nrp-c3.nysernet.org        1757.6666666666667         1020.3333333333334                       22
9     visualnet-api-6f4844fc74-7d8gr  rci-nrp-gpu-03.sdsu.edu         991.0666666666667                        623                     11.4
10      visualnet-ui-db64699c4-56kd8      gpu-12.nrp.mghpcc.org                    1359.4          721.6666666666666        19.433333333333334



--------------------------------------------------------------------------------

             Current Storage IO

--------------------------------------------------------------------------------

       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 02:21:18.661999872    gpu-01.nrp.mghpcc.org  openalto-frontend-57f8ff88df-8gtkj                    13.366667
1  2023-08-25 02:36:18.661999872    gpu-01.nrp.mghpcc.org  openalto-frontend-57f8ff88df-8gtkj                     8.066667
2  2023-08-25 02:51:18.661999872    gpu-01.nrp.mghpcc.org  openalto-frontend-57f8ff88df-8gtkj                    11.656953
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-11.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                    10.033333
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-wn9qv                     9.066667
21 2023-08-25 02:36:18.661999872     nrp-c12.nysernet.org              proxy-694466cbf7-wn9qv                    12.733333
22 2023-08-25 02:51:18.661999872     nrp-c12.nysernet.org              proxy-694466cbf7-wn9qv                    11.498467
23 2023-08-25 03:06:18.661999872     nrp-c12.nysernet.org              proxy-694466cbf7-wn9qv                    10.466667
24 2023-08-25 03:21:18.661999872     nrp-c12.nysernet.org              proxy-694466cbf7-wn9qv                    11.466667
25 2023-08-25 02:21:18.661999872     nrp-c12.nysernet.org      tcn-fts-mysql-65b54f9c48-z2frl                     0.000000
26 2023-08-25 02:36:18.661999872     nrp-c12.nysernet.org      tcn-fts-mysql-65b54f9c48-z2frl                     0.000000
27 2023-08-25 02:51:18.661999872     nrp-c12.nysernet.org      tcn-fts-mysql-65b54f9c48-z2frl                     0.000000
28 2023-08-25 03:06:18.661999872     nrp-c12.nysernet.org      tcn-fts-mysql-65b54f9c48-z2frl                     0.000000
29 2023-08-25 03:21:18.661999872     nrp-c12.nysernet.org      tcn-fts-mysql-65b54f9c48-z2frl                     0.000000
30 2023-08-25 02:21:18.661999872     nrp-c16.nysernet.org     openalto-agent-59b6c78877-zplf7                     0.000000
31 2023-08-25 02:36:18.661999872     nrp-c16.nysernet.org     openalto-agent-59b6c78877-zplf7                     0.000000
32 2023-08-25 02:51:18.661999872     nrp-c16.nysernet.org     openalto-agent-59b6c78877-zplf7                     0.000000
33 2023-08-25 03:06:18.661999872     nrp-c16.nysernet.org     openalto-agent-59b6c78877-zplf7                     0.000000
34 2023-08-25 03:21:18.661999872     nrp-c16.nysernet.org     openalto-agent-59b6c78877-zplf7                     0.000000
35 2023-08-25 02:21:18.661999872     nrp-c16.nysernet.org        openalto-db-7645f96cf6-gm69v                     0.000000
36 2023-08-25 02:36:18.661999872     nrp-c16.nysernet.org        openalto-db-7645f96cf6-gm69v                     0.000000
37 2023-08-25 02:51:18.661999872     nrp-c16.nysernet.org        openalto-db-7645f96cf6-gm69v                     0.000000
38 2023-08-25 03:06:18.661999872     nrp-c16.nysernet.org        openalto-db-7645f96cf6-gm69v                     0.000000
39 2023-08-25 03:21:18.661999872     nrp-c16.nysernet.org        openalto-db-7645f96cf6-gm69v                     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