Implementing Multi-User Cybersecurity Graph Analytics with xGT & ENSIGN on SDFX (8S/6TB)

1. HPE Lab Test environment Details

	Configuration
No. Of Chassis	2 (1 nPar)
Processor Architecture	Cascade Lake. Xeon-Platinum **8280**, (2.7 GHz/28-core/205 W)
Total Sockets	4 Socket / Chassis , 4 * 2 = 8 Socket
Total Cores	28 Core / Socket , 8 * 28 = 224 Cores
Total Memory	24 * 128 GB per chassis , 24 * 128 * 2 ~ 6 TB
Total Storage	2 NVMe per socket , 1.6 TB * 2 * 8 ~ 25 TB

CPU configuration details

```
[root@mdchawkeye ensign_dask]# lscpu
Architecture:          x86_64
CPU op-mode(s):        32-bit, 64-bit
Byte Order:            Little Endian
CPU(s):                448
On-line CPU(s) list:   0-447
Thread(s) per core:    2
Core(s) per socket:    28
Socket(s):             8
NUMA node(s):          8
Vendor ID:             GenuineIntel
CPU family:            6
Model:                 85
Model name:            Intel(R) Xeon(R) Platinum 8280 CPU @ 2.70GHz
Stepping:              7
CPU MHz:               2791.955
CPU max MHz:           4000.0000
CPU min MHz:           1000.0000
BogoMIPS:              5400.00
Virtualization:        VT-x
L1d cache:             32K
L1i cache:             32K
L2 cache:              1024K
L3 cache:              39424K
NUMA node0 CPU(s):     0-27,224-251
NUMA node1 CPU(s):     28-55,252-279
NUMA node2 CPU(s):     56-83,280-307
NUMA node3 CPU(s):     84-111,308-335
NUMA node4 CPU(s):     112-139,336-363
NUMA node5 CPU(s):     140-167,364-391
NUMA node6 CPU(s):     168-195,392-419
NUMA node7 CPU(s):     196-223,420-447
```

Memory Configuration

```
[root@mdchawkeye ensign_dask]# cat /proc/meminfo
MemTotal:       6239688620 kB
MemFree:        3973453232 kB
MemAvailable:   6138640760 kB
Buffers:           24608 kB
Cached:         2134108780 kB
SwapCached:            0 kB
Active:         639889056 kB
Inactive:       1548965328 kB
Active(anon):   54776196 kB
Inactive(anon):   252724 kB
Active(file):   585112860 kB
Inactive(file): 1548712604 kB
Unevictable:       10900 kB
Mlocked:           10944 kB
SwapTotal:       4194300 kB
SwapFree:        4194300 kB
Dirty:                16 kB
Writeback:             0 kB
AnonPages:      54733780 kB
Mapped:           458528 kB
Shmem:            305572 kB
Slab:           35663200 kB
SReclaimable:   33988928 kB
SUnreclaim:      1674272 kB
KernelStack:      103216 kB
PageTables:       168100 kB
NFS_Unstable:          0 kB
Bounce:                0 kB
WritebackTmp:          0 kB
CommitLimit:    3124038608 kB
Committed_AS:   83261464 kB
VmallocTotal:   34359738367 kB
VmallocUsed:    18070408 kB
VmallocChunk:   28704574460 kB
HardwareCorrupted:     0 kB
AnonHugePages:     10240 kB
CmaTotal:              0 kB
CmaFree:               0 kB
HugePages_Total:       0
HugePages_Free:        0
HugePages_Rsvd:        0
HugePages_Surp:        0
Hugepagesize:       2048 kB
DirectMap4k:      923672 kB
DirectMap2M:    36268032 kB
DirectMap1G:    6302990336 kB
```

Storage Disk Configuration

```
[root@mdchawkeye ensign_dask]# lsblk
NAME             MAJ:MIN RM  SIZE RO TYPE  MOUNTPOINT
sda                8:0    0  1.8T  0 disk
└─md126            9:126  0  1.7T  0 raid1
sdb                8:16   0  1.8T  0 disk
└─md126            9:126  0  1.7T  0 raid1
sdc                8:32   0  1.8T  0 disk
├─sdc1             8:33   0  200M  0 part
├─sdc2             8:34   0    1G  0 part
└─sdc3             8:35   0  1.1T  0 part
  ├─rhel-root    253:3    0  100G  0 lvm
  ├─rhel-home    253:4    0    1T  0 lvm
  └─rhel-swap    253:5    0    4G  0 lvm
sdd                8:48   0  1.8T  0 disk
├─sdd1             8:49   0  200M  0 part  /boot/efi
├─sdd2             8:50   0    1G  0 part  /boot
└─sdd3             8:51   0  704G  0 part
  ├─rhel00-root  253:0    0  200G  0 lvm   /
  ├─rhel00-swap  253:1    0    4G  0 lvm   [SWAP]
  └─rhel00-home  253:2    0  500G  0 lvm   /home
sr0               11:0    1 1024M  0 rom
nvme10n1         259:9    0  1.5T  0 disk  /nvme_datav1
nvme11n1         259:12   0  1.5T  0 disk  /nvme_data12
nvme12n1         259:5    0  1.5T  0 disk  /nvme_data13
nvme13n1         259:7    0  1.5T  0 disk  /nvme_data14
nvme14n1         259:8    0  1.5T  0 disk  /nvme_data15
loop0              7:0    0  4.2G  0 loop  /var/www/html/RHEL76
loop1              7:1    0  7.7M  0 loop  /var/www/html/HPEFS
nvme0n1          259:0    0  1.5T  0 disk  /nvme_data1
nvme1n1          259:13   0  1.5T  0 disk  /nvme_data2
nvme2n1          259:10   0  1.5T  0 disk  /nvme_data3
nvme3n1          259:14   0  1.5T  0 disk  /nvme_data4
nvme4n1          259:3    0  1.5T  0 disk  /nvme_data5
nvme5n1          259:4    0  1.5T  0 disk  /nvme_data6
nvme6n1          259:1    0  1.5T  0 disk  /nvme_data7
nvme7n1          259:2    0  1.5T  0 disk  /nvme_data8
nvme8n1          259:11   0  1.5T  0 disk  /nvme_data9
nvme9n1          259:6    0  1.5T  0 disk  /nvme_datav0
```

Environment planning for Co-Execution Test Environment

a. **Trovares**
   i. Dataset : LANL Dataset
   ii. Total Raw Data Volume : 1V , 2V, Netflow

LANL (xGT)	Data Volume(GB)
LANL 1V	116
LANL 2V	356
LANL Netflow	946
Total	**1418**

b. **Reservoir Lab**
   i. Dataset : CIC Dataset
   ii. Total RAW Data Volume :

CICDDoS2019 (ENSIGN)	Data Volume(GB)
csv	116
pcap	356
Total	**472**

c. **Resource Planning**
   i. **Master/Backup Copy of Dataset :** 3TB : 2NVMe Drives
   ii. **Working Dataset :**
      1. Trovares (xGT) : 6 * 1.5 TB = 9 TB (/nvme_data1 - /nvme_data6)
      2. RL (ENSIGN) : 6 * 1.5 TB = 9 TB (/nvme_data7 - /nvme_data12)
   iii. **Memory Requirement for xGT & ENSIGN (Single User):**
      1. xGT : 3TB
      2. ENSIGN : 3 TB
   iv. **Compute Allocation for xGT & ENSIGN (Single User):**
      1. xGT : 112 Cores (CPUID 0-111)
      2. ENSIGN : 112 Cores (CPUID 112-223)

d. **Key Configuration Files**
   i. **xGT :**
      1. **Install Location : "/opt/xgtd"**
      2. **Environment Variables :**

```
#setup xgt path
export PATH=$PATH:/opt/xgtd/bin
export PATH=$PATH:/opt/xgtd/lib
export PATH=$PATH:/opt/xgtd/util
export LSHOST="NO-NET"
export LSERVRC="/opt/xgtd/bin/lservrc"
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/usr/local/lib:/opt/xgtd/lib
```

      3. **Conf file : "/opt/xgtd/xgtd.conf"**

```
[root@mdchawkeye xgtd]# pwd
/opt/xgtd
[root@mdchawkeye xgtd]# cat xgtd.conf
{
  "system": {
    "worker_threads": 224,
    "io_threads": 224,
    "port": 4367,
    "max_memory": 3298534883328,
    "pin_threads": true
  },

  "system.locale": "en_US.UTF-8"
}
```

   ii. **ENSIGN :**
      1. **Install Location : "/root/ashish/reservoir_lab/ENSIGN-42"**
      2. **Environment Variables :**

```
#add following environment variables for ENSIGN from RESERVOIR LAB
export ENSIGN_BASE=/root/ashish/reservoir lab/ENSIGN-42
export ANACONDA_HOME=/root/anaconda3/
export PATH=$ENSIGN_BASE/ENSIGN_4.2/bin:$PATH
export PATH=$ENSIGN_BASE/ENSIGN_4.2/Ensign-Py3/bin:$PATH
export PATH=$ANACONDA_HOME/bin:$PATH
export PYTHONPATH=$ENSIGN_BASE/ENSIGN_4.2/Ensign-Py3
export LD_LIBARY_PATH=$LD_LIBRARY_PATH:$ENSIGN_BASE/ENSIGN_4.2/Ensign-CAPI/lib
```

      3. **Workflow Conf file : "/root/ashish/reservoir_lab/ENSIGN-42/workflow_cfg.yml"**


```
[root@mdchawkeye ENSIGN-42]# cd $ENSIGN_BASE
[root@mdchawkeye ENSIGN-42]# workflow.py /root/ashish/reservoir_lab/ENSIGN-42/workflow_cfg.yml | tee console_output_1.txt
```

End to end tensor decomposition workflow

tensor decomposition workflow config

2. Install and Configure xGT for Multi-User Graph Analytics for Known Threat Detection

a. Executing xGT Server with limited memory/cpu threads (**single-user**)

```
[root@mdchawkeye /]# cd /opt/
[root@mdchawkeye opt]# cd xgtd
[root@mdchawkeye xgtd]# pwd
/opt/xgtd
[root@mdchawkeye xgtd]# cat xgtd.conf
{
  "system": {
    "worker_threads": 224,
    "io_threads": 224,
    "port": 4367,
    "max_memory": 3298534883328,
    "pin_threads": true
  },

  "system.locale": "en_US.UTF-8"
}
[root@mdchawkeye xgtd]# cd /
[root@mdchawkeye /]# taskset -c 0-111,224-335 xgtd -c /opt/xgtd/xgtd.conf
[n0 w3097950528 0.000] Starting xGT version: 1.3.0
[n0 w3097950528 0.032] Available memory for xGT data: 3298534883328
[n0 w3097950528 0.032] TBB worker threads: 224
[n0 w3097950528 0.032] Pinning threads: true
```

b. Executing xGT Server with limited memory/cpu threads (multi-user)

User-1 Server

```
[root@mdchawkeye /]# cd /opt/
[root@mdchawkeye opt]# cd xgtd
[root@mdchawkeye xgtd]# pwd
/opt/xgtd
[root@mdchawkeye xgtd]# cat xgtd.conf
{
  "system": {
    "worker_threads": 224,
    "io_threads": 224,
    "port": 4367,
    "max_memory": 3298534883328,
    "pin_threads": true
  },

  "system.locale": "en_US.UTF-8"
}
[root@mdchawkeye xgtd]# cd /
[root@mdchawkeye /]# taskset -c 0-111,224-335 xgtd -c /opt/xgtd/xgtd.conf
[n0 w3097950528 0.000] Starting xGT version: 1.3.0
[n0 w3097950528 0.032] Available memory for xGT data: 3298534883328
[n0 w3097950528 0.032] TBB worker threads: 224
[n0 w3097950528 0.032] Pinning threads: true
```

User-1

User-1 Client

```
(Python_3_7) [root@mdchawkeye ~]# python
Python 3.7.4 (default, Nov 12 2019, 13:45:07)
[GCC 4.8.5 20150623 (Red Hat 4.8.5-36)] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import xgt
>>> from platform import python_version
>>> print (python_version())
3.7.4
>>> conn = xgt.Connection()
>>> conn_server_version
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'conn_server_version' is not defined
>>> conn.server_version
'1.3.0'
>>>
(Python_3_7) [root@mdchawkeye ~]#
```

User-1 Client

User-2 Server

```
[root@mdchawkeye /]# cat /opt/xgtd/xgtd_v2.conf
{
  "system": {
    "worker_threads": 224,
    "io_threads": 224,
    "port": 4368,
    "max_memory": 3298534883328,
    "pin_threads": true
  },

  "system.locale": "en_US.UTF-8"
}
[root@mdchawkeye /]# taskset -c 0-111,224-335 xgtd -c /opt/xgtd/xgtd_v2.conf
[n0 w10410304 0.000] Starting xGT version: 1.3.0
[n0 w10410304 0.033] Available memory for xGT data: 3298534883328
[n0 w10410304 0.033] TBB worker threads: 224
[n0 w10410304 0.033] Pinning threads: true
```

User-2

User-2 Client

```
(Python_3_7) [root@mdchawkeye ~]# python
Python 3.7.4 (default, Nov 12 2019, 13:45:07)
[GCC 4.8.5 20150623 (Red Hat 4.8.5-36)] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import xgt
>>> from platform import python_version
>>> print (python_version())
3.7.4
>>> conn = xgt.Connection(host='127.0.0.1', port=4368, flags=None)
>>> conn.server_version
'1.3.0'
>>>
```

User-2 Client

3. Install and Configure ENSIGN for Multi-User Unsupervised Machine Learning for Un-Known Thread Detection

a) Executing ENSIGN with memory/cpu threads limitations (single user)

num_threads = 112 (no. of physical core allocated)
mem_limit_gb = 3000 (memory allocated)
[root@mdchawkeye ENSIGN-42]# taskset -c 112-223,336-447 workflow.py /root/ashish/reservoir_lab/ENSIGN-42/workflow_cfg.yml | tee console_output_run_2

```
[root@mdchawkeye ENSIGN-42]# taskset -c 112-223,336-447 workflow.py /root/ashish/reservoir_lab/ENSIGN-42/workflow_cfg.yml | tee console_output_run_2
Converting CSV to tensor ...
Setting up Dask client ...
  Reading in /nvme_data14/CICDataset/CICDDoS2019/Dataset/csv/01-12-Backup/DrDoS_NetBIOS.csv ...
  Reading in /nvme_data14/CICDataset/CICDDoS2019/Dataset/csv/01-12-Backup/DrDoS_MSSQL.csv ...
  Reading in /nvme_data14/CICDataset/CICDDoS2019/Dataset/csv/01-12-Backup/DrDoS_LDAP.csv ...
  Reading in /nvme_data14/CICDataset/CICDDoS2019/Dataset/csv/01-12-Backup/DrDoS_DNS.csv ...
  Reading in /nvme_data15/CICDataset/DrDoS_NTP.csv ...
  Reading in /nvme_data15/CICDataset/DrDoS_SNMP.csv ...
  Reading in /nvme_data15/CICDataset/DrDoS_SSDP.csv ...
  Reading in /nvme_data15/CICDataset/DrDoS_UDP.csv ...
  Reading in /nvme_data15/CICDataset/Syn.csv ...
  Reading in /nvme_data15/CICDataset/TFTP.csv ...
  Reading in /nvme_data15/CICDataset/UDPLag.csv ...
Validating ...
Filtering ...
Casting column types ...
Binning ...
  Binning mode 0 (second)
  Binning mode 1 (none)
  Binning mode 2 (none)
  Binning mode 3 (none)
  Binning mode 4 (log10)
Fusing ...
Calculating tensor entries ...
Building sparse tensor ...
  csv2tensor took 217.18601036071777 seconds.
Closing dask ...
  Writing tensor files to disk ...
  Writing ' Timestamp' to /root/ashish/reservoir_lab/ENSIGN-42/output/map_mode_0.txt (size=19031)
  Writing ' Source IP' to /root/ashish/reservoir_lab/ENSIGN-42/output/map_mode_1.txt (size=550)
  Writing ' Destination IP' to /root/ashish/reservoir_lab/ENSIGN-42/output/map_mode_2.txt (size=598)
  Writing ' Destination Port' to /root/ashish/reservoir_lab/ENSIGN-42/output/map_mode_3.txt (size=65535)
  Writing 'Flow Bytes/s' to /root/ashish/reservoir_lab/ENSIGN-42/output/map_mode_4.txt (size=10)
  Writing 43730652 nonzeroes to /root/ashish/reservoir_lab/ENSIGN-42/output/tensor_data.txt
    tensor dump took: 203.35911107063293 seconds
Decomposing tensor w/ CP APR ...
  decomposition took 102.38855528831482 seconds.
  Writing decomposition files to disk ...
    decomp dump took: 1.9439828395843506 seconds
Starting visualization ...

  visualization took 63.75034785270691 seconds.
Generating textual report ...
  report generation took 7.125171899795532 seconds.
Starting detectors ...
Running portscan detector ...
Running network mapping detector ...
Running beacon detector ...
  detection took 1.412672758102417 seconds.
 ** Reached maximum iterations **
 *** Final Fit = 0.0659437066851436 ***
 *** Norm Scaling = 0.3569013742486944 ***
 *** Cosine Similarity = 0.3571258760156621 ***
 *** Final loglikelihood = -169833250.8345844 ***
 *** Final KKT violation = 532.6565289090041 ***
[root@mdchawkeye ENSIGN-42]#
```

b) Executing ENSIGN with memory/cpu threads limitations (multi user)

1. Create (per user) workflow configuration file
2. Create (per user) input file path and output directory
3. Execute (per user) workflow as indicated in step above

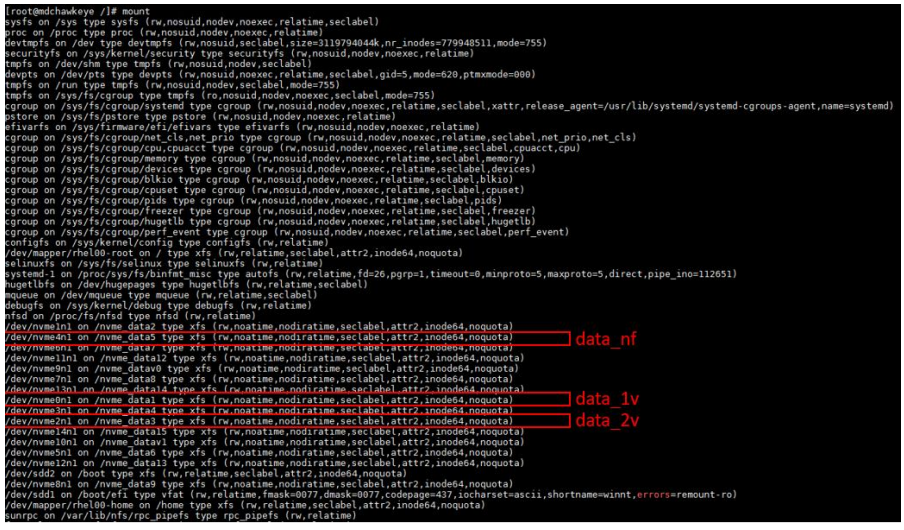
4. Setup Jupyter for Multi-User Execution

```
[root@mdchawkeye ENSIGN-42]# source env.sh
[root@mdchawkeye ENSIGN-42]# cat env.sh
#!/usr/bin/env bash

# SET THE FOLLOWING TWO ENVIRONMENT VARIABLES
export ENSIGN_BASE=/root/ashish/reservoir_lab/ENSIGN-42
export ANACONDA_HOME=/root/anaconda3/

export PATH=$ENSIGN_BASE/ENSIGN_4.2/bin:$ENSIGN_BASE/ENSIGN_4.2/Ensign-Py3/bin:$ANACONDA_HOME/bin/:$PATH
export PYTHONPATH=$ENSIGN_BASE/ENSIGN_4.2/Ensign-Py3
export LD_LIBRARY_PATH=$ENSIGN_BASE/ENSIGN_4.2/Ensign-CAPI/lib

[root@mdchawkeye ENSIGN-42]# source env.sh

[root@mdchawkeye ENSIGN-42]# jupyter kernelspec list
Available kernels:
  python_3_6    /root/.local/share/jupyter/kernels/python_3_6
  python_3_7    /root/.local/share/jupyter/kernels/python_3_7
  python3       /root/anaconda3/share/jupyter/kernels/python3
  python2       /usr/local/share/jupyter/kernels/python2
(reverse-i-search)`jup': ^Cpyter kernelspec list
[root@mdchawkeye ENSIGN-42]# jupyter notebook --ip 172.30.224.99 --allow-root
[W 15:09:38.054 NotebookApp] Error loading server extension jupyter_server_proxy
    Traceback (most recent call last):
      File "/root/anaconda3/lib/python3.7/site-packages/notebook/notebookapp.py", line 1615, in init_server_extensions
        mod = importlib.import_module(modulename)
      File "/root/anaconda3/lib/python3.7/importlib/__init__.py", line 127, in import_module
        return _bootstrap._gcd_import(name[level:], package, level)
      File "<frozen importlib._bootstrap>", line 1006, in _gcd_import
      File "<frozen importlib._bootstrap>", line 983, in _find_and_load
      File "<frozen importlib._bootstrap>", line 965, in _find_and_load_unlocked
    ModuleNotFoundError: No module named 'jupyter_server_proxy'
[I 15:09:38.075 NotebookApp] JupyterLab extension loaded from /root/anaconda3/lib/python3.7/site-packages/jupyterlab
[I 15:09:38.075 NotebookApp] JupyterLab application directory is /root/anaconda3/share/jupyter/lab
[I 15:09:38.076 NotebookApp] Serving notebooks from local directory: /root/ashish/reservoir_lab/ENSIGN-42
[I 15:09:38.076 NotebookApp] The Jupyter Notebook is running at:
[I 15:09:38.077 NotebookApp] http://172.30.224.99:8888/?token=e9a1b9333174ef0c0c9b385d41a983d4bb0f18b2e89eb5c0
[I 15:09:38.077 NotebookApp]  or http://127.0.0.1:8888/?token=e9a1b9333174ef0c0c9b385d41a983d4bb0f18b2e89eb5c0
[I 15:09:38.077 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 15:09:38.134 NotebookApp]

    To access the notebook, open this file in a browser:
        file:///root/.local/share/jupyter/runtime/nbserver-218983-open.html
    Or copy and paste one of these URLs:
        http://172.30.224.99:8888/?token=e9a1b9333174ef0c0c9b385d41a983d4bb0f18b2e89eb5c0
     or http://127.0.0.1:8888/?token=e9a1b9333174ef0c0c9b385d41a983d4bb0f18b2e89eb5c0
[W 15:09:42.569 NotebookApp] 404 GET /server-proxy/servers-info (172.30.224.99) 11.73ms referer=http://172.30.224.99:8888/tree
```

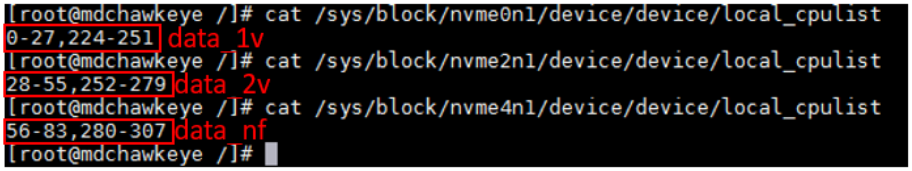
5. Parallel Test Execution

Executed xGT with limited core and memory

1. Identify NVMe drives for storing input data & distribute data across multiple NVMe drives for parallel loading



2. Identify mapping CPUs to be allocated for xGT execution

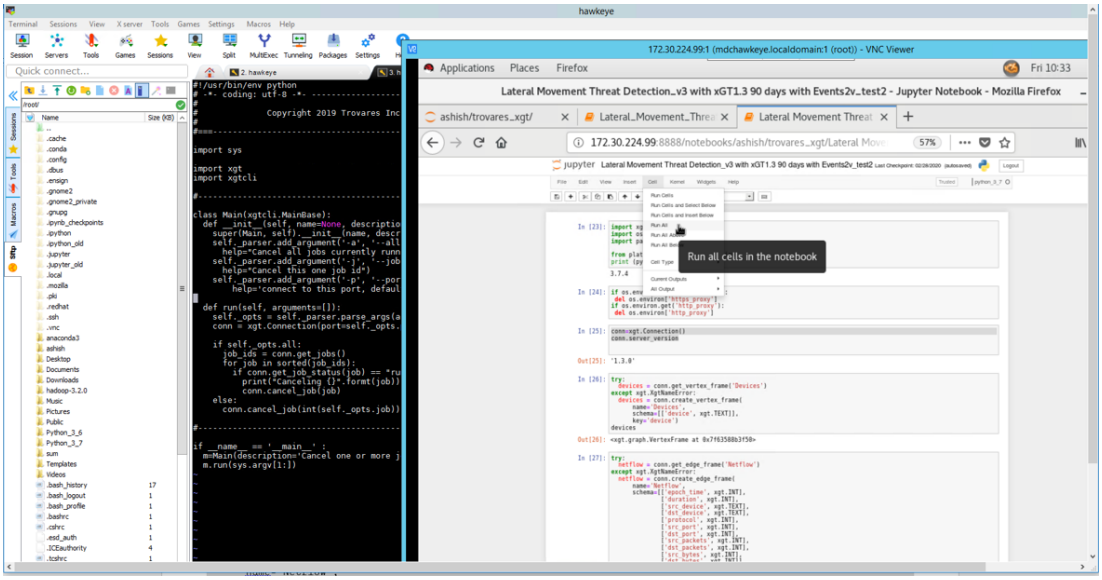


3. Start xGT server with pinned cores

```
[root@mdchawkeye /]# taskset -c 0-111,224-335 xgtd -c /opt/xgtd/xgtd.conf
[n0 w3989817664 0.000] Starting xGT version: 1.3.0
[n0 w3989817664 0.029] Available memory for xGT data: 3298534883328
[n0 w3989817664 0.029] TBB worker threads: 224
[n0 w3989817664 0.029] Pinning threads: true
[n0 w3682731776 38.600] User message:
```

4. Execute LANL detection Queries

a) Launch Jupyter Notebook Environment
b) Load Lateral Movement Cyber-Threat detection notebook
c) Execute the notebook End-2-End with performance data collection enabled



Execute ENSIGN with limited core and memory

1. Identify NVMe drives for storing input data & distribute data across multiple NVMe drives for parallel loading

```
[root@mdchawkeye /]# mount
sysfs on /sys type sysfs (rw,nosuid,nodev,noexec,relatime,seclabel)
proc on /proc type proc (rw,nosuid,nodev,noexec,relatime)
devtmpfs on /dev type devtmpfs (rw,nosuid,seclabel,size=3119794044k,nr_inodes=779948511,mode=755)
securityfs on /sys/kernel/security type securityfs (rw,nosuid,nodev,noexec,relatime)
tmpfs on /dev/shm type tmpfs (rw,nosuid,nodev,seclabel)
devpts on /dev/pts type devpts (rw,nosuid,noexec,relatime,seclabel,gid=5,mode=620,ptmxmode=000)
tmpfs on /run type tmpfs (rw,nosuid,nodev,seclabel,mode=755)
tmpfs on /sys/fs/cgroup type tmpfs (ro,nosuid,nodev,noexec,seclabel,mode=755)
cgroup on /sys/fs/cgroup/systemd type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,xattr,release_agent=/usr/lib/systemd/systemd-cgroups-agent,name=systemd)
pstore on /sys/fs/pstore type pstore (rw,nosuid,nodev,noexec,relatime)
efivarfs on /sys/firmware/efi/efivars type efivarfs (rw,nosuid,nodev,noexec,relatime)
cgroup on /sys/fs/cgroup/net_cls,net_prio type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,net_prio,net_cls)
cgroup on /sys/fs/cgroup/cpu,cpuacct type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,cpuacct,cpu)
cgroup on /sys/fs/cgroup/memory type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,memory)
cgroup on /sys/fs/cgroup/devices type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,devices)
cgroup on /sys/fs/cgroup/blkio type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,blkio)
cgroup on /sys/fs/cgroup/cpuset type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,cpuset)
cgroup on /sys/fs/cgroup/pids type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,pids)
cgroup on /sys/fs/cgroup/freezer type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,freezer)
cgroup on /sys/fs/cgroup/hugetlb type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,hugetlb)
cgroup on /sys/fs/cgroup/perf_event type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,perf_event)
configfs on /sys/kernel/config type configfs (rw,relatime)
/dev/mapper/rhel00-root on / type xfs (rw,relatime,seclabel,attr2,inode64,noquota)
selinuxfs on /sys/fs/selinux type selinuxfs (rw,relatime)
systemd-1 on /proc/sys/fs/binfmt_misc type autofs (rw,relatime,fd=26,pgrp=1,timeout=0,minproto=5,maxproto=5,direct,pipe_ino=112651)
hugetlbfs on /dev/hugepages type hugetlbfs (rw,relatime,seclabel)
mqueue on /dev/mqueue type mqueue (rw,relatime,seclabel)
debugfs on /sys/kernel/debug type debugfs (rw,relatime)
nfsd on /proc/fs/nfsd type nfsd (rw,relatime)
/dev/nvme1n1 on /nvme_data2 type xfs (rw,noatime,nodiratime,seclabel,attr2,inode64,noquota)
/dev/nvme4n1 on /nvme_data5 type xfs (rw,noatime,nodiratime,seclabel,attr2,inode64,noquota)
/dev/nvme6n1 on /nvme_data7 type xfs (rw,noatime,nodiratime,seclabel,attr2,inode64,noquota)
/dev/nvme11n1 on /nvme_data12 type xfs (rw,noatime,nodiratime,seclabel,attr2,inode64,noquota)
/dev/nvme9n1 on /nvme_data10 type xfs (rw,noatime,nodiratime,seclabel,attr2,inode64,noquota)
/dev/nvme7n1 on /nvme_data8 type xfs (rw,noatime,nodiratime,seclabel,attr2,inode64,noquota)
/dev/nvme13n1 on /nvme_data14 type xfs (rw,noatime,nodiratime,seclabel,attr2,inode64,noquota)
/dev/nvme0n1 on /nvme_data1 type xfs (rw,noatime,nodiratime,seclabel,attr2,inode64,noquota)
/dev/nvme3n1 on /nvme_data4 type xfs (rw,noatime,nodiratime,seclabel,attr2,inode64,noquota)
/dev/nvme2n1 on /nvme_data3 type xfs (rw,noatime,nodiratime,seclabel,attr2,inode64,noquota)
/dev/nvme14n1 on /nvme_data15 type xfs (rw,noatime,nodiratime,seclabel,attr2,inode64,noquota)
/dev/nvme10n1 on /nvme_data11 type xfs (rw,noatime,nodiratime,seclabel,attr2,inode64,noquota)
/dev/nvme5n1 on /nvme_data6 type xfs (rw,noatime,nodiratime,seclabel,attr2,inode64,noquota)
/dev/nvme12n1 on /nvme_data13 type xfs (rw,noatime,nodiratime,seclabel,attr2,inode64,noquota)
/dev/sdd2 on /boot type xfs (rw,relatime,seclabel,attr2,inode64,noquota)
/dev/nvme8n1 on /nvme_data9 type xfs (rw,noatime,nodiratime,seclabel,attr2,inode64,noquota)
/dev/sdd1 on /boot/efi type vfat (rw,relatime,fmask=0077,dmask=0077,codepage=437,iocharset=ascii,shortname=winnt,errors=remount-ro)
/dev/mapper/rhel00-home on /home type xfs (rw,relatime,seclabel,attr2,inode64,noquota)
sunrpc on /var/lib/nfs/rpc_pipefs type rpc_pipefs (rw,relatime)
```

DNS, LDAP, MSSQL, NETBIOS

NTP, SNMP, SSDP, UDP, SYN, TFTP, UDPLAG

2. Identify mapping CPUs to be allocated for ENSIGN execution

```
[root@mdchawkeye ENSIGN-42]# cat /sys/block/nvme13n1/device/device/local_cpulist
196-223,420-447   DNS, LDAP, MSSQL, NETBIOS
[root@mdchawkeye ENSIGN-42]# cat /sys/block/nvme14n1/device/device/local_cpulist
196-223,420-447   NTP, SNMP, SSDP, UDP,  SYN, TFTP, UDPLAG
[root@mdchawkeye ENSIGN-42]#
```

3. Start ENSIGN workflow with pinned cores

6. Performance Metrics Collection

Executed xGT with limited core and memory – Individual Process Monitoring

pidstat –C xgtd –u –t 5

```
[root@mdchawkeye sar_analyzer]# pidstat -C xgtd -u -t 5
Linux 3.10.0-957.el7.x86_64 (mdchawkeye.localdomain)    02/26/2020      _x86_64_        (448 CPU)

08:18:12 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
08:18:18 PM     0    449693         -    0.00    0.19    0.00    0.19     0  xgtdBinary
08:18:18 PM     0         -    449695    0.00    0.19    0.00    0.19    31  |__xgtd

08:18:18 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
08:18:23 PM     0    449693         -    0.20    0.00    0.00    0.20     0  xgtdBinary
08:18:23 PM     0         -    449716    0.00    0.20    0.00    0.20     0  |__xgtdBinary
08:18:23 PM     0    449695         -    0.00    0.20    0.00    0.20    31  xgtd

08:18:23 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command

08:18:28 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
08:18:33 PM     0    449693         -    0.00    0.20    0.00    0.20     0  xgtdBinary
08:18:33 PM     0         -    449716    0.20    0.00    0.00    0.20     0  |__xgtdBinary
08:18:33 PM     0    449695         -    0.00    0.20    0.00    0.20    31  xgtd
08:18:33 PM     0         -    449695    0.00    0.20    0.00    0.20    31  |__xgtd

08:18:33 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
08:18:38 PM     0    449693         -    0.00    0.20    0.00    0.20     0  xgtdBinary
08:18:38 PM     0         -    449716    0.00    0.20    0.00    0.20     0  |__xgtdBinary

08:18:38 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
08:18:43 PM     0    449693         -    0.20    0.00    0.00    0.20     0  xgtdBinary

08:18:43 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
08:18:48 PM     0    449693         -    0.00    0.20    0.00    0.20     0  xgtdBinary
08:18:48 PM     0         -    449716    0.00    0.20    0.00    0.20     0  |__xgtdBinary
08:18:48 PM     0         -    449717    0.20    0.00    0.00    0.20     0  |__xgtdBinary
08:18:48 PM     0    449695         -    0.00    0.20    0.00    0.20    32  xgtd
08:18:48 PM     0         -    449695    0.00    0.20    0.00    0.20    32  |__xgtd

08:18:48 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
08:18:53 PM     0         -    449716    0.00    0.20    0.00    0.20     0  |__xgtdBinary

08:18:53 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
08:18:58 PM     0    449693         -  100.00   72.40    0.00  100.00     0  xgtdBinary
08:18:58 PM     0         -    449716   63.20    2.00    0.00   65.20     0  |__xgtdBinary
08:18:58 PM     0         -    449717    2.80    1.00    0.00    3.80     0  |__xgtdBinary
08:18:58 PM     0         -    450328   62.60    0.20    0.00   62.80    91  |__xgtdBinary
08:18:58 PM     0         -    450329   58.00    0.20    0.00   58.20    50  |__xgtdBinary
08:18:58 PM     0         -    450330   64.00    0.40    0.00   64.40     3  |__xgtdBinary
08:18:58 PM     0         -    450331   65.40    0.40    0.00   65.80     6  |__xgtdBinary
08:18:58 PM     0         -    450332   66.20    0.40    0.00   66.60     4  |__xgtdBinary
08:18:58 PM     0         -    450334   65.00    0.40    0.00   65.40    16  |__xgtdBinary
08:18:58 PM     0         -    450333   65.60    0.20    0.00   65.80    18  |__xgtdBinary
08:18:58 PM     0         -    450335   65.60    0.40    0.00   66.00    19  |__xgtdBinary
08:18:58 PM     0         -    450336   66.40    0.20    0.00   66.60    13  |__xgtdBinary
08:18:58 PM     0         -    450337   65.40    0.40    0.00   65.80    15  |__xgtdBinary
08:18:58 PM     0         -    450338   65.00    0.40    0.00   65.40     5  |__xgtdBinary
08:18:58 PM     0         -    450339   63.40    0.20    0.00   63.60    12  |__xgtdBinary
08:18:58 PM     0         -    450340   66.40    0.20    0.00   66.60     7  |__xgtdBinary
08:18:58 PM     0         -    450342   65.60    0.40    0.00   66.00    10  |__xgtdBinary
```

Overall System Performance

dstat -tcmndylp

```
[root@mdchawkeye ~]# dstat -tcmndylp
---system---- ----total-cpu-usage---- ------memory-usage----- -net/total- -dsk/total- ---system-- ---load-avg--- ---procs---
     time     |usr sys idl wai hiq siq| used  buff  cach  free| recv  send| read  writ| int   csw | 1m   5m   15m |run blk new
19-03 16:17:56|  1   0  99   0   0   0| 102G 24.0M 2036G 3813G|   0     0 | 395k  664k|  48    82 | 221  132 56.7|0.0   0  8.2
19-03 16:17:57|  50   0  50   0   0   0| 102G 24.0M 2036G 3813G| 388k 2718B|   0     0 | 230   1538 | 221  132 56.7|225   0  8.0
19-03 16:17:58|  50   0  50   0   0   0| 102G 24.0M 2036G 3813G| 114k 1448B|   0     0 | 225   1463 | 221  134 57.6|224   0  7.0
19-03 16:17:59|  50   0  50   0   0   0| 102G 24.0M 2036G 3813G| 166k 5292B|   0     0 | 228   4923 | 221  134 57.6|224   0  6.0
19-03 16:18:00|  50   0  50   0   0   0| 103G 24.0M 2036G 3813G| 200k 5486B|   0     0 | 229   1543 | 221  134 57.6|224   0  7.0
19-03 16:18:01|  49   1  50   0   0   0| 103G 24.0M 2036G 3812G| 143k 1720B|   0     0 | 226   1305 | 221  134 57.6|224   0  7.0
19-03 16:18:02|  50   0  50   0   0   0| 103G 24.0M 2036G 3812G|  87k 1246B|   0     0 | 228   1548 | 221  134 57.6|224   0  7.0
19-03 16:18:03|  50   0  50   0   0   0| 103G 24.0M 2036G 3812G| 290k 1874B|   0     0 | 229   2118 | 221  135 58.5|224   0  7.0
19-03 16:18:04|  50   0  50   0   0   0| 103G 24.0M 2036G 3812G| 140k 6036B|   0     0 | 225   4821 | 221  135 58.5|224   0  6.0
19-03 16:18:05|  50   0  50   0   0   0| 103G 24.0M 2036G 3812G|  94k 4030B|   0     0 | 227   1818 | 221  135 58.5|224   0  7.0
19-03 16:18:06|  50   0  50   0   0   0| 103G 24.0M 2036G 3812G|  89k 3158B|   0    28k| 226   2481 | 221  135 58.5|224   0  7.0
19-03 16:18:07|  50   0  50   0   0   0| 103G 24.0M 2036G 3812G|  37k 9940B|   0     0 | 225   1318 | 221  135 58.5|224   0  7.0
19-03 16:18:08|  50   0  50   0   0   0| 103G 24.0M 2036G 3812G| 107k 8215B|   0     0 | 226   4525 | 222  137 59.4|225   0  15
19-03 16:18:09|  50   0  50   0   0   0| 103G 24.0M 2036G 3812G|  75k  346B|   0     0 | 229   8451 | 222  137 59.4|224   0  10
19-03 16:18:10|  50   0  50   0   0   0| 104G 24.0M 2036G 3811G|  53k  164B|   0     0 | 225   1880 | 222  137 59.4|224   0  7.0
19-03 16:18:11|  50   0  50   0   0   0| 104G 24.0M 2036G 3811G|  56k 1538B|   0     0 | 225   1688 | 222  137 59.4|224   0  12
19-03 16:18:12|  50   0  50   0   0   0| 104G 24.0M 2036G 3811G|  32k 1892B|   0     0 | 225   1908 | 222  137 59.4|224   0  7.0
19-03 16:18:13|  50   0  50   0   0   0| 104G 24.0M 2036G 3811G| 206k   14k|   0     0 | 230   7348 | 222  138 60.3|224   0  7.0
19-03 16:18:14|  50   0  50   0   0   0| 104G 24.0M 2036G 3811G| 156k   11k|   0     0 | 226   8732 | 222  138 60.3|224   0  7.0
19-03 16:18:15|  50   0  50   0   0   0| 104G 24.0M 2036G 3811G| 104k  225B|   0     0 | 230   3234 | 222  138 60.3|225   0  3.0
19-03 16:18:16|  50   0  50   0   0   0| 104G 24.0M 2036G 3811G| 306k  103B|   0     0 | 230   2429 | 222  138 60.3|224 1.0  31
19-03 16:18:17|  50   0  50   0   0   0| 104G 24.0M 2036G 3811G|  53k 1196B|   0  8192B| 227   2160 | 222  138 60.3|224   0  25
19-03 16:18:18|  50   0  50   0   0   0| 104G 24.0M 2036G 3811G| 211k 1624B|   0   388B| 227   1458 | 222  140 61.2|224   0  7.0
19-03 16:18:19|  50   0  50   0   0   0| 104G 24.0M 2036G 3811G| 193k 2750B|   0     0 | 226   5303 | 222  140 61.2|224   0  7.0
19-03 16:18:20|  50   0  50   0   0   0| 105G 24.0M 2036G 3810G| 101k   12k|   0     0 | 226   1822 | 222  140 61.2|224   0  7.0
19-03 16:18:21|  50   0  50   0   0   0| 105G 24.0M 2036G 3810G| 205k 1650B|   0     0 | 227   1358 | 222  140 61.2|224   0  6.0
19-03 16:18:22|  50   0  50   0   0   0| 105G 24.0M 2036G 3810G|  78k 1202B|   0     0 | 225   1329 | 222  140 61.2|225   0  7.0
19-03 16:18:23|  50   0  50   0   0   0| 105G 24.0M 2036G 3810G|  91k 1286B|   0     0 | 227   1433 | 222  141 62.1|224   0  7.0
19-03 16:18:24|  50   0  50   0   0   0| 105G 24.0M 2036G 3810G| 184k   23k|   0     0 | 228   5633 | 222  141 62.1|224   0  8.0
19-03 16:18:25|  50   0  50   0   0   0| 105G 24.0M 2036G 3810G|  36k   11k|   0     0 | 224   1274 | 222  141 62.1|224   0  7.0
19-03 16:18:26|  50   0  50   0   0   0| 105G 24.0M 2036G 3810G|  36k  868B|   0     0 | 226   1305 | 222  141 62.1|224   0  6.0
19-03 16:18:27|  50   0  50   0   0   0| 105G 24.0M 2036G 3810G|  93k   10k|   0     0 | 226   1372 | 222  141 62.1|224   0  7.0
19-03 16:18:28|  50   0  50   0   0   0| 105G 24.0M 2036G 3810G|  64k 6482B|   0     0 | 225   1335 | 222  143 62.9|224   0  7.0
19-03 16:18:29|  50   0  50   0   0   0| 106G 24.0M 2036G 3810G|  82k 2640B|   0     0 | 226   4880 | 222  143 62.9|224   0  7.0
19-03 16:18:30|  50   0  50   0   0   0| 106G 24.0M 2036G 3809G|  53k 6562B|   0     0 | 225   1213 | 222  143 62.9|224   0  7.0
19-03 16:18:31|  50   0  50   0   0   0| 106G 24.0M 2036G 3809G|  19k 1060B|   0     0 | 225   1251 | 222  143 62.9|224   0  6.0
19-03 16:18:32|  50   0  50   0   0   0| 106G 24.0M 2036G 3809G|  40k  992B|   0    24k| 225   1313 | 222  143 62.9|224   0  7.0
19-03 16:18:33|  49   0  51   0   0   0| 106G 24.0M 2036G 3809G|  34k  844B|   0     0 | 224   6738 | 223  144 63.8|225   0  7.0
19-03 16:18:34|  50   0  50   0   0   0| 106G 24.0M 2036G 3809G|  62k 6232B|   0     0 | 226   6253 | 223  144 63.8|225   0  7.0
19-03 16:18:35|  19  19  62   0   0   0| 109G 24.0M 2036G 3806G|  48k 9434B|   0     0 | 181   131  | 223  144 63.8|225   0  7.0
19-03 16:18:36|  50   0  50   0   0   0| 110G 24.0M 2036G 3806G|  26k 1926B|   0     0 | 226   3287 | 223  144 63.8|225   0  6.0
19-03 16:18:37|  9   3  88   0   0   0| 112G 24.0M 2036G 3803G|  34k   13k|   0    12k|  84   310  | 223  144 63.8| 24   0  7.0
19-03 16:18:38|  6  24  70   0   0   0| 113G 24.0M 2036G 3802G|  48k 1170B|   0     0 | 155   515  | 223  145 64.7|204   0  7.0
```

Execute ENSIGN with limited core and memory

Monitor dask scheduler

```
(Python_3_7) [root@mdchawkeye ~]# pidstat -C dask-scheduler -u -t 5
Linux 3.10.0-957.el7.x86_64 (mdchawkeye.localdomain)    03/19/2020      _x86_64_        (448 CPU)

05:35:23 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
05:35:29 PM     0    298961         -    3.88    0.55    0.00    4.44   136  dask-scheduler
05:35:29 PM     0         -    298961    2.96    0.37    0.00    3.33   136  |__dask-scheduler
05:35:29 PM     0         -    299320    0.37    0.18    0.00    0.55   178  |__dask-scheduler
05:35:29 PM     0         -    299443    0.55    0.00    0.00    0.55   218  |__dask-scheduler

05:35:29 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
05:35:34 PM     0    298961         -    6.80    0.60    0.00    7.40   136  dask-scheduler
05:35:34 PM     0         -    298961    5.20    0.40    0.00    5.60   136  |__dask-scheduler
05:35:34 PM     0         -    299320    0.20    0.00    0.00    0.20   183  |__dask-scheduler
05:35:34 PM     0         -    299443    1.20    0.20    0.00    1.40   205  |__dask-scheduler

05:35:34 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
05:35:39 PM     0    298961         -    5.80    1.00    0.00    6.80   136  dask-scheduler
05:35:39 PM     0         -    298961    4.40    0.60    0.00    5.00   136  |__dask-scheduler
05:35:39 PM     0         -    299320    0.40    0.20    0.00    0.60   183  |__dask-scheduler
05:35:39 PM     0         -    299443    1.00    0.20    0.00    1.20   213  |__dask-scheduler

05:35:39 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
05:35:44 PM     0    298961         -   14.60    2.40    0.00   17.00   137  dask-scheduler
05:35:44 PM     0         -    298961   11.40    1.80    0.00   13.20   137  |__dask-scheduler
05:35:44 PM     0         -    299320    1.20    0.20    0.00    1.40   191  |__dask-scheduler
05:35:44 PM     0         -    299443    1.80    0.40    0.00    2.20   412  |__dask-scheduler

05:35:44 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
05:35:49 PM     0    298961         -    7.20    0.60    0.00    7.80   418  dask-scheduler
05:35:49 PM     0         -    298961    6.40    0.40    0.00    6.80   418  |__dask-scheduler
05:35:49 PM     0         -    299320    0.20    0.00    0.00    0.20   193  |__dask-scheduler
05:35:49 PM     0         -    299443    0.60    0.20    0.00    0.80   181  |__dask-scheduler

05:35:49 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
05:35:54 PM     0    298961         -   10.00    1.20    0.00   11.20   169  dask-scheduler
05:35:54 PM     0         -    298961    8.80    0.80    0.00    9.60   169  |__dask-scheduler
05:35:54 PM     0         -    299320    0.40    0.20    0.00    0.60   180  |__dask-scheduler
05:35:54 PM     0         -    299443    1.20    0.20    0.00    1.40   178  |__dask-scheduler

05:35:54 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
05:35:59 PM     0    298961         -   36.00    3.80    0.00   39.80   160  dask-scheduler
05:35:59 PM     0         -    298961   30.60    2.80    0.00   33.40   160  |__dask-scheduler
05:35:59 PM     0         -    299320    1.40    0.20    0.00    1.60   169  |__dask-scheduler
05:35:59 PM     0         -    299443    3.80    0.80    0.00    4.60   367  |__dask-scheduler
```

Monitor dask worker

```
(Python_3_7) [root@mdchawkeye ~]# pidstat -C dask-worker -u -t 5
Linux 3.10.0-957.el7.x86_64 (mdchawkeye.localdomain)    03/19/2020      _x86_64_        (448 CPU)

05:35:27 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
05:35:32 PM     0    298963         -   15.53    5.18    0.00   20.70   173  dask-worker
05:35:32 PM     0         -    298963   14.05    5.18    0.00   19.22   173  |__dask-worker
05:35:32 PM     0         -    298966    1.29    0.00    0.00    1.29   146  |__dask-worker

05:35:32 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
05:35:37 PM     0    298963         -   15.20    6.20    0.00   21.40   173  dask-worker
05:35:37 PM     0         -    298963   13.80    6.20    0.00   20.00   173  |__dask-worker
05:35:37 PM     0         -    298966    1.40    0.20    0.00    1.60   149  |__dask-worker

05:35:37 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
05:35:42 PM     0    298963         -   28.20   18.60    0.00   46.80   196  dask-worker
05:35:42 PM     0         -    298963   24.00   18.00    0.00   42.00   196  |__dask-worker
05:35:42 PM     0         -    298966    4.40    0.40    0.00    4.80   151  |__dask-worker

05:35:42 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
05:35:47 PM     0    298963         -   26.60    9.80    0.00   36.40   161  dask-worker
05:35:47 PM     0         -    298963   25.00    9.80    0.00   34.80   161  |__dask-worker
05:35:47 PM     0         -    298966    1.80    0.00    0.00    1.80   146  |__dask-worker
```

Monitor Python Child Processes: pidstat –C python3.7 –u –t 5

```
05:35:41 PM   UID      TGID       TID    %usr %system  %guest    %CPU   CPU  Command
05:35:46 PM     0    299082         -   91.40   13.80    0.00  100.00   174  python3.7
05:35:46 PM     0         -    299082    2.40    0.60    0.00    3.00   174  |__python3.7
05:35:46 PM     0         -    299450    1.40    1.60    0.00    3.00   186  |__python3.7
05:35:46 PM     0         -    300419   43.00    6.00    0.00   49.00   178  |__python3.7
05:35:46 PM     0         -    300436   44.80    5.80    0.00   50.60   168  |__python3.7
05:35:46 PM     0    299084         -   89.80   14.60    0.00  100.00   130  python3.7
05:35:46 PM     0         -    299084    2.60    0.20    0.00    2.80   130  |__python3.7
05:35:46 PM     0         -    299489    1.00    0.80    0.00    1.80   132  |__python3.7
05:35:46 PM     0         -    301550   42.00    7.00    0.00   49.00   129  |__python3.7
05:35:46 PM     0         -    301556   44.00    6.60    0.00   50.60   359  |__python3.7
05:35:46 PM     0    299086         -   91.00   14.60    0.00  100.00   178  python3.7
05:35:46 PM     0         -    299086    2.20    0.60    0.00    2.80   178  |__python3.7
05:35:46 PM     0         -    299427    1.00    1.40    0.00    2.40   210  |__python3.7
05:35:46 PM     0         -    301522   46.40    6.80    0.00   53.20   198  |__python3.7
05:35:46 PM     0         -    301524   41.20    5.80    0.00   47.00   423  |__python3.7
05:35:46 PM     0    299088         -   88.40   16.60    0.00  100.00   197  python3.7
05:35:46 PM     0         -    299088    2.60    0.60    0.00    3.20   197  |__python3.7
05:35:46 PM     0         -    299483    0.60    2.80    0.00    3.40   210  |__python3.7
05:35:46 PM     0         -    301627   43.20    6.60    0.00   49.80   215  |__python3.7
05:35:46 PM     0         -    301634   41.60    6.60    0.00   48.20   222  |__python3.7
05:35:46 PM     0    299090         -   91.20   14.60    0.00  100.00   215  python3.7
05:35:46 PM     0         -    299090    2.40    0.40    0.00    2.80   215  |__python3.7
05:35:46 PM     0         -    299484    1.00    2.20    0.00    3.20   394  |__python3.7
05:35:46 PM     0         -    301694   46.00    6.40    0.00   52.40   190  |__python3.7
05:35:46 PM     0         -    301695   41.80    5.40    0.00   47.20   413  |__python3.7
05:35:46 PM     0    299092         -   92.20   11.20    0.00  100.00   163  python3.7
05:35:46 PM     0         -    299092    2.20    0.40    0.00    2.60   163  |__python3.7
05:35:46 PM     0         -    299580    1.20    1.40    0.00    2.60   144  |__python3.7
05:35:46 PM     0         -    300523   40.60    4.80    0.00   45.40   141  |__python3.7
05:35:46 PM     0         -    300526   48.00    4.80    0.00   52.80   385  |__python3.7
05:35:46 PM     0    299094         -   95.20   10.80    0.00  100.00   195  python3.7
05:35:46 PM     0         -    299094    2.20    0.60    0.00    2.80   195  |__python3.7
05:35:46 PM     0         -    299444    1.00    0.80    0.00    1.80   192  |__python3.7
05:35:46 PM     0         -    301669   43.20    4.60    0.00   47.80   194  |__python3.7
05:35:46 PM     0         -    301673   48.40    5.20    0.00   53.60   174  |__python3.7
05:35:46 PM     0    299096         -   91.60   13.60    0.00  100.00   202  python3.7
05:35:46 PM     0         -    299096    2.40    0.40    0.00    2.80   202  |__python3.7
05:35:46 PM     0         -    299428    1.00    1.60    0.00    2.60   202  |__python3.7
05:35:46 PM     0         -    301585   46.40    6.00    0.00   52.40   223  |__python3.7
05:35:46 PM     0         -    301587   41.60    6.00    0.00   47.60   433  |__python3.7
05:35:46 PM     0    299099         -   93.40   11.80    0.00  100.00   133  python3.7
05:35:46 PM     0         -    299099    2.60    0.40    0.00    3.00   133  |__python3.7
05:35:46 PM     0         -    299608    1.00    1.60    0.00    2.60   138  |__python3.7
05:35:46 PM     0         -    301601   42.20    5.20    0.00   47.40   340  |__python3.7
05:35:46 PM     0         -    301606   47.60    4.60    0.00   52.20   138  |__python3.7
05:35:46 PM     0    299102         -   94.40   12.20    0.00  100.00   208  python3.7
05:35:46 PM     0         -    299102    2.00    0.60    0.00    2.60   208  |__python3.7
05:35:46 PM     0         -    299675    1.00    1.40    0.00    2.40   197  |__python3.7
```

7. Troubleshooting Tips

A. Notebook Script to monitor/manage/troubleshoot during xgt execution

In [268]:
```
import xgt
import os
import pandas
from platform import python_version
```

In [269]:
```
if os.environ.get('https_proxy'):
 del os.environ['https_proxy']
if os.environ.get('http_proxy'):
 del os.environ['http_proxy']
```

xGT connection will fail if these environment variables are enabled

In [270]:
```
print (python_version())
conn=xgt.Connection()
conn.server_version
```

```
3.7.4
```

Out[270]: '1.3.0'

In [271]:
```
print("Job Monitoring")
all_jobs = conn.get_jobs()

#if len(all_jobs) == 0:
#    print("No Jobs!!!")

for j in all_jobs:
    print(j)

#Cancel all running jobs (if required)
for job in all_jobs:
    if job.status == 'running':
        conn.cancel_job(job)
        print ("calcelled JobId = {0}",job)
```

Long running jobs status monitoring and (if required can be cancelled)

```
Job Monitoring
id:2686, status:completed
id:2724, status:completed
id:2746, status:completed
id:2752, status:completed
id:2780, status:completed
id:2818, status:completed
```

In [272]:
```
#If your data is LANL Data then fetch details of lanl data loaded in xgt
print("Nodes/Edges Monitoring LANL")
devices = conn.get_vertex_frame('Devices')
netflow = conn.get_edge_frame('Netflow')
events1v = conn.get_edge_frame('HostEvents')
events2v = conn.get_edge_frame('AuthEvents')

print('Devices (vertices): {:,}'.format(devices.num_vertices))
print('Netflow (edges): {:,}'.format(netflow.num_edges))
print('Host-only event (edges): {:,}'.format(events1v.num_edges))
print('Host-communication event (edges): {:,}'.format(events2v.num_edges))
print('Total number of edges: {:,}'.format(netflow.num_edges + events1v.num_edges + events2v.num_edges))
```

Print edges and vertices (LANL Data)

```
Nodes/Edges Monitoring LANL
Devices (vertices): 137,812
Netflow (edges): 235,661,328
Host-only event (edges): 18,637,483
Host-communication event (edges): 47,790,045
Total number of edges: 302,088,856
```

In [273]:
```
#If your data is generic and not lanl data
print("Nodes/Edges Monitoring NON-LANL")
tables = conn.get_table_frames()
for table in tables:
    print("TableFrame {0} has {1:,} rows".format(table.name, table.num_rows))

vertices = conn.get_vertex_frames()
for vertex in vertices:
    print("VertexFrame {0} has {1:,} rows".format(vertex.name, vertex.num_rows))

edges = conn.get_edge_frames()
for edge in edges:
    print("EdgeFrame {0} has {1:,} rows".format(edge.name, edge.num_rows))
```

Print edges and vertices (Non-LANL Data)

```
Nodes/Edges Monitoring NON-LANL
TableFrame answers has 10,572 rows
VertexFrame Devices has 137,812 rows
EdgeFrame Netflow has 235,661,328 rows
EdgeFrame HostEvents has 18,637,483 rows
EdgeFrame AuthEvents has 47,790,045 rows
EdgeFrame RDPFlow has 9,732 rows
EdgeFrame HijackEvents has 166,889 rows
EdgeFrame PrivEscEvents has 135,214 rows
```

In [281]:
```
#Print specific table data
print("Print Specific Table Data")
tableName = "answers"
table = conn.get_table_frame(tableName)
numRows = table.num_rows
schema = table.schema

if numRows == 0:
    print("Zero Rows in Table")

#print schema
print("Print Table Schema for table {0} :",tableName)
result = ""
for columns in schema:
    result += "\n"
    colName = columns[0]
    if columns[1] == 'int':
        colType = 'xgt.INT'
    elif columns[1] == 'text':
        colType = 'xgt.TEXT'
    elif columns[1] == 'float':
        colType = 'xgt.FLOAT'
    elif columns[1] == 'datetime':
        colType = 'xgt.DATETIME'
    else:
        colType = 'xgt.UNKNOWN'
    if len(result) > 0:
        result += ", "
    result += "{0}:{1}".format(colName, colType)
print(result)

#print data
print("Print data for table {0} :",tableName)
data = table.get_data_pandas()
data[0:10]
data
```

Viewing sample table data or error tables

```
Print Specific Table Data
Print Table Schema for table {0} : answers

, rdp1_src_device:xgt.TEXT
, rdp1_dst_device:xgt.TEXT
, rdp1_epoch_time:xgt.INT
, rdp2_dst_device:xgt.TEXT
, rdp2_epoch_time:xgt.INT
Print data for table {0} : answers
```

Out[281]:

	rdp1_src_device	rdp1_dst_device	rdp1_epoch_time	rdp2_dst_device	rdp2_epoch_time
0	ActiveDirectory	EnterpriseAppServer	7290427	Comp503532	7291140
1	ActiveDirectory	EnterpriseAppServer	7290427	Comp503532	7291140

```
In [305]: #Data Pre-processing
input_file="/nvme_data1/data_2v/wls_day-86_2v.csv"
output_file="/nvme_data9/data_2v/wls_day-86_2v.csv"

print("Pre-processing specific field in input file : {0}",input_file)

LOGHOST = 2
SOURCE = 12
DEST = 14

with open(output_file, 'w') as outfile:
    writer = csv.writer(outfile, quoting=csv.QUOTE_MINIMAL)
    with open(input_file, 'r') as infile:
        reader = csv.reader(infile)
        for row in reader:
            if row[DEST] == '':
                row[DEST] = row[LOGHOST]
            if row[SOURCE] == '':
                row[SOURCE] = row[LOGHOST]
            writer.writerow(row)

print("Output file generation complete : {0}",output_file)
```

Data Preprocessing for specific fields

```
Pre-processing specific field in input file : {0} /nvme_data1/data_2v/wls_day-86_2v.csv
Output file : {0} /nvme_data9/data_2v/wls_day-86_2v.csv
```

B. Script to monitor/manage/troubleshoot during ENSIGN execution

```
[root@mdchawkeye ~]# cd ashish/
[root@mdchawkeye ashish]# cd $ENSIGN_BASE
[root@mdchawkeye ENSIGN-42]# ls
check_prereqs.sh      console_output_run_3  ENSIGN_install.sh        env.sh.bkp   output_run_1  output_run_4  output_run_7  README.txt
console_output_run_1  domain_map_3.tsv      ENSIGN_UserGuide_4.2.pdf  global.lock  output_run_2  output_run_5  output_run_8  workflow_cfg.yml
console_output_run_2  ENSIGN_4.2            env.sh                   output       output_run_3  output_run_6  purge.lock    workflow_cfg.yml.bkp
[root@mdchawkeye ENSIGN-42]# cd ENSIGN_4.2/
bin/          Ensign-C/     Ensign-CAPI/ Ensign-Py3/
[root@mdchawkeye ENSIGN-42]# cd ENSIGN_4.2/Ensign-Py3/
bin/     ensign/  test_data/
[root@mdchawkeye ENSIGN-42]# cd ENSIGN_4.2/Ensign-Py3/ensign/
cli/          cyber_detectors/ ensign_ctypes/   ensign_dask/     ensign_io/       __pycache__/
[root@mdchawkeye ENSIGN-42]# cd ENSIGN_4.2/Ensign-Py3/ensign/ensign_dask/
.dask_out/  __pycache__/
[root@mdchawkeye ENSIGN-42]# cd ENSIGN_4.2/Ensign-Py3/ensign/ensign_dask/
[root@mdchawkeye ensign_dask]# ls
dask_util.py  global.lock  __init__.py  purge.lock  __pycache__  start_dask.sh  start_dask.sh.bkp  start_dask.sh.bkp2  stop_dask.sh
[root@mdchawkeye ensign_dask]# cat start_dask.sh
#!/bin/bash

NUM_PROCS=$1   # Set to number of cores or threads available
DASK_OUT=$ENSIGN_BASE/ENSIGN_4.2/Ensign-Py3/ensign/ensign_dask/.dask_out

dask-scheduler --local-directory $DASK_OUT > /dev/null 2>&1 &
#dask-scheduler &
#taskset -c 140-223,364-447 dask-scheduler &
#taskset -c 112-223,336-447 dask-scheduler &
#taskset -c 112-223 dask-scheduler &

for i in `seq 1 1 1` ; do
    dask-worker 127.0.0.1:8786 --local-directory $DASK_OUT --nprocs $NUM_PROCS --memory-limit 0 > /dev/null 2>&1 &
    #dask-worker 127.0.0.1:8786 --nprocs $NUM_PROCS --memory-limit 0 --dashboard-address 8787 &
    #dask-worker 127.0.0.1:8786 --nprocs $NUM_PROCS --memory-limit 0 --dashboard-address 8787 &
    #taskset -c 140-223,364-447 dask-worker 127.0.0.1:8786 --nprocs $NUM_PROCS --memory-limit 0 &
    #taskset -c 112-223,336-447 dask-worker 127.0.0.1:8786 --nprocs $NUM_PROCS --memory-limit 0 &
    #taskset -c 112-223 dask-worker 127.0.0.1:8786 --nprocs $NUM_PROCS --memory-limit 0 &
done

sleep 10
```

start_dask.sh

```
[root@mdchawkeye ensign_dask]# cat stop_dask.sh
#!/bin/bash

killall dask-worker
killall dask-scheduler

#sleep 8
sleep 15
[root@mdchawkeye ensign_dask]#
```

stop_dask.sh

C. CICDDoS Dataset 2019

```
[root@mdchawkeye nvme_data14]# cat get_data.sh
#GET CIC DDOS2019 Dataset in pcap format
wget "http://205.174.165.80/CICDataset/CICDDoS2019/Dataset/PCAPs/03-11/PCAP-03-11.zip" -O "/nvme_data14/CICDataset/CICDDoS2019/Dataset/PCAPs/PCAP-03-11.zip"
wget "http://205.174.165.80/CICDataset/CICDDoS2019/Dataset/PCAPs/01-12/PCAP-01-12_0-0249.zip" -O "/nvme_data14/CICDataset/CICDDoS2019/Dataset/PCAPs/PCAP-01-12_0-0249.zip"
wget "http://205.174.165.80/CICDataset/CICDDoS2019/Dataset/PCAPs/01-12/PCAP-01-12_0250-0499.zip" -O "/nvme_data14/CICDataset/CICDDoS2019/Dataset/PCAPs/PCAP-01-12_0250-0499.zip"
wget "http://205.174.165.80/CICDataset/CICDDoS2019/Dataset/PCAPs/01-12/PCAP-01-12_0500-0749.zip" -O "/nvme_data14/CICDataset/CICDDoS2019/Dataset/PCAPs/PCAP-01-12_0500-0749.zip"
wget "http://205.174.165.80/CICDataset/CICDDoS2019/Dataset/PCAPs/01-12/PCAP-01-12_0750-0818.zip" -O "/nvme_data14/CICDataset/CICDDoS2019/Dataset/PCAPs/PCAP-01-12_0750-0818.zip"


#GET CIC DDOS2019 Dataset in csv format
wget "http://205.174.165.80/CICDataset/CICDDoS2019/Dataset/CSVs/CSV-01-12.zip" -O "/nvme_data14/CICDataset/CICDDoS2019/Dataset/csv/CSV-01-12.zip"
wget "http://205.174.165.80/CICDataset/CICDDoS2019/Dataset/CSVs/CSV-03-11.zip" -O "/nvme_data14/CICDataset/CICDDoS2019/Dataset/csv/CSV-03-11.zip"
[root@mdchawkeye nvme_data14]#
```

D. LANL Dataset 2019

```
[root@mdchawkeye nvme_data1]# cat get_data.sh
for i in {1..90}; do wget "https://datasets.trovares.com/LANL/xgt/wls_day-"$i"_2v.csv" -O "/nvme_data1/data_1v/wls_day-"$i"_2v.csv"; done
for i in {1..90}; do wget "https://datasets.trovares.com/LANL/xgt/wls_day-"$i"_1v.csv" -O "/nvme_data1/data_2v/wls_day-"$i"_1v.csv"; done
for i in {1..90}; do wget "https://datasets.trovares.com/LANL/xgt/nf_day-"$i".csv" -O "/nvme_data1/data_nf/nf_day-"$i".csv"; done
[root@mdchawkeye nvme_data1]#
```