Hewlett Packard
Enterprise

Performance Execution xGT & Ensign on SDFX

CYBERSECURITY GRAPH & AI – COMBINED EXECUTION CONFIGURATION AND PERFORMANCE – POST-OPTIMIZATION (MULTIPLE DATASET)

Test Run : 03/18/2020								
ENSIGN					xGT			
Environment Parameters	Combined Execution	Standalone Execution (limited core/memory)	Standalone Execution (full scale)		xGT 1.3 (LANL Attack Performance)	Combined Execution	Standalone Execution (limited core/memory)	Standalone Execution (full scale)
Python Version	3.7.4				Python Version	3.7.4		
Python Package Manager	Anaconda				Python Package Manager	PyPi		
S/W Version	4.2				S/W Version	1.3.0		
Input Data(GB) - CICDDoS2019 (DNS+LDAP+MSSQL+NetBIOS+NTP+SNMP+SSDP+UDP+Syn+TFTP+UDPLag)	21				Input Data(GB) - LANL Day 85 (1v+2v+nf)	19.7		
Total Cores Available	112(50%)	112(50%)	224(100%)		Total Cores Available	112(50%)	112 (50%)	224(100%)
Total Cores Allocated	112(50%)	112(50%)	224(100%)		Total Cores Allocated	112(50%)	112(50%)	224(100%)
Total Memory Allocated (TB)	3	3	6		Total Memory Allocated (TB)	3	3	6
Total No. of Components	50	50	50		Total No. of Edges	302,088,856	302,088,856	302,088,856
csv2tensor(sec) - Including disk write	421	421	423		RDP Hack Edges	10,572	10,572	10,572
Decomposing tensor w/ CP (sec) - Including disk write	103	103	103		Total Data Load Time (1v+2v+nf)	314.1	317.2	143
visualization (sec)	63	60	91		Extract Forward RDP Edges Time (sec)	1.02	1.01	1.04
Textual Report Generation (sec)	7	7	14		Extract Reverse RDP Edges Time (sec)	1.04	1.01	1.04
Total Time Taken (sec)	594	591	631		Lateral Movement Query Execution Time(sec)	1.43	1.47	1.52

Note :
a) rank: 50
b) num_threads: 112(half-scale)/224(full-scale)
c) mem_limit_gb: 3000
d) dump_tensor_files: True
e) dump_decomposition_files: True
f) data drives = /dev/nvme13n1 = vCPUs = 196-223,420-447
g) data drives = /dev/nvme14n1 = vCPUs = 196-223,420-447
h) taskset -c 112-223,336-447 workflow.py /root/ashish/reservoir_lab/ENSIGN-42/workflow_cfg.yml

Note :
a)"worker_threads": 224,
b)"io_threads": 224,
c)"port": 4367,
d)"max_memory": 3298534883328,
e)"pin_threads": true
f) data drives = 1V = /nvme_data1 = /dev/nvme0n1 = vCPUs = 0-27,224-251
g) data drives = 2V = /nvme_data3 = /dev/nvme2n1 = vCPUs = 28-55,252-279
h) data drives = nf = /nvme_data5 = /dev/nvme4n1 = vCPUs = 56-83,280-307
i) taskset -c 0-111,224-335 xgtd -c /opt/xgtd/xgtd.conf

BEST PRACTICES – COMBINED EXECUTION OF XGT & ENSIGN ON SDFX

- Distribute the input data across multiple NVMe data drives. Keep each input file upto a max of 2GB
- Execute XGT & ENSIGN against mapped CPU cores aligned to respective NVMe data drives
- Sequence / time shift the initial data load operation into xGT & Ensign (most costly operation).
- When running in parallel, properly setting CPU affinity has a big impact on system performance.
  - Independent vCPU and memory allocation for both xgtd and Ensign
  - Allocation should be based on use case, not necessarily a percentage of system resources. Starting point is 50%
- Leverage individual environment variables for Python Package Manager implementation.

- <u>xGT Specific:</u>
  - Consider distributing Host, Authentication and Netflow Logs data across different NVMe drives
  - Different resource consumption models based on log data type (e.g. Netflow = resource intensive. Host logs = lighter resource utilization).
  - Advantages seen when pinning processes to CPUs (e.g. Lateral Movement Query execution)
  - Python Package Manager: PyPi
  - Required Dataset for xGT: CSV

- <u>Ensign Specific:</u>
  - Data Load & Tensor Decomposition phase are most resource consuming.
  - Choosing the right number of components will have an impact. Based on our testing, we used 25 components (50% of 8S/6TB) as the starting point.
  - Configure OMP Threads proportionate to allocated cores, and set “NUM_THREADS” mapping to allocated physical cores
  - DO NOT configure any other specific OpenMP tunings like OMP_PLACES , OMP_PROC_BIND has it would impact the performance
  - Intermittent results dump of tensor decomposition should be disabled by default and enabled on need basis.
  - Python Package Manager: Anaconda
  - Required Dataset for Ensign: PCAP/CSV

PERFORMANCE VALIDATION WITH TROVARES XGT 1.3 ON SDFX

- Targeted threats :
  - Lateral Movement
  - Privilege Escalation
  - Hijack Event
- Datasets :
  - Search 90-days
  - Netflow and Log events
  - ~20 billion graph edges
  - ~212 billion edge properties
  - 3.2TB of RAM to hold graph