Sizing Guide - ClickHouse

FortiSIEM 6.7.0

FORTINET DOCUMENT LIBRARY
https://docs.fortinet.com

FORTINET VIDEO GUIDE
https://video.fortinet.com

FORTINET BLOG
https://blog.fortinet.com

CUSTOMER SERVICE & SUPPORT
https://support.fortinet.com

FORTINET TRAINING & CERTIFICATION PROGRAM
https://www.fortinet.com/training-certification

FORTINET TRAINING INSTITUTE
https://training.fortinet.com

FORTIGUARD CENTER
https://www.fortiguard.com

END USER LICENSE AGREEMENT
https://www.fortinet.com/doc/legal/EULA.pdf

FEEDBACK
Email: techdoc@fortinet.com

FORTINET®

10/02/2023
FortiSIEM 6.7.0 Sizing Guide - ClickHouse

TABLE OF CONTENTS

Change Log

Date	Change Description
03/30/2018	Initial release of FortiSIEM Sizing Guide.
04/12/2018	Revision 2 with updates to Storage Requirements for FortiSIEM EventDB and Elasticsearch Data Nodes sections.
11/20/2019	Sizing Guide released for 5.2.6.
03/30/2020	Sizing Guide release for 5.3.0.
09/08/2020	Sizing Guide release for 6.1.0.
03/23/2021	Sizing Guide release for 6.2.0.
04/12/2021	Sizing Guide updated with Sizing Online Deployments and Sizing Archive Deployments for 6.2.0.
05/06/2021	Sizing Guide release for 6.2.1.
06/15/2021	FSM-3500G information added for 6.2.x.
07/06/2021	Sizing Guide release for 6.3.0.
08/26/2021	Sizing Guide release for 6.3.1.
10/15/2021	Sizing Guide release for 6.3.2.
12/22/2021	Sizing Guide release for 6.3.3.
01/04/2022	Minimum Requirements Hardware section updated for 6.x Sizing guides.
01/05/2022	Spark / HDFS Resource Allocation Considerations added to HDFS Based Deployments section for 6.4.0.
01/18/2022	Sizing Guide release for 6.4.0.
03/09/2022	Spark / HDFS Resource Allocation Considerations section updated for 6.4.0 Sizing Guide.
05/09/2022	Sizing Guide release for 6.5.0.
05/31/2022	Added: Hardware Appliance EPS Test with ClickHouse, Cluster Wide Shard Count Limit (Elasticsearch), ClickHouse Based Deployment
07/26/2022	Sizing Guide release for 6.6.0. Virtual Appliance EPS Test with ClickHouse Database section added. Sizing Online Deployments - ClickHouse Based Deployment section updated.
08/23/2022	Added: Shard Count column for ClickHouse Software Based Deployments (Minimum Requirement) and ClickHouse Software Based Deployments (Recommended Requirement).

Date	Change Description
08/24/2022	Update to ClickHouse Software Based Deployments (Minimum Requirement) and ClickHouse Software Based Deployments (Recommended Requirement) tables.
09/12/2022	Sizing Guide release for 6.5.1.
09/14/2022	Sizing Guide release for 6.6.1.
09/19/2022	Sizing Guide release for 6.6.2.
02/07/2023	Sizing Guide - ClickHouse release for 6.7.0.
02/13/2023	Sizing Guide - ClickHouse release for 6.7.1.
02/21/2023	Updated ClickHouse Software Based Deployments section, and added table to Storage Requirements.
03/07/2023	Sizing Guide - ClickHouse release for 6.7.2.
03/28/2023	Sizing Guide - ClickHouse release for 6.7.3.
04/11/2023	Sizing Guide - ClickHouse release for 6.7.4.
05/12/2023	Updated Software Based Deployments table - ClickHouse Topology column.
05/22/2023	Sizing Guide - ClickHouse release for 6.7.5.
06/16/2023	Sizing Guide - ClickHouse release for 6.7.6.
07/13/2023	Sizing Guide - ClickHouse release for 6.7.7.
08/02/2023	Hardware table > Workers (Keeper Only Node) Local Disks updated.
09/12/2023	Sizing Guide - ClickHouse release for 6.7.8.
10/02/2023	Added OPT information under Minimum Requirements - Hardware.

FortiSIEM Sizing Guide - ClickHouse

This document provides information about the following topics:

Minimum Requirements

Hardware

Minimum hardware requirements for FortiSIEM nodes are as follows.

Node	vCPU	RAM	Local Disks
Supervisor (All in one)	Minimum – 12 Recommended - 32	Minimum • without UEBA – 24GB • with UEBA - 32GB Recommended • without UEBA – 32GB • with UEBA - 64GB	OS – 25GB OPT – 100GB CMDB – 60GB SVN – 60GB ClickHouse DB - based on EPS and retention
Supervisor (Cluster)	Minimum – 12 Recommended - 32	Minimum • without UEBA – 24GB • with UEBA - 32GB Recommended • without UEBA – 32GB • with UEBA - 64GB	OS – 25GB OPT – 100GB CMDB – 60GB SVN – 60GB ClickHouse DB - based on EPS and retention
Workers (Data Node)	Minimum – 16 Recommended - 32	Minimum – 32GB Recommended • without UEBA – 64GB • with UEBA - 64GB	OS – 25GB OPT – 100GB ClickHouse DB - based on EPS and retention

Node	vCPU	RAM	Local Disks
Workers (Keeper Only Node)	Minimum 8 Recommended 16	Minimum - 16GB Recommended 16 GB	OS – 25GB OPT – 100GB Data - 200GB
Collector	Minimum – 4 Recommended – 8 (based on load)	Minimum – 4GB Recommended – 8GB	OS – 25GB OPT – 100GB

- Supervisor VA needs more memory since it hosts many heavy-duty components such as Application Server (Java), PostGreSQL Database Server and Rule Master.
- For OPT - 100GB, the 100GB disk for /opt will consist of a single disk that will split into 2 partitions, /OPT and swap. The partitions will be created and managed by FortiSIEM when `configFSM.sh` runs.

Note that these are only the minimum requirements. The performance may improve by increasing vCPUs and RAM in certain situations. External storage depends on your EPS mix and the number of days of log storage needs. To provide more meaningful guidance, scalability tests were conducted as described below.

Internal Scalability Tests

FortiSIEM team performed several scalability tests described below.

Test Setup

- A specific set of events were sent repeatedly to achieve the target EPS.
- The target EPS was constant over time.
- A set of Linux servers were monitored via SNMP and performance monitoring data was collected.
- Events triggered many incidents.

Test Success Criteria

The following success criteria should be met on testing:

- Incoming EPS must be sustained without any event loss.
- Summary dashboards should be up to date and not fall behind.
- Widget dashboards should show data indicating that inline reporting is keeping up.
- Incidents should be up to date.
- Real-time search should show current data and trend chart should reflect incoming EPS.
- GUI navigation should be smooth.
- CPU, memory and IOPS are not maxed out. Load average must be less than the number of cores.

The tests were run for the following cases:

- All-in-one FSM Hardware Appliance: FSM-2000F and FSM-3500F with collectors FSM-500F sending events.

Hardware Appliance EPS Test with ClickHouse

The test bed is shown below. Scripts generated events on FSM-500F Collectors, which parsed those events and sent to the appliances.

	Event Sender				
FortiSIEM HW Appliance	**Hardware Spec**	**Collector Model**	**Count**	**EPS/Collector**	**Sustained EPS without Loss**
FSM-2000F	2000F - 12vCPU (1x6C2T), 32GB RAM, 12x3TB SATA (3 RAID Groups)	FSM-500F	3	5K	15K
FSM-2000G	2000G - 40vCPU (2x10C2T), 128GB RAM, 4x1TB SSD (RAID5), 8x4TB SAS (2 RAID50 Groups)	FSM-500F	6	7K	40K
FSM-3500G	3500G,48vCPU (2x12C2T),128GB RAM,24x4TBSATA (3 RAID50 groups)	FSM-500F	6	8K	40K

Notes:

1. Event Ingestion speed increased two fold in FSM-2000G with ClickHouse compared to FortiSIEM EventDB. ClickHouse event database made better utilization of the vCPUs in the system.
2. Since FSM-2000F has fewer vCPU compared to FSM-2000G, the performance of both FortiSIEM EventDB and ClickHouse are identical. The appliance is CPU bound.
3. For FortiSIEM 3500G, the insert performance of FortiSIEM EventDB and ClickHouse is identical as FortiSIEM EventDB could also use disk striping for better I/O.

Virtual Appliance EPS Test with ClickHouse Database

All tests were done in AWS. The following hardware was used.

Node Type	AWS Instance Type	Hardware Specification
Collector	c5.2xlarge	8 vCPU, 16 GB.
Worker as ClickHouse Keeper node	C6a.8xlarge	32 vCPU, 64 GB, SSD 125Mbps throughput
Worker as ClickHouse Data/Query Node	C6a.8xlarge	32 vCPU, 64 GB, SSD 1GBps throughput
Supervisor	m6a.8xlarge	32 vCPU, 128 GB, CMDB Disk 10K IOPS

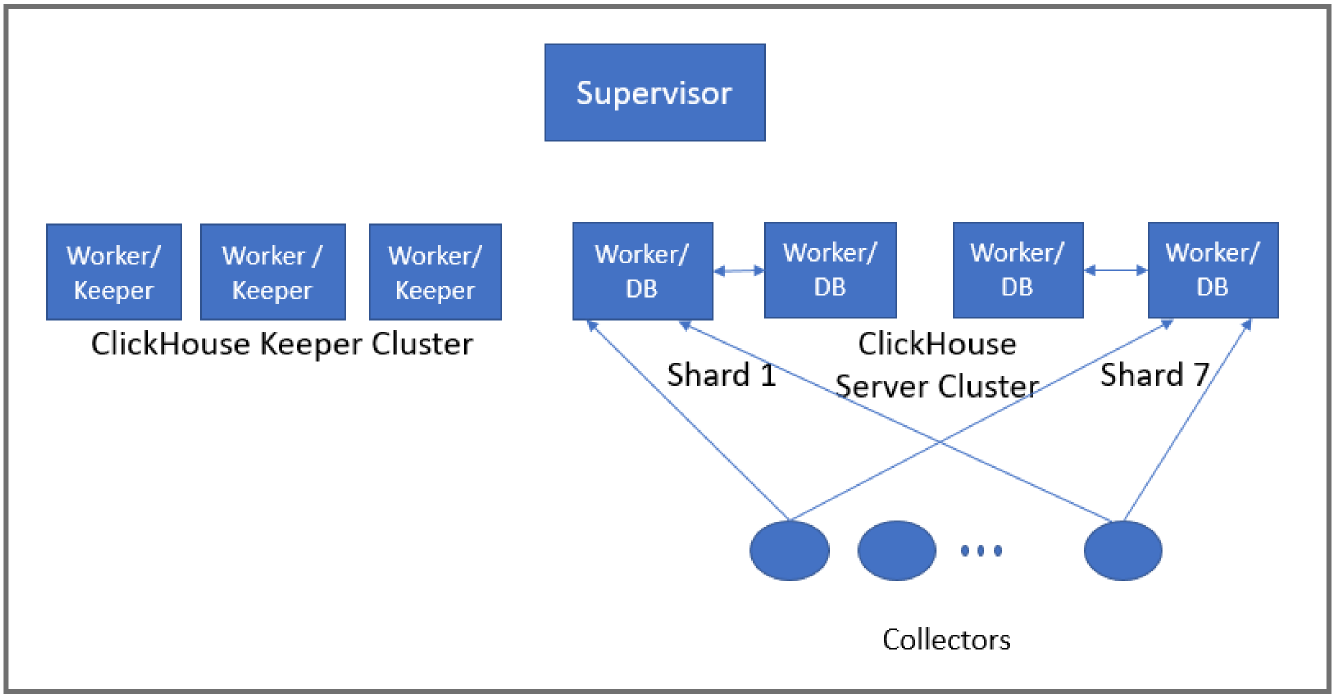
Based on the requirement to handle 500K EPS, the following setup was used:

- 1 Supervisor
- 3 Worker nodes as part of ClickHouse Keeper Cluster
- 14 Worker nodes as part of ClickHouse Server Cluster
  - 7 shards
  - 2 Workers in each shard. This means that 2 copies of each event were kept (Replication = 2).
- 150 Collectors, each sending 3.3K EPS to the 14 Workers in the ClickHouse Server Cluster, in a round robin fashion. Each Worker replicated its received events to the other Worker within the same shard.
- Collectors could also send events to the ClickHouse Keeper Cluster nodes, but this was not done. The ClickHouse Keeper Cluster nodes were dedicated to Replication management.
- Each Worker handles 35.7K EPS.

See ClickHouse Configuration in the latest Online Help for details on setting up ClickHouse Clusters.

See the testbed below. Scripts generated events on the Collectors, which were sent to the Workers. Service provider deployment was used. There were 150 Organizations and each Collector belonging to an Organization discovered and monitored the performance of 150 other Collectors in other Organization. This resulted in 22.5K devices in CMDB and each were being discovered using SNMP and monitored for basic performance metrics including CPU, Memory, Disk and Network interface utilization.

500K EPS were sustained without any event loss for over 2 days. 5 users logged on the system and ran queries and visited various parts of the user interface.



Sizing Online Deployment

Processing Requirement

- Hardware Appliance Deployments
- **Software Based Deployments**

Hardware Appliance Deployments

EPS	Deployment	Replication	Hardware Model	Network
0-20K	Hardware	1	2000F, 2000G, 3500G	1Gbps
20K-40K	Hardware	1	2000G, 3500G	1Gbps

Software Based Deployments

Software based deployments can be scaled out to handle more EPS by adding shards and adding Worker nodes in each shard. See ClickHouse Operational Overview for details. Follow these principles for a stable deployment:

1. Whenever possible, deploy separate ClickHouse Keeper nodes. This is true especially at medium to high EPS or you will run into many concurrent heavy-duty queries. In these cases, Keeper functionality may compete for CPU, Memory, and Disk I/O resources with Insert and Query. If Keeper does not get resources, replication will stop, database will become read only and event insertion stops. In the table below, Fortinet recommends **3 dedicated Keeper nodes** for 60K EPS and above. For 20K-60K, dedicated Keeper nodes is an option.
2. If more than 50% Keeper nodes are lost, then RAFT protocol quorum is lost and database may become read only, and event insertion stops. For this reason, Fortinet recommends **3 Keeper nodes** whenever possible as it can sustain 1 lost node.
   a. If you run 2 Keeper nodes, then loss of 1 node causes quorum to be lost and database may become read only.
   b. If you run 1 Keeper node, then loss of 1 node causes complete loss of Keeper cluster and database may become read only.

   In both these cases, follow the steps in Recovering from Losing Quorum to recover from lost quorum or complete keeper cluster loss. Using more than 3 Keeper nodes may lead to increased replication overhead.
3. **Use SSD for Hot Tier**, especially for medium to high EPS. This will speed up event insertion and queries.
4. If you need to handle more EPS, then add more shards, using the table below as a guide.
5. If you need to make queries run faster, there are two options:
   a. Add more shards
   or
   b. Add more Data + Query nodes in existing shards

   Both these approaches will spread out the data to more nodes.

Requirement		Configuration	
EPS	Replication	Supervisor/Worker Hardware	ClickHouse Topology
0-5K	1 (meaning 1 copy of events)	1 Supervisor – 16vCPU, 24GB RAM, 200MBps Disk	1 Shard with 1 Replica The Shard has Supervisor with Data and Query flag checked. Supervisor is also Keeper node
0-5K	2 (meaning 2 copies of events)	1 Supervisor – 16vCPU, 24GB RAM, 200MBps Disk 1 Worker – 16vCPU, 24GB RAM, 200MBps Disk 1 Gbps Network	1 Shard with 2 Replicas The Shard has Supervisor and Worker with both Data and Query flags checked. Supervisor is also Keeper Node
5K-10K	1	1 Supervisor – 32vCPU, 32GB RAM, 200MBps Disk	1 Shard with 1 Replica The Shard has Supervisor with both Data and Query flags checked. Supervisor is also Keeper node
5K-10K	2	1 Supervisor – 16vCPU, 32GB RAM, 200MBps Disk 1 Worker – 16vCPU, 32GB RAM, 200MBps Disk 1 Gbps Network	1 Shard with 2 Replicas The Shard has Supervisor and Worker with both Data and Query flags checked. Supervisor is also Keeper Node
10K-20K	1	1 Supervisor - 48vCPU, 64GB RAM, 200MBps Disk	1 Shard with 1 Replica The Shard has Supervisor with both Data and Query flags checked. Supervisor is also Keeper node
10K-20K	2	1 Supervisor – 32vCPU, 64GB RAM, 200MBps Disk 1 Worker – 32vCPU, 64GB RAM, 200MBps Disk 1 Gbps Network	1 Shard with 2 Replicas The Shard has Supervisor and Worker with both Data and Query flags checked. Supervisor is also Keeper Node
20K-30K	1	1 Supervisor – 48vCPU, 64GB RAM, 200MBps Disk 1 Worker – 32vCPU, 64GB RAM, 200MBps Disk 1 Gbps Network	1 Shard with 1 Replica The Shard has Supervisor with both Data and Query flags checked. Supervisor is also Keeper node
20K-30K	2	1 Supervisor – 32vCPU, 64GB RAM, 200MBps Disk 2 Workers – 32vCPU, 64GB RAM, 200MBps Disk	1 Shard with 2 Replicas The Shard has 2 Workers with both Data and Query flags checked. Supervisor is also Keeper Node

Requirement		Configuration	
EPS	Replication	Supervisor/Worker Hardware	ClickHouse Topology
		1 Gbps Network	
		1 Supervisor – 32vCPU, 64GB RAM, 200MBps Disk 2 Workers – 32vCPU, 64GB RAM, 200MBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 1 Gbps Network	1 Shard with 2 Replicas The Shard has 2 Workers with both Data and Query flags checked. 3 Workers (16vCPU) acting as Keeper only
30K-60K	2	1 Supervisor – 32vCPU, 64GB RAM, 200MBps Disk 2 Workers – 32vCPU, 64GB RAM, 500MBps Disk 1 Worker – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	1 Shard with 2 Replicas Each shard – 2 (32vCPU) Workers with both Data and Query flags checked. 1 Worker (16vCPU) acting as Keeper only
		1 Supervisor – 32vCPU, 64GB RAM, 200MBps Disk 2 Workers – 32vCPU, 64GB RAM, 500MBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	1 Shard with 2 Replicas Each shard – 2 (32vCPU) Workers with both Data and Query flags checked. 3 Workers (16vCPU) acting as Keeper only
60K-125K	2	1 Supervisor – 32vCPU, 64GB RAM, 200MBps Disk 4 Workers – 32vCPU, 64GB RAM, 500MBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	2 Shards with 2 Replicas per shard Each shard has 2 (32vCPU) Workers with both Data and Query flags checked. 3 (16vCPU) Workers acting as dedicated Keeper Nodes
125K-175K	2	1 Supervisor – 32vCPU, 64GB RAM, 200MBps Disk 6 Workers – 32vCPU, 64GB RAM, 500MBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	3 Shards with 2 Replicas per shard Each shard has 2 (32vCPU) Workers with both Data and Query flags checked. 3 (16vCPU) Workers acting as dedicated Keeper Nodes

Requirement		Configuration	
EPS	Replication	Supervisor/Worker Hardware	ClickHouse Topology
175K-250K	2	1 Supervisor – 32vCPU, 64GB RAM, 200MBps Disk 8 Workers – 32vCPU, 64GB RAM, 500MBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	4 Shards with 2 Replicas per shard Each shard has 2 (32vCPU) Workers with both Data and Query flags checked. 3 (16vCPU) Workers acting as dedicated Keeper Nodes
250K-300K	2	1 Supervisor – 32vCPU, 64GB RAM, 200MBps Disk 10 Workers – 32vCPU, 64GB RAM, 500MBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	5 Shards with 2 Replicas per shard Each shard has 2 (32vCPU) Workers with both Data and Query flags checked. 3 (16vCPU) Workers acting as dedicated Keeper Nodes
300K-360K	2	1 Supervisor – 32vCPU, 64GB RAM, 200MBps Disk 12 Workers – 32vCPU, 64GB RAM, 500MBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	6 Shards with 2 Replicas per shard Each shard has 2 (32vCPU) Workers with both Data and Query flags checked. 3 (16vCPU) Workers acting as dedicated Keeper Nodes
360K-420K	2	1 Supervisor – 32vCPU, 64GB RAM, 200MBps Disk 14 Workers – 32vCPU, 64GB RAM, 1GBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	7 Shards with 2 Replicas per shard Each shard has 2 (32vCPU) Workers with both Data and Query flags checked. 3 (16vCPU) Workers acting as dedicated Keeper Nodes
420K-500K	2	1 Supervisor – 32vCPU, 64GB RAM, 200MBps Disk 16 Workers – 32vCPU, 64GB RAM, 1GBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	8 Shards with 2 Replicas per shard Each shard has 2 (32vCPU) Workers with both Data and Query flags checked. 3 (16vCPU) Workers acting as dedicated Keeper Nodes
500K-550K	2	1 Supervisor – 32vCPU, 64GB RAM, 500MBps Disk	9 Shards with 2 Replicas per shard

Requirement		Configuration	
EPS	Replication	Supervisor/Worker Hardware	ClickHouse Topology
		18 Workers – 32vCPU, 64GB RAM, 1GBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	Each shard has 2 (32vCPU) Workers with both Data and Query flags checked. 3 (16vCPU) Workers acting as dedicated Keeper Nodes
550K-600K	2	1 Supervisor – 32vCPU, 64GB RAM, 500MBps Disk 20 Workers – 32vCPU, 64GB RAM, 1GBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	10 Shards with 2 Replicas per shard Each shard has 2 (32vCPU) Workers with both Data and Query flags checked. 3 (16vCPU) Workers acting as dedicated Keeper Nodes
600K-650K	2	1 Supervisor – 32vCPU, 64GB RAM, 500MBps Disk 22 Workers – 32vCPU, 64GB RAM, 1GBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	11 Shards with 2 Replicas per shard Each shard has 2 (32vCPU) Workers with both Data and Query flags checked. 3 (16vCPU) Workers acting as dedicated Keeper Nodes
650K-700K	2	1 Supervisor – 32vCPU, 64GB RAM, 500MBps Disk 24 Workers – 32vCPU, 64GB RAM, 1GBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	12 Shards with 2 Replicas per shard Each shard has 2 (32vCPU) Workers with both Data and Query flags checked. 3 (16vCPU) Workers acting as dedicated Keeper Nodes
700K-750K	2	1 Supervisor – 32vCPU, 64GB RAM, 500MBps Disk 26 Workers – 32vCPU, 64GB RAM, 1GBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	13 Shards with 2 Replicas per shard Each shard has 2 (32vCPU) Workers with both Data and Query flags checked. 3 (16vCPU) Workers acting as dedicated Keeper Nodes
750K-800K	2	1 Supervisor – 32vCPU, 64GB RAM, 500MBps Disk 28 Workers – 32vCPU, 64GB RAM, 1GBps Disk	14 Shards with 2 Replicas per shard Each shard has 2 (32vCPU) Workers with both Data and Query flags checked.

Requirement		Configuration	
EPS	Replication	Supervisor/Worker Hardware	ClickHouse Topology
		3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	3 (16vCPU) Workers acting as dedicated Keeper Nodes
850K-900K	2	1 Supervisor – 32vCPU, 64GB RAM, 500MBps Disk 30 Workers – 32vCPU, 64GB RAM, 1GBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	15 Shards with 2 Replicas per shard Each shard has 2 (32vCPU) Workers with both Data and Query flags checked. 3 (16vCPU) Workers acting as dedicated Keeper Nodes
900K-950K	2	1 Supervisor – 32vCPU, 64GB RAM, 500MBps Disk 32 Workers – 32vCPU, 64GB RAM, 1GBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	16 Shards with 2 Replicas per shard Each shard has 2 (32vCPU) Workers with both Data and Query flags checked. 3 (16vCPU) Workers acting as dedicated Keeper Nodes
950K-1M	2	1 Supervisor – 32vCPU, 64GB RAM, 500MBps Disk 34 Workers – 32vCPU, 64GB RAM, 1GBps Disk 3 Workers – 16vCPU, 16GB RAM, 200MBps Disk 10Gbps Network	17 Shards with 2 Replicas per shard Each shard has 2 (32vCPU) Workers with both Data and Query flags checked. 3 (16vCPU) Workers acting as dedicated Keeper Nodes

For more than 1 million EPS, contact FortiSIEM Professional Services.

See ClickHouse Usage Recommendations in References for more information.

Storage Requirement

FortiSIEM storage requirement depends on the following factors:

- EPS
- Bytes/event
- Compression ratio
- Retention period

Typically, EPS peaks during morning hours on weekdays and goes down dramatically after 2 pm on weekdays and also remains low on weekends. So the average EPS should be used to calculate storage needs.

Bytes/event depends on the event types and their rate in your environment. Typically, the Bytes/event is dominated by Windows AD network authentication logs, Firewall permit/deny logs, and Netflow, which are relatively short ~500 Bytes. General Windows Security logs tend to be a little larger (~1000 Bytes) and Cloud logs tend to be much larger (unto 10K Bytes sometimes).

Fortinet has chosen LZ4 compression algorithm in ClickHouse, which provides 4:1 compression ratio. Note that higher compression would require higher Worker CPU and memory to achieve the same insertion and Query rates as LZ4, so other higher compression algorithms were not chosen.

The storage requirement is simply EPS * Bytes/event * Compression ratio * Retention period (remember to normalize the units).

It is best for the user to estimate or measure the EPS and Bytes/event for their environment. If you have stored a sufficient mix of events in a file, then you can count Bytes/event as the file size divided by the number of lines in that file.

The following example illustrates the storage requirements.

- Suppose in your environment peak EPS is 10K and average EPS is 2K. An estimate may be 6K.
- Bytes/event is 500 Bytes
- Compression ratio 4:1
- Retention period 2 weeks (14 days) in Hot storage and 2.5 months (76 days) in Warm storage
- Replication = 2 (meaning 2 copies of data)

Then

- Storage per day: (2 * 6000 * 86400 * 500) / (4 * 1024 * 1024 * 1024) GB = 122GB
  The general formula is:
  Storage per day = (Replication * EPS * Seconds in a day * (Bytes/Event)) / (Compression * 1024 * 1024 * 1024) GB
- Hot storage requirement for 14 days
  - Cluster wide: 1.7TB
  - Assuming 2 Data/Query Nodes, per node storage is 854GB
- Warm storage requirement for 76 days
  - Cluster Wide : 9.4TB
  - Assuming 2 Data/Query Nodes, per node storage is 4.7TB

The following table provides storage requirements for a range of EPS, under the following assumptions:

- Bytes/event is 500 Bytes
- Storage compression ratio 4:1

This leads to Storage need to be 10.5GB per day, per 1K Average EPS and replication = 1 (meaning 1 copy of data).

Requirement		Storage Configuration per Data Node	
Average EPS	Replication	Hot = 1 week; Warm = 6 months	Hot = 2 weeks; Warm = 1 year
1K	1 (meaning 1 copy of events)	Supervisor is the Data node. Hot Tier: 73.5GB Warm Tier: 1.85TB	Supervisor is the Data node. Hot Tier: 147GB Warm Tier: 3.7 TB

Requirement		Storage Configuration per Data Node	
Average EPS	Replication	Hot = 1 week; Warm = 6 months	Hot = 2 weeks; Warm = 1 year
1K	2 (meaning 2 copies of events)	Supervisor is the Data node. Hot Tier: 73.5GB Warm Tier: 9.23TB	Supervisor is the Data node. Hot Tier: 147GB Warm Tier: 3.7 TB
5K	2	Supervisor and Worker are the Data nodes. Storage for each node: Hot Tier: 367.5GB Warm Tier: 9.23TB	Supervisor and Worker are the Data nodes. Storage for each node: Hot Tier: 735GB Warm Tier: 18.5TB
10K	2	Supervisor and Worker are the Data nodes. Storage for each node: Hot Tier: 735GB Warm Tier: 18.45TB	Supervisor and Worker are the Data nodes. Storage for each node: Hot Tier: 1.43TB Warm Tier: 37TB
25K	2	Supervisor and Worker are the Data nodes. Storage for each node: Hot Tier: 1.8TB Warm Tier: 46.2TB	Supervisor and Worker are the Data nodes. Storage for each node: Hot Tier: 3.59GB Warm Tier: 92.3TB
50K	2	Supervisor and Worker are the Data nodes. Storage for each node: Hot Tier: 3.59GB Warm Tier: 92.3TB	Supervisor and Worker are the Data nodes. Storage for each node: Hot Tier: 7.18GB Warm Tier: 184.58TB
75K	2	4 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 2.7TB Warm Tier: 69.2TB	4 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 5.4TB Warm Tier: 138.4TB
100K	2	4 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 5.4TB Warm Tier: 138.4TB	4 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 7.18TB Warm Tier: 184.6TB
125K	2	6 (32vCPU) Workers are the Data nodes. Storage for each node:	6 (32vCPU) Workers are the Data nodes.

Requirement		Storage Configuration per Data Node	
Average EPS	Replication	Hot = 1 week; Warm = 6 months	Hot = 2 weeks; Warm = 1 year
		Hot Tier: 3TB Warm Tier: 77TB	Storage for each node: Hot Tier: 6TB Warm Tier: 153.8TB
150K	2	6 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 3.59TB Warm Tier: 92.3TB	6 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 7.18TB Warm Tier: 184.5TB
175K	2	6 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 4.19TB Warm Tier: 107.67TB	6 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 8.38TB Warm Tier: 215.33TB
200K	2	8 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 3.59TB Warm Tier: 92.3TB	8 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 7.18TB Warm Tier: 184.5TB
250K	2	8 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 4.49TB Warm Tier: 115.33TB	8 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 8.98TB Warm Tier: 230.7TB
300K	2	10 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 4.31TB Warm Tier: 110.75TB	10 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 8.61TB Warm Tier: 221.49TB
400K	2	14 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 4.11TB Warm Tier: 105.47TB	14 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 8.21TB Warm Tier: 210.94TB
500K	2	18 (32vCPU) Workers are the Data nodes. Storage for each node:	18 (32vCPU) Workers are the Data nodes.

Requirement		Storage Configuration per Data Node	
Average EPS	Replication	Hot = 1 week; Warm = 6 months	Hot = 2 weeks; Warm = 1 year
		Hot Tier: 3.99TB Warm Tier: 102.54TB	Storage for each node: Hot Tier: 7.98TB Warm Tier: 205.08TB
600K	2	20 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 4.31TB Warm Tier: 110.75TB	20 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 8.62TB Warm Tier: 221.49TB
700K	2	24 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 4.19TB Warm Tier: 107.67TB	24 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 8.38TB Warm Tier: 215.34TB
800K	2	28 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 4.11TB Warm Tier: 105.47TB	28 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 8.21TB Warm Tier: 210.94 TB
900K	2	30 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 4.31TB Warm Tier: 110.75TB	30 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 8.62TB Warm Tier: 221.49TB
1M	2	34 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 4.23TB Warm Tier: 108.57TB	34 (32vCPU) Workers are the Data nodes. Storage for each node: Hot Tier: 8.45TB Warm Tier: 217.15TB

References

ClickHouse Usage Recommendations

https://clickhouse.com/docs/en/operations/tips/

FORTINET®

www.fortinet.com