VMware Carbon Black EDR 7.x OER Frequently Asked Questions (FAQ)

Why was the 6.x OER revised?

VMware Carbon Black noticed that in certain edge cases, customers were experiencing storage/performance issues. Engineering, Cloud Operations, and the Technical Services Organization launched a project to triage and perform a root cause assessment of what was discovered to be an issue related to the resource sizing and configuring of VMware Carbon Black EDR environments.

How is the 7.x OER different from the 6.x OER?

The 7.x OER provides much more detail to explain how resource sizing is accomplished. In the 7.x OER, Carbon Black EDR environment resources can be sized in a tailored fashion based on the number of endpoints, data volume, and required days of retention. Each of these variables can be tuned to the specific technical and business needs of a customer. The 7.x OER implements flexible resource sizing methodology to align data volume and retention with the appropriate resources (Sensor Data Volume X Days of Retention = Resources Required).

What types of environments are most likely to be impacted by this updated OER?

The updated OER contains significantly more detail regarding sensor data accumulation and how this data volume affects resources. For customers with sensors that are primarily installed on Windows endpoints, there are typically no significant changes. The analysis produced a much more robust sizing model for macOS and Linux sensors.

What versions of Carbon Black EDR does the 7.x OER apply to?

The Carbon Black EDR 7.x OER is retroactive and applies to 6.x and 7.x installations of Carbon Black EDR.

Will I experience a performance / retention impact by upgrading to Carbon Black EDR 7.x?

No. Carbon Black EDR 6.x & 7.x use roughly the same system resources. While there are performance efficiencies and improvements in Carbon Black EDR 7.x, both versions use a similar processing/data storage architecture. Performance and retention are not significantly impacted by simply upgrading from 6.x to 7.x.

How is Sensor Data Volume Calculated?

Sensor data volume is calculated by taking the daily number of processes (documents) that are generated by a certain sensor and multiplying this by the size of these process documents. A significant amount of analysis has gone into building and validating the models found in the 7.x OER related to sensor data collection.

If I am experiencing performance issues and cannot upgrade system resources, what are my options?

The 7.x OER has introduced flexible resource sizing; therefore, you can make certain adjustments in an existing environment's configuration to align current data volumes with existing resources. *Resources Required = Sensor Data Volume X Days of Retention*. If you cannot adjust the resources in an environment, you can bring an environment into compliance by either reducing the Days of Retention or the Sensor Data Volume. Both items can be accomplished with a Health Check from Professional Services.

What are the ramifications if the OER is not adhered to?

You will experience similar performance / retention as with Carbon Black EDR 6.x. The 7.x OER was written specifically to address niche cases where customers were experiencing performance / retention issues. If an environment is experiencing performance / retention issues, and resources are not compliant with the OER, the environment must be brought into compliance from a hardware (or configuration perspective) before it is eligible for support assistance. **Note**: OER compliance status does not affect eligibly for Carbon Black support to assist with product issues that are not related to performance/retention.

How were the various percentile tables gathered?

The percentage tables represent average levels of activity categorized by the sensor operating system. Examples of the range of system activity levels include: kiosks, labs, workstations, high activity users, IT admins, servers, etc.

Are there suggested percentiles to use when sizing a specific type of endpoint?

When estimating the sensor process document for a traditional enterprise, it is typically recommended to evenly split the number of Windows workstations between the median and 75th percentile. Due to their higher levels of activity, it is recommended to size Windows servers, macOS workstations, and Linux servers using the 90th percentile due to provide a better assurance that edge cases are covered. In most cases, one can use the median when calculating the process document size (3.6 KB).

An example for estimating 1 day of sensor data for 100 endpoints (50 Windows Workstations, 15 Windows servers, 25 Macs, and 10 Linux servers) is as follows:

Count	OS	Process Docs	Doc Size	Count x Docs x Doc Size
25	Windows Workstations	7,800 (Median)	3.6 KB (Median)	(25 x 7,800 x 3.6) / 1,024 = 685.55 MB/Day
25	Windows Workstations	10,750 (75th)	3.6 KB (Median)	(25 x 10,750 x 3.6) / 1,024 = 944.82 MB/Day
15	Windows Servers	16,000 (90th)	3.6 KB (Median)	(15 x 16,000 x 3.6) / 1,024 = 843.75 MB/Day
25	Mac	25,500 (90th)	3.6 KB (Median)	(25 x 25,500 x 3.6) / 1,024 = 2,241.21 MB/Day
10	Linux	195,750 (90th)	3.6 KB (Median)	(10 x 195,750 x 3.6) / 1,024 = 6,881.84 MB/Day
100	All	N/A	N/A	11.597 GB/Day

Can sensor data volume be collected from existing environments?

The 7.x OER provides an estimate for the volume of data in a net new environment. If an environment already has sensors deployed, you can run the `/usr/share/cb/cbdiag` command to collect the data aggregation statistics that are located on each server, that process and store sensor data. (This is typically the minion servers). The following process can be used to calculate the current sensor data volume:

1. For each data server:
    a. Run the `/usr/share/cb/cbdiag` (defaults to `/tmp` directory).
    b. Extract the `/cb_services/cbfs-http/datastore_stats.txt` file from the archive.
    c. Navigate to the sections named **EventStoreStats**.
        - Note the **CoreIndexSize** (Bytes)
        - Note the **ShardName**
    d. Determine the average CoreIndexSize per day:
        - To determine the number of day's data within a specific shard, compare the date in the ShardName with the date in the previous (or following) shard. (Typically, this is approximately 3 days).
        - Convert the CoreIndexSize into Gigabytes by dividing the CoreIndex Size by (1024^3)
        - Divide the CoreIndexSize by the number of days calculated in step 4.
        - To increase accuracy you can repeat and average this across each of the EventStoreStats sections
2. After the average CoreIndexSize per day has been calculated for each data server:
    a. Calculate and combine the Daily Average EventStoreStats for each data server to get the Cumulative Daily Average EventStoreStats. This Cumulative Daily Average EventStoreStats can be used as a reference for estimating retention based on the current sensor data volume.