

# Open Virtual Appliance (OVA) configuration requirements

## Application stack

- vCPU:2
- Memory: 4 GB
- Storage:
  - Primary disk (vmdisk1): 21 GB
  - Data disk (vmdisk2): 10 GB

## Gateway server

- vCPU:4
- Memory: 8 GB
- Storage:
  - Primary disk (vmdisk1): 100 GB
  - Data disk (vmdisk2): 75 GB
  - Swap disk (vmdisk3): 40 - 100 GB

## Dataserver - Classic

- vCPU:4

Even though increasing the number of vCPUs increases performance, the actual benefits depend on whether the specific host is oversubscribed or not.


- Memory: 16 GB

While the minimum value works under light-load condition, as the load increases, the data server quickly starts using swap space. For high load situations, increasing RAM beyond 16 GB can benefit performance. Monitoring swap usage can provide insight.

- Storage:
  - Primary disk (vmdisk1, SCSI 0:0): Default is 150 GB

This virtual disk has an associated VMDK that contains the IBM StoredIQ operating code. Do not change its size.

- 

⚠ CAUTION	
	<p><b>Attention</b></p> <p>If you delete the primary disk, you delete the operating system, and the IBM StoredIQ software; the virtual m</p>

- Data disk (vmdisk2, SCSI 0:1): Default is 2 TB

This virtual disk can be resized according to expectations on the amount of harvest data to be stored. For purposes of estimation, the index storage requirement for metadata is about 30 GB per TB of managed source data. Full-text indexing requires an extra 170 GB per TB. Therefore, the default data disk size is targeted for managing 10 TB of source information.

- Swap disk (vmdisk3, SCSI 0:2): Default is 40 GB

When under load, the data server can use much RAM; therefore, having ample swap space is prudent. The minimum swap size is equal to the amount of RAM configured for the virtual machine. For best performance under load, place this disk on the highest speed data store available to the host.

The general size limits for a data server are 150 million objects or 500 defined volumes, whichever limit is reached first. Assuming an average object size of 200 KB equals about 30 TB of managed storage across 30 volumes of 5 million objects each, the index storage requirement for metadata on ~30 TB of storage that contains uncompressed general office documents is ~330 GB (11 GB per TB). Add 100 GB per TB of managed storage for full-text or snippet index. For example, to support 30 TB of storage that is indexed for metadata, you need 8 TB indexed for full-text search and extracted text (snippet cache) of 8 TB for auto-classification. A total of 1.9 TB of storage is required (metadata: 330 GB, full-text: 800 GB, snippet cache: 800 GB).

Data-server performance is impacted by the IOPS available from the storage subsystem. For each data server under maximum workload, at least 650 IOPS generally delivers acceptable performance. In the situations where there is a high load on the system, the IOPS that is used can reach up to 7000 with main write operations.

### **DataServer - Distributed**

- vCPU:4
- Memory: 16 GB
- Storage:
  - Primary disk (vmdisk1, SCSI 0:0): Default is 150 GB
  - Data disk (vmdisk2, SCSI 0:1): Default is 2 TB
  - Swap disk (vmdisk3): (vmdisk3, SCSI 0:2): Default is 40 GB

If you deploy this type of data server, you must also deploy an Elasticsearch cluster with at least one node. If you deploy a cluster with more nodes, each of the Elasticsearch nodes must meet the listed requirements.

### **Each Elasticsearch node**

- vCPU:1
- Memory: 32 GB
- Storage:
  - Primary disk (vmdisk1): 100 GB
  - Data disk (vmdisk2): 1 TB

The required memory depends on the index size that is handled by the data node. For a better performance, consider increasing the memory. The total memory available to the node must be distributed between the host operating system, the Elasticsearch java heap space, and the kernel file system caches. For example, if the system has 32 GB memory, 2 GB must be allocated for the host operating system, 15 GB for the java heap space, and the remaining 15 GB for the file system caches.

The data disk (vmdisk2) can be resized according to expectations on the amount of harvest data to be stored.