CouchBase

Srinivas Maddali

Source: CouchBase documentation

Supported Versions of Operating Systems

Platform	Version	32 / 64 bit	Supported	Recommended Version
Red Hat Enterprise Linux	5	32 and 64 bit	Developer and Production	RHEL 5.8
Red Hat Enterprise Linux	6	32 and 64 bit	Developer and Production	RHEL 6.3
CentOS	5	32 and 64 bit	Developer and Production	CentOS 5.8
CentOS	6	32 and 64 bit	Developer and Production	CentOS 6.3
Amazon Linux	2011.09	32 and 64 bit	Developer and Production	
Ubuntu Linux	10.04	32 and 64 bit	Developer and Production	
Ubuntu Linux	12.04	32 and 64 bit	Developer and Production	Ubuntu 12.04
Windows 2008	R2 with SP1	64 bit	Developer and Production	Windows 2008
Windows 2012		64 bit	Developer and Production	
Windows 7		32 and 64 bit	Developer only	
Windows 8		32 and 64 bit	Developer only	
Mac OS	10.7	64 bit	Developer only	
Mac OS	10.8	64 bit	Developer only	Mac OS 10.8

Resource Requirements

The following hardware requirements are recommended for installation:
01. Quad-core for key-value store, 64-bit CPU running at 3GHz
02. Six cores if you use XDCR and views.
03. 16GB RAM (physical)
04. Block-based storage device (hard disk, SSD, EBS, iSCSI).
05. **Network filesystems (e.g. CIFS, NFS) are not supported**.

A minimum specification machine should have the following characteristics:

01. Dual-core CPU running at 2GHz for key-value store
02. 4GB RAM (physical)

Note

For development and testing purposes a reduced CPU and RAM than the minimum specified can be used. This can be as low as 1GB of free RAM beyond operating system requirements and a single CPU core. However, you should not use a configuration lower than that specified in production. Performance on machines lower than the minimum specification will be significantly lower and should not be used as an indication of the performance on a production machine.

View performance on machines with less than 2 CPU cores will be significantly reduced.

You must have enough memory to run your operating system and the memory reserved for use by Couchbase Server.

For example, if you want to dedicate 8GB of RAM to Couchbase Server you must have enough RAM to host your operating system. If you are running additional applications and servers, you will need additional RAM. For smaller systems, such as those with less than 16GB you should allocate at least 40% of RAM to your operating system.

You must have the following amount of storage available:

- 1GB for application logging
- At least twice the disk space to match your physical RAM for persistence of information

For information and recommendations on server and cluster sizing

Server Sizing Guide Lines

Here are the primary considerations when sizing your Couchbase Server cluster:

- How many nodes do I need?
- How large (RAM, CPU, disk space) should those nodes be?

To answer the first question, consider following factors:

- RAM
- Disk throughput and sizing
- Network bandwidth
- Data distribution and safety

Due to the in-memory nature of Couchbase Server, RAM is usually the determining factor for sizing. But ultimately, how you choose your primary factor will depend on the data set and information that you are storing.

- If you have a very small data set that gets a very high load, you'll need to base your size more off of network bandwidth than RAM.
- If you have a very high write rate, you'll need more nodes to support the disk throughput needed to persist all that data (and likely more RAM to buffer the incoming writes).
- Even with a very small dataset under low load, you may want three nodes for proper distribution and safety.

With Couchbase Server, you can increase the capacity of your cluster (RAM, Disk, CPU, or network) by increasing the number of nodes within your cluster, since each limit will be increased linearly as the cluster size is increased.

RAM sizing

01. Leaving 40% to
    a. the OS

b. inbound connection management until they reach database

c. other application that work on the server for the database

d. and monitoring the server and database

60% is to be allocated. This is driven by the working set of the application. Working set is the data set of the application. Ideally at any given point in time what amount of data is used by the application to manage::

a. its job/work and

b. user requests

02. the application variables that determine the memory requirements

Variable	Description
documents_num	The total number of documents you expect in your working set
ID_size	The average size of document IDs
value_size	The average size of values
number_of_replicas	The number of copies of the original data you want to keep
working_set_percentage	The percentage of your data you want in memory
per_node_ram_quota	How much RAM can be assigned to Couchbase

Use the following items to calculate how much of memory you need:

Constant	Description
Metadata per document (metadata_per_document)	This is the amount of memory that Couchbase needs to store metadata per document. Metadata uses 56 bytes. All the metadata needs to live in memory while a node is running and serving data.
SSD or Spinning	SSDs give better I/O performance.
headroom	The cluster needs additional overhead to store metadata. That space is called the headroom. This requires approximately 25-30% more space than the raw RAM requirements for your dataset. Since SSDs are faster than spinning (traditional) hard disks, you should set aside 25% of memory for SSDs and 30% of memory for spinning hard disks.
High Water Mark (high_water_mark)	By default, the high water mark for a node's RAM is set at 85%.

This is a rough guideline to size your cluster:

Variable	Calculation
no_of_copies	`1 + number_of_replicas`
total_metadata All the documents need to live in the memory.	`(documents_num) * (metadata_per_document + ID_size) * (no_of_copies)`
total_dataset	`(documents_num) * (value_size) * (no_of_copies)`
working_set	`total_dataset * (working_set_percentage)`
Cluster RAM quota required	`(total_metadata + working_set) * (1 + headroom) / (high_water_mark)`
number of nodes	`Cluster RAM quota required / per_node_ram_quota`

You will need at least the number of replicas + 1 nodes regardless of your data size.

Here is a sample sizing calculation:

Input Variable	value
documents_num	1,000,000
ID_size	100
value_size	10,000
number_of_replicas	1
working_set_percentage	20%

Constants	value
Type of Storage	SSD
overhead_percentage	25%
metadata_per_document	56 for 2.1 and higher, 64 for 2.0.x
high_water_mark	85%

Variable	Calculation
no_of_copies	= 1 for original and 1 for replica
total_metadata	= 1,000,000 * (100 + 56) * (2) = 312,000,000
total_dataset	= 1,000,000 * (10,000) * (2) = 20,000,000,000
working_set	= 20,000,000,000 * (0.2) = 4,000,000,000
Cluster RAM quota required	= (440,000,000 + 4,000,000,000) * (1+0.25)/(0.7) = 7,928,000,000

For example, if you have 8GB machines and you want to use 6 GB for Couchbase…

```
Number of nodes =

    Cluster RAM quota required/per_node_ram_quota =

    7.9 GB/6GB = 1.3 or 2 nodes
```

RAM quota

You will not be able to allocate all your machine RAM to the per_node_ram_quota as there may be other programs running on your machine.