

Recommended Environments for Clearspace Deployments

This document outlines recommended hardware and environmental configurations for Clearspace and Clearspace X deployments.

General Requirements

Internal Deployments (Clearspace) – Resources Based on Number of Users

Note: The number of users represents the total number of users, not "active" users. It's assumed that around 20 percent of the total users are active.

Users	CPUs	RAM	Location of App Server and Database	Clustered?
100	2 dual core CPUs (at least 2 GHz)	2GB, 1GB Java heap	Share a box	No
500	2 dual core CPUs (at least 2 GHz)	2GB, 1GB Java heap	Separate boxes	No
1000	2 dual core CPUs (at least 2 GHz)	2GB, 2GB Java heap	Separate boxes	Yes
5000	2 dual core CPUs (at least 2 GHz)	4GB, 2GB Java heap	Separate boxes	Yes
10000	4 dual core CPUs (at least 2 GHz)	4GB, 2GB Java heap	Separate boxes	Yes

External Deployments (Clearspace X) – Resources Based on Page Views

For external deployments, the configurations can be similar but the activity is measured in different ways. Also, the association between page views and content is not a hard rule. For example, it would be possible to have more reads (page views) than content creation (content / month) on a busy site.

Page Views per Month	Content Created per Month	CPUs	RAM	Location of App Server and Database	Clustered?
500,000	20,000	2 dual core CPUs (at least 2 GHz)	2GB, 1GB Java heap	Share a box	No
1,000,000	20,000	2 dual core CPUs (at least 2 GHz)	2GB, 1GB Java heap	Separate boxes	No

5,000,000	50,000	2 dual core CPUs (at least 2 GHz)	2GB, 1.5GB Java heap	Separate boxes	Yes
10,000,000	50,000	2 dual core CPUs (at least 2 GHz)	4GB, 2GB Java heap	Separate boxes	Yes
25,000,000	100,000	4 dual core CPUs (at least 2 GHz)	4GB, 2GB Java heap	Separate boxes	Yes

Example JVM Settings

It's important to use the latest JVM (Java Virtual Machine) for your platform as well as the correct GC (garbage collection) settings. Clearspace requires at least a 1.5 VM but runs best with 1.6.

Heap	Startup Parameters
1 GB	-Xms512m -Xmx1024m -XX:MaxPermSize=256m -XX:+UseParNewGC -XX:+UseConcMarkSweepGC
2 GB	-Xms512m -Xmx2048m -XX:MaxPermSize=256m -XX:+UseParNewGC -XX:+UseConcMarkSweepGC

In both cases this will allocate a 512 MB heap at minimum as well as allocate 256 MB to the "PermGen" space of the VM heap. Finally, we configure the "Concurrent Mark Sweep" GC algorithm. We've found this to be the optimal GC setting for deployments on multiple core processor configurations.

If you want to debug the contents of your VM, use the following extra startup parameters:

Debug Parameters
-verbose:gc -XX:+PrintGC -XX:+PrintGCDetails -XX:+PrintGCTimeStamps -Xloggc:/path/to/log/file/here.log

Please refer to your application server's documentation for information on setting these startup parameters.

Time Configuration

For clustered configurations, it is absolutely essential that all cluster nodes have system time with as little drift as possible. This is generally adjusted on a regular interval by syncing with an NTP server inside the firewall. For servers under higher load, it may be necessary to increase the clock sync interval as loaded servers will be more likely to experience clock drift.

Network Topology

It's important to keep your network as simple as possible to minimize environmental interference. For example, a bad network architecture would contain multiple firewalls, proxies and network hops between

your users, the application server and the database. Keep the following things in mind:

- **A proxy in front of the application server is not required, but can be useful in some cases.** For example, a web proxy can show a nice "site is down" message if you restart your application server and it can assist with network port translations. If you can run your application without a web proxy then that is recommended.
- **Keeping the boxes on the same network is important.** You can have too much latency in your environment if your application servers and/or database live on different physical networks. It's best to configure those machines on the same network.
- **Evaluate the quality of switches and other network gear in and around your deployment.** Bad switches will drop packets and reduce the overall quality of your network. Further, you must use gigabit Ethernet and switches.

Example Hardware

- Dell PowerEdge 2950 2U Server
 - 2 2.33 GHz Dual Core Xeon CPUs
 - 4 GB RAM
 - Linux OS
 - RAID 5 controller with 3 hard drives