

PortageII 4.0

Recommended System Configuration

Updated June 2018

For training PortageII 4.0

Strict minimum configuration: 32 GB RAM, 4 cores, 1 TB disk space, plus 64 GB swap space on disk.

Recommended configuration: 128 GB RAM, 16 cores, 1 GPU, 2+ TB disk space, 256 GB swap space on disk.

RAM: a strict minimum of 32 GB is required, 64 GB or 128 GB is recommended, and even more can still be useful. For multi-core machines, you should have around 8 GB of RAM per core (or virtual core, if you use hyperthreading), so you can maximize parallelism.

Number of CPUs and cores: it's worth having many cores (4 to 16). Technically, there is no minimum requirement, but anything below 4 will be painfully slow; 16 is recommended for good performance. PortageII can take advantage of all available cores, and can also take advantage of clustered environments with multiple machines sharing the same file system.

GPU: training NNJMs requires a GPU with at least 6GB of RAM.

Disk space: this depends on how many systems you plan to train, but you should plan no less than 1 TB, while several to many TBs will likely be useful. This storage does not need to be local; it can be on an NFS server, on a SAN or some other shared storage solution.

Swap space: in your disk purchases, allow for twice as much local disk space for swap as you have RAM on each machine. This is on top of the disk space you need for your work (discussed above), and on top of other local disk space required for the OS and /tmp. The 2:1 swap:RAM ratio is important to respect : having less swap space than that will cause problems.

OS: PortageII only runs on Linux 64 bits. We have used Scientific Linux, Red Hat, OpenSUSE, Ubuntu; PortageII is compatible with most Linux distros. The preferred choices are CentOS or Red Hat 6 or 7.

Virtualization: compatible but not recommended for training PortageII.

For translating using PortageII 4.0

Recommended configuration: this depends entirely on the size of your corpora, the number of models, and the number of users.

Approximate recommendations based on the number of professional translators using the system daily:

- **10 professional translators:** 16 GB RAM, 4 cores, 100 GB disk space, plus 32 GB swap space on disk.
- **100 professional translators:** 64 GB RAM, 16 Cores, 1 TB disk space, plus 128 GB swap space on disk.

RAM: you should plan for enough RAM to hold all the models in memory, plus a bit more. Depending on the size of your corpora, each model should be in the 2-8 GB range. If you plan to allow parallel access to your translation server, allow some more RAM, but not linearly: concurrent requests that use the same models will share the memory required for these models, so only the memory needed for the computations increases; however, doing concurrent requests using different models means adding the memory requirements for both models and computation space.

Number of CPUs and cores: PortageII typically translates about 1 sentence per second. You should calculate the number of cores based on the number of concurrent requests you want to submit and the total volume of text you need to translate during the peak hours of work. PortageII will run with just 1 core, but we don't recommend fewer than 4 if you have multiple users, or fewer than 16 if you have many users.

Disk space: the requirements are quite small, with the system only needing enough space for the models and the temporary files. In a virtualized environment with tight resources, systems could be deployed with just 10's of GBs, or more if you plan to have many systems on the same server, but 100GB is recommended if you have just one server for several users, 1TB for many users. Periodic clean-up of the translation server is recommended, as temporary files are not automatically removed.

Swap space: follow the 2:1 swap:RAM ratio, as discussed above.

OS: Linux 64 bits (see above for details).

Virtualization: PortageLive allows for easily deploying and running translation servers running Linux in a virtual computing environment (e.g., using VMWare on a PC or a centralized data centre), although a multi-core physical server will allow you to better exploit the benefits of memory-mapped IO and obtain faster performance.

Parallelization and Load balancing: PortageLive servers can be duplicated and placed behind a load balancer only if the asynchronous methods are not used (the `translateFile*()` methods in the SOAP and REST APIs or the ability to translate a whole file in the cgi and html demo pages).

For using Incremental Document Adaptation with PortageII 4.0

Incremental Document Adaptation involves doing model training directly on your PortageLive server, and therefore requires significantly more hardware resources. One update will take between 10 seconds and 1 minute. The application is configured to run at most one update process at a time per document, so the requirements will depend on how many different document models are being used and updated simultaneously.

CPU cores and RAM: You should plan at least one dedicated CPU core with 4 GB of RAM per active document model in use. For example, if you have 16 professional translators each working with his or her own distinct document model, you would need a server with 16 cores and 64 GB of RAM. However, you can control what a “document” refers to: for example, if you use the translation project number as “document id” and an average of four translators are working on each project, the 16 core + 64 GB machine would suffice for 64 translators.

Disk space: plan 1 GB of additional disk space per document model you plan to keep on the server, on top of the disk space required by PortageLive itself, as described on the previous page.

Swap space: follow the 2:1 swap:RAM ratio, as discussed above.

OS: Linux 64 bits (see above for details).

Virtualization: PortageLive can still be run on a virtual server if all of the above requirements are met, even if Incremental Document Adaptation is used.

Parallelization and Load balancing: Incremental Document Adaptation is incompatible with server duplication since the document model is stored on the instance of PortageLive where training is launched. If you need to scale to multiple servers and use Incremental Document Adaptation, you should place different models on different servers, or assign individual translators to a specific server.