

10. Instance Cost Reports

PECE depends on a set of Free and Open Source technologies that constitute the Drupal framework: *nix system tools (such as cron, drush, df, awk, bash, and other multimedia manipulation tools, such as FFmpeg), database server (such as MariaDB), scripting languages (such as PHP and Javascript), and a set of contributed libraries that are used for data manipulation, management, and security purposes.

Given the level of complexity of the system as a whole, we recommend PECE users to rely on Drupal managed hosting services offered by web hosting companies. This option is recommended to PECE administrators who are not experienced in *nix system administration. For experienced administrators, we suggest contracting a virtual private server (VPS) that matches the size and the needs of your project as described below.

In order to provide PECE administrators with data on monthly usage for calculating costs, PECE relies on basic descriptive statistics that are generated by the Drupal core module “statistics” as well as information about disk usage that is gathered in the back-end at every `cron` run. This information is very useful when estimating data transfers and calculating incurring hosting costs. Fully automated gathering and reporting of the usage of computational resources (such as CPU time, IO, individual artifact sizes) is a functionality that is being planned for the version 2.0 of the platform. It is not currently supported on PECE itself, since this information can be easily obtained on a monthly basis when contracting a Drupal managed hosting company. Please observe that this is one of the benefits of having a managed *versus* an unmanaged host: the ability to obtain fine-grained information on usage of the platform and not having to dedicate considerable time administering it (since the administrative tasks are performed regularly by the hosting company).

For calculating the cost of running and maintaining a PECE instance, we collected estimates from more than twelve web companies that are specialized in Drupal and described their services along three tiers (small, medium, large) and four of the most important variables for assessing hosting costs: number of authenticated users, disk consumption for both the file system and the database (in GB or TB), system memory (in GB), data transfer allowance (in GB or TB) and available bandwidth (in Gbps), and vCPU (per number of allocated virtual CPU core units) as demonstrated in the table below:

Instance	Users	vCPU	Disk	RAM	Data Allowance
Small	10	2	10 GB	2 GB	100 GB
Medium	50	4	100 GB	4 GB	1 TB
Large	100	8	1 TB	8 GB	10 TB

These numbers represent a rough estimate of the recommended specs for the virtual private host (VPS) in cloud services of hosting companies. It is important to bear in mind that these numbers can be quite different depending on the nature of the data that is hosted on PECE: audio and video files, for example, would create a different need in respect to the usage of disk, disk I/O, and RAM with substantial increase in the data transfer, therefore creating the need for bigger transfer allowances, if not for dedicated hosts and content delivery networks

(according to the geographical distribution of users in a particular research collaboration).

Another important factor to take into consideration is the number of published artifacts on the platform, which impacts overall performance and determines the need for more or less computational resources, making it difficult to estimate with precision. This estimation was informed by the web hosting market research conducted by the PECE team throughout the summer of 2015.

For estimating with more precision, PECE automatically generates access reports for individual artifact pages on a monthly basis (and comes with built-in modules to assess database and storage usage). Please observe that the total bandwidth usage is monitored and generated monthly by your hosting service provider, represented in the example image below, and notified to the administrator by email for managed and unmanaged VPS instances:



Coupled with basic statistics, PECE comes pre-configured with the “diskfree” module to run the `df` command and obtain information on disk usage for a particular instance. Whenever the disk usage reaches 90%, the administrator is informed by email that the disk is almost full:

	Usage on /	29% in-use; 4.3G free
	Usage on /	29% in-use; 4.3G free
	Usage on /dev	0% in-use; 254M free
	Usage on /run	1% in-use; 51M free
	Usage on /run/lock	0% in-use; 5.3M free
	Usage on /var/gandi	0% in-use; 25k free

The general orientation for administrators obtaining the PECE distribution via release package file, public repository, or one of our pre-configured virtual machine images is to dedicate one or more instances per project, that is, if a new project is created out of an ongoing project, it is recommended for one or more PECE instances to be created in addition. Using the PECE Open API, it is possible for users and administrators to share and harvest data from different PECE instances. Another important suggestion is for PECE administrators to rely on Drupal managed hosting companies in order to use their backup and system usage reporting capabilities. These services are important for redundant backup purposes as described in this document on the “backup” section.

Information for developers: the core module “statistics” must be enabled and configured as such: “Enable access log”, “Discard access logs older than 3 months and 3 weeks”, “Count content views” and “Use Ajax to increment the counter” must be marked ON. The extension module “statistics_counter” (7.x-1.4) has to be installed in order to provide more detailed data access information. The disk usage report functionality relies on the “diskfree” module (7.x-1.3). Diskfree relies on the following *nix programs: `df`, `grep`, and `awk`. They must be available on the server `$PATH` to be used by the PHP module. The fine-grained database usage report depends on the module “dba” which is explained in the next section of this document.