12 Most Active Modules in the OpenStack Nova Project

Fatemeh Maddahzadeh

This project aims to identify the 12 most active modules in a GitHub repository by processing the commits. The number of churns, number of commits, and time of commits have been collected per module within the nova subdirectory. We can identify the 12 most active modules based on these collected data.

At the first step, all commits from the last six months within the nova subdirectory are retrieved from GitHub. Only 22 modules have commits within the nova subdirectory in the last six months. Data related to these modules is shown in detail in the figures below. Next, data related to each module is aggregated to the module level. Also, in this project, a caching mechanism is used to increase the speed of the test process. The retrieved data from GitHub is written into a file rather than sending many requests to GitHub each time the code runs.

According to figure 1 and figure 2, we identify that the 12 most active modules within the nova subdirectory are: **tests, db, objects, virt, api, compute, scheduler, network, cmd, conf, policies, conductor**.

In the future, to better identify the most active modules, we can include the commit time parameter in our measurement to identify which modules have been updated constantly. For example, in a five-month study period, there is a difference between the activity of a module with 100 commits in the first month and no more commits in the following four months compared to a module with 20 commits in each month. The commit time is also collected in the code's final report for this issue.

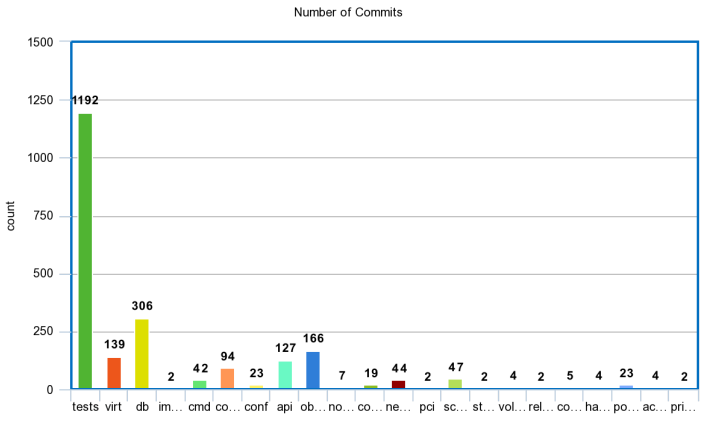

Figure 1 - Number of Commits per Module

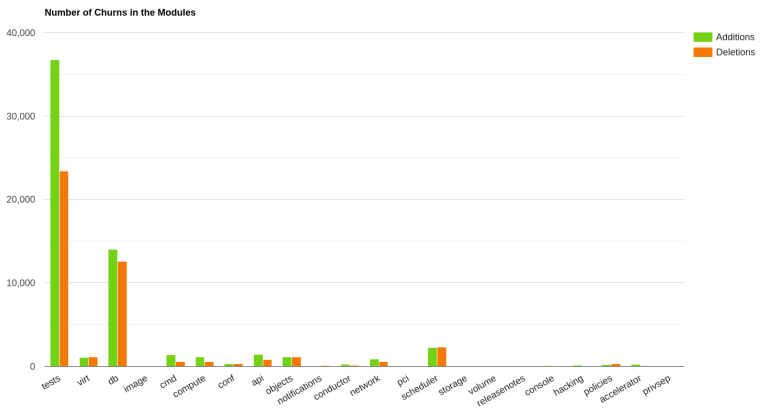

Figure 2 - Number of Churns per Module