Report on "Finding the most actively modified modules of the OpenStack Nova project"

Madhusudan Basak

November 15, 2021

1 Result

Considering the studied period of last six months (180 days), the following result has been obtained:

- **Total number of commits during the studied period:** 2411
- **Total number of churns during the studied period:** 89948
- **12 most actively modified modules (considering commit count)**: tests (1400), db (175), objects (170), virt (150), api (128), compute (95), notes (58), scheduler (48), network (47), cmd (39), conf (28) and policies (24).
- **12 most actively modified modules (considering churn count)**: tests (50049), db (20199), scheduler (4583), objects (2274), api (2265), compute (2212), virt (2200), cmd (1902), network (1423), notes (670), policies (566) and conf (507).

2 Summary Observations

Following are the key observations

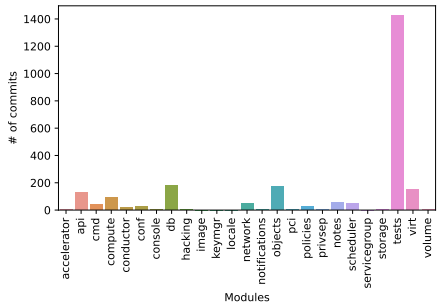

Figure 1: Bar plot for module-wise commit

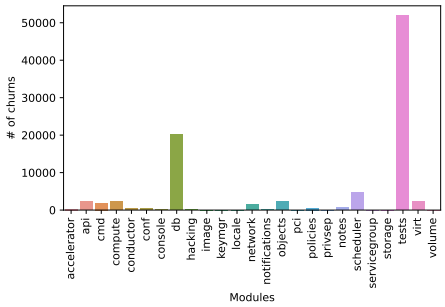

Figure 2: Bar plot for module-wise churn

- **Commit and churn count highly vary among the modules:** From the study of the last six months, it is evident that commits do not occur uniformly for the modules. During the studied period, the most number of commits (1400 commits) occurred for "tests" module, whereas no commit happened for some other modules (e.g., "accelerator" module). The high imbalance of commit count can also be found from median (19) and median absolute deviation (93.28) values. From Figure **??**, it is also visible that the number of commits for one module ("tests") is significantly higher than other modules. The same observation holds for module-wise churn count (Figure 2) too, where the median and median absolute deviation values are 378 and 3525.2, respectively.

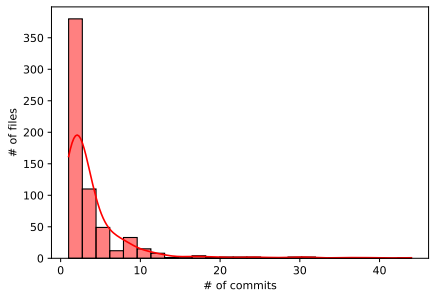

Figure 3: Histogram for file level commit distribution

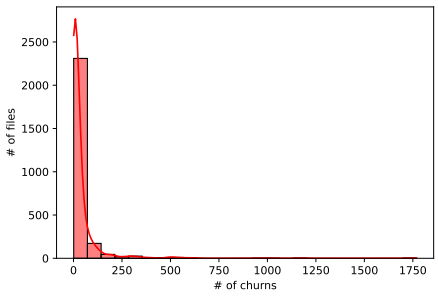

Figure 4: Histogram for file level churn distribution

- **Skewed distribution of commit and churn count (file level)**: If we plot the histogram of the commits (and churns) for all the files under the modules, we find a left skewed distribution of commits (and churns), i.e., most of the files have rarely been modified during the studied period. The file-level commit and churn distributions are represented in Figure 3 and Figure 4 respectively.