

Alireza Tavakkoli  
Coding task - Nova Project  
November 14, 2021

# Identifying 12 most modified modules in OpenStack Nova

## My report

Github Address.

[https://github.com/sinatav/msr\\_nova](https://github.com/sinatav/msr_nova)

Suggestion.

I suggest using .ipynb file to see the results and take a look at the code, but using the code is okay too, the instructions on how to run the code are in the README.md file.

Tool.

I used Github API to complete this task. The libraries that I used are also listed in the requirements.txt file.

Code.

For extracting data from the repository we need to make a request and we will need GitHub username and token. After setting them we will make the connection to crawl all the commits in the repository. This can take a while since there are about 60k commits within the project. Then we will extract the date (last 6 months) and exclude all other commits. There will be about 600 commits left. The commit hashes in this new data frame will be used to generate urls and with them we will make another connection to commits information themselves and from them, we will understand every files and number of codes altered in this commit. Then by just counting number of changes in each file (lines of code added and deleted) and how many times a file was changed during that period, we will find 12 most modified modules in this project.

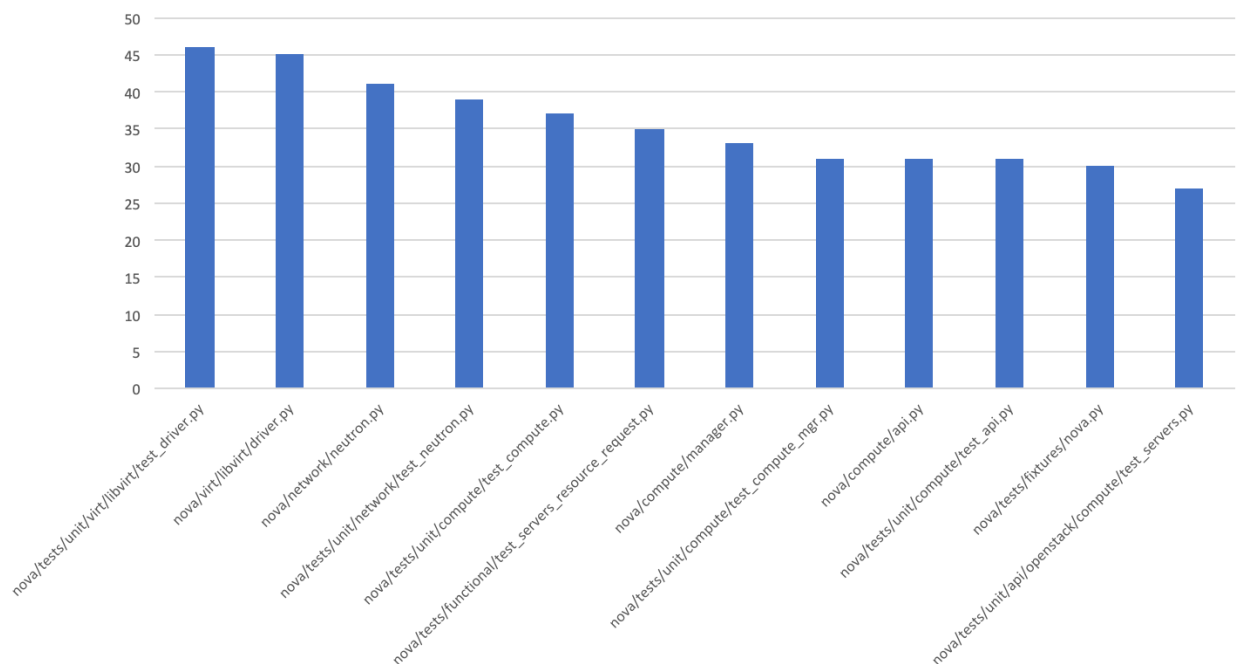
Results.

## Number of commits.

12 most modified modules indicated by how many commits occurred that involved this file, are as bellow:

```
{'nova/tests/unit/virt/libvirt/test_driver.py': 46,  
 'nova/virt/libvirt/driver.py': 45,  
 'nova/network/neutron.py': 41,  
 'nova/tests/unit/network/test_neutron.py': 39,  
 'nova/tests/unit/compute/test_compute.py': 37,  
 'nova/tests/functional/test_servers_resource_request.py': 35,  
 'nova/compute/manager.py': 33,  
 'nova/tests/unit/compute/test_compute_mgr.py': 31,  
 'nova/compute/api.py': 31,  
 'nova/tests/unit/compute/test_api.py': 31,  
 'nova/tests/fixtures/nova.py': 30,  
 'nova/tests/unit/api/openstack/compute/test_servers.py': 27}
```

The number in front of each module indicates the number of commits. This bar graph is also showing the same result.

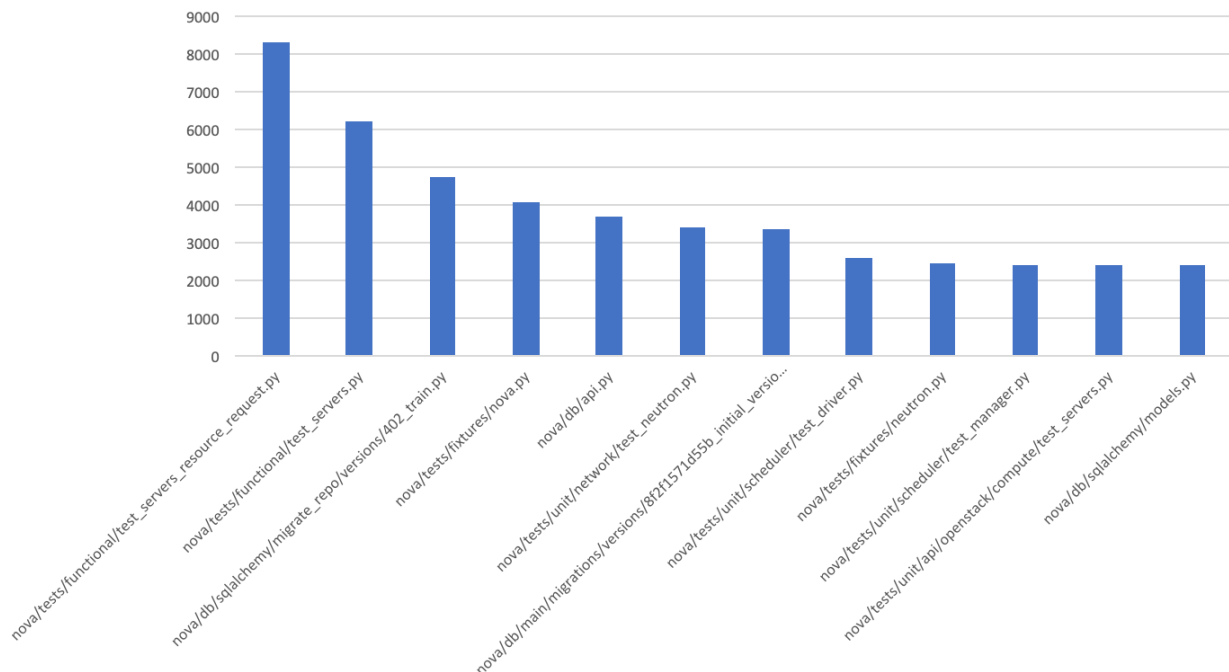


## Sum of added and removed lines.

12 most modified modules indicated by how much churn occurred during this period, are as bellow:

```
{'nova/tests/functional/test_servers_resource_request.py': 8302,  
'nova/tests/functional/test_servers.py': 6200,  
'nova/db/sqlalchemy/migrate_repo/versions/402_train.py': 4713,  
'nova/tests/fixtures/nova.py': 4068,  
'nova/db/api.py': 3678,  
'nova/tests/unit/network/test_neutron.py': 3397,  
'nova/db/main/migrations/versions/8f2f1571d55b_initial_version.py': 3358,  
'nova/tests/unit/scheduler/test_driver.py': 2588,  
'nova/tests/fixtures/neutron.py': 2454,  
'nova/tests/unit/scheduler/test_manager.py': 2382,  
'nova/tests/unit/api/openstack/compute/test_servers.py': 2381,  
'nova/db/sqlalchemy/models.py': 2372}
```

The number in front of each module shows the number of lines added and removed. This bar graph is also showing the same result.



I suggest checking the notebook I wrote for more thorough explanations on the code. Thank you so much for considering me.