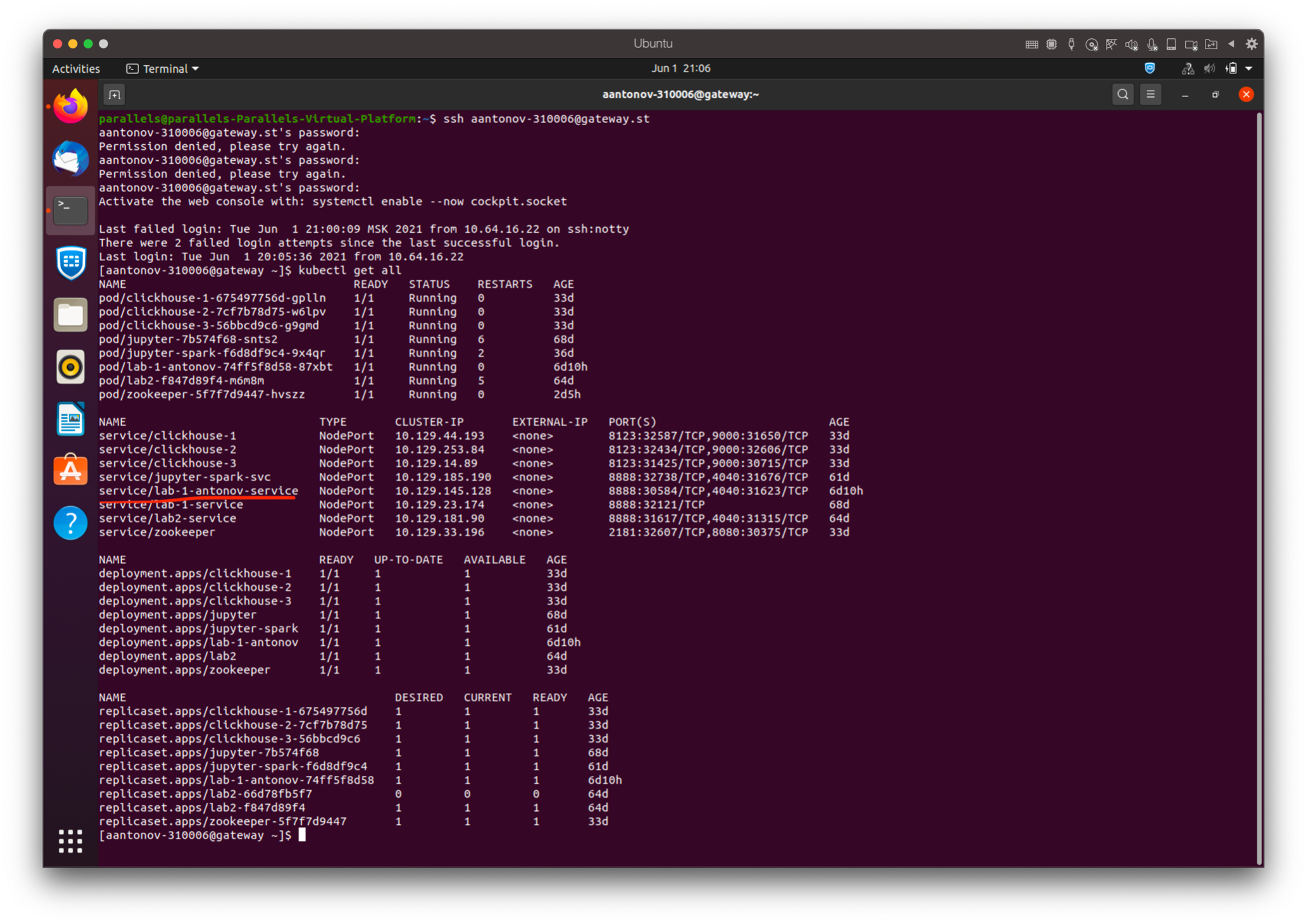
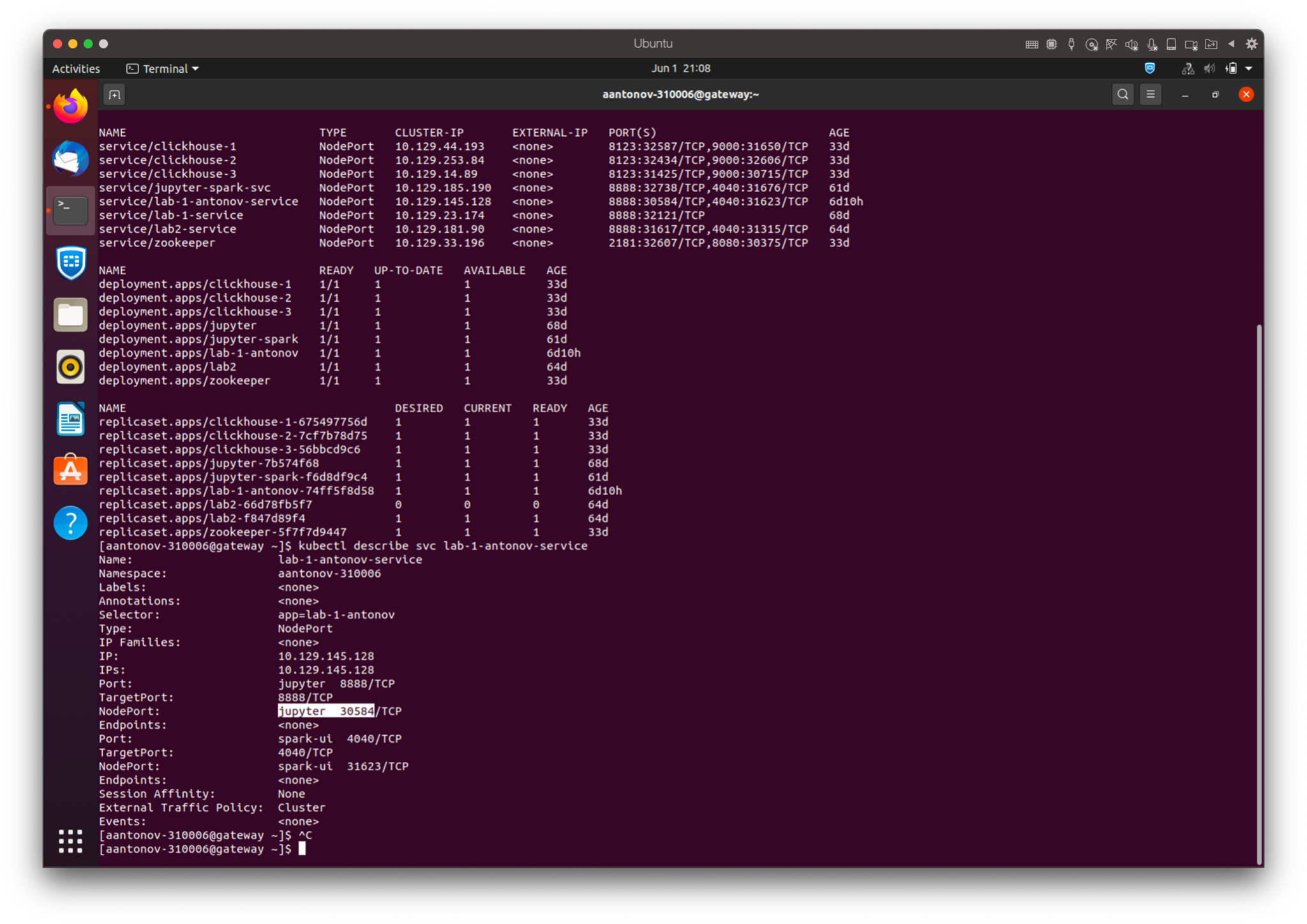
Файл x\_lab1 мой, LAB1 ‒ Андрея Антонова

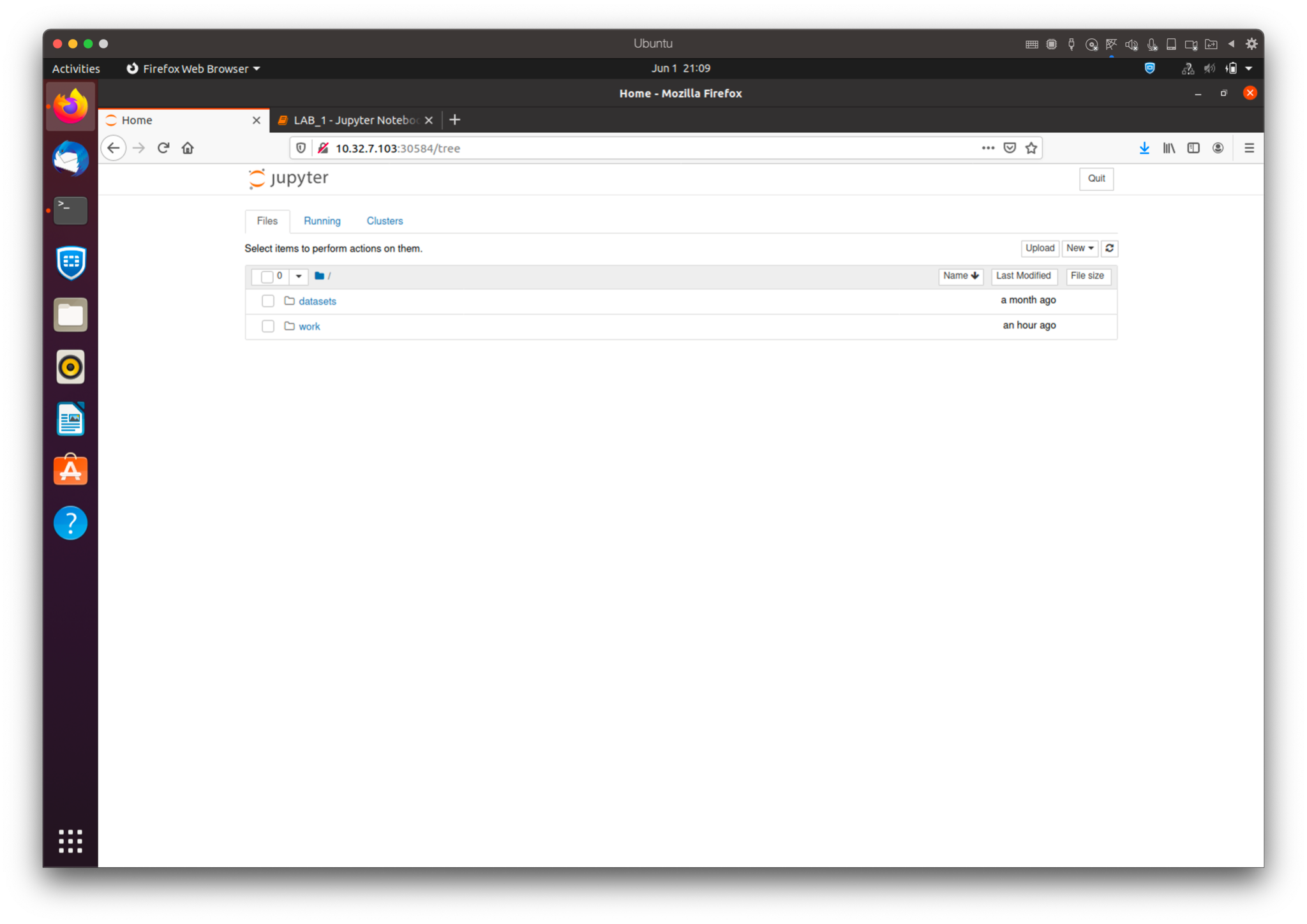
1. Запустить VPN
2. $ ssh <username>@gateway.st
3. $ nano pyspark-template.yaml (если у вас нет в домашней директории файла, то он есть в этой папке в гугле. Я не знаю где его можно взять на кластере, так что если что просто загрузите с компьютера туда)
4. В файле поменять мое имя и номер на свои (смотрите в .yaml в этой гугл папке)
5. Сохраните файл
6. $ kubectl create -f pyspark-template.yaml
7. $ kubectl get all
8. В сервисах должен появится ваш новый сервис (его имя вы задавали в файле .yaml, у меня это “lab-1-antonov-service”)



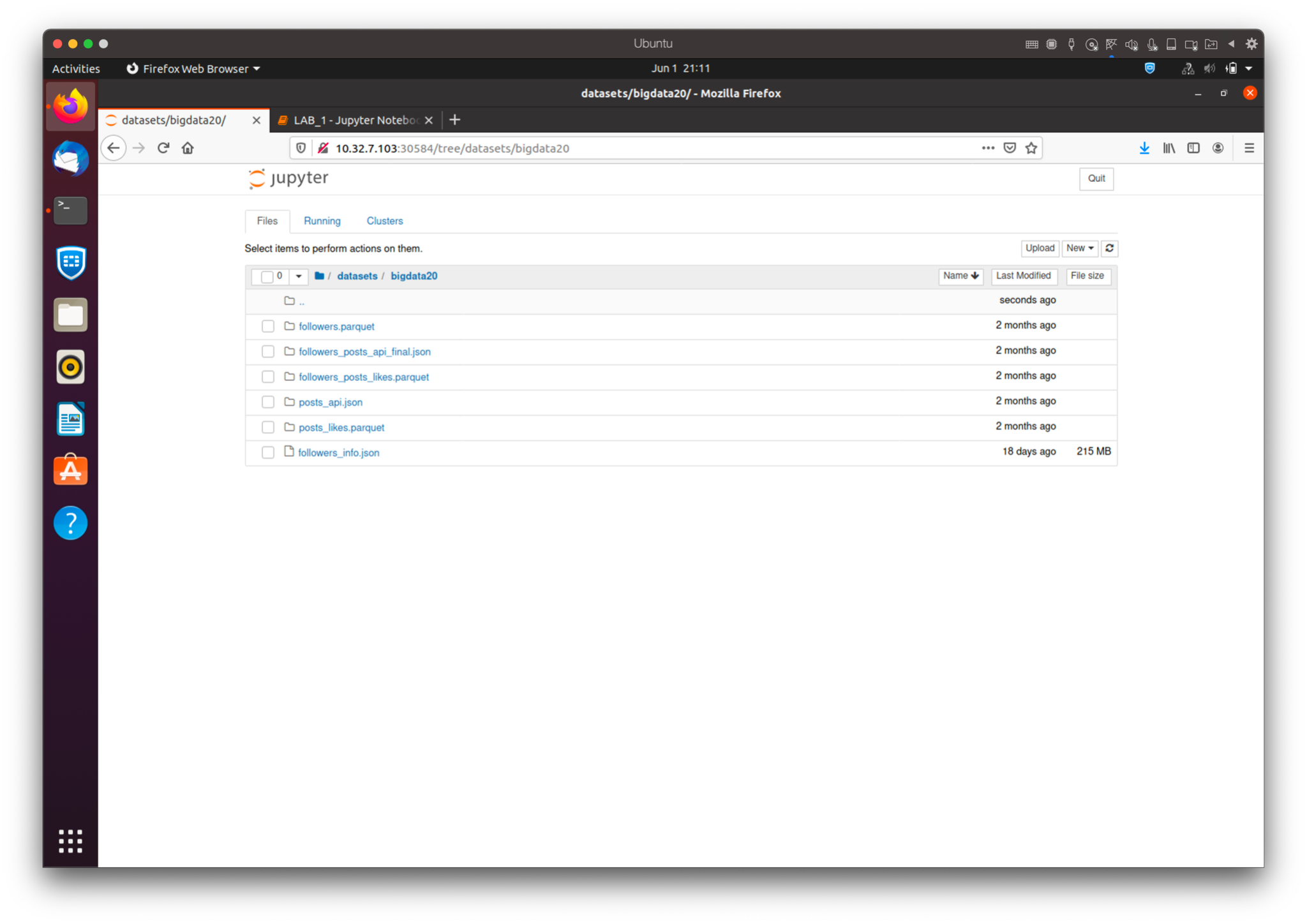
1. $ kubectl describe svc <название сервиса>



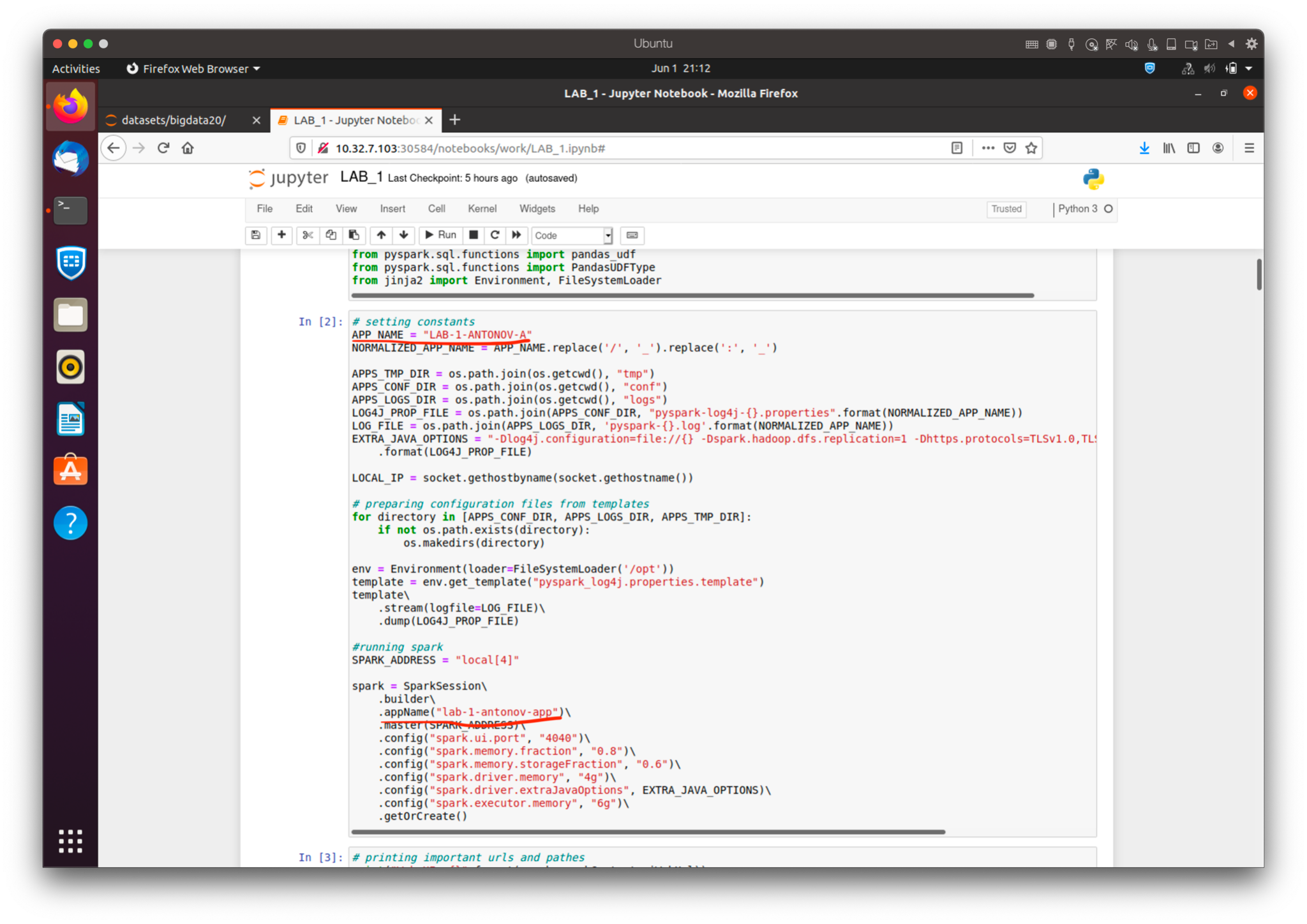
1. Находим Jupyter node port, идем по адресу 10.32.7.103:<порт>. Должен открыться ваш юпитер



1. В папку work кладем файл LAB\_1.ipynb из этой гугл папки
2. Убедитесь что датасеты на месте. У меня они лежат в папке /datasets/bigdata20, и выглядят вот так:



1. Запускайте юпитер ноутбук. В нем поменяйте переменную APP\_NAME и SparkSession.appName(<имя>)



1. Проверьте в ноутбуке путь к датасетам. Чтение данных происходит в Step 0
2. Запускайте ноутбук, все должно выполнится
3. Что нибудь поменяйте, чтобы не выглядело как списывание