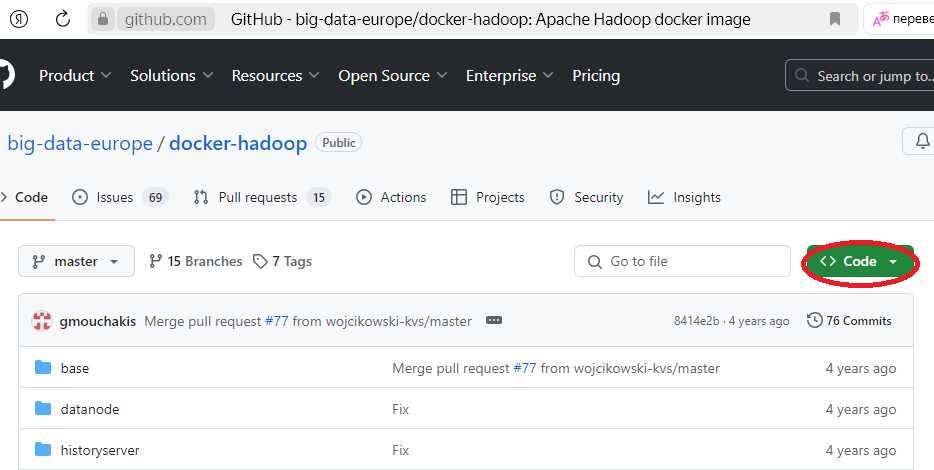
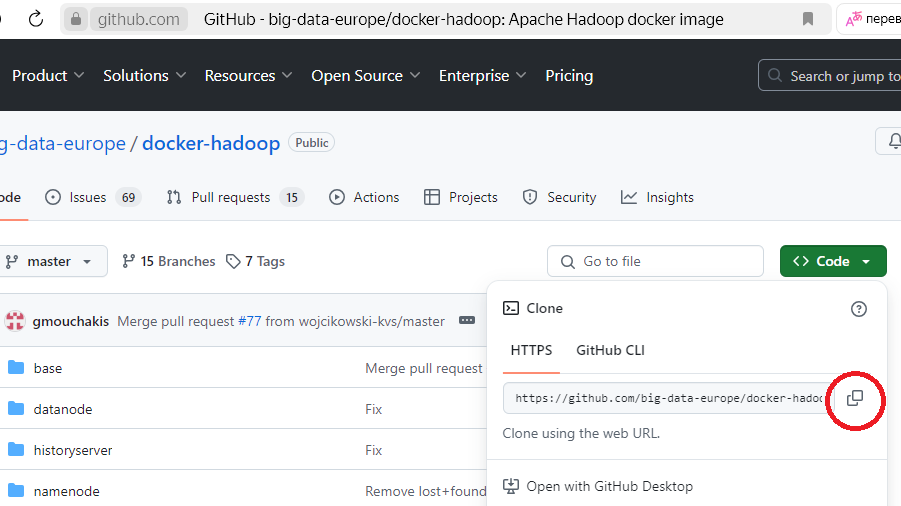
Установка Hadoop в Ubuntu

Разворачивать Hadoop и др. будем в контейнерах используя docker-compose.

https://github.com/WtCrow/docker-hadoop-spark-hive2-jupyter

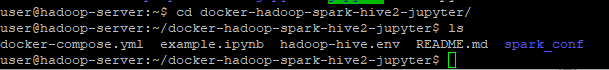




Клонируем репозиторий на Ubuntu server:

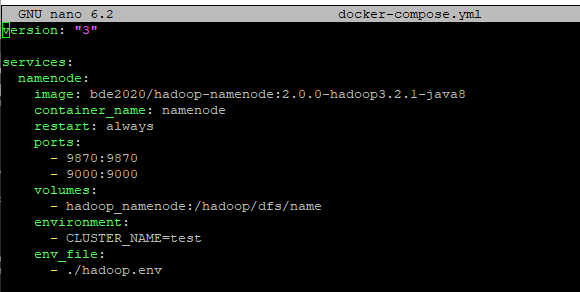


Переходим в скачанную папку:

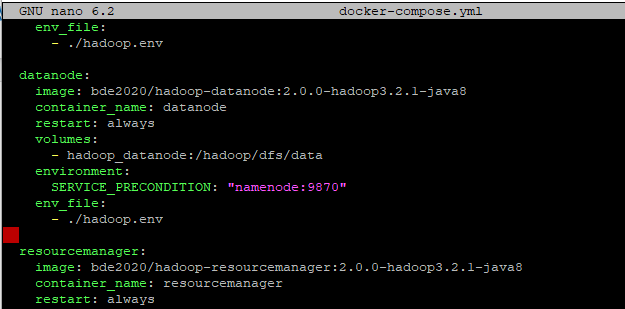


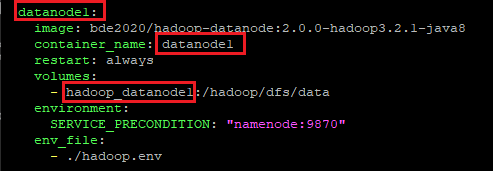
В папке находятся директории для каждого компонента Hadoop и **docker-compose.yml** – файл настройки многоконтейнерного приложения, настроим его открыв в редакторе.

Настройки нейм-ноды не трогаем:



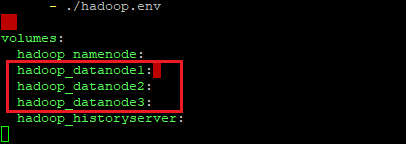
Дата-нода по умолчанию одна,

добавим еще две и дадим им уникальные имена и назначим свои volume:

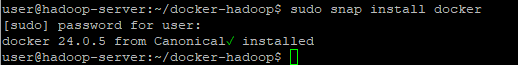


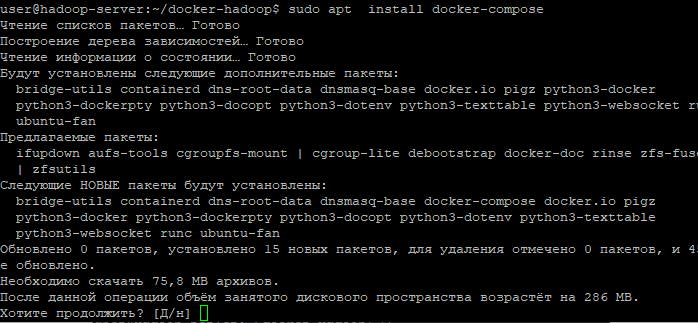


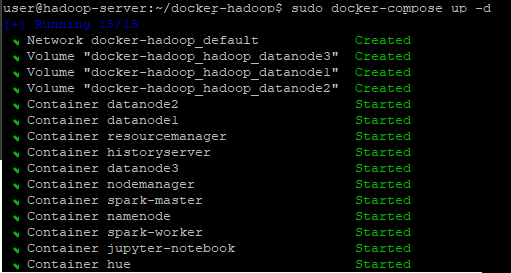
И добавим в конце файла наши volume:



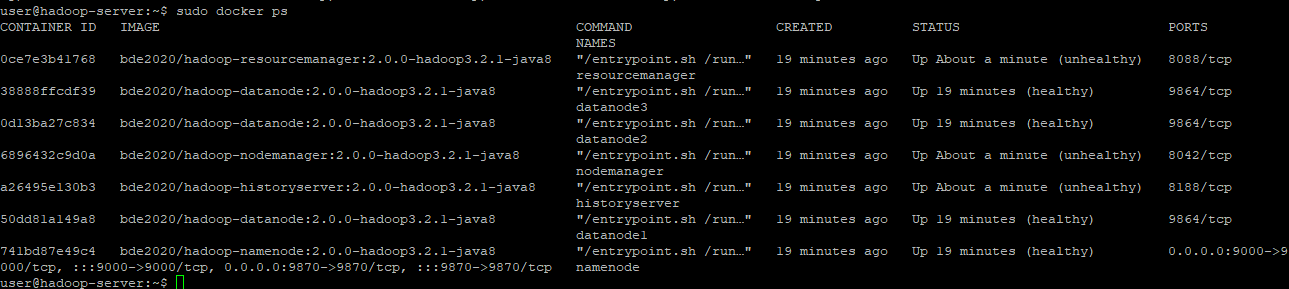
Сохраняем изменения в файле и стартуем, предварительно установив docker и docker-compose:

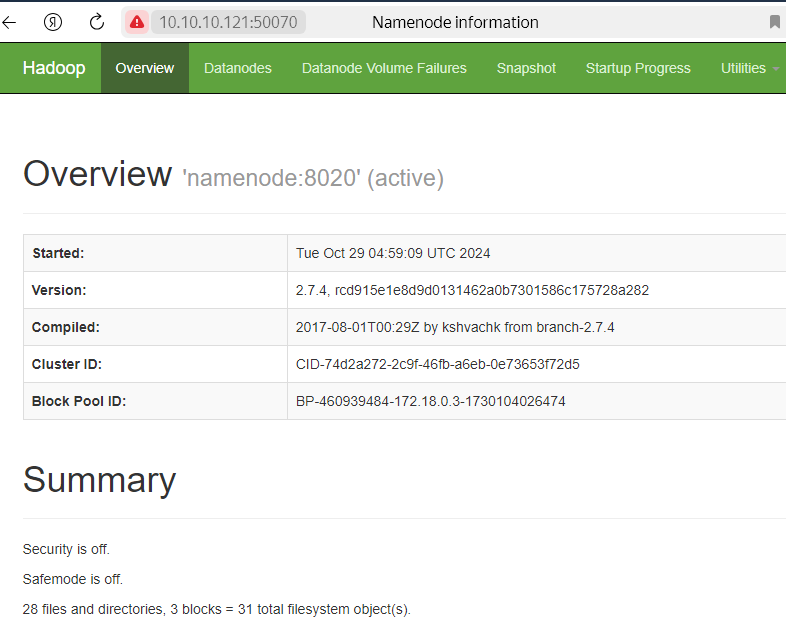


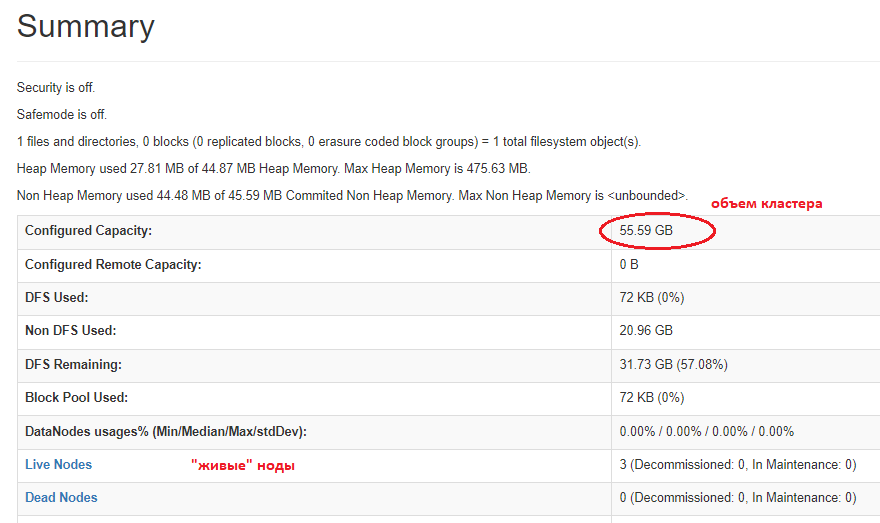




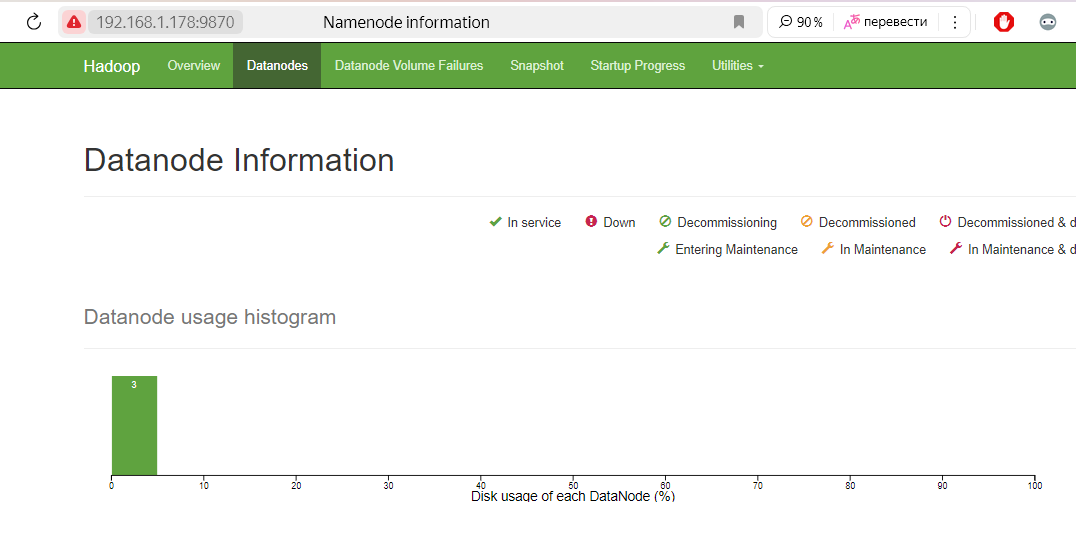
Дожидаемся запуска всех контейнеров и открываем в браузере страницу админки Hadoop:

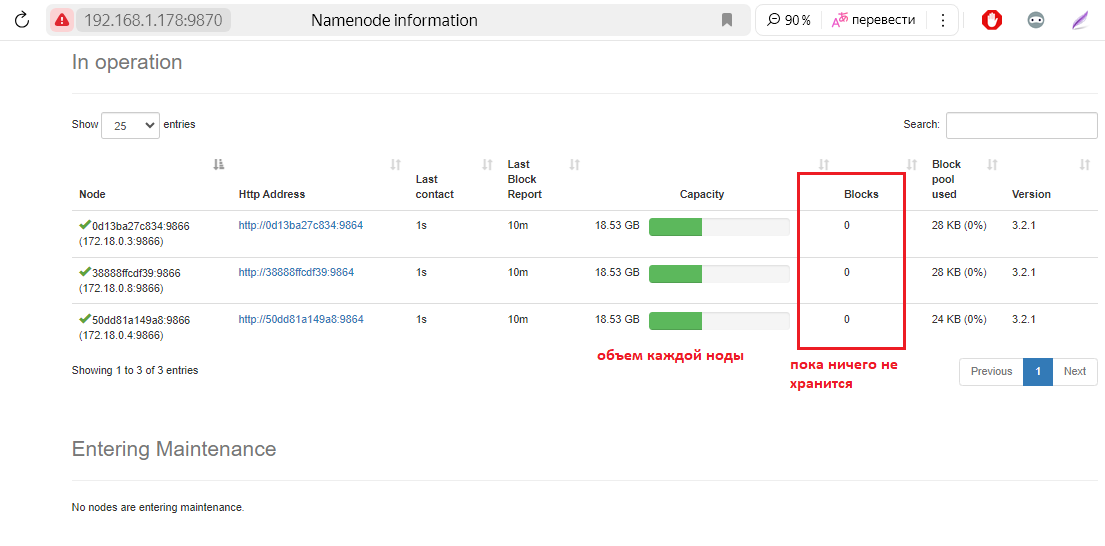






Во вкладке Datanodes, можно посмотреть на ноды:

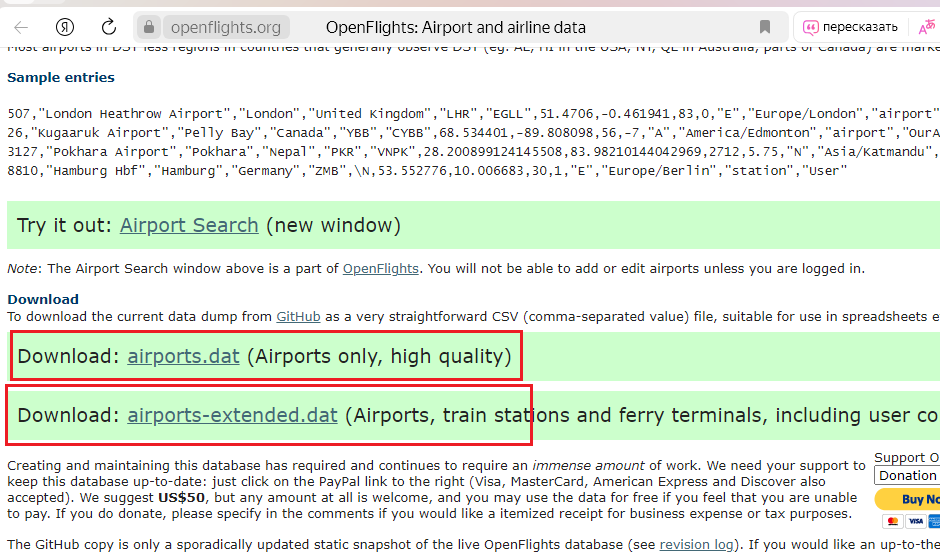




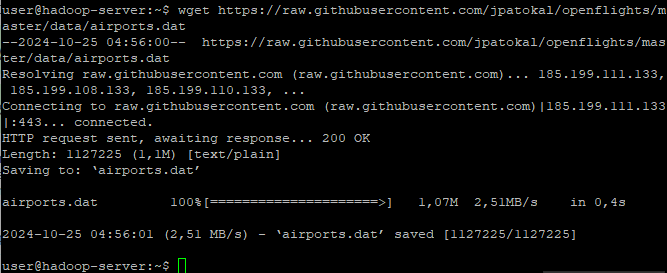
Теперь загрузим данные, например, откроем в браузере <https://openflights.org/data>

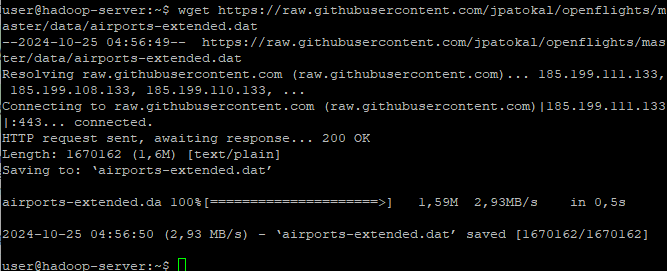


Копируем ссылки:

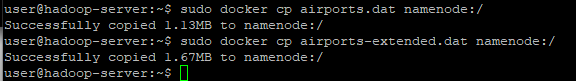


И скачиваем на сервер:

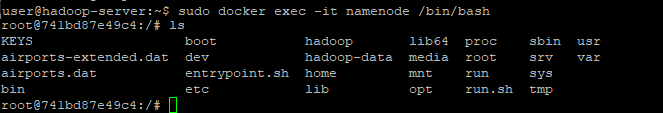




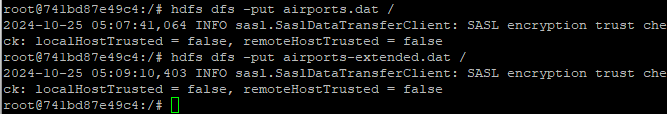
Затем копируем файлы в контейнер:



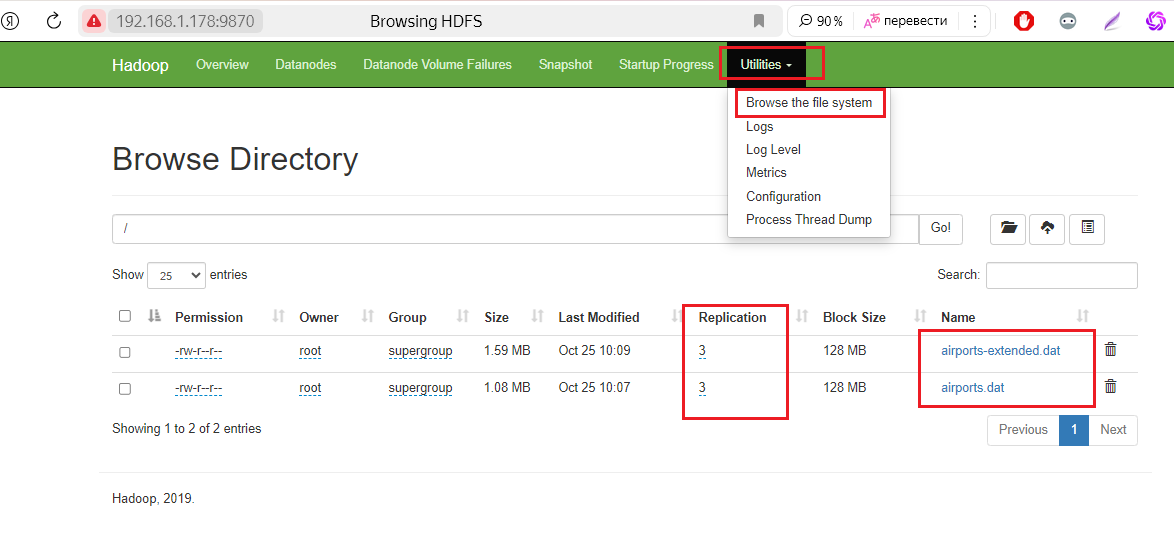
Проверяем, файлы внутри контейнера:



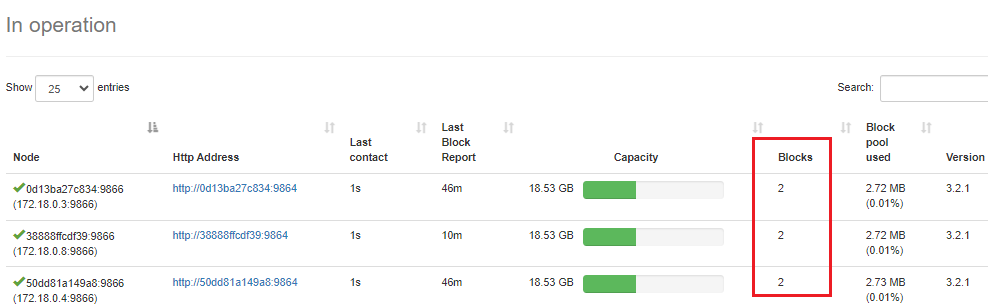
Затем кладем файлы в HDFS:



Проверяем в админке:

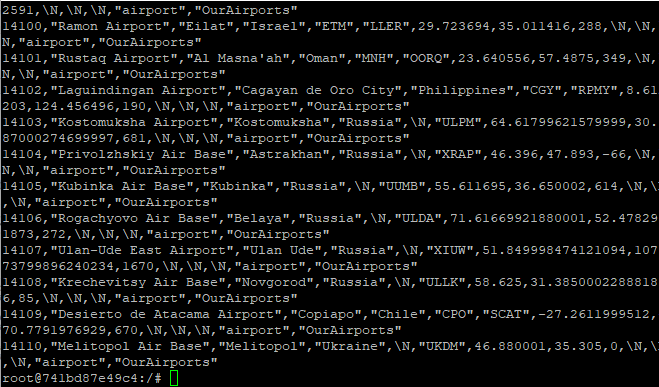


Фактор репликации 3, значит файл должен быть на всех трех нодах:



Можно прочитать содержимое файла:





Установка и настройка распределенной файловой системы завершена.