# **Инструкция по развертыванию виртуальной машины с Hadoop**

1. **Установка виртуальной машины с** **Hadoop от Cloudera**

Скачать образ виртуалки можно по ссылкам:

<https://drive.google.com/open?id=1PLI_s0147TDN3GkGH9t6BaPkFuu8U8mv>

<https://drive.google.com/open?id=18NTLRa6nEJt3Snf8-g2fpEARA5jIzPeG>

<https://yadi.sk/d/f6IYF8zoWOOLOw>

После окончания загрузки, необходимо разархивировать скачанный архив куда-нибудь и запомнить это место.

Запускаем VirtualBox. Далее нажимаем Файл->Импорт конфигураций. Должно открыться такое окно (рис. 4):

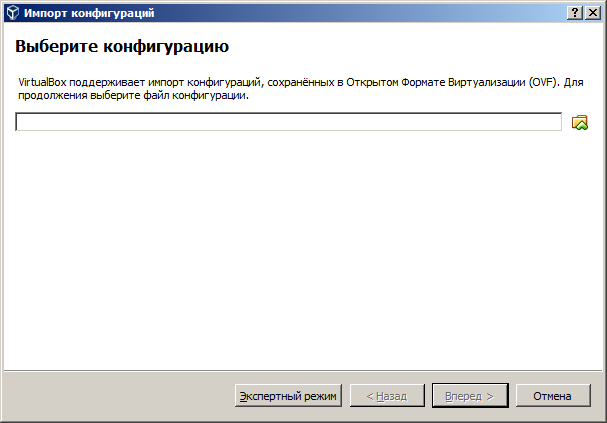


Рис. 1.

Далее нажимаем на значок папки (желтый с зеленой стрелкой) и выбираем папку в которую мы ранее разархивировали скачанный архив. В этой папке должен быть виден только один файл:

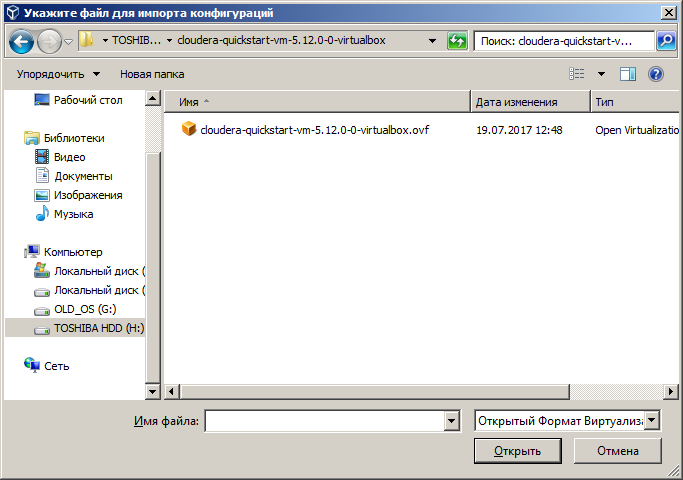


Рис.2.

Открываем этот файл. (Также это можно сделать непосредственно из графического интерфейса VirtualBox Файл->Импорт конфигураций)Нажимаем клавишу «Вперед»

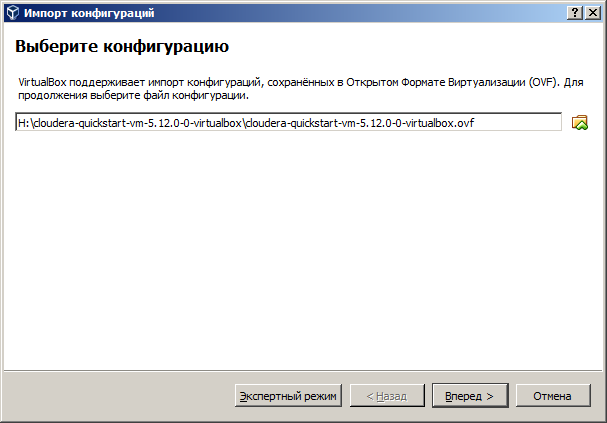


Рис.3.

Должно открыться окно (рис. 7):

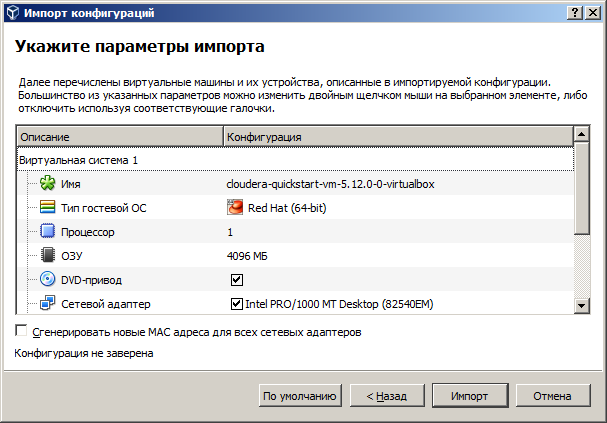


Рис.4.

По возможности необходимо установить наибольшее возможное значение ОЗУ, но не более ¾ всей доступной, тогда виртуальная машина будет работать быстрее.

Также рекомендуется сменить путь к виртуальному образу диска, так как он довольно объемный: 64 Гб, выберите тот диск, где у вас достаточно места. Для этого необходимо два раза щелкнуть по полю с путем к виртуальному образу диска (рис. 8), указать необходимую папку и дополнительно оставить название диска, то есть:

C:\Users\Kolychev\VirtualBox VMs\cloudera-quickstart-vm-5.12.0-0-virtualbox**\cloudera-quickstart-vm-5.12.0-0-virtualbox-disk1.vmdk**

То, что выделено жирным шрифтом, оставляется, остальное меняется на то, что вам удобно

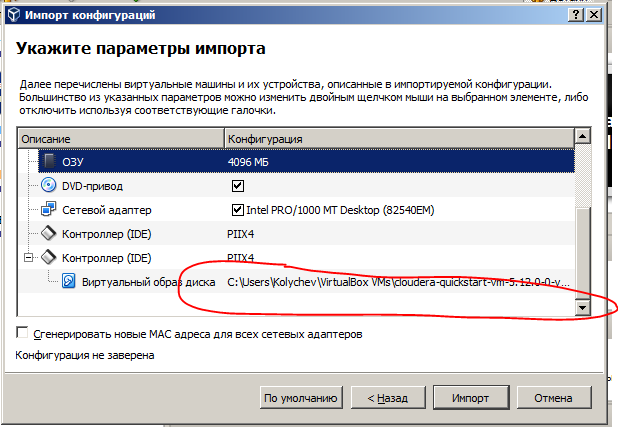


Рис.5.

После нажатия кнопки «Импорт» запустится импорт конфигурации (процесс займет около 10 минут):

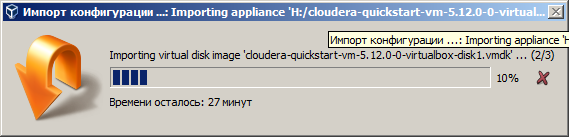


Рис.6.

После окончания виртуальная машина будет готова к использованию. Запускаться виртуальная машина будет довольно долго (5-10 минут), это абсолютно нормально, так как со стартом Linux запускаются все демоны Hadoop, Hive и других программных продуктов экосистемы, включенных в данный дистрибутив Hadoop.

Для удобства работы рекомендуется включить двунаправленный буфер, для того чтобы можно было копировать и вставлять из гостевой ОС в основную и наоборот: Устройства->Общий буфер обмена -> Двунаправленный

# **3**. **Работа с Hive**

После того как запуститься виртуальная машина и загрузится гостевая система, автоматически откроется браузер (Если этого не произойдет, запустите его самостоятельно). Для начала работы с Hive (рис. 10) необходимо нажать на закладку Hue, логин/пароль cloudera/cloudera

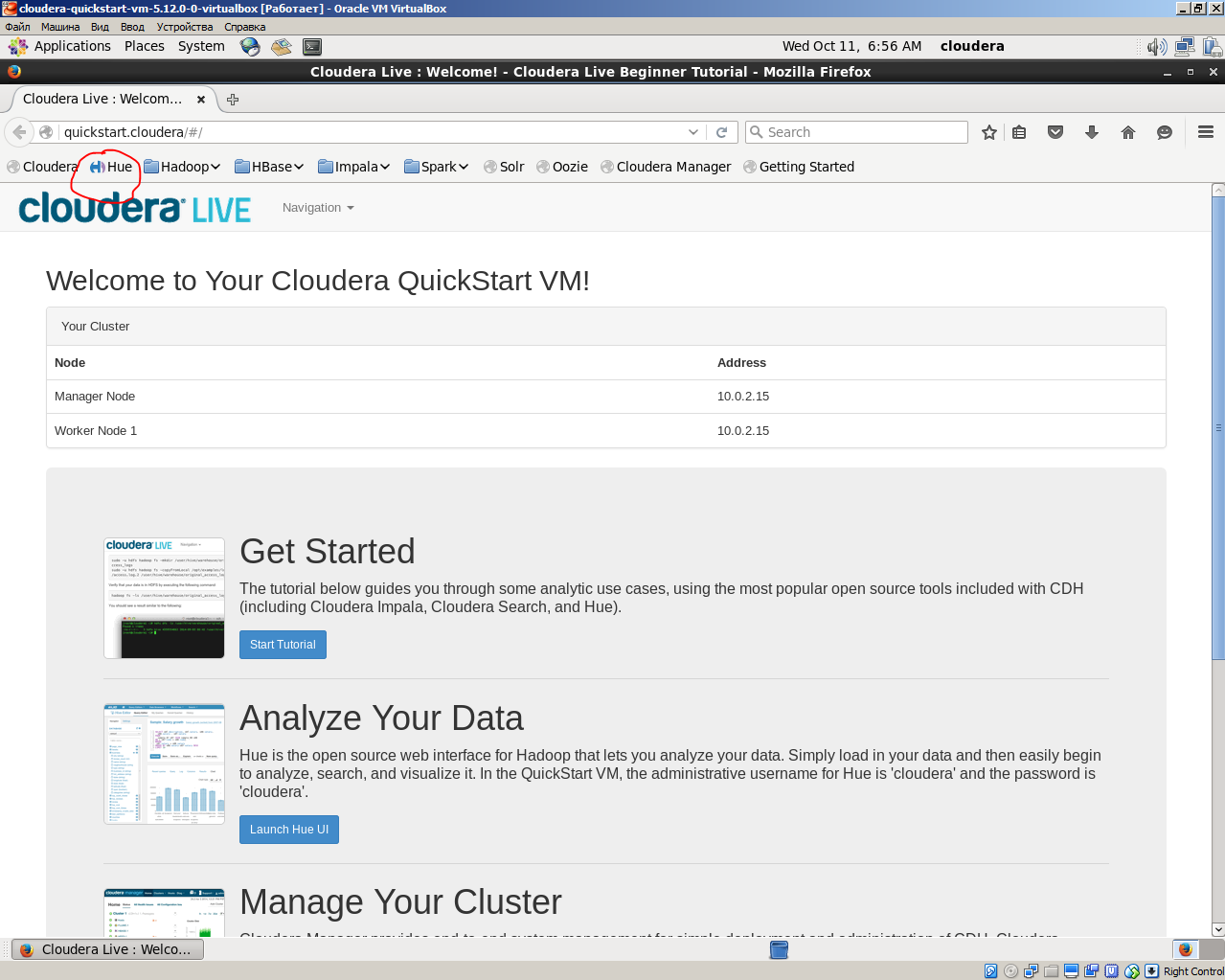


Рис.7.

Далее необходимо выбрать (рис. 11) Query Editor -> Hive и Вы у цели.

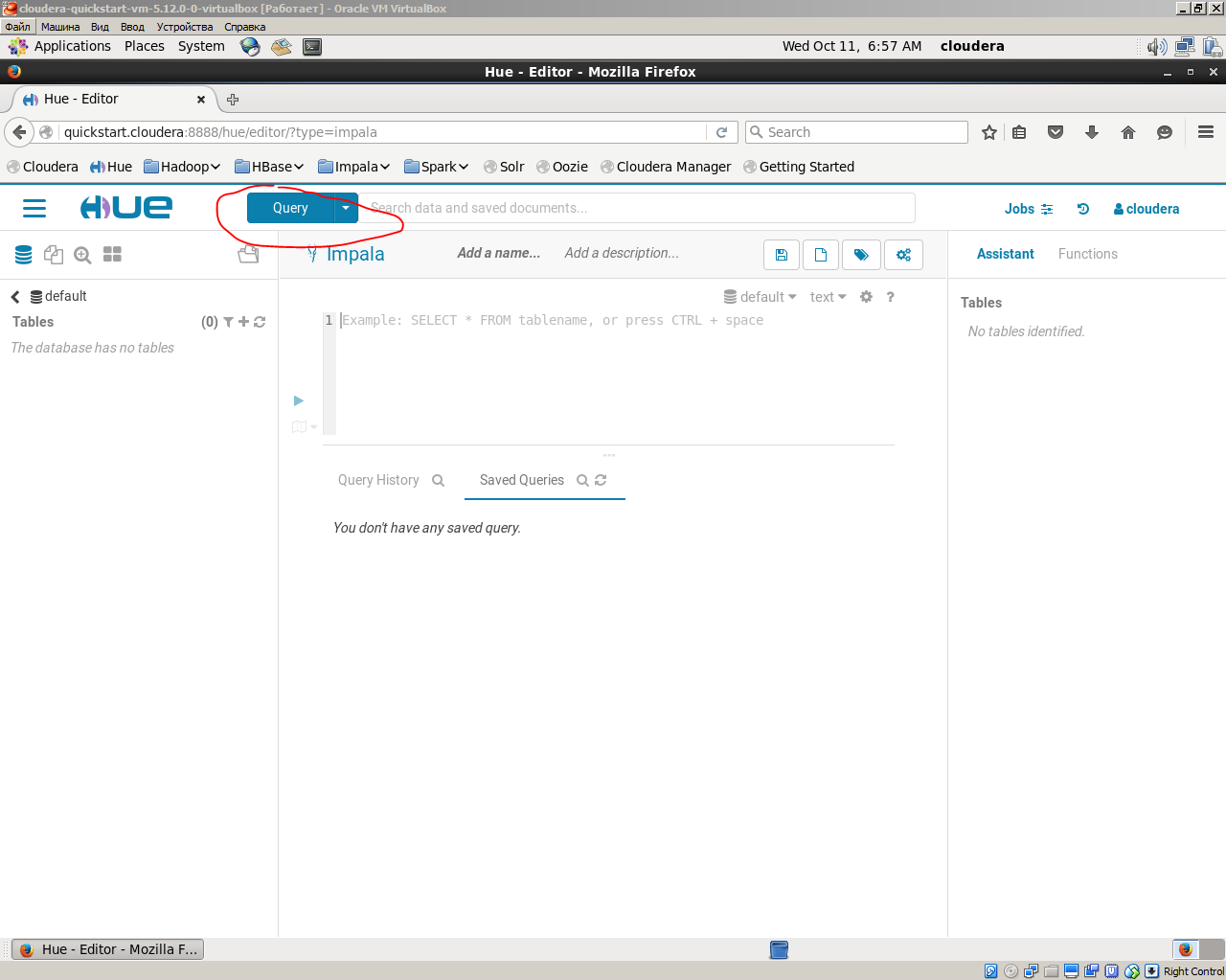


Рис.8.

Для того чтобы перенести какой-нибудь файл в хранилище Hive для начала работы с ним, необходимо создать директорию в **Hive** командой:

hadoop fs -mkdir /user/hive/warehouse/<имя директории>/

Для копирования файла из локальной файловой системы в хранилище Hive используется команда:

hadoop fs -copyFromLocal <путь к файлу в локальной файловой системе> <путь к каталогу созданному в Hive>

***Пример:***

hadoop fs -copyFromLocal /home/cloudera/Desktop/price /user/hive/warehouse/price/

Здесь файл расположенный на рабочем столе скопируется в каталог price в хранилище Hive.