…

Закрыть браузер.

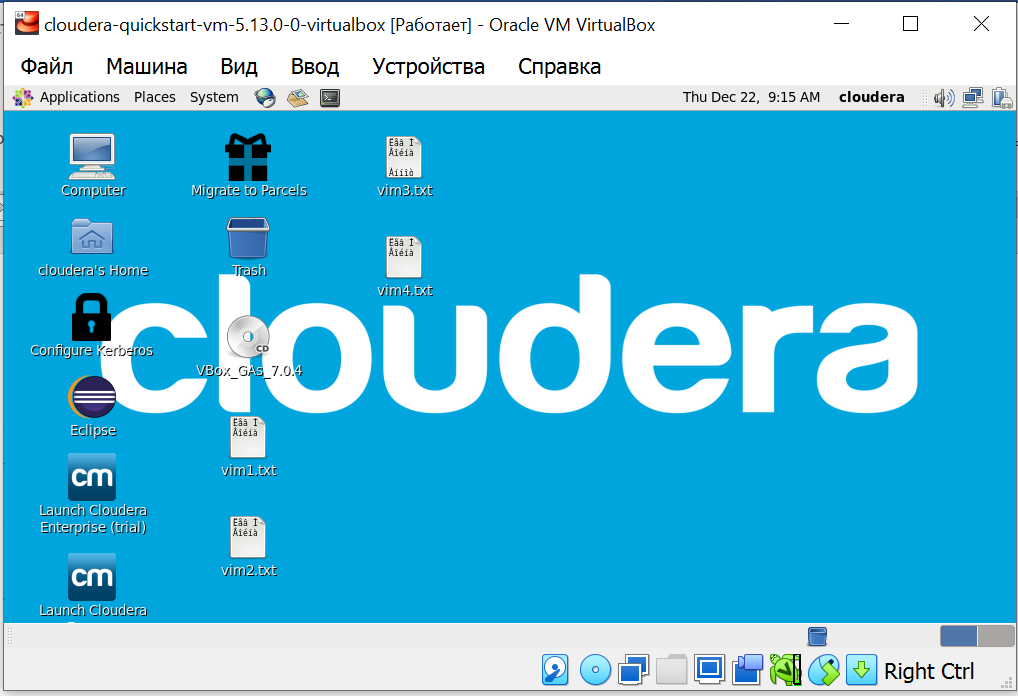
Устройства-Подключить образ диска дополнительной гостевой ОС.

Запустить виртуальный диск (пароль cloudera).

Перезагрузить ВМ.

Устройства-Ф-ция Drag and Drop – Двунаправленный.

Перетащить файлы с основной ОС в ВМ (на рабочий стол).



Есть HDFS и есть ОС той машины, на которой HDFS развернут. И они друг друга не видят. Они изолированы.

Есть кластер (несколько машин), на которых установлена файловая система Hadoop. У каждой из этих машин есть своя файловая система. Когда пользователь взаимодействует с Hadoop, он не взаимодействует с ним напрямую, он взаимодействует изначально с файловой системой той машины/сервера, на котором Hadoop развернут.

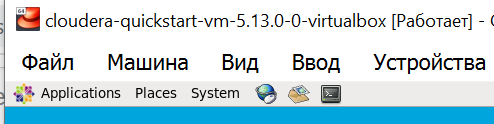
В данном случае сервером является ВМ.

У ВМ есть своя ОС.

**Файлы Война и мир мы перебросили в файловую систему ВМ**, НЕ в HDFS.

Чтобы перебросить файлы в HDFS, нужно открыть терминал и написать команду копирования файлов с указанием пути, где эти файлы находятся (чтобы посмотреть путь, в контекстном меню файлов выберите Свойства (Properties)):

hdfs dfs -copyFromLocal /home/cloudera/Desktop/vim1.txt /user/cloudera



Вызов терминала

Изображение выглядит как текст

Автоматически созданное описание

Просмотр пути файла vim1.txt

Изображение выглядит как текст

Автоматически созданное описание

Содержимое папки cloudera в hue

Перетащите все 4 тома Войны и мир в HDFS.

Изображение выглядит как стол

Автоматически созданное описание

Для вывода содержимого папки введите в терминале команду:

hdfs dfs -ls

Изображение выглядит как стол

Автоматически созданное описание

Для изменения прав по отношению к файлу введите в терминале команду:

hdfs dfs -chmod 764 /user/cloudera/vim1.txt

При просмотре содержимого папки увидим, что права для файла vim1 изменились на -rwxrw-r--.

Изображение выглядит как текст, стол

Автоматически созданное описание

Для вывести на экран информации о том, сколько места на диске занимает файл введите в терминале команду:

hdfs dfs -du vim1.txt (для просмотра размера файла в байтах)

hdfs dfs -du -h vim1.txt (для просмотра размера файла в килобайтах)

Изображение выглядит как текст

Автоматически созданное описание

Одна из сильных сторон Hadoop – это отказоустойчивость. При выводе из строя одной/нескольких машин данные сохраняются за счет репликации.

В данном случае фактор репликации = 1, т.к. размер файла одинаков в обоих отображениях.

Для изменения фактора репликации на 2 введите в терминале команду:

hdfs dfs -setrep -R 2 /user/cloudera/vim1.txt

Теперь он будет отличаться в 2 раза:

hdfs dfs -du -h vim1.txt

Изображение выглядит как текст

Автоматически созданное описание

Для подсчета количества строк в файле введите в терминале команду:

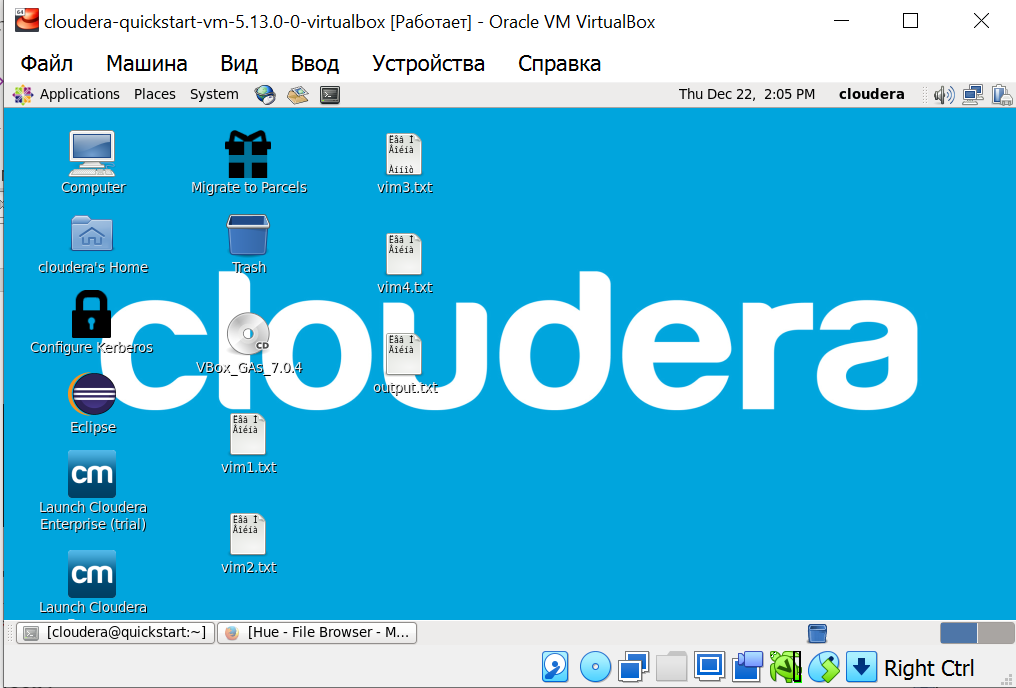
hdfs dfs -cat /user/cloudera/vim1.txt | wc -l



Для слияния 4-х файлов в один введите в терминале команду:

hdfs dfs -getmerge -nl /user/cloudera/vim1.txt /user/cloudera/vim2.txt /user/cloudera/vim3.txt /user/cloudera/vim4.txt /home/cloudera/Desktop/output.txt

Новый объединенный файл создастся в локальной файловой системе (на рабочем столе ВМ):



Для переноса объединенного файла из локальной файловой системы в hdfs введите в терминале команду:

hdfs dfs -copyFromLocal /home/cloudera/Desktop/output.txt /user/cloudera

Изображение выглядит как текст

Автоматически созданное описание