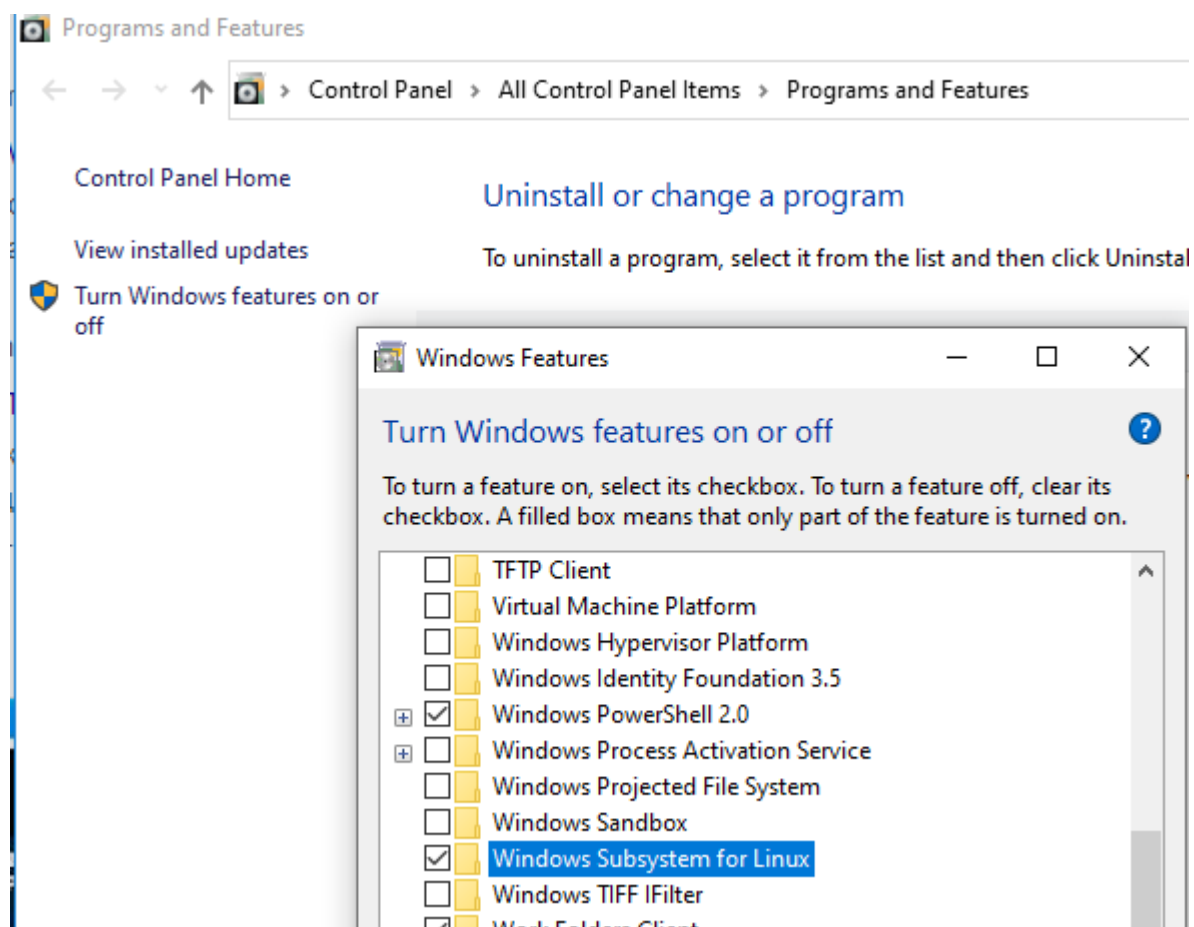


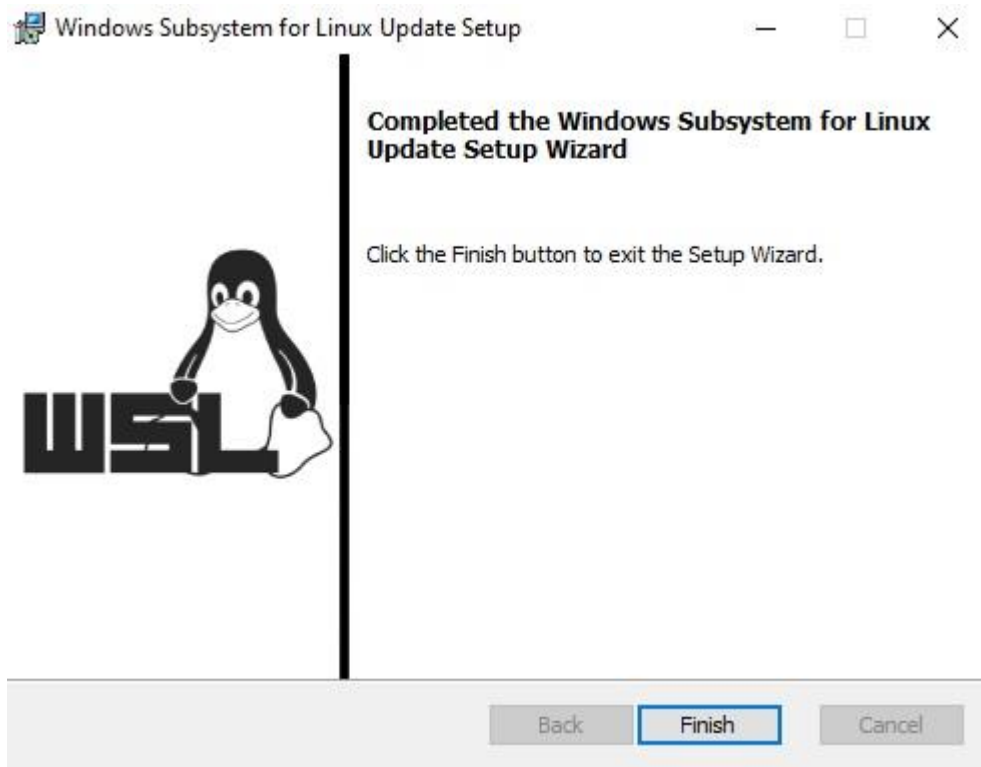
Шаг 1: Установка и использование подсистемы Linux (WSL 2) в Windows 10

Изначально компонент WSL отключен. Чтобы его включить, вам необходимо зайти в **Пуск -> Панель управления -> Программы и компоненты -> Включение и отключение компонентов Windows (Control Panel\All Control Panel Items\Programs and Features\Turn Windows features on or off)**, активировать галочку Подсистема Windows для Linux (Windows Subsystem for Linux), нажать кнопку ОК, и перезагрузить компьютер.

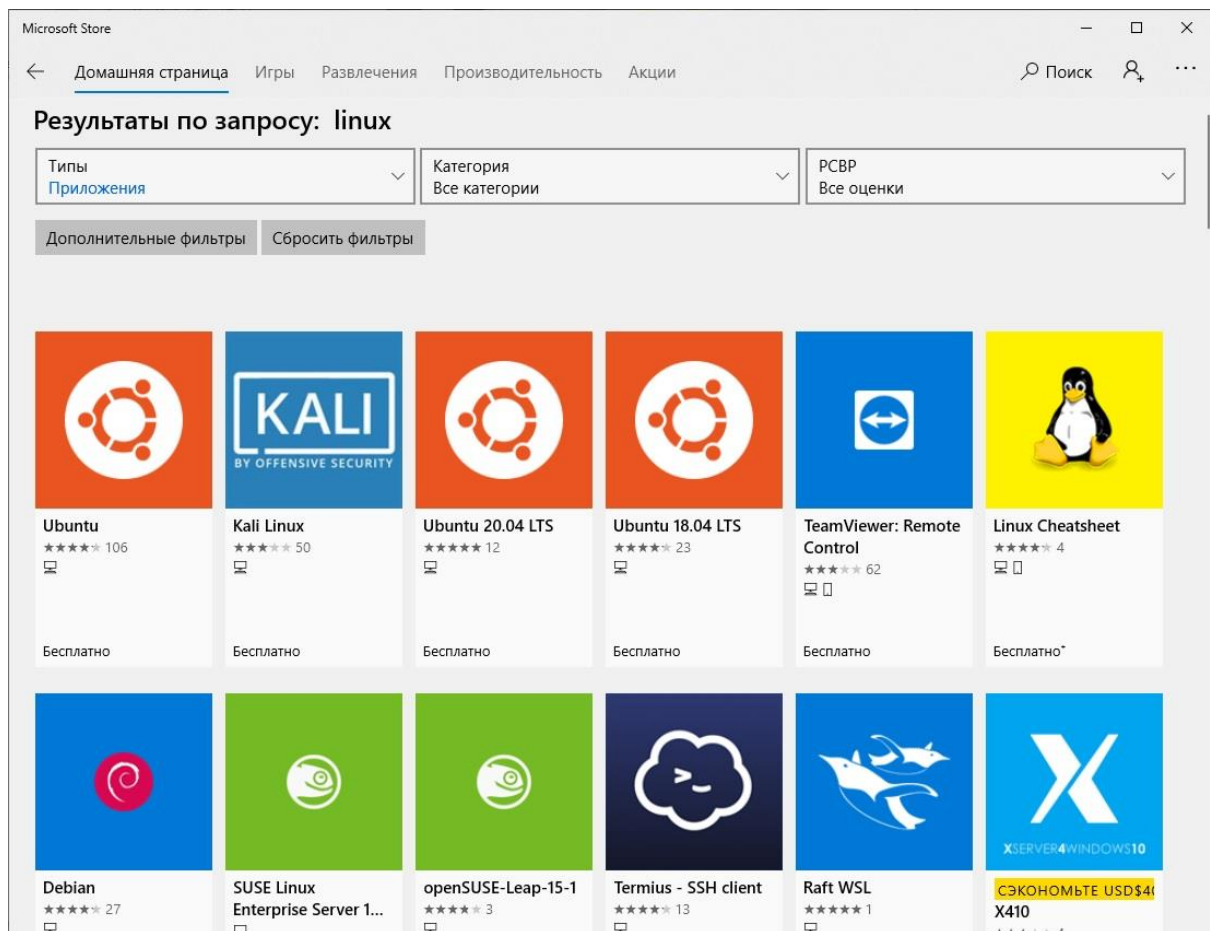


После этого также необходимо перезагрузить компьютер.

Теперь вам нужно выполнить обновление WSL до версии 2. Для этого требуется зайти на сайт <https://docs.microsoft.com/ru-ru/windows/wsl/wsl2-kernel>, скачать файл wsl_update_x64.msi, установить его. По завершении увидите картинку



Далее вам необходимо открыть Microsoft Store, в поиске ввести слово "Linux". В появившемся списке выберите нужный дистрибутив. Доступные Ubuntu, Debian, Kali Linux, Linux Cheatsheet, SUSE Linux Enterprise Server15, openSUSE Leap 15-1, Penguin Enterprise, Fedora Remix for WSL или другие. В нашем примере мы будем использовать Ubuntu 20.04 LTS, нажмите кнопку Получить.



После установки можно проверить используемую версию WSL с помощью команды

```
wsl --list --verbose
```

```
Администратор: Windows PowerShell

Windows PowerShell
(C) Корпорация Майкрософт (Microsoft Corporation). Все права защищены.

Попробуйте новую кроссплатформенную оболочку PowerShell (https://aka.ms/pscore6)

PS C:\WINDOWS\system32> wsl --list --verbose
NAME                STATE              VERSION
* Ubuntu-20.04      Stopped            2
PS C:\WINDOWS\system32>
```

Если на этом экране увидите VERSION 1, тогда команда:

```
wsl --set-version Ubuntu-20.04 2
```

Чтобы сделать WSL2 архитектурой по умолчанию для новых дистрибутивов, в PowerShell выполните команду:

```
wsl --set-default-version 2
```

Шаг 2: Установка и настройка Airflow

Перед началом работы с Apache Airflow необходимо убедиться, что установлен компилятор gcc и необходимые инструменты для разработчика:

```
sudo apt-get update
```

```
sudo apt-get install build-essential python3-dev libpq-dev
```

Создаём виртуальное окружение и активируем его:

```
sudo apt-get install python3-venv
```

```
python3 -m venv .venv
```

```
source .venv/bin/activate
```

Обновляем версию пакетного менеджера pip:

```
pip install pip -U
```

Всё готово для того, чтобы начать установку Apache Airflow. На момент написания этого текста самой последней версией Airflow является 2.0.1.

Я устанавливаю ее такой командой:

```
pip install apache-airflow[postgres,aws]==2.0.1 --constraint
```

```
https://raw.githubusercontent.com/apache/airflow/constraints-2.0.1/constraints-3.8.txt
```

У этой команды есть объяснение, пока не будем его затрагивать. Просто установите.

Apache Airflow использует переменную окружения AIRFLOW_HOME для назначения пути, где будут лежать логи, конфигурационные файлы, python код и прочая вспомогательная информация. По умолчанию путь ~/airflow, переопределите эту переменную, если хотите изменить путь:

```
export AIRFLOW_HOME=~/airflow-data
```

Перед началом работы необходимо инициализировать настройки, подготовить базу данных. Для этого выполните команду:

```
airflow db init
```

По-молчанию в домашней директории должна быть создана папка `airflow`, выполните:

```
cd /root/airflow-data
```

Если что-то с директорией здесь у вас не так, тогда проверьте ее с помощью команды:

```
echo $AIRFLOW_HOME
```

и если директория как-то отличается, то укажите ее в команде

```
cd /root/airflow-data
```

Посмотрите содержимое этой папки командой:

```
ls -la
```

Внутри этой папки должен быть файл `airflow.cfg`, его нужно открыть и сверху будет опция `dags_folder`, вот туда надо указать путь до даг-файлов.

Я открываю файл конфигурации с помощью встроенной в Linux программы для редактирования файлов:

```
nano airflow.cfg
```

Нам необходимо изменить параметр:

```
dags_folder =/mnt/c/Users/vandr/PyCharmProjects/apache-airflow-course-materials/dags
```

и

```
load_examples = False
```

Если ранее этот параметр был `True` и вы инициализировали БД, то после изменения на `False` необходимо сбросить всю информацию о пайплайнах командой:

```
airflow db reset
```

В Airflow 2.0 появилась авторизация по умолчанию, поэтому перед запуском сервера нам необходимо создать пользователя-админа.

```
airflow users create \  
  --username airflow \  
  --firstname Airflow \  
  --lastname Apache \  
  --role Admin \  
  --email airflow@example.com \  
  --password airflow
```

Теперь можно запустить веб-сервер, чтобы поглядеть на пользовательский интерфейс Apache Airflow:

```
airflow webserver -p 8080
```

Помимо всего прочего, для того, чтобы Airflow увидел ваши даги, необходимо включить планировщик.

Для этого открываем новый терминал и делаем следующее:

снова активируем окружение

```
source .venv/bin/activate
```

снова указываем путь к настройкам airflow так как новый терминал ничего об этом еще не знает:

```
export AIRFLOW_HOME=~/.airflow-data
```

Запускаем планировщик:

```
airflow scheduler
```

3. Настройка дополнительных программ

Если мне нужно поработать с Total Commander в Ubuntu:

```
sudo apt-get install mc
```

4. Удаление WSL Ubuntu из вашего компьютера

Для начала идем снова в PowerShell.

Список запущенных программ:

```
wsl --list --all
```

Завершение работы всех дистрибутивов:

```
wsl --shutdown
```

Идальше удалить через список всех программ