Урок 5. Обзор возможностей Airflow, установка и настройка

- 1. Установите AirFlow как показано на семинаре, соберите список команд установки, скриншоты GUI и работы AirFlow из мобы. Соберите все данные в один pdf файл.
- 2. Проверить гипотезу будет ли достаточно установки sudo apt install -y python3 python3-pip python3-venv без sudo apt install python3.8
- 3. Создайте мануал со списком команд для установки.

!!!BAЖHO!!! Все команды из PowerShell выполняются с правами Администратора, ваша учётная запись тоже должна быть с правами администратора, у вас должен быть обеспечен доступ в интеренет.

Все действия выполнялись под ЛИЦЕНЗИОННУЮ операционную систему Windows 10 Pro, разрядность x64, инструкция написана под версию x64, никакой совместимости с 32x не гарантирую, Windows 10 Home не поддерживает WSL.

Так же не гарантируется работоспособность на Wondows 10 Repack by Васяна, выполняйте действия на свой страх и риск, поскольку репаки by Васян часто любят чудить с слётом регистрации и установкой компонентов, когда пытаешься делать, что-то за гранью обычного пользователя.

Это же касается Linux, работоспособность репаков Linux от Васяна на wsl не гарантируется, благо пингвины бесплатны, нужно лишь поставить НОРМАЛЬНЫЙ Linux(к примеру Ubuntu) без всякого гурманства типо Astra Linux который впервую очередь написан для Товарища Майора.

- 1. Для начала установки AirFlow на платформу Windows 10 нужно включить встроеную виртуальную машину Linux.
- 1.1 Вы можете воспользоваться офицальной инструкцией от Microsoft доступной по данным ссылкам:

https://learn.microsoft.com/ru-ru/windows/wsl/install - в случае если данная инструкция выводит ошибки то нужно воспользоваться более старой версией данной иструкции, про неё так же написано в данной иснструкции но на крайний случай ссылку также прилагаю.

https://learn.microsoft.com/ru-ru/windows/wsl/install-manual - как написано более старая инструкция.

https://learn.microsoft.com/ru-ru/windows/wsl/basic-commands#install - список основных команд инструкция.

Краткая инструкция от меня для кому лень читать тонны текста от Microsoft.

- 1.2 Увы, нужно обновить Windows 10 до последней версии, в которой полно рекламы, очень нужный помощьник Сири и куча другой нужной лабуды и шлака(да обновления безопасности процессоров, я о вас, обычному пользователю вы нахер не нужны, пожиратели производительности), но таков путь!
- 1.2.1 Для начала вам надо зайти в биос (кнопка зависит от вашего производителя материнской платы вот некоторые из них: F2, F10, F12, DEL) и включить виртуализацию вашего процессора, в случае если процессор не поддерживает виртуализацию то установка виртуальной машины не возможна.
- 1.2.2 В системе Windows 10 Зайти в "Панель управления" -> "Программы" или "Программы и компоненты" -> "Включение и отключение компонентов Windows".

- 1.2.3 Далее вам нужно либо включить либо удостовериться что включены следующие компонены (поставьте галочки):
 - Платформа для виртуальной машины;
 - Подсистема Windows для Linux;
 - 1.2.4 Перезагрузите компьютер.
 - 1.2.5 Откройте PowerShell и ввидите следующию команду:
- wsl --set-default-version 2 (Данной командой мы обозначим запуск второй версии wsl которая нам нужна)
- 1.2.6 Если в консоле не будет ошибки перезагрузите компьютер и далее перейдите к пункту запуска и настройки Linux.
- В случае той или иной ошибки в консоле то Google вам в помощь и инструкция от Microsoft.
 - 1.2.7 Список команд для работы с wsl:
- wsl install автоматическая устанвока wsl (лично у меня не сработало пришлось делать ручками)
 - wsl --update обновление wsl (скачивание новой версии программы)
 - wsl --version проверить версию установленой и используемой wsl
 - wsl -l список установленых дистрибутивов Linux
- wsl --unregister <name> удаление дистрибутива Linux из wsl из списка установленых пишите имя вашего дистрибутива за место <name>
- wsl --list --running лист запущенных Linux на текущий моент в среде wsl
- wsl -t <name> команда для выключения Linux который запущен название можно взять с листа и вписать в за место поля <name>
 - wsl --shutdown команда которая завершает все виртуальные машины wsl
- Restart-Service LxssManager для перезапуска службы Диспетчера подсистемы Windows для Linux (LxssManager), может потребоваться для решения проблем с WSL или виртуальными машинами Linux, а также для применения обновлений или изменений конфигурации.
- 1.2.8 Здесь я напишу ошибки с которыми, я встречался после выполнения всех действий выше, т.е они возникнут даже если вы все сделали правильно по моей инструкции:

Ошибка:

WslRegisterDistribution failed with error: 0x800701bc ????????????? (Вопросики это кирилица или другой язык которую не может отобразить консоль, вопросиков будет много я не стал их полностью копировать)
Лечение:

Я смог обойти с помощью команды в POwerShell: wsl --update

Ниже оригинал с английской версии сайта:

"I typed wsl --update and the new version of wsl is installed. Then I installed Ubuntu by wsl --install -d ubuntu and it worked."

Ошибка:

Вы вдруг такой решили версию wsl, ввели команду wsl --version а консооль напишет, что такой команды нет.

Лечение:

He переживайте если команда wsl --set-default-version 2 прошла успешно то у вас всё хорошо и Linux должен запуститься после настройки и перезагрузки а после инсталяции Linux команда wsl --version у вас зарабоатет.

Ошибка:

Error code: Wsl/Service/CreateInstance/MountVhd/ERROR_FILE_NOT_FOUND Лечение:

У меня это вылезло когда я захотел переставить свой Linux лечиться командой wsl -l затем командой wsl --unregister Ubuntu всё можете приступать к работе с чистого листа.

Ниже оригинал с английской версии сайта:

List the distributions installed, by running following in PowerShell. wsl -l

Unregister the distribution. Replace the "Ubuntu" below with your distribution name found in Step #1:

NOTE: THIS COMMAND WILL COMPLETELY UNINSTALL YOUR WSL DISTRO.

wsl --unregister Ubuntu

Launch the Ubuntu (or other distribution) which was installed using Microsoft Store

- 2. Клиент Linux можно установить на вашу wsl несколькими способами:
- 2.1 Заходим в магазин Microsoft Store, пишим в поиске Linux, выбираем который понравился и нажимаем скачать, после скачивания и установки, нажимаем запустить, начётся процесс инициализации и установки Linux куда мы вводим свой логин и пароль.
- 2.2 Скачиваем дистрибутив с офицального сайта Linux который вам понравился, распоковываем и запускаем из папки, начнётся процесс инициализации и установки Linux куда мы вводим свой логин и пароль.
- 2.3 Так же образ можно скачать с помощью самой консоли PowerShell для этого:
- 2.3.1 cd C:\ с помощью данной команды перейдите в каталог в котором вы хотите создать папку
- 2.3.2 mkdir <name> с помощью данной команды создайте папку, название папки за место <name>
- 2.3.3 Invoke-WebRequest -Uri https://aka.ms/wslubuntu2004 -OutFile Ubuntu.appx -UseBasicParsing для загрузки образа Linux предварительно не забудьте перейти в созданую папку спомощью команды cd
- Uri: Этот параметр указывает URL-адрес, с которого загружается файл.
- OutFile: Этот параметр указывает путь и имя файла, в который будет сохранен загруженный файл.
- UseBasicParsing: Этот параметр указывает PowerShell не использовать Internet Explorer для анализа веб-страницы при загрузке.

- 2.3.4 move .\Ubuntu.appx .\Ubuntu.zip данная команда используется для того, чтобы переименовать файл в .zip
- 2.3.5 Expand-Archive .\Ubuntu.zip данная команда нужна для распоковки .zip в папку.
- 3. Теперь обновим Linux, добавим репозиторий DeadSnakes PPA (Personal Package Archive), что позволит нам устанавливать различные версии Python, используя "apt" или "apt-get" и установим за одним Python
 - 3.1 sudo apt update получаем пакеты обновлений для Linux
- 3.2 sudo add-apt-repository ppa:deadsnakes/ppa подключаем нужный нам репозиторий
- 3.3 sudo apt install -y python3 python3-pip python3-venv команда устанавливает Python, pip и venv последней версии доступной в поключённом репозитории.
- python3: Это интерпретатор языка программирования Python версии 3.х. Он позволяет запускать и выполнять Python-скрипты и приложения.
- python3-pip: Это инструмент управления пакетами для Python. Он используется для установки, обновления и управления сторонними библиотеками и модулями Python.
- python3-venv: Этот пакет предоставляет виртуальное окружение Python, которое позволяет изолировать проекты Python друг от друга. Виртуальные окружения полезны для управления зависимостями проекта и предотвращения конфликтов между версиями пакетов.
- sudo apt install -y python3.8 python3.8-pip python3.8-venv пример команды если нам нужна конкретная версия Python, в данном случае Python 3.8
 - 3.4 sudo apt update получаем пакеты обновлений для Linux
 - 3.5 sudo apt upgrade обновляем все до последних версий
- 4. Далее мы устанавливаем Mysql server и настроим его, для простоты понимания логов лучше запустить через программу MobaXterm_Personal, однако если у вас нет проблем орентирования в командной строке можно продолжить работу в ней:
 - 4.1 sudo apt-get install mysql-server устанавливаем Mysql server
 - 4.2 sudo /etc/init.d/mysql start запускаем Mysql
- 4.3 sudo mysql_secure_installation теперь запускаем настройку сервиса везде ставим нет кроме последнего пункта:
 - VALIDATE PASSWORD COMPONENT... ставим No
 - By default, a MySQL installation has an anonymous user... ставим No
- Normally, root should only be allowed to connect from
- 'localhost'... ставим No
- By default, MySQL comes with a database named 'test' that... ставим No
- Reloading the privilege tables will ensure that all changes made so far will take effect immediately anyone can access... ставим Yes
- 4.5 sudo mysql -u root -p заходим в программу Mysql, если спросит пароль нажимаем Enter далее вводим данную строку:

ALTER USER 'root'@'localhost' IDENTIFIED WITH mysql_native_password BY '1';

строка выше установит пароль "1", затем выходим командой exit

4.6 sudo nano /etc/mysql/my.cnf данный файл нужно подредактировать добавив в конец все строчки:

[mysqld]

port = 33061

сохранить изменения в файле

- 4.7 sudo service mysql restart перезапускаем Mysql сервер
- 4.8 Теперь вам надо зайти в MySQL оболочку, которая у вас установлина на Windows 10 это может быть HeidiSQL или любая другая программа, вам нужно создать новое поключение к базе данных ввести логин: root(который мы установили), пароль: 1(который мы установили) и порт: 33061(который мы установили) сделать следующий SQL запрос:

CREATE DATABASE Airflow CHARACTER SET UTF8mb3 COLLATE utf8_general_ci; CREATE USER 'Airflow'@'%' IDENTIFIED BY '1';

- в результате у вас появиться база данных Airflow, затем зайти в пользователей и выдать пользователю Airflow все права доступа.
 - 4.9 Чтобы быть более увереным в работе можно ввести команду: SELECT NOW() на выходе должно в запросе получиться "2024-04-24

20:02:06" будет ваше время и дата

- 4.10 Далее надо создать базу данных "spark"
- 5. Это последний шаг а именно установка программы AirFlow
 - 5.1 python3 -m venv airflow-venv && source airflow-
- venv/bin/activate данной командой мы заходим в виртуальную среду airflow
 - 5.2 pip install apache-airflow==2.7.3 --constraint
- "https://raw.githubusercontent.com/apache/airflow/constraints-2.7.3/constraints-3.8.txt" --force-reinstall --upgrade данной командой мы устанавливаем airflow
- 5.3 export AIRFLOW_HOME=/home/airflow && airflow db init теперь делаем пробный запуск airflow

если у вас возникли проблемы с правами доступа тогда сделайте следующие:

- cd /home/ перейдите в данную директорию
- sudo chmod -R 777 ./ и выполните данную команду, затем снова сделайт пробный запуск airflow
- 5.4 Теперь у вас по пути "/home/" появится папка "airflow" там есть файл airflow.cfg который, мы должны заменить нашим файлом airflow.cfg
 - 5.5 Теперь нужно в каталоге "/home/airflow/" создать папку "dags"
- 5.6 export AIRFLOW_HOME=/home/airflow && airflow db init делаем повторную иницализацию

ModuleNotFoundError: No module named 'MySQLdb' - данная ошибка говорит отом, что нам нужно установить ещё один модуль, чтобы его установить введите команду:

sudo apt-get install python3.8-dev libmysqlclient-dev && sudo apt install
pkg-config && pip install mysqlclient

- pip install python3.8 установит дополнительно питон нужной версии
- pip install pkg-config установит дополнительную библиотеку
- pip install mysqlclient установит нужный mysqlclient
- 5.7 Выключаем wsl и включаем его через команду PoserHell wsl -shutdown, затем запускаем заново.

- 5.8 Снова инициализируем виртуальную среду шаг 5.1 и инициализируем airflow шаг 5.3
- 5.9 airflow users create --role Admin --username admin --email admin --firstname admin --lastname admin --password admin данной командой создаём пользователя для airflow
- 5.10 export AIRFLOW_HOME=/home/airflow && (airflow scheduler & airflow webserver -p 8080) данной командой запускаем сервер airflow, чтобы к нему поключиться в адресной страке браузера нужно ввести localhost:8080 в результате чего появиться админ сайт airflow логин/пароль: admin

Поздравляю вы полностью установили и запустили AirFlow в Windows 10 с помощью WSL

























