



Машинное обучение и
высоконагруженные системы

Москва, осень 2023

MLOps. Начало

Гаврилова Елизавета,
Senior ML-engineer





22.09 Что будем обсуждать сегодня?

- Узнаем, зачем нужны виртуальные машины и виртуальные окружения
- Настроим **виртуальное окружение**
- Установим и настроим **AirFlow**
- Установим **MLFlow**





Виртуальная машина

Это программное обеспечение, эмулирующее аппаратное обеспечение компьютера.

Простыми словами, это программа, симулирующая компьютер внутри компьютера со своей полноценной операционной системой.

Технология виртуализации позволяет использовать существующее «железо» для создания его виртуальных копий.



Виртуальная машина: основные термины

Виртуализация – процесс создания виртуальных копий.

Гипервизор – служба, отвечающая за создание и управление виртуальными машинами.

Хостовая система – физическое «железо», на котором создаются и запускаются виртуальные машины.

Гостевая система – виртуальная машина, запущенная на хостовой системе.



Виртуальная машина: какие задачи решает

Тестирование ПО на
различных ОС с
различными
характеристиками

Создание
виртуального сервера

Безопасное
тестирование

Использование
уникального для ОС
софта



Виртуальная машина **V/S** Контейнер

Виртуальные машины имитируют полноценный компьютер со своей операционной системой и аппаратурой.

Контейнеры служат для запуска приложений и их компонент в изолированной среде, имитирующей часть ОС.



Виртуальное окружение

Виртуальное окружение – это инструмент, позволяющий создавать изолированные среды(да, да, опять они).

Виртуальное окружение – это по сути директория, в которой хранятся необходимые файлы для запуска самого Python определенной версии и какого-либо проекта со всеми зависимостями.



Пакетный менеджер

Пакетный менеджер — это программа, позволяющая управлять процессом установки, удаления, настройки и обновления различных компонентов программного обеспечения.

В некоторых ОС есть уже установленные пакетные менеджеры, например **brew** в MacOS и **apt** в Ubuntu.

Пакетные менеджеры могут использоваться для работы с ОС целиком, а могут для разработки на конкретном языке программирования. Самые популярные пакетные менеджеры для работы с python – это **pip** и **conda**.



Как установить AirFlow?

Source

pip

Docker

Helm



Как установить MLFlow?

Source

pip



Основы командной строки

sudo – substitute user & do – замена пользователя для исполнения команды. По умолчанию применяет пользователя **root** (используется для повышения прав), но может быть использована и для выполнения команд от имени любого другого пользователя.

Для windows: аналог запуска чего-либо от имени администратора.

sh – shell – вызов командного интерпретатора shell

source – выполнение скрипта (в текущем процессе оболочки bash)

cat – вывод содержимого файла

cp – копирование файла

mkdir – создание новой директории

ls – перечисление файлов в директории

chmod – выдача прав на папки или файлы

wget – загрузка файлов из интернета

apt-get install – установка какого-либо пакета (с помощью пакетного менеджера apt-get)

apt-get remove – удаление какого-либо пакета

apt-get update – обновление локального индекса пакетов



Основы conda

conda create – создание нового виртуального окружения

conda activate – активация виртуального окружения

conda install – установка пакета в активированное виртуальное окружение

conda remove – удаление пакета из виртуального окружения

conda list – перечисление пакетов, установленных в активированное виртуальное окружение

conda env list – перечисление созданных виртуальных окружений

conda env remove – удаление виртуального окружения



Разобрали сегодня

- Что такое VM и BO
- Как настроить виртуальное окружение
- Как установить и настроить AirFlow

ДЗ

- Прodelать весь сетап руками☺
- Настроить свою почту для получения сообщений с AirFlow

В следующей серии

- Мой первый DAG