Инструкция по установке MongoDB, Metabase, PostgreSQL, Apache (Airflow, Hadoop, Spark, Hive)

**Инструкция по установке**

[Настройка WSL 2 6](#_Toc190544738)

[Примечание: 6](#_Toc190544739)

[Установка и обновление: 6](#_Toc190544740)

[Проверка установки: 6](#_Toc190544741)

[Установка WSL 2 по умолчанию: 6](#_Toc190544742)

[Установка Ubuntu-24.04 (последняя версия на данный момент) 6](#_Toc190544743)

[Как узнать доступные версии: 6](#_Toc190544744)

[Как узнать установленные версии: 7](#_Toc190544745)

[MongoDB 8.0 8](#_Toc190544746)

[Перед началом установки: 8](#_Toc190544747)

[Установка: 8](#_Toc190544748)

[Проверка установки: 8](#_Toc190544749)

[Запуск MongoDB Server: 8](#_Toc190544750)

[Проверка запуска MongoDB Server: 9](#_Toc190544751)

[Запуск MongoDB Shell для работы с MongoDB: 9](#_Toc190544752)

[Восстановление данных из дампа MongoDB: 9](#_Toc190544753)

[PostgreSQL 17: 10](#_Toc190544754)

[Установка PostgreSQL 17: 10](#_Toc190544755)

[Создание БД для Airflow: 10](#_Toc190544756)

[Создание БД для Metabase: 10](#_Toc190544757)

[Создание БД для Apache Hive: 11](#_Toc190544758)

[Установка драйвера JDBC для Apache Hive: 11](#_Toc190544759)

[Metabase v0.53.2 12](#_Toc190544760)

[Предварительная настройка: 12](#_Toc190544761)

[Установка 12](#_Toc190544762)

[WEB UI: 13](#_Toc190544763)

[Apache Airflow latest 14](#_Toc190544764)

[Предварительно: 14](#_Toc190544765)

[Создание пользователя apache: 14](#_Toc190544766)

[Изменение прав в директории /home/apache: 14](#_Toc190544767)

[Добавление пользователя в группу пользователя apache: 14](#_Toc190544768)

[Вход в пользователя apache: 14](#_Toc190544769)

[Создание python venv для airflow: 14](#_Toc190544770)

[Подключение к python venv: 14](#_Toc190544771)

[Установка Airflow: 15](#_Toc190544772)

[Инициализация файлов Airflow: 15](#_Toc190544773)

[Настройка Airflow для работы с PSQL: 15](#_Toc190544774)

[Ищем строчку sql\_alchemy\_conn и заменяем на: 15](#_Toc190544775)

[Ищем строчку executor и заменяем на: 15](#_Toc190544776)

[Ищем строчку dag\_dir\_list\_interval и заменяем на: 15](#_Toc190544777)

[Выходим из vim: 15](#_Toc190544778)

[Запуск инициализации БД: 15](#_Toc190544779)

[Создание пользователя в airflow: 15](#_Toc190544780)

[Создание папки хранения dag-ов: 16](#_Toc190544781)

[WEB UI: 16](#_Toc190544782)

[Apache Hadoop 3.4.1 17](#_Toc190544783)

[Установка Hadoop: 17](#_Toc190544784)

[Настройка .bachrc: 17](#_Toc190544785)

[Применение изменений .bashrc: 18](#_Toc190544786)

[Настройка Hadoop (Single-Node): 18](#_Toc190544787)

[Запуск: 19](#_Toc190544788)

[Настройка YARN Single-Node: 20](#_Toc190544789)

[Запуск: 21](#_Toc190544790)

[Проверка статуса сервисов: 21](#_Toc190544791)

[WEB UI: 21](#_Toc190544792)

[VS Code 22](#_Toc190544793)

[Подключение к WSL: 22](#_Toc190544794)

[Установка расширения для VS Code: 23](#_Toc190544795)

[Выбор интерпретатора: 23](#_Toc190544796)

[Apache Spark 3.5.4 24](#_Toc190544797)

[Входим в пользователя: 24](#_Toc190544798)

[Установка: 24](#_Toc190544799)

[Добавляем в .bashrc: 24](#_Toc190544800)

[Apache Hive 4.0.1 25](#_Toc190544801)

[Установка: 25](#_Toc190544802)

[Настройка .bashrc: 25](#_Toc190544803)

[Применяем изменеия: 25](#_Toc190544804)

[Настройка Hive: 25](#_Toc190544805)

[Настройка Hadoop: 27](#_Toc190544806)

[Инициализация БД: 27](#_Toc190544807)

[Подключение (Локальное): 28](#_Toc190544808)

[Показать базы данных: 28](#_Toc190544809)

[Удалённое подключение: 28](#_Toc190544810)

[WEB UI: 28](#_Toc190544811)

[Службы systemd 29](#_Toc190544812)

[Как запустить службы: 29](#_Toc190544813)

[Как посмотреть статус службы: 29](#_Toc190544814)

[Как остановить службу: 29](#_Toc190544815)

[После добавления новой службы: 29](#_Toc190544816)

[Metabase 30](#_Toc190544817)

[Apache Airflow 31](#_Toc190544818)

[airflow-scheduler 31](#_Toc190544819)

[airflow-webserver 32](#_Toc190544820)

[Apache Hadoop 33](#_Toc190544821)

[Apache Spark 34](#_Toc190544822)

[Apache Hive 35](#_Toc190544823)

[Дополнение 36](#_Toc190544824)

[Если не заходит в WEB UI 36](#_Toc190544825)

# Настройка WSL 2

## Примечание:

Перед началом работы убедитесь, что у вас стоит версия Windows 10 22H2 или более новая версия иначе могут возникнуть проблемы

## Установка и обновление:

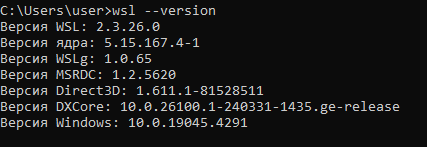
CMD от админа: wsl --install

CMD от админа: wsl --update

----------------------------------------------------------------------------------------------------

## Проверка установки:

CMD: wsl --version



----------------------------------------------------------------------------------------------------

## Установка WSL 2 по умолчанию:

CMD: wsl --set-default-version 2

----------------------------------------------------------------------------------------------------

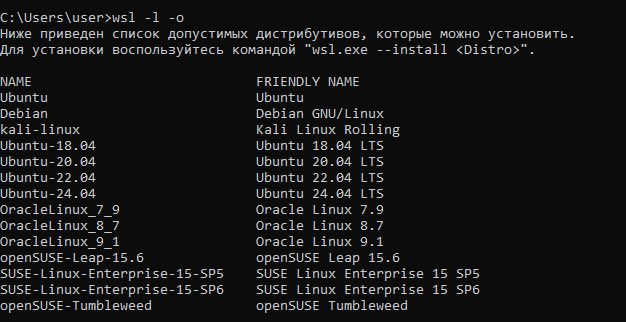
## Установка Ubuntu-24.04 (последняя версия на данный момент)

CMD: wsl --install -d Ubuntu-24.04

----------------------------------------------------------------------------------------------------

## Как узнать доступные версии:

CMD: wsl -l -o



## Как узнать установленные версии:

CMD: wsl -l

# MongoDB 8.0

## Перед началом установки:

sudo apt-get update -y && sudo apt-get upgrade -y

----------------------------------------------------------------------------------------------------

## Установка:

sudo apt install -y gnupg curl

----------------------------------------------------------------------------------------------------

curl -fsSL https://pgp.mongodb.com/server-8.0.asc | sudo gpg -o /usr/share/keyrings/mongodb-server-8.0.gpg --dearmor

----------------------------------------------------------------------------------------------------

echo "deb [ arch=amd64,arm64 signed-by=/usr/share/keyrings/mongodb-server-8.0.gpg ] https://repo.mongodb.org/apt/ubuntu jammy/mongodb-org/8.0 multiverse" | sudo tee /etc/apt/sources.list.d/mongodb-org-8.0.list

----------------------------------------------------------------------------------------------------

sudo apt update

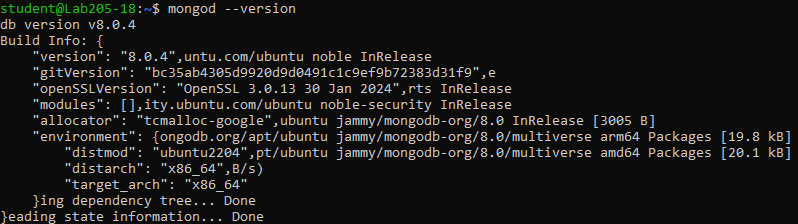
----------------------------------------------------------------------------------------------------

sudo apt install -y mongodb-org

----------------------------------------------------------------------------------------------------

## Проверка установки:

mongod --version



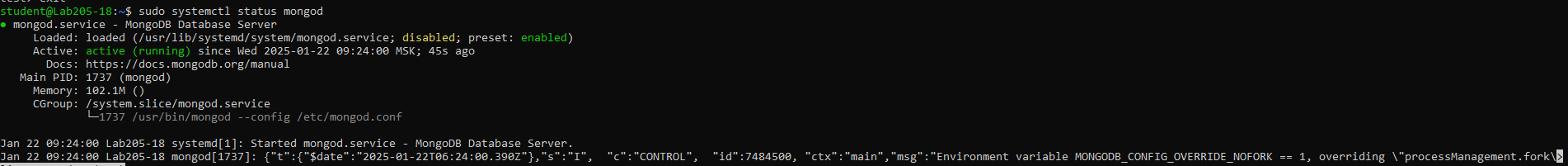
----------------------------------------------------------------------------------------------------

## Запуск MongoDB Server:

sudo systemctl start mongod

## Проверка запуска MongoDB Server:

sudo systemctl status mongod

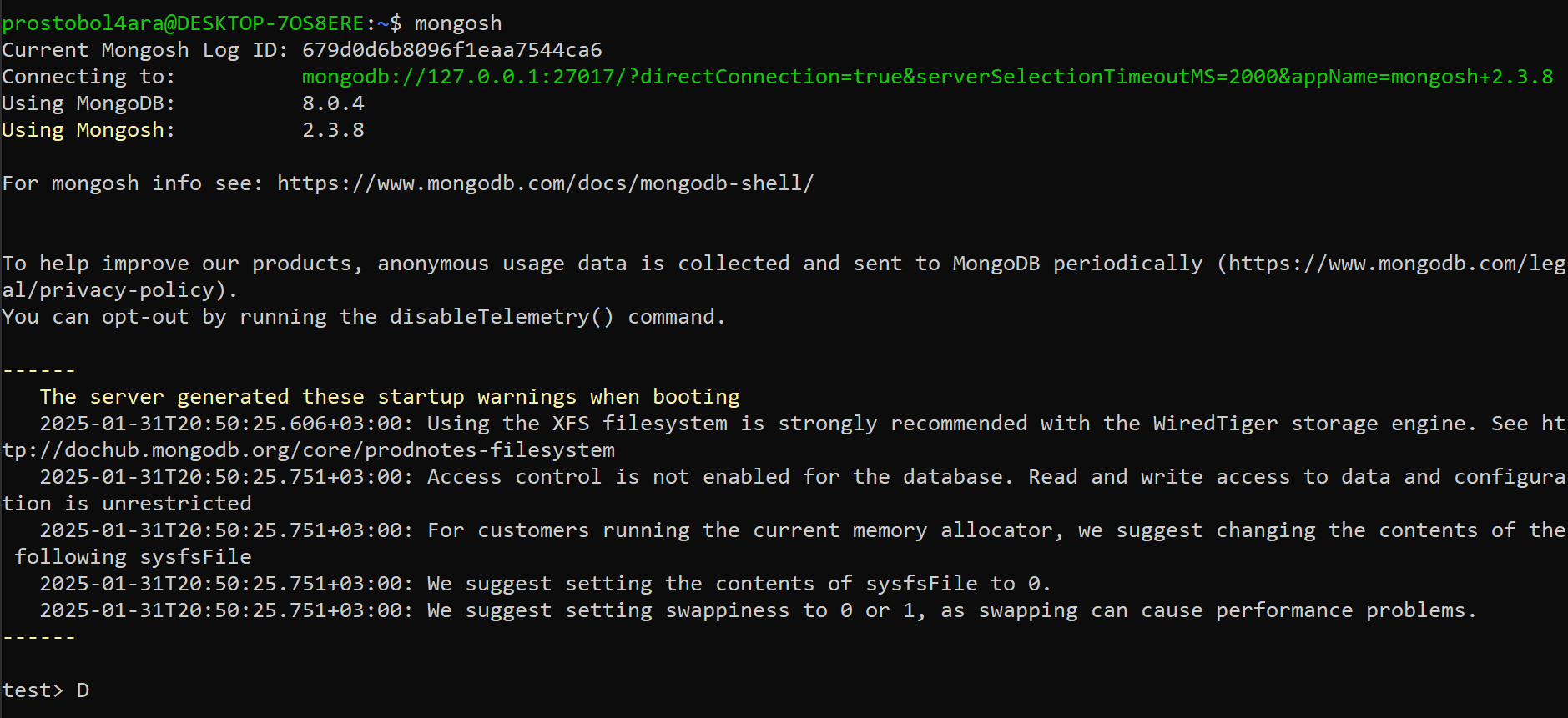


Для выхода из меню службы CTRL + C

----------------------------------------------------------------------------------------------------

## Запуск MongoDB Shell для работы с MongoDB:

mongosh



----------------------------------------------------------------------------------------------------

## Восстановление данных из дампа MongoDB:

mongorestore -d [Название БД] [Путь к файлу]

# PostgreSQL 17:

## Установка PostgreSQL 17:

sudo sh -c 'echo "deb http://apt.postgresql.org/pub/repos/apt $(lsb\_release -cs)-pgdg main" > /etc/apt/sources.list.d/pgdg.list'

----------------------------------------------------------------------------------------------------

curl -fsSL https://www.postgresql.org/media/keys/ACCC4CF8.asc | sudo gpg --dearmor -o /etc/apt/trusted.gpg.d/postgresql.gpg

----------------------------------------------------------------------------------------------------

sudo apt update

----------------------------------------------------------------------------------------------------

sudo apt install -y postgresql-17 postgresql-contrib

----------------------------------------------------------------------------------------------------

## Создание БД для Airflow:

sudo -u postgres psql

----------------------------------------------------------------------------------------------------

CREATE USER airflow\_user WITH PASSWORD 'your\_password';

----------------------------------------------------------------------------------------------------

CREATE DATABASE airflow\_db;

----------------------------------------------------------------------------------------------------

ALTER DATABASE airflow\_db OWNER TO airflow\_user;

----------------------------------------------------------------------------------------------------

\q

----------------------------------------------------------------------------------------------------

## Создание БД для Metabase:

sudo -u postgres psql

----------------------------------------------------------------------------------------------------

CREATE USER metabase\_user WITH PASSWORD 'your\_password';

----------------------------------------------------------------------------------------------------

CREATE DATABASE metabase\_db;

----------------------------------------------------------------------------------------------------

ALTER DATABASE metabase\_db OWNER TO metabase\_user;

----------------------------------------------------------------------------------------------------

\q

----------------------------------------------------------------------------------------------------

## Создание БД для Apache Hive:

sudo -u postgres psql

----------------------------------------------------------------------------------------------------

CREATE USER metastore\_user WITH PASSWORD 'your\_password';

----------------------------------------------------------------------------------------------------

CREATE DATABASE metastore\_db;

----------------------------------------------------------------------------------------------------

ALTER DATABASE metastore\_db OWNER TO metastore\_user;

----------------------------------------------------------------------------------------------------

\q

----------------------------------------------------------------------------------------------------

## Установка драйвера JDBC для Apache Hive:

wget https://jdbc.postgresql.org/download/postgresql-42.7.5.jar

----------------------------------------------------------------------------------------------------

sudo mv postgresql-42.7.5.jar /usr/share/java/postgresql-jdbc.jar

----------------------------------------------------------------------------------------------------

sudo chmod 644 /usr/share/java/postgresql-jdbc.jar

----------------------------------------------------------------------------------------------------

sudo ln -s /usr/share/java/postgresql-jdbc.jar /home/apache/hive/lib/postgresql-jdbc.jar

# Metabase v0.53.2

## Предварительная настройка:

----------------------------------------------------------------------------------------------------

sudo apt install -y openjdk-21-jre-headless

----------------------------------------------------------------------------------------------------

sudo useradd -p your\_password -s /bin/false -U -m metabase

----------------------------------------------------------------------------------------------------

sudo vim /etc/default/metabase

----------------------------------------------------------------------------------------------------

**Вставляем это:**

**Важно перед вставкой нажмите I иначе вставка может произойти не корректно!**

MB\_DB\_TYPE=postgres

MB\_DB\_DBNAME=metabase\_db

MB\_DB\_PORT=5432

MB\_DB\_USER=metabase\_user

MB\_DB\_PASS=[your\_db\_user\_password](#_Создание_БД_для_1)

MB\_DB\_HOST=localhost

----------------------------------------------------------------------------------------------------

sudo chmod 640 /etc/default/metabase

----------------------------------------------------------------------------------------------------

## Установка

wget https://downloads.metabase.com/v0.53.2/metabase.jar

Прим: Перед установкой зайдите проверить последнюю доступную версию metabase на github сайт: <https://github.com/metabase/metabase/releases>



Достаточно скопировать ссылку JAR download и написать: wget [ссылка на jar]

----------------------------------------------------------------------------------------------------

sudo mv metabase.jar /home/metabase/

----------------------------------------------------------------------------------------------------

Готово!  
Для запуска смотрите раздел [systemd](#_Metabase)

## WEB UI:

<http://localhost:3000/>

[Что делать если не заходит в WEB UI?](#_Если_не_заходит)

# Apache Airflow latest

## Предварительно:

sudo apt update

----------------------------------------------------------------------------------------------------

sudo apt install -y python3-pip python3-dev libpq-dev build-essential python3-venv

----------------------------------------------------------------------------------------------------

## Создание пользователя apache:

sudo useradd -p your\_password -s /bin/bash -U -m apache

----------------------------------------------------------------------------------------------------

## Изменение прав в директории /home/apache:

sudo chmod 770 -R /home/apache/

----------------------------------------------------------------------------------------------------

## Добавление пользователя в группу пользователя apache:

sudo usermod -aG apache [Имя пользователя]

После этого стоит перезапустить сессию wsl:

1. Закройте все окна wsl
2. Откройте cmd
3. Введите wsl --shutdown
4. Можете заново открывать wsl

----------------------------------------------------------------------------------------------------

## Вход в пользователя apache:

sudo su - apache

----------------------------------------------------------------------------------------------------

## Создание python venv для airflow:

python3 -m venv apache\_venv

----------------------------------------------------------------------------------------------------

## Подключение к python venv:

source ~/apache\_venv/bin/activate

----------------------------------------------------------------------------------------------------

## Установка Airflow:

pip install apache-airflow

pip install apache-airflow[postgres]

----------------------------------------------------------------------------------------------------

## Инициализация файлов Airflow:

airflow db init

----------------------------------------------------------------------------------------------------

## Настройка Airflow для работы с PSQL:

vim ~/airflow/airflow.cfg

----------------------------------------------------------------------------------------------------

## Ищем строчку sql\_alchemy\_conn и заменяем на:

sql\_alchemy\_conn = postgresql+psycopg2://airflow\_user:your\_password@localhost/airflow\_db

## Ищем строчку executor и заменяем на:

executor = LocalExecutor

## Ищем строчку dag\_dir\_list\_interval и заменяем на:

dag\_dir\_list\_interval = 20

**Ищем строчку load\_examples = True и заменяем на (Опционально):**

load\_examples = False

----------------------------------------------------------------------------------------------------

## Выходим из vim:

:wq!

----------------------------------------------------------------------------------------------------

## Запуск инициализации БД:

airflow db init

----------------------------------------------------------------------------------------------------

## Создание пользователя в airflow:

airflow users create --username admin --password admin --firstname Firstname --lastname Lastname --role Admin --email admin@example.com

----------------------------------------------------------------------------------------------------

## Создание папки хранения dag-ов:

mkdir ~/airflow/dags

----------------------------------------------------------------------------------------------------

## WEB UI:

<http://localhost:8080/>

[Что делать если не заходит в WEB UI?](#_Если_не_заходит)

Для создание служб и запуска Apache Airflow смотрите раздел [Службы systemd](#_Службы_systemd)

# Apache Hadoop 3.4.1

## Установка Hadoop:

sudo apt update && sudo apt install openjdk-8-jdk -y

----------------------------------------------------------------------------------------------------

sudo apt install ssh -y

----------------------------------------------------------------------------------------------------

sudo su - apache

----------------------------------------------------------------------------------------------------

ssh-keygen -t rsa -P '' -f ~/.ssh/id\_rsa

----------------------------------------------------------------------------------------------------

cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

----------------------------------------------------------------------------------------------------

chmod 640 ~/.ssh/authorized\_keys

----------------------------------------------------------------------------------------------------

wget https://dlcdn.apache.org/hadoop/common/hadoop-3.4.1/hadoop-3.4.1.tar.gz

----------------------------------------------------------------------------------------------------

tar -xvzf hadoop-3.4.1.tar.gz

----------------------------------------------------------------------------------------------------

mv hadoop-3.4.1 hadoop

----------------------------------------------------------------------------------------------------

## Настройка .bachrc:

vim ~/.bashrc

----------------------------------------------------------------------------------------------------

**Вставляем в конец файла это:**

**Важно перед вставкой нажмите I иначе вставка может произойти не корректно!**

export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64

export HADOOP\_HOME=/home/apache/hadoop

export HADOOP\_INSTALL=$HADOOP\_HOME

export HADOOP\_MAPRED\_HOME=$HADOOP\_HOME

export HADOOP\_COMMON\_HOME=$HADOOP\_HOME

export HADOOP\_HDFS\_HOME=$HADOOP\_HOME

export HADOOP\_YARN\_HOME=$HADOOP\_HOME

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_HOME/lib/native

export PATH=$PATH:$HADOOP\_HOME/sbin:$HADOOP\_HOME/bin

export HADOOP\_OPTS=”-Djava.library.path=$HADOOP\_HOME/lib/native”

P.S Совет проверьте (”) т.к они могут плохо вставляться в консоль так что ручками заменяем на (“)

----------------------------------------------------------------------------------------------------

## Применение изменений .bashrc:

source ~/.bashrc

----------------------------------------------------------------------------------------------------

## Настройка Hadoop (Single-Node):

**Вместо vim советую использовать nano**

----------------------------------------------------------------------------------------------------

**Открываем hadoop-env.sh:**

vim $HADOOP\_HOME/etc/hadoop/hadoop-env.sh

----------------------------------------------------------------------------------------------------

**Найдите линию (export JAVA\_HOME=) и вставьте это:**

export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64

----------------------------------------------------------------------------------------------------

**Открываем core-site.xml:**

vim $HADOOP\_HOME/etc/hadoop/core-site.xml

----------------------------------------------------------------------------------------------------

**Вставляем вот это:**

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>

</configuration>

----------------------------------------------------------------------------------------------------

**Открываем hdfs-site.xml:**

vim $HADOOP\_HOME/etc/hadoop/hdfs-site.xml

----------------------------------------------------------------------------------------------------

**Вставляем вот это:**

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:///home/apache/hadoopdata/hdfs/namenode</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>file:///home/apache/hadoopdata/hdfs/datanode</value>

</property>

</configuration>

----------------------------------------------------------------------------------------------------

## Запуск:

hdfs namenode -format

start-dfs.sh

----------------------------------------------------------------------------------------------------

**Проверка:**

<http://localhost:9870/>

[Что делать если не заходит в WEB UI?](#_Если_не_заходит)

----------------------------------------------------------------------------------------------------

## Настройка YARN Single-Node:

----------------------------------------------------------------------------------------------------

**Открываем файл mapred-site.xml:**

vim $HADOOP\_HOME/etc/hadoop/mapred-site.xml

----------------------------------------------------------------------------------------------------

**Вставляем это:**

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

<property>

<name>mapreduce.application.classpath</name>

<value>$HADOOP\_MAPRED\_HOME/share/hadoop/mapreduce/\*:$HADOOP\_MAPRED\_HOME/share/hadoop/mapreduce/lib/\*</value>

</property>

</configuration>

----------------------------------------------------------------------------------------------------

**Открываем файл yarn-site.xml:**

vim $HADOOP\_HOME/etc/hadoop/yarn-site.xml

**Вставляем это:**

<configuration>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.env-whitelist</name>

<value>JAVA\_HOME,HADOOP\_COMMON\_HOME,HADOOP\_HDFS\_HOME,HADOOP\_CONF\_DIR,CLASSPATH\_PREPEND\_DISTCACHE,HADOOP\_YARN\_HOME,HADOOP\_HOME,PATH,LANG,TZ,HADOOP\_MAPRED\_HOME</value>

</property>

</configuration>

## Запуск:

hdfs namenode -format

start-all.sh

----------------------------------------------------------------------------------------------------

## Проверка статуса сервисов:

Jps

----------------------------------------------------------------------------------------------------

## WEB UI:

<http://localhost:8088/>

<http://localhost:9870/>

[Что делать если не заходит в WEB UI?](#_Если_не_заходит)

----------------------------------------------------------------------------------------------------

Для создание служб и запуска Apache Hadoop смотрите раздел [Службы systemd](#_Службы_systemd):

# VS Code

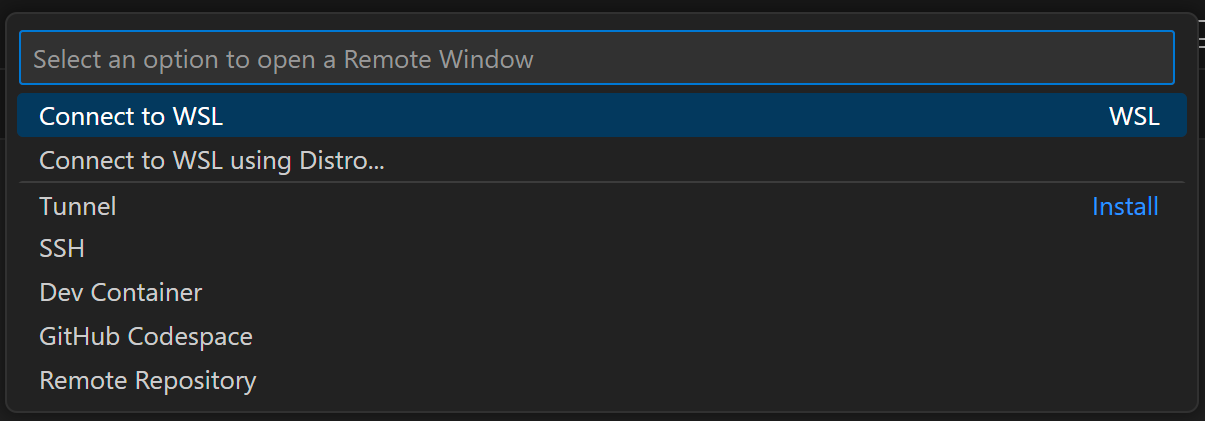
## Подключение к WSL:

Находите такой символ снизу слева окна VS Code:



Нажимаете на него

После чего появляется примерно такое окно ищите WSL/Connect to WSL/Connect to WSL using Disto…:



В первом случае (WSL): Нажмите на него VS Code скачает необходимое расширение для работы с WSL  
Во втором случае (Connect to WSL): Вы сразу подключитесь к дистрибутиву Linux который стоит у вас по умолчанию

В третьем случае (Connect to WSL using Distro): Вам будет дан выбор к какому именно дистрибутиву WSL вы хотите подключиться

## Установка расширения для VS Code:



----------------------------------------------------------------------------------------------------

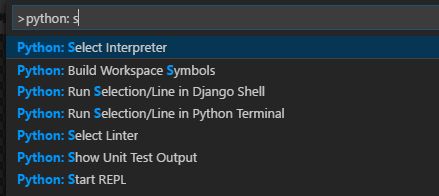
## Выбор интерпретатора:

Ctrl + Shift + P

----------------------------------------------------------------------------------------------------

**В окошке пишем:**

Python: Select Interpreter



----------------------------------------------------------------------------------------------------

**Потом находим Python в Python venv:**

/home/apache/apache\_venv/bin/python

# Apache Spark 3.5.4

## Входим в пользователя:

sudo su - apache

----------------------------------------------------------------------------------------------------

## Установка:

wget https://dlcdn.apache.org/spark/spark-3.5.4/spark-3.5.4-bin-hadoop3.tgz

----------------------------------------------------------------------------------------------------

tar -xzvf spark-3.5.4-bin-hadoop3.tgz

----------------------------------------------------------------------------------------------------

mv spark-3.5.4-bin-hadoop3 spark

----------------------------------------------------------------------------------------------------

cd spark

----------------------------------------------------------------------------------------------------

export SPARK\_HOME=`pwd`

----------------------------------------------------------------------------------------------------

export PYTHONPATH=$(ZIPS=("$SPARK\_HOME"/python/lib/\*.zip); IFS=:; echo "${ZIPS[@]}"):$PYTHONPATH

----------------------------------------------------------------------------------------------------

Добавляем в .bashrc:  
export SPARK\_HOME=/home/apache/spark

----------------------------------------------------------------------------------------------------

export PYTHONPATH=$(ZIPS=("$SPARK\_HOME"/python/lib/\*.zip); IFS=:; echo "${ZIPS[@]}"):$PYTHONPATH

Для создание служб и запуска Apache Spark смотрите раздел [Службы systemd](#_Службы_systemd):

# Apache Hive 4.0.1

## Установка:

wget https://dlcdn.apache.org/hive/hive-4.0.1/apache-hive-4.0.1-bin.tar.gz

----------------------------------------------------------------------------------------------------

tar -xzvf apache-hive-4.0.1-bin.tar.gz

----------------------------------------------------------------------------------------------------

mv apache-hive-4.0.1-bin hive

----------------------------------------------------------------------------------------------------

## Настройка .bashrc:

vim ~/.bashrc

----------------------------------------------------------------------------------------------------

**Вставляем в конец файла это:**

**Важно перед вставкой нажмите I иначе вставка может произойти не корректно!**

export HIVE\_HOME=/home/apache/hive

export PATH=$PATH:$HIVE\_HOME/sbin:$HIVE\_HOME/bin

export CLASSPATH=$CLASSPATH:$HADOOP\_HOME/lib/\*:$HIVE\_HOME/lib/\*

----------------------------------------------------------------------------------------------------

## Применяем изменеия:

source ~/.bashrc

----------------------------------------------------------------------------------------------------

## Настройка Hive:

cd $HIVE\_HOME/conf

----------------------------------------------------------------------------------------------------

cp hive-default.xml.template hive-site.xml

----------------------------------------------------------------------------------------------------

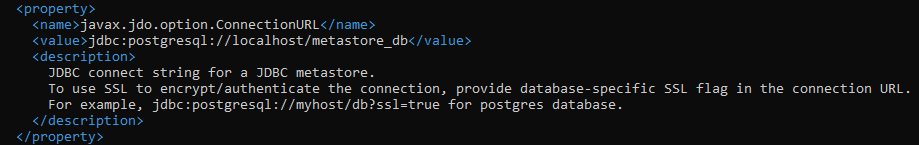
nano hive-site.xml

Замените все ${system:java.io.tmpdir} на /tmp/hive

Замените все ${system:user.name} на apache

Так же в свойсте javax.jdo.option.ConnectionURL:

jdbc:postgresql://localhost/metastore\_db



Так же javax.jdo.option.ConnectionDriverName:

org.postgresql.Driver

Так же javax.jdo.option.ConnectionUserName:

metastore\_user

Так же javax.jdo.option.ConnectionPassword:

[your\_password](#_Создание_БД_для)

----------------------------------------------------------------------------------------------------  
cp hive-env.sh.template hive-env.sh

----------------------------------------------------------------------------------------------------

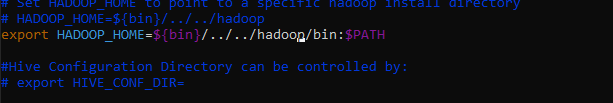
vim hive-env.sh

ищем # HADOOP\_HOME=${bin}/../../hadoop

**Вставляем это:**

**Важно перед вставкой нажмите I иначе вставка может произойти не корректно!**

export HADOOP\_HOME=${bin}/../../hadoop



----------------------------------------------------------------------------------------------------

nano $HADOOP\_HOME/etc/hadoop/core-site.xml

**Добавляем:**

<property>

<name>hadoop.proxyuser.apache.hosts</name>

<value>\*</value>

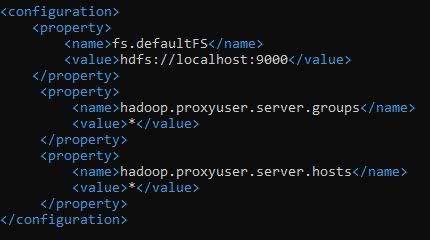
</property>

<property>

<name>hadoop.proxyuser.apache.groups</name>

<value>\*</value>

</property>



----------------------------------------------------------------------------------------------------

## Настройка Hadoop:

hdfs dfs -mkdir /user

----------------------------------------------------------------------------------------------------

hdfs dfs -mkdir /user/hive

----------------------------------------------------------------------------------------------------

hdfs dfs -mkdir /user/hive/warehouse

----------------------------------------------------------------------------------------------------

hdfs dfs -mkdir /user/tmp

----------------------------------------------------------------------------------------------------

hdfs dfs -chmod g+w /user/tmp

----------------------------------------------------------------------------------------------------

hdfs dfs -chmod g+w /user/hive/warehouse

----------------------------------------------------------------------------------------------------

## Инициализация БД:

schematool -dbType postgres –initSchema

----------------------------------------------------------------------------------------------------

## Подключение (Локальное):

beeline -u jdbc:hive2:// -n new\_username -p password

## Показать базы данных:

show databases;

## Удалённое подключение:

beeline -u jdbc:hive2://localhost:10000 apache

----------------------------------------------------------------------------------------------------

## WEB UI:

<http://localhost:10002/>

[Что делать если не заходит в WEB UI?](#_Если_не_заходит)

----------------------------------------------------------------------------------------------------

Для создание служб и запуска Apache Hive смотрите раздел [Службы systemd](#_Службы_systemd):

# Службы systemd

**systemd** — подсистема инициализации и управления службами в Linux, фактически вытеснившая в 2010-е годы традиционную подсистему init. Основная особенность — интенсивное распараллеливание запуска служб в процессе загрузки системы, что позволяет существенно ускорить запуск операционной системы. Основная единица управления — модуль, одним из типов модулей являются «службы» — аналог демонов — наборы процессов, запускаемые и управляемые средствами подсистемы и изолируемые контрольными группами.

----------------------------------------------------------------------------------------------------

## Как запустить службы:

sudo systemctl start [Название службы]

**Пример:**

sudo systemctl start airflow-scheduler.service

----------------------------------------------------------------------------------------------------

## Как посмотреть статус службы:

sudo systemctl status [Название службы]

**Пример:**

sudo systemctl status airflow-scheduler.service

----------------------------------------------------------------------------------------------------

## Как остановить службу:

sudo systemctl stop [Название службы]

**Пример:**

sudo systemctl stop airflow-scheduler.service

## После добавления новой службы:

sudo systemctl daemon-reload

Требуется для корректной работы systemd. Иначе будет вылезать Warning.

## Metabase

----------------------------------------------------------------------------------------------------

**metabase**

**Перед создание службы убедитесь что вы находитесь под суперпользователем (тот что не apache/metabase)**

sudo vim /etc/systemd/system/metabase.service

----------------------------------------------------------------------------------------------------

Вставляем это:

Важно перед вставкой нажмите I иначе вставка может произойти не корректно!

[Unit]

Description=Metabase server

After=network.target

[Service]

WorkingDirectory=/home/metabase

ExecStart=/usr/lib/jvm/java-21-openjdk-amd64/bin/java --add-opens java.base/java.nio=ALL-UNNAMED -jar /home/metabase/metabase.jar

EnvironmentFile=/etc/default/metabase

User=metabase

Type=simple

Restart=on-failure

[Install]

WantedBy=multi-user.target

## Apache Airflow

----------------------------------------------------------------------------------------------------

### airflow-scheduler

sudo vim /etc/systemd/system/airflow-scheduler.service

**Перед создание службы убедитесь что вы находитесь под суперпользователем (тот что не apache)**

----------------------------------------------------------------------------------------------------

Вставляем это:

**Важно перед вставкой нажмите I иначе вставка может произойти не корректно!**

[Unit]

Description=Apache Airflow Scheduler

After=network.target

[Service]

Environment="AIRFLOW\_HOME=/home/apache/airflow"

User=apache

Group=apache

Type=simple

ExecStart=/home/apache/apache\_venv/bin/airflow scheduler

Restart=always

RestartSec=5s

[Install]

WantedBy=multi-user.target

### airflow-webserver

sudo vim /etc/systemd/system/airflow-webserver.service

**Перед создание службы убедитесь что вы находитесь под суперпользователем (тот что не apache)**

----------------------------------------------------------------------------------------------------

**Вставляем это:**

**Важно перед вставкой нажмите I иначе вставка может произойти не корректно!**

[Unit]

Description=Apache Airflow Web Server

After=network.target

[Service]

Environment="AIRFLOW\_HOME=/home/apache/airflow"

User=apache

Group=apache

Type=simple

ExecStart=/home/apache/apache\_venv/bin/airflow webserver --port 8080

Restart=always

RestartSec=5s

[Install]

WantedBy=multi-user.target

## Apache Hadoop

----------------------------------------------------------------------------------------------------

**hadoop**

vim /etc/systemd/system/hadoop.service

**Перед создание службы убедитесь что вы находитесь под суперпользователем (тот что не apache)**

----------------------------------------------------------------------------------------------------

**Вставляем это:**

**Важно перед вставкой нажмите I иначе вставка может произойти не корректно!**

[Unit]

Description=Hdfs service

After=network.target

[Service]

Type=forking

User=apache

Group=apache

ExecStart=/home/apache/hadoop/sbin/start-all.sh

ExecStop=/home/apache/hadoop/sbin/stop-all.sh

ExecReload=/bin/kill -HUP $MAINPID

Restart=on-failure

[Install]

WantedBy=multi-user.target

## Apache Spark

----------------------------------------------------------------------------------------------------

**spark**

sudo vim /etc/systemd/system/spark.service

**Перед создание службы убедитесь что вы находитесь под суперпользователем (тот что не apache)**

----------------------------------------------------------------------------------------------------

**Вставляем это:**

**Важно перед вставкой нажмите I иначе вставка может произойти не корректно!**

[Unit]

Description=Spark service

After=network.target

[Service]

Type=forking

User=apache

Group=apache

ExecStart=/home/apache/spark/sbin/start-master.sh

ExecStop=/home/apache/spark/sbin/stop-master.sh

ExecReload=/bin/kill -HUP $MAINPID

Restart=on-failure

[Install]

WantedBy=multi-user.target

## Apache Hive

----------------------------------------------------------------------------------------------------

**hive-server**

sudo vim /etc/systemd/system/hive-server.service

**Перед создание службы убедитесь что вы находитесь под суперпользователем (тот что не apache)**

----------------------------------------------------------------------------------------------------

**Вставляем это:**

**Важно перед вставкой нажмите I иначе вставка может произойти не корректно!**

[Unit]

Description=hiveserver2 service

After=syslog.target

After=network.target

[Service]

Type=simple

User=apache

Group=apache

ExecStart=/home/apache/hive/bin/hiveserver2

[Install]

WantedBy=multi-user.target

# Дополнение

Отличный сайт с дополнительной инфой и примерами по Spark, PySpark, Pandas, Hive:

https://sparkbyexamples.com/

## Если не заходит в WEB UI

Иногда подключение к localhost может не работать! Пробуйте подключаться по IP WSL!

Пример: <http://172.29.80.12:8088>

