УСТАНОВКА DOCKER НА UBUNTU (https://github.com/BosenkoTM/Distributed\_systems/blob/main/d-zadanie/03-mq/3%D0%A1%D0%B5%D0%BC%D0%B8%D0%BD%D0%B0%D1%80RabbitMQ.pdf)

$ sudo apt update

$ sudo apt install apt-transport-https ca-certificates curl software-properties-common

$ curl -fsSL https://download.docker.com/linux/ubuntu/gpg | sudo apt-key add –

$ sudo add-apt-repository "deb [arch=amd64] https://download.docker.com/linux/ubuntu focal stable"

$ sudo apt update $ apt-cache policy docker-ce

$ sudo apt install docker-ce

$ sudo systemctl status docker

$ sudo curl -L "https://github.com/docker/compose/releases/download/1.25.0/dockercompose-$(uname -s)-$(uname -m)" -o /usr/local/bin/docker-compose

$ sudo chmod +x /usr/local/bin/docker-compose

$ docker-compose --version

$ docker-compose ps

Для запуска Hadoop в Docker контейнере можно использовать готовый образ, например, sequenceiq/hadoop-docker. Для запуска word count на этом контейнере необходимо выполнить следующие действия:

Запустить контейнер:

docker run -it sequenceiq/hadoop-docker /etc/bootstrap.sh -bash

Проверить, что Hadoop запущен:

hadoop version

Создать входные данные для word count:

echo "hello world" > input.txt

Скопировать входные данные в HDFS:

hadoop fs -mkdir /input

hadoop fs -put input.txt /input

Запустить word count:

hadoop jar $HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.1.jar wordcount /input /output

Проверить результат:

hadoop fs -cat /output/\*

Теперь вы можете увидеть результат выполнения word count в консоли.

Подсчет определенного слова:

hadoop jar $HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.1.jar grep /input /output «hello»

Проверить результат:

hadoop fs -cat /output/\*