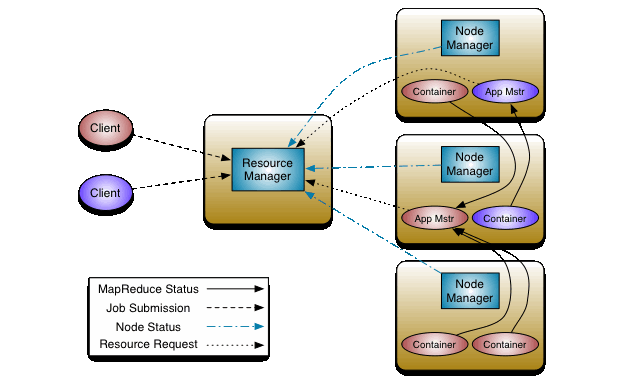
**HADOOP & YARN**

**APACHE HADOOP YARN**

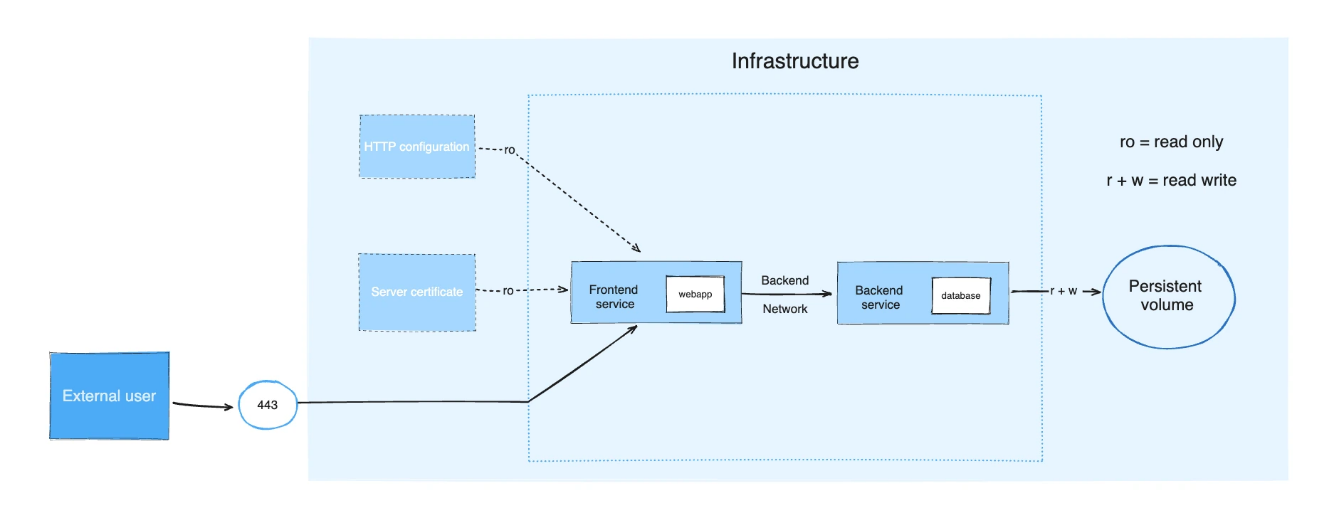
* Docs: <https://hadoop.apache.org/docs/stable/hadoop-yarn/hadoop-yarn-site/YARN.html>

[](file:///T:\Hadoop-yarn-html\Lab3-Hadoop-yarn-html\Hadoop%20&%20Yarn%200710e74db3724817a6e5c5035ec0f577\image.png)

**RUN HADOOP YARN VỚI DOCKER COMPOSE**

1. **Docker compose**

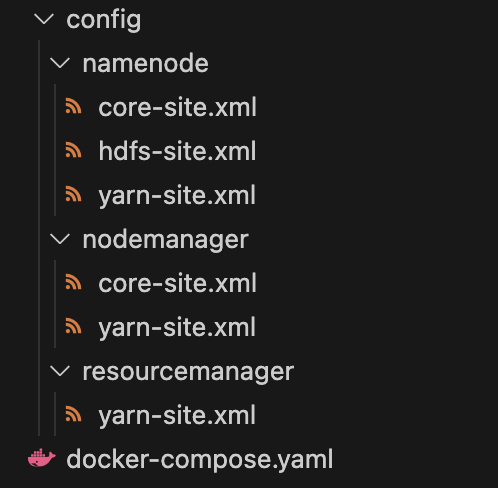
* Docs: <https://docs.docker.com/compose/>

[](file:///T:\Hadoop-yarn-html\Lab3-Hadoop-yarn-html\Hadoop%20&%20Yarn%200710e74db3724817a6e5c5035ec0f577\image%201.png)

* Volumn
* Network

1. **Hadoop Yarn với docker compose**

Cấu trúc thư mục lab3:

[](file:///T:\Hadoop-yarn-html\Lab3-Hadoop-yarn-html\Hadoop%20&%20Yarn%200710e74db3724817a6e5c5035ec0f577\dir.png)

Docker-compose.yaml file

services:

namenode:

image: bde2020/hadoop-namenode:2.0.0-hadoop2.7.4-java8

container\_name: namenode

environment:

- CLUSTER\_NAME=test-cluster

volumes:

- namenode:/hadoop/dfs/name

- ./config/namenode/core-site.xml:/etc/hadoop/core-site.xml

- ./config/namenode/hdfs-site.xml:/etc/hadoop/hdfs-site.xml

- ./config/namenode/yarn-site.xml:/etc/hadoop/yarn-site.xml

ports:

- 9870:9870

- 9000:9000

networks:

- hadoop

datanode:

image: bde2020/hadoop-datanode:2.0.0-hadoop2.7.4-java8

container\_name: datanode

environment:

- CORE\_CONF\_fs\_defaultFS=hdfs://namenode:9000

- CLUSTER\_NAME=test-cluster

volumes:

- datanode:/hadoop/dfs/data

networks:

- hadoop

resourcemanager:

image: bde2020/hadoop-resourcemanager:2.0.0-hadoop2.7.4-java8

container\_name: resourcemanager

environment:

- CORE\_CONF\_fs\_defaultFS=hdfs://namenode:9000

volumes:

- ./config/resourcemanager/yarn-site.xml:/etc/hadoop/yarn-site.xml

ports:

- 8088:8088

networks:

- hadoop

nodemanager:

image: bde2020/hadoop-nodemanager:2.0.0-hadoop2.7.4-java8

container\_name: nodemanager

environment:

- CORE\_CONF\_fs\_defaultFS=hdfs://namenode:9000

volumes:

- ./config/nodemanager/yarn-site.xml:/etc/hadoop/yarn-site.xml

networks:

- hadoop

depends\_on:

- resourcemanager

historyserver:

image: bde2020/hadoop-historyserver:2.0.0-hadoop2.7.4-java8

container\_name: historyserver

environment:

- CORE\_CONF\_fs\_defaultFS=hdfs://namenode:9000

- YARN\_CONF\_yarn\_resourcemanager\_hostname=resourcemanager

- YARN\_CONF\_yarn\_resourcemanager\_address=resourcemanager:8050

networks:

- hadoop

volumes:

namenode:

datanode:

networks:

hadoop:

* **Khởi động docker compose**

docker compose up -d --build

Output

🍎 ❯ docker compose up -d --build

[+] Running 5/5

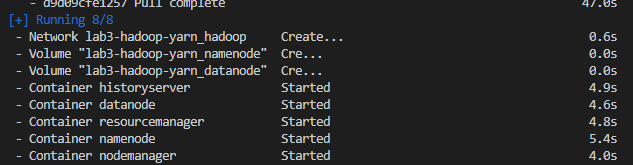
✔ Container resourcemanager Started 0.2s

✔ Container namenode Started 0.2s

✔ Container historyserver Started 0.2s

✔ Container datanode Started 0.2s

✔ Container nodemanager Started 0.4s

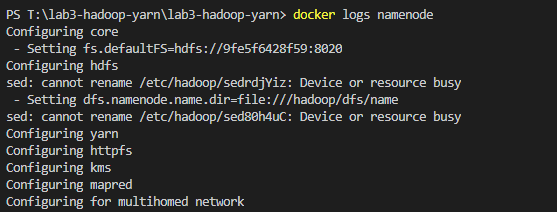


* **Kiểm tra log từng service bằng lệnh**

docker logs <container\_name>

ví dụ xem log ở container *namenode*

docker logs namenode

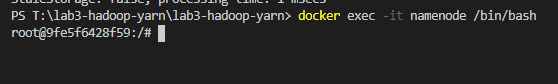


* **Kiểm tra cluster hadoop**
  + **NameNode UI:** [http://localhost:9870](http://localhost:9870/)
  + **ResourceManager UI:** http://localhost:8088

**Một số lệnh yarn cơ bản**

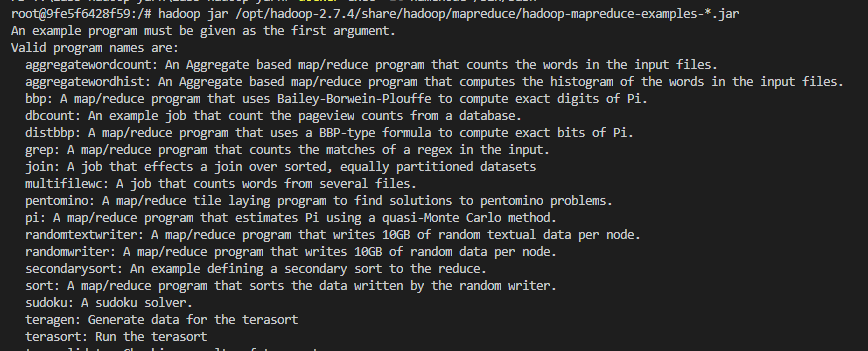
Truy cập vào cluser bằng lệnh:

docker exec -it namenode /bin/bash



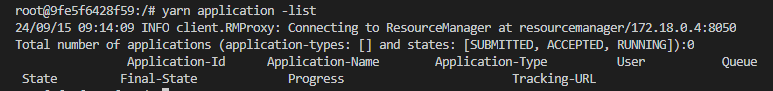
* **Danh sách các ứng dụng được cài đặt sẵn:**

hadoop jar /opt/hadoop-2.7.4/share/hadoop/mapreduce/hadoop-mapreduce-examples-\*.jar



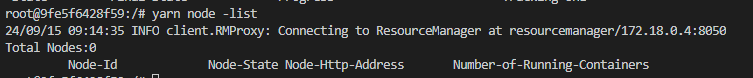
* **Hiển thị các ứng dụng đang chạy và đã hoàn thành.**

yarn application –list



* **Xem danh sách các node**

yarn node –list



**WordCount**

* Tạo thư mục lab2

mkdir lab3 && cd lab3



* Tạo file input như sau và lưu với tên iuh.txt

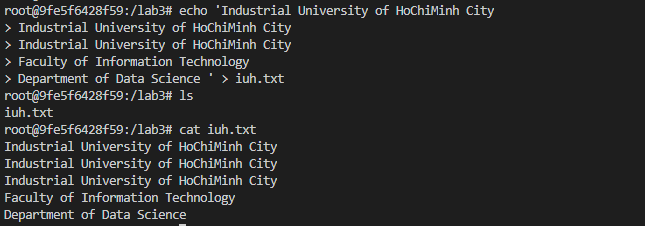
echo 'Industrial University of HoChiMinh City

Industrial University of HoChiMinh City

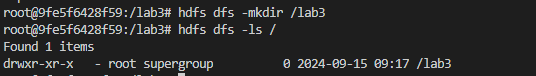
Industrial University of HoChiMinh City

Faculty of Information Technology

Department of Data Science ' > iuh.txt



* Bài tập:
  + **Trên dfs tạo 1 thư mục tên lab3 và tải file iuh.txt vào thư mục đó**

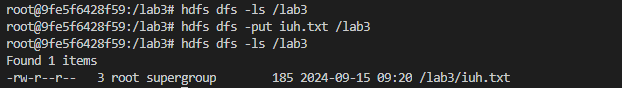


* Kiểm tra file trên dfs

root@2bb70ff7de7f:/lab2# hdfs dfs -ls /lab2

Found 1 items

-rw-r--r-- 3 root supergroup 190 2024-09-03 07:41 /lab2/iuh.txt

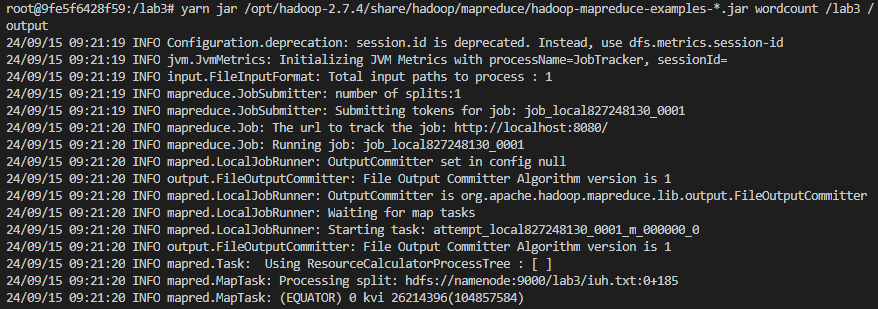
****

* Chạy ứng dụng workcout được cài sẵn và xem kết quả

yarn jar <JARs> <application\_name> <input\_dir> <output\_dir>

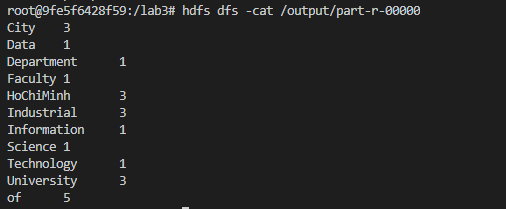
ví dụ run với ứng dụng workcout

yarn jar /opt/hadoop-2.7.4/share/hadoop/mapreduce/hadoop-mapreduce-examples-\*.jar wordcount /lab2 /output



* Kiểm tra kết quả tại thư mục output

root@2bb70ff7de7f:/lab2# hdfs dfs -cat /output/part-r-00000



**Sudoku**

* Chuẩn bị input có tên ‘sudoku\_input.txt' với nội dung như sau

echo '8 5 ? 3 9 ? ? ? ?

? ? 2 ? ? ? ? ? ?

? ? 6 ? 1 ? ? ? 2

? ? 4 ? ? 3 ? 5 9

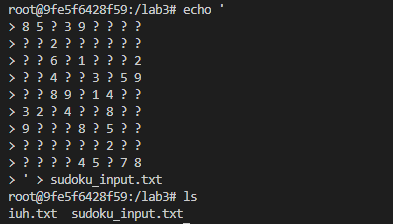
? ? 8 9 ? 1 4 ? ?

3 2 ? 4 ? ? 8 ? ?

9 ? ? ? 8 ? 5 ? ?

? ? ? ? ? ? 2 ? ?

? ? ? ? 4 5 ? 7 8' > sudoku\_input.txt



* Hãy dùng ứng dụng sudoku để giải bài toán trên.cat

yarn jar /opt/hadoop-2.7.4/share/hadoop/mapreduce/hadoop-mapreduce-examples-\*.jar sudoku /lab3/sudoku\_input.txt

Output

