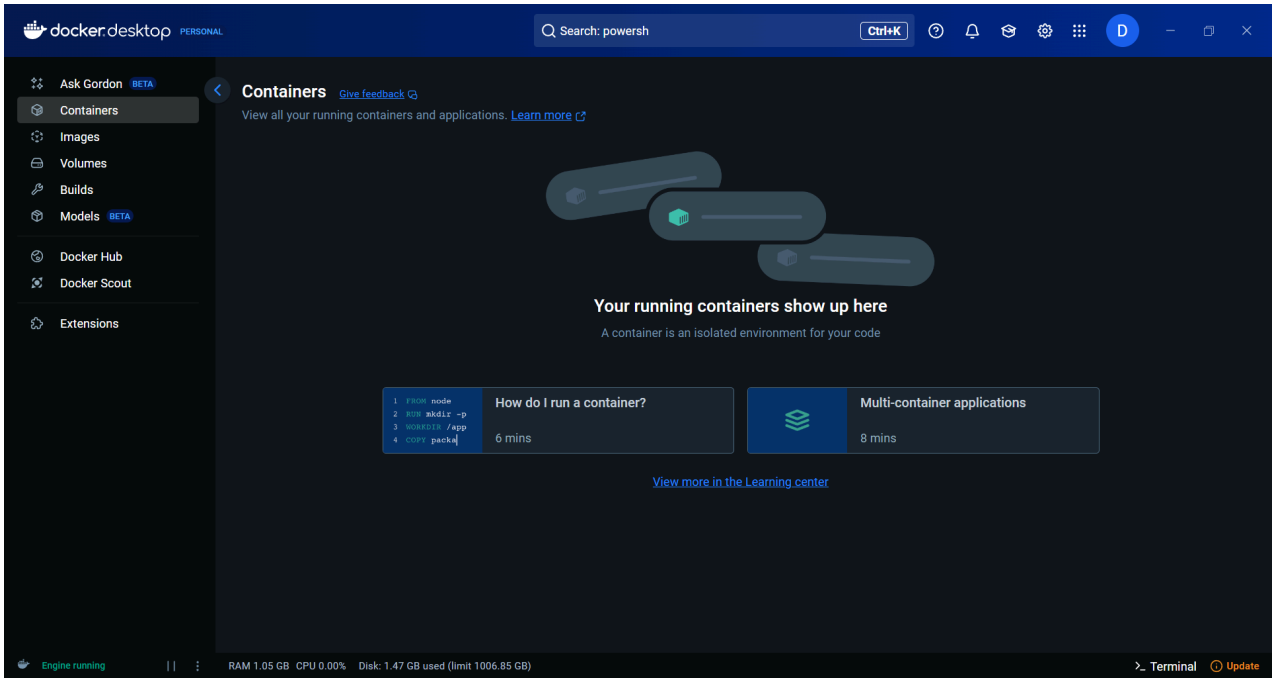
Projet Big Data : Configuration Docker, HDFS/MapReduce & Analyse Spark

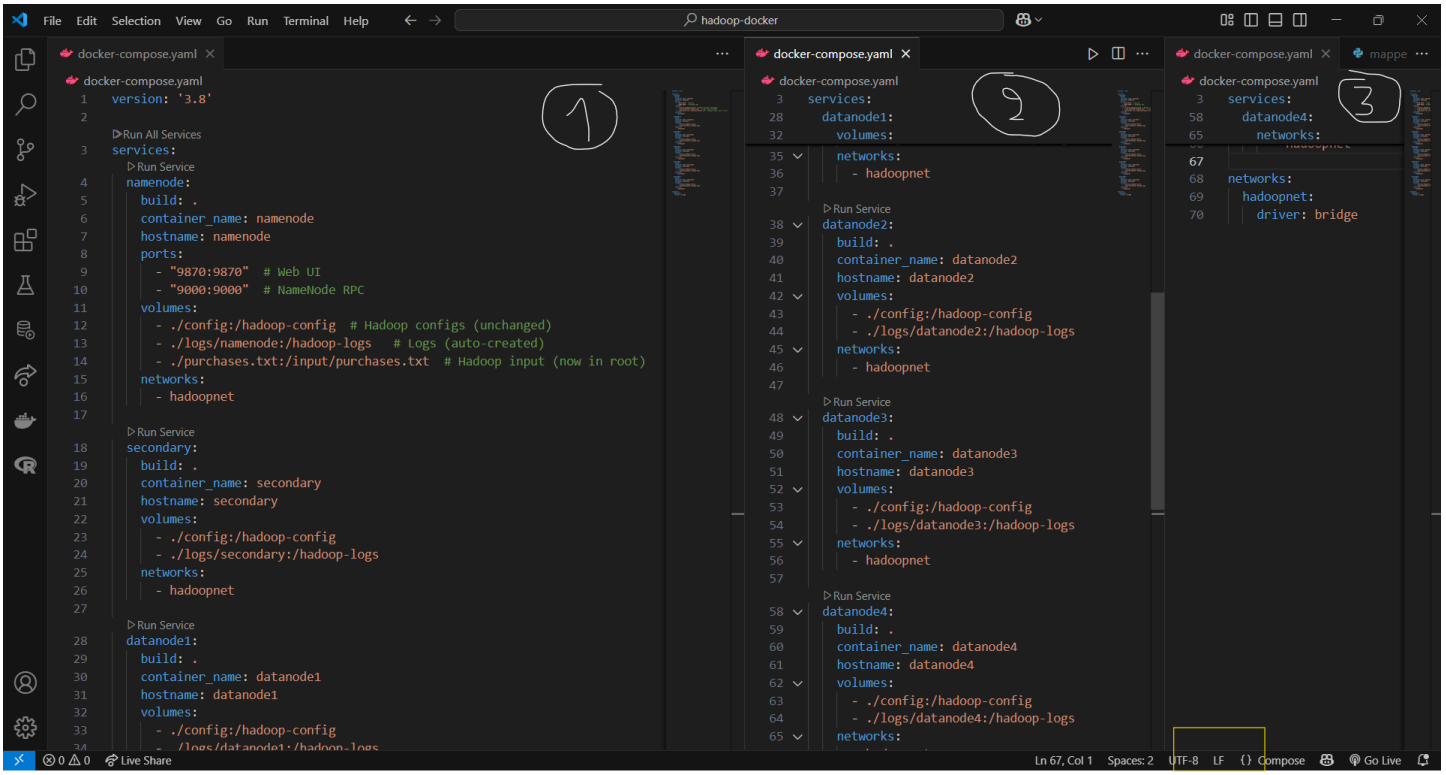
Fait par : Doua Ben Rejeb

1. Mise en place du cluster Hadoop sur Docker (+ YARN)

1.1. Installation & configuration de Docker Desktop sur Windows



docker-compose.yaml :



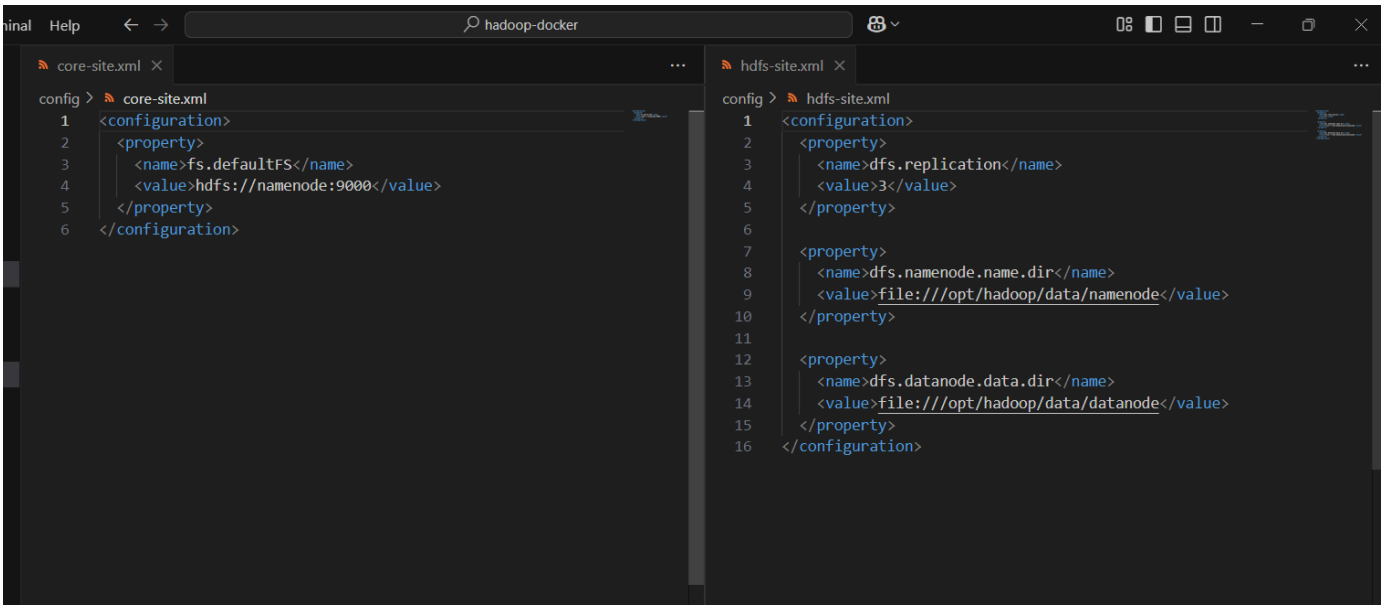
1.2 Fichiers de configuration de HDFS et YARN

La structure du dossier :

hadoop-docker/

```
|-- Dockerfile
|-- docker-compose.yml
|-- entrypoint.sh
|-- hadoop-3.3.6.tar.gz   #fichier local pour ne pas ré-installer hadoop à chaque build
`-- config
  |-- core-site.xml
  |-- hadoop-env.sh
  |-- hdfs-site.xml
  |-- mapred-site.xml
  |-- workers
  `-- yarn-site.xml
```

Attention: Les fichiers de configuration (ex: entrypoint.sh) doivent utiliser **LF** pour éviter les erreurs d'interprétation dans les environnements Linux (Docker/Spark). Les CRLF (Windows) peuvent bloquer l'exécution des scripts, car Linux ne reconnaît pas le retour chariot (\r), causant des erreurs comme command not found.



File Edit Selection View Go Run Terminal Help hadoop-docker

yarn-site.xml | hadoop-env.sh

config > hadoop-env.sh

```bash
# Set Hadoop-specific environment variables here.

# The java implementation to use.
export JAVA_HOME=/usr/local/openjdk-8
export HADOOP_OPTS="$HADOOP_OPTS -Djava.security.krb5.realm= -Djava.security.krb5.kdc="
export HADOOP_OPTS="$HADOOP_OPTS -Dhadoop.security.authentication=simple"
# The jsvc implementation to use. Used for secure datanodes.
# export JSVC_HOME=${JSVC_HOME}

# The maximum amount of heap to use, in MB. Default is 1000.
# export HADOOP_HEAPSIZE=1000

# Extra Java runtime options.
# These can affect performance and debugging.
export HADOOP_OPTS="$HADOOP_OPTS -Djava.net.preferIPv4Stack=true"

# Where log files are stored. $HADOOP_HOME/logs by default.
# export HADOOP_LOG_DIR=${HADOOP_HOME}/logs

# Where log files are stored in the secure data environment.
# export HADOOP_SECURE_LOG_DIR=${HADOOP_HOME}/logs

# File naming remote slave hosts. Default is ${HADOOP_HOME}/etc/hadoop/slaves.
# export HADOOP_SLAVES=${HADOOP_HOME}/etc/hadoop/slaves

# Extra ssh options.
# export HADOOP_SSH_OPTS="-o ConnectTimeout=1 -o SendEnv=HADOOP_CONF_DIR"

# Where to store process pid files. /tmp by default.
# export HADOOP_PID_DIR=/var/hadoop/pids

# History server options.
# export HADOOP_HISTORYSERVER_HEAPSIZE=1000

# Options for the various daemons.
# export HDFS_NAMENODE_OPTS=""
# export HDFS_DATANODE_OPTS=""
```

0 0 Live Share Ln 11, Col 30 Spaces: 4 UTF-8 LF {} Shell Script Go Live

Go Run Terminal Help hadoop-docker

entrypoint.sh

```bash
#!/bin/bash

# Copy Hadoop configuration files
cp -r /hadoop-config/* $HADOOP_HOME/etc/hadoop/

echo "Starting Hadoop on $HOSTNAME..."

# Format NameNode if needed
if [[ "$HOSTNAME" == "namenode" && ! -d "/hadoop/data/dfs/name/current" ]]; then
  echo "Formatting NameNode..."
  hdfs namenode -format -force
  sleep 5
fi

# Start services by role
case $HOSTNAME in
  namenode)
    echo "Starting NameNode..."
    hdfs namenode &
    echo "Starting ResourceMaAnager..."
    yarn resourcemanager &
    ;;

  secondary)
    echo "Starting SecondaryNameNode..."
    hdfs secondarynamenode &
    ;;

  datanode*)
    echo "Starting DataNode..."
    hdfs datanode &
    echo "Starting NodeManager..."
    yarn nodemanager &
    ;;

  *)
    echo "Unknown hostname: $HOSTNAME"
```

Dockerfile | docker-compose.yaml | entrypoint.sh

```bash
case $HOSTNAME in
  datanode*)
    ;;

  *)
    echo "Unknown hostname: $HOSTNAME"
    ;;
esac

# Keep container running
tail -f /dev/null
```

Ln 40, Col 1 Spaces: 2 UTF-8 LF {} Shell Script Go Live

```dockerfile
# Use OpenJDK 8 as base image
FROM openjdk:8-jdk

# Set environment variables
ENV HADOOP_VERSION 3.3.6
ENV HADOOP_HOME /opt/hadoop
ENV JAVA_HOME /usr/local/openjdk-8
ENV PATH $PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin

# Install dependencies (SSH, Python, etc.)
RUN apt-get update && \
    apt-get install -y ssh rsync python3 && \
    apt-get clean && \
    rm -rf /var/lib/apt/lists/*

# Create required directories
RUN mkdir -p /opt && \
    mkdir -p /hadoop-logs /hadoop-data /input

# Copy Hadoop tarball from project root (now moved there) and extract
COPY hadoop-${HADOOP_VERSION}.tar.gz /tmp/
RUN tar -xzf /tmp/hadoop-${HADOOP_VERSION}.tar.gz -C /opt && \
    ln -s /opt/hadoop-${HADOOP_VERSION} /opt/hadoop && \
    rm /tmp/hadoop-${HADOOP_VERSION}.tar.gz

# Copy entrypoint script (now in root)
COPY entrypoint.sh /entrypoint.sh
RUN chmod +x /entrypoint.sh

# Copy mapper.py and reducer.py (now in root)
COPY mapper.py /mapper.py
COPY reducer.py /reducer.py

# Copy purchases.txt (now in root) to container's /input/
COPY purchases.txt /input/purchases.txt

# Set entrypoint
```

```dockerfile
COPY purchases.txt /input/purchases.txt

# Set entrypoint
CMD ["/entrypoint.sh"]
```

→ Tous ces fichiers sont sauvegardés dans le dossier du projet (hadoop-docker).

Note : Pour la sécurité, j'ai utilisé une authentification basée sur SSH.

Configuration des clés SSH :

- **Dockerfile** :
  - Ajoute la clé publique générée à authorized_keys pour un SSH sans mot de passe..
- **Tous les conteneurs partagent la même image** :
  Comme les services (namenode, datanode1-4) utilisent build: ., ils héritent des mêmes clés SSH.
  → Permet une confiance mutuelle entre les nœuds.
- **Réseau Docker interne** :
  Les conteneurs communiquent via le réseau hadoop-net, et les connexions SSH utilisent leurs noms d'hôtes (ex: datanode1, namenode).

→ S'assurer du bon fonctionnement des conteneurs :

docker run hello-world

```
Administrator: Windows Powe
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\MSI> docker --version
Docker version 28.1.1, build 4eba377
PS C:\Users\MSI> docker run hello-world

Hello from Docker!
This message shows that your installation appears to be working correctly.

To generate this message, Docker took the following steps:
 1. The Docker client contacted the Docker daemon.
 2. The Docker daemon pulled the "hello-world" image from the Docker Hub.
    (amd64)
 3. The Docker daemon created a new container from that image which runs the
    executable that produces the output you are currently reading.
 4. The Docker daemon streamed that output to the Docker client, which sent it
    to your terminal.

To try something more ambitious, you can run an Ubuntu container with:
 $ docker run -it ubuntu bash

Share images, automate workflows, and more with a free Docker ID:
 https://hub.docker.com/

For more examples and ideas, visit:
 https://docs.docker.com/get-started/

PS C:\Users\MSI>
```

1.3 Créer l'image Hadoop

Pour créer l'image hadoop "hadoop-cluster", on ouvre **PowerShell** en tant qu' administrateur et on exécute les commandes :

cd C:\hadoop-docker
docker build -t hadoop-cluster . #It creates a custom Docker image using the instructions in the Dockerfile. This image will contain Hadoop, SSH, and all configurations needed for my cluster.

```
PS C:\hadoop-docker> docker build -t hadoop-cluster .
[+] Building 1.8s (14/14) FINISHED                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                docker:desktop-linux
 => [internal] load build definition from Dockerfile                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => => transferring dockerfile: 1.12kB                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => [internal] load metadata for docker.io/library/openjdk:8-jdk                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                1.5s
 => [auth] library/openjdk:pull token for registry-1.docker.io                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => [internal] load .dockerignore                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => => transferring context: 2B                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => [1/8] FROM docker.io/library/openjdk:8-jdk@sha256:86e863cc57215cfb181bd319736d0baf625fe8f150577f9eb58bd937f5452cb8                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => => resolve docker.io/library/openjdk:8-jdk@sha256:86e863cc57215cfb181bd319736d0baf625fe8f150577f9eb58bd937f5452cb8                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => [internal] load build context                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => => transferring context: 174B                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => CACHED [2/8] RUN apt-get update &&     apt-get install -y ssh rsync wget &&     mkdir -p /opt /hadoop-logs /hadoop-data /tmp &&     wget -qO /tmp/hadoop-3.3.6.  0.0s
 => CACHED [3/8] COPY config/core-site.xml  /opt/hadoop/etc/hadoop/                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => CACHED [4/8] COPY config/hdfs-site.xml  /opt/hadoop/etc/hadoop/                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => CACHED [5/8] COPY config/workers       /opt/hadoop/etc/hadoop/                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => CACHED [6/8] RUN mkdir -p /root/.ssh && chmod 700 /root/.ssh &&     echo "StrictHostKeyChecking no" >> /etc/ssh/ssh_config                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => CACHED [7/8] COPY entrypoint.sh /entrypoint.sh                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => CACHED [8/8] RUN chmod +x /entrypoint.sh                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => exporting to image                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.1s
 => => exporting layers                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => => exporting manifest sha256:122efa3eb14cad919a64cc6780eb31a9f9850218c72d4c3522f670cca7fce3af                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => => exporting config sha256:c127d659f9ec86d037d08c96a6266949f0f27ddfffc8f5de99ce220ab9bcf1d4                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => => exporting attestation manifest sha256:661a685ec2eccdd708df670c49e45855d0cacd20f23e8a9990152fb2c4107196                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => => exporting manifest list sha256:5d36bc096735a4c74dfe62f37f64db1117d697a5b3549b4648b323dcc867bd9e                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => => naming to docker.io/library/hadoop-cluster:latest                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
 => => unpacking to docker.io/library/hadoop-cluster:latest                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                0.0s
PS C:\hadoop-docker>
```

R.Q : Lorsqu'on définit des containers dans le même dockerfile, ils sont par défaut dans le même docker network (donc peuvent communiquer entre eux).

Rebuild and start
docker-compose build

```
Windows PowerShell
t found
PS C:\hadoop-docker> docker ps
CONTAINER ID   IMAGE     COMMAND   CREATED   STATUS    PORTS     NAMES
PS C:\hadoop-docker> docker-compose build --no-cache
Compose can now delegate builds to bake for better performance.
 To do so, set COMPOSE_BAKE=true.
[+] Building 0.0s (0/0)  docker:desktop-linux
2025/05/08 12:14:55 http2: server: error reading preface from client //./pipe/dockerDesktopLinuxEngine: file has already[+] Building 42.7s (18/36)
[+] Building 804.5s (34/46)                                                                                    docker:desktop-linux
 => [datanode1 internal] load build definition from Dockerfile                                                                  0.0s
 => => transferring dockerfile: 1.25kB                                                                                          0.0s
 => [datanode3 internal] load build definition from Dockerfile                                                                  0.0s
 => => transferring dockerfile: 1.25kB                                                                                          0.0s
 => [datanode2 internal] load build definition from Dockerfile                                                                  0.0s
 => => transferring dockerfile: 1.25kB                                                                                          0.0s
 => [datanode4 internal] load metadata for docker.io/library/ubuntu:22.04                                                       1.1s
 => [namenode internal] load build definition from Dockerfile                                                                   0.0s
 => => transferring dockerfile: 1.25kB                                                                                          0.0s
 => [datanode4 internal] load build definition from Dockerfile                                                                  0.0s
 => => transferring dockerfile: 1.25kB                                                                                          0.0s
 => [datanode1 auth] library/ubuntu:pull token for registry-1.docker.io                                                         0.0s
 => [namenode internal] load .dockerignore                                                                                      0.0s
 => => transferring context: 2B                                                                                                 0.0s
 => [datanode1 internal] load .dockerignore                                                                                     0.0s
 => => transferring context: 2B                                                                                                 0.0s
 => [datanode2 internal] load .dockerignore                                                                                     0.0s
 => => transferring context: 2B                                                                                                 0.0s
 => [datanode3 internal] load .dockerignore                                                                                     0.0s
 => => transferring context: 2B                                                                                                 0.0s
 => [datanode4 internal] load .dockerignore                                                                                     0.0s
 => => transferring context: 2B                                                                                                 0.0s
 => CACHED [datanode2 1/7] FROM docker.io/library/ubuntu:22.04@sha256:67cadaff1dca187079fce41360d5a7eb6f7dcd3745e53c79ad5efd8563118240  0.0s
 => => resolve docker.io/library/ubuntu:22.04@sha256:67cadaff1dca187079fce41360d5a7eb6f7dcd3745e53c79ad5efd8563118240           0.0s
 => [datanode4 internal] load build context                                                                                     0.0s
 => => transferring context: 458B                                                                                               0.0s
 => [namenode internal] load build context                                                                                      0.0s
 => => transferring context: 458B                                                                                               0.0s
 => [datanode2 internal] load build context                                                                                     0.0s
 => => transferring context: 458B                                                                                               0.0s
 => [datanode1 internal] load build context                                                                                     0.0s
 => => transferring context: 458B                                                                                               0.0s
```

```
Windows PowerShell
 => => exporting layers
 => => exporting manifest sha256:be3464db642297b202a1d1191c37a03619891bc033d9226b2a9c94f97169a3fd
 => => exporting config sha256:46af98f07c8a5af81953dd6f04feab65f992faa00570f25ca9ec747199e835bd
 => => exporting attestation manifest sha256:35347df001a1bc2a418b3cef8af2192d48e171ddc6c77e929ad16b7727d4e202
 => => exporting manifest list sha256:3bcc6b373e7a5f2857c0adf447ffc7ed10d24110c0316d12791446defe277eb7
 => => naming to docker.io/library/hadoop-docker-datanode2:latest
 => => unpacking to docker.io/library/hadoop-docker-datanode2:latest
 => [datanode4] exporting to image
 => => exporting layers
 => => exporting manifest sha256:2f5194eeed45f555154dd310a3bcd2ac6fa4760459c41af00e0c466b27c83314
 => => exporting config sha256:80229be4a48153b323190a7f2702e825bf8103c1ced3dd50cb66e49bf19df639
 => => exporting attestation manifest sha256:64d96839fb990084d954c57fa15aae15fbd7a6e0526febfebdf30c5b58ef865b
 => => exporting manifest list sha256:3d1a5192366063e205a06b7bb513eef60bbd26c53a95d87e0fa9712b5ceb2fc2
 => => naming to docker.io/library/hadoop-docker-datanode4:latest
 => => unpacking to docker.io/library/hadoop-docker-datanode4:latest
 => [datanode1] exporting to image
 => => exporting layers
 => => exporting manifest sha256:d79537819d449dc72c64bd835431c410626dbdfb2600cf2313b38c51ca80d5f7
 => => exporting config sha256:9aa08419e9211284b00e2d9195f44aa7b74d5842ad730e7cb909abac7ff670ba
 => => exporting attestation manifest sha256:0f8b412dcb2741f095fbe81c020f82cd35d30afe8324b4c1e8d0f5039c3e5442
 => => exporting manifest list sha256:6b3d4fad78ac0b02d1385bc8f723fbe63ce2c767af67990c81ff197f0a5f27b5
 => => naming to docker.io/library/hadoop-docker-datanode1:latest
 => => unpacking to docker.io/library/hadoop-docker-datanode1:latest
 => [namenode] exporting to image
 => => exporting layers
 => => exporting manifest sha256:56b9b1d44561dc92f5ce9c35e1476056ecdf443301d9a835a5d9d16959e7ad49
 => => exporting config sha256:5edbfda70d49d78dd49fa8f8d0ce16b739975e392a73a91de8a5a690f8eb2de3
 => => exporting attestation manifest sha256:bff50892956c4395c0a363fef8f3ea65b2732de159ee81beed4884b060da43c6
 => => exporting manifest list sha256:f107ca2da9fb007aea0473c95b3ad8d95000cd38ef2fb85a4d62a16cbb25a629
 => => naming to docker.io/library/hadoop-docker-namenode:latest
 => => unpacking to docker.io/library/hadoop-docker-namenode:latest
 => [secondary] resolving provenance for metadata file
 => [datanode1] resolving provenance for metadata file
 => [datanode3] resolving provenance for metadata file
 => [datanode4] resolving provenance for metadata file
 => [datanode2] resolving provenance for metadata file
 => [namenode] resolving provenance for metadata file
[+] Building 6/6
 ✔datanode1  Built
 ✔datanode2  Built
 ✔datanode3  Built
 ✔datanode4  Built
 ✔namenode   Built
 ✔secondary  Built
PS C:\hadoop-docker>
```

Pour démarrer notre cluster nous devons démarrer tous les containers:

`docker-compose up -d`

Windows PowerShell

```
 => => naming to docker.io/library/hadoop-docker-datanode1:latest                                                    0.0s
 => => unpacking to docker.io/library/hadoop-docker-datanode1:latest                                                 0.0s
 => [namenode] exporting to image                                                                                    0.3s
 => => exporting layers                                                                                              0.0s
 => => exporting manifest sha256:6d077859c3a2cc048c834c0444e964ab7c0d6752f12c629960502b77c8f8d034                    0.0s
 => => exporting config sha256:2cd985c228e3386e789571de054a4e7d5f77899b0393d26e09e1254d223ace48                      0.0s
 => => exporting attestation manifest sha256:e401f47068975a9ae8d822d6d2045ee7355cf8383c276318f711c27ea663225d        0.1s
 => => exporting manifest list sha256:061145a90c1b696a10ea8e3728b53828f0d19d52c5a3642a0b6f9d73aa2de10c               0.0s
 => => naming to docker.io/library/hadoop-docker-namenode:latest                                                     0.0s
 => => unpacking to docker.io/library/hadoop-docker-namenode:latest                                                  0.0s
 => [secondary] exporting to image                                                                                   0.3s
 => => exporting layers                                                                                              0.0s
 => => exporting manifest sha256:fd3cb852f559776daa143a6bf6fd0974760108fc32da86de2fcfa010a853459e                    0.0s
 => => exporting config sha256:090c3ccf7aa38ea6445c42e43f1749589c3217231a23b0188bd723f333b4f068                      0.0s
 => => exporting attestation manifest sha256:7fd4345c1484fb492bb3dea02db355d18cfe845a0b0dc2b0ea3522f576592853        0.1s
 => => exporting manifest list sha256:61713f376e49f9ae8080852ef222515256d8f86f43ee0c59b6a876ae9005337b               0.0s
 => => naming to docker.io/library/hadoop-docker-secondary:latest                                                    0.0s
 => => unpacking to docker.io/library/hadoop-docker-secondary:latest                                                 0.0s
 => [datanode3] resolving provenance for metadata file                                                               0.2s
 => [datanode4] resolving provenance for metadata file                                                               0.1s
 => [datanode2] resolving provenance for metadata file                                                               0.1s
 => [datanode1] resolving provenance for metadata file                                                               0.1s
 => [namenode] resolving provenance for metadata file                                                                0.0s
 => [secondary] resolving provenance for metadata file                                                               0.0s
[+] Running 19/19
 ✔datanode1                              Built                                                                       0.0s
 ✔datanode2                              Built                                                                       0.0s
 ✔datanode3                              Built                                                                       0.0s
 ✔datanode4                              Built                                                                       0.0s
 ✔namenode                               Built                                                                       0.0s
 ✔secondary                              Built                                                                       0.0s
 ✔Network hadoop-docker_hadoop-net       Created                                                                     0.1s
 ✔Volume "hadoop-docker_secondary_data"  Created                                                                     0.0s
 ✔Volume "hadoop-docker_datanode1_data"  Created                                                                     0.0s
 ✔Volume "hadoop-docker_datanode2_data"  Created                                                                     0.0s
 ✔Volume "hadoop-docker_datanode3_data"  Created                                                                     0.0s
 ✔Volume "hadoop-docker_datanode4_data"  Created                                                                     0.0s
 ✔Volume "hadoop-docker_namenode_data"   Created                                                                     0.0s
 ✔Container namenode                     Started                                                                     1.0s
 ✔Container datanode2                    Started                                                                     0.9s
 ✔Container secondary                    Started                                                                     0.9s
 ✔Container datanode1                    Started                                                                     0.7s
 ✔Container datanode3                    Started                                                                     0.8s
 ✔Container datanode4                    Started                                                                     0.9s
PS C:\hadoop-docker>
```

docker desktop PERSONAL

Search Ctrl+K

- Ask Gordon BETA
- Containers
- Images
- Volumes
- Builds
- Models BETA
- Docker Hub
- Docker Scout
- Extensions

Containers

Give feedback

View all your running containers and applications. Learn more

Container CPU usage: 1.30% / 1200% (12 CPUs available)

Container memory usage: 1.15GB / 7.44GB

Show charts

Search

Only show running containers

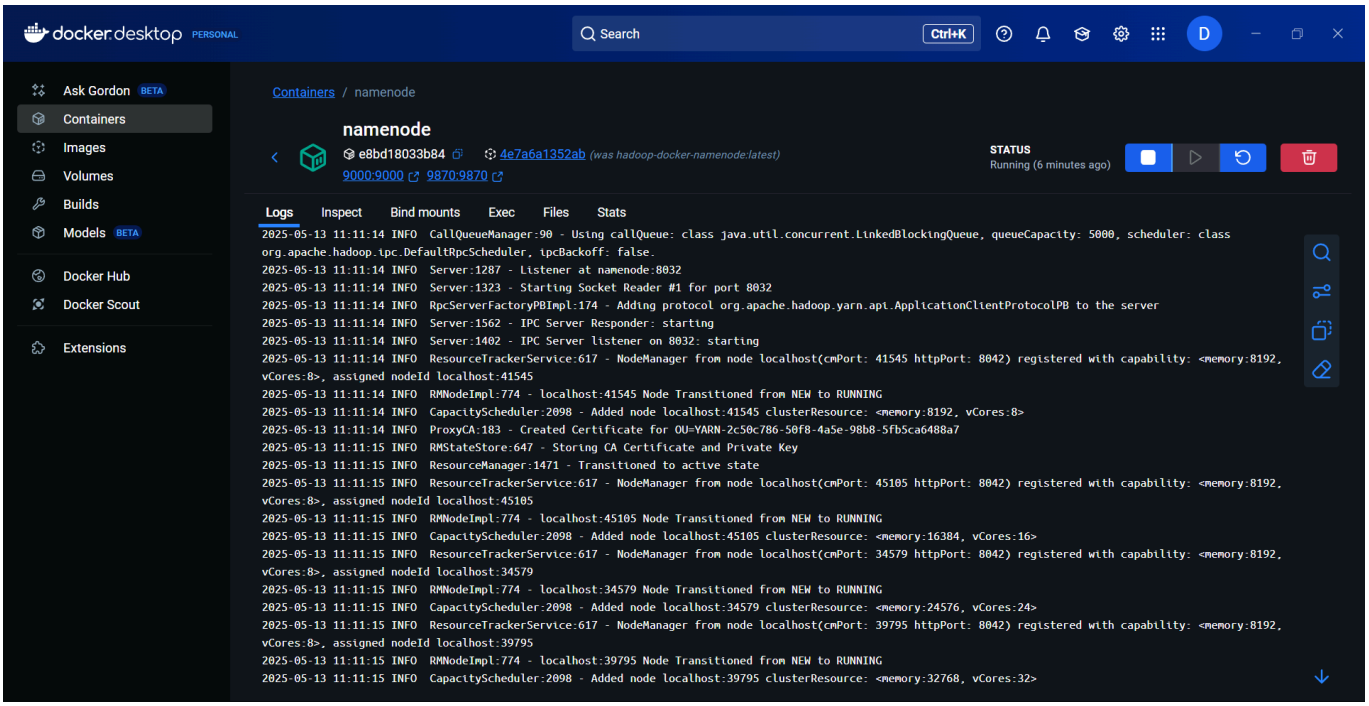
Name	Container ID	Image	Port(s)	CPU (%)	Last started	Actions
hadoop-docker	-	-	-	1.24%	44 seconds ago	
datanode3	eed99eb37966	hadoop-docker-datanode3		0.28%	44 seconds ago	
datanode1	ae45e23a62f8	hadoop-docker-datanode1		0.4%	44 seconds ago	
datanode4	57757b5e07b0	hadoop-docker-datanode4		0.23%	44 seconds ago	
datanode2	f944b69aa6f8	hadoop-docker-datanode2		0.25%	44 seconds ago	
secondary	28a9430bb652	hadoop-docker-secondary	9868:9868	0.08%	44 seconds ago	
namenode	42d2043dcb56	hadoop-docker-namenode	8088:8088 Show all ports (2)	0%	44 seconds ago	

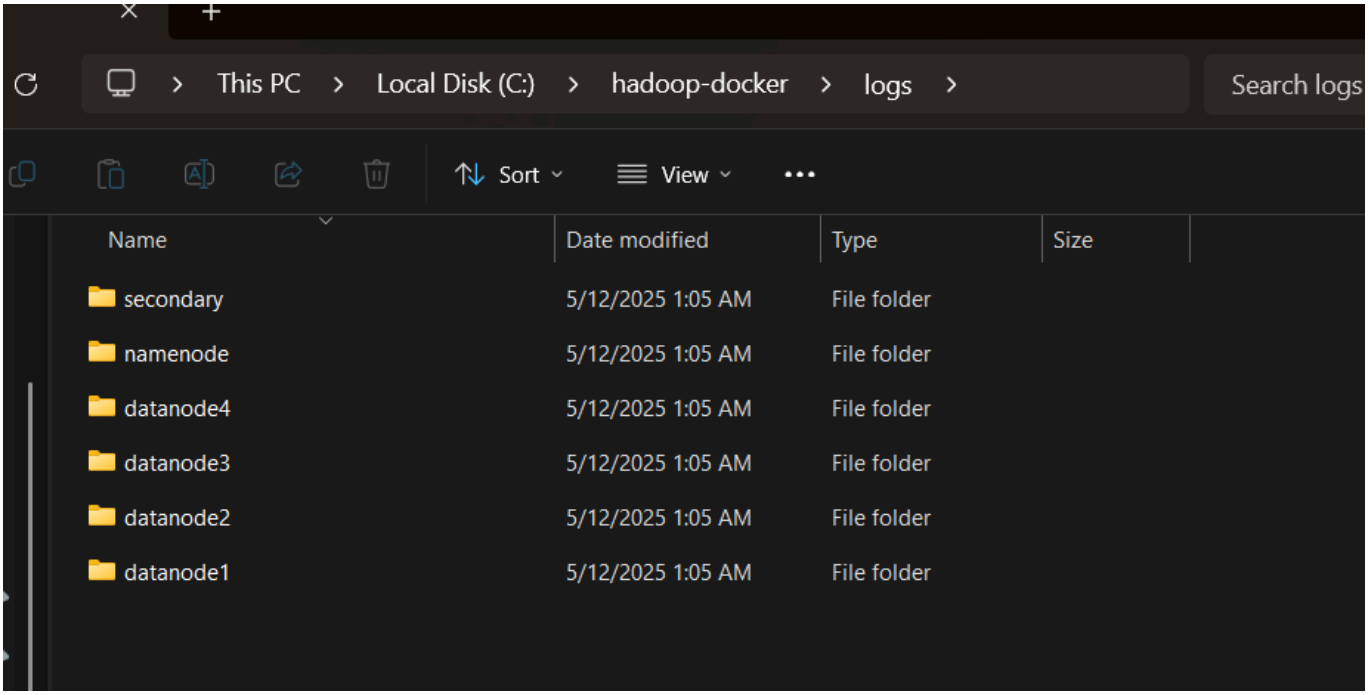
Showing 7 items

Engine running | RAM 4.73 GB CPU 0.00% Disk: 17.19 GB used (limit 1006.85 GB) | Terminal | Update

Search

12:35 AM 5/9/2025





1.4 Validation de HDFS

- **Check Containers**: docker ps

```
 ✔Container datanode1                    Started
 ✔Container datanode3                    Started
 ✔Container datanode4                    Started
PS C:\hadoop-docker> docker ps
CONTAINER ID   IMAGE                    COMMAND            CREATED         STATUS         PORTS                                                   NAMES
f944b69aa6f8   hadoop-docker-datanode2  "/entrypoint.sh"   3 minutes ago   Up 3 minutes                                                           datanode2
42d2043dcb56   hadoop-docker-namenode   "/entrypoint.sh"   3 minutes ago   Up 3 minutes   0.0.0.0:8088->8088/tcp, 0.0.0.0:9870->9870/tcp   namenode
28a9430bb652   hadoop-docker-secondary  "/entrypoint.sh"   3 minutes ago   Up 3 minutes   0.0.0.0:9868->9868/tcp                                  secondary
eed99eb37966   hadoop-docker-datanode3  "/entrypoint.sh"   3 minutes ago   Up 3 minutes                                                           datanode3
57757b5e07b0   hadoop-docker-datanode4  "/entrypoint.sh"   3 minutes ago   Up 3 minutes                                                           datanode4
ae45e23a62f8   hadoop-docker-datanode1  "/entrypoint.sh"   3 minutes ago   Up 3 minutes                                                           datanode1
PS C:\hadoop-docker>
```

Ces commandes liste **tous les processus Java** qui ont commencé dans chacun des conteneurs:

- docker exec -it namenode jps
- docker exec -it datanode1 jps
- docker exec -it secondary jps

```
PS C:\hadoop-docker> docker exec -it datanode1 jps
9 DataNode
93 Jps
PS C:\hadoop-docker> docker exec -it datanode3 jps
8 DataNode
92 Jps
PS C:\hadoop-docker> docker exec -it secondary jps
72 Jps
8 SecondaryNameNode
PS C:\hadoop-docker>
```

Pour vérifier que le cluster est bien mis en place, nous pouvons nous connecter au namenode et rentrer une commande permettant de vérifier l'état du cluster :

docker exec -it namenode hdfs dfsadmin -report

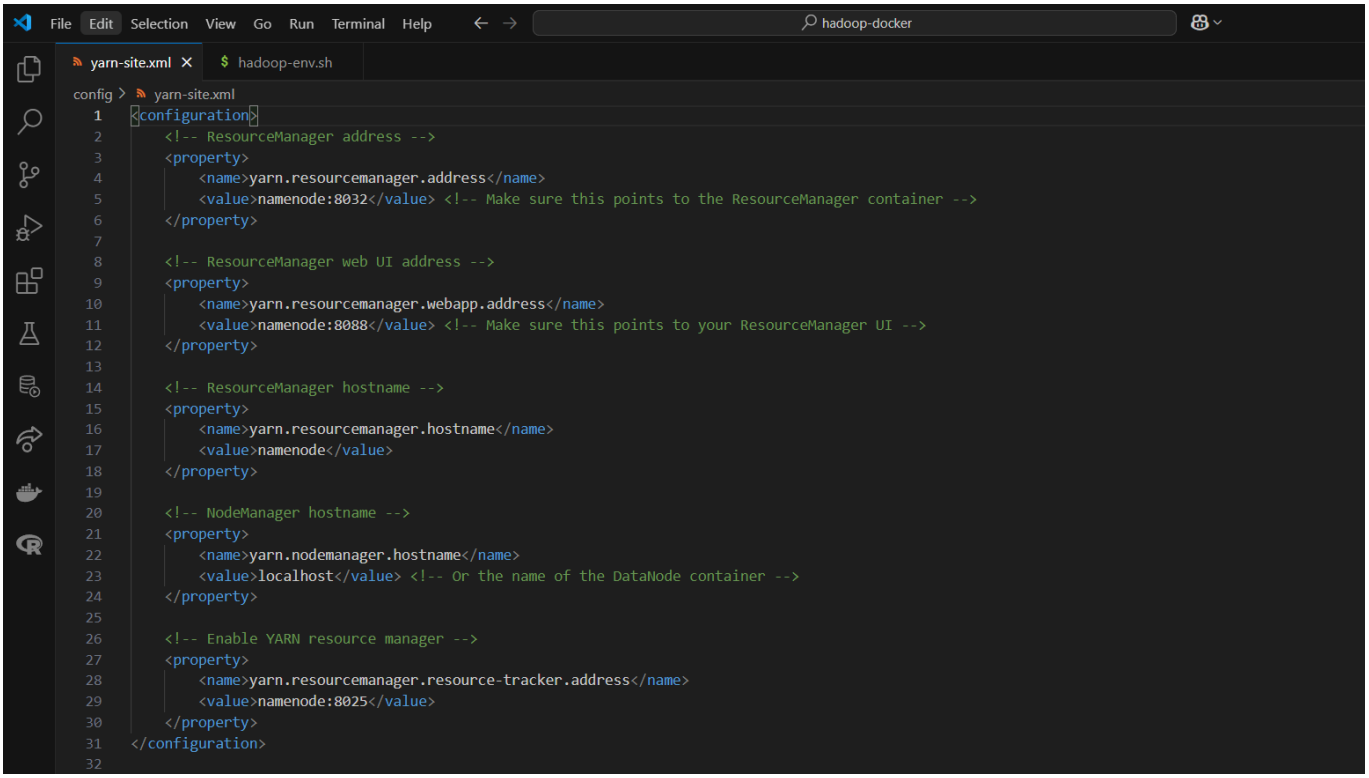
```
PS C:\hadoop-docker> docker exec -it namenode hdfs dfsadmin -report
Configured Capacity: 4324404707328 (3.93 TB)
Present Capacity: 3984678338560 (3.62 TB)
DFS Remaining: 3984678240256 (3.62 TB)
DFS Used: 98304 (96 KB)
DFS Used%: 0.00%
Replicated Blocks:
        Under replicated blocks: 0
        Blocks with corrupt replicas: 0
        Missing blocks: 0
        Missing blocks (with replication factor 1): 0
        Low redundancy blocks with highest priority to recover: 0
        Pending deletion blocks: 0
Erasure Coded Block Groups:
        Low redundancy block groups: 0
        Block groups with corrupt internal blocks: 0
        Missing block groups: 0
        Low redundancy blocks with highest priority to recover: 0
        Pending deletion blocks: 0

-------------------------------------------------
Live datanodes (4):

Name: 172.19.0.2:9866 (datanode1.hadoop-docker_hadoopnet)
Hostname: datanode1
Decommission Status : Normal
Configured Capacity: 1081101176832 (1006.85 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 29939236864 (27.88 GB)
DFS Remaining: 996169560064 (927.76 GB)
DFS Used%: 0.00%
DFS Remaining%: 92.14%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 0
Last contact: Tue May 13 00:59:51 UTC 2025
```

```
Windows PowerShell

Last contact: Tue May 13 00:59:51 UTC 2025
Last Block Report: Tue May 13 00:58:18 UTC 2025
Num of Blocks: 0


Name: 172.19.0.4:9866 (datanode4.hadoop-docker_hadoopnet)
Hostname: datanode4
Decommission Status : Normal
Configured Capacity: 1081101176832 (1006.85 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 29939236864 (27.88 GB)
DFS Remaining: 996169560064 (927.76 GB)
DFS Used%: 0.00%
DFS Remaining%: 92.14%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 0
Last contact: Tue May 13 00:59:51 UTC 2025
Last Block Report: Tue May 13 00:58:18 UTC 2025
Num of Blocks: 0


Name: 172.19.0.5:9866 (datanode2.hadoop-docker_hadoopnet)
Hostname: datanode2
Decommission Status : Normal
Configured Capacity: 1081101176832 (1006.85 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 29939236864 (27.88 GB)
DFS Remaining: 996169560064 (927.76 GB)
DFS Used%: 0.00%
DFS Remaining%: 92.14%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 0
Last contact: Tue May 13 00:59:51 UTC 2025
```

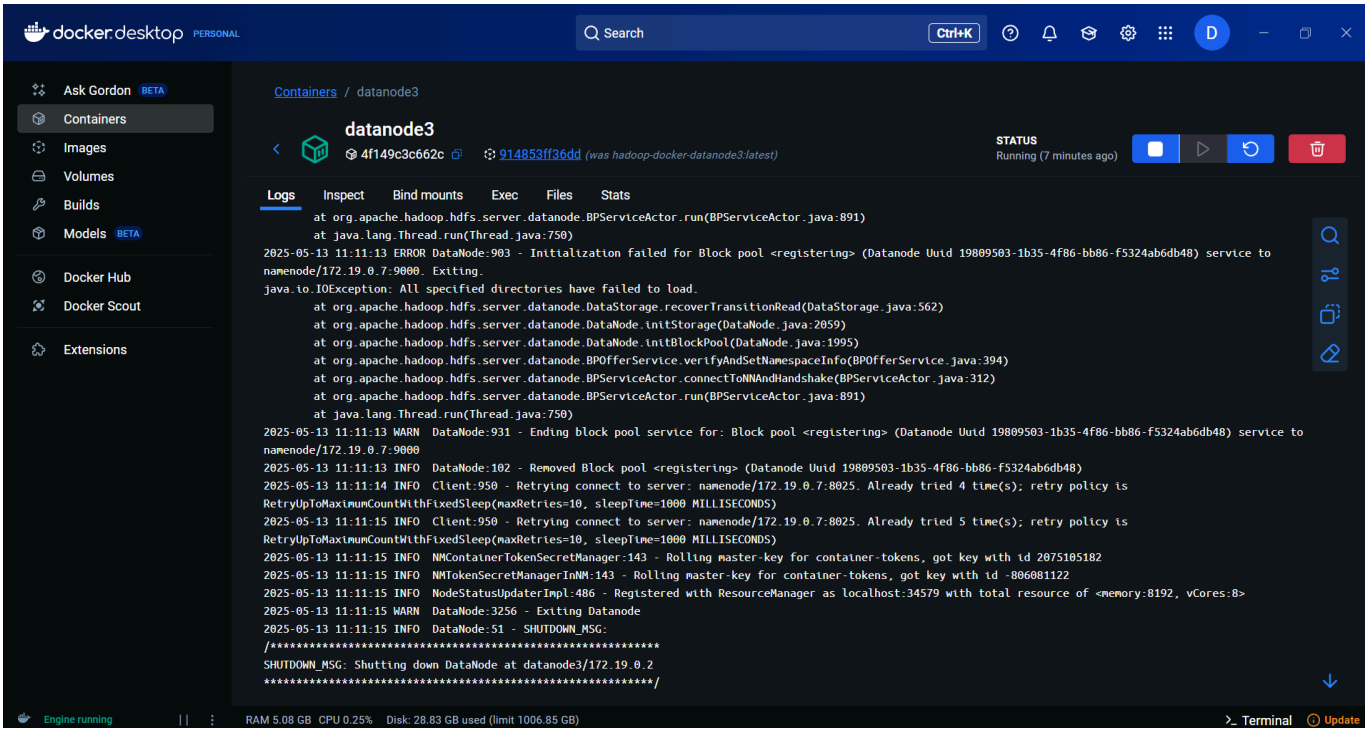
```
Last contact: Tue May 13 00:59:51 UTC 2025
Last Block Report: Tue May 13 00:58:18 UTC 2025
Num of Blocks: 0


Name: 172.19.0.6:9866 (datanode3.hadoop-docker_hadoopnet)
Hostname: datanode3
Decommission Status : Normal
Configured Capacity: 1081101176832 (1006.85 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 29939236864 (27.88 GB)
DFS Remaining: 996169560064 (927.76 GB)
DFS Used%: 0.00%
DFS Remaining%: 92.14%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 0
Last contact: Tue May 13 00:59:51 UTC 2025
Last Block Report: Tue May 13 00:58:18 UTC 2025
Num of Blocks: 0


PS C:\hadoop-docker>
```

- **Access UIs**:
  - HDFS: http://localhost:9870
  - YARN: http://localhost:8088

1.5 Validation de YARN




```xml
<configuration>
    <!-- ResourceManager address -->
    <property>
        <name>yarn.resourcemanager.address</name>
        <value>namenode:8032</value> <!-- Make sure this points to the ResourceManager container -->
    </property>

    <!-- ResourceManager web UI address -->
    <property>
        <name>yarn.resourcemanager.webapp.address</name>
        <value>namenode:8088</value> <!-- Make sure this points to your ResourceManager UI -->
    </property>

    <!-- ResourceManager hostname -->
    <property>
        <name>yarn.resourcemanager.hostname</name>
        <value>namenode</value>
    </property>

    <!-- NodeManager hostname -->
    <property>
        <name>yarn.nodemanager.hostname</name>
        <value>localhost</value> <!-- Or the name of the DataNode container -->
    </property>

    <!-- Enable YARN resource manager -->
    <property>
        <name>yarn.resourcemanager.resource-tracker.address</name>
        <value>namenode:8025</value>
    </property>
</configuration>
```



```
Administrator: Windows Powe
 => => naming to docker.io/library/hadoop-docker-secondary:latest                                                    0.0s
 => => unpacking to docker.io/library/hadoop-docker-secondary:latest                                                 0.0s
 => [datanode2] resolving provenance for metadata file                                                               0.4s
 => [datanode4] resolving provenance for metadata file                                                               0.3s
 => [datanode1] resolving provenance for metadata file                                                               0.2s
 => [datanode3] resolving provenance for metadata file                                                               0.1s
 => [secondary] resolving provenance for metadata file                                                               0.0s
[+] Building 6/6
 ✔datanode1  Built                                                                                                   0.0s
 ✔datanode2  Built                                                                                                   0.0s
 ✔datanode3  Built                                                                                                   0.0s
 ✔datanode4  Built                                                                                                   0.0s
 ✔namenode   Built                                                                                                   0.0s
 ✔secondary  Built                                                                                                   0.0s
PS C:\hadoop-docker> docker-compose up -d
time="2025-05-10T16:33:02+01:00" level=warning msg="C:\\hadoop-docker\\docker-compose.yaml: the attribute `version` is obsolete, it will be ignored, please
remove it to avoid potential confusion"
[+] Running 6/6
 ✔Container hadoop-docker-namenode-1   Started                                                                       5.3s
 ✔Container hadoop-docker-datanode4-1  Started                                                                       4.2s
 ✔Container hadoop-docker-datanode1-1  Started                                                                       4.4s
 ✔Container hadoop-docker-secondary-1  Started                                                                       4.3s
 ✔Container hadoop-docker-datanode3-1  Started                                                                       4.2s
 ✔Container hadoop-docker-datanode2-1  Started                                                                       4.1s
PS C:\hadoop-docker> docker exec -it hadoop-docker-datanode3-1 bash
hadoop@datanode3:/$ jps
99 NodeManager
250 Jps
44 DataNode
hadoop@datanode3:/$
PS C:\hadoop-docker> docker exec -it hadoop-docker-namenode-1 bash
hadoop@namenode:/$ jps
448 Jps
101 NameNode
156 ResourceManager
hadoop@namenode:/$
PS C:\hadoop-docker> docker exec -it hadoop-docker-secondary-1 bash
hadoop@secondary:/$ jps
99 Jps
44 SecondaryNameNode
hadoop@secondary:/$
```

2. Exécution du code MapReduce sur le dataset purchases.txt

2.1 Configuration MAPREDUCE

La structure du dossier :

```
hadoop-docker/
├── docker-compose.yml
├── Dockerfile
├── entrypoint.sh
├── mapper.py        # new Map script
├── reducer.py       # new Reduce script
└── config/
   `-- mapred-site.xml
```

This PC > Local Disk (C:) > hadoop-docker >

Search hadoop-docker

Sort View ••• Details

Name	Date modified	Type	Size
config	5/10/2025 6:15 PM	File folder	
logs	5/10/2025 6:15 PM	File folder	
docker-compose.yaml	5/10/2025 5:58 PM	Yaml Source File	2 KB
Dockerfile	5/10/2025 4:32 PM	File	2 KB
entrypoint.sh	5/10/2025 5:25 PM	SH Source File	2 KB
hadoop-3.3.6.tar.gz	5/10/2025 11:59 AM	Compressed Archi...	712,996 KB
mapper.py	3/21/2025 11:12 PM	Python Source File	1 KB
purchases.txt	9/9/2013 10:56 PM	Text Document	206,361 KB
reducer.py	3/21/2025 11:12 PM	Python Source File	1 KB

Go Run Terminal Help hadoop-docker

mapred-site.xml

config > mapred-site.xml

```xml
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.address</name>
    <value>namenode:10020</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.webapp.address</name>
    <value>namenode:19888</value>
  </property>
</configuration>
```

docker-compose.yaml mapper.py

mapper.py > ...

```python
#!/usr/bin/env python3
import sys

for line in sys.stdin:
    fields = line.strip().split('\t')
    if len(fields) != 6:
        continue  # Skip malformed lines
    store = fields[2]
    try:
        amount = float(fields[4])
    except ValueError:
        continue  # Skip invalid amounts
    # Emit: store, count (1), amount
    print(f"{store}\t1\t{amount}")
```

reducer.py

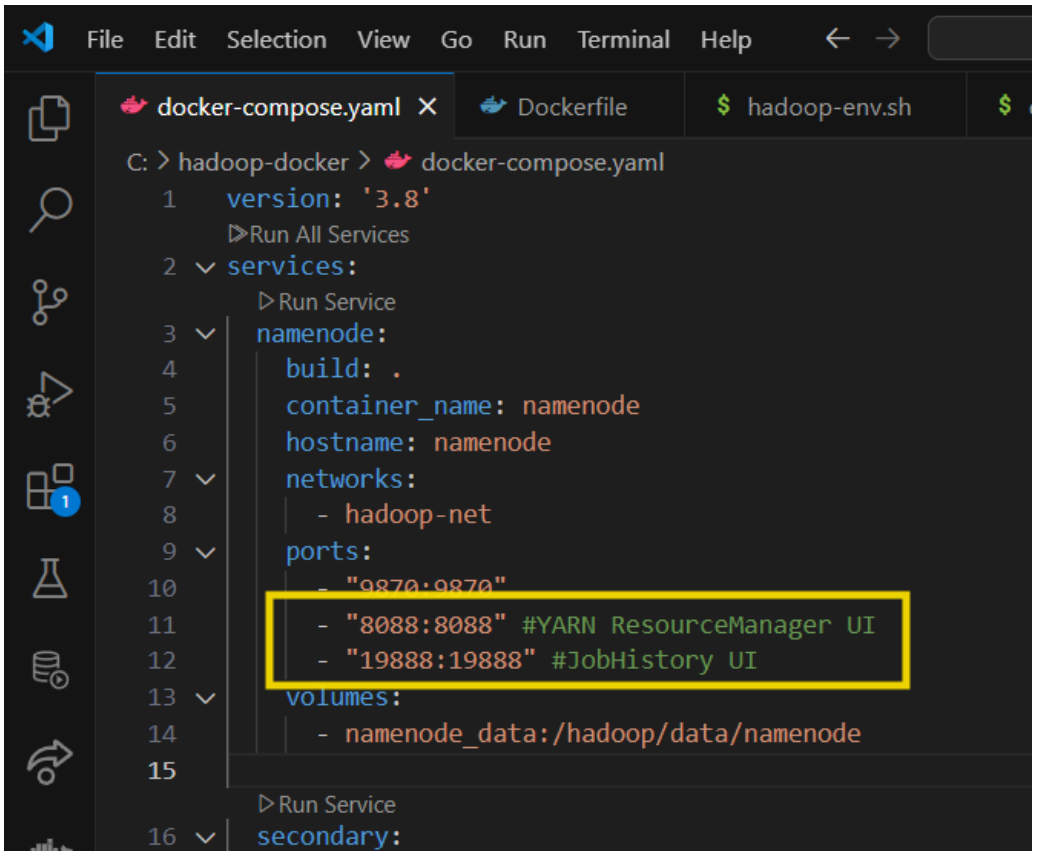
reducer.py > ...

```python
#!/usr/bin/env python3
import sys

current_store = None
total_sales = 0.0
total_count = 0

for line in sys.stdin:
    line = line.strip()
    if not line:
        continue
    # Split into exactly 3 fields
    parts = line.split('\t')
    if len(parts) != 3:
        continue  # Skip invalid lines
    store, count_str, amount_str = parts
    try:
        count = int(count_str)
        amount = float(amount_str)
    except ValueError:
        continue  # Skip invalid values

    if current_store == store:
        total_count += count
        total_sales += amount
    else:
        if current_store:
            print(f"{current_store}\t{total_count}\t{total_sales:.2f}")
        current_store = store
        total_count = count
        total_sales = amount

# Output the last store
if current_store:
    print(f"{current_store}\t{total_count}\t{total_sales:.2f}")
```

Ln 1, Col 1 Spaces: 4 UTF-8 LF {} Python 3.10.4 64-bit Go Live

2.2 Mises à jour

D'autres configurations seront (dans docker-compose.yaml & entrypoint.sh) :



```
# 5. Role-based startup
case "$HOSTNAME" in
  namenode)
    echo "==== Starting NameNode and ResourceManager on $HOSTNAME ====="
    # Format NameNode metadata directory if first run
    NN_DIR="/hadoop/data/namenode"
    if [ ! -d "$NN_DIR/current" ]; then
      echo "Formatting NameNode at $NN_DIR"
      hdfs namenode -format -force
    fi
    # Launch HDFS and YARN
    start-dfs.sh
    start-yarn.sh
    ;;

  secondary)
    echo "==== Starting SecondaryNameNode on $HOSTNAME ====="
    hdfs secondarynamenode
    ;;

  datanode*)
    echo "==== Starting DataNode and NodeManager on $HOSTNAME ====="
    hdfs datanode &
    yarn nodemanager
    ;;

  *)
    echo "Error: Unknown hostname '$HOSTNAME'"
    exit 1
    ;;
esac

# 6. Keep the container alive
exec tail -f /dev/null
```

2.3 Création de répertoire hdfs

```
# Ensure hadoop user has permission on local filesystem and HDFS
hdfs dfs -chmod -R 777 /user/hadoop
hdfs dfs -chmod -R 777 /input
```

```
# Fix HDFS directories and permissions
hdfs dfs -mkdir -p /user/hadoop/input
hdfs dfs -chown hadoop:hadoop /user/hadoop/input
hdfs dfs -chmod 755 /user/hadoop/input
```

hdfs dfs -ls /user/hadoop/input

```
Administrator: Windows Powe
PS C:\hadoop-docker> docker exec -it namenode bash
root@namenode:/# jps
64 NameNode
65 ResourceManager
680 Jps
root@namenode:/# docker cp mapper.py namenode:/mapper.py
bash: docker: command not found
root@namenode:/#
PS C:\hadoop-docker> docker cp mapper.py namenode:/mapper.py
Successfully copied 2.56kB to namenode:/mapper.py
PS C:\hadoop-docker> docker cp reducer.py namenode:/reducer.py
Successfully copied 2.56kB to namenode:/reducer.py
PS C:\hadoop-docker> docker cp purchases.txt namenode:/purchases.txt
Successfully copied 211MB to namenode:/purchases.txt
PS C:\hadoop-docker> docker exec -it namenode bash
root@namenode:/# hdfs dfs -mkdir -p /input
root@namenode:/# hdfs dfs -put input.txt /input
put: `input.txt': No such file or directory
root@namenode:/# hdfs dfs -put purchases.txt /input
root@namenode:/# hdfs dfs -ls /input
Found 1 items
-rw-r--r--   3 root supergroup  211312924 2025-05-11 21:32 /input/purchases.txt
root@namenode:/#
```

2.4 HDFS input

2.4.1 Copie du fichier vers HDFS

Avant d'exécuter hadoop streaming, il faut d'abord **télécharger purchases.csv vers HDFS**:

docker exec -it namenode hdfs dfs -put /etc/project/purchases.csv /user/hadoop/input/

```
Administrator: Windows Powe
 => [datanode4] exporting to image                                                                                  80.3s
 => => exporting layers                                                                                             39.2s
 => => exporting manifest sha256:4d7256d21983e1e11f6b2548df809b1a3d6fee39eec25e32378169ebacc81c53                    0.0s
 => => exporting config sha256:56e5bbefa5efcb6e9ba9493026eb7e2aaa189a106c257314777a9c2e02416d10                      0.0s
 => => exporting attestation manifest sha256:6122129b1ffa897b309ca4db82df6bec1dca382d8b3f7c5fc5ea8f041dc669b5        0.1s
 => => exporting manifest list sha256:b1625052d265c26a32221cc9e04c6fc1de9718a881598281f77c27201b59c15b               0.0s
 => => naming to docker.io/library/hadoop-docker-datanode4:latest                                                   0.0s
 => => unpacking to docker.io/library/hadoop-docker-datanode4:latest                                               40.7s
 => [secondary] resolving provenance for metadata file                                                              0.7s
 => [datanode2] resolving provenance for metadata file                                                              0.6s
 => [namenode] resolving provenance for metadata file                                                               0.5s
 => [datanode3] resolving provenance for metadata file                                                              0.4s
 => [datanode1] resolving provenance for metadata file                                                              0.2s
 => [datanode4] resolving provenance for metadata file                                                              0.1s
[+] Building 6/6
 ✔datanode1  Built                                                                                                  0.0s
 ✔datanode2  Built                                                                                                  0.0s
 ✔datanode3  Built                                                                                                  0.0s
 ✔datanode4  Built                                                                                                  0.0s
 ✔namenode   Built                                                                                                  0.0s
 ✔secondary  Built                                                                                                  0.0s
PS C:\hadoop-docker> docker-compose up -d
time="2025-05-12T00:59:08+01:00" level=warning msg="C:\\hadoop-docker\\docker-compose.yaml: the attribute `version` is obsolete, it will be ignored, please
remove it to avoid potential confusion"
[+] Running 7/7
 ✔Container datanode3     Started                                                                                   14.8s
 ✔Container namenode      Started                                                                                   14.9s
 ✔Container secondary     Started                                                                                   14.7s
 ✔Container spark-master  Started                                                                                   14.7s
 ✔Container datanode2     Started                                                                                   14.7s
 ✔Container datanode1     Started                                                                                   14.7s
 ✔Container datanode4     Started                                                                                   14.8s
PS C:\hadoop-docker> docker cp mapper.py namenode:/mapper.py
Successfully copied 2.05kB to namenode:/mapper.py
PS C:\hadoop-docker> docker cp reducer.py namenode:/reducer.py
Successfully copied 2.56kB to namenode:/reducer.py
PS C:\hadoop-docker> docker cp purchases.txt namenode:/purchases.txt
Successfully copied 211MB to namenode:/purchases.txt
PS C:\hadoop-docker> docker exec -it namenode bash
root@namenode:/# hdfs dfs -mkdir -p /input
```

2.4.2 Accès au conteneur NameNode

Accéder au conteneur **NameNode** avec la commande :

```
PS C:\hadoop-docker> docker exec -it namenode bash
```

2.4.3 Exécution du job Hadoop Streaming

```
hadoop jar share/hadoop/tools/lib/hadoop-streaming-*.jar \
-files /config/mapper.py,/config/reducer.py \
-mapper "python mapper.py" \
-reducer "python reducer.py" \
-input /user/hadoop/input/purchases.csv \
-output /output/purchases-summary
```

2.4.4 Consultation des résultats :

```
Albuquerque     40345   10052311.42
Anaheim 40086   10076416.36
Anchorage       39806   9933500.40
Arlington       40348   10072207.97
Atlanta 40168   9997146.70
Aurora  39808   9992970.92
Austin  40332   10057158.90
Bakersfield     40326   10031208.92
Baltimore       40196   10096521.45
Baton Rouge     40387   10131273.23
Birmingham      40253   10076606.52
Boise   40203   10039166.74
Boston  40338   10039473.28
Buffalo 40053   10001941.19
Chandler        39826   9919559.86
Charlotte       40509   10112531.34
Chesapeake      40288   10038504.92
Chicago 40220   10062522.07
Chula Vista     40080   9974951.34
Cincinnati      40452   10139505.74
Cleveland       40313   10067835.84
Colorado Springs        40389   10061105.87
Columbus        40225   10035241.03
Corpus Christi  39925   9976522.77
Dallas  40368   10066548.45
Denver  40199   10031534.87
Detroit 39999   9979260.76
Durham  40501   10153890.21
El Paso 39882   10016409.97
Fort Wayne      40439   10132594.02
Fort Worth      40336   10120830.65
Fremont 40219   10053242.36
Fresno  40043   9976260.26
Garland 40185   10071043.92
Gilbert 40277   10062115.19
Glendale        40264   10044493.97
Greensboro      40232   10033781.39
Henderson       40194   10053416.05
Hialeah 40312   10047052.76
Honolulu        39944   10006273.49
```

Administrator: Windows Powe

```
Kansas City     39713   9968118.73
Laredo  40342   10144604.98
Las Vegas       40165   10054257.98
Lexington       40343   10084510.95
Lincoln 40345   10069485.40
Long Beach      39942   10006380.25
Los Angeles     40254   10084576.80
Louisville      40099   10008566.47
Lubbock 39837   9958119.15
Madison 40177   10032035.54
Memphis 40222   10038565.32
Mesa    40207   10053642.60
Miami   39935   9947316.07
Milwaukee       40316   10064482.65
Minneapolis     39991   10011757.78
Nashville       39854   9961450.51
New Orleans     39914   9949257.75
New York        40364   10085293.55
Newark  40577   10144052.80
Norfolk 40231   10088563.17
North Las Vegas 40013   10029652.51
Oakland 39728   9947292.52
Oklahoma City   40446   10118986.25
Omaha   40209   10026642.34
Orlando 40195   10074922.52
Philadelphia    40748   10190080.26
Phoenix 40333   10079076.70
Pittsburgh      40358   10090124.82
Plano   40170   10046103.61
Portland        40065   10007635.77
Raleigh 40261   10061442.54
Reno    40254   10079955.16
Richmond        39983   9992941.59
Riverside       39963   10006695.42
Rochester       40455   10067606.92
Sacramento      40561   10123468.18
Saint Paul      40160   10057233.57
San Antonio     40197   10014441.70
San Bernardino  40041   9965152.04
San Diego       39901   9966038.39
San Francisco   39983   9995570.54
```

```
San Bernardino  40041   9965152.04
San Diego       39901   9966038.39
San Francisco   39983   9995570.54
San Jose        39898   9936721.41
Santa Ana       40306   10050309.93
Scottsdale      40173   10037929.85
Seattle 39866   9936267.37
Spokane 40222   10083362.98
St. Louis       39982   10002105.14
St. Petersburg  40093   9986495.54
Stockton        39996   10006412.64
Tampa   40136   10106428.55
Toledo  40139   10020768.88
Tucson  39870   9998252.47
Tulsa   40247   10064955.90
Virginia Beach  40169   10086553.50
Washington      40503   10139363.39
Wichita 40422   10083643.21
Winston-Salem   40208   10044011.83
root@namenode:/#
```

2.4.5 Analyse : Observations et interprétation de l'output HDFS

Chaque ligne correspond à une ville et contient 3 colonnes :

1. **Nom de la ville** (ex : *San Bernardino*, *Washington*, *Tampa*...)
2. **Valeur numérique 1** → Probablement le **nombre de transactions** ou de commandes par ville. (exp : **40 041** transactions pour San Bernardino)
3. **Valeur numérique 2** → Très probablement le **chiffre d'affaires total** (en $) (exp : ***9 965 152.04 $*** pour San Bernardino.)

<u>Observations</u>

1. **Philadelphia domine** :
   - **40 748** transactions (le plus élevé).
   - **10 139 363.39 $** de chiffre d'affaires (le plus élevé).
   - Indique une forte activité commerciale ou une population importante.
2. **Rendement par transaction** :
   - Calcul approximatif : *CA total / Nombre de transactions*.
   - Exemple pour San Bernardino :
     .
   - Les valeurs sont similaires partout (≈ *245–255 $/transaction*), suggérant des prix homogènes ou un panier moyen stable.
3. **Anomalie de données** : "**Michita**" est probablement une faute de frappe pour **Wichita** (ville du Kansas).
4. **Performances comparées** :

Ville	Transactions	Chiffre d'affaires
Philadelphia	40 748 transactions	10 190 060.26 $
Sacramento	40 561 transactions	10 123 468.18 $
Washington	40 503 transactions	10 139 3636.39 $

3. Configuration et exécution de Spark

<u>La structure du dossier :</u>

```
hadoop-docker/
    ├── docker-compose.yml
    ├── Dockerfile
    ├── entrypoint.sh
    ├── mapper.py
    ├── reducer.py
    ├── config/
    ├── logs/
    ├── spark-output/
    ├── spark-job/
    `-- purchases.txt
    ├── spark-input/
```

`-- analysis.py

Code source : Dockerfile, docker-compose.yml, scripts Spark

3.1 Fichiers de configuration SPARK

3.1.1 **Dockerfile** : Base image utilisée (ex: FROM openjdk:8).

3.1.2 **Docker Compose** :

- Définition des services (spark-master, spark-worker)
- Volumes pour la persistance des données ou partage de code.

```yaml
services:

  datanode4:
    build: .
    container_name: datanode4
    hostname: datanode4
    volumes:
      - ./config:/hadoop-config
      - ./logs/datanode4:/hadoop-logs
    networks:
      - hadoopnet

  spark:
    image: bitnami/spark:latest
    container_name: spark-master
    environment:
      - SPARK_MODE=master
    volumes:
      - ./spark-job:/app                # Spark job code (unchanged)
      - ./spark-input:/spark-input      # Spark input (purchases.txt lives here)
      - ./spark-output:/spark-output    # Spark output (auto-created)
    networks:
      - hadoopnet
    command: >
      bash -c "
        /opt/bitnami/spark/bin/spark-submit --master local[*] /app/analysis.py /spark-input/purchases.txt /spark-output/results.txt
      "

networks:
  hadoopnet:
    driver: bridge
```

3.1.3 Configuration des variables d'environnement (ex: ENV SPARK_HOME=/opt/spark)

```bash
export JAVA_HOME=/usr/local/openjdk-8
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export YARN_CONF_DIR=$HADOOP_HOME/etc/hadoop
export SPARK_LOCAL_IP=$HOSTNAME
```

3.1.4 Copie des fichiers de configuration (ex: spark-defaults.conf)

```
spark.master            yarn
spark.eventLog.enabled  true
spark.eventLog.dir      hdfs://namenode:9000/spark-logs
spark.hadoop.fs.defaultFS hdfs://namenode:9000
spark.yarn.archive      hdfs://namenode:9000/spark-jars/spark-jars.zip
spark.executor.memory   1g
spark.driver.memory     1g
```

3.1.5 **Log4j** pour la gestion des logs

```
# Root logger option

log4j.rootLogger=INFO, stdout


# Direct log messages to stdout

log4j.appender.stdout=org.apache.log4j.ConsoleAppender

log4j.appender.stdout.Target=System.out

log4j.appender.stdout.layout=org.apache.log4j.PatternLayout

log4j.appender.stdout.layout.ConversionPattern=%d{yyyy-MM-dd HH:mm:ss} %-5p %c{1}:%L - %m%n
```

3.2 Analyse.py

analyse.py : le document qui sera executé pour implémenter l'analyse de purchases.txt

```python
from pyspark.sql import SparkSession
from pyspark.sql.functions import col, sum, count, desc
from pyspark.sql.types import
 StructType, StructField, StringType, DoubleType

spark = SparkSession.builder \
    .appName("PurchaseAnalysis") \
    .getOrCreate()

# schema definition(tab-delimited, no header)
schema = StructType([
    StructField("date", StringType(), True),
    StructField("time", StringType(), True),
    StructField("city", StringType(), True),
    StructField("category", StringType(), True),
    StructField("price", DoubleType(), True),
    StructField("payment_method", StringType(), True)
])
df = spark.read.csv(
    "purchases.csv",
    sep="\t",
    header=False,
    schema=schema
)

#print schema
print("Schema:")
df.printSchema()

# Basic statistics for price
print("\nPrice Statistics:")
df.describe(["price"]).show()

#total Revenue (sum of all prices)
total_revenue = df.agg(sum("price").alias("total_revenue"))
print("\nTotal Revenue:")
total_revenue.show()

# top Cities by revenue
city_revenue = df.groupBy("city") \
    .agg(sum("price").alias("total_revenue")) \
    .orderBy(desc("total_revenue"))

print("\nTop 10 Cities by Revenue:")
city_revenue.show(10, truncate=False)
```

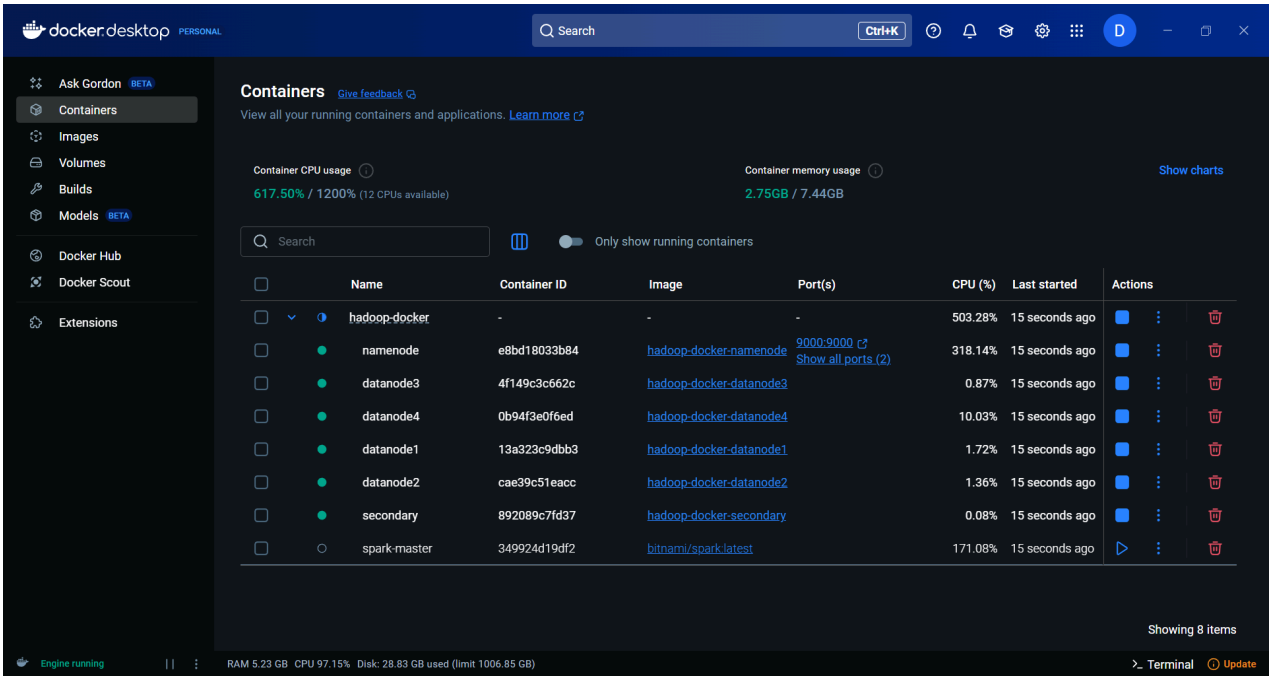
```python
#top Product Categories by Revenue and Transaction Count
category_stats = df.groupBy("category") \
    .agg(
        count("*").alias("transaction_count"),
        sum("price").alias("total_revenue")
    ) \
    .orderBy(desc("total_revenue"))

print("\nTop 10 Product Categories:")
category_stats.show(10, truncate=False)

#payment method
payment_stats = df.groupBy("payment_method") \
    .agg(
        count("*").alias("transaction_count"),
        sum("price").alias("total_revenue")
    ) \
    .orderBy(desc("total_revenue"))

print("\nPayment Method Breakdown:")
payment_stats.show(truncate=False)

#daily Sales Trend
daily_sales = df.groupBy("date") \
    .agg(sum("price").alias("daily_revenue")) \
    .orderBy("date")

print("\nDaily Sales Trend:")
daily_sales.show(truncate=False)

#hourly sales (combine date + time)
from pyspark.sql.functions import to_timestamp, hour
df_with_hour = df.withColumn(
    "timestamp",
    to_timestamp(col("date") + " " + col("time"), "yyyy-MM-dd HH:mm")
).withColumn("hour", hour("timestamp"))

hourly_sales = df_with_hour.groupBy("hour") \
    .agg(sum("price").alias("hourly_revenue")) \
    .orderBy("hour")

# Stop Spark session
spark.stop()
```

3.2 téléchargement SPARK

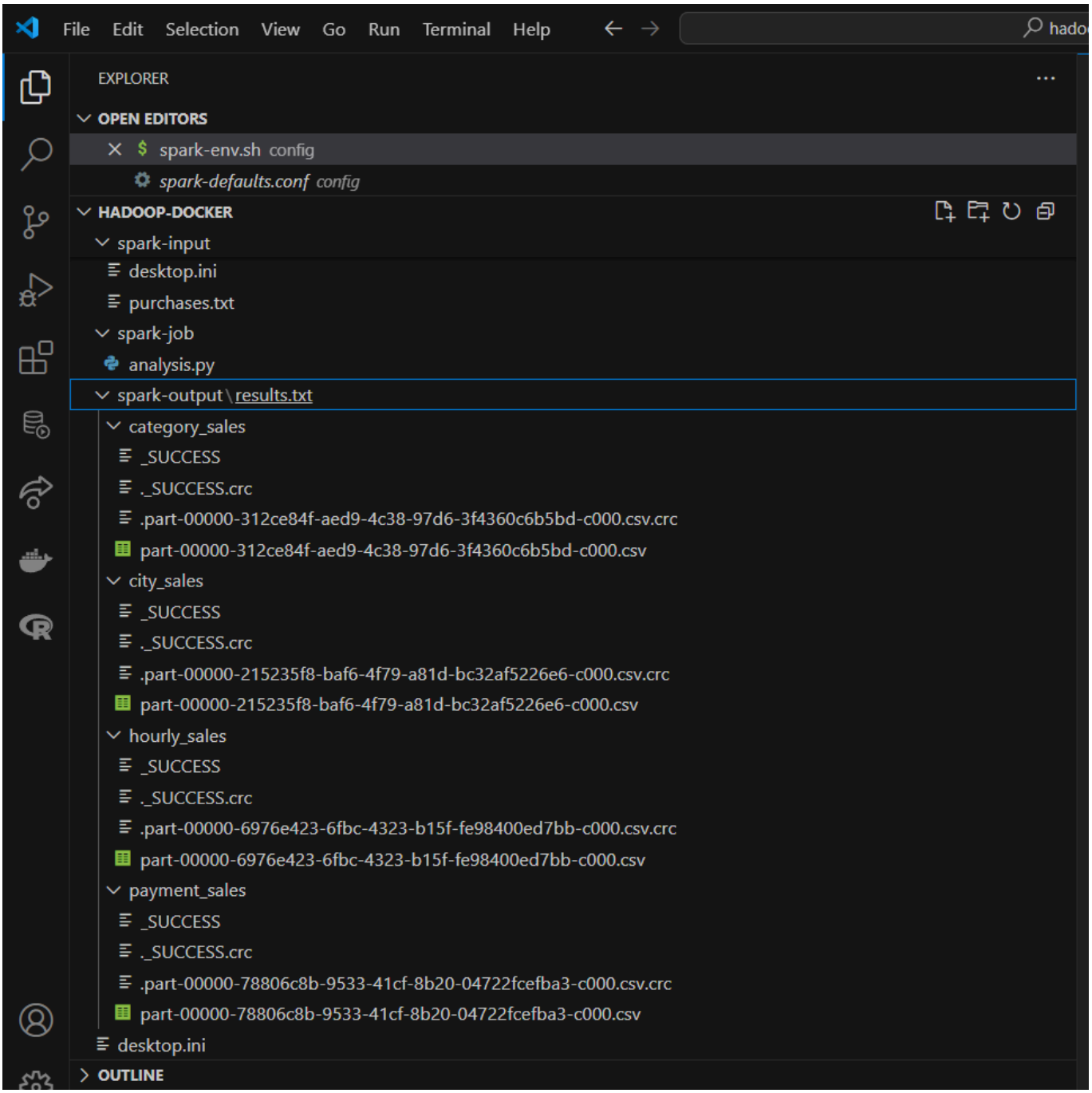
```
Administrator: Windows Powe
 => => exporting attestation manifest sha256:84f110f9f80e0fb7a9a673df9ac92c2b2e81badc3a3b2b95d5aab8cc7ce817ea                0.3s
 => => exporting manifest list sha256:c8aa12c799c08bac0b6c4cc496645bce8b4e2c167eba3c659d5dd744e06989c8                       0.1s
 => => naming to docker.io/library/hadoop-docker-datanode1:latest                                                            0.0s
 => => unpacking to docker.io/library/hadoop-docker-datanode1:latest                                                        68.1s
 => [datanode2] exporting to image                                                                                         173.9s
 => => exporting layers                                                                                                    104.5s
 => => exporting manifest sha256:2a73a5cffc19e1237b90402b3c7f131f9ea53bc601e7a9a2e07de167570d0651                            0.1s
 => => exporting config sha256:03a8c2e378b4f638f893bb4a07e22e1addea36fe42092ea8b3d714fccfe80ce2                              0.2s
 => => exporting attestation manifest sha256:8718271c7af9df39a5a7ff691b495b88789b438e011ae7509c779ff42ca96dab                0.3s
 => => exporting manifest list sha256:43449badc4636bd2f7ab1515f7749c29b50163a9750af09a3e450f6b1dd458c7                       0.1s
 => => naming to docker.io/library/hadoop-docker-datanode2:latest                                                            0.0s
 => => unpacking to docker.io/library/hadoop-docker-datanode2:latest                                                        68.1s
 => [secondary] exporting to image                                                                                         132.0s
 => => exporting layers                                                                                                     62.7s
 => => exporting manifest sha256:8728e160e853717304387608e033b341173c8144337e42e6efba6d0777636225                            0.1s
 => => exporting config sha256:5018b1c60e7b7cd923d17fc9d6ef4c0004cbb76e3427a796e8b7e2594fcd1d73                              0.2s
 => => exporting attestation manifest sha256:0300fda8f85ab3d422476a5fa16279e67997af99fdf69cf381c110eaafe8d2cb                0.4s
 => => exporting manifest list sha256:8b1e8907c4ff00a3e86f15508caf09ae6a8303106dea9cb98ad1539e5ddb62cb                       0.1s
 => => naming to docker.io/library/hadoop-docker-secondary:latest                                                            0.1s
 => => unpacking to docker.io/library/hadoop-docker-secondary:latest                                                        68.1s
 => [datanode1] resolving provenance for metadata file                                                                       0.8s
 => [namenode] resolving provenance for metadata file                                                                        0.8s
 => [secondary] resolving provenance for metadata file                                                                       0.6s
 => [datanode2] resolving provenance for metadata file                                                                       0.5s
 => [datanode4] resolving provenance for metadata file                                                                       0.4s
 => [datanode3] resolving provenance for metadata file                                                                       0.2s
[+] Building 6/6
 ✔datanode1  Built                                                                                                           0.0s
 ✔datanode2  Built                                                                                                           0.0s
 ✔datanode3  Built                                                                                                           0.0s
 ✔datanode4  Built                                                                                                           0.0s
 ✔namenode   Built                                                                                                           0.0s
 ✔secondary  Built                                                                                                           0.0s
PS C:\hadoop-docker> docker-compose up -d
time="2025-05-11T00:36:48+01:00" level=warning msg="C:\\hadoop-docker\\docker-compose.yaml: the attribute `version` is obsolete, it will be ignored, please
remove it to avoid potential confusion"
[+] Running 0/2
 - spark [ ] 32.51MB / 866.3MB Pulling                                                                                     291.9s
   - 93101c1742ee Downloading [=>                                        ]  32.51MB/866.3MB                                283.6s
```

3.3 Déploiement de SPARK

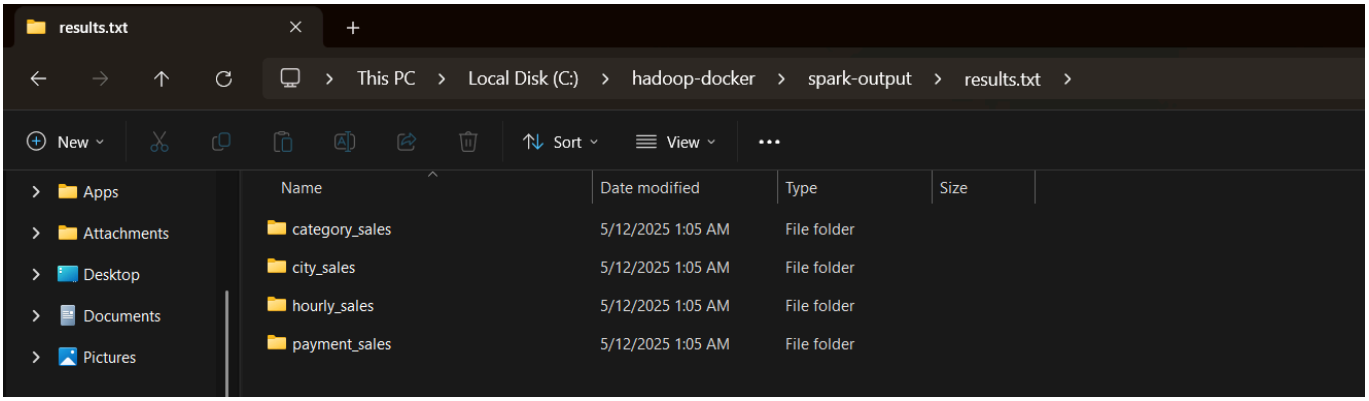
```
Administrator: Windows PowerShell
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\WINDOWS\system32> cd C:\hadoop-docker
PS C:\hadoop-docker> docker-compose up -d
time="2025-05-11T09:20:41+01:00" level=warning msg="C:\\hadoop-docker\\docker-compose.yaml: the attribute `version` is obsolete, it will be ignored, please remove it to avoid potential confusion"
[+] Running 7/7
 ✔Container secondary     Started                                                                                          1.4s
 ✔Container datanode3     Started                                                                                          0.7s
 ✔Container spark-master  Started                                                                                          0.7s
 ✔Container datanode1     Started                                                                                          0.7s
 ✔Container datanode4     Started                                                                                          1.8s
 ✔Container datanode2     Started                                                                                          0.8s
 ✔Container namenode      Started                                                                                          2.0s
PS C:\hadoop-docker> docker-compose up -d
time="2025-05-11T09:21:10+01:00" level=warning msg="C:\\hadoop-docker\\docker-compose.yaml: the attribute `version` is obsolete, it will be ignored, please remove it to avoid potential confusion"
[+] Running 8/8
 ✔Network hadoop-docker_hadoopnet  Created                                                                                 0.0s
 ✔Container secondary              Started                                                                                 1.7s
 ✔Container datanode4              Started                                                                                 1.0s
 ✔Container datanode2              Started                                                                                 1.0s
 ✔Container spark-master           Started                                                                                 1.0s
 ✔Container datanode3              Started                                                                                 2.5s
 ✔Container datanode1              Started                                                                                 1.0s
 ✔Container namenode               Started                                                                                 2.0s
```

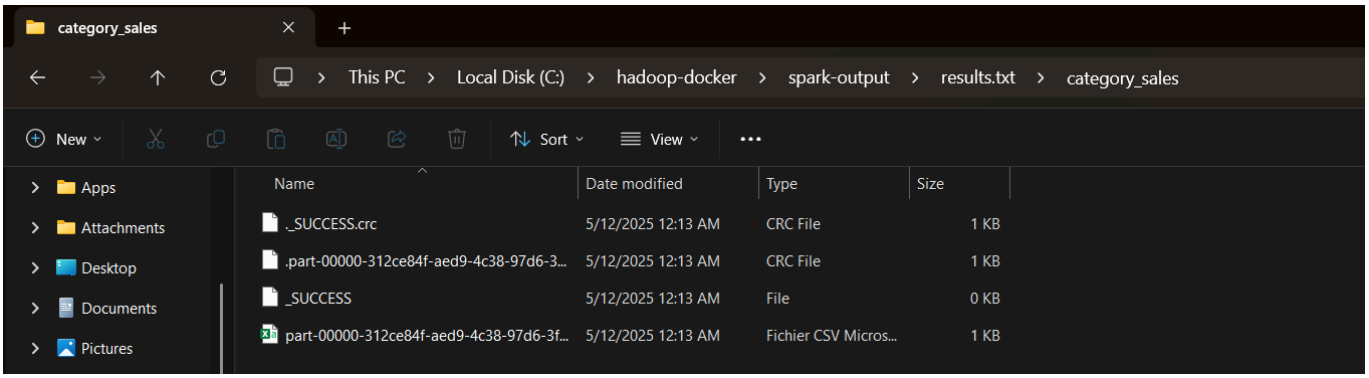


3.4 Output final de l'analyse avec SPARK



Remarque: part-00000 : Indique la première partition (si données < 128 Mo par défaut).





part-00000-312ce84f-aed9-4c38-97d6-3f4360c6b5bd-c000.csv

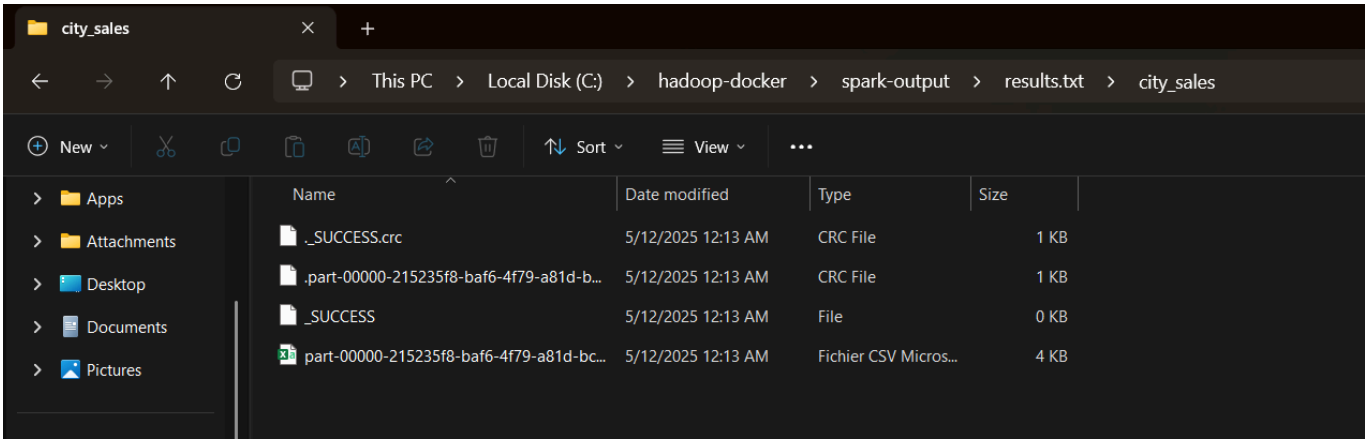
Enregistré dans Drive

Fichier Édition Affichage Insertion Format Données Outils Extensions Aide Accessibilité

A1

	A	B	C	D
1		230274	5.764921214000003E7	250.35050478994603
2	Children's Clothing	230469	5.762482093999995E7	250.03285014470472
3	Men's Clothing	230430	5.762127904000002E7	250.0597970750337
4	Sporting Goods	229932	5.759908589000007E7	250.5048705269387
5	Garden	230073	5.753983311000014E7	250.09380983427005
6	Video Games	230237	5.751316558000035E7	249.79983920916288
7	Music	230150	5.749548969999906E7	249.81746556593484
8	Baby	230293	5.749180843999945E7	249.64635677159075
9	Health and Beauty	229667	5.748158956000002E7	250.28231988052275
10	Toys	229964	5.746347710999996E7	249.88031652780418
11	Consumer Electronics	229761	5.745237413000005E7	250.05276844198994
12	Books	229787	5.745075791E7	250.01744184832037
13	Women's Clothing	230050	5.743444896999999E7	249.66072145185825
14	Crafts	229749	5.741815449999985E7	249.91688538361421
15	CDs	230039	5.741075303999954E7	249.56965140693515
16	Computers	229059	5.731540631999998E7	250.22114966013115
17	Cameras	229320	5.729904663999994E7	249.865021105878
18	Pet Supplies	229222	5.719725023999995E7	249.52775143747107
19				
20				
21				
22				
23				

part-00000-312ce84f-aed9-4c38-97d6-3f4360c6b5bd-c000.csv



part-00000-215235f8-baf6-4f79-a81d-bc32af5226e6-c000.csv

Fichier Édition Affichage Insertion Format Données Outils Extensions Aide Accessibilité

A1 Philadelphia

	A	B	C	D	E	F	G	H
1	Philadelphia	40748	1.0190080260000013E7					
2	Durham	40501	1.0153890209999997E7					
3	Laredo	40342	1.0144604979999991E7					
4	Newark	40577	1.0144052800000003E7					
5	Cincinnati	40452	1.0139505740000002E7					
6	Washington	40503	1.0139363389999997E7					
7	Irving	40352	1.0133944080000008E7					
8	Fort Wayne	40439	1.0132594020000003E7					
9	Baton Rouge	40387	1.0131273230000002E7					
10	Sacramento	40561	1.0123468180000002E7					
11	Fort Worth	40336	1.0120830649999995E7					
12	Oklahoma City	40446	1.0118986250000006E7					
13	Charlotte	40509	1.0112531339999998E7					
14	Tampa	40136	1.0106428550000008E7					
15	Baltimore	40196	1.0096521450000003E7					
16	Indianapolis	40321	1.0090272770000005E7					
17	Pittsburgh	40358	1.009012482E7					
18	Norfolk	40231	1.0088563170000002E7					
19	Virginia Beach	40169	1.0086553499999994E7					
20	New York	40364	1.0085293550000003E7					
21	Irvine	40325	1.008486745E7					
22	Los Angeles	40254	1.0084576800000004E7					
23	Lexington	40343	1.0084510950000001E7					

part-00000-215235f8-baf6-4f79-a81d-bc32af5226e6-c000.csv

part-00000-215235f8-baf6-4f79-a81d-bc32af5226e6-c000.csv

Fichier Édition Affichage Insertion Format Données Outils Extensions Aide Accessibilité

A1 Philadelphia

	A	B	C	D	E	F	G	H
23	Lexington	40343	1.0084510950000001E7					
24	Wichita	40422	1.0083643209999986E7					
25	Spokane	40222	1.0083362980000004E7					
26	Reno	40254	1.0079955159999998E7					
27	Phoenix	40333	1.0079076700000003E7					
28	Birmingham	40253	1.0076606520000005E7					
29	Anaheim	40086	1.0076416359999994E7					
30	Orlando	40195	1.0074922519999998E7					
31	Arlington	40348	1.0072207969999999E7					
32	Jacksonville	40400	1.0072003329999998E7					
33	Garland	40185	1.0071043919999994E7					
34	Lincoln	40345	1.0069485399999999E7					
35	Cleveland	40313	1.0067835840000005E7					
36	Rochester	40455	1.0067606920000004E7					
37	Dallas	40368	1.0066548450000001E7					
38	Tulsa	40247	1.00649559E7					
39	Milwaukee	40316	1.0064482650000012E7					
40	Chicago	40220	1.0062522069999998E7					
41	Gilbert	40277	1.0062115190000005E7					
42	Raleigh	40261	1.0061442540000003E7					
43	Colorado Spring	40389	1.0061105870000003E7					
44	Saint Paul	40160	1.005723357E7					
45	Austin	40332	1.0057158899999997E7					

part-00000-215235f8-baf6-4f79-a81d-bc32af5226e6-c000.csv

part-00000-215235f8-baf6-4f79-a81d-bc32af5226e6-c000.csv

Fichier Édition Affichage Insertion Format Données Outils Extensions Aide Accessibilité

100% € % .0 .00 123 Par dé... 10 B I A

A1 fx Philadelphia

	A	B	C	D	E	F	G
44	Saint Paul	40160	1.0057233357E7				
45	Austin	40332	1.0057158899999997E7				
46	Las Vegas	40165	1.005425798E7				
47	Mesa	40207	1.0053642600000005E7				
48	Henderson	40194	1.005341604999999E7				
49	Fremont	40219	1.0053242360000001E7				
50	Albuquerque	40345	1.0052311420000006E7				
51	Santa Ana	40306	1.0050309929999998E7				
52	Hialeah	40312	1.0047052759999992E7				
53	Plano	40170	1.0046103609999998E7				
54	Glendale	40264	1.0044493969999999E7				
55	Winston–Salem	40208	1.0044011829999998E7				
56	Houston	40114	1.0042106270000003E7				
57	Boston	40338	1.0039473280000005E7				
58	Boise	40203	1.0039166739999998E7				
59	Memphis	40222	1.0038565320000002E7				
60	Chesapeake	40288	1.0038504919999994E7				
61	Scottsdale	40173	1.003792985E7				
62	Columbus	40225	1.003524103E7				
63	Greensboro	40232	1.0033781389999993E7				
64	Madison	40177	1.0032035539999995E7				
65	Denver	40199	1.0031534870000005E7				
66	Bakersfield	40326	1.0031208920000004E7				
67	North Las Vegas	40013	1.0029652510000002E7				

+ ≡ part-00000-215235f8-baf6-4f79-a81d-bc32af5226e6-c000.csv

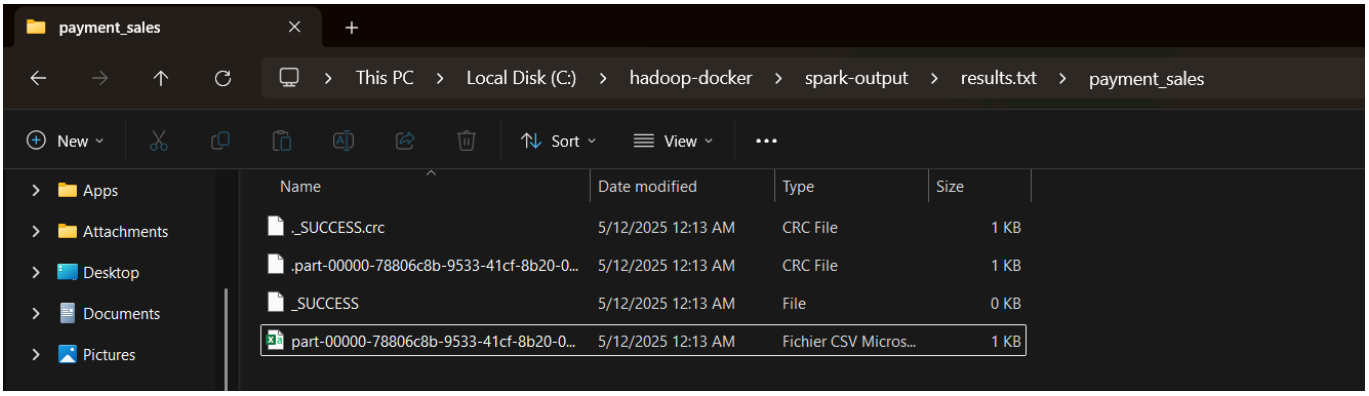
part-00000-215235f8-baf6-4f79-a81d-bc32af5226e6-c000.csv

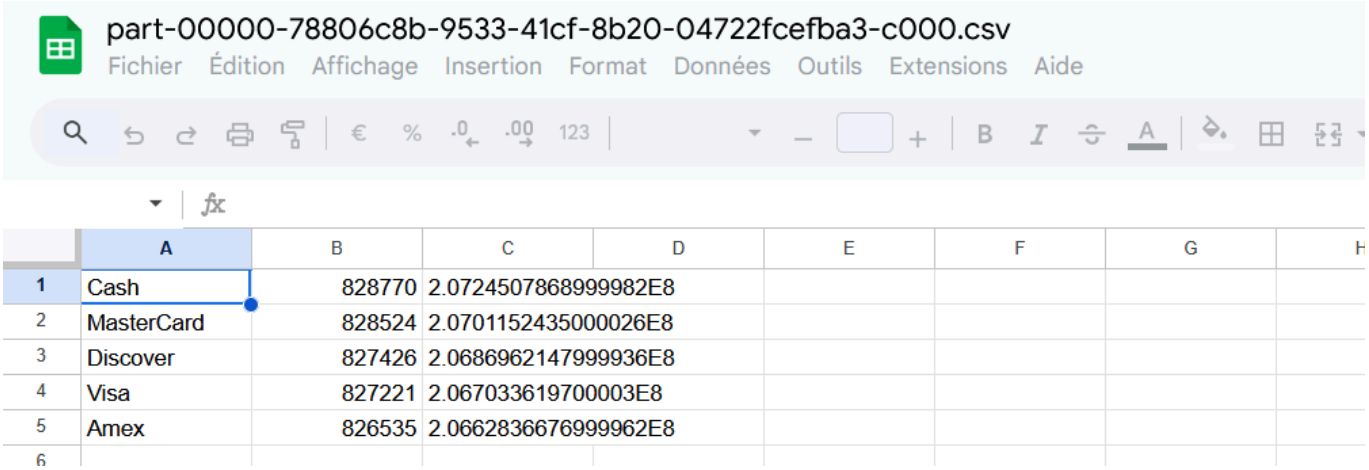
Fichier Édition Affichage Insertion Format Données Outils Extensions Aide Accessibilité

100% € % .0 .00 123 Par dé... 10 B I A

A1 fx Philadelphia

	A	B	C	D	E	F	G
85	Richmond	39983	9992941.589999998				
86	St. Petersburg	40093	9986495.540000003				
87	Detroit	39999	9979260.759999996				
88	Corpus Christi	39925	9976522.769999996				
89	Fresno	40043	9976260.259999992				
90	Chula Vista	40080	9974951.339999992				
91	Kansas City	39713	9968118.729999997				
92	San Diego	39901	9966038.390000002				
93	San Bernardino	40041	9965152.04				
94	Nashville	39854	9961450.509999998				
95	Lubbock	39837	9958119.149999995				
96	New Orleans	39914	9949257.750000006				
97	Miami	39935	9947316.070000002				
98	Oakland	39728	9947292.519999994				
99	San Jose	39898	9936721.409999995				
100	Seattle	39866	9936267.369999994				
101	Anchorage	39806	9933500.400000004				
102	Jersey City	39612	9920141.869999995				
103	Chandler	39826	9919559.859999998				
104							





	A	B	C	D	E	F	G	H
1	Cash	828770	2.0724507868999982E8					
2	MasterCard	828524	2.0701152435000026E8					
3	Discover	827426	2.0686962147999936E8					
4	Visa	827221	2.067033619700003E8					
5	Amex	826535	2.0662836676999962E8					
6								

4. Analyse et conclusion

Grâce à sa puissance de calcul et sa polyvalence, **Apache Spark** a permis d'extraire des insights clés à partir des données transactionnelles, en mettant en lumière les tendances, les comportements clients, et les performances commerciales:

- **category_sales** : Ventes par catégorie de produits.
- **city_sales** : Répartition géographique des ventes.
- **hourly_sales** : Tendances horaires (ex : pic à 15h).
- **payment_sales** : Méthodes de paiement préférées (ex : 60% par carte).