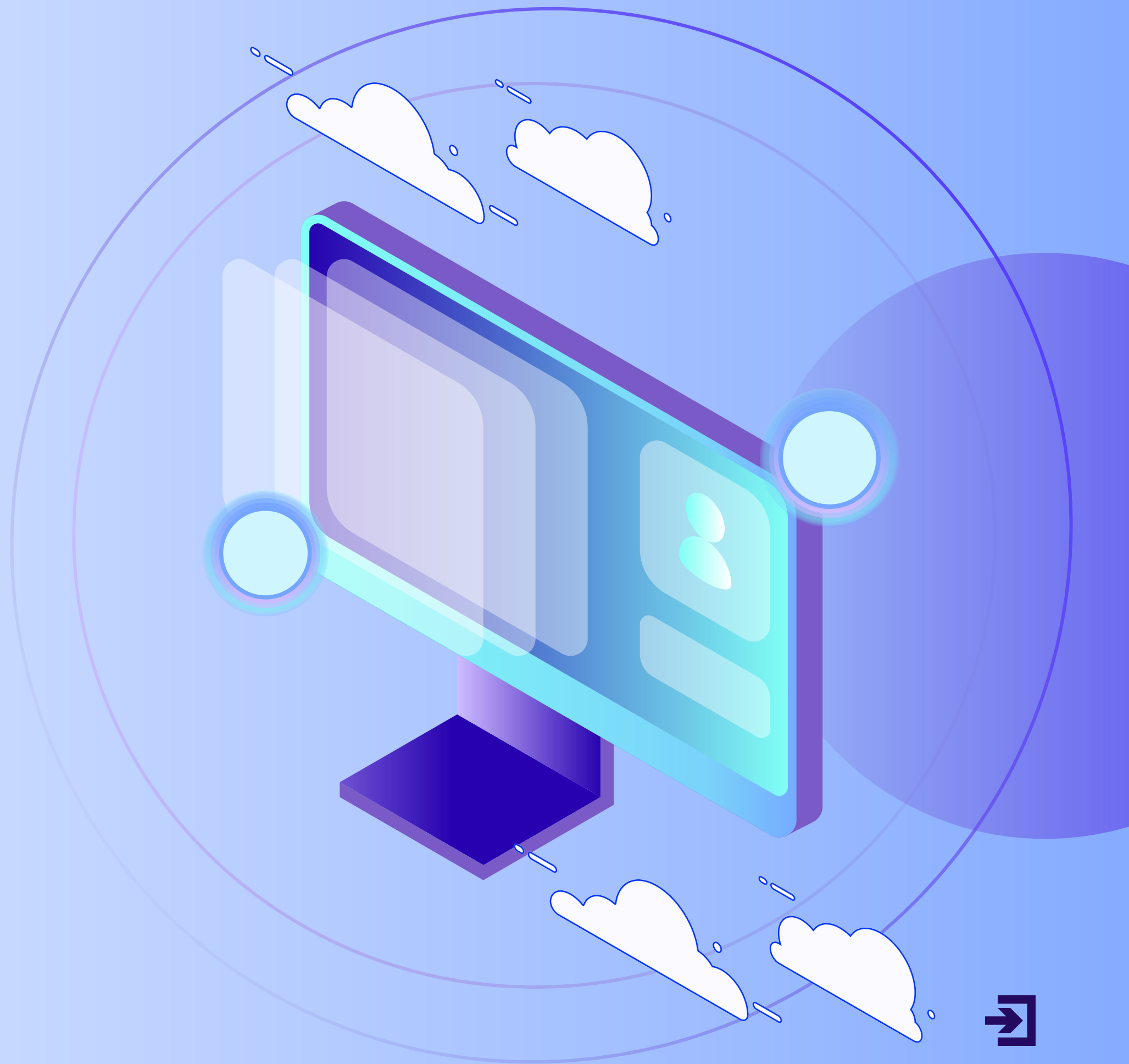


# HADOOP JAVA

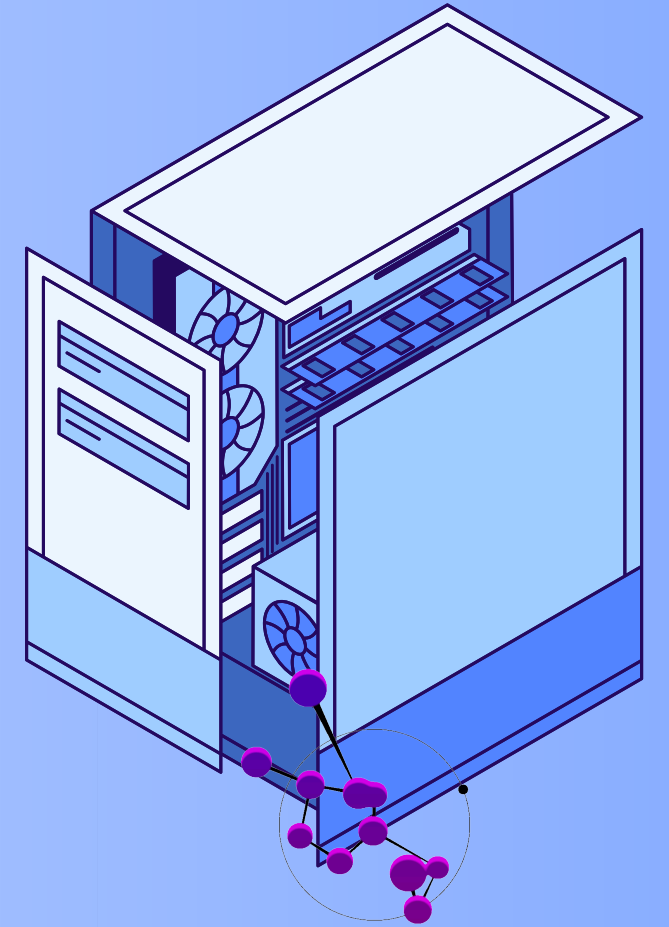
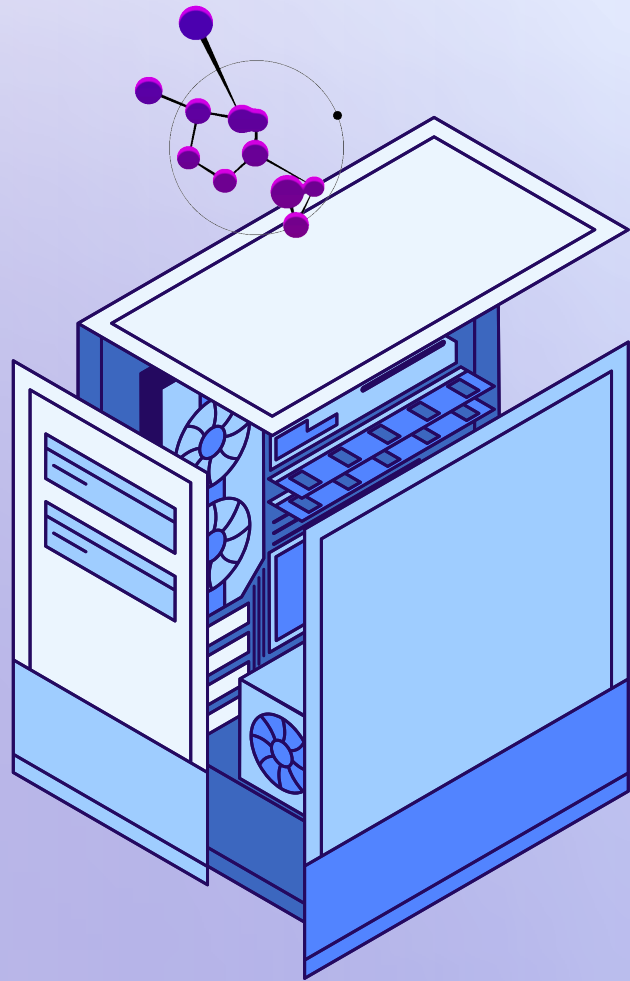


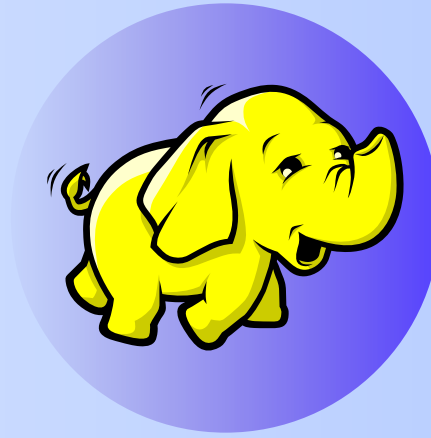


# DEFINITION JAVA

---

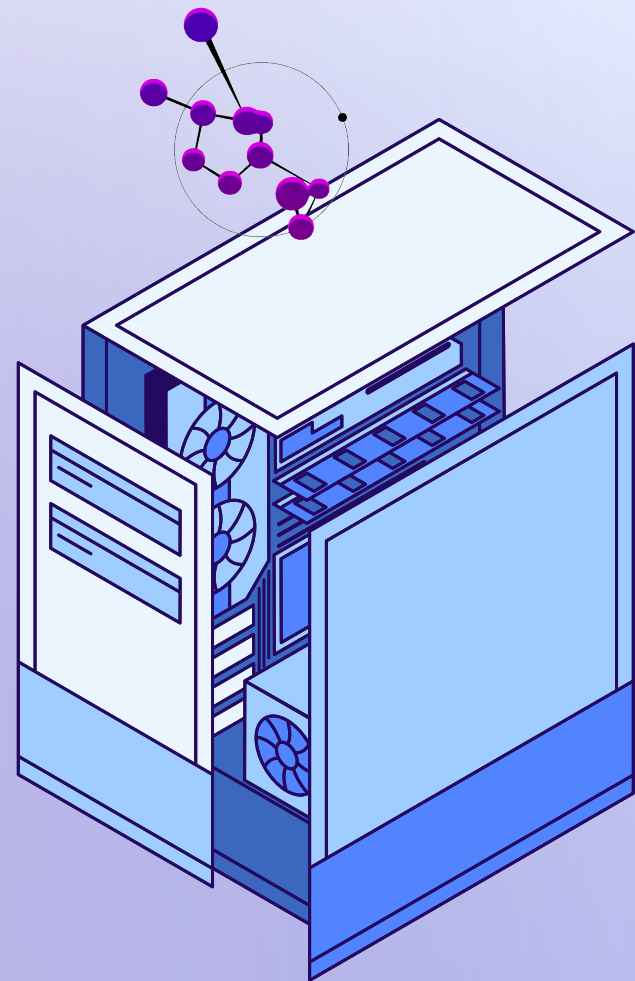
- Orienté objet, simple, portable.
- Écrit une fois, partout (JVM).
- Web, mobile, entreprise.
- Mémoire auto-gérée.
- Sécurisé et fiable.
- Multiplateforme.
- Écosystème riche (Spring, Hibernate)



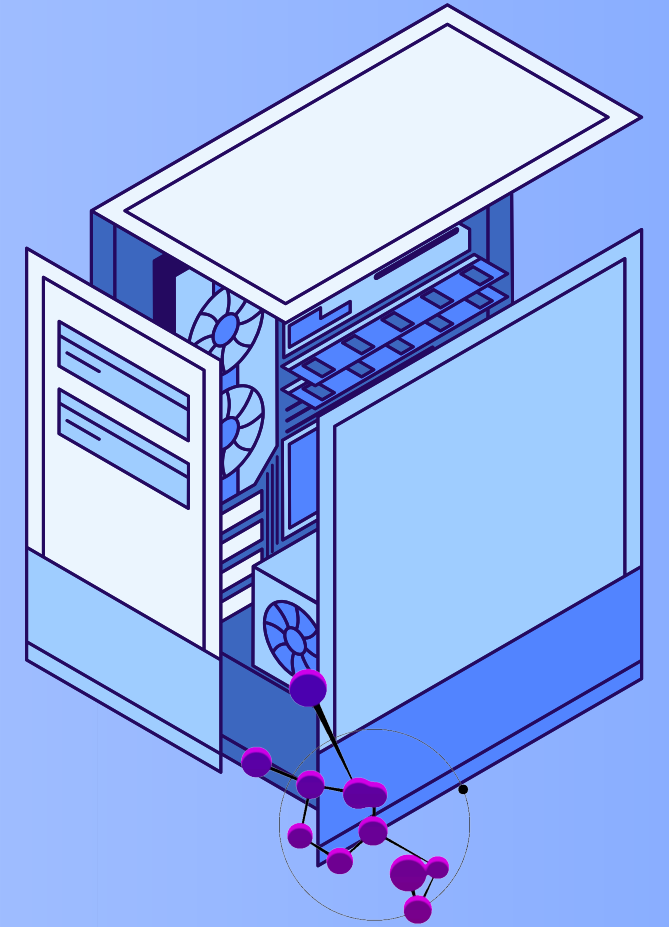


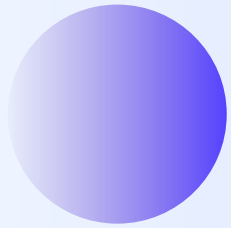
# DEFINITION HADOOP

---




- Framework open-source pour Big Data.
- Stockage distribué (HDFS).
- Traitement parallèle (MapReduce).
- Scalable et résilient.
- Gestion de grandes quantités de données.
- Écosystème riche (Hive, Pig, HBase).
- Utilisé pour analyses massives.



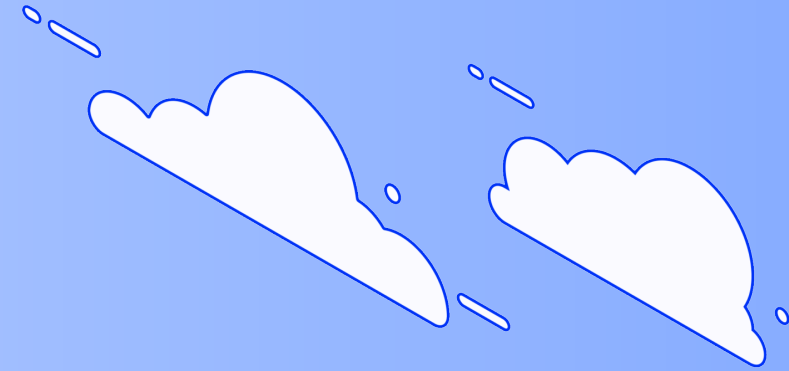
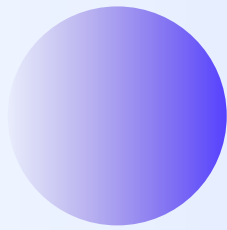


# ETAPE PROJET



- Étape 1 : Job MapReduce
- Traite donnees.json (input) → produit processed.json (output).
- Étape 2 : Transfert des données
- Transfert de processed.json du répertoire output vers input.
- Étape 3 : Automatisation
- Script pour répéter le cycle de traitement et transfert.





# ETAPE PROJET



- Créer un nouveau projet
- Ajouter trois fichiers
- Créer un fichier pom.xml avec les plugins et la version Java 1.7
- Changer l'extension en .jar

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	<b>1</b>		hdfs	supergroup	drwxr-xr-x	October 11, 2024 02:52 AM
<input type="checkbox"/>	<b>.</b>		cloudera	supergroup	drwxr-xr-x	October 11, 2024 08:00 AM
<input type="checkbox"/>	<b>input</b>		cloudera	supergroup	drwxr-xr-x	October 11, 2024 06:49 AM
<input type="checkbox"/>	<b>output</b>		cloudera	supergroup	drwxr-xr-x	October 11, 2024 08:05 AM
Show 45 of 2 items						
Page 1 of 1						

monPremierMapReduce

ID

job\_1728476012470\_0004

NAME

monPremierMapReduce

TYPE

MAPREDUCE

STATUS

SUCCEEDED

USER

cloudera

PROGRESS

100%

MAP

100% 1 / 1

REDUCE

100% 10 / 10

DURATION

4m:10s

SUBMITTED

10/11/24 08:00:51

LogsTasksMetadataCounters

Filter by name

Succeeded

Running

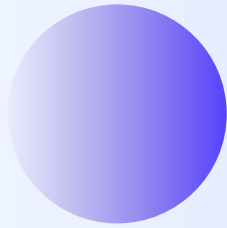
Failed

Map

Reduce

Type	Id	Elapsed Time	Progress	State	Start Time	Successful Attempt	Finish Time
REDUCE	task_1728476012470_0004_r_000009	1m, 27s	100	SUCCEEDED	October 11, 2024 8:03 AM	attempt_1728476012470_0004_r_000009_0	October 11, 2024 8:05 AM
REDUCE	task_1728476012470_0004_r_000005	2m, 15s	100	SUCCEEDED	October 11, 2024 8:01 AM	attempt_1728476012470_0004_r_000005_0	October 11, 2024 8:03 AM
MAP	task_1728476012470_0004_m_000000	18.67s	100	SUCCEEDED	October 11, 2024 8:00 AM	attempt_1728476012470_0004_m_000000_0	October 11, 2024 8:01 AM
REDUCE	task_1728476012470_0004_r_000001	2m, 8s	100	SUCCEEDED	October 11, 2024 8:01 AM	attempt_1728476012470_0004_r_000001_0	October 11, 2024 8:03 AM
REDUCE	task_1728476012470_0004_r_000007	1m, 32s	100	SUCCEEDED	October 11, 2024 8:03 AM	attempt_1728476012470_0004_r_000007_0	October 11, 2024 8:04 AM
REDUCE	task_1728476012470_0004_r_000002	2m, 19s	100	SUCCEEDED	October 11, 2024 8:01 AM	attempt_1728476012470_0004_r_000002_0	October 11, 2024 8:03 AM
REDUCE	task_1728476012470_0004_r_000006	1m, 25s	100	SUCCEEDED	October 11, 2024 8:03 AM	attempt_1728476012470_0004_r_000006_0	October 11, 2024 8:04 AM
REDUCE	task_1728476012470_0004_r_000003	2m, 12s	100	SUCCEEDED	October 11, 2024 8:01 AM	attempt_1728476012470_0004_r_000003_0	October 11, 2024 8:03 AM
REDUCE	task_1728476012470_0004_r_000004	1m, 59s	100	SUCCEEDED	October 11, 2024 8:01 AM	attempt_1728476012470_0004_r_000004_0	October 11, 2024 8:03 AM



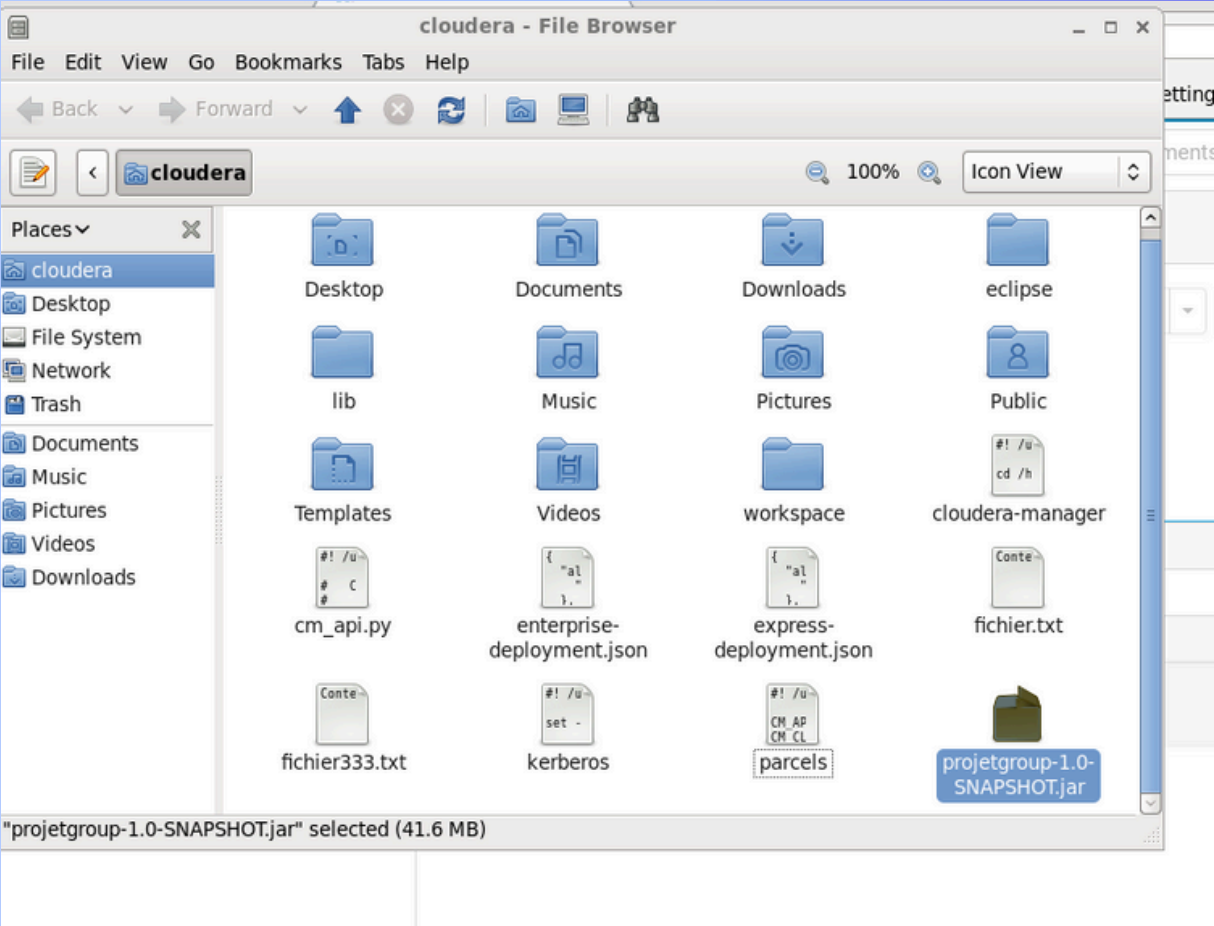
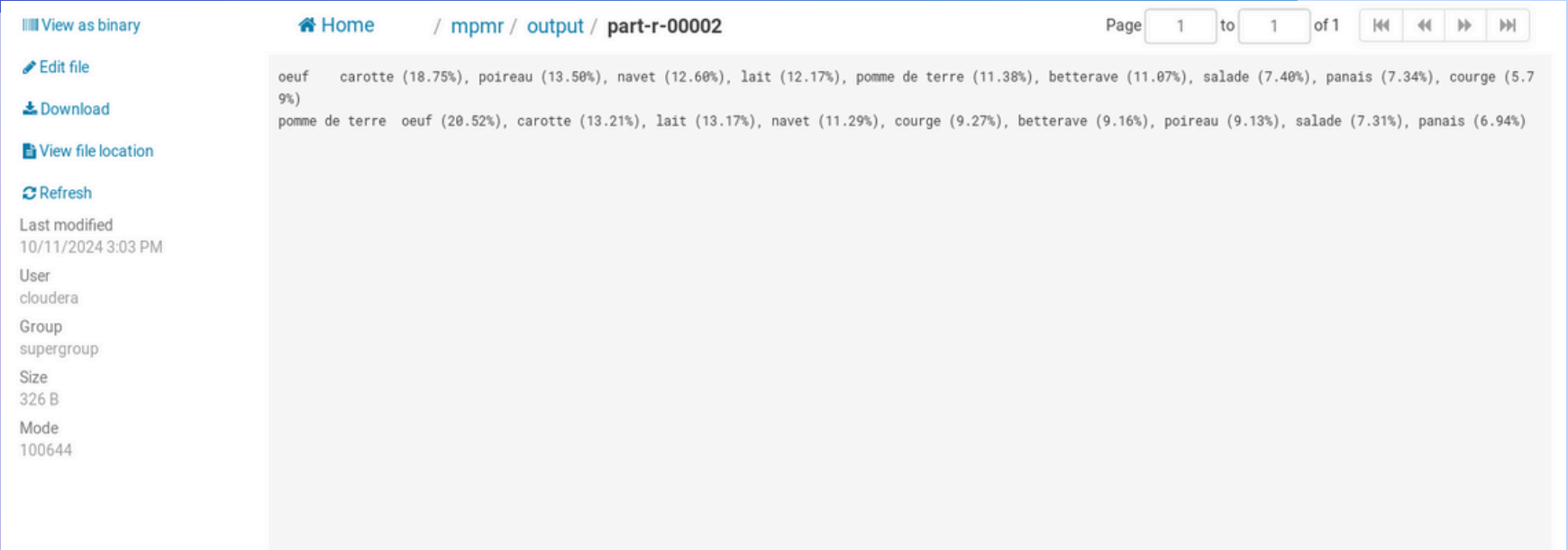


# ETAPE PROJET

<input type="checkbox"/>	Id	Name	User	Type	Status	Progress	Group	Started	Duration
<input type="checkbox"/>	application_1728476012470_0004	monPremierMapReduce	cloudera	MAPREDUCE	SUCCEEDED	100%	root.cloudera	October 11, 2024 8:00 AM	4m, 29s
<input type="checkbox"/>	application_1728476012470_0003	monPremierMapReduce	cloudera	MAPREDUCE	SUCCEEDED	100%	root.cloudera	October 11, 2024 7:33 AM	5m, 25s
<input type="checkbox"/>	application_1728476012470_0002	monPremierMapReduce	cloudera	MAPREDUCE	SUCCEEDED	100%	root.cloudera	October 11, 2024 6:51 AM	3m, 58s
<input type="checkbox"/>	application_1728476012470_0001	monPremierMapReduce	cloudera	MAPREDUCE	SUCCEEDED	100%	root.cloudera	October 11, 2024 6:41 AM	4m, 24s

4 jobs

- Ajouter le fichier sur HUE
- Exécuter la commande `hdfs dfs -get /mpmr/projetgroup-1.0-SNAPSHOT.jar ./`
- Lancer le JAR avec `hadoop jar projetgroup-1.0-SNAPSHOT.jar`
- Le fichier output ne doit pas être créé



# ETAPE PROJET

```
24/10/11 08:00:30 INFO mapreduce.JobSubmitter: number of splits:1
24/10/11 08:00:31 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1728476012470_0004
24/10/11 08:00:32 INFO impl.YarnClientImpl: Submitted application application_1728476012470_0004
24/10/11 08:00:32 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1728476012470_0004/
24/10/11 08:00:32 INFO mapreduce.Job: Running job: job_1728476012470_0004
24/10/11 08:00:54 INFO mapreduce.Job: Job job_1728476012470_0004 running in uber mode : false
24/10/11 08:00:54 INFO mapreduce.Job:  map 0% reduce 0%
24/10/11 08:01:14 INFO mapreduce.Job:  map 100% reduce 0%
^[[S24/10/11 08:02:56 INFO mapreduce.Job:  map 100% reduce 7%
24/10/11 08:03:05 INFO mapreduce.Job:  map 100% reduce 10%
24/10/11 08:03:17 INFO mapreduce.Job:  map 100% reduce 17%
24/10/11 08:03:18 INFO mapreduce.Job:  map 100% reduce 24%
24/10/11 08:03:21 INFO mapreduce.Job:  map 100% reduce 40%
24/10/11 08:03:23 INFO mapreduce.Job:  map 100% reduce 41%
24/10/11 08:03:24 INFO mapreduce.Job:  map 100% reduce 44%
24/10/11 08:03:25 INFO mapreduce.Job:  map 100% reduce 51%
24/10/11 08:03:28 INFO mapreduce.Job:  map 100% reduce 54%
24/10/11 08:03:30 INFO mapreduce.Job:  map 100% reduce 57%
24/10/11 08:03:32 INFO mapreduce.Job:  map 100% reduce 58%
24/10/11 08:03:38 INFO mapreduce.Job:  map 100% reduce 60%
24/10/11 08:04:34 INFO mapreduce.Job:  map 100% reduce 70%
24/10/11 08:04:43 INFO mapreduce.Job:  map 100% reduce 77%
24/10/11 08:04:49 INFO mapreduce.Job:  map 100% reduce 83%
24/10/11 08:04:50 INFO mapreduce.Job:  map 100% reduce 87%
24/10/11 08:04:53 INFO mapreduce.Job:  map 100% reduce 93%
24/10/11 08:04:56 INFO mapreduce.Job:  map 100% reduce 95%
24/10/11 08:05:00 INFO mapreduce.Job:  map 100% reduce 100%
24/10/11 08:05:03 INFO mapreduce.Job: Job job_1728476012470_0004 completed successfully
24/10/11 08:05:04 INFO mapreduce.Job: Counters: 50
```

```
courge  lait (15.36%), navet (14.21%), oeuf (13.54%), carotte (12.28%), pomme de terre (10.20%), poireau (10.03%), panais (8.51%), betterave (8.44%), salade (7.43%)
)
poireau carotte (16.91%), lait (16.53%), navet (12.46%), betterave (10.92%), oeuf (10.54%), pomme de terre (10.03%), salade (9.51%), panais (8.20%), courge (4.91%)
```





**THANK YOU!**