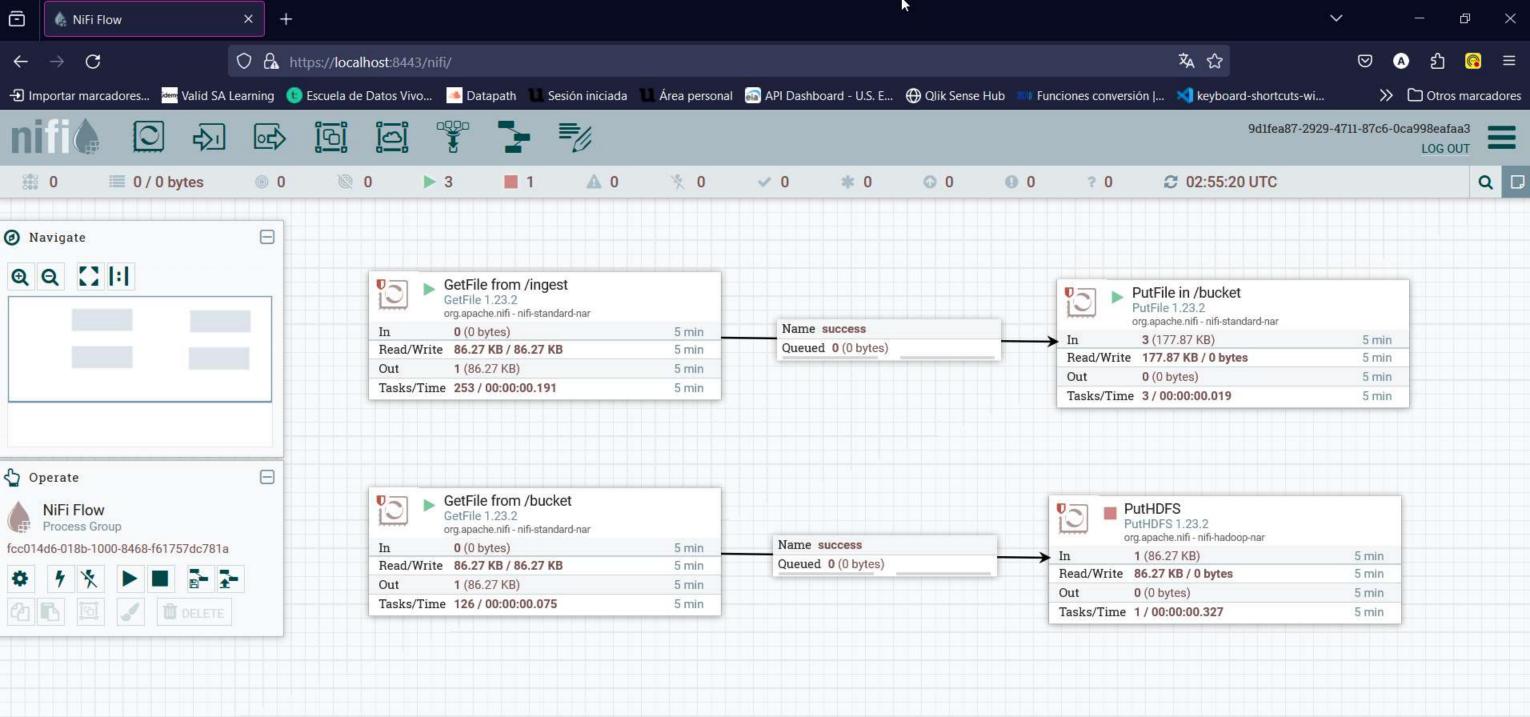```
nifi@9e71bc35ef8b:~/hadoop$ pwd
/home/nifi/hadoop
nifi@9e71bc35ef8b:~/hadoop$ cat > ingest.sh
FILE=/home/nifi/ingest/titanic.csv

if test -f "$FILE"; then
    echo "$FILE already exists"
    rm /home/nifi/ingest/titanic.csv
    echo "$FILE was removed"
else
    echo "$FILE didnt exist."
fi

wget -P /home/nifi/ingest https://raw.githubusercontent.com/fpineyro/homework-0/master/titanic.csv -o titanic.csv
echo "$FILE was downloaded"
nifi@9e71bc35ef8b:~/hadoop$ chmod 777 ingest.sh
nifi@9e71bc35ef8b:~/hadoop$ ./ingest.sh
/home/nifi/ingest/titanic.csv didnt exist.
/home/nifi/ingest/titanic.csv was downloaded
nifi@9e71bc35ef8b:~/hadoop$ ls -lh /home/nifi/ingest/
total 60K
-rw-r--r-- 1 nifi nifi 59K Dec 28 03:43 titanic.csv
nifi@9e71bc35ef8b:~/hadoop$
```

# Configure Processor | GetFile 1.23.2
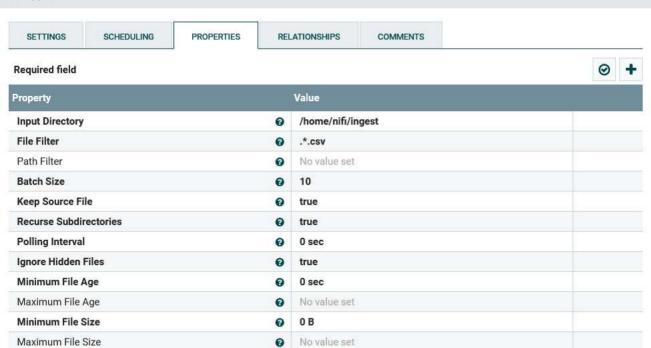
**Stopped**

| SETTINGS | SCHEDULING | **PROPERTIES** | RELATIONSHIPS | COMMENTS |
|----------|------------|----------------|---------------|----------|

**Required field**                                                    ⊘  +

| Property | | Value |
|----------|---|-------|
| **Input Directory** | ❓ | /home/nifi/ingest |
| **File Filter** | ❓ | .*.csv |
| Path Filter | ❓ | No value set |
| **Batch Size** | ❓ | 10 |
| **Keep Source File** | ❓ | true |
| **Recurse Subdirectories** | ❓ | true |
| **Polling Interval** | ❓ | 0 sec |
| **Ignore Hidden Files** | ❓ | true |
| **Minimum File Age** | ❓ | 0 sec |
| Maximum File Age | ❓ | No value set |
| **Minimum File Size** | ❓ | 0 B |
| Maximum File Size | ❓ | No value set |

CANCEL     APPLY

# Configure Processor | PutFile 1.23.2

**Stopped**

| SETTINGS | SCHEDULING | PROPERTIES | RELATIONSHIPS | COMMENTS |
|----------|------------|------------|---------------|----------|

**Required field**

| Property | | Value |
|----------|---|-------|
| **Directory** | ❓ | /home/nifi/bucket |
| **Conflict Resolution Strategy** | ❓ | replace |
| **Create Missing Directories** | ❓ | true |
| Maximum File Count | ❓ | No value set |
| Last Modified Time | ❓ | No value set |
| Permissions | ❓ | No value set |
| Owner | ❓ | No value set |
| Group | ❓ | No value set |

CANCEL    APPLY

# Configure Processor | PutFile 1.23.2

**Stopped**

| SETTINGS | SCHEDULING | PROPERTIES | **RELATIONSHIPS** | COMMENTS |

Automatically Terminate / Retry Relationships ❓

**failure**

☑ terminate  ☑ retry

Files that could not be written to the output directory for some reason are transferred to this relationship

**success**

☑ terminate  ☐ retry

Files that have been successfully written to the output directory are transferred to this relationship

Number Of Retry Attempts ❓

10

Retry Back Off Policy ❓

● Penalize  ○ Yield

Retry Maximum Back Off Period ❓

10 mins

CANCEL   APPLY

# Configure Processor | GetFile 1.23.2

**Stopped**

| SETTINGS | SCHEDULING | PROPERTIES | RELATIONSHIPS | COMMENTS |
|---|---|---|---|---|

**Required field**

| Property | | Value |
|---|---|---|
| **Input Directory** | ❓ | /home/nifi/bucket |
| **File Filter** | ❓ | .*.csv |
| Path Filter | ❓ | No value set |
| **Batch Size** | ❓ | 10 |
| **Keep Source File** | ❓ | false |
| **Recurse Subdirectories** | ❓ | true |
| **Polling Interval** | ❓ | 0 sec |
| **Ignore Hidden Files** | ❓ | true |
| **Minimum File Age** | ❓ | 0 sec |
| Maximum File Age | ❓ | No value set |
| **Minimum File Size** | ❓ | 0 B |
| Maximum File Size | ❓ | No value set |

CANCEL     APPLY

# Configure Processor | PutHDFS 1.23.2

Stopped

| SETTINGS | SCHEDULING | **PROPERTIES** | RELATIONSHIPS | COMMENTS |

**Required field**                                                                    ⊘  ＋

| Property | | Value |
|---|---|---|
| Hadoop Configuration Resources | ❷ | /home/nifi/hadoop/core-site.xml, /home/nifi/hadoop/hdfs-site.xml |
| Kerberos Credentials Service | ❷ | No value set |
| Kerberos User Service | ❷ | No value set |
| Kerberos Principal | ❷ | No value set |
| Kerberos Keytab | ❷ | No value set |
| Kerberos Password | ❷ | No value set |
| Kerberos Relogin Period | ❷ | 4 hours |
| Additional Classpath Resources | ❷ | No value set |
| **Directory** | ❷ | **/nifi** |
| **Conflict Resolution Strategy** | ❷ | **replace** |
| **Writing Strategy** | ❷ | **Write and rename** |
| Block Size | ❷ | No value set |

CANCEL     APPLY

# Configure Processor | PutHDFS 1.23.2

**Stopped**

| SETTINGS | SCHEDULING | PROPERTIES | RELATIONSHIPS | COMMENTS |
|----------|------------|------------|---------------|----------|

**Automatically Terminate / Retry Relationships** ❓

**failure**

☑ terminate  ☑ retry

Files that could not be written to HDFS for some reason are transferred to this relationship

**success**

☑ terminate  ☐ retry

Files that have been successfully written to HDFS are transferred to this relationship

**Number Of Retry Attempts** ❓

    10

**Retry Back Off Policy** ❓
● Penalize   ○ Yield

**Retry Maximum Back Off Period** ❓

    10 mins

CANCEL    APPLY

```
hadoop@0e8a607bbfab:/$ hdfs dfs -head /nifi/titanic.csv
PassengerId,Survived,Pclass,Name,Sex,Age,SibSp,Parch,Ticket,Fare,Cabin,Embarked
1,0,3,"Braund, Mr. Owen Harris",male,22,1,0,A/5 21171,7.25,,S
2,1,1,"Cumings, Mrs. John Bradley (Florence Briggs Thayer)",female,38,1,0,PC 17599,71.2833,C85,C
3,1,3,"Heikkinen, Miss. Laina",female,26,0,0,STON/O2. 3101282,7.925,,S
4,1,1,"Futrelle, Mrs. Jacques Heath (Lily May Peel)",female,35,1,0,113803,53.1,C123,S
5,0,3,"Allen, Mr. William Henry",male,35,0,0,373450,8.05,,S
6,0,3,"Moran, Mr. James",male,,0,0,330877,8.4583,,Q
7,0,1,"McCarthy, Mr. Timothy J",male,54,0,0,17463,51.8625,E46,S
8,0,3,"Palsson, Master. Gosta Leonard",male,2,3,1,349909,21.075,,S
9,1,3,"Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)",female,27,0,2,347742,11.1333,,S
10,1,2,"Nasser, Mrs. Nicholas (Adele Achem)",female,14,1,0,237736,30.0708,,C
11,1,3,"Sandstrom, Miss. Marguerite Rut",female,4,1,1,PP 9549,16.7,G6,S
12,1,1,"Bonnell, Miss. Elizabeth",female,58,0,0,113783,26.55,C103,S
13,0,3,"Saundercock, Mr. William Henry",male,20,0,0,A/5. 2151,8.05,,S
14,0,3,"hadoop@0e8a607bbfab:/$
```

Airflow   **DAGs**   **Security**   **Browse**   **Admin**   **Docs**   05:45 UTC   AA

Triggered ingest-transform-titanic-andresleal, it should start any moment now.   ✕

DAG: ingest-transform-titanic-andresleal   queued   Schedule: 0 0 * * *   Next Run: 2023-12-28, 00:00:00

⊞ Grid   🔲 Graph   📅 Calendar   ⌛ Task Duration   ⇄ Task Tries   ⬐ Landing Times   ☰ Gantt   ⚠ Details   <> Code   📄 Audit Log   ▶   🗑

📅 2023-12-28T05:45:00Z   Runs   25 ▾   Run   manual__2023-12-28T05:44:59.160551+00:00 ▾   Layout   Left > Right ▾   Update   Find Task...

BashOperator   DummyOperator   queued   running   success   failed   up_for_retry   up_for_reschedule   upstream_failed   skipped   scheduled   deferred   no_status

Auto-refresh   ↻

inicia_proceso → transform_titanic → finaliza_proceso

DBeaver 23.2.5 - <tripdata_andres_leal_db> Script

File   Edit   Navigate   Search   SQL Editor   Database   Window   Help

Auto   tripdata_andres_leal_db   tripdata_andres_leal_db

Database Navigator   Projects

Enter a part of object name here

- northwind - localhost:5432
- tripdata_andres_leal_db - localhost:10000
  - default
  - northwind_analytics
  - tripdata
  - tripdata_andres_leal_db
    - Tables
      - airport_trips
      - congestion
      - distance
      - passengers
      - payments
      - titanic_transformed
      - tolls
    - Views
    - Indexes
    - Procedures
    - Data Types

*<tripdata_andres_leal_db> Script

```
SELECT *
FROM titanic_transformed tt
;
```

Results 1

SELECT * FROM titanic_transformed tt   Enter a SQL expression to filter results (use Ctrl+Space)

	passengerid	survived	pclass	name	sex	age	ticket	fare	embarked	sex_avg_age	cabin
1	1	0	3	Braund, Mr. Ow	male	22	A/5 21171	7.25	S	30.7266445916	0
2	2	1	1	Cumings, Mrs.	female	38	PC 17599	71.2833	C	27.9157088123	C85
3	3	1	3	Heikkinen, Mis	female	26	STON/O2. 310	7.925	S	27.9157088123	0
4	4	1	1	Futrelle, Mrs. Ji	female	35	113803	53.1	S	27.9157088123	C123
5	5	0	3	Allen, Mr. Willi	male	35	373450	8.05	S	30.7266445916	0
6	6	0	3	Moran, Mr. Jan	male	[NULL]	330877	8.4583	Q	30.7266445916	0
7	7	0	1	McCarthy, Mr.	male	54	17463	51.8625	S	30.7266445916	E46
8	8	0	3	Palsson, Maste	male	2	349909	21.075	S	30.7266445916	0
9	9	1	3	Johnson, Mrs. (	female	27	347742	11.1333	S	27.9157088123	0
10	10	1	2	Nasser, Mrs. Ni	female	14	237736	30.0708	C	27.9157088123	0
11	11	1	3	Sandstrom, Mi	female	4	PP 9549	16.7	S	27.9157088123	G6
12	12	1	1	Bonnell, Miss. I	female	58	113783	26.55	S	27.9157088123	C103
13	13	0	3	Saundercock, I	male	20	A/5. 2151	8.05	S	30.7266445916	0
14	14	0	3	Andersson, Mr	male	39	347082	31.275	S	30.7266445916	0
15	15	0	3	Vestrom, Miss.	female	14	350406	7.8542	S	27.9157088123	0
16	16	1	2	Hewlett, Mrs. (	female	55	248706	16	S	27.9157088123	0
17	17	0	3	Rice, Master. E	male	2	382652	29.125	Q	30.7266445916	0
18	18	1	2	Williams, Mr. C	male	[NULL]	244373	13	S	30.7266445916	0
19	19	0	3	Vander Planke	female	31	345763	18	S	27.9157088123	0
20	20	1	3	Masselmani, N	female	[NULL]	2649	7.225	C	27.9157088123	0
21	21	0	2	Fynney, Mr. Jos	male	35	239865	26	S	30.7266445916	0
22	22	1	2	Beesley, Mr. La	male	34	248698	13	S	30.7266445916	D56
23	23	1	3	"McGowan, Mi	female	15	330923	8.0292	Q	27.9157088123	0
24	24	1	1	Sloper, Mr. Wil	male	28	113788	35.5	S	30.7266445916	A6
25	25	0	3	Palsson, Miss.	female	8	349909	21.075	S	27.9157088123	0
26	26	1	3	Asplund, Mrs. (	female	38	347077	31.3875	S	27.9157088123	0
27	27	0	3	Emir, Mr. Farre	male	[NULL]	2631	7.225	C	30.7266445916	0
28	28	0	1	Fortune, Mr. Ch	male	19	19950	263	S	30.7266445916	C23 C25 C27
29	29	1	3	"O'Dwyer, Miss	female	[NULL]	330959	7.8792	Q	27.9157088123	0
30	30	0	3	Todoroff, Mr. L	male	[NULL]	349216	7.8958	S	30.7266445916	0
31	31	0	1	Uruchurtu, Dor	male	40	PC 17601	27.7208	C	30.7266445916	0
32	32	1	1	Spencer, Mrs. \	female	[NULL]	PC 17569	146.5208	C	27.9157088123	B78
33	33	1	3	Glynn, Miss. M.	female	[NULL]	335677	7.75	Q	27.9157088123	0
34	34	0	2	Wheadon, Mr.	male	66	C.A. 24579	10.5	S	30.7266445916	0
35	35	0	1	Meyer, Mr. Edç	male	28	PC 17604	82.1708	C	30.7266445916	0
36	36	0	1	Holverson, Mr.	male	42	113789	52	S	30.7266445916	0
37	37	1	3	Mamee, Mr. H.	male	[NULL]	2677	7.2292	C	30.7266445916	0
38	38	0	3	Cann, Mr. Erne	male	21	A./5. 2152	8.05	S	30.7266445916	0
39	39	0	3	Vander Planke	female	18	345764	18	S	27.9157088123	0

Refresh   Save   Cancel   200   200+   — 200 row(s) fetched - 115ms (24ms fetch), on 2023-12-28 at 00:45:15

Project - General

Name   DataSource
- Bookmarks
- Diagrams
- Scripts

```sql
SELECT *
FROM titanic_transformed tt
;
--a) Cuántos hombres y cuántas mujeres sobrevivieron

SELECT sex, SUM(survived) as supervivientes
FROM titanic_transformed tt
GROUP BY sex
;

--b) Cuántas personas sobrevivieron según cada clase (Pclass)

SELECT pclass, SUM(survived) as supervivientes
FROM titanic_transformed tt
GROUP BY pclass
;

--c) Cuál fue la persona de mayor edad que sobrevivió

SELECT max(age) as max_age
FROM titanic_transformed tt
WHERE survived = 1
;

--d) Cuál fue la persona más joven que sobrevivió

SELECT min(age) as min_age
FROM titanic_transformed tt
WHERE survived = 1
;
```

Results 1 ×

SELECT sex, SUM(survived) as supervivientes FROM titanic_transforme · Enter a SQL expression to filter results (use Ctrl+Space)

	sex	supervivientes
1	female	233
2	male	109

```sql
SELECT *
FROM titanic_transformed tt
;
--a) Cuántos hombres y cuántas mujeres sobrevivieron

SELECT sex, SUM(survived) as supervivientes
FROM titanic_transformed tt
GROUP BY sex
;

--b) Cuántas personas sobrevivieron según cada clase (Pclass)

SELECT pclass, SUM(survived) as supervivientes
FROM titanic_transformed tt
GROUP BY pclass
;

--c) Cuál fue la persona de mayor edad que sobrevivió

SELECT max(age) as max_age
FROM titanic_transformed tt
WHERE survived = 1
;

--d) Cuál fue la persona más joven que sobrevivió

SELECT min(age) as min_age
FROM titanic_transformed tt
WHERE survived = 1
;
```

Results 1 ×

SELECT pclass, SUM(survived) as supervivientes FROM titanic_transform | Enter a SQL expression to filter results (use Ctrl+Space)

	pclass	supervivientes
1	1	136
2	2	87
3	3	119

```sql
SELECT *
FROM titanic_transformed tt
;
--a) Cuántos hombres y cuántas mujeres sobrevivieron

SELECT sex, SUM(survived) as supervivientes
FROM titanic_transformed tt
GROUP BY sex
;

--b) Cuántas personas sobrevivieron según cada clase (Pclass)

SELECT pclass, SUM(survived) as supervivientes
FROM titanic_transformed tt
GROUP BY pclass
;

--c) Cuál fue la persona de mayor edad que sobrevivió

SELECT max(age) as max_age
FROM titanic_transformed tt
WHERE survived = 1
;

--d) Cuál fue la persona más joven que sobrevivió

SELECT min(age) as min_age
FROM titanic_transformed tt
WHERE survived = 1
;
```

Results 1 ×

⊕T SELECT max(age) as max_age FROM titanic_transformed tt WHERE su | Enter a SQL expression to filter results (use Ctrl+Space)

	123 max_age
1	80

```sql
SELECT *
FROM titanic_transformed tt
;
--a) Cuántos hombres y cuántas mujeres sobrevivieron

SELECT sex, SUM(survived) as supervivientes
FROM titanic_transformed tt
GROUP BY sex
;

--b) Cuántas personas sobrevivieron según cada clase (Pclass)

SELECT pclass, SUM(survived) as supervivientes
FROM titanic_transformed tt
GROUP BY pclass
;

--c) Cuál fue la persona de mayor edad que sobrevivió

SELECT max(age) as max_age
FROM titanic_transformed tt
WHERE survived = 1
;

--d) Cuál fue la persona más joven que sobrevivió

SELECT min(age) as min_age
FROM titanic_transformed tt
WHERE survived = 1
;
```

Results 1 ×

SELECT min(age) as min_age FROM titanic_transformed tt WHERE sur | Enter a SQL expression to filter results (use Ctrl+Space)

	min_age
1	0