



## Jad SAMAIN - TP Hive

La base de données “films” n’existe pas.

 default

 foodmart

✓

On crée la nouvelle base de données.

**+ Create Database**

New Database Name

films

Create

Cancel

Ensuite on recrée la table evaluation

TABLE > CREATE TABLE 

UPLOAD TABLE





Name

evaluation

COLUMNS

ADVANCED

TABLE PROPERTIES

COLUMN NAME	DATA TYPE	SIZE	ADVANCED	ACTION
<input type="text" value="user_id"/>	<input type="text" value="INT"/>			<div>Delete</div>
<input type="text" value="movie_id"/>	<input type="text" value="INT"/>			<div>Delete</div>
<input type="text" value="rating"/>	<input type="text" value="INT"/>			<div>Delete</div>
<input type="text" value="time_stamp"/>	<input type="text" value="INT"/>			<div>Delete</div>

+ Add New Column

On recrée la lecture des données

Select File Format

File type

CSV

×

Field Delimiter

TAB - (horizontal tab)

×

Escape Character

ETX - (end of text)

×

Quote Character

NUL - (null)

×

Is first row header?

☐

Contains endlines?

☐

Après l'importation des données, on peut vérifier si les données sont importées correctement :

1 select \* from evaluation

2 limit 5;

✓ Execute

Save As

Insert UDF ▾

Visual Explain

RESULTS

LOG

VISUAL EXPLAIN

TEZ UI





Filter columns





×


evaluation.user_id	evaluation.movie_id	evaluation.rating	evaluation.time_stamp
0	50	5	881250949
0	172	5	881250949
0	133	1	881250949
196	242	3	881250949
186	302	3	891717742

Combien d'enregistrements ?

```
1 select count(*) as Compteur from evaluation;
```

 Execute  Save As  Insert UDF  Visual Explain





 **RESULTS**  LOG  VISUAL EXPLAIN  TEZ UI





Filter columns 


compteur
100003

Combien d'enregistrements par films :

```
1 SELECT rating, COUNT(*) AS occurrence
2 FROM evaluation
3 GROUP BY rating;
```

 Execute  Save As  Insert UDF  Visual Explain

 **RESULTS**  LOG  VISUAL EXPLAIN  TEZ UI

Filter columns 

rating	occurrence
1	6111
2	11370
3	27145
4	34174
5	21203