

EXERCISE-5

Import the data

The screenshot shows the Hue Metastore Manager interface. The top navigation bar includes links for HUE, Home, Query Editors, Metastore Manager, Workflows, File Browser, Job Browser, and a user profile for Phoebe_Wu. The Metastore Manager section is active. On the left, a sidebar titled 'Metastore Manager' shows the 'DATABASE' dropdown set to 'default' and two actions: 'Create a new table from a file' and 'Create a new table manually'. The main content area is titled 'Databases > default > Create a new table from a file' and is divided into three steps: Step 1: Choose File (which is selected), Step 2: Choose Delimiter, and Step 3: Define Columns. A large blue header box says 'Name Your Table and Choose A File'. Below it, the 'Table Name' field contains 'ex_5', which is highlighted with a yellow background. A descriptive text states: 'Name of the new table. Table names must be globally unique. Table names tend to correspond to the directory where the data will be stored.' The 'Description' field is labeled 'Optional' and has a placeholder text: 'Use a table comment to describe the table. For example, note the data's provenance and any caveats users need to know.' The 'Input File' field contains '/ngraminput', also highlighted with a yellow background. A note below it says: 'The HDFS path to the file on which to base this new table definition. It can be compressed (gzip) or not.' The 'Import data from file' checkbox is checked, with a note: 'Check this box to import the data in this file after creating the table definition. Leave it unchecked to define an empty table.' A warning message in a yellow box says: 'Warning: The selected file is going to be moved during the import.' and a note: 'Note: To create a table from a file in s3 follow the instructions [here](#)'. At the bottom is a 'Next' button.

HUE Home Query Editors Metastore Manager Workflows File Browser Job Browser Phoebe_Wu ? Help

Metastore Manager

DATABASE

default

ACTIONS

Create a new table from a file

Create a new table manually

Databases > default > Create a new table from a file

Step 1: Choose File Step 2: Choose Delimiter Step 3: Define Columns

Choose a Delimiter

Beeswax has determined that this file is delimited by tabs.

Delimiter: Tab (\t) Preview

Enter the column delimiter which must be a single character. Use syntax like "\001" or "\t" for special characters.

Table preview	col_1	col_2	col_3	col_4	col_5
	! \$17.95	1985	1	1	1
	! \$17.95	1987	1	1	1
	! \$17.95	1990	1	1	1
	! \$17.95	1991	1	1	1
	! \$17.95	1992	5	5	5
	! \$17.95	1993	2	2	2
	! \$17.95	1995	1	1	1
	! \$17.95	1996	4	2	2
	! \$17.95	1997	6	5	5
	! \$17.95	1998	4	3	3

Create the table (Columns)

The screenshot shows the Hue Metastore Manager interface. The top navigation bar includes links for Home, Query Editors, Metastore Manager (which is selected), Workflows, File Browser, Job Browser, and a user profile for Phoebe_Wu. The main area is titled "Metastore Manager" and displays the structure of a table named "ex_5".

The left sidebar, under the "ACTIONS" heading, lists options: Import Data, Browse Data, Drop Table, and View File Location.

The central panel shows the table structure under "Databases > default > ex_5". It has three tabs: Columns (selected), Sample, and Properties.

	Name	Type	Comment
0	ngram	string	
1	year	smallint	
2	count	bigint	
3	page	bigint	
4	book	bigint	

The Table

HUE Query Editors ▾ Metastore Manager Workflows ▾ Phoebe_Wu ▾

Metastore Manager

ACTIONS

- Import Data
- Browse Data
- Drop Table
- View File Location

Databases > default > ex_5

Columns Sample Properties

	ngram	year	count	page	book
0	\$17.95	1985	1	1	1
1	\$17.95	1987	1	1	1
2	\$17.95	1990	1	1	1
3	\$17.95	1991	1	1	1
4	\$17.95	1992	5	5	5
5	\$17.95	1993	2	2	2
6	\$17.95	1995	1	1	1
7	\$17.95	1996	4	2	2
8	\$17.95	1997	6	5	5
9	\$17.95	1998	4	3	3
10	\$17.95	1999	11	10	10
11	\$17.95	2000	11	9	9
12	\$17.95	2001	5	4	4
13	\$17.95	2002	5	5	5
14	\$17.95	2003	2	2	2
15	\$17.95	2004	14	14	14
16	\$17.95	2005	13	13	13
17	\$17.95	2006	5	5	5

THE MOST POPULAR BIGRAM (OVER ALL THE YEARS)

HUE [Query Editors](#) [Metastore Manager](#) [Workflows](#) [File Browser](#) [Job Browser](#) [Phoebe_Wu](#) [?](#) [Print](#) [Help](#)

[Hive Editor](#) [Query Editor](#) [My Queries](#) [Saved Queries](#) [歴史](#)

Assist [Settings](#)

DATABASE
[default](#)

Table name...

[ex_5](#)

```
1 select ngram, sum (count) over (partition by ngram) as count_total from default.ex_5 order by count_total desc limit 1;
```

[Execute](#) [儲存](#) [Save as...](#) [Explain](#) or create a [New query](#)

Output:

[Recent queries](#) [Query](#) [Log](#) [Columns](#) [Results](#) [Chart](#)

	ngram	count_total
0	"	141164328

EXERCISE-6

THE TOP 10 BIGRAMS IN EACH YEAR

HUE Home Query Editors Metastore Manager Workflows File Browser Job Browser Phoebe_Wu ? X ↗

Hive Editor Query Editor My Queries Saved Queries 歷史

Assist Settings

DATABASE C default

Table name... ex_5

```
1 select year, ngram, count
2 from (
3   select year, ngram, rank() over (partition by year order by count desc) as year_rank,
4   count from default.ex_5 ) rankt
5 where rankt.year_rank < 11
6 order by year desc;
```

Execute 儲存 Save as... Explain or create a New query

Output:

	year	ngram	count
0	2008	the air	885834
1	2008	"	977978
2	2008	cannot be	1014440
3	2008	A.	1390058
4	2008	1.	2154711
5	2008	"place	934601
6	2008	the year	765532
7	2008	other .	811209
8	2008	off the	877036
9	2008	up a	684661
10	2007	: (703266

Hue Job Browser

(Exercise 5 & 6)

HUE [Query Editors](#) ▾ [Metastore Manager](#) [Workflows](#) ▾ [File Browser](#) [Job Browser](#) Phoebe_Wu ▾

Job Browser

使用者名稱: Phoebe_Wu 文字: Search for text [Succeeded](#) [Running](#) [Failed](#) [Killed](#)

Logs	ID	Name	Status	使用者	Maps	Reduces	Queue	Priority	Duration	Submitted
	1541823808471_0004	select year, ngram, count from (se...desc(Stage-2)	SUCCEEDED	Phoebe_Wu	100%	100%	default	N/A	21s	11/09/18 20:59:07
	1541823808471_0003	select year, ngram, count from (se...desc(Stage-1)	SUCCEEDED	Phoebe_Wu	100%	100%	default	N/A	2m:27s	11/09/18 20:56:37
	1541823808471_0002	select ngram, sum (count) over (partition...1(Stage-2)	SUCCEEDED	Phoebe_Wu	100%	100%	default	N/A	2m:43s	11/09/18 20:40:26
	1541823808471_0001	select ngram, sum (count) over (partition...1(Stage-1)	SUCCEEDED	Phoebe_Wu	100%	100%	default	N/A	2m:52s	11/09/18 20:37:32

Showing 1 to 4 of 4 entries [← Previous](#) [1](#) [Next →](#)