

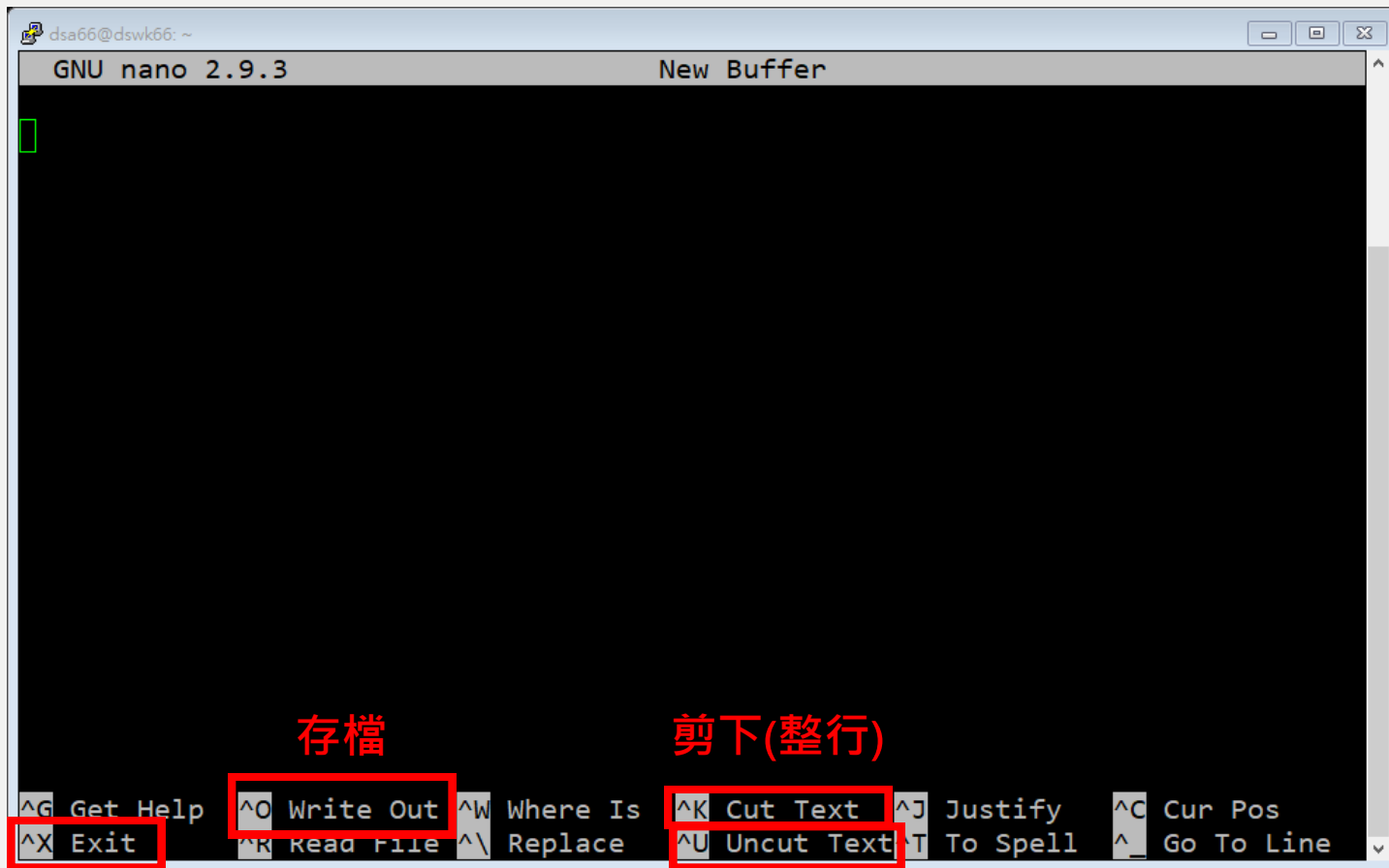
03-實做練習

2018金象盃

講者: 林葳秦

Linux 文字編輯器 - nano

\$ nano



離開

貼上(整行)



Linux 文字編輯器 - nano

編輯完內容後，按 **ctrl + o** 存檔

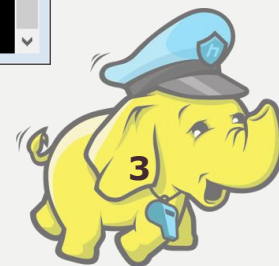


```
dsa66@dswk66: ~
GNU nano 2.9.3          New Buffer          Modified
abc

File Name to Write: abc.pig
^G Get Help      ^M-D DOS Format  ^M-A Append      ^M-B Backup File
^C Cancel        ^M-M Mac Format  ^M-P Prepend     ^M-T To Files
```

輸入檔名後，按 [Enter]

存檔完後，按 **ctrl + x** 離開 nano 文字編輯器

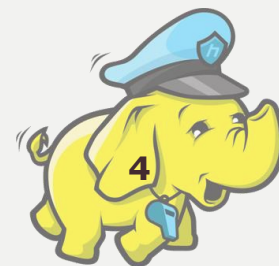


Linux 文字編輯器 - nano

查看剛剛編輯的檔案

\$ ls

```
dsa66@dswk66:~$ ls  
abc.pig  derby.log  metastore_db  
dsa66@dswk66:~$
```



查看 movies_data.csv 資料內容

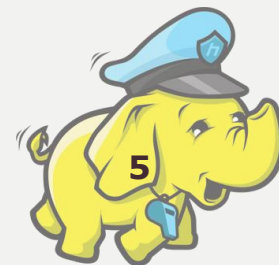
撰寫pig程式

```
$ nano cat_movies.pig
```

```
a = load '/dataset/movies_data/movies_data.csv' using  
PigStorage(',');  
dump a;
```

執行pig程式

```
$ pig cat_movies.pig
```



資料集: 線上音樂串流網站

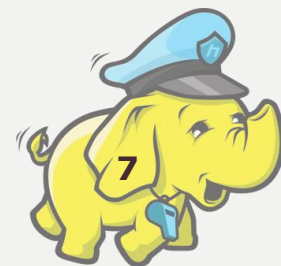
userprofile.tsv, userdemand.tsv

認識 userprofile.tsv 資料

使用者個人資料

user_000003	m	22	United States	Oct 30, 2005
user_000008	m	23	Slovakia	Sep 28, 2006
user_000009	f	19	United States	Jan 13, 2007

欄位名稱	資料型態	欄位說明
userid	chararray	使用者 ID
gender	chararray	性別
age	int	年齡
country	chararray	國籍
registered	chararray	註冊日期

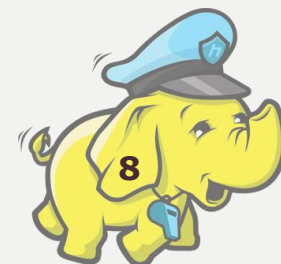


認識 userdemand.tsv 資料

使用者點聽紀錄

```
user_000010 2008-04-25T14:59:58Z Supermode Tell Me Why (Radio Edit)
user_000010 2008-04-25T15:02:50Z Sun Liquide Celebrate The Light
user_000010 2008-04-25T15:05:44Z Linkin Park Wake
user_000010 2008-04-25T15:07:24Z 4 Clubbers (Feat. Silvy) Time (The
Hitmen Remix Cut)
user_000010 2008-04-25T15:10:41Z Mims Ft Junior Reid & Baby Cham This Is
Why Im Hot
```

欄位名稱	資料型態	欄位說明
userid	chararray	使用者 ID
time	chararray	點聽時間
artname	chararray	歌手名稱
traname	chararray	歌曲名稱



查看 userprofile.tsv 資料

```
$ nano cat_userprofile.pig
```

```
a = load '/dataset/userprofile/userprofile.tsv' using PigStorage('\t') as  
(userid:chararray,gender:chararray,age:int,country:chararray,registered:ch  
ararray);  
dump a;
```

```
$ pig cat_userprofile.pig
```

使用者 ID	性別	年齡	國籍	註冊日期
--------	----	----	----	------

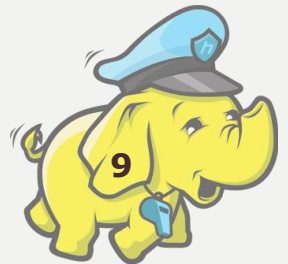
(user_000359,m,31,United States,Aug 21, 2005)

(user_000361,f,26,Poland,Jan 16, 2007)
--

(user_000364,f,29,Norway,Mar 16, 2006)
--

(user_000366,m,24,Finland,Dec 12, 2004)

(user_000533,m,27,Japan,Nov 5, 2009)



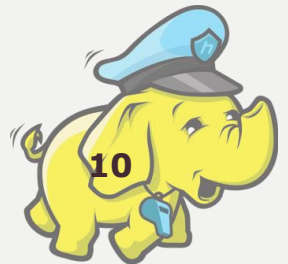
儲存 userprofile.tsv 資料

```
$ nano cat_userprofile.pig
```

```
a = load '/dataset/userprofile/userprofile.tsv' using PigStorage('\t') as  
(userid:chararray,gender:chararray,age:int,country:chararray,registered:ch  
ararray);  
STORE a into 'dsa@@.csv' using PigStorage(',');
```

```
$ pig cat_userprofile.pig
```

```
$ hdfs dfs -cat dsa@@.csv
```



```
$ hdfs dfs -ls
```

Found 1 items

```
drwxr-xr-x  - dsa66 dsa66          0 2018-10-25 01:51 dsa66.csv
```

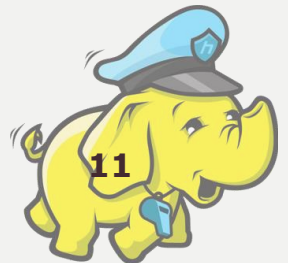
進入**pig**交談模式

```
$ pig
```

```
grunt> ls
```

```
hdfs://queen:8020/user/dsa66/dsa66.csv  <dir>
```

```
grunt> cat dsa@@.csv
```



查看 userdemand.tsv 資料

```
grunt> cat /dataset/userdemand
```

使用者 ID	點聽時間	歌手名稱	歌曲名稱
(user_000010,	2008-04-25T14:59:58Z,	Supermode,	Tell Me Why (Radio Edit))
(user_000010,	2008-04-25T15:02:50Z,	Sun Liquide,	Celebrate The Light)
(user_000010,	2008-04-25T15:05:44Z,	Linkin Park,	Wake)
(user_000010,	2008-04-25T15:07:24Z,	4 Clubbers (Feat. Silvy),	Time (The Hitmen Remix Cut))
(user_000010,	2008-04-25T15:10:41Z,	Mims Ft Junior Reid & Baby Cham,	This Is Why Im Hot)

```
grunt> quit
```

