

The background of the slide is a light gray gradient. It is decorated with numerous realistic water droplets of various sizes. Some droplets are at the top left, some are scattered in the middle, and a larger cluster of droplets is on the right side. The droplets have highlights and shadows, giving them a three-dimensional appearance.

IBM Quest Synthetic Data Generator 使用教學

1. 將 MOODLE 上之執行檔下載並解壓縮

主題 2

Data Mining Fundamentals	<input type="checkbox"/>
 1 introduction	<input type="checkbox"/>
 DataMining introduction - PowerCam上課錄音	<input type="checkbox"/>
 Association Analysis (10/11 updated)	<input type="checkbox"/>
 Association Analysis - PowerCam上課錄音	<input type="checkbox"/>
 Sequential Pattern	<input type="checkbox"/>
 Sequential Pattern - PowerCam上課錄音	<input type="checkbox"/>
 Suffix Tree	<input type="checkbox"/>
 Suffix Tree - PowerCam上課錄音	<input type="checkbox"/>
 Project 1	<input type="checkbox"/>
 IBM-Quest-Data-Generator	<input type="checkbox"/>

2. 使用 cmd 執行此程式

- 使用 -help 查看可下哪些參數

cmd 命令提示字元

```
I:\DataMining\Project1\IBM-Quest-Data-Generator.exe>"IBM Quest Data Generator.exe" -help
Synthetic Data Generation, Version dated July 22, 1997
Usage:  IBM Quest Data Generator.exe lit|tax|seq [options]
        IBM Quest Data Generator.exe lit|tax|seq -help    For more detailed list of options

I:\DataMining\Project1\IBM-Quest-Data-Generator.exe>
```

3. 使用LIT模式產生資料

- 目前僅有 lit 模式可正常運行
- 使用 -help 指令查看可下哪些參數
- 可依據自己的需求調整需要的資料量

```
I:\DataMining\Project1\IBM-Quest-Data-Generator.exe>"IBM Quest Data Generator.exe" lit -help
Command Line Options:
-ntrans number_of_transactions_in_000s (default: 1000)
-tlen avg_items_per_transaction (default: 10)
-nitems number_of_different_items_in_000s (default: 100)

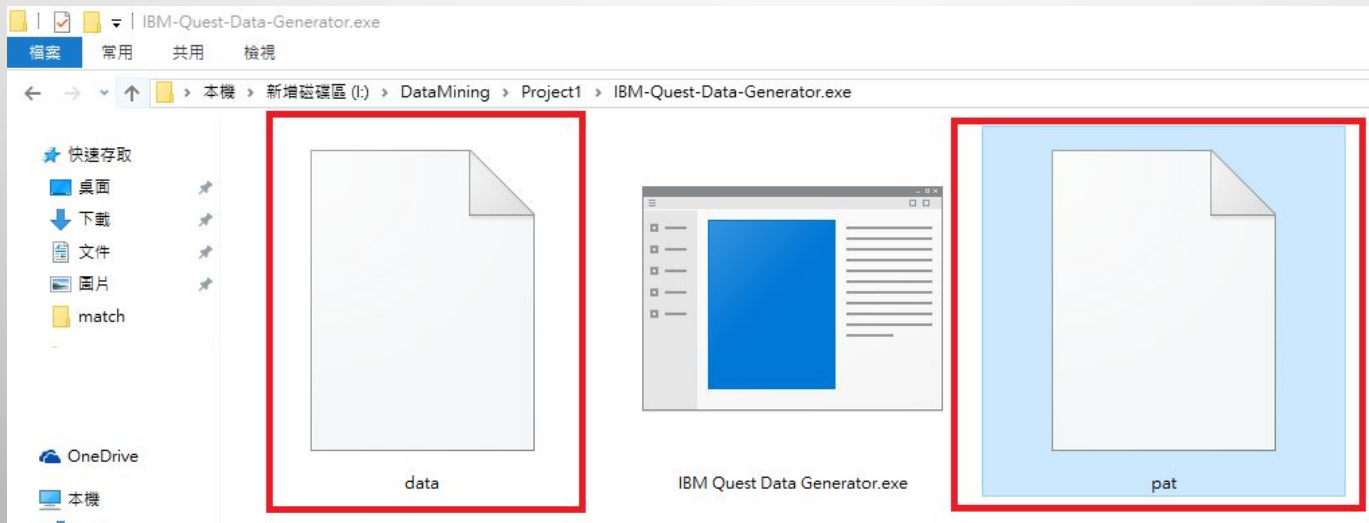
-npats number_of_patterns (default: 10000)
-patlen avg_length_of_maximal_pattern (default: 4)
-corr correlation_between_patterns (default: 0.25)
-conf avg_confidence_in_a_rule (default: 0.75)

-fname <filename> (write to filename.data and filename.pat)
-ascii (default: True)
-randseed # (reset seed used generate to x-acts; must be negative)
-version (to print out version info)

I:\DataMining\Project1\IBM-Quest-Data-Generator.exe>
```

4. 產生資料

- 執行完成之後即會看到程式產生之 data 檔及 pat 檔
- data 檔為所需資料, pat 檔可不必要理會
- 請自行做檔案前處理或轉檔成 .txt .csv 等其他檔案類型後再做使用



5. Data 檔介紹

- Data 中的三個欄位由左至右分別代表:

- 1. Customer ID

- 2. Transaction ID

- 3. Item ID

1	1	21
1	1	4512
1	1	4785
1	1	4886
1	1	4890
1	1	5185
1	1	6338
1	1	6395
1	1	9484
1	1	9994