交易資料處理

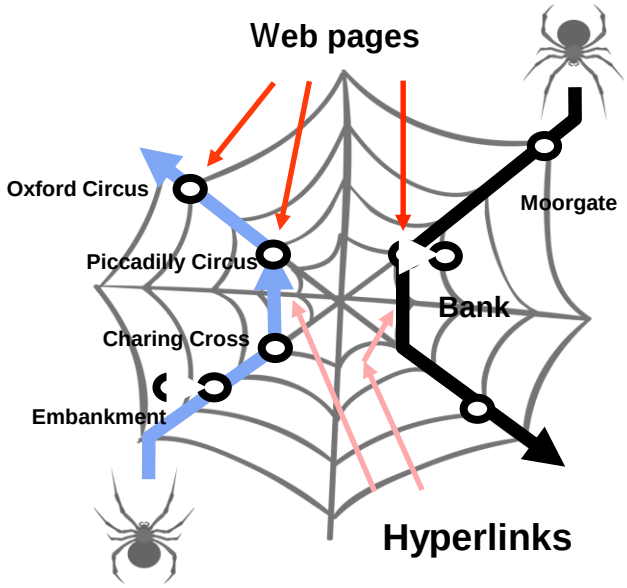
林佳緯

程式交易實務 – 使用 R 語言（一）

財團法人中華民國
證券暨期貨市場發展基金會
SECURITIES & FUTURES INSTITUTE

課程綱要

1. 期交所tick資料處理技巧
2. dplyr, reshape2 等資料清整套件介紹與使用
3. R 與 SQL關聯資料庫
4. 歷史資料匯入排程設計



期交所tick資料處理技巧

下載期交所盤後資料

http://www.taifex.com.tw/chinese/index.asp

交易資訊 -> 盤後資訊

前30日期每筆成交資料

https://www.taifex.com.tw/DailyDownload/DailyDownloadCSV/Daily_2017_04_05.zip

下載CSV檔案

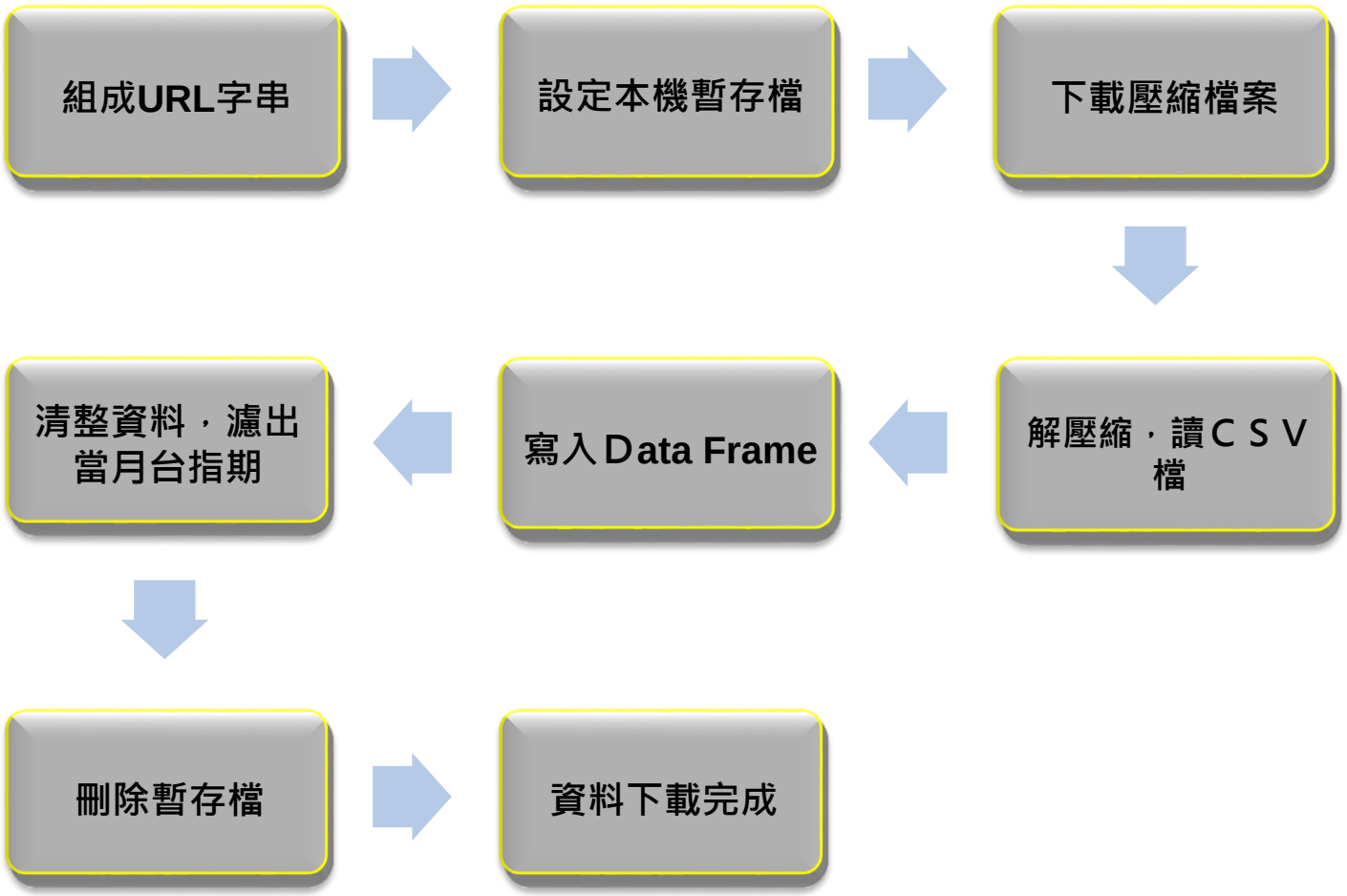
下載檔案位置

https://www.taifex.com.tw/DailyDownload/DailyDownloadCSV/Daily_2017_04_05.zip

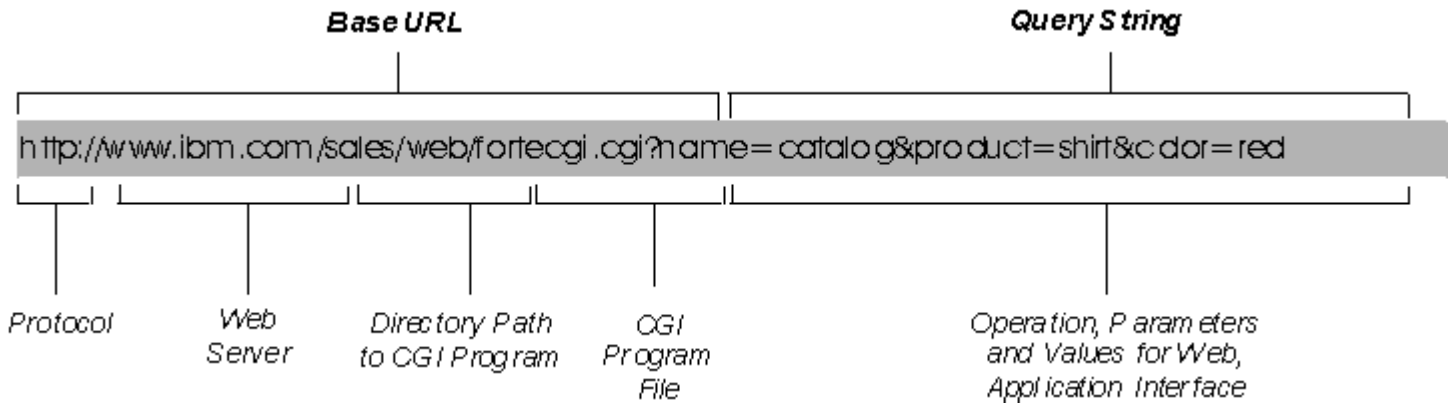
```
63              <td><input name="button7" type="button" class="btn_orange" id="button7" onClick="window.open('../../DailyDownload/DailyDownloadCSV/help/檔案開啟說明.doc')"
value="     檔案開啟說明     " ></td>
64              </tr>
65              </table>
66                註：本公司「前30個交易日期貨每筆成交資料(不含鉅額)」分別以*.rpt與*.csv兩種檔案格式提供下載。    <br><br>
67
68                  <table width="100%" border="0" cellpadding="0" cellspacing="0" class="table_c">
69                      <tr>
70                          <th width="60"> 時間 </th>
71                          <th width="60"> 日期 </th>
72                          <th width="60">下載(*.rpt)</th>
73                          <th width="60">下載(*.csv)</th>
74                      </tr>
75
76                      <tr>
77                          <td align="center">2016/7/12 下午 04:48:49</td>
78                          <td align="center">2016/07/12</td>
79
81                          <input name="button7" type="button" class="btn_orange" id="button7" onClick="window.open('../../DailyDownload/DailyDownload/Daily_2016_07_12.zip')" value="下載" ></td>
82                          <td align="center">
83                          <input name="button7" type="button" class="btn_orange" id="button7" onClick="window.open('../../DailyDownload/DailyDownloadCSV/Daily_2016_07_12.zip')" value="下載" ></td>
84
85
86
87                      </tr>
88
89                      <tr class='color'>
90                          <td align="center">2016/7/11 下午 04:45:40</td>
91                          <td align="center">2016/07/11</td>
92
93                          <td align="center">
94                          <input name="button7" type="button" class="btn_orange" id="button7" onClick="window.open('../../DailyDownload/DailyDownload/Daily_2016_07_11.zip')" value="下載" ></td>
95                          <td align="center">
96                          <input name="button7" type="button" class="btn_orange" id="button7" onClick="window.open('../../DailyDownload/DailyDownloadCSV/Daily_2016_07_11.zip')" value="下載" ></td>
```

Here you are!

程式流程圖


組成URL字串
設定本機暫存檔
下載壓縮檔案
解壓縮，讀ＣＳＶ檔
寫入Data Frame
清整資料，濾出當月台指期
刪除暫存檔
資料下載完成

組成URL字串



證交所目標URL:
https://www.taifex.com.tw/DailyDownload/DailyDownloadCSV/Daily_2017_04_05.zip

置換查詢字串:
tfeurl <- sprintf("https://www.taifex.com.tw/DailyDownload/
DailyDownloadCSV/Daily_%s_%s_%s.zip", qyryear, qrymonth, qrydate)

建立供下載之暫存檔

```
# 暫存檔名
tmpf <- sprintf("Daily_%s_%s_%s.csv", qyryear, qrymonth, qrydate)


# 解壓縮暫存檔
temp <- tempfile()
```

解壓縮，讀取CSV並清理暫存

```
# 解壓縮暫存檔
temp <- tempfile()

# 下載壓縮檔
download.file(tfeurl, temp)

tx <- read.csv(unz(temp, tmpf), header = FALSE, skip = 1, sep = ",")

unlink(temp)
```

後續資料清整過濾

```
# 清整資料
# Trim spaces
tx$V2 =trimws(tx$V2)
tx$V3 =trimws(tx$V3)

tx <- tx[(tx$V3 == "201607" & tx$V2 == "TX"),]

tx <- subset(tx, select = c("V1", "V4", "V5", "V6"))

colnames(tx) <- c("日期", "時間", "成交價", "成交量")

View(tx)
```

結果畫面

	日期	時間	成交價	成交量
72238	20160712	84500	8751	1140
72239	20160712	84500	8750	2
72240	20160712	84500	8751	10
72241	20160712	84500	8751	2
72242	20160712	84500	8751	2
72243	20160712	84500	8752	2
72244	20160712	84500	8752	4
72245	20160712	84500	8750	2
72246	20160712	84500	8750	2
72247	20160712	84500	8750	6
72248	20160712	84500	8752	2
72249	20160712	84500	8751	2
72250	20160712	84500	8750	4
72251	20160712	84500	8752	4
72252	20160712	84500	8750	2
72253	20160712	84500	8750	2
72254	20160712	84500	8751	2
72255	20160712	84500	8751	2
72256	20160712	84500	8750	2
72257	20160712	84500	8752	4
72258	20160712	84500	8751	2
72259	20160712	84500	8752	14
72260	20160712	84500	8752	2
72261	20160712	84500	8753	8
72262	20160712	84500	8753	4
72263	20160712	84500	8751	2
72264	20160712	84500	8751	2
72265	20160712	84500	8751	2
72266	20160712	84500	8751	2
72267	20160712	84500	8750	6

台指期tick資料

完整程式碼

```r
# 整理 期交所 期貨歷史資料CSV 檔
# qryear: 查詢年份
# qrymonth: 查詢月份
# qrydate: 查詢日期
#
# 期交所網站下載網址:
# "https://www.taifex.com.tw/chinese/3/3_1_3.asp"
#
# 期交所實際目標URL:
# "https://www.taifex.com.tw/DailyDownload/DailyDownloadCSV/Daily_2016_07_12.zip"
#
#
# 僅限 證基會課程學員教學使用 2016/07/25
#
# Julian Lin

symbolid <- 'TXF'; qryear <- '2016'; qrymonth <- '07';qrydate <- '12'

# 組成動態查詢字串
tfeurl <- sprintf("https://www.taifex.com.tw/DailyDownload/DailyDownloadCSV/Daily_%s_%s_%s.zip",
        qryear, qrymonth, qrydate)

# 暫存檔名
tmpf <- sprintf("Daily_%s_%s_%s.csv", qryear, qrymonth, qrydate)


# 解壓縮暫存檔
temp <- tempfile()

# 下載壓縮檔
download.file(tfeurl, temp)

tx <- read.csv(unz(temp, tmpf), header = FALSE, skip = 1, sep = ",")

unlink(temp)

# 清整資料
# Trim spaces
tx$V2 =trimws(tx$V2)
tx$V3 =trimws(tx$V3)

tx <- tx[(tx$V3 == "201607" & tx$V2 == "TX"),]

tx <- subset(tx, select = c("V1", "V4", "V5", "V6"))

colnames(tx) <- c("日期", "時間", "成交價", "成交量")

View(tx)

```

dplyr, reshape2 等資料清整套件介紹與使用

資料 1-2-3

- 小費(tips)統計資料隨機取樣

```
tips <- read.csv("http://www.ggobi.org/book/data/tips.csv")
n <- 15 #sample size
rows <- nrow(tips)
indx <- sample(seq(rows),n)
tips[indx,]
```

	obs	totbill	tip	sex	smoker	day	time	size
58	58	26.41	1.50	F	No	Sat	Night	2
30	30	19.65	3.00	F	No	Sat	Night	2
195	195	16.58	4.00	M	Yes	Thu	Day	2
224	224	15.98	3.00	F	No	Fri	Day	3
230	230	22.12	2.88	F	Yes	Sat	Night	2
210	210	12.76	2.23	F	Yes	Sat	Night	2
177	177	17.89	2.00	M	Yes	Sun	Night	2
120	120	24.08	2.92	F	No	Thu	Day	4
167	167	20.76	2.24	M	No	Sun	Night	2
235	235	15.53	3.00	M	Yes	Sat	Night	2
170	170	10.63	2.00	F	Yes	Sat	Night	2
188	188	30.46	2.00	M	Yes	Sun	Night	5
204	204	16.40	2.50	F	Yes	Thu	Day	2
77	77	17.92	3.08	M	Yes	Sat	Night	2
226	226	16.27	2.50	F	Yes	Fri	Day	2

資料 1-2-3

- 選取欄位

```
myColumns <- c("tip", "day", "size")
tips[indx, myColumns]
```

```
     tip day size
58  1.50 Sat    2
30  3.00 Sat    2
195 4.00 Thu    2
224 3.00 Fri    3
230 2.88 Sat    2
210 2.23 Sat    2
177 2.00 Sun    2
120 2.92 Thu    4
167 2.24 Sun    2
235 3.00 Sat    2
170 2.00 Sat    2
188 2.00 Sun    5
204 2.50 Thu    2
77  3.08 Sat    2
226 2.50 Fri    2
```

資料 1-2-3

- R subset 函數

```
subset(tips, size > 5, select = myColumns)
```

```
    tip day size
126 4.2 Thu    6
142 6.7 Thu    6
144 5.0 Thu    6
157 5.0 Sun    6
```

- [] + which 函數

```
tips[which(tips$size > 5),myColumns]
```

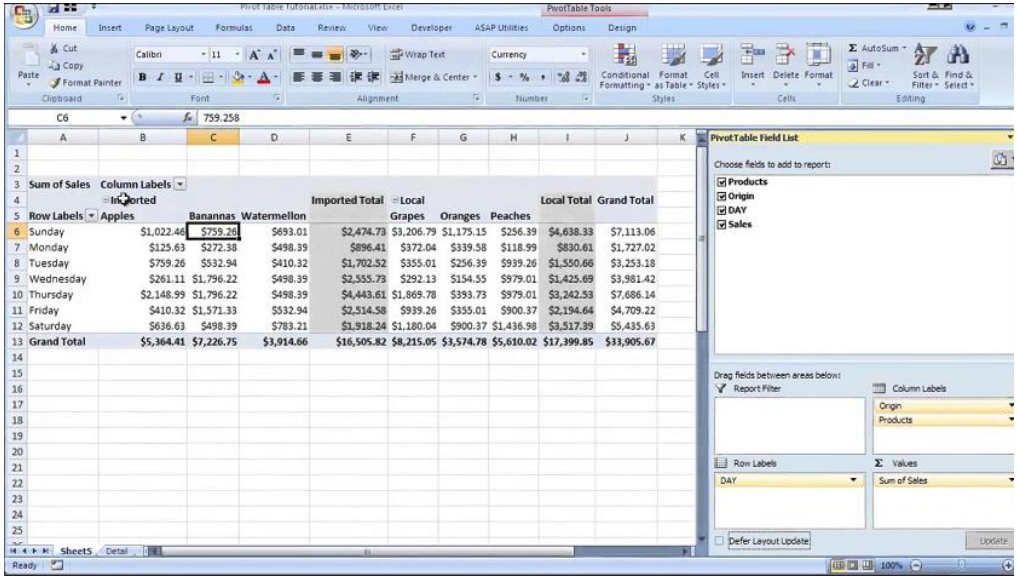
```
    tip day size
126 4.2 Thu    6
142 6.7 Thu    6
144 5.0 Thu    6
157 5.0 Sun    6
```

資料表呈現轉換 - reshape2

- 安裝套件

```
install.packages("reshape2")
library(reshape2)
```

- 重點函式
  - melt() 資料由寬變長
  - cast() 資料由長變寬



Sum of Sales	Column Labels								
	Imported			Imported Total	Local			Local Total	Grand Total
Row Labels	Apples	Bananas	Watermellon		Grapes	Oranges	Peaches		
Sunday	$1,022.46	$759.26	$693.01	$2,474.73	$3,206.79	$1,175.15	$256.39	$4,638.33	$7,113.06
Monday	$125.63	$272.38	$498.39	$896.41	$372.04	$339.58	$118.99	$830.61	$1,727.02
Tuesday	$759.26	$532.94	$410.32	$1,702.52	$355.01	$256.39	$939.26	$1,550.66	$3,253.18
Wednesday	$261.11	$1,796.22	$498.39	$2,555.73	$292.13	$154.55	$979.01	$1,425.69	$3,981.42
Thursday	$2,148.99	$1,796.22	$498.39	$4,443.61	$1,869.78	$393.73	$979.01	$3,242.53	$7,686.14
Friday	$410.32	$1,571.33	$532.94	$2,514.58	$939.26	$355.01	$900.37	$2,194.64	$4,709.22
Saturday	$636.63	$498.39	$783.21	$1,918.24	$1,180.04	$900.37	$1,436.98	$3,517.39	$5,435.63
Grand Total	$5,364.41	$7,226.75	$3,914.66	$16,505.82	$8,215.05	$3,574.78	$5,610.02	$17,399.85	$33,905.67

Excel 的樞紐功能(pivot table)!

資料表呈現轉換 - reshape2

- 空氣品質資料 airquality 範例
- melt() 實作

```
names(airquality) <- tolower(names(airquality))
head(airquality)
```

```
  ozone solar.r wind temp month day
1    41     190  7.4   67     5   1
2    36     118  8.0   72     5   2
3    12     149 12.6   74     5   3
4    18     313 11.5   62     5   4
5    NA      NA 14.3   56     5   5
6    28      NA 14.9   66     5   6
```

```
aql <- melt(airquality) # [a]ir [q]uality [l]ong format
head(aql)
tail(aql)
```

```
  variable value
1    ozone    41
2    ozone    36
3    ozone    12
4    ozone    18
5    ozone    NA
6    ozone    28
```

```
     variable value
913       day    25
914       day    26
915       day    27
916       day    28
917       day    29
918       day    30
```

資料表呈現轉換 - reshape2

```
aql <- melt(airquality, id.vars = c("month", "day"))
head(aql))
```

```
  month day variable value
1     5   1    ozone    41
2     5   2    ozone    36
3     5   3    ozone    12
4     5   4    ozone    18
5     5   5    ozone    NA
6     5   6    ozone    28
```

```
aql <- melt(airquality, id.vars = c("month", "day"),
  variable.name = "氣候變數",
  value.name = "氣候數值")
head(aql)
```

```
  month day 氣候變數 氣候數值
1     5   1    ozone       41
2     5   2    ozone       36
3     5   3    ozone       12
4     5   4    ozone       18
5     5   5    ozone       NA
6     5   6    ozone       28
```

資料表呈現轉換 - reshape2

- cast() 實作

```
aql <- melt(airquality, id.vars = c("month", "day"))
aqw <- dcast(aql, month + day ~ variable)
head(aqw)
```

```
  month day ozone solar.r wind temp
1     5   1    41     190  7.4   67
2     5   2    36     118  8.0   72
3     5   3    12     149 12.6   74
4     5   4    18     313 11.5   62
5     5   5    NA      NA 14.3   56
6     5   6    28      NA 14.9   66
```

```
head(airquality)
```

```
  ozone solar.r wind temp month day
1    41     190  7.4   67     5   1
2    36     118  8.0   72     5   2
3    12     149 12.6   74     5   3
4    18     313 11.5   62     5   4
5    NA      NA 14.3   56     5   5
6    28      NA 14.9   66     5   6
```

cast = melt 的資料轉置矩陣(data transposition)

資料把玩工具 - dplyr

- 安裝套件

```
install.packages("dplyr")
library(dplyr)
```

- 重點函式
  - filter()
  - select()
  - arrange()
  - mutate()
  - summarize()
- 其它函式
  - groub_by()
  - summarise_each()

資料把玩工具 - dplyr

- R 內建資料過濾

```
tips[tips$size==3 & tips$tip > 4, ]
```

```
    obs totbill   tip sex smoker day  time size
40   40   31.27  5.00   M     No Sat Night    3
171 171   50.81 10.00   M    Yes Sat Night    3
215 215   28.17  6.50   F    Yes Sat Night    3
239 239   35.83  4.67   F     No Sat Night    3
240 240   29.03  5.92   M     No Sat Night    3
```

- dplyr資料過濾

```
filter(tips, size==3, tip > 4)
```

```
  obs totbill   tip sex smoker day  time size
1  40   31.27  5.00   M     No Sat Night    3
2 171   50.81 10.00   M    Yes Sat Night    3
3 215   28.17  6.50   F    Yes Sat Night    3
4 239   35.83  4.67   F     No Sat Night    3
5 240   29.03  5.92   M     No Sat Night    3
```

資料把玩工具 - dplyr

- dplyr資料過濾

```
df <- head(tips, n=3)
select(df,tip,sex,size)
```

```
   tip sex size
1 1.01   F    2
2 1.66   M    3
3 3.50   M    3
```

```
select(df,sex:size)
```

```
  sex smoker day  time size
1   F     No Sun Night    2
2   M     No Sun Night    3
3   M     No Sun Night    3
```

資料把玩工具 - dplyr

- dplyr資料排序

```
df <- arrange(tips, totbill)
head(df, n=5)
```

```
  obs totbill  tip sex smoker day  time size
1  68    3.07 1.00   F    Yes Sat Night    1
2  93    5.75 1.00   F    Yes Fri Night    2
3 112    7.25 1.00   F     No Sat Night    1
4 173    7.25 5.15   M    Yes Sun Night    2
5 150    7.51 2.00   M     No Thu   Day    2
```

```
df <- arrange(tips, totbill, desc(tip))
head(df, n=5)
```

```
  obs totbill  tip sex smoker day  time size
1  68    3.07 1.00   F    Yes Sat Night    1
2  93    5.75 1.00   F    Yes Fri Night    2
3 173    7.25 5.15   M    Yes Sun Night    2
4 112    7.25 1.00   F     No Sat Night    1
5 150    7.51 2.00   M     No Thu   Day    2
```

資料把玩工具 - dplyr

- dplyr 新增計算欄位

```
df <- mutate(tips, tiprate = tip/totbill)
head(df)
```

```
  obs totbill  tip sex smoker day  time size    tiprate
1   1   16.99 1.01   F     No Sun Night    2 0.05944673
2   2   10.34 1.66   M     No Sun Night    3 0.16054159
3   3   21.01 3.50   M     No Sun Night    3 0.16658734
4   4   23.68 3.31   M     No Sun Night    2 0.13978041
5   5   24.59 3.61   F     No Sun Night    4 0.14680765
6   6   25.29 4.71   M     No Sun Night    4 0.18623962
>
```

- R 內建函式做法

```
tips$tiprate <- with(tips, tip/totbill)
head(tips)
```

```
  obs totbill  tip sex smoker day  time size    tiprate
1   1   16.99 1.01   F     No Sun Night    2 0.05944673
2   2   10.34 1.66   M     No Sun Night    3 0.16054159
3   3   21.01 3.50   M     No Sun Night    3 0.16658734
4   4   23.68 3.31   M     No Sun Night    2 0.13978041
5   5   24.59 3.61   F     No Sun Night    4 0.14680765
6   6   25.29 4.71   M     No Sun Night    4 0.18623962
```

資料把玩工具 - dplyr

- dplyr 管線%>%符號

```
tips %>%
  filter(size > 5) %>%
  select(totbill, tip, sex, day, size) %>%
  arrange(tip) %>%
  mutate(tiprate = tip/totbill)
```

```
  totbill tip sex day size   tiprate
1   29.80 4.2   F Thu    6 0.1409396
2   27.05 5.0   F Thu    6 0.1848429
3   48.17 5.0   M Sun    6 0.1037990
4   34.30 6.7   M Thu    6 0.1953353
```

資料把玩工具 - dplyr

- dplyr 複雜鎖鍊式資料計算

```
tips %>%
  filter(size > 3) %>%
  select(totbill, tip, sex, day, size) %>%
  arrange(tip) %>%
  mutate(tiprate=tip/totbill) %>%
    select(totbill, tip, sex, day, tiprate) %>%
    group_by(sex, day) %>%
    summarise_each(funs(mean)) %>%
    filter(tiprate > 0.12)
```

```
Source: local data frame [5 x 5]
Groups: sex [2]

     sex    day  totbill      tip   tiprate
  <fctr> <fctr>    <dbl>    <dbl>     <dbl>
1      F    Sun 29.90200 4.540000 0.1520376
2      F    Thu 31.77400 4.458000 0.1422925
3      M    Sat 30.47583 4.255833 0.1402830
4      M    Sun 27.06176 4.001176 0.1517739
5      M    Thu 30.80500 4.925000 0.1646549
```

R 與 SQL關聯資料庫

SQL是甚麼 ？

- SQL 是一種用來處理事實集以及其關係的電腦語言。
- 關聯式資料庫程式(例如Microsoft Office Access) 使用SQL 來處理資料。
- SQL 和許多電腦語言不同，它容易閱讀及理解，即使初學者也容易上手。 和許多電腦語言相同的是，SQL 是一種國際標準，由像是ISO 和ANSI 這種標準機構所認可。
- 常用動詞: SELECT, UPDATE, DELETE, INSERT
- 管理動詞: CREATE, ALTER, DROP
- 常用語法保留字: FROM, WHERE, GROUP BY, ORDER BY
- 常用SQL命令:

  SELECT <column names seperated by comma>
  FROM <database table>
  WHERE <condition>
  GROUP BY <column name>
  ORDER BY <column name>
- SELECT 內還可包含巢狀 SELECT 敘述

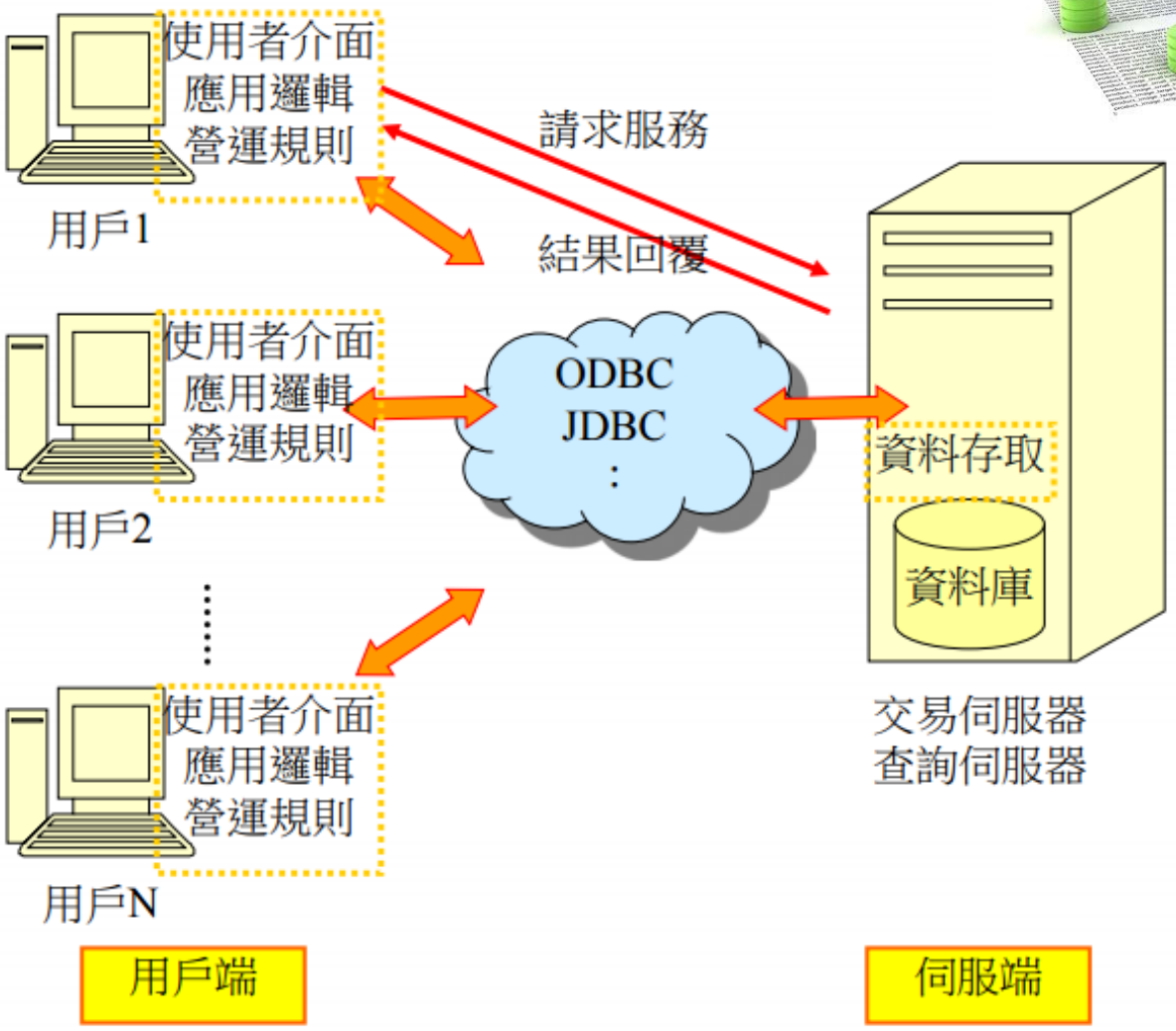
SQL 資料庫

- 資料庫系統（Database System）

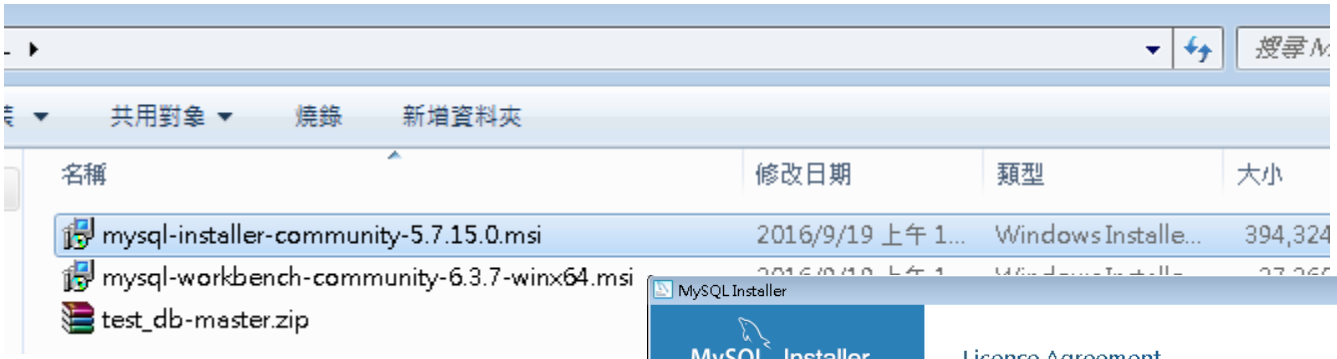
  是電腦化的資料儲存系統，使用者則透過各種應用程式來存取其中的資料。

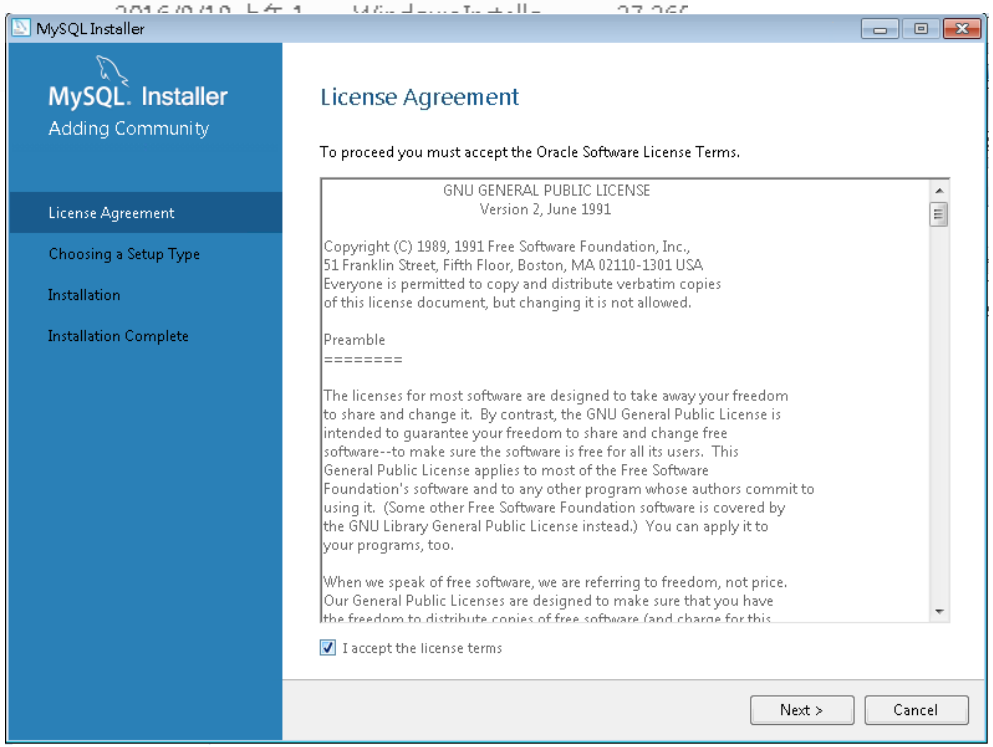
- 資料庫系統又可分為兩個部份：
  - 資料庫（Database）與
  - 資料庫管理系統（Database Management System, DBMS）。

資料庫的中央處理、分散式架構

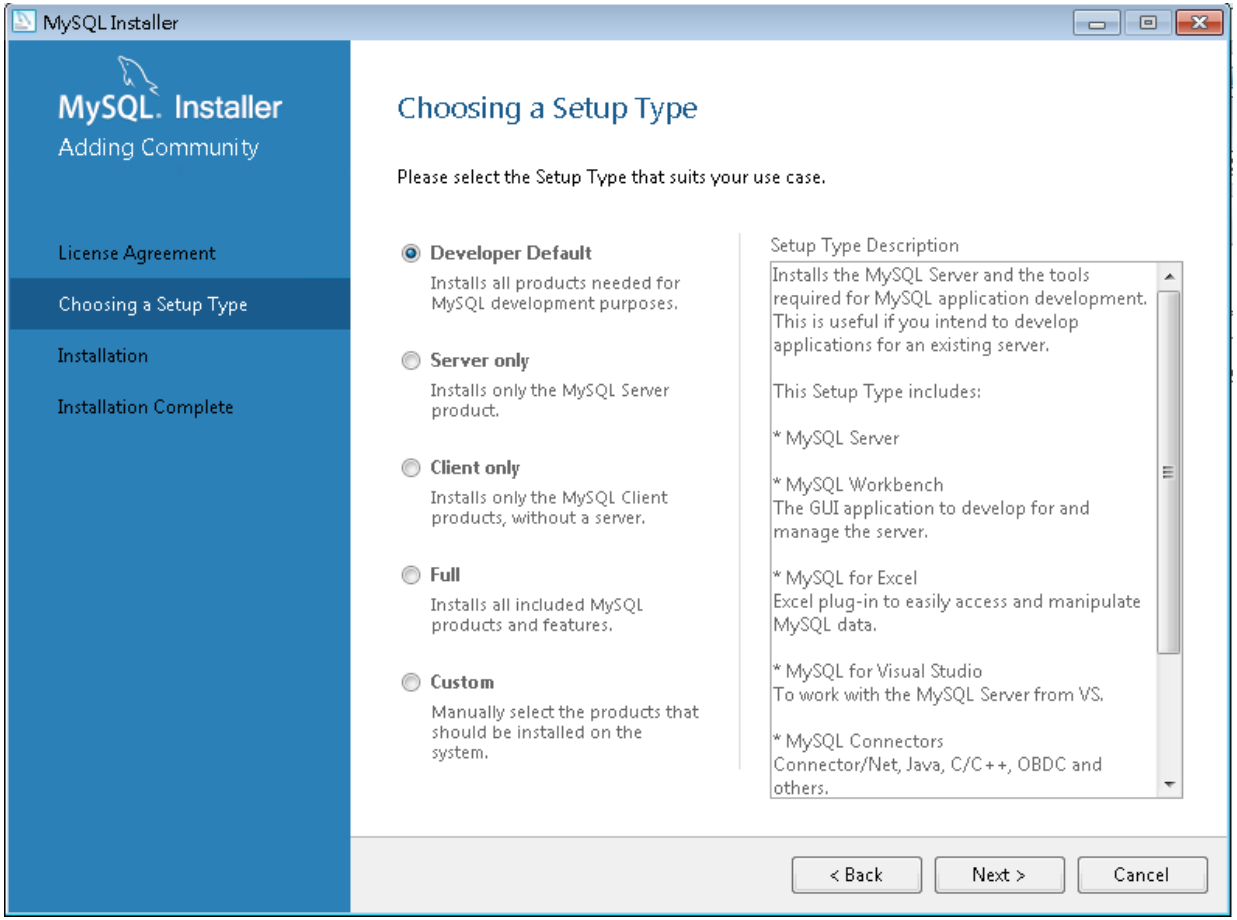


MySQL 資料庫安裝

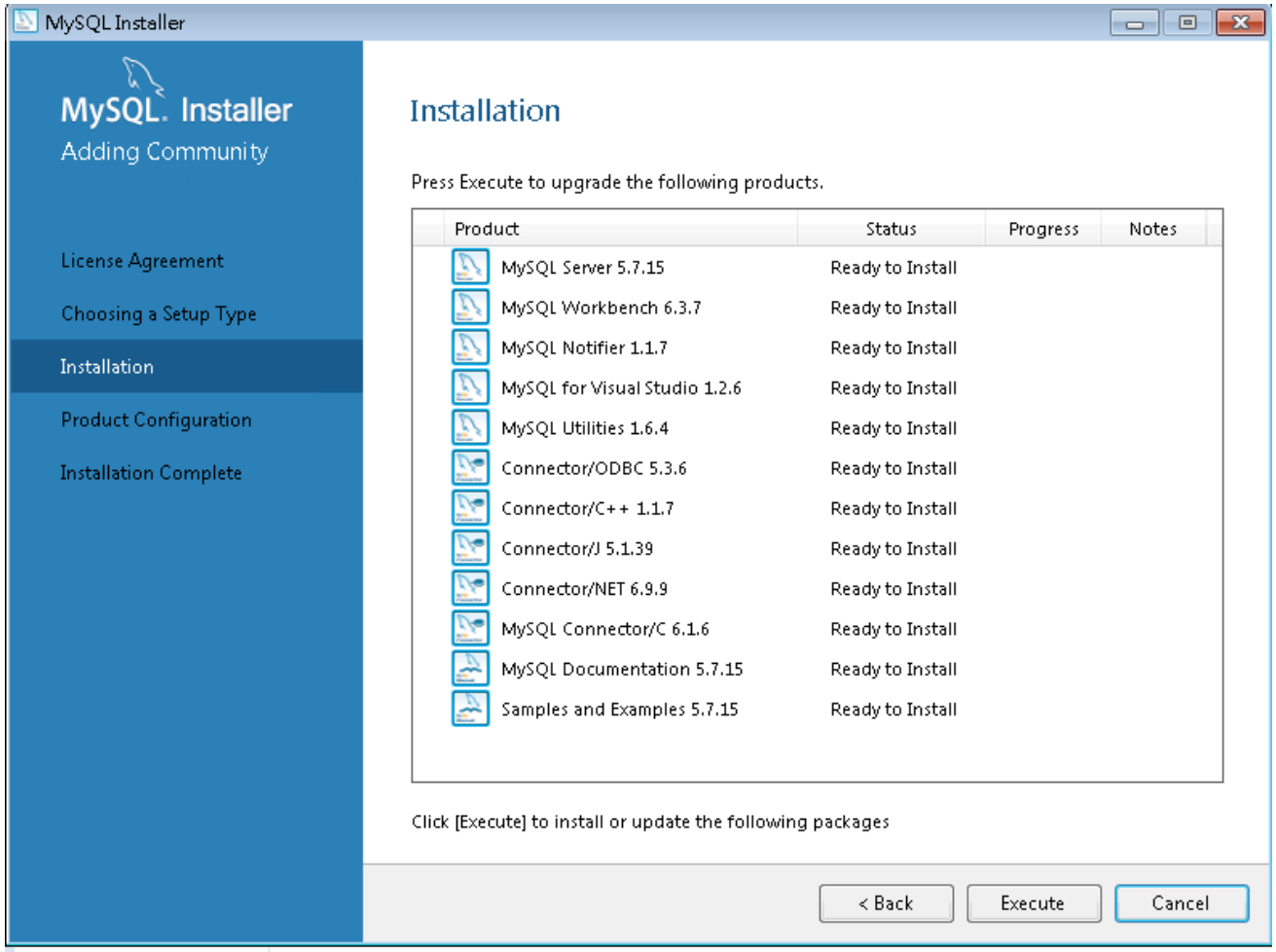

共用對象
燒錄
新增資料夾
名稱
修改日期
類型
大小
mysql-installer-community-5.7.15.0.msi
2016/9/19 上午 1...
Windows Installe...
394,324
mysql-workbench-community-6.3.7-winx64.msi
test_db-master.zip


MySQL Installer
MySQL. Installer
Adding Community
License Agreement
Choosing a Setup Type
Installation
Installation Complete
License Agreement
To proceed you must accept the Oracle Software License Terms.
GNU GENERAL PUBLIC LICENSE
Version 2, June 1991
Copyright (C) 1989, 1991 Free Software Foundation, Inc.,
51 Franklin Street, Fifth Floor, Boston, MA 02110-1301 USA
Everyone is permitted to copy and distribute verbatim copies
of this license document, but changing it is not allowed.
Preamble
The licenses for most software are designed to take away your freedom
to share and change it. By contrast, the GNU General Public License is
intended to guarantee your freedom to share and change free
software--to make sure the software is free for all its users. This
General Public License applies to most of the Free Software
Foundation's software and to any other program whose authors commit to
using it. (Some other Free Software Foundation software is covered by
the GNU Library General Public License instead.) You can apply it to
your programs, too.
When we speak of free software, we are referring to freedom, not price.
Our General Public Licenses are designed to make sure that you have
I accept the license terms
Next >
Cancel

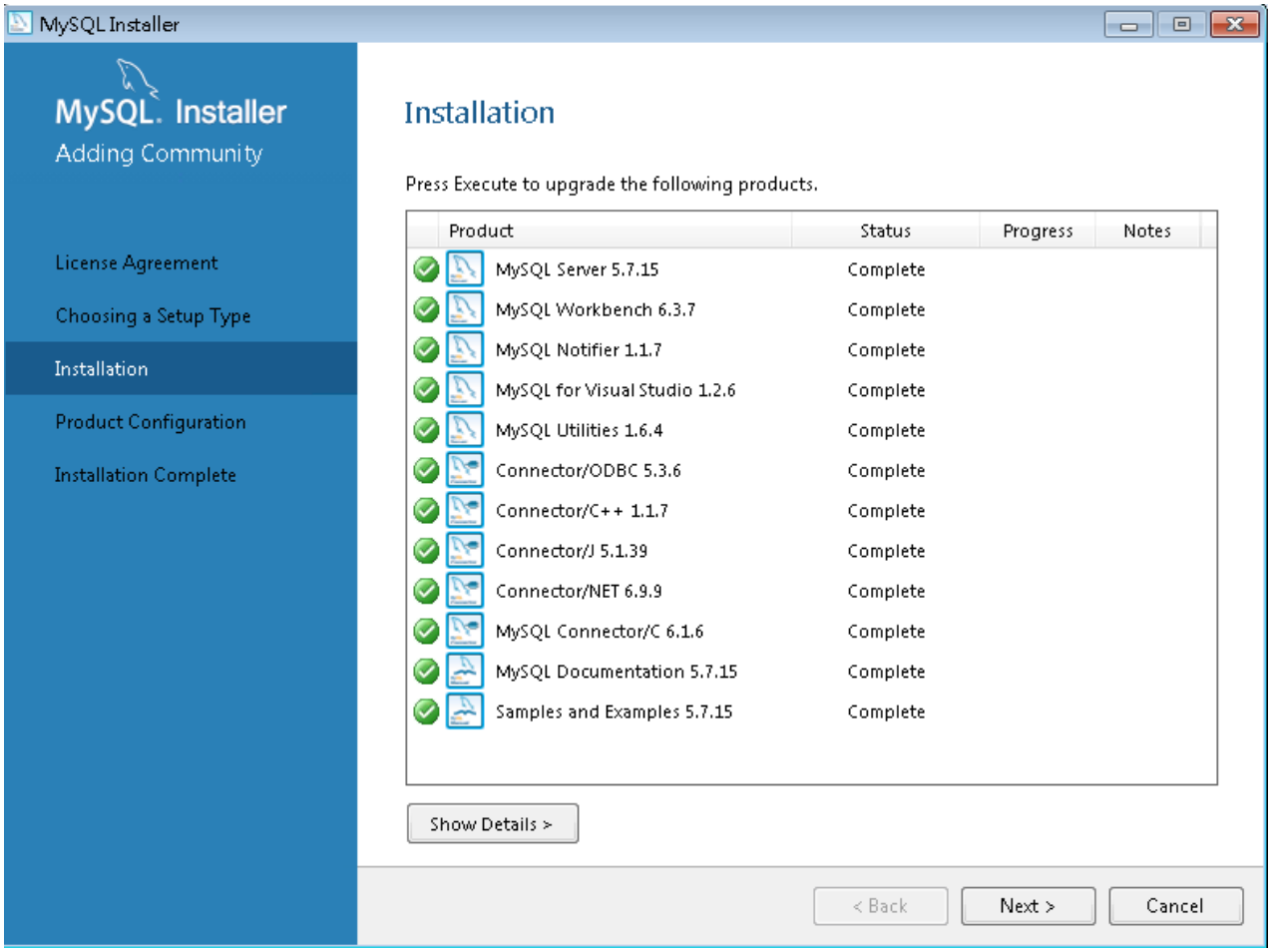
MySQL 安裝


MySQL Installer
MySQL. Installer
Adding Community
License Agreement
Choosing a Setup Type
Installation
Installation Complete
Choosing a Setup Type
Please select the Setup Type that suits your use case.
Developer Default
Installs all products needed for MySQL development purposes.
Server only
Installs only the MySQL Server product.
Client only
Installs only the MySQL Client products, without a server.
Full
Installs all included MySQL products and features.
Custom
Manually select the products that should be installed on the system.
Setup Type Description
Installs the MySQL Server and the tools required for MySQL application development. This is useful if you intend to develop applications for an existing server.
This Setup Type includes:
* MySQL Server
* MySQL Workbench
The GUI application to develop for and manage the server.
* MySQL for Excel
Excel plug-in to easily access and manipulate MySQL data.
* MySQL for Visual Studio
To work with the MySQL Server from VS.
* MySQL Connectors
Connector/Net, Java, C/C++, OBDC and others.
< Back
Next >
Cancel

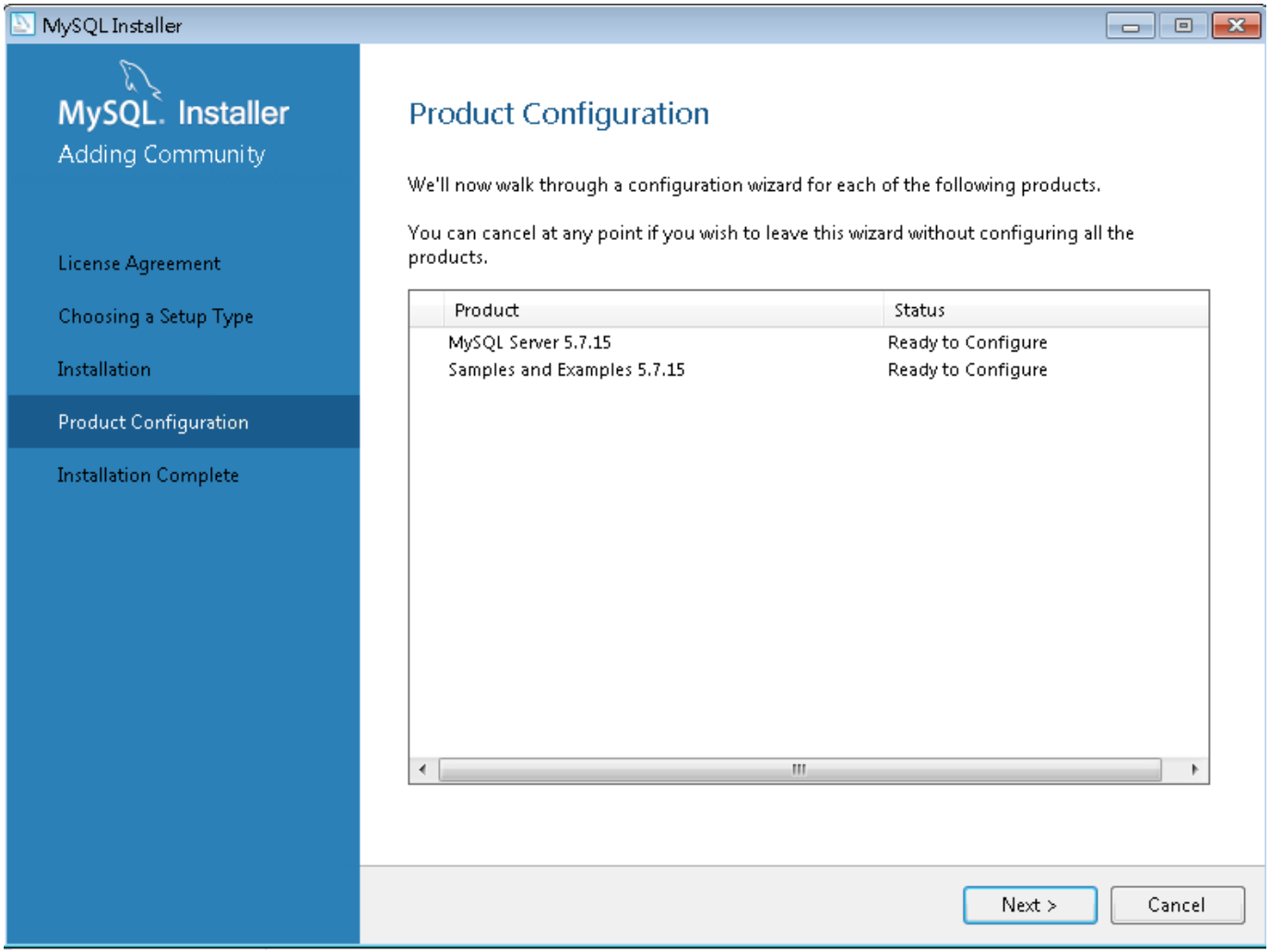
MySQL 安裝


MySQL Installer
MySQL. Installer
Adding Community
License Agreement
Choosing a Setup Type
Installation
Product Configuration
Installation Complete
Installation
Press Execute to upgrade the following products.
Product
Status
Progress
Notes
MySQL Server 5.7.15
Ready to Install
MySQL Workbench 6.3.7
Ready to Install
MySQL Notifier 1.1.7
Ready to Install
MySQL for Visual Studio 1.2.6
Ready to Install
MySQL Utilities 1.6.4
Ready to Install
Connector/ODBC 5.3.6
Ready to Install
Connector/C++ 1.1.7
Ready to Install
Connector/J 5.1.39
Ready to Install
Connector/NET 6.9.9
Ready to Install
MySQL Connector/C 6.1.6
Ready to Install
MySQL Documentation 5.7.15
Ready to Install
Samples and Examples 5.7.15
Ready to Install
Click [Execute] to install or update the following packages
< Back
Execute
Cancel

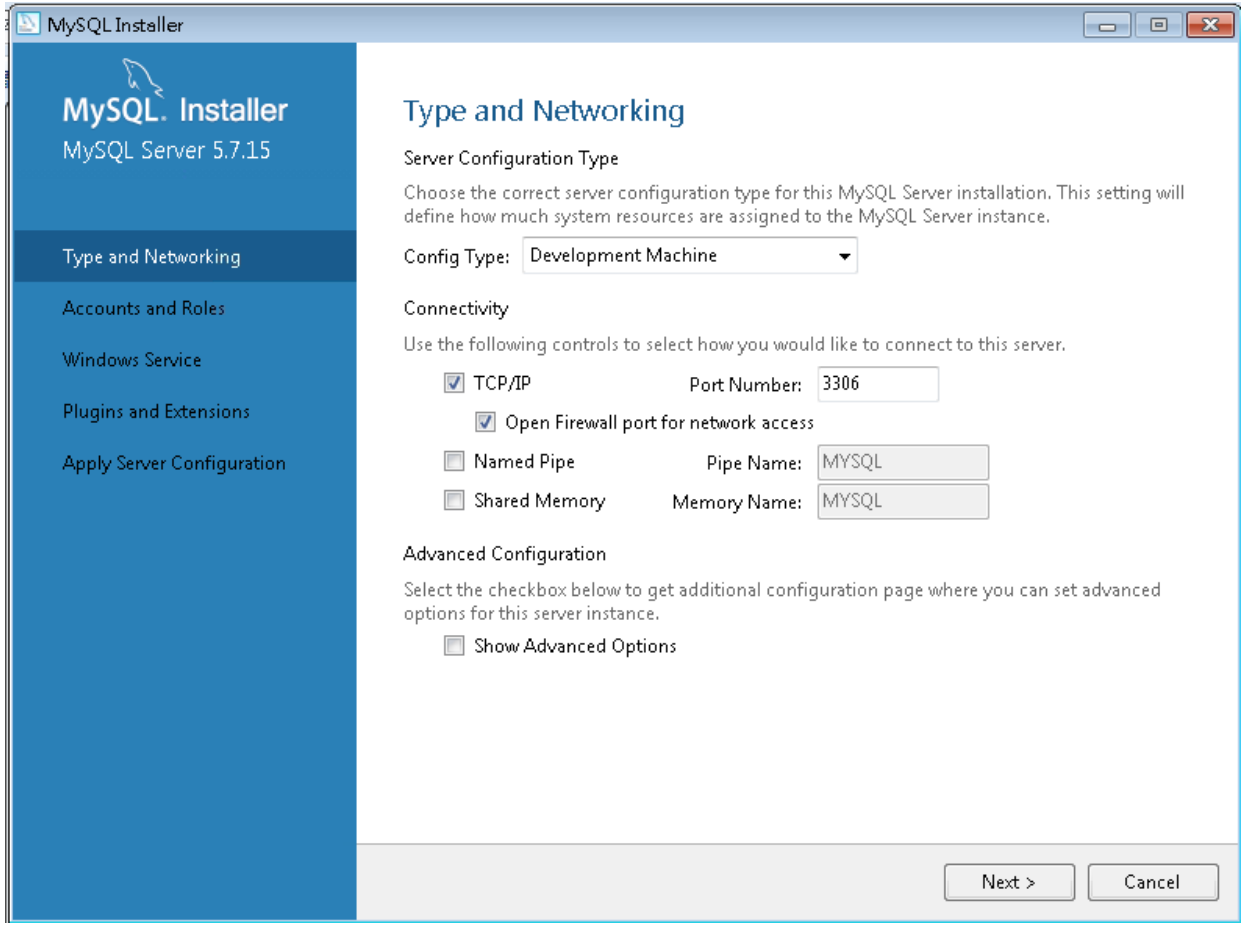
MySQL 安裝


MySQL Installer
MySQL. Installer
Adding Community
License Agreement
Choosing a Setup Type
Installation
Product Configuration
Installation Complete
Installation
Press Execute to upgrade the following products.
Product
Status
Progress
Notes
MySQL Server 5.7.15 Complete
MySQL Workbench 6.3.7 Complete
MySQL Notifier 1.1.7 Complete
MySQL for Visual Studio 1.2.6 Complete
MySQL Utilities 1.6.4 Complete
Connector/ODBC 5.3.6 Complete
Connector/C++ 1.1.7 Complete
Connector/J 5.1.39 Complete
Connector/NET 6.9.9 Complete
MySQL Connector/C 6.1.6 Complete
MySQL Documentation 5.7.15 Complete
Samples and Examples 5.7.15 Complete
Show Details >
< Back
Next >
Cancel

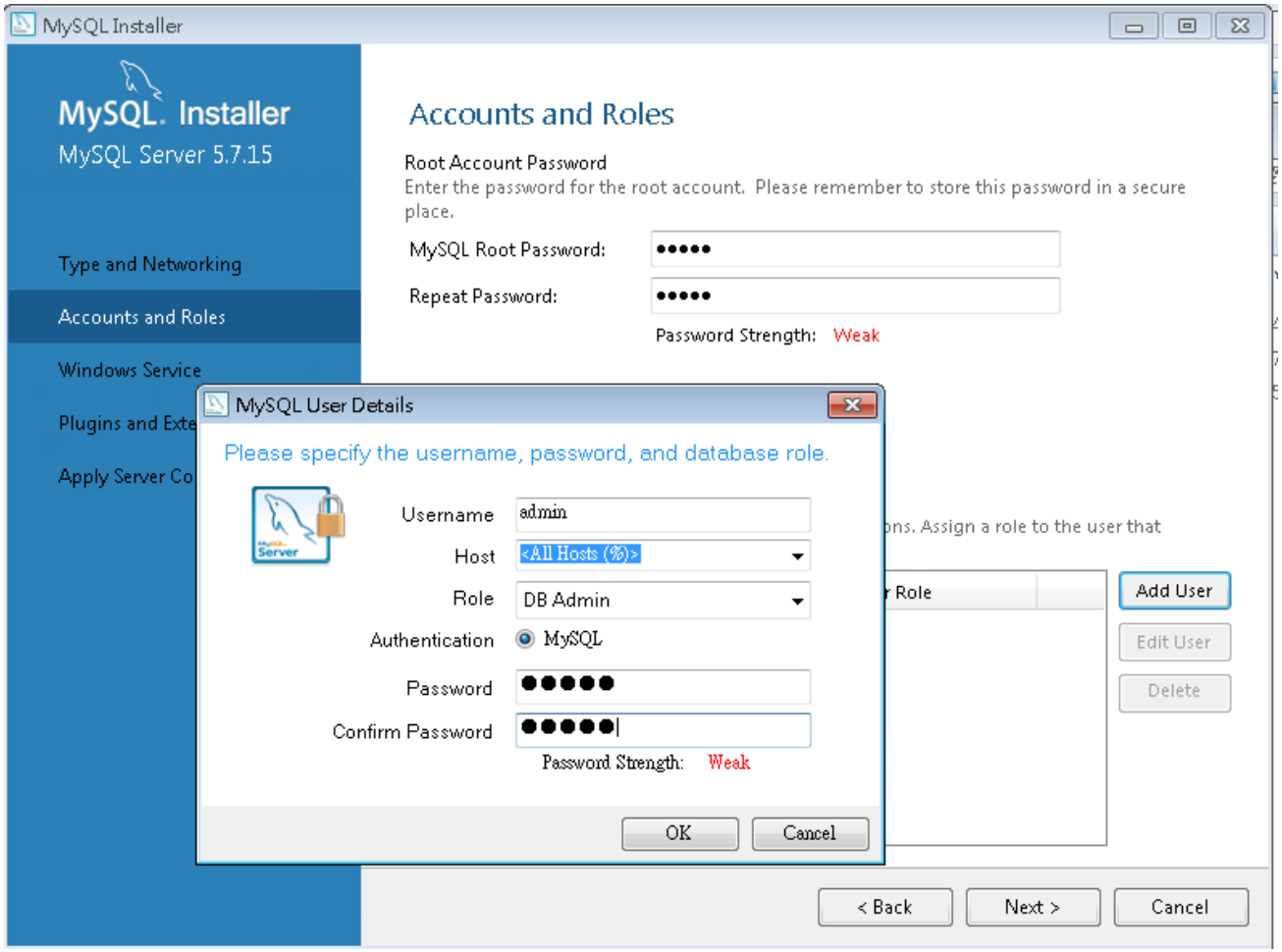
MySQL 安裝


MySQL Installer
MySQL. Installer
Adding Community
License Agreement
Choosing a Setup Type
Installation
Product Configuration
Installation Complete
Product Configuration
We'll now walk through a configuration wizard for each of the following products.
You can cancel at any point if you wish to leave this wizard without configuring all the products.
Product
Status
MySQL Server 5.7.15
Ready to Configure
Samples and Examples 5.7.15
Ready to Configure
Next >
Cancel

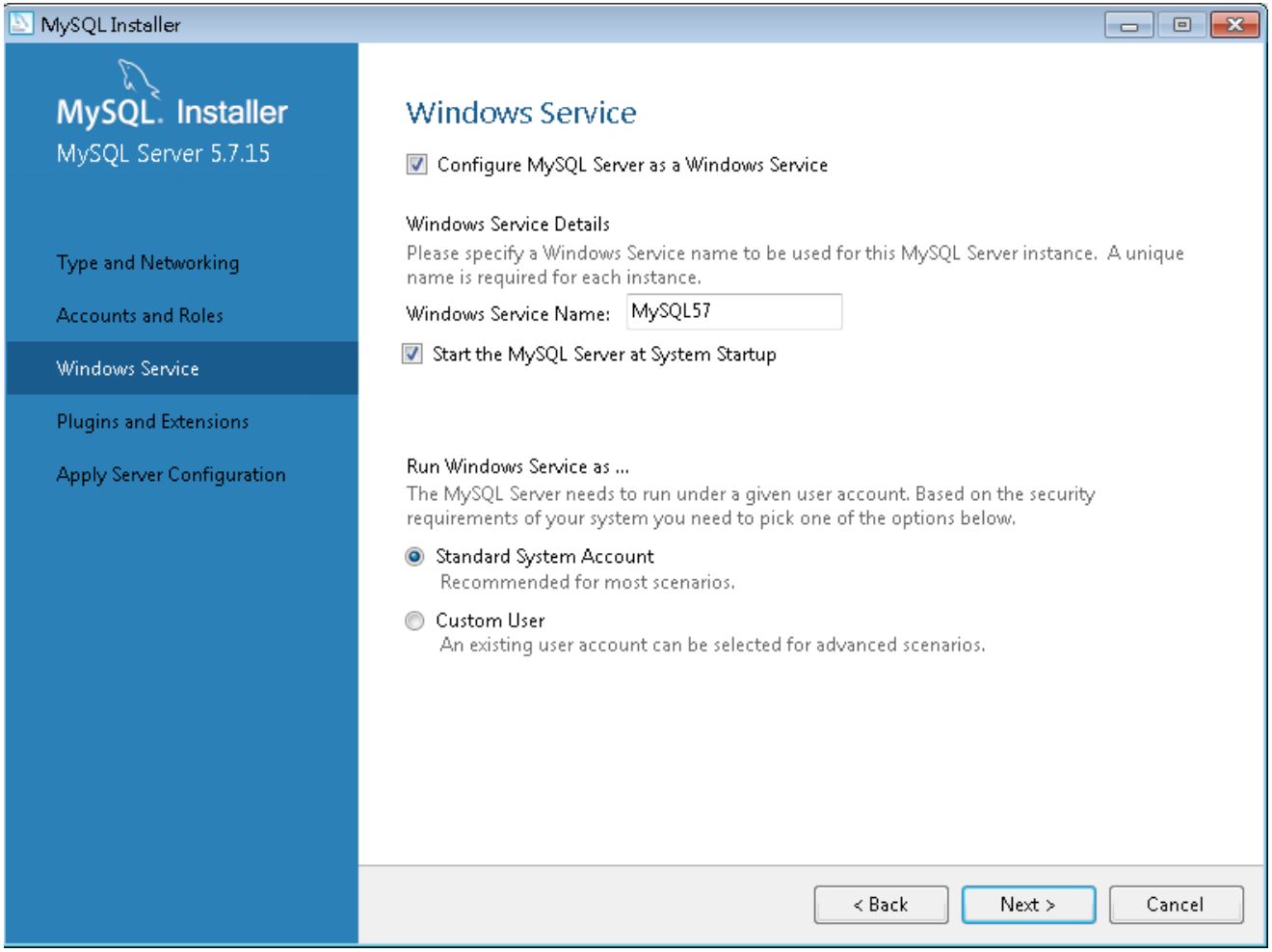
MySQL 安裝


MySQL Installer
MySQL. Installer
MySQL Server 5.7.15
Type and Networking
Accounts and Roles
Windows Service
Plugins and Extensions
Apply Server Configuration
Type and Networking
Server Configuration Type
Choose the correct server configuration type for this MySQL Server installation. This setting will define how much system resources are assigned to the MySQL Server instance.
Config Type: Development Machine
Connectivity
Use the following controls to select how you would like to connect to this server.
TCP/IP
Port Number: 3306
Open Firewall port for network access
Named Pipe
Pipe Name: MYSQL
Shared Memory
Memory Name: MYSQL
Advanced Configuration
Select the checkbox below to get additional configuration page where you can set advanced options for this server instance.
Show Advanced Options
Next >
Cancel

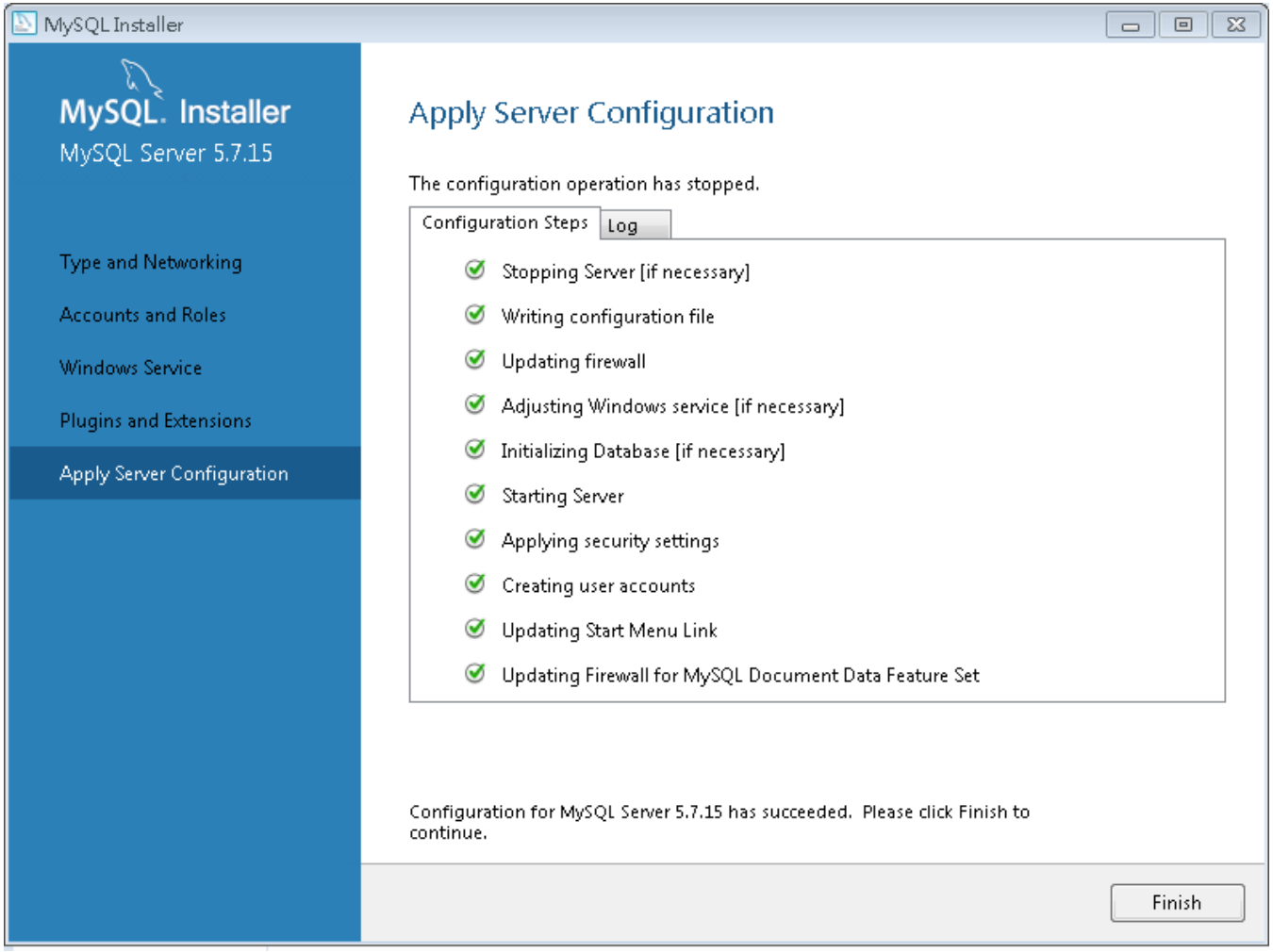
MySQL 安裝


MySQL Installer
MySQL. Installer
MySQL Server 5.7.15
Type and Networking
Accounts and Roles
Windows Service
Plugins and Exte
Apply Server Co
Accounts and Roles
Root Account Password
Enter the password for the root account. Please remember to store this password in a secure place.
MySQL Root Password:
Repeat Password:
Password Strength: Weak
MySQL User Details
Please specify the username, password, and database role.
Username
admin
Host
<All Hosts (%)>
Role
DB Admin
Authentication
MySQL
Password
Confirm Password
Password Strength: Weak
OK
Cancel
ons. Assign a role to the user that
Role
Add User
Edit User
Delete
< Back
Next >
Cancel

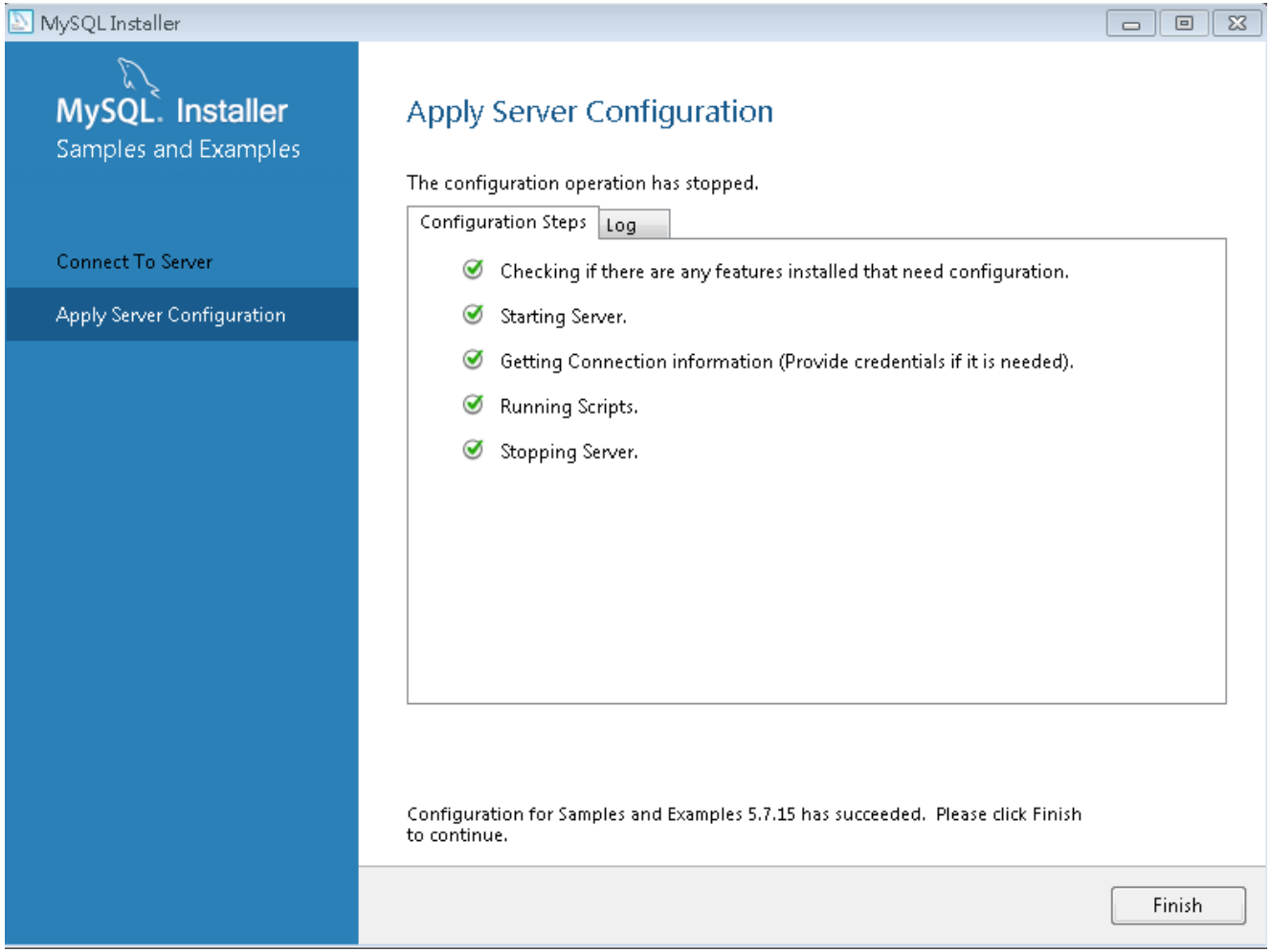
MySQL 安裝


MySQL Installer
MySQL. Installer
MySQL Server 5.7.15
Type and Networking
Accounts and Roles
Windows Service
Plugins and Extensions
Apply Server Configuration
Windows Service
Configure MySQL Server as a Windows Service
Windows Service Details
Please specify a Windows Service name to be used for this MySQL Server instance. A unique name is required for each instance.
Windows Service Name: MySQL57
Start the MySQL Server at System Startup
Run Windows Service as ...
The MySQL Server needs to run under a given user account. Based on the security requirements of your system you need to pick one of the options below.
Standard System Account
Recommended for most scenarios.
Custom User
An existing user account can be selected for advanced scenarios.
< Back
Next >
Cancel

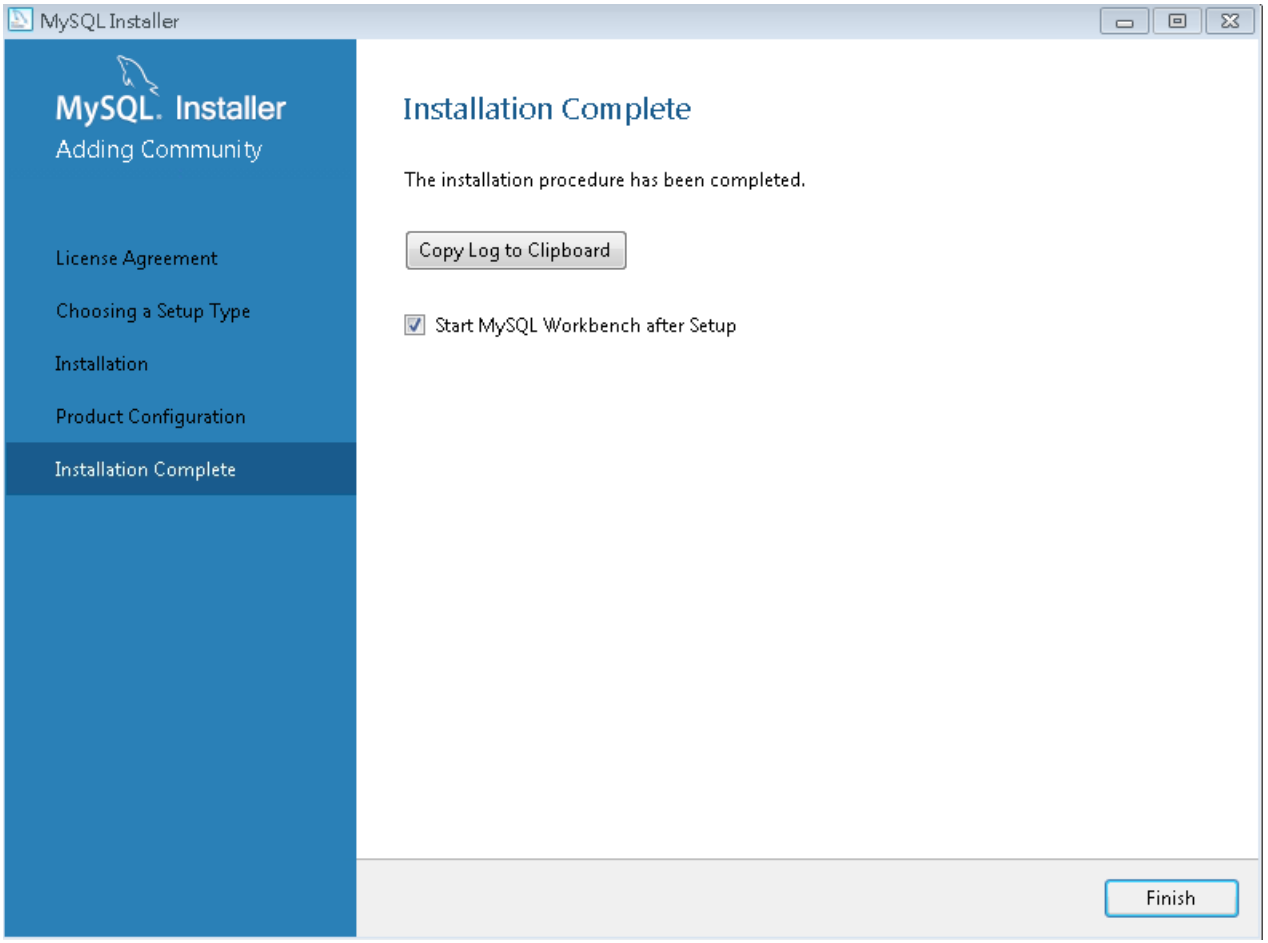
MySQL 安裝


MySQL Installer
MySQL. Installer
MySQL Server 5.7.15
Type and Networking
Accounts and Roles
Windows Service
Plugins and Extensions
Apply Server Configuration
Apply Server Configuration
The configuration operation has stopped.
Configuration Steps
Log
Stopping Server [if necessary]
Writing configuration file
Updating firewall
Adjusting Windows service [if necessary]
Initializing Database [if necessary]
Starting Server
Applying security settings
Creating user accounts
Updating Start Menu Link
Updating Firewall for MySQL Document Data Feature Set
Configuration for MySQL Server 5.7.15 has succeeded. Please click Finish to continue.
Finish

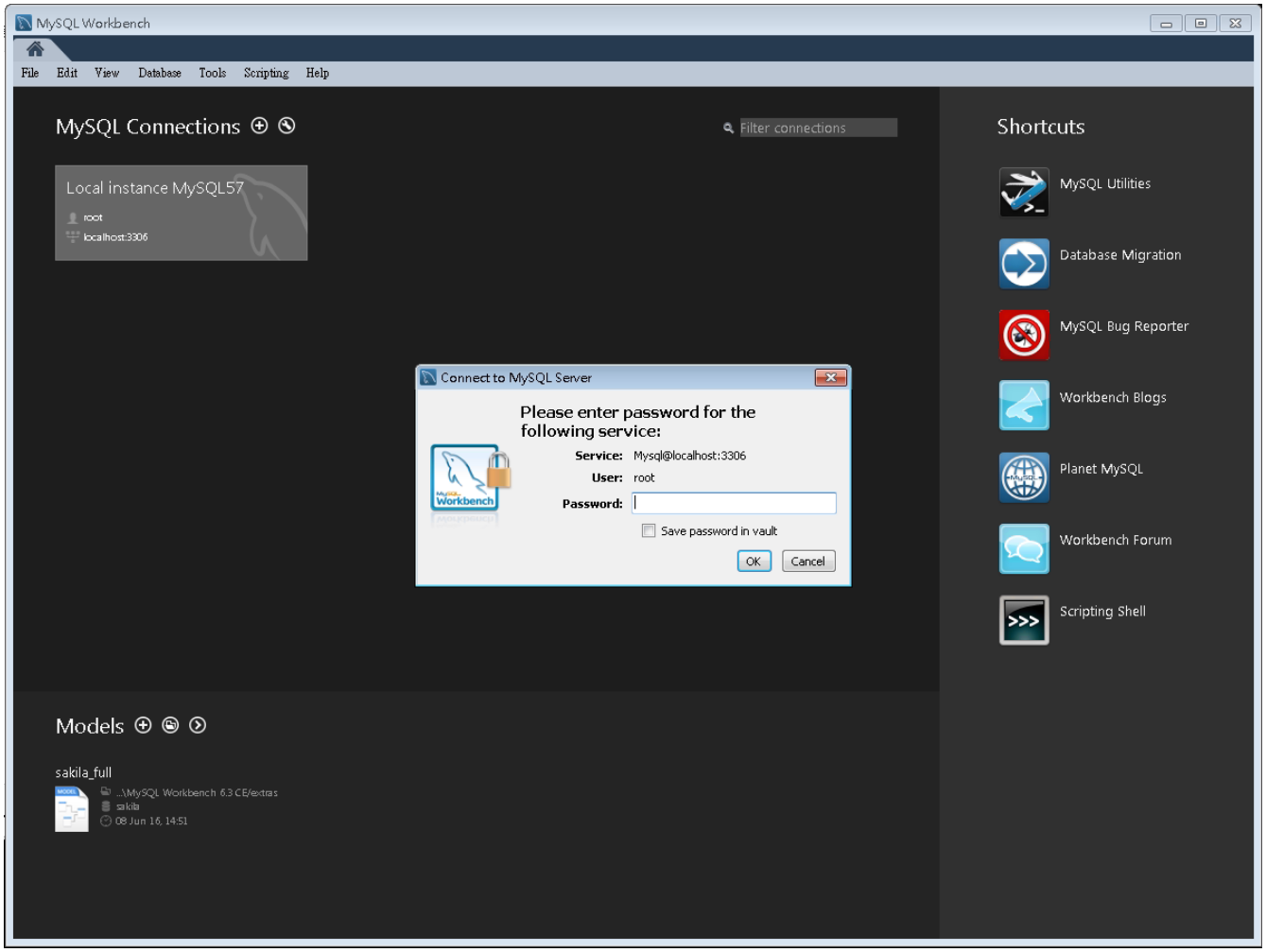
MySQL 安裝


MySQL Installer
MySQL. Installer
Samples and Examples
Connect To Server
Apply Server Configuration
Apply Server Configuration
The configuration operation has stopped.
Configuration Steps
Log
Checking if there are any features installed that need configuration.
Starting Server.
Getting Connection information (Provide credentials if it is needed).
Running Scripts.
Stopping Server.
Configuration for Samples and Examples 5.7.15 has succeeded. Please click Finish to continue.
Finish

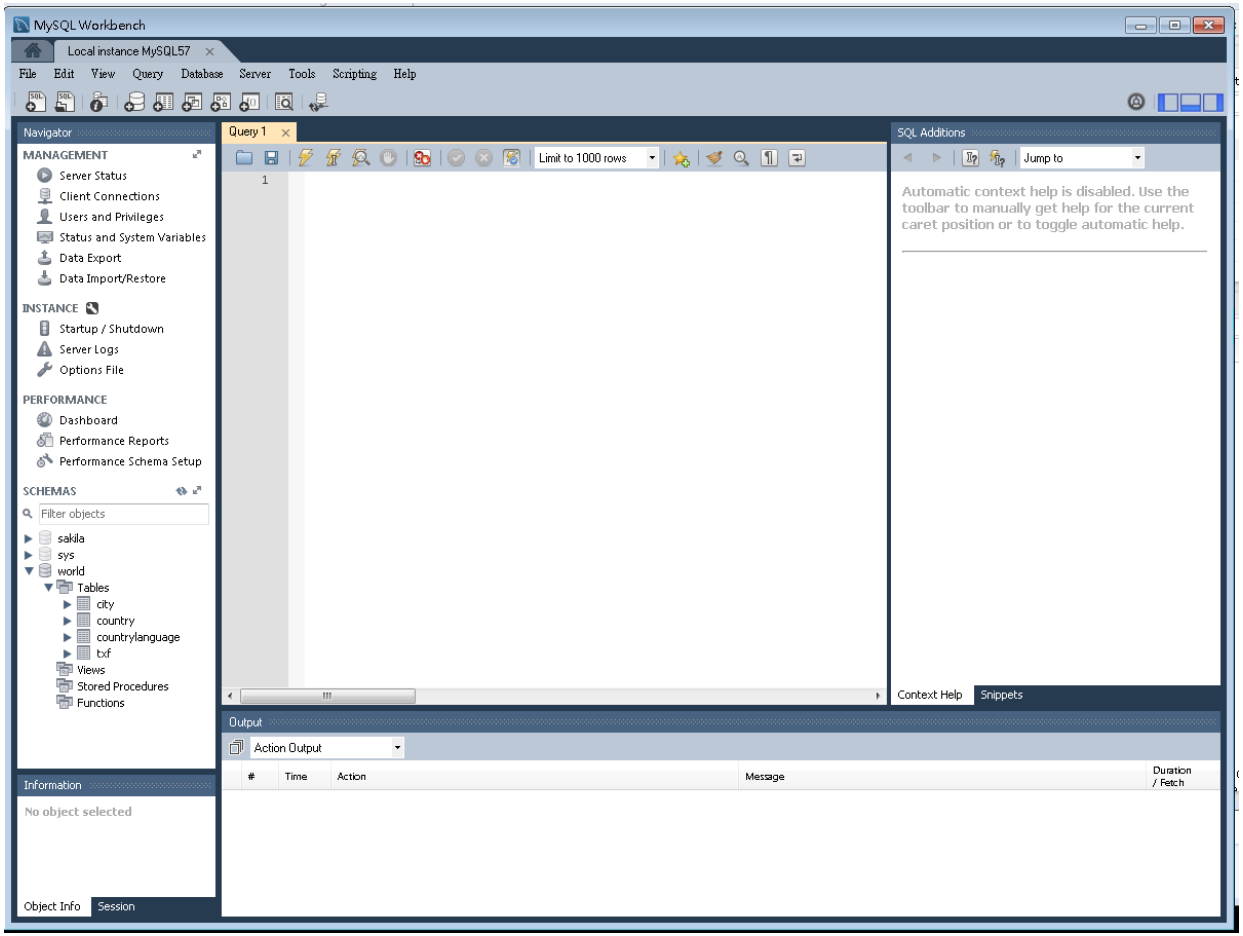
MySQL 安裝


MySQL Installer
MySQL. Installer
Adding Community
License Agreement
Choosing a Setup Type
Installation
Product Configuration
Installation Complete
Installation Complete
The installation procedure has been completed.
Copy Log to Clipboard
Start MySQL Workbench after Setup
Finish

MySQL Workbench 啟動

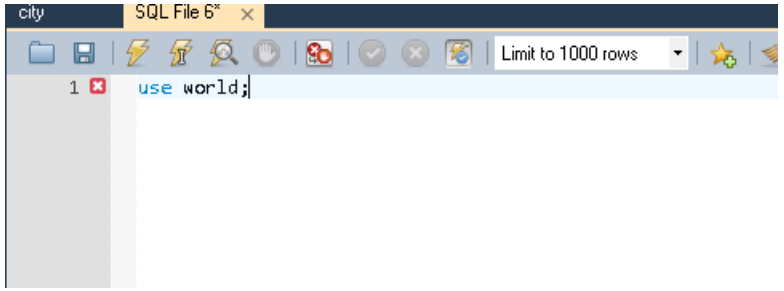


MySQL Workbench 工作區



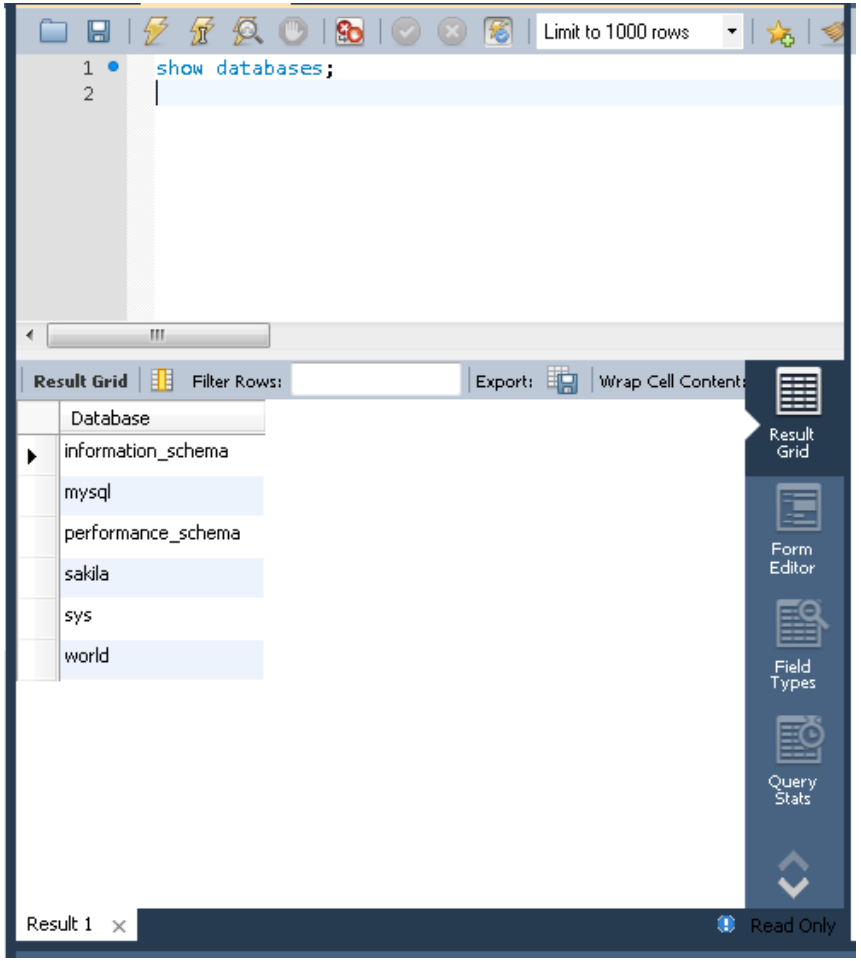
SQL 基礎指令介紹

指定目前使用的資料庫



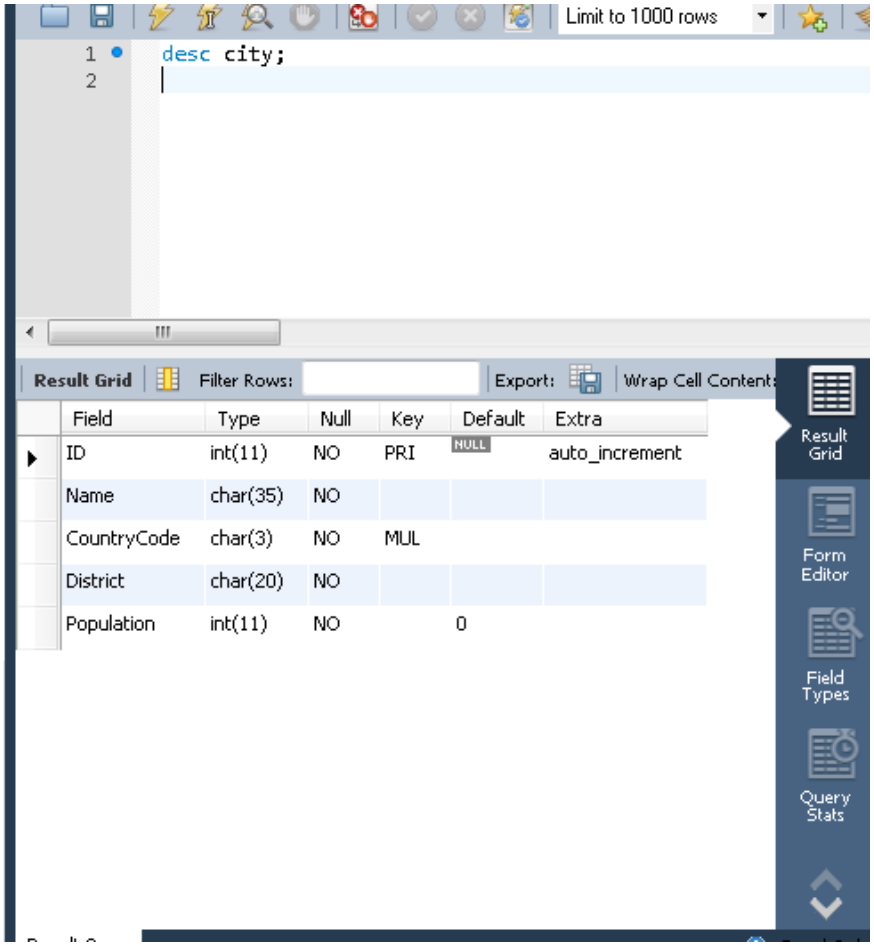
SQL 基礎指令介紹

顯示所有資料庫



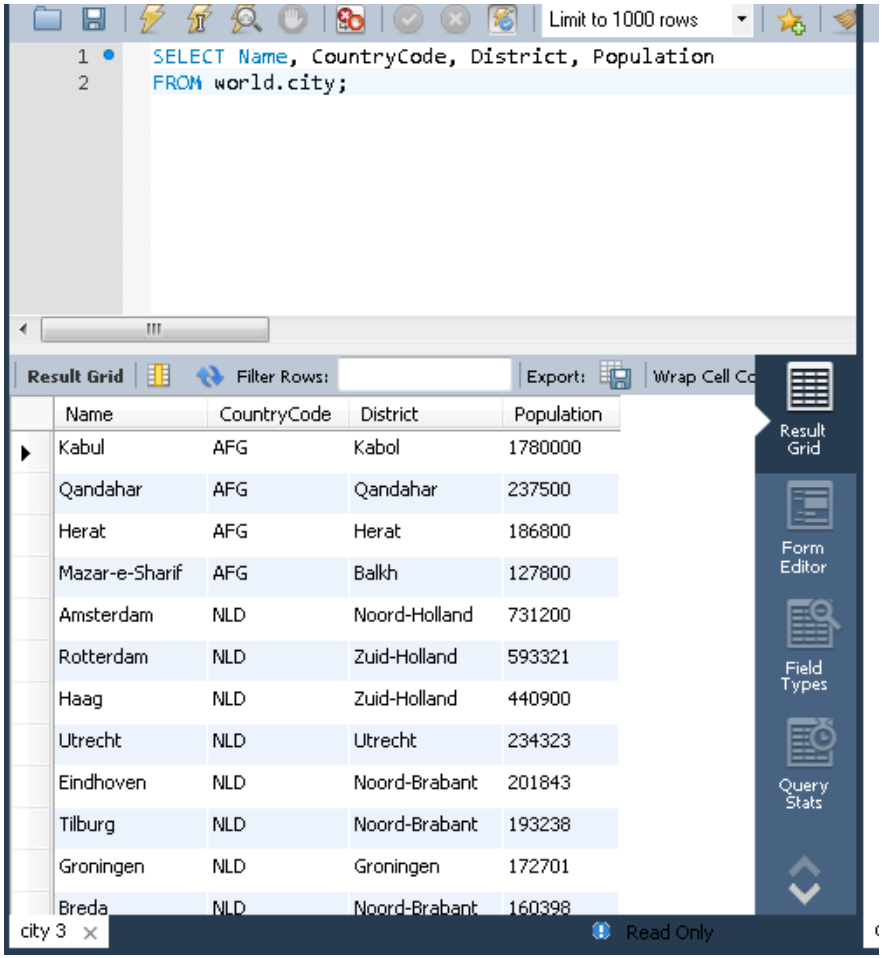
SQL 基礎指令介紹

顯示資料表結構



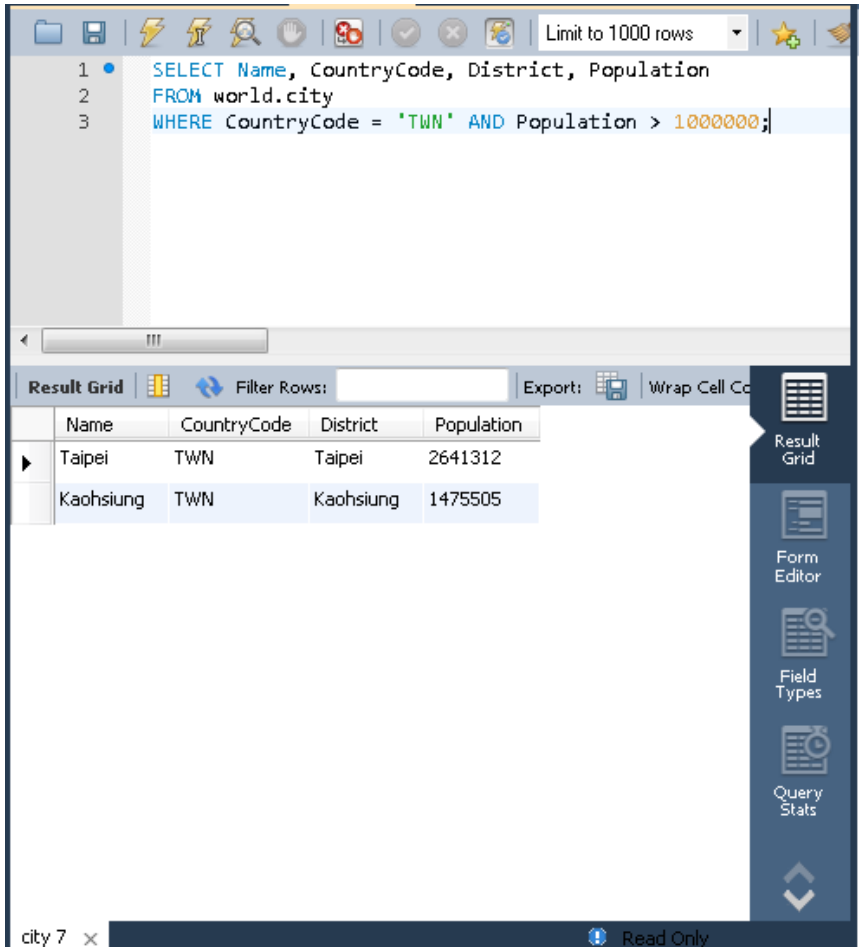
SQL 基礎指令介紹

選取及過濾資料庫資料

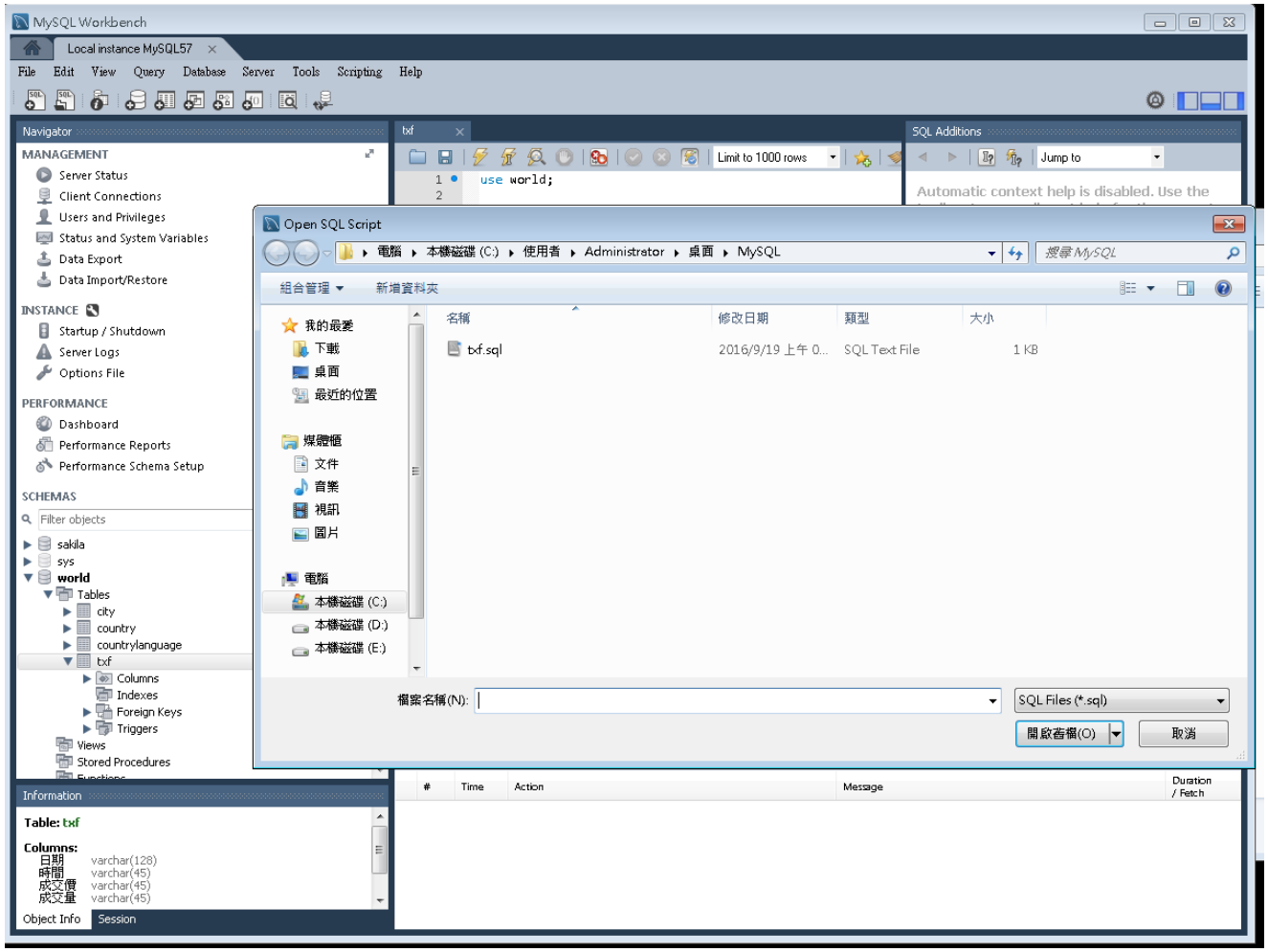


SQL 基礎指令介紹

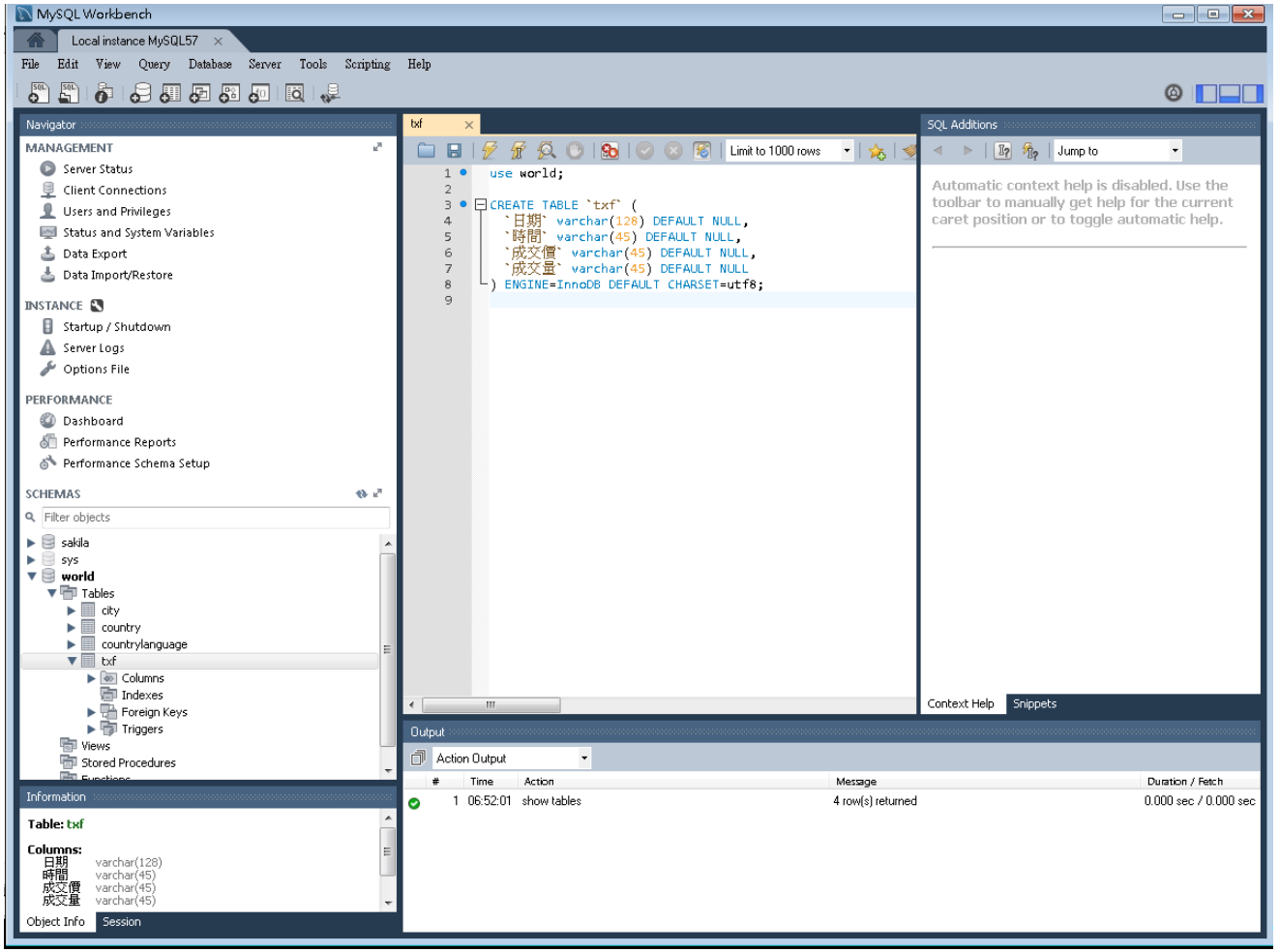
選取及過濾:
國家為 TWN 且 人口數 大於 100萬 的資料庫資料



新建立 TXF Table



TXF Table 新增


MySQL Workbench
Local instance MySQL57
File Edit View Query Database Server Tools Scripting Help
Navigator
MANAGEMENT
Server Status
Client Connections
Users and Privileges
Status and System Variables
Data Export
Data Import/Restore
INSTANCE
Startup / Shutdown
Server Logs
Options File
PERFORMANCE
Dashboard
Performance Reports
Performance Schema Setup
SCHEMAS
Filter objects
sakila
sys
world
Tables
city
country
countrylanguage
txf
Columns
Indexes
Foreign Keys
Triggers
Views
Stored Procedures
Information
Table: txf
Columns:
日期 varchar(128)
時間 varchar(45)
成交價 varchar(45)
成交量 varchar(45)
Object Info
Session
txf
Limit to 1000 rows
use world;
CREATE TABLE `txf` (
`日期` varchar(128) DEFAULT NULL,
`時間` varchar(45) DEFAULT NULL,
`成交價` varchar(45) DEFAULT NULL,
`成交量` varchar(45) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
SQL Additions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Context Help
Snippets
Output
Action Output
Time Action Message Duration / Fetch
1 06:52:01 show tables 4 row(s) returned 0.000 sec / 0.000 sec

R 與 MySQL資料庫連線

安裝RMySQL 套件

```
install.packages("RMySQL")
install.packages("dplyr")
```

RMySQL 連線MySQL資料庫

```
library(RMySQL)

con = dbConnect(MySQL(),user="admin", password="admin",
                dbname="world", host="localhost")

query <- "SELECT Name, CountryCode, District, Population
FROM world.city
WHERE CountryCode = 'TWN' AND Population > 500000;"

myData <- dbGetQuery(con, query)

View(myData)
```

查詢結果:

Filter

	Name	CountryCode	District	Population
1	Taipei	TWN	Taipei	2641312
2	Kaohsiung	TWN	Kaohsiung	1475505
3	Taichung	TWN	Taichung	940589
4	Tainan	TWN	Tainan	728060
5	Panchiao	TWN	Taipei	523850

R 與 MySQL資料庫連線

dplyr 連線MySQL資料庫

```
library(dplyr)
conDplyr = src_mysql(dbname = "world", user = "admin",
                     password = "admin", host = "localhost")

myData <- conDplyr %>%
  tbl("city") %>%
  select(Name, CountryCode, District, Population) %>%
  filter(CountryCode == 'TWN', Population > 500000 ) %>%
  collect()

View(myData)
```

查詢結果:

Filter

	Name	CountryCode	District	Population
1	Taipei	TWN	Taipei	2641312
2	Kaohsiung	TWN	Kaohsiung	1475505
3	Taichung	TWN	Taichung	940589
4	Tainan	TWN	Tainan	728060
5	Panchiao	TWN	Taipei	523850

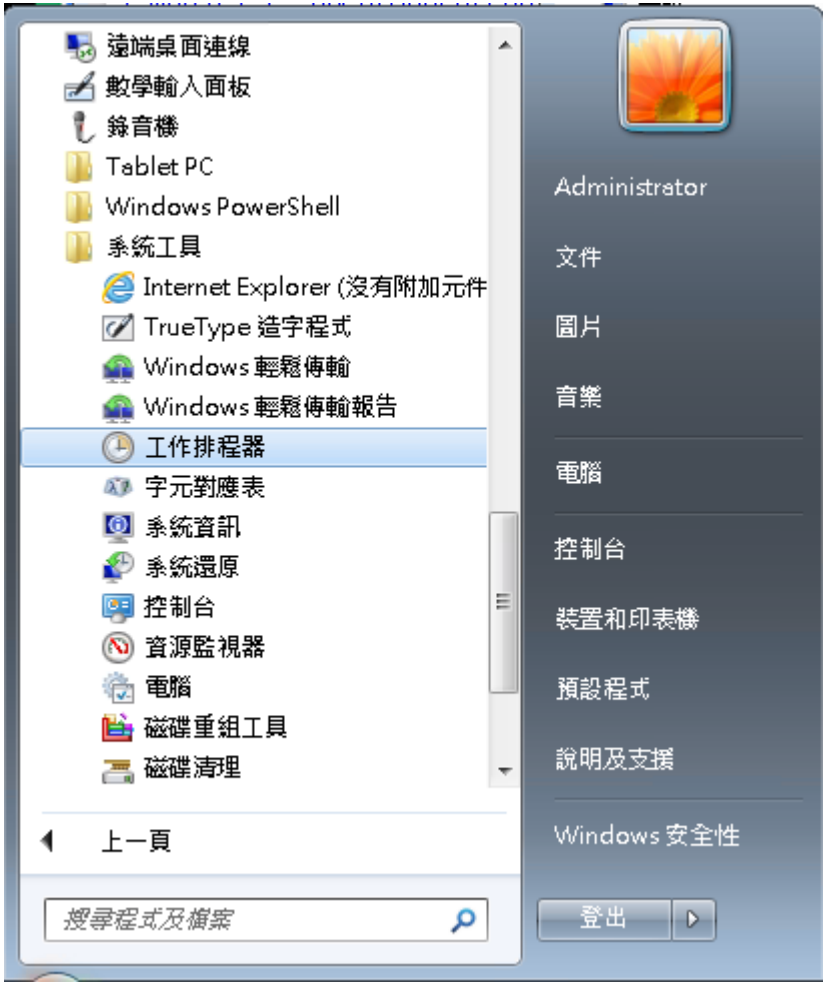
歷史資料匯入與排程設計

Windows 排程工具

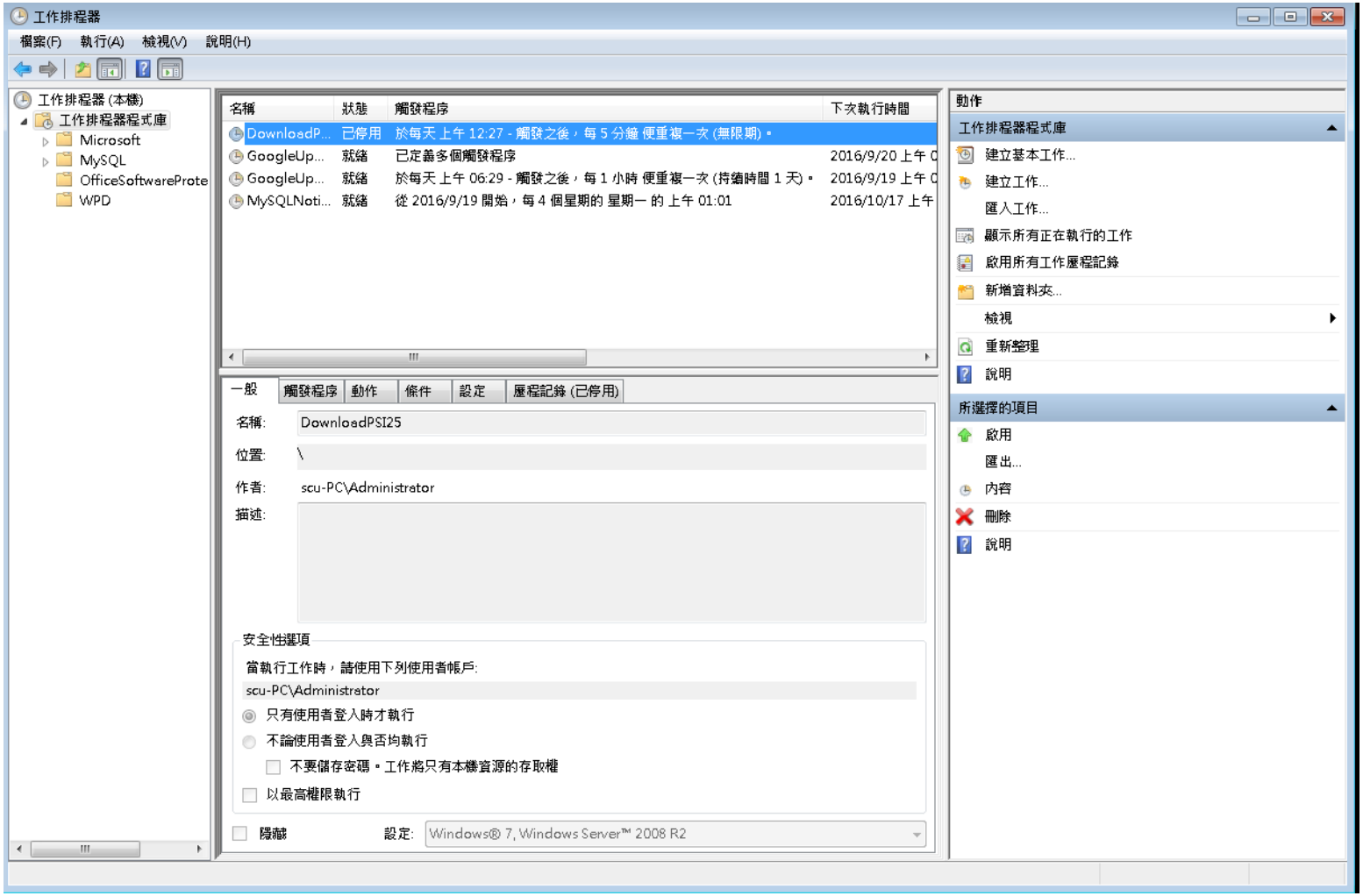
工作排程器

位置:
附屬應用程式 -> 系統工具

指令:
taskschd.msc

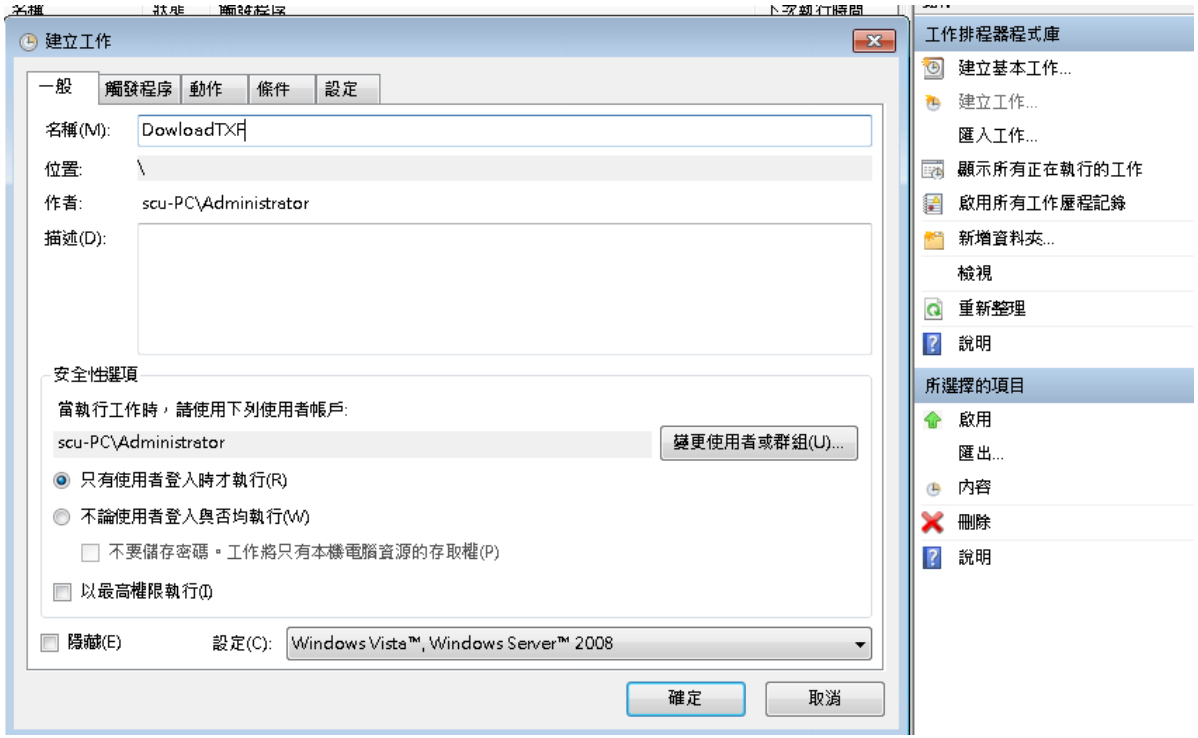


Windows 排程設定畫面



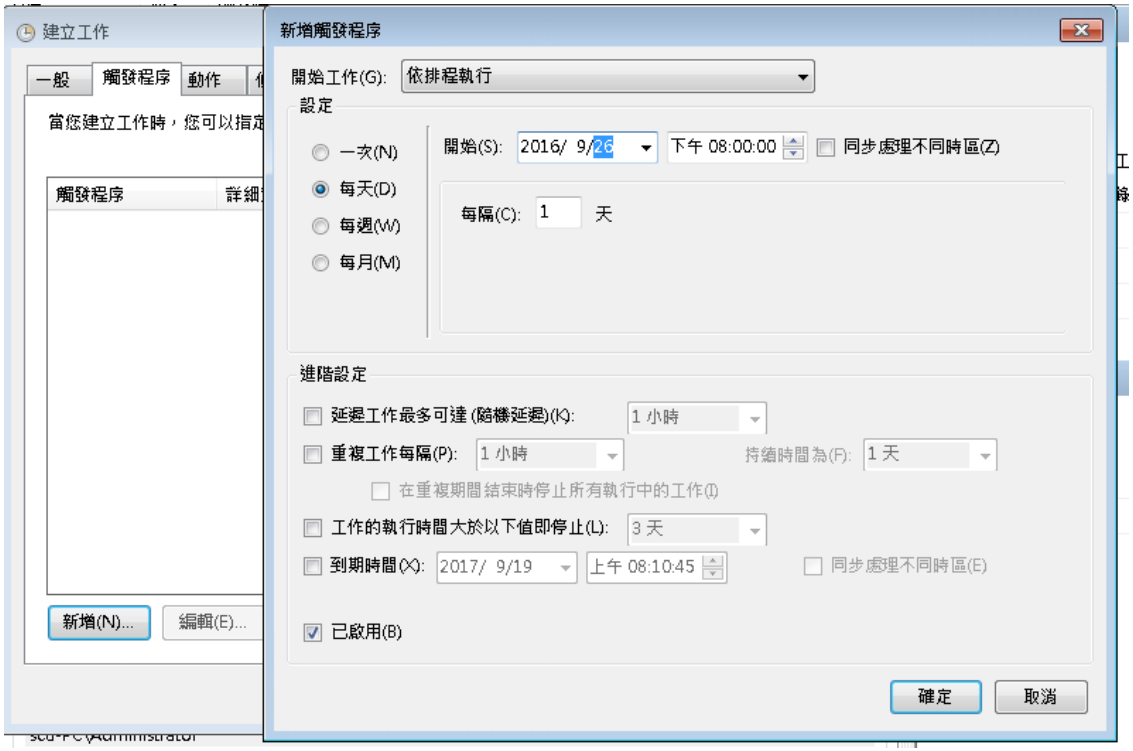
Windows 排程設定畫面

設定名稱:
DownloadTXF



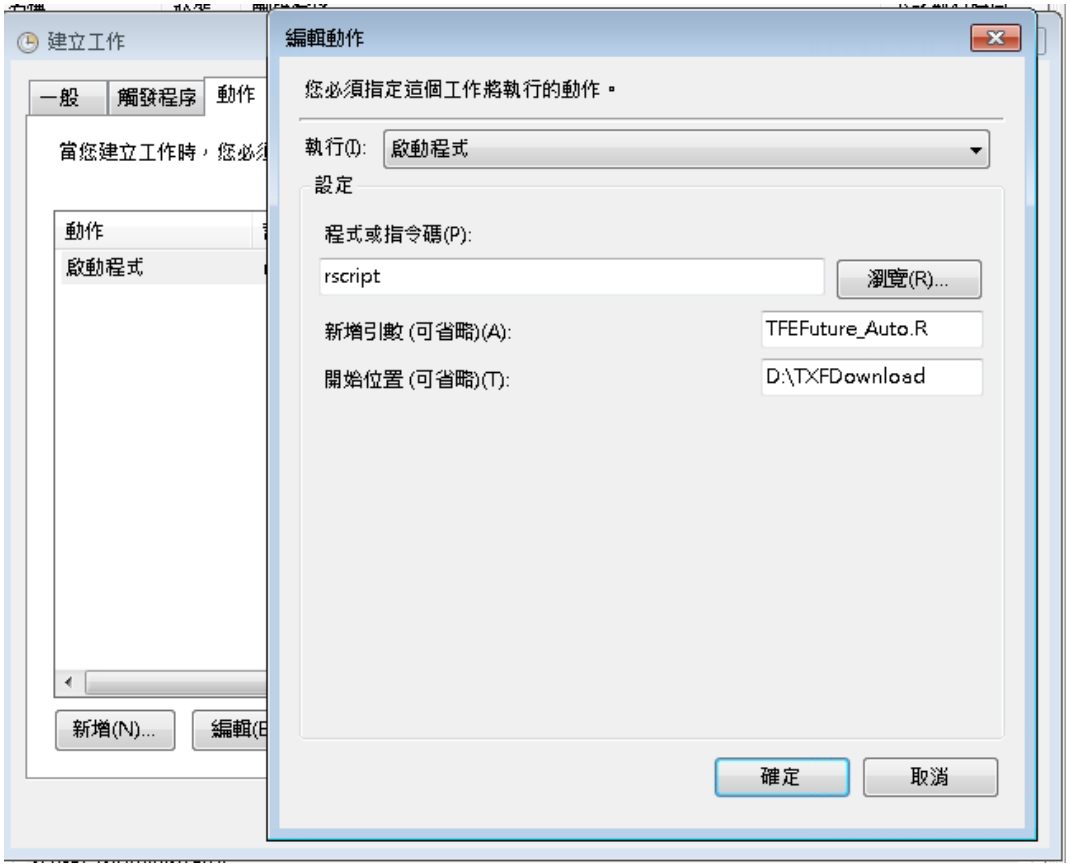
Windows 排程設定畫面

觸發程序:
從2016/9/26 下午8點啟動
每天執行



Windows 排程設定畫面

啟動程式:
程式或指令碼: rscript
引數: TFEFuture_Auto.R
開始位置: D:\TXFDownload

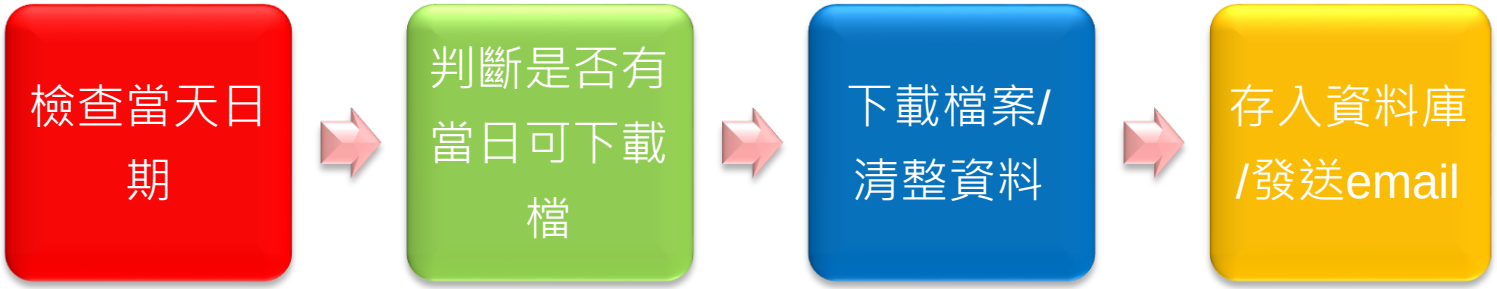


Windows 排程設定畫面

設定完成!

R 下載排程程式

- 程式流程



判斷當日之結算月

- 程式說明
- 安裝 timeDate 套件　　install.packages("timeDate")

```
library(timeDate)


#-------------------------------------------------------------------------------------------
#  判斷最近結算月
#

curDateStr <- unlist(strsplit(as.character(Sys.Date()), '-'))

# 測試用only
symbolid <- 'TXF'; qryyear <- '2016'; qrymonth <- '09';qryday <- '09'
#symbolid <- 'TXF'; qryyear <- curDateStr[1]; qrymonth <- curDateStr[2];qryday <- curDateSt


# 判斷當日屬於哪一近月結算月
settleMonth <- timeNthNdayInMonth(paste(curDateStr[1], curDateStr[2],'01', sep ='-'), 3, 3)
if(Sys.Date() > as.Date(settleMonth["Data"])){
  print("leap")
  curDateStr <- unlist(strsplit(as.character(Sys.Date()+30), '-'))

  curStlMth <- paste(curDateStr[1], curDateStr[2], sep ='')
}else{
  print("not leap")
  curStlMth <- paste(qryyear, qrymonth, sep ='')
}
```

判斷當日之結算月

- 程式說明
- 安裝 rvest 套件　　install.packages("rvest")

```
library(rvest)

page <- read_html("http://www.taifex.com.tw/chinese/3/3_1_3.asp")

rst <- page %>%
  html_nodes(xpath='//*[@id="printhere"]/table[2]') %>%
  html_table(fill = TRUE)
rst <- rst[[1]]

if(nrow(rst[rst$日期 == paste(qryyear, qrymonth, qryday, sep ='/'),]) == 1){
```

下載檔案/清整資料

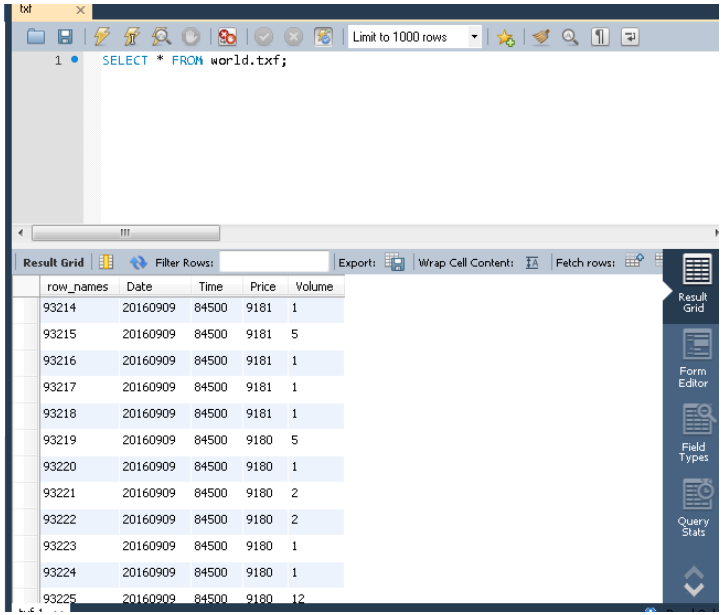
- 程式說明

```
# 暫存檔名
tmpf <- sprintf("Daily_%s_%s_%s.csv", qryyear, qrymonth, qryday)


# 解壓縮暫存檔
temp <- tempfile()

# 下載壓縮檔
download.file(tfeurl, temp)

tx <- read.csv(unz(temp, tmpf), header = FALSE, skip = 1, sep = ",")

unlink(temp)

# 清整資料
# Trim spaces
tx$V2 =trimws(tx$V2)
tx$V3 =trimws(tx$V3)

tx <- tx[(tx$V3 == curStlMth & tx$V2 == "TX"),]

tx <- subset(tx, select = c("V1", "V4", "V5", "V6"))

colnames(tx) <- c("日期", "時間", "成交價", "成交量")
tx$成交價 <- as.integer(tx$成交價)
tx$成交量 <- as.integer(tx$成交量)/2
```

存入資料庫

- 程式說明

```
# 連結MySQL 寫入資料庫
library(RMySQL)

con <- dbConnect(MySQL(), user="admin", password="admin",
                 dbname="world", host="localhost",client.flag=CLIENT_MULTI_STATEMENTS)

dbWriteTable(con,"txf", tx, overwrite=T)

dbDisconnect(con)
```



傳送email

- 安裝R mail 套件

```
install.packages("mailR")
```

- 先確認有沒有安裝Java JRE 套件
- https://java.com/zh_TW/download/

R 傳送mail 程式碼

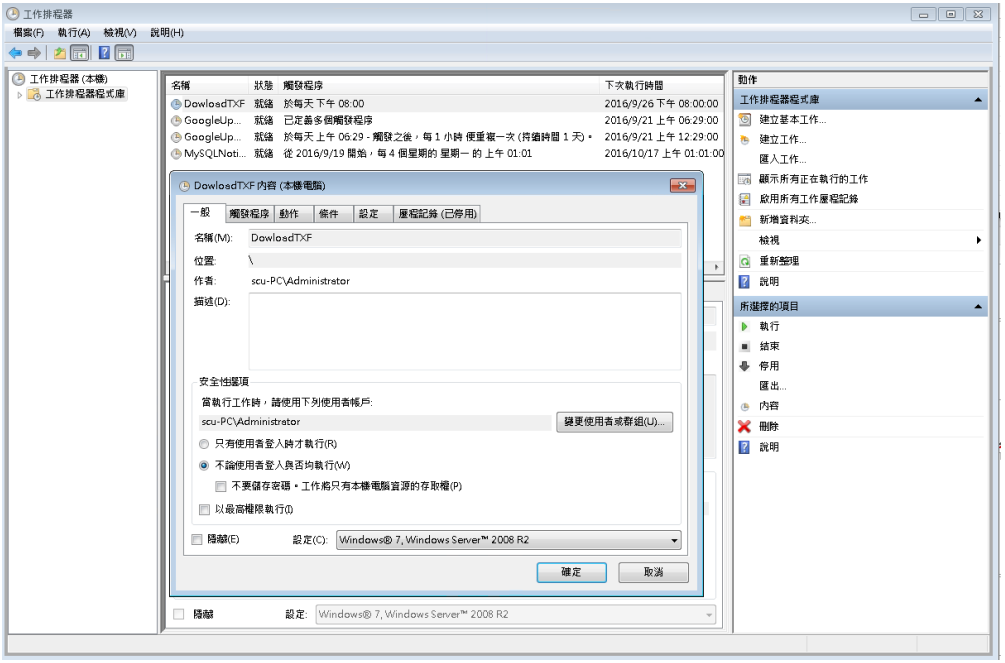
寄送email結果

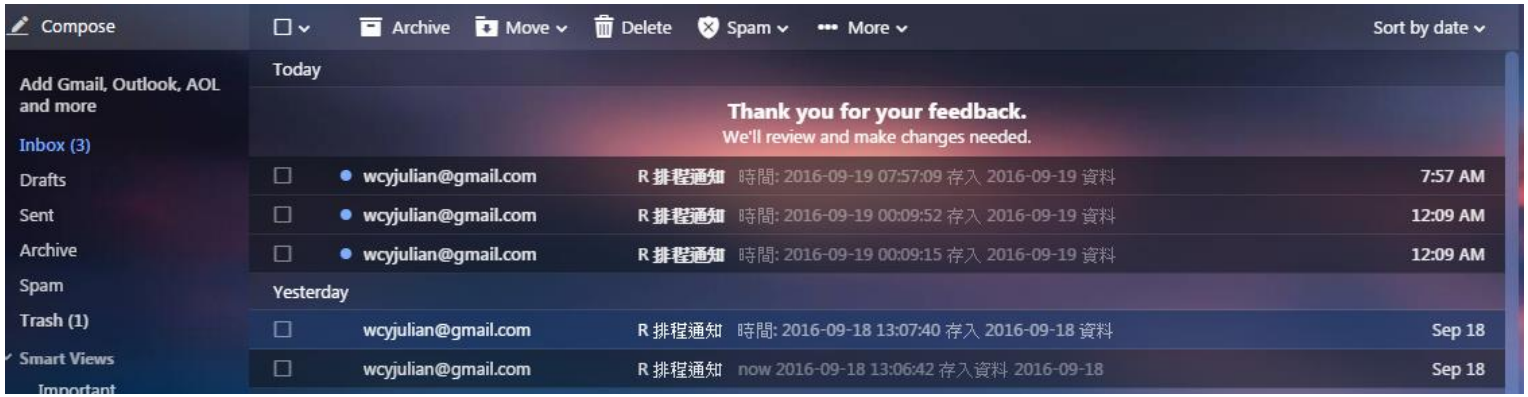
```
library(mailR)

sender <- xxx@gmail.com
recipients <- c("xxx@yahoo.com")

send.mail(from = sender,
          to = recipients,
          subject = "R 排程通知",
          body = sprintf("時間: %s 存入 %s 資料",
          Sys.time(),
          as.character(Sys.Date())
          smtp = list(host.name = "smtp.gmail.com",
                      port = 465,
                      user.name = "你的帳號",
                      passwd = "你的密碼", ssl = TRUE),
          authenticate = TRUE,
          encoding = 'utf-8',
          send = TRUE)
```

R 傳送mail 設定與結果

設定

結果





課後討論