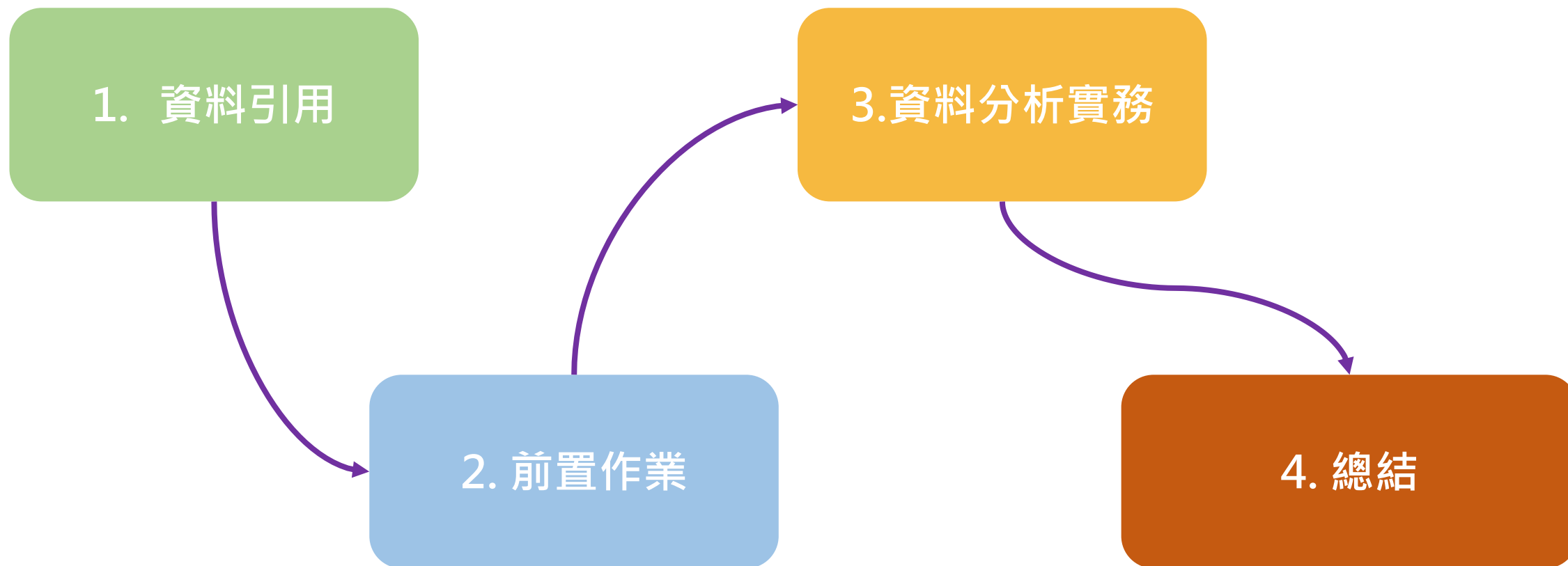


教育數據分析

SDPMLab, NKTU
軟體開發與流程管理實驗室
國立高雄師範大學

課程大綱



資料引用

SDPMLab, NKNU
軟體開發與流程管理實驗室
國立高雄師範大學

資料引用

檔案下載連結：<https://shorturl.at/gluzC>

本章節簡報於lesson 2資料夾中。

檔案說明：

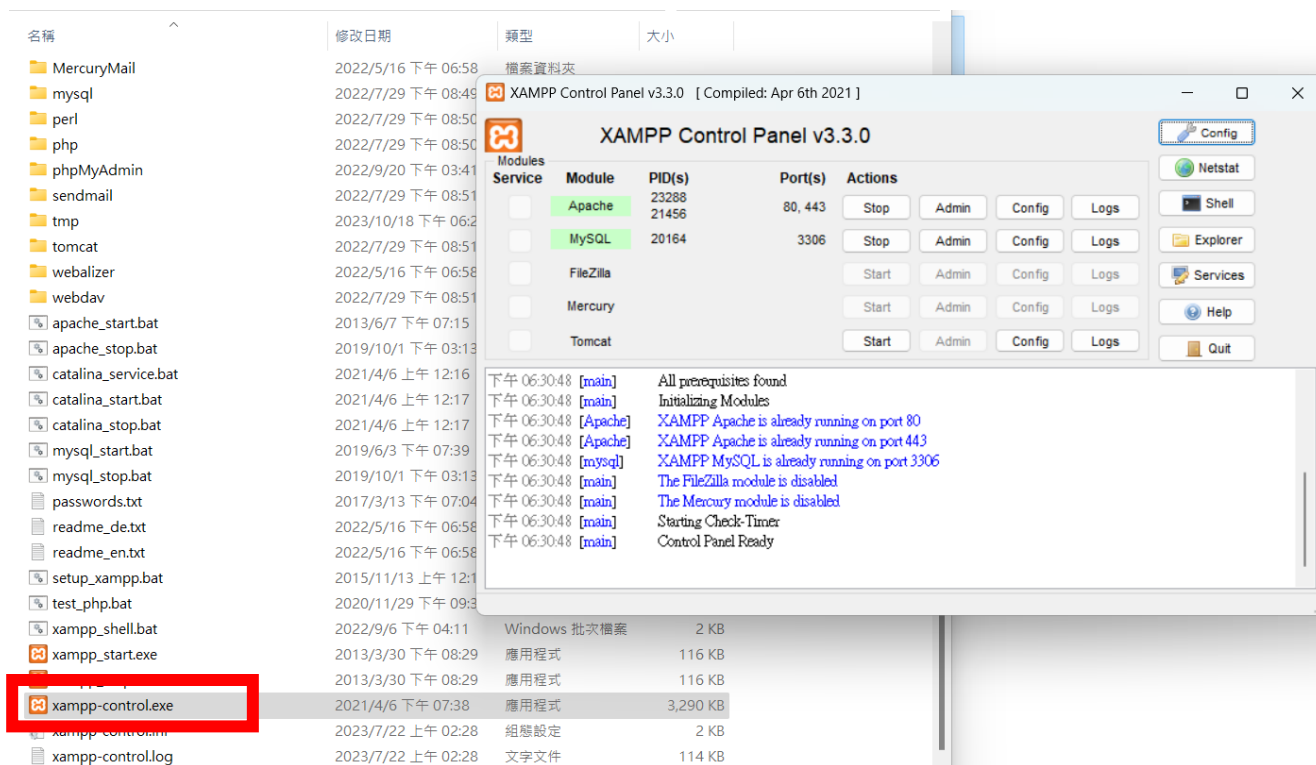
- 教學用開放資料說明.pdf
- edu_bigdata_imp1資料檔練習題.pdf
- edu_bigdata_imp1.sql
- Function資料夾

前置作業

SDPMLab, NKTU
軟體開發與流程管理實驗室
國立高雄師範大學

前置作業

- XAMPP的資料夾，進入目錄後找尋檔案名為 “**xampp-control.exe**” 的執行檔並執行。
- 執行後，將會出現如右圖之視窗。
- 將Apache 開啟，**不需要開啟MySQL**。



前置作業

- 於XAMPP安裝目錄找尋 “**htdocs**” 目錄並進入，該目錄主要讓我們放置欲執行的php檔案。
- 將先前下載的“ **code**” 資料夾放入此目錄。
- 使用任何習慣的編輯器，開啟該目錄。
- **請於function資料夾的同級目錄創建一支php檔案。**
- 請命名為**Exam.php**。

| > 本機 > 新增磁碟區 (E:) > xampp | | |
|---------------------------|---------------------|-------|
| 名稱 | 修改日期 | 類型 |
| folder apache | 2022/7/29 下午 08:49 | 檔案資料夾 |
| folder cgi-bin | 2022/7/29 下午 08:49 | 檔案資料夾 |
| folder contrib | 2022/7/29 下午 08:49 | 檔案資料夾 |
| folder FileZillaFTP | 2022/5/16 下午 06:58 | 檔案資料夾 |
| folder htdocs | 2023/10/18 下午 06:36 | 檔案資料夾 |
| folder install | 2022/7/29 下午 08:49 | 檔案資料夾 |
| | | |
| 名稱 | 修改日期 | |
| folder function | 2023/12/5 | |
| file Exam.php | 2023/12/5 | |

前置作業

- function資料夾中的DBConnectionHanlder.php已將資料庫連線邏輯進行抽離，並提供兩個方法給予各位使用，分別是：
- **setConnection**：將資料庫的host、userName、password、db名稱傳入後，即可建立資料庫連線。
- **getConnection**：取得資料庫連線。

```
3 references | 0 overrides
public static function setConnection(
    $serverName,
    $userName,
    $password,
    $db
)
{
    static::$serverName = $serverName;
    static::$userName = $userName;
    static::$password = $password;
    static::$db = $db;

    $connectionStr = sprintf(
        "mysql:host=%s;dbname=%s",
        static::$serverName,
        static::$db
    );

    try {
        static::$connection = new PDO($connectionStr, static::$userName, static::$password);
        // set the PDO error mode to exception
        static::$connection->setAttribute(PDO::ATTR_ERRMODE, PDO::ERRMODE_EXCEPTION);
    } catch(PDOException $e) {
        // echo "Connection failed: " . $e->getMessage();
        throw $e;
    }
}

3 references | 0 overrides
public static function getConnection()
{
    if (is_null(static::$connection)) {
        throw new InvalidArgumentException();
    }
    return static::$connection;
}
```


前置作業

- 紅框處為資料庫伺服器資訊。
- 請於Exam.php中寫入右圖中的程式碼，以確認目前是否能正確連接資料庫。
- 撰寫完畢後，請開啟瀏覽器並於搜尋框中搜尋 <http://localhost/code/Exam.php>。
- 如連接成功，你將於瀏覽器中看到“ success”。

```
1  <?php
2  require_once 'function/DBConnectionHandler.php';
3
4
5  $serverName = "140.127.74.201:9001";
6  $userName = "root";
7  $password = "root";
8  $db = 'bigdata';
9
10 try {
11     DBConnectionHandler::setConnection(
12         $serverName,
13         $userName,
14         $password,
15         $db
16     );
17     $connection = DBConnectionHandler::getConnection();
18     echo "success";
19 } catch (Exception $e) {
20     echo " ERROR ". $sql . "<br>" . $e->getMessage();
21 }
22 ?>
```

資料分析實務

SDPMLab, NKTU
軟體開發與流程管理實驗室
國立高雄師範大學

資料分析實務

各位可先閱讀下列兩個檔案，分別為**資料庫欄位說明**以及後續我們將**實作的練習題**。

- 教學用開放資料說明.pdf
- edu_bigdata_imp1資料檔練習題.pdf

資料分析實務 - 程式撰寫

我們將以 edu_bigdata_imp1資料檔練習題.pdf 中的題目進行解題，並使用剛剛創建的 Exam.php 撰寫解答程式碼。

簡報中將以第一大題為範例，帶各位進行初步分析。

| 題號 | 題目 | 答案 |
|-----------------------|---|----|
| 請依據 39 號學生的行為紀錄回答下列問題 | | |
| 1.1 | 於 dp001 平臺總共進行幾次 <u>不重複</u> 的影片瀏覽的學習紀錄? | |
| 1.2 | 於 dp001 平臺瀏覽影片時，總共進行幾次 <u>不重複</u> 的檢核點作答? | |

資料分析實務 – 程式撰寫

問題 1-1

將問題進行簡化即為，**39號學生看過幾部不同的學習影片**，並用影片編號和學生編號進行實作。

解題思路

1. 建立資料庫連線，並取得連線。
2. 撰寫SQL(搭配 “DISTINCT” 關鍵字)，針對39號學生搜尋出不重複資料。
3. 使用PDO預處理與資料綁定語法。
4. 即得所求。

資料分析實務 - 1-1解答

- 建立資料庫連線，並取得連線。
- 使用COUNT、DISTINCT等關鍵字針對“ dp001_review_sn” 欄位撰寫SQL。
- 使用PDO語法取得搜尋結果。
- 印出答案，使用瀏覽器搜尋“<http://localhost/code/Exam.php>”，即可看到答案。

```
1  <?php
2  require_once 'function/DBConnectionHandler.php';
3
4
5  $serverName = "140.127.74.201:9001";
6  $userName = "root";
7  $password = "root";
8  $db = 'bigdata';
9
10 DBConnectionHandler::setConnection(
11     $serverName,
12     $userName,
13     $password,
14     $db
15 );
16 $connection = DBConnectionHandler::getConnection();
17
18 // 1.1
19 $sql = "SELECT COUNT(DISTINCT dp001_review_sn) AS result FROM edu_bigdata_imp1 WHERE PseudoID=:ID";
20 $stmt = $connection->prepare($sql);
21 $stmt->bindValue(':ID',39);
22 $stmt->execute();
23 $r = $stmt->fetchAll(PDO::FETCH_ASSOC);
24 print_r($r);
```

資料分析實務 – 程式撰寫

問題 1-2

將問題進行簡化即為，**39號學生做過幾個不同的檢核作答**，並用檢核點試題編號和學生編號進行實作。

解題思路

1. 建立資料庫連線，並取得連線。
2. 撰寫SQL(搭配 “DISTINCT” 關鍵字)，針對39號學生搜尋出不重複資料。
3. 使用WHERE、AND關鍵字排除欄位中包含NA的無效資料
4. 即得所解。

資料分析實務 – 1-2解答

- 建立資料庫連線，並取得連線。
- 使用COUNT、DISTINCT等關鍵字針對“ dp001_question_sn” 欄位撰寫SQL。
- 並且透過“ **WHERE PseudoID=:ID AND dp001_question_sn != :VAL;** ”搜尋出39號學生的資料且“ dp001_question_sn” 欄位不為**NA**。
- 使用PDO語法取得搜尋結果。
- 印出答案，使用瀏覽器搜尋“ <http://localhost/code/Exam.php> ”，即可看到答案。

```
$sql = "SELECT COUNT(DISTINCT dp001_question_sn) AS result FROM edu_bigdata_imp1 WHERE PseudoID=:ID AND dp001_question_sn != :VAL;";
$stmt = $connection->prepare($sql);
$stmt->bindValue(':ID',39);
$stmt->bindValue(':VAL','NA');
$stmt->execute();
$r = $stmt->fetchAll(PDO::FETCH_ASSOC);
print_r($r);
```


資料分析實務 實作時間

SDPMLab, NKTU
軟體開發與流程管理實驗室
國立高雄師範大學

資料分析實務-實作時間

- 請實作第一大題並將解答紀錄下來。
- 可參考第一大題的實作過程，接續實作第二大題。



資料分析實務 – 程式撰寫

解題思路 2-1

- 建立資料庫連線，並取得連線。
- 使用GROUP BY關鍵字針對“ dp001_video_item_sn” 欄位進行分群，並利用“ DISTINCT” 排除重複資料，針對281號學生進行搜尋。
- 使用PDO語法取得搜尋結果。
- 使用瀏覽器搜尋“ <http://localhost/code/Exam.php> ”，即可看到答案。

請依據 281 號學生的行為紀錄回答下列問題

2.1 於 dp001 平臺總共瀏覽過哪些不重複的影片且對應的能力指標為何？

2.2 於 dp001 平臺共有幾次的練習題作答紀錄正確率是 100？

資料分析實務 – 程式撰寫

解題思路 2-2

- 建立資料庫連線，並取得連線。
- 使用COUNT關鍵字針對“ dp001_prac_score_rate” 欄位進行聚合，並利用“ WHERE” 針對281號學生與“ ” dp001_prac_score_rate” 進行搜尋。
- 使用PDO語法取得搜尋結果。
- 使用瀏覽器搜尋“ <http://localhost/code/Exam.php> ”，即可看到答案。

請依據 281 號學生的行為紀錄回答下列問題

2.1 於 dp001 平臺總共瀏覽過哪些不重複的影片且對應的能力指標為何?

2.2 於 dp001 平臺共有幾次的練習題作答紀錄正確率是 100?

資料分析實務-實作時間

- 請實作第三大題並將解答紀錄下來。



資料分析實務 – 程式撰寫

解題思路 3-1

- 建立資料庫連線，並取得連線。
- 使用COUNT關鍵字針對“ dp001_record_plus_view_action” 欄位進行聚合，並利用“ WHERE” 針對71號學生與“ dp001_record_plus_view_action” 進行搜尋。
- 使用PDO語法取得搜尋結果。
- 使用瀏覽器搜尋“ <http://localhost/code/Exam.php> ”，即可看到答案。

請依據 71 號學生的行為紀錄回答下列問題

3.1 於 dp001 平臺的瀏覽影片時，操作行為名稱為「暫停」總共有幾次？

3.2 分別於哪幾天進入 dp001 平臺？

資料分析實務 – 程式撰寫

解題思路 3-2

- 建立資料庫連線，並取得連線。
- 使用GROUP BY關鍵字針對" dp001_review_start_time" 與" dp001_review_end_time" 欄位進行分群，並利用" DISTINCT" 排除重複資料，針對71號學生進行搜尋。
- 使用PDO語法取得搜尋結果。
- 使用瀏覽器搜尋" <http://localhost/code/Exam.php> "，即可看到答案。

請依據 71 號學生的行為紀錄回答下列問題

3.1 於 dp001 平臺的瀏覽影片時，操作行為名稱為「暫停」總共有幾次？

3.2 分別於哪幾天進入 dp001 平臺？

資料分析實務 - 第3大題解答

```
// 3.1
$sql = "SELECT COUNT(dp001_record_plus_view_action) AS result FROM edu_bigdata_imp1 WHERE dp001_record_plus_view_action=:VAL AND PseudoID=:ID";
$stmt = $connection->prepare($sql);
$stmt->bindValue(':VAL','paused');
$stmt->bindValue(':ID',71);
$stmt->execute();
$r = $stmt->fetchAll(PDO::FETCH_ASSOC);
print_r($r);

// 3.2 |
$sql = "SELECT DISTINCT DATE(dp001_review_start_time),DATE(dp001_review_end_time) FROM edu_bigdata_imp1 WHERE PseudoID=:ID GROUP BY dp001_review_start_time,dp001_review_end_time";
$stmt = $connection->prepare($sql);
$stmt->bindValue(':ID',71);
$stmt->execute();
$r = $stmt->fetchAll(PDO::FETCH_ASSOC);
print_r($r);
```


資料分析實務-實作時間

- 請實作第四大題並將解答紀錄下來。



資料分析實務 – 程式撰寫

解題思路 4-1

1. 使用影片瀏覽序號欄位(“dp001_review_sn”)，找出瀏覽次數最高的影片，最終結果須包含影片瀏覽序號以及對應的總瀏覽次數。
2. 在這過程中，可能需要用到GROUP BY、ORDER BY、COUNT、LIMIT等關鍵字實作。
3. 使用瀏覽器搜尋” <http://localhost/code/Exam.php> ”，即可看到答案。

請依據全體學生的行為紀錄回答下列問題

- 4.1 請找出在 dp001 平臺中，最多影片瀏覽行為的影片序號
- 4.2 請找出在 dp002 平臺中，操作資源的知識架構分類中為「十二年國民基本教育類」總共有幾筆？
- 4.3 請找出在 dp002 平臺中，前 3 個最常發生的操作行為名稱
- 4.4 請找出在 dp002 平臺中，操作資源的知識架構分類中為「校園職業安全」總共有幾筆？

資料分析實務 – 程式撰寫

解題思路 4-2

1. 過濾出操作資源的知識架構分類("dp002_extensions_alignment")為十二年國民基本教育類的資料。
2. 使用瀏覽器搜尋" <http://localhost/code/Exam.php> "，即可看到答案。
3. 備註：請注意資料中操作資源的知識架構分類的值，欄位中的值有標點符號存在，過濾時需注意。

請依據全體學生的行為紀錄回答下列問題

4.1 請找出在 dp001 平臺中，最多影片瀏覽行為的影片序號

4.2 請找出在 dp002 平臺中，操作資源的知識架構分類中為「十二年國民基本教育類」總共有幾筆？

4.3 請找出在 dp002 平臺中，前 3 個最常發生的操作行為名稱

4.4 請找出在 dp002 平臺中，操作資源的知識架構分類中為「校園職業安全」總共有幾筆？

資料分析實務 – 程式撰寫

解題思路 4-3

1. 從操作行為名稱欄位("dp002_verb_display_zh_TW") 中排除NA的值。
2. 為欄位中的值進行計數，並透過GROUP BY 、ORDER BY 、COUNT、LIMIT等關鍵字實作。
3. 使用瀏覽器搜尋" <http://localhost/code/Exam.php> "，即可看到答案。

請依據全體學生的行為紀錄回答下列問題

4.1 請找出在 dp001 平臺中，最多影片瀏覽行為的影片序號

4.2 請找出在 dp002 平臺中，操作資源的知識架構分類中為「十二年國民基本教育類」總共有幾筆？

4.3 請找出在 dp002 平臺中，前 3 個最常發生的操作行為名稱

4.4 請找出在 dp002 平臺中，操作資源的知識架構分類中為「校園職業安全」總共有幾筆？

資料分析實務 – 程式撰寫

解題思路 4-4

1. 此題解題方法與4-2相同。
2. 過濾出操作資源的知識架構分類(“dp002_extensions_alignment”)為**校園職業安全**的資料。
3. 使用瀏覽器搜尋“ <http://localhost/code/Exam.php> ”，即可看到答案。
4. 備註：請注意資料中操作資源的知識架構分類的值，欄位中的值有標點符號存在，過濾時需注意。

請依據全體學生的行為紀錄回答下列問題

4.1 請找出在 dp001 平臺中，最多影片瀏覽行為的影片序號

4.2 請找出在 dp002 平臺中，操作資源的知識架構分類中為「十二年國民基本教育類」總共有幾筆？

4.3 請找出在 dp002 平臺中，前 3 個最常發生的操作行為名稱

4.4 請找出在 dp002 平臺中，操作資源的知識架構分類中為「校園職業安全」總共有幾筆？

總結

SDPMLab, NKNU
軟體開發與流程管理實驗室
國立高雄師範大學

總結

關於今日課程我們做到了...

- 架設本地Web Server，使PHP能夠良好的與資料庫進行交互。
- PHP連接資料庫語法。
- 了解常用的資料庫操作。
- 實際操作20萬+資料量的資料庫。
- 實作貼近現實需求的練習題。

或許你可以...

- 嘗試挑戰處理OPEN DATA的巨量資料集。
- 結合OPEN DATA資料並實作一套系統。
- 參加類似競賽。

教育數據分析

SDPMLab, NKTU
軟體開發與流程管理實驗室
國立高雄師範大學