



# 大數據資料處理 – Hadoop MapReduce Programming

## 『交通部高速公路闡道資料』 資訊擷取為例

授課教授：王經篤 (Jing-Doo Wang)

Email: [jdwang@asia.edu.tw](mailto:jdwang@asia.edu.tw)



亞洲大學 資訊工程系 (CSIE, Asia University)

# 目的 (Goal)

- 快速進入Hadoop開發練習 (壓縮檔案下載解開, 即可使用)
  - Windoop 程式開發環境(MS-Window)
    - Hadoop
    - Java
    - Eclipse
- 利用Hadoop MapReduce程式, 練習巨量資料整理。
  - 利用線上開放資料-『交通部高速公路闢道資料』
  - 完整的資料前處理(Data Preprocessing)與資訊擷取(Information Extraction)

# 課程大綱 (Outline)

- Hadoop 介紹 (Hadoop Introduction)
- Windoop 開發環境介紹
  - Windoop 安裝 (MS-Windows + Hadoop)
  - Hadoop File System (HDFS) 系統操作 (HDFS Operations)
- Windoop 練習
  - WordCount 程式載入修改與執行
- 網路資料自動下載 (Web Robot)
  - 交通部高速公路闖道資料 (<http://tisvcloud.freeway.gov.tw/>)
- 巨量資料處理案例 (Case Study)
  - 高速公路闖道資料之資訊擷取 (Information Extraction from Freeway Gantry Records)

# 作業與評分 (Assignment and Score)

1. (10%) Windoop 單機系統安裝
2. (10%) Hadoop Java 程式載入 修改與執行
3. (10%) 網路資料自動下載
4. (30%) 高速公路闢道資料之資訊擷取

# 參與課程教材準備人員

- 系統操作測試與畫面截圖
  - 呂宏仁
- 操作投影片製作與使用者模擬
  - 王家恩
- 系統操作測試與錄影
  - 王寬之
- 教學影片錄製與剪接
  - 朱凱聞