

銘傳大學

應用統計與資料科學學系碩士班
碩士論文

交通 ETC 大數據探討與計算
-以國道五號為例

指導教授：林真真、李御璽 教授
研究生：曾韋翔

中 華 民 國 106 年 06 月

交通 ETC 大數據探討與計算-以國道五號為例

摘要

大數據隨著雲端技術的進步，由單主機處理，成長成透過多台機器的串聯，提高運算能力及儲存空間，讓所有資料都能完整的存入大數據雲端分析平台，其中自由軟體 Hortonworks 結合 Hadoop，可以達到巨量資料管理與分析的功能。可以在此平台進行資訊的篩選，找出關聯性並做預測防範等，若運用在交通旅行上，可以提供資訊給用路人，避開壅塞路段、減少資源的浪費。

臺灣的高速公路 ETC 電子收費系統於 2014 年全面啟用，至今累積了龐大的數據。本研究將巨量的高速公路行駛數據匯入雲端分析平台 Hortonworks，並結合 Rhadoop 等雲端平台套件進行資料探勘，同時在 RStudio Server 與 Hortonworks 上的 Hive 資料庫進行資料篩選，針對不同的路況時段(平假日、時段等)比較不同車種的旅行時間，為用路人提供更有效益的行車建議。同時，本研究也將在 RStudio Server 上透過多核心套件 parallel 以及 foreach，比較不同大小的資料量在單核心與多核心計算之效能差異。

針對高速公路旅行時間的研究結果可以發現平日的上班車潮與星期六上午的出遊車潮是車流量最多的時段，其中以南港至頭城的路段最為嚴重；清晨以及晚間行駛國道五號時要多注意大型車輛，以策安全；另外也建議用路人上路前要多關注即時路況與廣播，以避開突發的壅塞狀況。在多核心計算的部分則可明顯發現使用多核心技術能有要節省運算所花費的時間，根據硬體環境的不同，使用的套件功能也會有不同的效能差異。

關鍵詞：大數據、Hortonworks、Rhadoop、資料探勘、多核心計算

Application of Big Data Analysis to ETC Data case of Freeway No.5

Abstract

The advancement of cloud computing has grown up to multi-machine, more access space and computing power let almost every information can save into the data platform, the open source Hadoop software Hortonworks Data Platform (HDP) can manage and compute big data efficiently, filter useful information, find the association to prevent risk, while using on traffic data can provide driving advice to driver, avoid traffic jam, and reduce the waste of resources.

Electronic Toll Collection (ETC) had enabled since 2014 on three main freeway in Taiwan, lots of traffic data had been collected. This research will upload two years traffic data into Hortonworks data platform (HDP), use Data Mining by RHadoop packages in RStudio server, filter ETC data with RStudio Server and Hive, then connect Hive database with RStudio Server, compare different sections' status, car type, and time period, provide driver an efficiently driving advice. Also, this study use parallel and foreach package of RStudio Server, compare the operational efficiency by different size of data, different quantity of cores, and different methods.

For the research of travel time, we can find out that weekday morning's working crowds and weekend morning's travel crowds are most congestion time, where Nankang to Toucheng interchange is the most serious section. Traveler at night and early morning need to alert the trucks and bus, to ensure the safety of themselves. Also, we suggest every traveler get attention to the traffic information on the broadcast, to avoid emergencies. In the case of multi-cores computing, we can find that multi-cores' efficacy is better than single-core's environment, different package function on different quantity of cores also can reduce different time spend.

Keywords: Big data, Hortonworks, RHadoop, Data Mining, parallel computing

目錄

第一章、緒論	1
第一節、研究背景與動機.....	1
第二節、研究目的	1
第三節、研究範圍	2
第四節、研究流程	3
第二章、文獻探討	4
第一節、交通資料分析.....	4
第二節、大數據平台與 R.....	6
第三節、資料探勘(Data Mining).....	10
第三章、研究方法	13
第一節、需求軟體	13
第二節、研究環境架構概念圖.....	15
第三節、Hortonworks 架構與範例資料模擬.....	16
第四節、Rstudio Server 安裝.....	35
第五節、連接埠之彙整.....	41
第六節、RStudio Server 之多核心計算比較	42
第七節、K-Means(K-平均演算法).....	44
第八節、資料描述	45
第四章、實證分析	48
第一節、基本敘述統計.....	48
第二節、集群分析	58
第三節、RStudio Server 之多核心計算	90
第五章、結論與建議	96
第一節、系統架構與效能.....	96
第二節、資料分析	96
第三節、研究限制	97
第四節、未來研究之建議.....	97
附錄(一)-Anaconda 安裝與 Python 程式碼解說	98
附錄(二)-多核心計算時間比較之程式碼解說	100
附錄(三)-RStudio Server 集群分析程式碼	101
參考文獻	104

圖目錄

圖 1.1 研究流程架構圖.....	3
圖 3.1 研究環境結構圖.....	15
圖 3.2 下載 Hortonworks Sandbox 2.5(VirtualBox 版本).....	16
圖 3.3 匯入 Hortonworks Sandbox 2.5(一).....	17
圖 3.4 匯入 Hortonworks Sandbox 2.5(二).....	17
圖 3.5 匯入 Hortonworks Sandbox 2.5(三).....	18
圖 3.6 新增虛擬網路卡.....	18
圖 3.7 指定新增的虛擬網路卡給 Hortonworks	19
圖 3.8 於 Oracle VM VirtualBox 啟動 Hortonworks.....	19
圖 3.9 成功啟動後之畫面.....	19
圖 3.10 Hortonworks 歡迎介面	20
圖 3.11 Hortonworks 登入首頁.....	20
圖 3.12 Putty 上登入管理者帳號密碼	21
圖 3.13 Putty 上登入管理者帳號密碼	21
圖 3.14 範例資料下載位置.....	22
圖 3.15 從使用者介面進入 HDFS	22
圖 3.16 HDFS 介面	23
圖 3.17 HDFS 中建立新資料夾	23
圖 3.18 上傳檔案至 HDFS(一).....	23
圖 3.19 上傳檔案至 HDFS(二).....	24
圖 3.20 設定資料權限.....	24
圖 3.21 於首頁進入 Hive 介面.....	25
圖 3.22 Hive 介面介紹	25
圖 3.23 設定資料讀取格視為 TEZ 壓縮檔	26
圖 3.24 HIVE 中 Upload Table 頁面(本機端匯入/HDFS 匯入)	26
圖 3.25 儲存資料表之相關介面與設定.....	27
圖 3.26 HIVE 瀏覽資料.....	27
圖 3.27 計算變數另存之 truck_mileage 資料表.....	28
圖 3.28 DriverMileage 資料表輸出結果	29
圖 3.29 建立 PIG 程式碼	30
圖 3.30 PIG 基本介面介紹	30
圖 3.31 PIG 執行完成畫面	31
圖 3.32 riskfactor 資料表結果瀏覽	32
圖 3.33 Zeppelin 主頁	33
圖 3.34 於 Zeppelin 執行 Hive 程式呼叫資料表.....	34
圖 3.35 Zeppelin 中選擇變數與繪圖	34

圖 3.36 設定 rstudioserver 專用連接埠.....	35
圖 3.37 Rstudio Server 啟動畫面.....	36
圖 3.38 Rstudio Server 登入畫面.....	36
圖 3.39 命令提示字元合併 CSV 檔.....	45
圖 3.40 匯入 ETC 資料至 HDFS.....	45
圖 4.1 12 月車流量時間數列圖(小客車).....	49
圖 4.2 各路徑集群分析時速圖.....	89
圖 4.3 mclapply 運算時間比較折線圖.....	94
圖 4.4 parLapply 運算時間比較折線圖.....	94
圖 4.5 Foreach 運算時間比較折線圖.....	95



表目錄

表 3.1 連接埠功能介紹.....	41
表 3.2 資料變數說明.....	46
表 3.3 ETC 測站說明與代號.....	46
表 3.4 整理過後之變數說明.....	47
表 3.5 選用之旅行資料整理.....	47
表 4.1 各車種與月份之資料筆數(欄比例；列比例).....	48
表 4.2 各車種於平假日之平均車流量(欄比例；列比例).....	49
表 4.3 各星期與時段之平均旅行時間及車速(南港-石碇-坪林).....	50
表 4.4 各星期與時段之平均旅行時間及車速(石碇-坪林-頭城).....	51
表 4.5 各星期與時段之平均旅行時間及車速(坪林-頭城-宜蘭).....	52
表 4.6 各星期與時段之平均旅行時間(頭城-宜蘭-羅東).....	53
表 4.7 各星期與時段之平均旅行時間(宜蘭-羅東-蘇澳).....	54
表 4.8 平假日與車種間的平均旅行時間(南港-石碇-坪林).....	55
表 4.9 平假日與車種間的平均旅行時間(石碇-坪林-頭城).....	55
表 4.10 平假日與車種間的平均旅行時間(坪林-頭城-宜蘭).....	56
表 4.11 平假日與車種間的平均旅行時間(頭城-宜蘭-羅東).....	56
表 4.12 平假日與車種間的平均旅行時間(宜蘭-羅東-蘇澳).....	57
表 4.13 ETC_KA1 分群結果與平均旅行時間(時速).....	58
表 4.14 ETC_KA1 各集群於不同星期之資料比例.....	58
表 4.15 ETC_KA1 各集群於不同時段之資料比例.....	59
表 4.16(集群 1)假日白天出遊之車潮.....	59
表 4.17(集群 2)下班後前往宜蘭地區之車潮.....	60
表 4.18(集群 3)上午時段上班之車潮.....	61
表 4.19 ETC_KA1 各級群命名與用路人建議.....	62
表 4.20 ETC_KA2 分群結果與平均旅行時間(時速).....	63
表 4.21 ETC_KA2 各集群於不同星期之資料比例.....	63
表 4.22 ETC_KA2 各集群於不同時段之資料比例.....	64
表 4.23(集群 1)白天之車潮.....	64
表 4.24(集群 2)週六白天出遊之車潮.....	65
表 4.25(集群 3)清晨以外之車潮.....	66
表 4.26 ETC_KA2 各級群命名與用路人建議.....	67
表 4.27 ETC_KA3 分群結果與平均旅行時間(時速).....	68
表 4.28 ETC_KA3 各集群於不同星期之資料比例.....	69
表 4.29 ETC_KA2 各集群於不同時段之資料比例.....	69
表 4.30(集群 1)假日白天之車潮.....	70
表 4.31(集群 2)清晨以外之車潮.....	71

表 4. 32(集群 3)全天之車潮	72
表 4. 33 ETC_KA3 各級群命名與用路人建議	73
表 4. 34 ETC_KB1 分群結果與平均旅行時間(時速).....	74
表 4. 35 ETC_KB1 各集群於不同星期之資料比例	74
表 4. 36 ETC_KB2 各集群於不同時段之資料比例	75
表 4. 37(集群 1)凌晨以外之車潮	75
表 4. 38(集群 2)白天之車潮	76
表 4. 39(集群 3)週六白天之車潮	77
表 4. 40 ETC_KB1 各級群命名與用路人建議	78
表 4. 41 ETC_KB2 分群結果與平均旅行時間(時速).....	79
表 4. 42 ETC_KB2 各集群於不同星期之資料比例	79
表 4. 43 ETC_KB2 各集群於不同時段之資料比例	80
表 4. 44(集群 1)週六白天之車潮	80
表 4. 45(集群 2)白天之車潮	81
表 4. 46(集群 3)一般車潮	82
表 4. 47 ETC_KB2 各級群命名與用路人建議	83
表 4. 48 ETC_KC1 分群結果與平均旅行時間(時速).....	84
表 4. 49 ETC_KC1 各集群於不同星期之資料比例	84
表 4. 50 ETC_KC1 各集群於不同時段之資料比例	85
表 4. 51(集群 1)假日凌晨時段之車潮	85
表 4. 52(集群 2)假日白天之車潮	86
表 4. 53(集群 3)少數壅塞車潮	87
表 4. 54 ETC_KB2 各級群命名與用路人建議	88
表 4. 55 集群分析彙整表.....	89
表 4. 56 lapply 方法與 parallel 套件之程式碼.....	90
表 4. 57 lapply 方法之多核心計算效能比較表(單位：秒(省時比例))	91
表 4. 58 lapply 方法與 parallel 套件之程式碼.....	92
表 4. 59 for 迴圈方法之多核心計算效能比較表(單位：秒(省時比例)).....	93

第一章、緒論

第一節、研究背景與動機

在大數據的浪潮下，收集、整理資料，分析、預測結果已然成為社會重視的一個技術，透過過去所累積的各式資料，去蕪存菁後轉換成對研究目標有用的資訊，再透過獲得的資訊預測未來、預防風險。這是大數據帶給我們的便利，但要能夠實現這些目標，需要有足夠強大的硬體支援，面對數據量的迅速膨脹，單一機器的分析能力與效能逐漸無法負荷，雲端分析應運而生。Apache 所開發的開源雲端計算平台 Hadoop 可將多台機器串接運行、分散式儲存資料，這些單主機無法做到的功能，讓使用者可以從大量結構與非結構化的資訊中撈出所需要的。

國道高速公路是台灣民眾進行中長途旅行必選的道路之一，每逢上下班時間及例假日就會出現龐大的車潮，過去在收費站拆除之前，每個收費點與匝道皆是塞車的主要發生路段，用路人易發生危險事故，且花費大量旅行時間在高速公路上，過去的改善方式是擴展國道車道數量，或是另闢新的替代道路，但已近乎飽和狀態，要新建或者擴建道路的可行性越來越低，在這樣的情況下，透過現有的巨量資料找出減少壅塞的方法，是努力的重點目標之一。

第二節、研究目的

高速公路電子收費系統所收集的資料不斷更新，自 2014 年初全面開通起，已收集超過兩年半的國道旅行資料，資料量有一定的水準，而本研究決定使用 Hortonworks 公司所開發的雲端平台 Hortonworks Sandbox 作為資料存放、整理的平台，並結合 R 語言的統計分析能力，運用集群分析與資料探勘演算法探討高速公路旅行時間的狀況，找出影響大小客車旅行時間的影響因子，進一步的，為了區別[單核心指令的執行]與[多核心計算的效能]差異，本研究會將資料切割成不同大小，分別以單核心與多核心進行運算，比較其花費時間的差異，提供給使用者，可依照所使用的資料量，分配多少硬體規格來進行資料分析。

第三節、 研究範圍

本研究的研究目的是為透過雲端計算平台對交通資料進行分析，透過集群分析與資料探勘演算法分析旅行時間，並比較多核心計算與一般程式的運算時間差異。

過去計算旅行時間所使用之資料多為車輛偵測器(Vehicle Detector, VD)、自動車輛辨識(Automatic Vehicle Identification, AVI)、移動式探針車(Probe Vehicle)，也有使用到高速公路電子收費系統(Electronic Toll Collection, ETC)，但過去所使用的是第二代的紅外線感應式系統，當時的普及率不及合約訂定之 65%，因此於 2014 年起全面改用第三代無線射頻辨識系統(Radio Frequency Identification, RFID)，目前欲上高速公路的車輛都已強制要貼 E-Tag，因此回收率近乎 100%，相較於過去的資料回收有著明顯的進步，因此本研究將使用第三代 ETC 所收集之資料進行旅行時間的預測。所使用之 ETC 資料來自交通部臺灣區國道高速公路局「交通資料庫」的各旅次路徑原始資料(M06A)，選用 2015 年 2 月至 2016 年 12 月之資料進行旅行時間的分群。(註：2015 年一月之資料有誤，因此不採用)

雲端分析平台的部分將在 Hortonworks 上架設 R 語言，資料透過 Hortonworks 進行篩選與整理之後，在 R 的介面上進行集群分析 K-Means 等資料探勘方法，為了比較運行時間的差異，將在雲端平台上架構之 RStudio-Server 選用不同函數、不同資料量及不同核心數量進行運算時間的比較。

第四節、 研究流程

本研究的研究流程如下，確立研究主題後，進行相關的文獻整理與回顧，接著在電腦上架構雲端分析平台結合統計分析軟體，將資料匯入進行整理後，透過統計分析軟體進行分析，並且在最後比較運算時間的差異，提出結論與具體建議。

圖 1.1 研究流程架構圖



第二章、文獻探討

本研究針對我國高速公路 ETC 資料進行旅行時間預測，並透過雲端分析平台存取資料後，在統計分析軟體 RStudio 中進行資料探勘，文獻探討的部分將分成交通旅行資料、大數據平台與 R 語言、資料探勘三個重點。

第一節、交通資料分析

2.1.1. 高速公路電子收費系統(Electronic Toll Collection, ETC)

中華民國國道的電子收費系統是世界上最大的公路電子化收費系統，亦為我國國道唯一的公路收費系統，其演變可分成三個階段：

第一階段，交通部臺灣區國道高速公路局(簡稱高公局)委託中華電信使用奧地利公司 EFKON 的紅外線系統在國道三號樹林、龍潭兩收費站試辦電子收費，為計次電子收費服務之系統。

第二階段沿用第一階段之系統，以民間興建營運後轉移模式(Build-operate-transfer, BOT)委託遠通電收於 2006 年 2 月 10 日正式於全線收費站啟用，將收費門架合併於各個收費站，使用者須於車上安裝車上單元裝置(OBU)並插入儲值金額之 e 通卡方可通行，但由於安裝率僅 40%，不及合約所簽訂之 65% 因而被第三代系統”RFID 無線射頻辨識系統”取代。

第三階段所採用的 RFID 系統是於 2011 年 9 月開始推廣，有別於前一代系統，使用者僅需於車體上黏貼電子標籤(eTag)即可感應收費，且收費方式由先前的計次收費轉變為計程收費，會依照所駕駛之車種、實際里程計費，儲值金若低於新台幣 120 元，也會以手機簡訊通知駕駛人，且於主線道上安裝之感應閘門也較國際慣用於交流道出入口的設置或我國舊式收費站來的簡便，於 2014 年 1 月 2 號建置完成且已正式啟用，裝設路段包含國道一號(含高架)、國道三號(含國三甲)及國道五號。

2.1.2. 高速公路旅行時間預測之相關研究

林士傑、李穎與魏建宏(民 92)使用類神經網路針對國道一號新竹往台北的客運進行旅行時間預測，蒐集之資料包含了 GPS、VD、事件、流量等資料，選用六種預測模式，並採用未來時間模式模擬出各路段旅行時間，利用 error 、 $|\text{error}|$ 、MPE 與 MAPE 作為驗證指標來判定哪一種模式的誤差最小，結果發現使用總體資料融合空間考量模式所產生的誤差較其餘五種預測模式低，也發現群組路段的長度會影響旅行時間驗證的績效。

汪進財與邱孟佑(民 99)以車流狀態為基礎預測旅行時間，收集楊梅到泰山之間共八個偵測器的 ETC 資料，使用 2006 年 8 月份的資料做模式建構，並採用 2006 年 10 月 15 日至 2006 年 10 月 21 日之資料做模式的驗證。先使用集群分析對資料做分類，再以迴歸分析建構不同交通狀況的旅行時間模式，最後以 ETC 行駛資料計算旅行時間。研究中也發現偵測器設置之位置對預測的精準度有很大的影響，如設置在匝道出入口或上下坡的地方，會使資料的變異變大。

邱逸彥(民 103)利用貝氏理論來推估旅行時間，收集高速公路的車輛偵測器資料作為分析目標，過濾掉正負三倍標準差之外的異常資料後，利用貝氏更新法(SD、VAR)與歷史資料的原始預測方法對高速公路行車進行時間預測，最後用平均絕對誤差百分比(MAPE)、均方根誤差(RMSE)、與平均絕對誤差(MAE)作為評估指標，判斷是否需要更新歷史速度資料，結果顯示歷史資料庫的預測模式擁有較佳的預測能力，且兩種貝氏更新法的預測結果並無明顯差異，且新舊方法之間並沒有存在明顯落差，表示對於我國高速公路的旅行狀況，兩個貝氏更新法的預測能力仍為優秀的。

Brain L. Smith(1994) 與 Michael J. Demetsky(1994)利用類神經網路對短程交通資料進行時間預測，Brain L. Smith 等人將預測旅行時間分為三種：第一種是以歷史資料為基礎做預測；第二種是透過時間數列的方式建模；第三種則是使用模擬方法，也就是利用類神經網路進行建模，Brain L. Smith 等人提到類神經網

路是一個非常適合用在交通時間預測的方法，他能夠為很多未定義、複雜或是非線性的模型建模。比較倒傳遞類神經網路與前兩個方法後可以發現到傳遞類神經網路方法的預測結果更為優秀。

針對我國高速公路的相關研究目前多為第三代 ETC 收費系統實裝前之研究分析，會受到資料收集不完整之影響，且在新一代的 ETC 上路之後，收費站也就全面廢除了，對於行駛時間之影響因子也會有明顯的轉變，因此本研究利用第三代 ETC 收費系統全面實裝後所蒐集到之資料進行旅行時間預測，期望能做出夠準確的預測結果，夠符合用路人與管理者之期望。

第二節、大數據平台與 R

2.2.1. Apache Hadoop 架構

Apache 所開發的 Hadoop 是一個可在普通機器上執行資料分散儲存與大數據處理的架構，可讓使用者從大量的結構與非結構資料中抓取所需的重要資訊，其重點架構可分為以下幾個部分：

1. Hadoop Distributed File System(HDFS)：

基於 Java 的分散式檔案管理系統，當資料集合的大小超過單台電腦的儲存能力時，就需要將其分割儲存到不同電腦中，分散儲存就是必須要引入網路串接的概念，更有可能在某一個節點產生資料遺失的狀況，而 HDFS 就是一個可以完美解決以上問題的重要套件。

HDFS 具有以下優點：

- i. 可儲存的檔案大小可達 GB、TB 甚至 PB。
- ii. 不需要昂貴的伺服器，普通的 PC Server 即可安裝使用，且當資料節點出現問題時，HDFS 也可以繼續運行。
- iii. 採用一次寫入，多次讀取的串流式資料存取，每次分析時都會讀取大部分的資料。

除了上述優點，也包含了以下的缺點：

- iv. 由於 HDFS 針對大量資料的吞吐做了優化，因此犧牲了即時讀取資料的速度，如追求速度可考慮使用 HBase。
- v. 會產生大量小檔案，對於硬體容量較小的電腦容易達到存取上限。
- vi. 存在 HDFS 中的檔案只能有一個使用者寫入，不支援多使用者寫入、修改檔案。

2. Hadoop MapReduce：

MapReduce 是由 Google 所提出的一個軟體框架，目的是針對電腦上的大型資料及進行分散式運算。藉由使用者所輸入的鍵值(Key value)給予 map 指令，MapReduce 便會透過 reduce 找尋相同鍵值，並抓出相關資料給使用者。

3. Hadoop Yarn：

Hadoop 的核心之一，用於管理及調度平台的資源，擴展了 MapReduce 的能力，結合了其他套件以支援非 MapReduce 之工作負載。

4. Hadoop common：包含了上述三個架構以外的工具及套件。

另外還有五大核心套件讓 Hadoop 達到企業級的應用級別，分別為：

1. 數據管理(Data Management)

有效地利用 HDFS 管理資料，進行分散式儲存，並在 Hadoop Yarn 的架構下進行有效的資源管理等服務。

2. 數據獲取(Data Collection)

以多種方式與套件對資料做存取、計算、查詢等功能，如：Hive 提供了用類 SQL 語法做基礎的篩選；Pig 的可編程能力能進行進階的變數計算；MapReduce 協助處理海量結構與非結構化的資料；Spark 提供更加進階且操作簡便的分析計算方法……等功能。

3. 數據的掌握與集合 (Data Governance and Integration)

對資料進行載入與策略管理，包含了 Apache Falcon 數據管理框架、Apache Flume 與 Sqoop 讓使用者進行簡單的資料抓取。

4. 安全

提供帳戶管理、安全認證與使用授權等保護， Apache Knox 為每一層套件皆提供了安全措施；Apache Ranger 則提供了更加易於操作的管理方式。

5. 操作

提供管理、規劃與監控整個 Hadoop 集群套件的功能，如 Apache Ambari 可監控並管理各個套件的運行狀態、Apache Oozie 則用來對應 JAVA Web、Apache ZooKeeper 則是用來協調分散式儲存所需進行處理的高效能系統。

2.2.2. Hortonworks Sandbox

Hortonwork 是由雅虎出資 2300 萬美金，於 2011 年六月成立的獨立公司，開發一個整合了 Apache Hadoop 與 BDAS 兩巨量資料分析元件的大數據平台(Big Data Platform)，他們將其產品 OS 包裝並簡化啟用之步驟，讓使用者可以透過 VMplayer 或是 VituralBox 等虛擬機器直接啟動已整合好的 Hortonworks Sandbox 平台，亦可多機串接，提升運算速度。

2.2.3. R 語言與 Rstudio

R 語言是一個免費的統計分析軟體，由紐西蘭奧克蘭特大學的 George Ross Ihaka 和 Robert Clifford Gentleman 所開發，主要功能包含統計分析、繪圖、資料探勘等，除了內建之基本功能外，亦可安裝額外套件，執行更多樣更複雜的指令與功能。亦有許多開發商在 R 語言的基礎下開發出圖型化介面，Rstudio 就是其中之一，Rstudio 為使用者提供一個更加人性化的介面，提升 R 語言的製圖能力與套件安裝、管理的功能，Rstudio 有開發伺服器版本，的使用介面，可以多人同時進行操作，亦可掛載 Hadoop 作業平台上，透過 RHadoop 的相關套件連結雲端平台，讓大數據平台更加強大，功能更完整。

2.2.4. 雲端分析平台之相關研究

Rakesh Kumar, Neha Gupya, Shilpi Charu, Somya Bansal 與 Kusum Yadav(2014) 針對 SQL 語法與 HiveQL 語法做比較，比較兩種語法上的優缺點與特性差異。SQL 語法是基於 SQL-92 標準建立的一個宣告式程式語言，他能夠讓使用者透過簡單的語法如 SELECT、INSERT、CREATE 等直觀的函數跟資料庫互動、管理；面對現今分散式雲端分析平台崛起，為了讓使用者在雲端平台上也能夠與資料庫進行互動而發明了 Hive query language(HiveQL)，這是一種類 SQL 語法，與傳統 SQL 的架構不完全一樣，功能上也較為陽春，無法支援到較進階的函數，但透過 HiveQL 就能夠對雲端平台的資料庫進行查詢、篩選等功能，且支援開放資料庫互連（Open Database Connectivity, ODBC）與 Java 資料庫連線(Java Database Connectivity, JDBC)等資料存取方式，完美結合了 MapReduce 的架構。

高諤亨(民 102)研究雲端心電圖資料的分析與顯示，透過 Hadoop 雲端分析系統將心電圖資料儲存至 HBase 中，HBase 是基於 HDFS 的一個分散式儲存資料庫，再利用 MapReduce 分析大量心電圖資料，顯示 Hadoop 具備對應大量資料的儲存與分析功能。

在 HDP 平台上可進行多種資料整理、分析及存取的動作，數據操作平台為 YARN，用於資料的維護、分析及管理，並透過 HIVE、PIG 等程式語言進行處理；可再利用 Zeppelin 將資料視覺化，亦可掛載 R 語言、R Studio 等統計軟體，進行進階的統計分析操作。本研究將採用 Hortonworks 作為巨量資料的儲存平台，並透過結合 R 語言提高統計分析能力。

第三節、 資料探勘(Data Mining)

2.3.1. 資料探勘的定義

從企業的角度，資料探勘就是希望從大量的企業日常資料交易資料中抽取潛在、不為人知的有用資訊、模式及趨勢走向，期望可以提高決策能力，預警異常狀態並預言未來趨勢。

而從資訊科技的角度來說，資料探勘就是建立一個電腦程式來自動篩選整個資料庫，尋找資料的樣式(Pattern)，分析樣式、檢驗、解釋並使用於幫助決策之結構，即可轉換為能用的知識。

而對統計學家而言，資料探勘其實已經不是一個新的概念了，畢竟從資料中抓取有用的資訊原本就是統計學家的職責所在，因此對統計學家而言，辨識出資料中的系統(System)與隨機(Random)是主要的工作。另一個重點為推論(Inference)，與統計觀點不同的是，資料探勘的推論是對未來之預測或危險之預防。若將模式建立、估計、檢驗及各種資料整合後並視為一遞迴程式，統計方法與資料探勘就非常類似了，實際上在資料科學家的資料探勘工具中，統計分析工具也扮演著重要的角色。

資料探勘的技術大致可區分為以下幾種

1. 特徵與鑑別(Characterization and Discrimination)

將原始概念層級的資料概括起來表示更高層級的資訊。

2. 關聯規則分析(Association Rule Analysis)

這是一個重要的資料探勘議題，將有興趣的所有項目的關聯找出來，並同時會引發下一層關聯而挖掘出更多資料。

3. 分類及預測(Classification and Prediction)

基於資料的屬性對資料做切分即為分類，並利用分類後的法則進行預測。

4. 集群分析(Cluster Analysis)

根據集群的原則將資料分群，讓群組內相似度大，群組間相異度高。

5. 演化分析(Evolution Analysis)

時間及空間資料佔了電腦資料的絕大多數，從時間資料中尋找樣式(Pattern)，預測並發現危險、因果關係與特定型態之趨勢。

2.3.2. 應用資料探勘之相關文獻

顏嘉惠(民 91)應用資料探勘方法於圖書館行銷與顧客關係管理，對圖書館使用者使用分類分析，針對使用情況與借閱次數做分類，利用 C4.5 建立決策樹；對非圖書館使用者則採用群集分析，找出不上圖書館的族群並分析其特徵，在根據結果來討論應如何吸引他們上圖書館；推薦書單的部分使用連結分析，從點擊過、借閱過的書籍的關鍵字去連結其他相關書集，計算 support 與 confidence，找出高相關的書單。亦可使用次序相關分析，從 support 與 confidence 計算 pattern 來做推薦書單的編輯。從以上四種資料探勘的方法可以找出圖書館的使用人數、滿意度、書籍借閱率是否有提高。

梁衍忠(民 101)利用資料探勘技術分析學生國語學習的成就，使用由下而上(button-up)的決策樹方法對 2003 年與 2005 年的國語學習成績進行分析，主要分析的重點為家長背景與學生背景，研究發現對學生國語成績影響最高的分別是父親的職業與家中的藏書量，父親的職業樂屬於社會高層、家中的藏書量越多，對國語成績的影響越高。其他因素包含學生性別、母親的祖籍、校外補習等，都對國語成績有明顯的影響。

蔡宗諺、楊志平、彭克仲、陳仕偉(民 98)運用資料探勘的關聯規則對超市做研究，探討商品之間的關聯性，過濾掉無效的關聯規則後，將商品分為乾貨、生鮮以及日用品三類許多結果都沒有顯示非常高的關聯，主要是由於資料量較少，生鮮類產品的消費數量較多，因此多數產品跟他的關聯性較高，應以生鮮產品為搭配重點，提高其他品項的銷售量，以改善超市的狀況。

陳垂呈、利益多(民 91)對網拍競標者進行資料探勘分析挖掘最適性，使用 ID3 決策樹分類法對會員資料，如性別、年齡、居住地等，進行分類與比較；關聯分析用來分析交易資料庫，找出交易商品之間的關聯性，以挖掘競標者的最適性。期望以主動行銷的方式來獲取新的消費者，鞏固原本的客人回流，並提升會員對產品與交易服務的滿意度。

張效娟(2008)運用 K-Means 演算法於保險業 CRM，提到 K-Means 是一種最著名、最常用的劃分方法，他最吸引人的地方在於處理巨量資料時的高效能，以距離作為劃分依據，使得同一群中的資料最相似，不同群間的資料最相異。張效娟也提到 K-Means 的缺點，初始種子與 K 值的數量對演算結果的影響非常大，在做這部分的決策時需較為謹慎。否則聚類的結果會較差。

本研究將採用集群分析作為車輛分類的方法，分類出不同的行駛狀況後，檢視資料中的車種差異，以及在不同路段所花費的旅行時間，最後希望可以為用路人提出有效的行車建議，避免在尖峰時間堵車，分散車流量。

第三章、研究方法

本章將針對 Hortonworks 系統的架構、R 語言與 Rstudio Server 的安裝、RHadoop 套件安裝以及資料整理流程進行詳述，並透過簡單的範例讓使用者了解 Hortonworks 的基本功能操作與使用流程，所使用之軟體皆為免費之開源(Open Source)軟體。。

第一節、需求軟體

3.1.1. Oracle Vm VirtualBox (ver 5.1.22)

Oracle Vm VirtualBox 於 2007 年由德國軟體公司 Inno Tek 所研發的虛擬機器軟體，目前由甲骨文公司(Oracle)進行開發維護，提供使用者在 32 位元或 64 位元的 Windows、Solaris 與 Linux 作業系統上模擬其他 X86 的作業系統，且可多個環境同時運行，彼此之間可做連結或是獨立運作，且支援快照、遠端桌面、共享資料夾等功能。

3.1.2. Hortonworks Sandbox 2.5

Hortonworks Sandbox 2.5 平台，提供使用者一個整合的 Hadoop 系統，透過分散式儲存系統 HDFS，可幫助使用者存取巨量資料；HIVE 支援使用者使用 SQL 語法對分散儲存的資料進行查詢、篩選；PIG、Spark 等套件可進行更高階的分析計算；Zepplin 提供了繪圖功能，可將資料視覺化；如需進行更進階之統計分析，可以外掛 R 語言、SAS 等統計軟體，進行進階的統計分析操作。

3.1.3. Rstudio Server (ver 0.99.893)

R 語言是一個免費的統計分析系統，在統計分析、繪圖與資料探勘的能力非常的強大，另外 Rstudio 是基於 R 語言所設計的整合介面，提供使用者更為人性化的介面。Rstudio 也提供了一般單機版與伺服器版本，伺服器版本的 Rstudio 可支援多人使用，亦可外掛於 Hortonworks 上與資料庫串聯，為雲端平台提供更強大的統計分析能力

3.1.4. Putty (ver 0.70)

Putty 是一款讓使用者在 Windows 作業系統上連結 SSH 伺服器的工具。本研究將透過 Putty 連結至 Hortonworks，可以在 Linux 環境使用複製、貼上等快捷功能，增加使用者上的方便度。

3.1.5. Google Chrome (ver 59.0.3071.115)

Google Chrome 是一款由 Google 開發的免費瀏覽器，相容性較其他瀏覽器(如：Internet Explorer、Firefox)佳，由於本研究使用的資料平台 Hortonworks 多數的使用需在瀏覽器上進行，因此選擇使用 Google Chrome。

3.1.6. Anaconda(ver2-4.4)

Anaconda 是一款輔助使用者學習、操作 Python 的免費軟體，他幫你裝好許多資料分析、科學計算的 Python 套件，且包含 IPython Notebook，可以讓使用者在一個非常好用的介面上進行操作。亦支援 Linux、Mac、Windows 等作業系統，本研究所需使用之資料於高速公路國道資料庫以每小時一個檔案的方式存放，因此需透過 Anaconda 之 Python 語言快速下載大量資料。

第二節、 研究環境架構概念圖

本研究所使用之硬體為 64 位元之電腦，系統為 Windows 7，並在 Oracle Vm VirtualBox 中安裝 Hortonworks 所開發之 Linux 系統的 Hadoop 雲端分析平台，最後再結合 Rstudio Server 進行後續的統計分析研究，整體環境架構圖如圖 3.1。

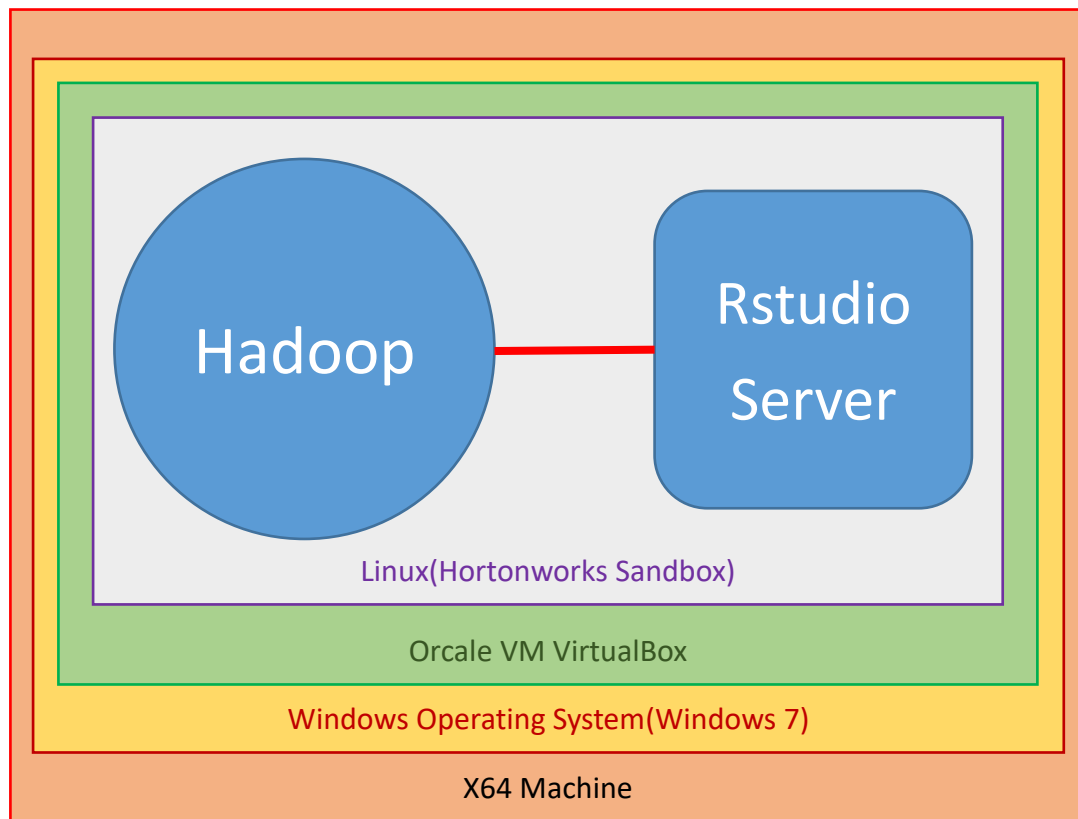


圖 3.1 研究環境結構圖

第三節、 Hortonworks 架構與範例資料模擬

本研究將使用 Hortonworks 與 Rstudio Server 作為巨量資料整理及統計分析環境，詳細設置步驟如下：

3.3.1. 安裝 Hortonworks Sandbox 2.5

(1) 下載軟體

Hortonworks Sandbox 2.5 有對應三種不同虛擬機環境的版本，在 Hortonworks 官方網站下載時要下載本研究所對應 Oracle VM VirtualBox 的 Hortonworks Sandbox 2.5，下載頁面如圖 3.2，檔案 HDP_2.5_virtualbox.ova 的大小大約 10G。

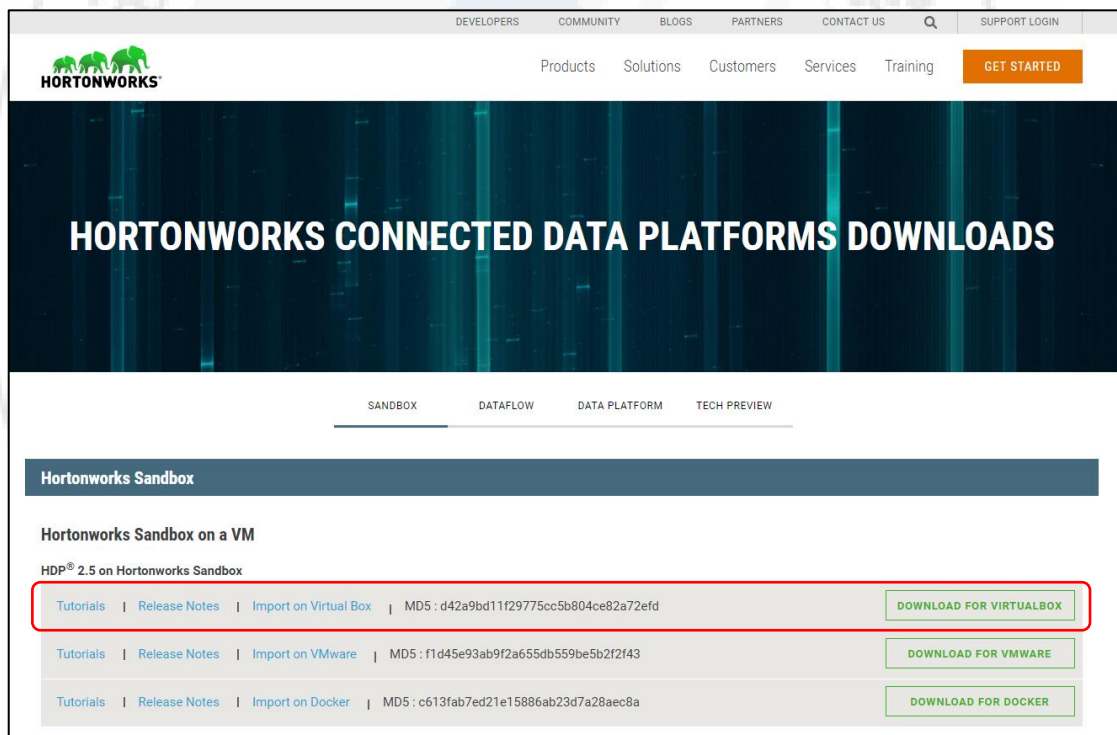


圖 3.2 下載 Hortonworks Sandbox 2.5(VirtualBox 版本)

(2) 安裝 Hortonworks Snadbox 2.5 於 VirtualBox 虛擬機

Hortonworks Snadbox 2.5 下載完成後，開啟 Oracle VM VirtualBox，點選檔案→匯入應用裝置，選擇下載完成後的 ova 檔 Hortonworks Snadbox 2.5 並匯入，如圖 3.3 與圖 3.4。



圖 3.3 匯入 Hortonworks Sandbox 2.5(一)

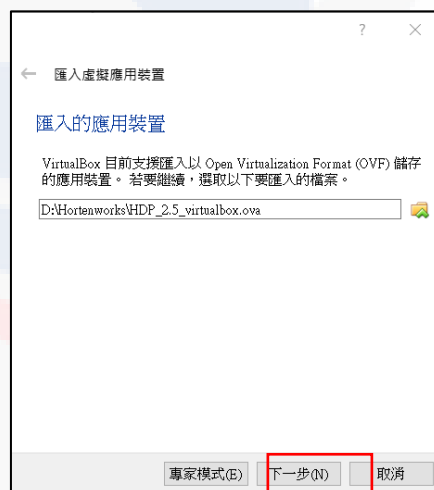


圖 3.4 匯入 Hortonworks Sandbox 2.5(二)

在硬體設置的部分(如圖 3.5)，預設為 CPU 四顆、RAM 記憶體 8192 MB，可依照電腦規格調整至最低兩顆 CPU，RAM 記憶體的部分則建議最低 8192 MB，低於以上標準將影響整體系統的運行效能。

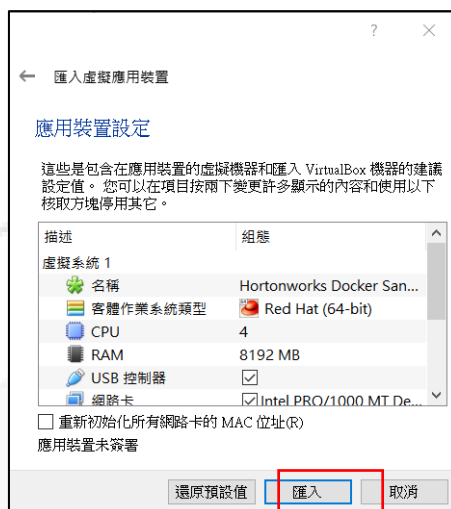


圖 3.5 匯入 Hortonworks Sandbox 2.5(三)

(3) 設定網路卡

匯入成功後要設定額外網路卡，供虛擬機使用。點擊檔案→喜好設定→網路→[僅限主機]網路，新增一張虛擬網路卡。新增完成後進入安裝好的 Hortonworks 設定值，將新增的虛擬網路卡指定給 Hortonworks 使用，如圖 3.6 與圖 3.7。

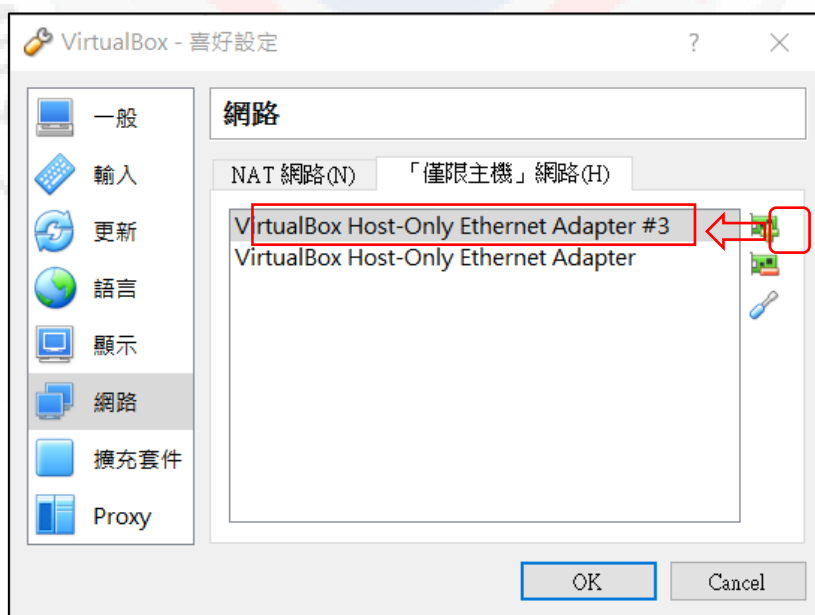


圖 3.6 新增虛擬網路卡

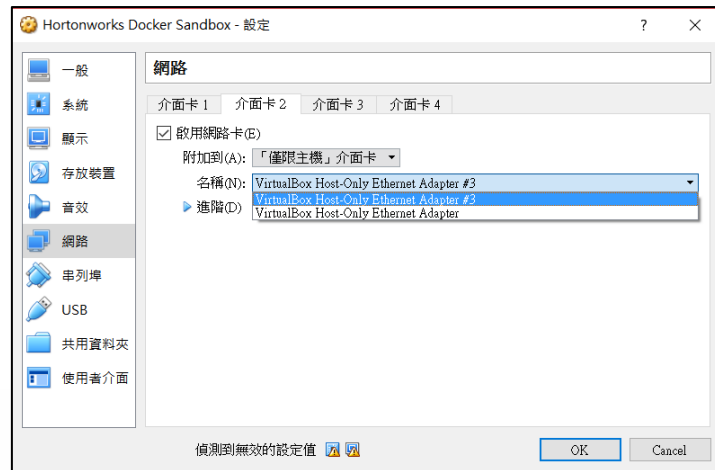


圖 3.7 指定新增的虛擬網路卡給 Hortonworks

(4) 啟動 Hortonworks Sandbox 確認運行狀態

在 Oracle VM VirtualBox 點選 Hortonworks 後點擊啟動鈕(如圖 3.8)。成功啟動後的 HDP 系統如(圖 3.9)。確認後，即可開啟 Google Chrome 瀏覽器，登入使用者介面後即可操作各式套件，如 HDFS、HIVE 等。



圖 3.8 於 Oracle VM VirtualBox 啟動 Hortonworks

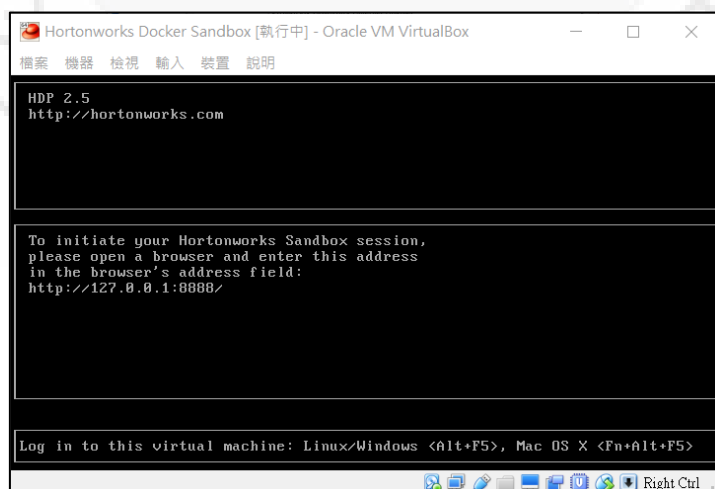


圖 3.9 成功啟動後之畫面

當開啟 Google Chrome 瀏覽器後，輸入網址 <http://127.0.0.1:8888>，即可進入使用者歡迎介面(如圖 3.10)，左邊的 LUNCH DASHBOARD 按鈕可登入使用者介面，右邊的 QUICK LINK 按鈕進入之後可連結至各個套件的連接埠，如 Zeppelin、Ranger 等。輸入網址 <http://127.0.0.1:8080> 或是點選 LUNCH DASHBOARD 可進入登入畫面(如圖 3.11)，在此我們使用系統預設之操作員身分的 maria_dev 作為登入帳號密碼登入系統。帳號密碼同樣為 maria_dev。登入後便可看到系統首頁，所呈現的各個圖是為各類套件的運行狀態。

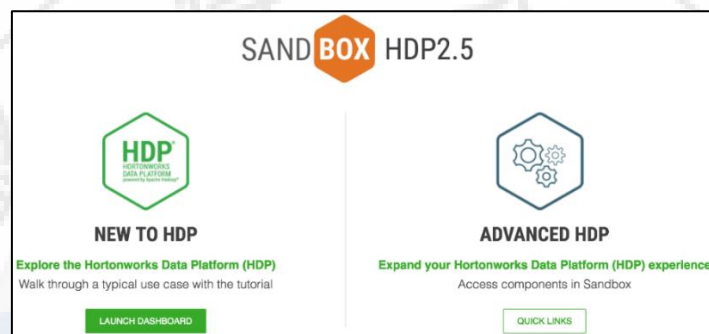


圖 3.10 Hortonworks 歡迎介面

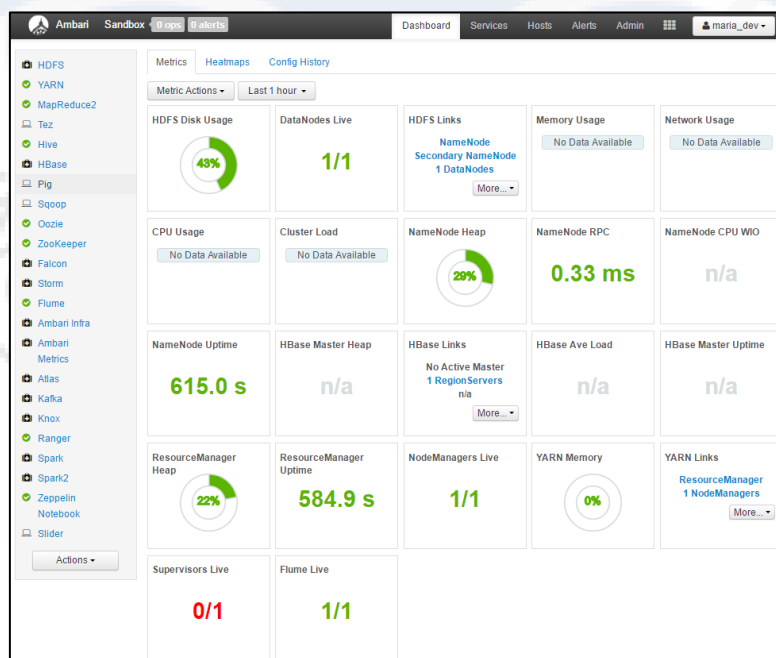


圖 3.11 Hortonworks 登入首頁

(5) 透過 Putty 設定 HDP 管理者密碼

開啟 Putty 後，在 Host Name 的位置輸入 Hortonworks 的 IP 位置 127.0.0.1，Port 的位置輸入連接埠位置 2222，如圖 3.12，點選開啟即可進入 HDP 管理者畫面，管理者帳號為 root，預設密碼為 hadoop(輸入時不會顯示)，如圖 3.13。

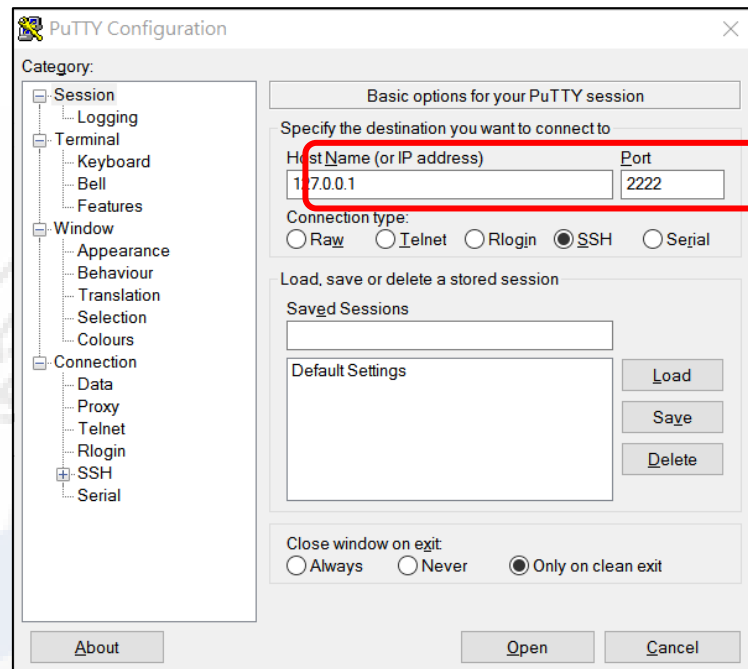


圖 3.12 Putty 上登入管理者帳號密碼

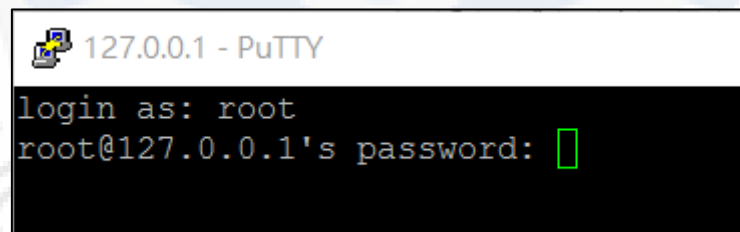


圖 3.13 Putty 上登入管理者帳號密碼

首次登入會要求修改密碼，登入後再次輸入原密碼 hadoop，接著輸入兩次新的密碼即可修改成功。

3.3.2. Hortonworks 基本功能與範例

本節將介紹 Hortonworks 的基本功能操作，主要介紹重點有 HDFS 存取資料、HIVE 操作(從 HDFS 存取資料與查詢功能)、PIG 的進階計算能力以及 Zeppelin 繪圖功能。

範例使用的資料為 Hortonworks 官方教學中所提供之車輛行駛資料，geolocation.csv 與 trucks.csv，Geolocation 資料中的變數包含車輛編號、行駛狀況(正常與否)、所在城市的相關資料以及車速；Trucks 的資料中包含了每輛車的廠牌以及兩年半中，每個月的行駛里程與耗油量，下載網址為 http://hortonworks.com/hadoop-tutorial/hello-world-an-introduction-to-hadoop-hcatalog-hive-and-pig/#section_1，下載位置如圖 3.14。本研究將透過兩個行駛資料作為範例，計算貨車這兩年半的異常駕駛狀態，並計算風險因子，探索 Hortonworks 對資料處理的基本功能。

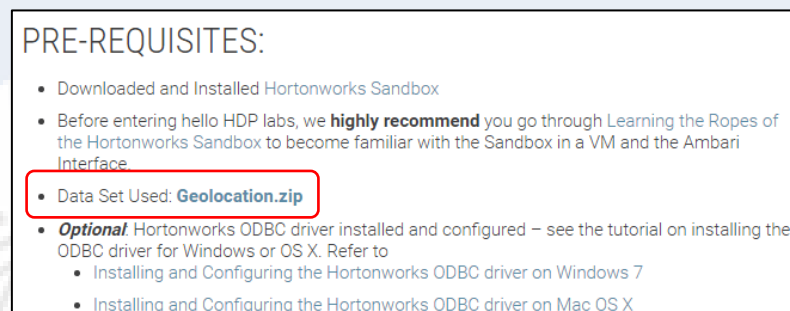


圖 3.14 範例資料下載位置

1. HDFS 存取資料

輸入網址 <http://127.0.0.1:8080> 後，點選右上角的選單紐並點擊 File Views 進入 HDFS。如圖 3.15。



圖 3.15 從使用者介面進入 HDFS

進入 HDFS 後的介面呈現就如同檔案總管一樣(如圖 3.16)，選擇使用者身分 maria_dev 的資料夾，進入後點選 New Folder 建立存放資料用的資料夾，命名為 data(如圖 3.17)。

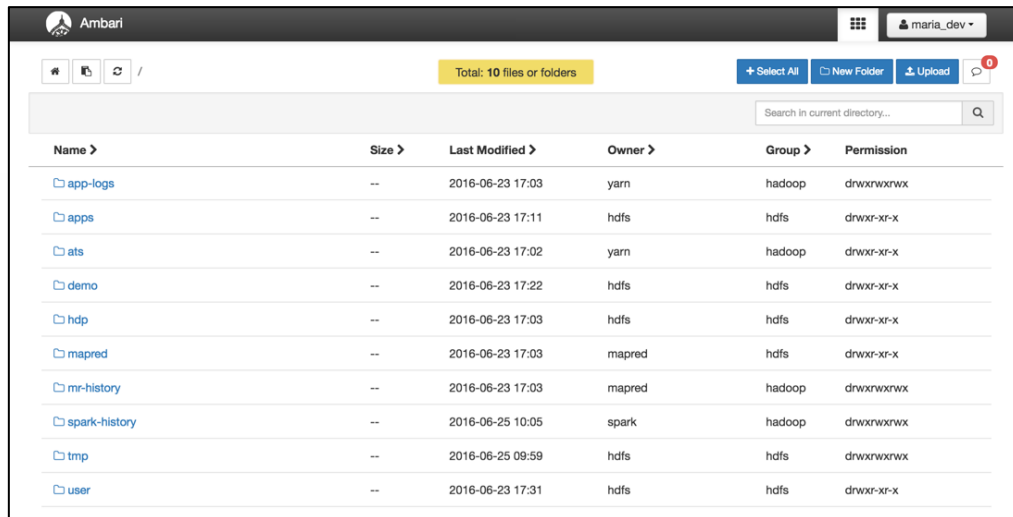


圖 3.16 HDFS 介面

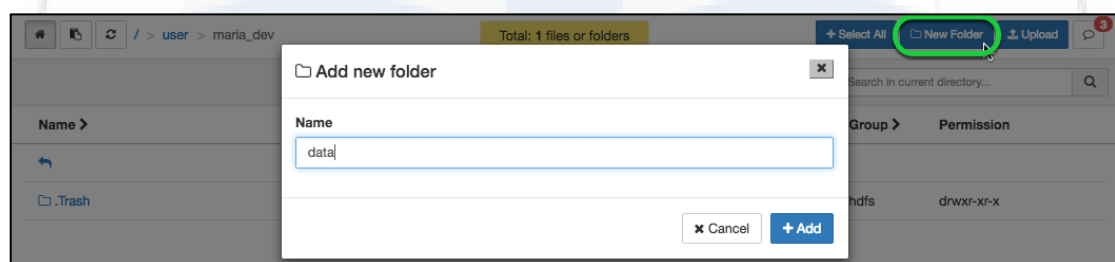


圖 3.17 HDFS 中建立新資料夾

進入建立的 data 資料夾後，點選 Upload 上傳檔案(如圖 3.18)，可以點擊圖示開啟瀏覽資料夾，或是直接將檔案拖曳進視窗中，在此將 geolocation.csv 與 trucks.csv 上傳至 HDFS 中，上傳資料後圖如 3.19。

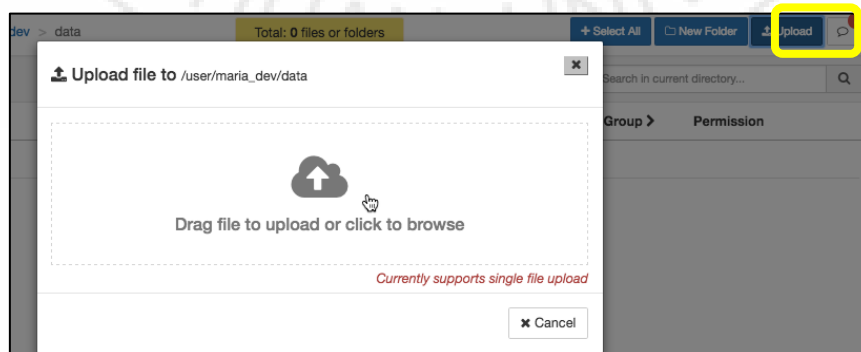


圖 3.18 上傳檔案至 HDFS(一)

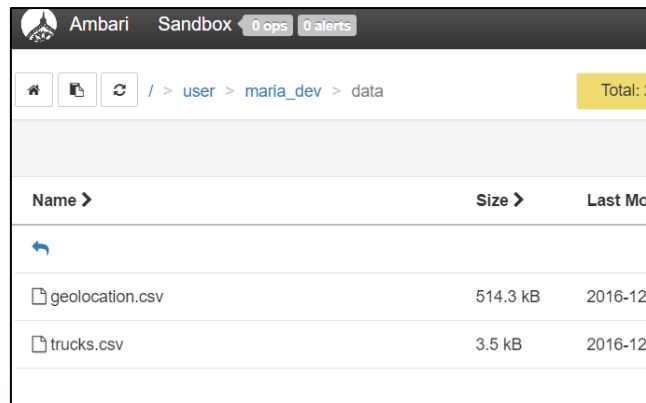


圖 3.19 上傳檔案至 HDFS(二)

回到 maria_dev 的資料夾，點選 DATA 資料夾後點擊 Permissions 設定權限，將 Group 與 Other 的 Write 啟動(藍色)，如圖 3.20。其他套件(如：HIVE、SPARK.....等)便可讀寫存取至 HDFS 中的資料。

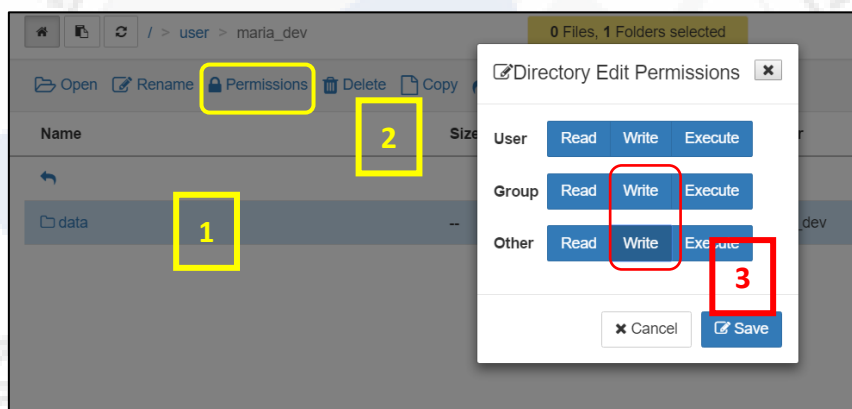


圖 3.20 設定資料權限

2. HIVE 操作(從 HDFS 存取資料與查詢功能)

HIVE 讓使用者可以從本機端或是 HDFS 中讀取資料轉存成資料表，其轉檔後的隔視為 ORC 格式，將資料從原本的行儲存改為列儲存，存取資料時可減少需花費的時間，於首頁點擊右上角選單並選擇 Hive View 進入 Hive 介面，如圖 3.21。

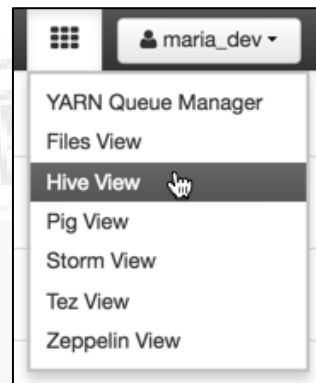


圖 3.21 於首頁進入 Hive 介面

進入 Hive 介面後如圖 3.22，左邊是資料庫及資料表的列表，中間是 HiveQL 語法輸入的地方，輸入視窗下方為輸出結果的位置，上方列表的功能由左至右分別為選擇輸入語法、檢視儲存的語法、檢視歷史紀錄、定義 UDF 介面與上傳資料表，右邊長條圖示則可以進行各種設定、簡單的圖形化資料、執行資訊通知等。



圖 3.22 Hive 介面介紹

在匯入資料表之前，需先設定資料的匯入形式，點選右側齒輪進入設定頁面，按下+ADD 按鈕後選擇 hive.execution.engine，並設定其 Value 為 TEZ，將資料匯入後，用 TEZ 格式壓縮讀取。設定完成後按下 save default settings 儲存，如圖 3.23。

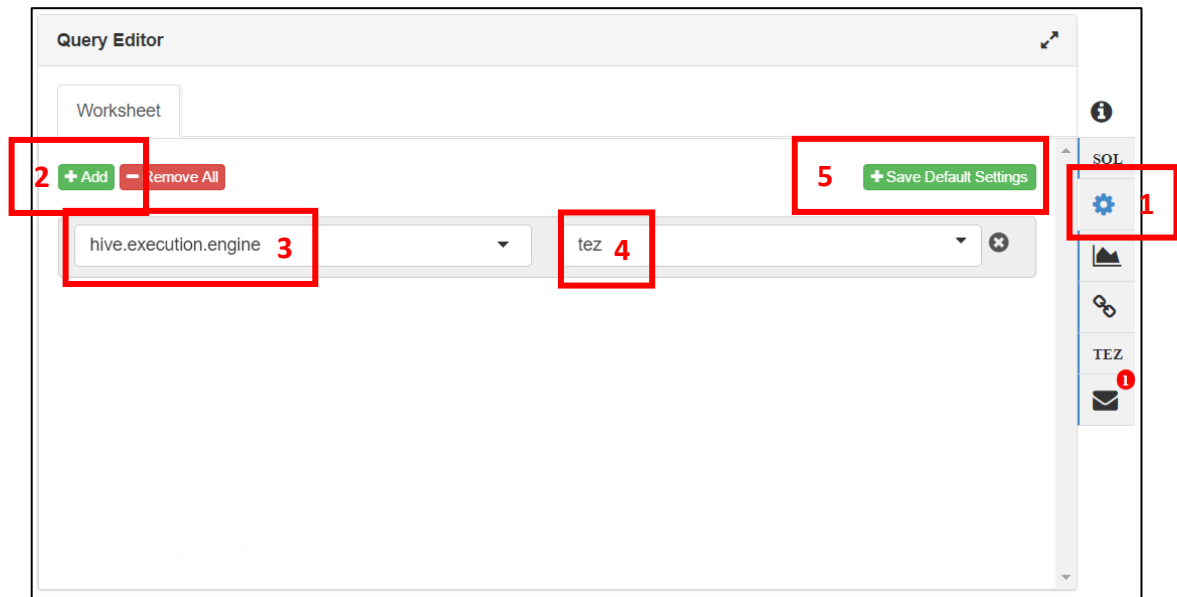


圖 3.23 設定資料讀取格視為 TEZ 壓縮檔

設定完成之後，選擇 Upload Table 即可匯入檔案，有兩種匯入方式：從本機端或是從 HDFS，如圖 3.24。從 HDFS 匯入資料時需要手動輸入資料的路徑，在此先從 HDFS 匯入 turcks.csv，路徑為/user/maria_dev/data/trucks.csv。如圖 3.25。

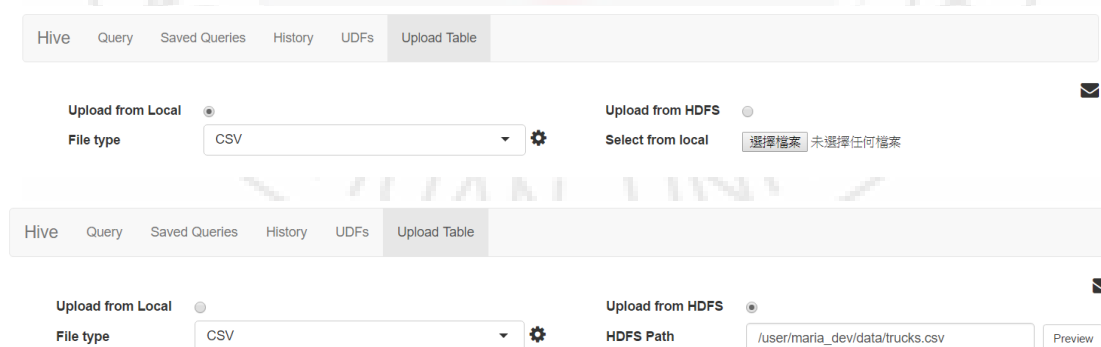


圖 3.24 HIVE 中 Upload Table 頁面(本機端匯入/HDFS 匯入)

指定完檔案路徑點選 preview 後可預覽資料，並選擇儲存的資料庫、儲存格式、資料表名稱、齒輪選項中可選擇是否要指定第一列為變數名稱等，設定完成後點選 Upload Table 即可將資料表存入 Hive 中，如圖 3.25。

Upload from Local 點選進入可設定第一列資料為變數名稱

File type: CSV

Database: default 選擇儲存資料庫

Stored as: ORC

Upload from HDFS

HDFS Path: /user/maria_dev/data/trucks.csv

Table name: trucks 資料表名稱

Contains endlines? ☐

上傳至 HIVE Upload Table

driverid	truckid	model	jun13_miles
A1	A1	Freightliner	9217
A2	A2	Ford	12058

圖 3.25 儲存資料表之相關介面與設定

上傳完成後點選上方選單的 Hive 可回到 Hive 介面主頁，從左邊資料庫列表選擇剛剛指定的資料庫，點選上傳的資料表名稱旁邊的圖示即可瀏覽前一百筆資料存入後的狀態，並在中間語法輸入的地方輸入 HiveQL 語法做資料查詢，如圖 3.26。

Hive 回到 Hive 主頁 Query Saved Queries History UDFs Upload Table

Database Explorer

default

Search tables...

Databases

- default
- sample 07
- sample 08
- trucks 點選瀏覽前 100 筆資料
- foodmart
- xademo

Query Editor

Worksheet trucks sample * x

1 SELECT * FROM trucks LIMIT 100;

Execute Explain Save as... New Worksheet

Query Process Results (Status: SUCCEEDED) Save results...

Logs Results

Filter columns...

trucks.driverid	trucks.truckid	trucks.model	trucks.jun13_miles	trucks.jun13_gas	trucks.may13_miles	trucks.may13_gas
A1	A1	Freightliner	9217	1914	8769	189
A2	A2	Ford	12058	2335	14314	264
A3	A3	Ford	13652	2899	12075	260

圖 3.26 HIVE 瀏覽資料

輸入以下 SQL 語法可建立名為 truck_mileage 的新資料表，並從 truck 資料表中篩選、計算變數，將變數 jun13_miles 至 jan09_gas 依照日期轉置儲存至 miles 與 gas 變數中。結果如圖 3.27。

```
CREATE TABLE truck_mileage
STORED AS ORC AS
SELECT truckid, driverid, rdate, miles, gas, miles / gas mpg
FROM trucks
LATERAL VIEW stack(54,
'jun13'.jun13 miles.jun13 gas.'may13'.may13 miles.may13 gas.'apr13'.apr13 miles.apr13 gas.'mar13'.mar13 miles.mar13 gas,
'feb13'.feb13 miles.feb13 gas.'jan13'.jan13 miles.jan13 gas.'dec12'.dec12 miles.dec12 gas.'nov12'.nov12 miles.nov12 gas.'oct12'.oct12 miles.oct12 gas.'sep12'.sep12 miles.sep12 gas.'aug12'.aug12 miles.aug12 gas.'jul12'.jul12 miles.jul12 gas.'jun12'.jun12 miles.jun12 gas.'may12'.may12 miles.may12 gas.'apr12'.apr12 miles.apr12 gas.'mar12'.mar12 miles.mar12 gas.'feb12'.feb12 miles.feb12 gas.'jan12'.jan12 miles.jan12 gas.'dec11'.dec11 miles.dec11 gas.'nov11'.nov11 miles.nov11 gas.'oct11'.oct11 miles.oct11 gas.'sep11'.sep11 miles.sep11 gas.'aug11'.aug11 miles.aug11 gas.'jul11'.jul11 miles.jul11 gas.'jun11'.jun11 miles.jun11 gas.'may11'.may11 miles.may11 gas.'apr11'.apr11 miles.apr11 gas.'mar11'.mar11 miles.mar11 gas.'feb11'.feb11 miles.feb11 gas.'jan11'.jan11 miles.jan11 gas.'dec10'.dec10 miles.dec10 gas.'nov10'.nov10 miles.nov10 gas.'oct10'.oct10 miles.oct10 gas.'sep10'.sep10 miles.sep10 gas.'aug10'.aug10 miles.aug10 gas.'jul10'.jul10 miles.jul10 gas.'jun10'.jun10 miles.jun10 gas.'may10'.may10 miles.may10 gas.'apr10'.apr10 miles.apr10 gas.'mar10'.mar10 miles.mar10 gas.'feb10'.feb10 miles.feb10 gas.'jan10'.jan10 miles.jan10 gas.'dec09'.dec09 miles.dec09 gas.'nov09'.nov09 miles.nov09 gas.'oct09'.oct09 miles.oct09 gas.'sep09'.sep09 miles.sep09 gas.'aug09'.aug09 miles.aug09 gas.'jul09'.jul09 miles.jul09 gas.'jun09'.jun09 miles.jun09 gas.'may09'.may09 miles.may09 gas.'apr09'.apr09 miles.apr09 gas.'mar09'.mar09 miles.mar09 gas.'feb09'.feb09 miles.feb09 gas.'jan09'.jan09 miles.jan09 gas ) dummyalias AS rdate, miles, gas;
```

The screenshot shows a database interface with a 'Database Explorer' on the left and a 'Query Editor' on the right. The 'Query Editor' displays a SQL query: `SELECT * FROM truck_mileage LIMIT 100;`. Below the query editor, the 'Query Process Results (Status: SUCCEEDED)' are shown. The results are displayed in a table with columns: truck_mileage.truckid, truck_mileage.driverid, truck_mileage.rdate, truck_mileage.miles, truck_mileage.gas, and truck_mileage.mpg. The first three rows of data are visible.

truck_mileage.truckid	truck_mileage.driverid	truck_mileage.rdate	truck_mileage.miles	truck_mileage.gas	truck_mileage.mpg
A1	A1	jun13	9217	1914	4.81
A1	A1	may13	8769	1892	4.63
A1	A1	apr13	14234	3008	4.73

圖 3.27 計算變數另存之 truck_mileage 資料表

下一步要計算各車輛行駛里程加總，輸入以下程式並將結果儲存至 DriverMileage 資料表，執行後之結果如圖 3.28。

```
CREATE TABLE DriverMileage
STORED AS ORC AS
SELECT driverid, sum(miles) totmiles
FROM truck_mileage
GROUP BY driverid;
```

Query Process Results (Status: SUCCEEDED)

Logs Results

Filter columns...

drivermileage.driverid	drivermileage.totmiles
A1	628507.0
A10	675377.0
A100	634338.0
A11	652452.0
A12	668241.0

圖 3.28 DriverMileage 資料表輸出結果

以上是 Hive 的基本資料存取與查詢功能，可以發現它的操作方式與 SQL 資料庫非常類似，不過進階的計算功能就沒有 SQL 那麼便利，因此我們會要用到其他套件來協助我們進行計算，下一段將介紹 PIG，讓我們可以進行進階統計值的計算。

3. PIG 的進階計算

進入 PIG 進行操作之前，我們需要在 HIVE 預先建立好一個資料表給 PIG 執行後的結果進行儲存，先在 HIVE 輸入以下 SQL 語法建立空白資料表。

```
CREATE TABLE riskfactor (driverid string,events bigint,totmiles bigint,riskfactor float) STORED AS ORC;
```

如同 HDFS 與 HIVE，先從右上角選單點選 PIG view 開啟 PIG 介面後，要先建立一個空白程式檔，點選 new script 後輸入編譯檔案名稱為 riskfactor.pig 按下 create 即可建立。如圖 3.29。

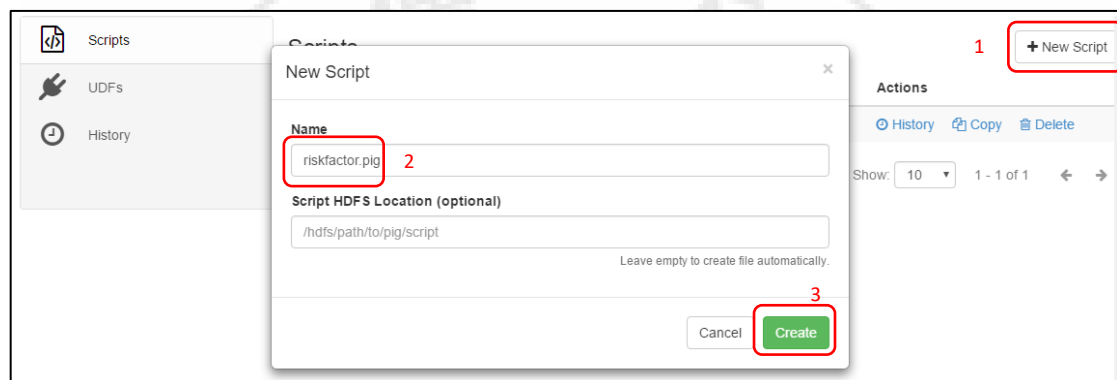


圖 3.29 建立 PIG 程式碼

建立好的畫面如圖 3.30，介面中點選 pig Helper 可選擇函數，協助使用者進行程式碼的編寫，Execute 可執行編寫的程式碼，History 可看過去的程式執行紀錄。下方的 Argument 可載入所需插件，讓 PIG 可連結其他 Hadoop 套件，在此先輸入-useHCatalog 點選+Add。

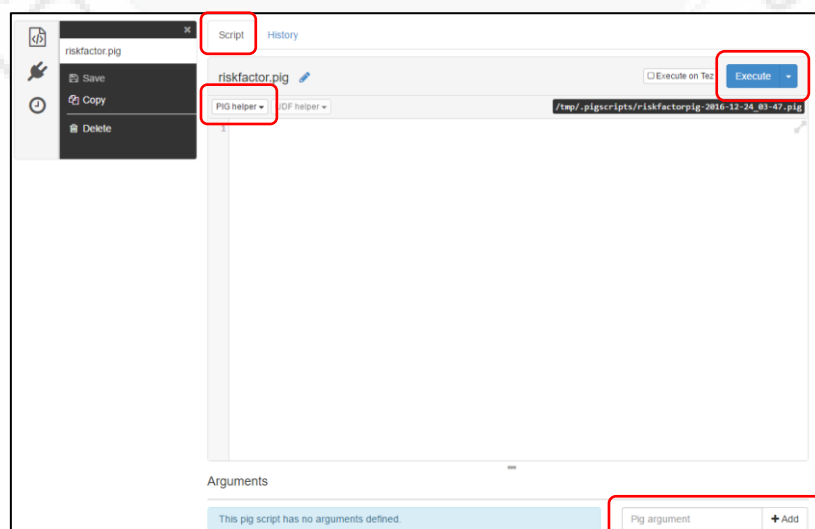


圖 3.30 PIG 基本介面介紹

在 PIG 輸入下列語法，目的在於進行變數的計算，統計每輛車的非正常行駛狀況後，計算風險因子(Risk Factor)並回存至 HIVE 中。

PIG 程式碼	說明
a = LOAD 'geolocation' using org.apache.hive.hcatalog.pig.HCatLoader();	從 Hive 讀取 geolocation 資料表，存至 a。
b = filter a by event != 'normal';	從 a 篩選變數 event 中，不為”normal”的資料存至 b。
c = foreach b generate driverid, event, (int) '1' as occurrence;	新增一變數，名稱為 occurrence，每筆資料都給予一數值 1，並將變數 driverid、event、occurrence 篩選儲存至 c。
d = group c by driverid;	將 c 以 driverid 為群組存至 d。
e = foreach d generate group as driverid, SUM(c.occurrence) as t_occ;	將以 driverid 為群組後的 d 與計算 occurrence 加總的 c 合併為 e。
g = LOAD 'drivermileage' using org.apache.hive.hcatalog.pig.HCatLoader();	從 HIVE 載入 drivermileage 資料表存至 g。
h = join e by driverid, g by driverid;	將 g 與 e 以 driverid 並聯為 h。
final_data = foreach h generate \$0 as driverid, \$1 as events, \$3 as totmiles, (float) \$3/\$1 as riskfactor;	於 h 中選出以下變數 driverid、events、totmiles，totmiles 除以 events 命名為 riskfactor。將資料存至 final_data。
store final_data into 'riskfactor' using org.apache.hive.hcatalog.pig.HCatStorer();	將 final_data 回存至 HIVE 中。

點選 Execute 執行以上 PIG 程式碼，待出現圖 3.31 顯示完成後，即可回到 HIVE 檢視結果。

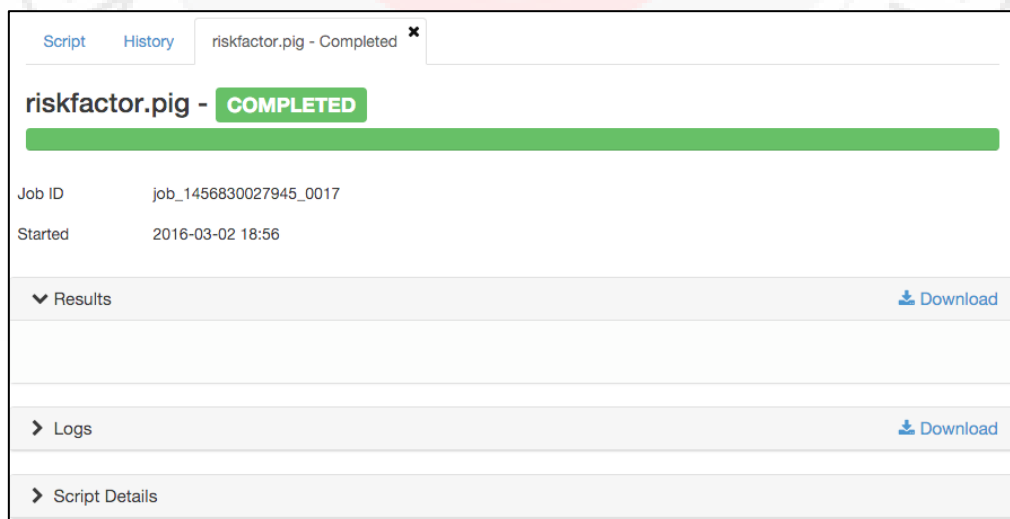
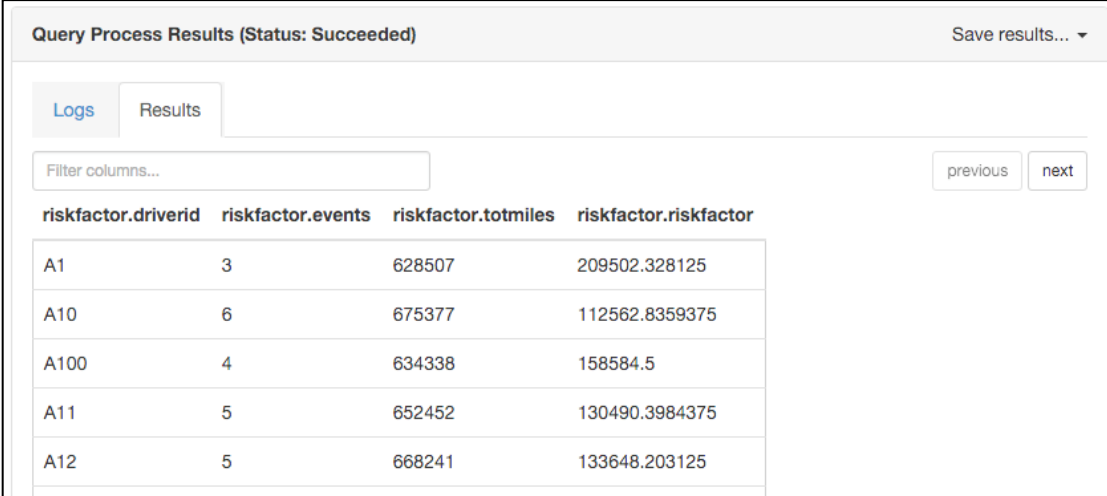


圖 3.31 PIG 執行完成畫面

回到 HIVE 後便可發現剛剛回存的資料表 riskfactor 已出現在資料庫中，點選後可看到計算結果。如圖 3.32。



The screenshot displays the 'Query Process Results (Status: Succeeded)' window. It features a 'Save results...' dropdown in the top right. Below the status bar, there are tabs for 'Logs' and 'Results', with 'Results' being the active tab. A 'Filter columns...' input field is located below the tabs. To the right of the filter field are 'previous' and 'next' navigation buttons. The main area shows a table with four columns: 'riskfactor.driverid', 'riskfactor.events', 'riskfactor.totmiles', and 'riskfactor.riskfactor'. The table contains six rows of data, with the first row being the header and the subsequent five rows representing individual driver records.

riskfactor.driverid	riskfactor.events	riskfactor.totmiles	riskfactor.riskfactor
A1	3	628507	209502.328125
A10	6	675377	112562.8359375
A100	4	634338	158584.5
A11	5	652452	130490.3984375
A12	5	668241	133648.203125

圖 3.32 riskfactor 資料表結果瀏覽

至此，PIG 的計算就告一段落，下一段將介紹如何用 Zeppelin 繪圖，檢視圖形化資料。

4. Zeppelin 繪圖功能

Zeppelin 是一個 Web 版的記事本，可以在上面執行許多種類的程式語言，並檢視輸出結果，結果為資料表，亦可以圖形方式呈現。

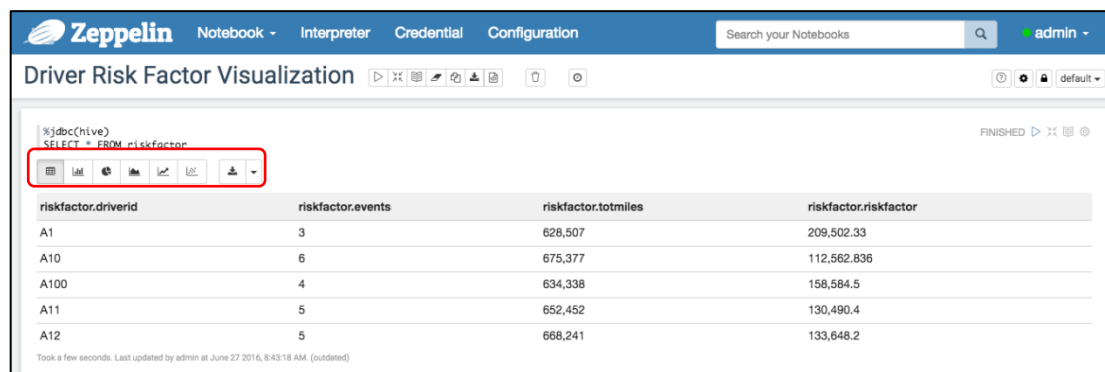


圖 3.33 Zeppelin 主頁

在網址輸入 <http://127.0.0.1:9995> 即可進入 Zeppelin。如圖 3.33。進入 Zeppelin 後，點選上方 notebook 在按下 create new note 建立新的記事本，名稱為 Driver Risk Factor。如圖 3.34，建立後便可在視窗中輸入以下程式呼叫資料表。

```
%jdbc(hive)
SELECT * FROM riskfactor
```

在 Zeppelin 中要執行某程式語言前(如：R、SQL Query、SPARK 等)，必須先指定其名稱，使用%(套件名稱)後，第二行開始即可輸入該程式語言的指令。編寫完成後於鍵盤按下 Shift+enter 執行該段落，結果如圖 3.34。



The screenshot shows the Zeppelin Notebook interface with a Hive query executed. The query is `%jdbc(hive) SELECT * FROM riskfactor`. The result is displayed as a table with the following data:

riskfactor.driverid	riskfactor.events	riskfactor.totmiles	riskfactor.riskfactor
A1	3	628,507	209,502.33
A10	6	675,377	112,562.836
A100	4	634,338	158,584.5
A11	5	652,452	130,490.4
A12	5	668,241	133,648.2

圖 3.34 於 Zeppelin 執行 Hive 程式呼叫資料表

輸出結果如同在 HIVE 看到的一樣，可以檢視資料表，不過在上方的選單中亦可以不同圖形呈現資料，點選第二個長條圖的圖示即可以長條圖的方式瀏覽，按下 setting 便可以設定所使用的變數，如圖 3.35。

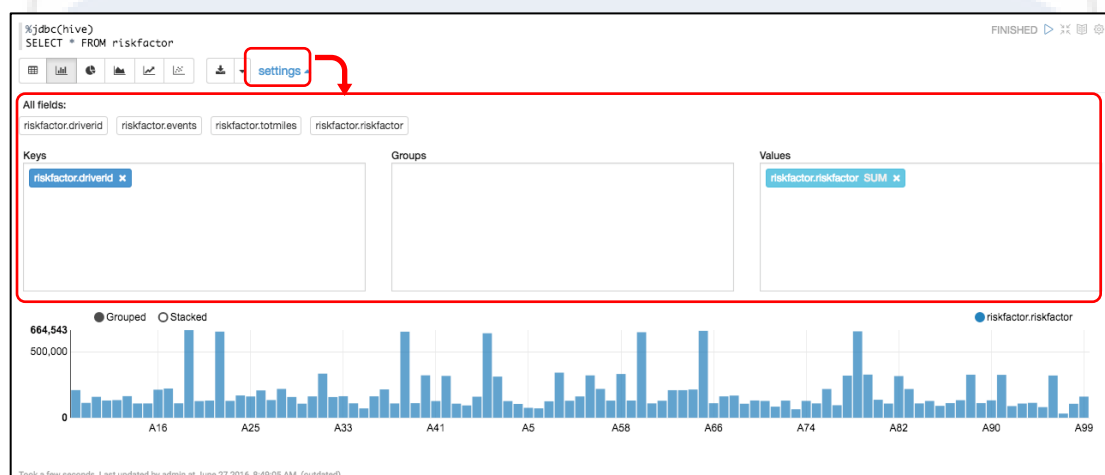


圖 3.35 Zeppelin 中選擇變數與繪圖

以上便是 Hortonworks 的基本操作介紹，可以發現在此平台上不但可以解決很多基本統計分析的問題，亦可讓使用者儲存大量資料，學習使用類 SQL 語法的 HiveQL，並做基本的繪圖。但如需進行進階統計分析如集群分析、資料探勘等功能，我們需要額外掛載其他統計軟體，下一個章節將介紹在 Hortonworks 中安裝伺服器版的 Rstudio - Rstudio Server。

第四節、 Rstudio Server 安裝

本節將講解如何在 Hortonworks 中安裝 Rstudio Server 以及 Rstudio 連結 HDP 所需的套件，讓大數據平台能夠擁有更強大的統計分析與繪圖能力，另外，由於我們要安裝的是伺服器版的 Rstudio，因此在做設定與套件的安裝時，透過 Putty 直接安裝進系統中，才可避免日後設定出錯。

3.4.1. 在 Hortonworks 環境中安裝 Rstudio Server

(1) 設定 RstudioServer 專用連接埠

在 Oracle VM VirtualBox 中打開 Hortonworks 設定值，點選網路→介面卡 1→進階→連接埠傳送，將 Custom4 名稱修改為 rstudioserver，指定 rstudioserver 的 port 為 8090，做為日後登入 Rstudio 使用，如圖 3.36。

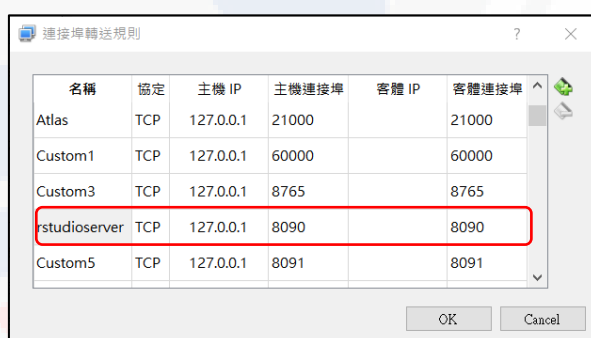


圖 3.36 設定 rstudioserver 專用連接埠

(2) 安裝 Rstudio Server 與相關軟體

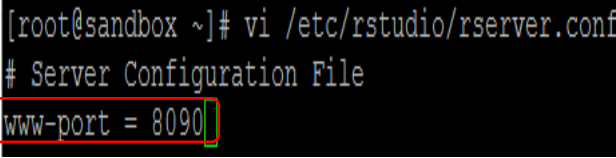
從 Putty 安裝 Rstudio Server 之過程，依序輸入以下 Linux 程式碼。

Linux 程式碼	說明
1 wget https://download2.rstudio.org/rstudio-server-rhel-0.99.893-x86_64.rpm	從網路下載 Rstudio server 至 HDP。
2 sudo yum install --nogpgcheck rstudio-server-rhel-0.99.893-x86_64.rpm	安裝 Rstudio Server。
3 ln -s /usr/lib64/libssl.so.10 /usr/lib64/libssl.so.6	將兩個指定之路徑資料夾進行連結。
4 ln -s /usr/lib64/libcrypto.so.10 /usr/lib64/libcrypto.so.6	將兩個指定之路徑資料夾進行連結。
5 yum install dpkg	安裝 dpkg 安裝工具包。(用途：用於安裝、解除安裝和供給和 .deb 軟體包相關的資訊)
6 dpkg-divert --local --rename --add /sbin/initctl	設定 dpkg 工具的版本、路徑與名稱。
7 ln -s /bin/true /sbin/initctl	將兩個指定之路徑資料夾進行連結。
8 yum install R	下載安裝最新版本之 R 語言。

9 rstudio-server verify-installation	對安裝好的 Rstudio Server 進行啟動確認
10 sudo adduser rstudio	建立 Rstudio 使用者，名稱為 Rstudio
11 sudo passwd rstudio	為新建立的使用者 Rstudio 設定密碼，輸入此程式碼後會要求輸入兩次密碼盡情確認。

(3) 於 config 檔中新增連接埠位置

將步驟(1)指定給 Rstudio 的連接埠 8090 儲存至 config 文件中，並退出 putty。

Linux 程式碼	說明
1 vi /etc/rstudio/rserver.conf	開啟 etc/rstudio 資料夾中的 rserver.conf 文件
	開啟文件後按下 insert 鍵即可進入編輯模式，用方向鍵移至文件最末行並輸入 www-port = 8090，指定 Rstudio Server 的連接埠，呼應步驟 2.1。輸入完成後按下 Esc 鍵並輸入:wq 即可儲存並退出文件。
2 exec /usr/lib/rstudio-server/bin/rserver	將上述所有指令傳送給 Rstudio 並退出 Putty。

完成以上步驟後即可回到 Google Chrome 瀏覽器中，輸入網址 <http://127.0.0.1:8090> 即可連結至 Rstudio Server 登入畫面，輸入步驟(2) Linux 程式碼第 10 與 11 行所建立的使用者帳號密碼即可登入，如圖 3.37 與圖 3.38。

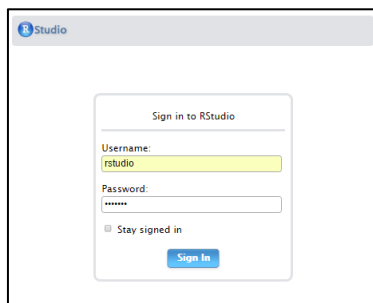


圖 3.37 Rstudio Server 啟動畫面

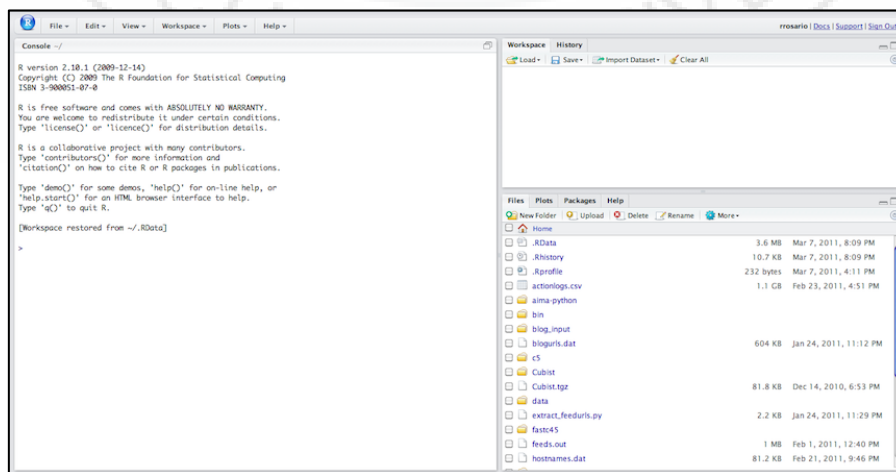


圖 3.38 Rstudio Server 登入畫面

3.4.2. 安裝連結 R 與 Hadoop 的相關套件(rmr2、rhive)

(1) 檢查 Hadoop 版本編號

在 Putty 中，重新開啟 HDP 並輸入帳密後，檢查所安裝的 Hadoop 版本編號，
以方便後續的編輯文件。

Linux 程式碼	說明
<code>1 find / -name 'hadoop-streaming*.jar'</code>	從系統中尋找檔名開頭為 hadoop-streaming 的 jar 檔
<pre>..... /usr/hdp/2.5.0.0-1245/hadoop-mapreduce/hadoop-streaming- 2.7.3.2.5.0.0-1245.jar /usr/hdp/2.5.0.0-1245/hadoop-mapreduce/hadoop-streaming.jar /usr/hdp/2.5.0.0-1245/oozie/share/lib/mapreduce- streaming/hadoop-streaming-2.7.3.2.5.0.0-1245.jar</pre>	執行結果會列出許多”未發現檔案或資料夾”的結果，從列表 中找出左側這三個有 hadoop-streaming 為開頭的 jar 檔 名的路徑並複製到記事本中，以利後續編輯文件使用。

(2) 編輯環境設定文件

將系統所查出的版本指定給 Rstudio Server 做連結，才可正確讀取資料庫中的資料，以下為相關文件的編輯內容。

Linux 程式碼	說明
<pre>1 sudo vi /usr/lib64/R/etc/Renviron</pre> <pre>### END: ### R_LIBS_SITE=\${R_LIBS_SITE-'/usr/local/lib/R/site- library:/usr/local/lib/R/library:/usr/lib64/R/lib rary:/usr/share/R/library'}</pre> <pre>HADOOP_CMD='/usr/bin/hadoop' HADOOP_STREAMING='/usr/hdp/2.5.0.0-1245/hadoop-map reduce/hadoop-streaming.jar' -- INSERT --</pre>	<p>開啟 <code>usr/lib64/R/etc</code> 資料夾中的 <code>Renviron</code> 檔案</p> <p>開啟文件後按下 <code>insert</code> 鍵即可進入編輯模式，用方向鍵移至文件最末行，指定 <code>HADOOP_CMD</code> 與 <code>HADOOP_STREAMING</code> 的位置：</p> <p><code>HADOOP_CMD='/usr/bin/hadoop'</code></p> <p><code>HADOOP_STREAMING='/usr/hdp/2.5.0.0-1245/hadoop-mapreduce/hadoop-streaming.jar'</code></p> <p>其中，<code>HADOOP_STREAMING</code> 所指定的路徑為步驟(1)中所查到的第二個檔案路徑。輸入完成後按下 <code>Esc</code> 鍵並輸入：<code>wq</code> 即可儲存並退出文件。</p>
<pre>2 sudo vi /etc/profile</pre> <pre>unset i unset -f pathmunge</pre> <pre>export HADOOP_CMD='/usr/bin/Hadoop' export HADOOP_STREAMING='/usr/hdp/2.5.0.0-1245/hadoop-mapreduce/hadoop-streaming.jar' -- INSERT --</pre>	<p>開啟 <code>etc</code> 資料夾中的 <code>profile</code> 檔案</p> <p>開啟文件後按下 <code>insert</code> 鍵即可進入編輯模式，用方向鍵移至文件最末行，輸出 <code>HADOOP_CMD</code> 與 <code>HADOOP_STREAMING</code> 的位置：</p> <p><code>export HADOOP_CMD='/usr/bin/Hadoop'</code></p> <p><code>export HADOOP_STREAMING='/usr/hdp/2.5.0.0-1245/hadoop-mapreduce/hadoop-streaming.jar'</code></p> <p>其中，<code>HADOOP_STREAMING</code> 所指定的路徑為步驟(1)中所查到的第二個檔案路徑。輸入完成後按下 <code>Esc</code> 鍵並輸入：<code>wq</code> 即可儲存並退出文件。</p>
<pre>3 sudo vi /etc/init/rstudio-server.conf</pre> <pre># (note that embedding a script t actions are supported) # start on runlevel [345] stop on runlevel [!345] expect fork</pre>	<p>開啟 <code>etc/init</code> 資料夾中的 <code>rstudio-server.conf</code> 檔案</p> <p>原程式碼為</p> <p><code>start on runlevel [2345]</code>與 <code>stop on runlevel [!2345]</code></p> <p>將其[]中的數字 2 刪除，修改為</p> <p><code>start on runlevel [345]</code>與 <code>stop on runlevel [!345]</code></p>

完成以上檔案之編輯後，即可開始安裝 RHadoop 相關套件。

安裝 rmr2 套件與其相關套件

安裝 rmr2 的相關套件之前，需先安裝以下的相關套件，rmr2 套件才可正常運作。要注意的是，rmr2 無法從 R 語言中的 library 直接載入，必需先從 linux server 環境中下載套件再安裝至 R。

Linux 與 R 程式碼	說明
1 cmd R >install.packages(c("Rcpp","RJSONIO","bitops","digest","functional","iterators","itertools","reshape2","stringr","plyr","caTools"))	在 HDP 環境中(Putty 介面)進入 R 的編輯介面，並安裝 rmr2 的相關套件。
2 wget -O /tmp/rmr2_3.3.0.tar.gz https://github.com/RevolutionAnalytics/rmr2/raw/master/build/rmr2_3.3.0.tar.gz	下載 rmr2 套件
3 R CMD INSTALL /tmp/rmr2_3.3.0.tar.gz	在 Putty 進入 R 的編輯介面，並安裝 rmr2 套件
4 mkdir -p /var/log/hadoop/rstudio/	新增資料夾
5 chmod -R 777 /var/log/hadoop/rstudio	修改資料夾(含子資料夾)之讀取權限。
6 R	從 linux 進入 R 語言的編輯畫面
➤ Sys.setenv(HADOOP_STREAMING="/usr/hdp/2.5.0.0-1245/hadoop-Mapreduce/hadoop-streaming.jar")	設定 HADOOP_STREAMING 的路徑，如步驟 3.2
➤ Sys.setenv(HADOOP_CMD="/usr/bin/hadoop")	設定 HADOOP_CMD 的路徑，如步驟 3.2
➤ library(Rcpp) ➤ library(RJSONIO) ➤ library(bitops) ➤ library(digest) ➤ library(functional) ➤ library(iterators) ➤ library(itertools) ➤ library(reshape2) ➤ library(stringr) ➤ library(plyr) ➤ library(caTools) ➤ library(rmr2)	載入所有安裝好之套件，確認安裝成功
➤ q()	退出 R 的編輯介面，回到 linux 環境。

(4)安裝 rhive 與其相關套件

在 HDP 環境中安裝 rhive，未來可從 Rstudio 中讀取 HDP 之 hive 資料表，讓使用者在 Rstudio 中從 HDP 資料庫篩選資料，並進行統計分析。

Linux 與 R 程式碼	說明
1 cmd R install.packages(c("DBI","rJava","RJDBC","Rserve"))	在 HDP 環境中(Putty 介面)進入 R 的編輯介面，並安裝 rhive 的相關套件。
2 wget -O /tmp/ RHive_2.0-0.2.tar.gz https://cran.r-project.org/src/contrib/Archive/RHive/RHive_2.0-0.2.tar.gz	下載 rhive
3 R CMD INSTALL /tmp/ RHive_2.0-0.2.tar.gz	在 Putty 進入 R 的編輯介面，並安裝 rhive 套件
4 R	從 linux 進入 R 語言的編輯畫面
<ul style="list-style-type: none">➤ library(DBI)➤ library(rJava)➤ library(RJDBC)➤ library(Rserve)➤ library(RHive)	載入安裝的套件，並確認是否正常
<ul style="list-style-type: none">➤ options(java.parameters = "-Xmx8g")➤ cp = c("/usr/hdp/current/hive-client/lib/hive-jdbc.jar","/usr/hdp/current/hadoop-client/hadoop-common.jar")➤ .jinit(classpath=cp)➤ drv <- JDBC("org.apache.hive.jdbc.HiveDriver","/usr/hdp/current/hive-client/lib/hive-jdbc.jar",identifier.quote="`")	輸入程式碼，設定環境變數的路徑
<ul style="list-style-type: none">➤ conn <- dbConnect(drv,"jdbc:hive2://127.0.0.1:10000/default","maria_dev","maria_dev")	使用 maria_dev 的帳號密碼連接 Hive 的 default 資料庫
<ul style="list-style-type: none">➤ show_databases <- dbGetQuery(conn,"show databases")	列出資料庫清單，成功運行代表以上套件均已安裝成功。

以上安裝完成後，即可進行測試，下一章將使用 ETC 資料做為測試，嘗試將資料匯入 HDP 後，再由 Rstudio 讀取進行分析。

第五節、 連接埠之彙整

前幾節的安裝設定流程中，使用到許多套件，有些套件擁有自己的獨立連接埠，本節將彙整所有所需套件之連接埠，如表 3.1。

表 3.1 連接埠功能介紹

工具名稱	連接埠	功能簡介
Welcome Page	8888	HDP 的歡迎介面，提供許多套件、教學以及官方討論區的快速連結。
Ambari	8080	初始登入介面，可觀察系統運行狀態與各個程式所出現的問題，亦可連接至其他套件中。
HDFS	8080	雲端分析平台的儲存空間，採用分散式儲存方式，提供平台一個穩定的存取空間
Hive	8080	Hadoop 上的類 SQL 語法工具，提供使用者在雲端平台上操作 SQL 語法進行資料查詢。
Pig	8080	進階的計算工具，輔助 Hive 處理進階的計算與篩選。
Zeppelin	9995	網頁版的記事本，可提供許多程式語言進行編輯，亦有圖形化資料的功能。
Rstudio Server	8090	額外掛載於 HDP 上的統計分析工具，提供專業統計分析能力，讓雲端平台更加強大。
SSH Client	4200	HDP 上專用的 SSH 介面，亦可進程式碼的複製貼上，功能較 putty 陽春。

第六節、 RStudio Server 之多核心計算比較

R 語言提供了不同的套件讓使用者可以運用多核心的計算方式對資料進行分析，如 Parallel、Foreach 等套件，本研究在 RStudio Server 上進行多核心套件的計算時間的比較，比較不同資料量在不同核心數量上的運算時間差異，另外，不同方法間的效能差異也會進行比較。本研究透過以下三種函數進行運算時間的比較，所使用之完整詳細的程式碼於附錄(二)

1. Parallel 套件介紹

Parallel 套件於 R 版本 2.14.0 起，被設定為預設的套件，其中包含了 snow、multicores 等套件，使用者可透過函數如 mclapply 及 parLapply 將程式碼包覆，要求軟體依照指定的核心數量進行運算，其中 mclapply 僅支援在 linux 系統上使用，parLapply 則額外支援 Windows、Mac 等作業系統。

mclapply 基本語法：

```
mclapply( FUN, ..., mc.cores)
```

選取四核心環境與執行 K-Means 集群分析之範例：

```
mclapply(FUN = function(x) { ETC.cluster(etc1, 3)},mc.cores =4)
```

其中 FUN 是自行設計的函數 ETC.cluster，mc.cores 表示選取的核心數量為四核心，主要細節如下：

- FUN

進行運算所設計的函數，本研究中所設計的函數是以 K-Means 集群分析為主的函數。函數設計程式碼範例如下(名稱：ETC.cluster)：

程式碼	說明
ETC.cluster <- function(data, N) { set.seed(278613) K3=kmeans(x=data,centers=N) return(c('iter','totss'))}	1. 定義函數名稱與輸入變數 2. 設定集群種子(seed) 3. 使用 K-Means 函數進行分群，資料集為 data，共分為 N 群 4. 回傳 K-Means 之分析結果

- mc.cores：設定需要使用的核心數量，可透過數字設定或 getOption 函數抓取系統所具備的核心數量。

parLapply 基本語法：

```
parLapply(cl, FUN)
```

選取四核心環境與執行 K-Means 集群分析之範例：

```
cl<-4
```

```
parLapply(cl, function(x) {ETC.cluster(etc1, 3) })
```

其中 FUN 是自行設計的函數 ETC.cluster，cl 表示選取的核心數量為四核心，主要細節如下：

- cl：使用的核心數量，透過 detectCores 或 makeCluster 函數設定核心數量。
- FUN：

進行運算所設計的函數與 mclapply 相同，與 mclapply 不同的是 parLapply 需先用 clusterExport 把計算所需的資料和函數分配給各個計算組。

2. Foreach 套件介紹

Foreach 套件是一個支援多核心運算的套件，亦可支援叢集式電腦進行多 node 的計算，Foreach 跟 For 迴圈類似，基本運算上沒有太大差異。

Foreach 基本語法：

```
foreach(..., .combine) %dopar% FUN
```

選取四核心環境與執行 K-Means 集群分析之範例：

```
foreach(i=1:4, .combine='rbind') %dopar% ETC.cluster(etc1, 3)
```

其中 %dopar% 後面是自行設計的函數 ETC.cluster，%dopar% 前面代表讀取後方函數的方式與環境的定義，主要細節如下：

- foreach(..., .combine)

讀取 %dopar% 後面的函數的方式與環境的規格定義，combine 為結果的呈現方式指定。

- %dopar%；連接後方函數與前方讀取方式的語法，加快系統的運算效能。
- FUN
進行運算所設計的函數，亦與 mclapply 相同，本研究中所設計的函數是以 K-Means 集群分析為主的函數。

第七節、 K-Means(K-平均演算法)

K-Means 演算法是一個將各個資料點劃分到 k 個聚類中，使每個資料點與屬於他的平均值距離最接近，此演算法源於訊號處理的一種像量化方法，現在多用於資料探勘中的聚類分析。演算法步驟如下：

1. 隨機指派群集中心

從群組中隨機選出 k 個資料點作為初始種子。

2. 產生初始群集

計算每一筆資料到各個種子的距離，選擇距離最近的種子聚類產生初始集合。

3. 產生新的質量中心

根據各集合邊界內的資料點計算新的質量中心，並取代先前的隨機種子作為該群的中心。

4. 變動群集邊界

重新計算每一筆資料到新的種子之距離，重新分配所屬及群。

5. 持續反覆 3, 4 的步驟，一直執行到群集成員不再變動為止

本研究使用 K-Means 演算法針對旅行時間資料進行集群分析，將資料分為 3 群，並探討集群間的差異性與集群內的特性。透過分類結果找出適當的建議，提出有益的行駛建議。

第八節、 資料描述

ETC 的資料主要存放於交通部臺灣區國道高速公路局「交通資料庫」，網址為 <http://tisvcloud.freeway.gov.tw/>，本研究主要利用資料庫中的各旅次路徑原始資料（M06A）作為分析目標，資料以每小時一個資料檔的方式存放於資料庫。

8.1. 資料檔案下載與合併

本研究利用 Anaconda 將 2015 年與 2016 年之資料下載，下載方式與程式碼詳細內容請參考 P.97 附錄(一)。下載後須透過 Windows 的命令提示字元輸入程式碼：

```
copy *.csv 年分_M06A.csv
```

將 csv 檔合併(如圖 3.39)，之後再匯入 HDP 中，檔案名稱分別為 2015_M06A.csv 與 2016_M06A.csv。

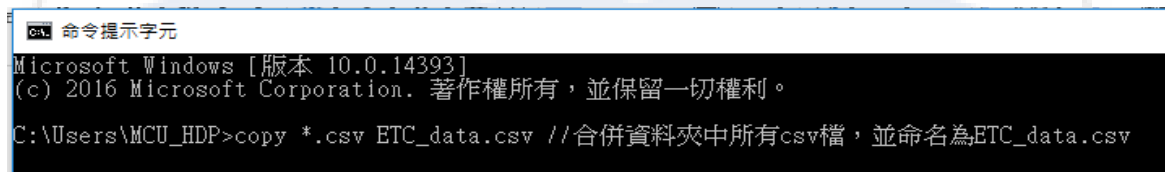


圖 3.39 命令提示字元合併 CSV 檔

8.2. 將資料儲存至 HDP 之 HDFS

透過第三章研究方法 3.3.2 章節中(p.22)HDP 之 HDFS 存取資料，匯入資料至 HDFS 的步驟將資料匯入 (如圖 3.40)，以便後續使用 HIVE 及 RStudio Server 使用，資料存放位置為/user/maria_dev/etc。

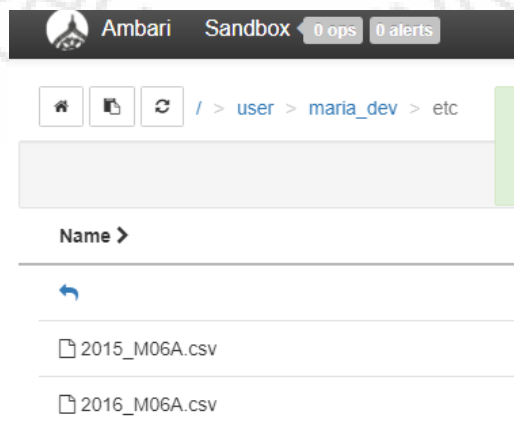


圖 3.40 匯入 ETC 資料至 HDFS

8.3. 資料內容

原始資料變數共有 7 個，中英文名稱及說明如表 3.2，其中，高速公路的 ETC 測站安裝在兩兩交流道之間，如石碇-南港之間，國道五號南下路段中，南港至蘇澳出口間共有 6 個 ETC 測站。本研究主要將各測站資料從詳細資料中切割出來，進行旅行時間的計算並編號，如表 3.3，透過計算之旅行時間進行集群分析，探討一群內的特性與集群間的差異。

表 3.2 資料變數說明

中文名稱	英文名稱	說明
車輛種類	cartype	車種資料(31：小客車；32：小貨車；41：大客車；42：大貨車；5：聯結車)
起始時間	starttime	通過第一個 ETC 測站的時間
起始測站	startgate	通過的第一個 ETC 測站編號
終點時間	endtime	通過最後一個 ETC 測站的時間
終點測站	endgate	通過的最後一個 ETC 測站編號
行駛距離	length	行駛距離(km)
資料勘誤	error	資料紀錄是否有誤
詳細資料	middle	通過的 ETC 測站編號以及時間

表 3.3 ETC 測站說明與代號

編號	說明
A1	南港-石碇之間之 ETC 測站
A2	石碇-坪林之間之 ETC 測站
A3	坪林-頭城之間之 ETC 測站
A4	頭城-宜蘭之間之 ETC 測站
A5	宜蘭-羅東之間之 ETC 測站
A6	羅東-蘇澳之間之 ETC 測站

彙整過後之資料透過 SQL 資料庫及 R 語言進行篩選與變數計算，計算出星期、尖峰時段、平假日等變數。另外，由於大客車、大貨車、聯結車的資料量較少，因此將車種資料進行合併，整理過後之資料變數如表 3.4。

表 3.4 整理過後之變數說明

變數名稱	說明
車輛種類	包含小客車(31)、小貨車(32)、大型車(41、42、5)
上路時間	車輛起始之上路時間
起始測站	通過的第一個 ETC 測站編號
終點測站	通過的最後一個 ETC 測站編號
星期資料	星期資料(星期天-星期六)
上班時段	07:00-9:59 & 16:00-18:59 為尖峰時段，其他時間為離峰時段
平假日	定義星期六與星期日為假日，其他星期為平日
時間區段	將一天分為上午(6-12)、下午(12-18)、晚上(18-24)、凌晨(0-6)
A21	石碇-坪林(A2)與南港-石碇(A1)間之旅行時間(秒)
A32	坪林-頭城(A3)與石碇-坪林(A2)間之旅行時間(秒)
A43	頭城-宜蘭(A4)與坪林-頭城(A3)間之旅行時間(秒)
A54	宜蘭-羅東(A5)與頭城-宜蘭(A4)間之旅行時間(秒)
A65	羅東-蘇澳(A6)與宜蘭-羅東(A5)間之旅行時間(秒)

由於 2015/01 的資料不完整，因此本研究使用 2015/02 至 2016/12 共 23 個月之之資料進行分析，針對國道五號南下路段進行研究，目標旅行資料的則選取通過路段數量不小於三段(通過四個測站以上)的車輛作為分析目標，主要分為六個旅行資料如表 3.5，經整理與篩選後，共有 12,953,088 筆資料。第四章將針對資料進行分析與探討，並針對多核心計算的部分做效能比較。

表 3.5 選用之旅行資料整理

路段編號	路徑說明	資料量	比例
ETC_A1	南港-石碇-坪林-頭城-宜蘭	5,101,445	39.4%
ETC_A2	南港-石碇-坪林-頭城-宜蘭-羅東	4,476,414	34.6%
ETC_A3	南港-石碇-坪林-頭城-宜蘭-羅東-蘇澳	2,056,563	15.9%
ETC_B1	石碇-坪林-頭城-宜蘭-羅東	656,133	5.1%
ETC_B2	石碇-坪林-頭城-宜蘭-羅東-蘇澳	279,576	2.2%
ETC_C1	坪林-頭城-宜蘭-羅東-蘇澳	26,328	0.2%
總計		12,953,088	100%

第四章、實證分析

本研究將交通資料透過集群分析 K-Means 方法分為三群，比較集群間的差異與及群內的特性，針對行駛時間、車種等變數進行探討；另外，將在 RStudio Server 上進行多核心計算的運算時間比較，比較不同核心數量在資料分析上的運算效能。

第一節、基本敘述統計

表 4.1 各車種與月份之資料筆數(欄比例；列比例)

年 月	小客車	小貨車	大型車	月份總計
1 月	759626 (72.1% ; 8.3%)	246174 (23.4% ; 8.5%)	48194 (4.6% ; 8.1%)	1053994 (100% ; 8.4%)
2 月	848175 (74.1% ; 9.3%)	250646 (21.9% ; 8.6%)	46384 (4.1% ; 7.8%)	1145205 (100% ; 9.1%)
3 月	699490 (72.0% ; 7.7%)	224375 (23.1% ; 7.7%)	47599 (4.9% ; 8.0%)	971464 (100% ; 7.7%)
4 月	782822 (72.2% ; 8.6%)	247614 (22.8% ; 8.5%)	53328 (4.9% ; 9.0%)	1083764 (100% ; 8.6%)
5 月	788475 (72.1% ; 8.7%)	249603 (22.8% ; 8.6%)	55844 (5.1% ; 9.4%)	1093922 (100% ; 8.7%)
6 月	785739 (72.1% ; 8.6%)	253283 (23.3% ; 8.7%)	50314 (4.6% ; 8.5%)	1089336 (100% ; 8.6%)
7 月	869887 (72.4% ; 9.6%)	277046 (23.1% ; 9.5%)	54320 (4.5% ; 9.2%)	1201253 (100% ; 9.5%)
8 月	823534 (72.3% ; 9.0%)	265009 (23.3% ; 9.1%)	50998 (4.5% ; 8.6%)	1139541 (100% ; 9.0%)
9 月	746263 (72.0% ; 8.2%)	241919 (23.3% ; 8.3%)	48426 (4.7% ; 8.2%)	1036608 (100% ; 8.2%)
10 月	789266 (71.7% ; 8.7%)	256367 (23.3% ; 8.8%)	55489 (5.0% ; 9.4%)	1101122 (100% ; 8.7%)
11 月	766974 (71.8% ; 8.4%)	248077 (23.2% ; 8.5%)	52827 (4.9% ; 8.9%)	1067878 (100% ; 8.5%)
12 月	821492 (72.1% ; 9.0%)	264800 (23.2% ; 9.1%)	53077 (4.7% ; 9.0%)	1139369 (100% ; 9.0%)
車種總計	9101930 (72.3% ; 100%)	2901826 (23.0% ; 100%)	592703 (4.7% ; 100%)	12596459 (100% ; 100%)

(Level：綠，低於 8%；黃，8%至 8.9%；紅，9%以上)

表 4.1 為各車種與月份資料筆數交叉表，一月之資料僅包含 2016 年，此表中將其加倍以便於做比較，將資料依照月份比例做分級，小客車車流量較高的月份受到 2 月寒假及春節返鄉、7、8 月寒暑假出遊以及 12 月底跨年(如圖 4.1)之車潮的影響，這幾個月份之車流量比例皆高於 9%；車流量最少的月份為 3 月，其餘月份之車流量介在 8% 至 9% 之間。小貨車的狀態與小客車類似，唯獨 2 月之車潮沒有明顯增加。大型車車流量較高的月份為 4、5、10、12 月，比例皆超過 9%；2 月反而較少，僅占了 7.8%；其餘月分則落在 8% 至 9% 之間。整體來看，由於小客車在各個月份的比例都超過 70%，受小客車的影響，整體各月的車量輛狀態與小客車相似。



圖 4.1 12 月車流量時間數列圖(小客車)

表 4.2 為各車種於平假日之平均車流量，假日車流量中可以看出小客車的車流量佔 75.93%，其次為小貨車 19.15%，大型車佔了 4.91%；平日車流量中，小客車佔了 70.54%，小貨車 24.85%，大型車僅 4.61%。小客車與大型車於假日的車流量較平日高，分別為 55.2%與 55.4%，平假日的比值分別為 0.797:1 與 0.804:1。不論是自行開車或搭乘客運，假日出遊前往宜蘭地區的人潮較平日多，小貨車的部分則是平日的車流量較假日高，平日平均車流量佔了 52.7%，平假日比值為 1.112:1。

表 4.2 各車種於平假日之平均車流量(欄比例；列比例)

平假日 車種	平日	假日	平假日 比值	列總計
小客車	1211930 (70.54% ; 44.0%)	1521141 (75.93% ; 55.2%)	0.797 : 1	2733071 (73.05% ; 100%)
小貨車	426896 (24.85% ; 52.7%)	383673 (19.15% ; 47.3%)	1.112 : 1	810569 (22.70% ; 100%)
大型車	79166 (4.61% ; 44.6%)	98436 (4.91% ; 55.4%)	0.804 : 1	177602 (5.97% ; 100%)
欄總計	1717992 (100% ; 45.5%)	2003249 (100% ; 53.1%)	0.858 : 1	3721241 (100% ; 100%)

(單位：輛)

表 4.3 至表 4.7 為各路段於各個星期與時段之平均旅行時間及車速，受到上班車潮的影響，南港-石碇-坪林與石碇-坪林-頭城兩個路段在星期六上午及下午時段的車速較慢，上午的車速小於 50 公里，下午時速約 50 公里，皆低於各時段平均時速許多，各個時段於晚間與清晨的行駛車速都很順暢，南港-石碇-坪林路段平均都有 70 公里以上、石碇-坪林-頭城與坪林-頭城-宜蘭都有 80 公里以上，頭城-宜蘭-羅東與宜蘭-羅東-蘇澳晚間平均時速都有 90 公里，宜蘭-羅東-蘇澳凌晨的平均時速有 84 公里。各星期平均時速中，受到上午及下午時段車速較低的影響，星期六前兩個路段的平均時速也較其他星期低。顯示週六上午及下午的出遊車潮較多，車速較慢，用路人應盡量避免這個時段出發前往宜蘭地區。

表 4.3 各星期與時段之平均旅行時間及車速(南港-石碇-坪林)

星期 \ 時段		上午	下午	晚間	凌晨	星期平均
Mon	旅行時間(秒)	351.62	300.93	270.51	270.32	313.28
	時速(km)	56.31	65.8	73.2	73.25	63.2
Tues	旅行時間(秒)	320.41	294.52	269.94	269.37	298.13
	時速(km)	61.8	67.23	73.35	73.5	66.41
Wed	旅行時間(秒)	326.73	294.02	275.59	277.48	301.8
	時速(km)	60.6	67.34	71.85	71.36	65.61
Thurs	旅行時間(秒)	339.35	303.16	282.77	283.87	311.27
	時速(km)	58.35	65.31	70.02	69.75	63.61
Fri	旅行時間(秒)	352.53	306.93	290.52	281.32	314.01
	時速(km)	56.17	64.51	68.15	70.38	63.06
Sat	旅行時間(秒)	437.18	383.74	279.46	290.41	368.9
	時速(km)	45.29	51.6	70.85	68.18	53.67
Sun	旅行時間(秒)	316.63	304.91	274.78	273.72	299.74
	時速(km)	62.53	64.94	72.06	72.34	66.06
時段平均	旅行時間(秒)	351.24	315.38	279.96	279.31	317.44
	時速(km)	56.37	62.78	70.73	70.89	62.37

表 4.4 各星期與時段之平均旅行時間及車速(石碇-坪林-頭城)

星期 \ 時段		上午	下午	晚間	凌晨	星期平均
Mon	旅行時間(秒)	1206.1	1079.79	988.92	980.18	1107.72
	時速(km)	69.25	77.35	84.46	85.21	75.4
Tues	旅行時間(秒)	1147.93	1058.25	988.22	984.02	1075.82
	時速(km)	72.76	78.92	84.52	84.88	77.63
Wed	旅行時間(秒)	1159.8	1053.36	1021.07	995.63	1085.46
	時速(km)	72.01	79.29	81.8	83.89	76.94
Thurs	旅行時間(秒)	1190.43	1103	1046.04	1009.58	1119.16
	時速(km)	70.16	75.72	79.84	82.73	74.63
Fri	旅行時間(秒)	1211.81	1144.68	1059.7	1029.04	1132.49
	時速(km)	68.92	72.96	78.81	81.16	73.75
Sat	旅行時間(秒)	1678.06	1549.26	1011.06	1027.46	1427.23
	時速(km)	49.77	53.91	82.61	81.29	58.52
Sun	旅行時間(秒)	1129.15	1085.08	987.29	985.06	1071.26
	時速(km)	73.97	76.97	84.6	84.79	77.96
時段平均	旅行時間(秒)	1258.21	1170.77	1005.65	1019.77	1157.38
	時速(km)	66.38	71.34	83.05	81.9	72.16

表 4.5 各星期與時段之平均旅行時間及車速(坪林-頭城-宜蘭)

星期 \ 時段		上午	下午	晚間	凌晨	星期平均
Mon	旅行時間(秒)	96.32	92.98	90.26	92.4	93.79
	時速(km)	82.22	85.18	87.74	85.71	84.45
Tues	旅行時間(秒)	96.39	93.38	90.51	92.52	93.96
	時速(km)	82.16	84.82	87.5	85.6	84.29
Wed	旅行時間(秒)	95.23	93.33	91.08	92.29	93.56
	時速(km)	83.16	84.86	86.96	85.81	84.66
Thurs	旅行時間(秒)	95.98	93.69	91.97	92.43	94.14
	時速(km)	82.52	84.53	86.12	85.69	84.13
Fri	旅行時間(秒)	96.49	93.77	92.5	92.88	94.17
	時速(km)	82.08	84.46	85.62	85.27	84.1
Sat	旅行時間(秒)	99.57	95.49	91.63	93.98	95.97
	時速(km)	79.54	82.94	86.44	84.28	82.52
Sun	旅行時間(秒)	95.63	91.09	90.46	93.15	92.78
	時速(km)	82.82	86.95	87.55	85.02	85.36
時段平均	旅行時間(秒)	96.6	93.47	93.02	91.37	94.13
	時速(km)	81.98	84.73	85.14	86.68	84.14

表 4.6 各星期與時段之平均旅行時間(頭城-宜蘭-羅東)

星期 \ 時段		上午	下午	晚間	凌晨	星期平均
Mon	旅行時間(秒)	494.44	490.34	491.81	488.24	492.21
	時速(km)	94.65	95.44	95.16	95.86	95.08
Tues	旅行時間(秒)	492.69	490.27	492.08	490.09	491.62
	時速(km)	94.99	95.46	95.11	95.49	95.2
Wed	旅行時間(秒)	492.93	490.44	493.01	491.37	492.02
	時速(km)	94.94	95.42	94.93	95.24	95.12
Thurs	旅行時間(秒)	494.54	490.92	493.65	491.17	492.96
	時速(km)	94.63	95.33	94.8	95.28	94.94
Fri	旅行時間(秒)	494.43	491.68	497.04	490.59	494.1
	時速(km)	94.65	95.18	94.16	95.39	94.72
Sat	旅行時間(秒)	496.42	496.46	493.92	486.95	494.81
	時速(km)	94.27	94.27	94.75	96.11	94.58
Sun	旅行時間(秒)	490.09	488.2	493.34	485.27	489.71
	時速(km)	95.49	95.86	94.86	96.44	95.57
時段平均	旅行時間(秒)	493.75	491.44	488.57	494.04	492.65
	時速(km)	94.79	95.23	95.79	94.73	95

表 4.7 各星期與時段之平均旅行時間(宜蘭-羅東-蘇澳)

星期 \ 時段		上午	下午	晚間	凌晨	星期平均
Mon	旅行時間(秒)	231.05	238.03	234.04	217.42	231.69
	時速(km)	85.7	83.18	84.6	91.07	85.46
Tues	旅行時間(秒)	229.93	237.55	236.97	216.41	231.47
	時速(km)	86.11	83.35	83.55	91.49	85.54
Wed	旅行時間(秒)	230.35	240.45	238.22	217.79	232.71
	時速(km)	85.96	82.34	83.12	90.91	85.08
Thurs	旅行時間(秒)	230.1	237.16	231.99	217.68	230.79
	時速(km)	86.05	83.49	85.35	90.96	85.79
Fri	旅行時間(秒)	230.1	235.46	231.52	215.58	230.09
	時速(km)	86.05	84.09	85.52	91.85	86.05
Sat	旅行時間(秒)	233.38	239.7	236.63	215.57	231.56
	時速(km)	84.84	82.6	83.67	91.85	85.51
Sun	旅行時間(秒)	229.22	235.72	236.88	217.69	230.73
	時速(km)	86.38	84	83.59	90.95	85.82
時段平均	旅行時間(秒)	230.69	237.65	216.6	234.62	231.22
	時速(km)	85.83	83.32	91.41	84.39	85.63

表 4.8 為南港-石碇-坪林平假日與車種間的平均旅行時間及車速表，此路段的車速較其他路段相比偏低，平日平均車速僅 58.7 公里，假日為 62.24 公里。不同車種間，小客車與小貨車的平均時速大約 62 公里，大型車的平均車速為 60.7 公里，較其他車種低。

表 4.8 平假日與車種間的平均旅行時間(南港-石碇-坪林)

平假日 車種		平日	假日	平假日 總平均
小客車	旅行時間(秒)	337.22	307.84	317.64
	時速(km)	58.71	64.32	62.33
小貨車	旅行時間(秒)	333.19	308.42	314.94
	時速(km)	59.43	64.2	62.87
大型車	旅行時間(秒)	354.68	312.57	326.2
	時速(km)	55.82	63.35	60.7
車種總平均	旅行時間(秒)	337.32	308.21	317.44
	時速(km)	58.7	64.24	62.37

表 4.9 為石碇-坪林-頭城平假日與車種間的平均旅行時間及車速表，平日平均車速 65.91 公里，假日為 75.5 公里，顯示平日上班車潮對整體車速的影響較假日車潮高。不同車種間的车速在此路段並沒有差異，平均 72 公里左右。

表 4.9 平假日與車種間的平均旅行時間(石碇-坪林-頭城)

平假日 車種		平日	假日	平假日 總平均
小客車	旅行時間(秒)	1271.4	1105.44	1160.86
	時速(km)	65.69	75.55	71.95
小貨車	旅行時間(秒)	1261.25	1105.8	1146.84
	時速(km)	66.22	75.53	72.83
大型車	旅行時間(秒)	1225.02	1121.06	1155.59
	時速(km)	68.18	74.5	72.27
車種總平均	旅行時間(秒)	1267.18	1106.25	1157.38
	時速(km)	65.91	75.5	72.16

表 4.10 為坪林-頭城-宜蘭平假日與車種間的平均旅行時間及車速表，平日平均車速 83.78 公里，假日為 84.31 公里，從此路段開始，行駛時間較不受上班車潮的影響，不同車種間，小客車與小貨車的平均時速分別為 83.83 與 82.25 公里，大型車的平均車速為 78.39 公里，花費的旅行時間較其他車種高。

表 4.10 平假日與車種間的平均旅行時間(坪林-頭城-宜蘭)

平假日		平日	假日	平假日 總平均
車種				
小客車	旅行時間(秒)	93.93	93.08	93.36
	時速(km)	84.32	85.09	84.83
小貨車	旅行時間(秒)	95.02	95.18	95.13
	時速(km)	83.35	83.21	83.25
大型車	旅行時間(秒)	102.1	100.5	101.03
	時速(km)	77.57	78.81	78.39
車種總平均	旅行時間(秒)	94.54	93.94	94.13
	時速(km)	83.78	84.31	84.14

表 4.11 為頭城-宜蘭-羅東平假日與車種間的平均旅行時間及車速表，平日平均車速 95.01 公里，假日為 94.99 公里，行駛時間較不受上班車潮的影響，不同車種間，小客車與小貨車的平均時速分別為 95.71 與 94.11 公里，大型車的平均車速為 89.34 公里，花費的旅行時間較其他車種高。

表 4.11 平假日與車種間的平均旅行時間(頭城-宜蘭-羅東)

平假日		平日	假日	平假日 總平均
車種				
小客車	旅行時間(秒)	489.58	488.64	488.96
	時速(km)	95.59	95.78	95.71
小貨車	旅行時間(秒)	495.5	497.95	497.29
	時速(km)	94.45	93.99	94.11
大型車	旅行時間(秒)	524.67	523.38	523.82
	時速(km)	89.2	89.42	89.34
車種總平均	旅行時間(秒)	492.56	492.7	492.6514
	時速(km)	95.01	94.99	95

表 4.12 為宜蘭-羅東-蘇澳平假日與車種間的平均旅行時間及車速表，平日平均車速 85.64 公里，假日為 85.63 公里，行駛時間較不受上班車潮的影響，不同車種間，小客車與小貨車的平均時速分別為 86.01 與 85.79 公里，大型車的平均車速為 74.5 公里，花費的旅行時間較其他車種高。

表 4.12 平假日與車種間的平均旅行時間(宜蘭-羅東-蘇澳)

平假日		平日	假日	平假日 總平均
車種	旅行時間(秒)	230.43	230.1	230.21
	時速(km)	85.93	86.05	86.01
小客車	旅行時間(秒)	228.67	231.62	230.8
	時速(km)	86.59	85.48	85.79
小貨車	旅行時間(秒)	274.17	261.12	265.78
	時速(km)	72.22	75.83	74.5
大型車	旅行時間(秒)	231.2	231.24	231.22
	時速(km)	85.64	85.63	85.63
車種總平均				

整體來看，前兩路段受到上下班時段車潮的影響，各路段在平日的平均車速大都低於假日車速，而大型車的平均車速大多比小客車及小貨車低。用路人應注意南港-頭城兩個路段之上路時間，避開易壅塞路段。

第二節、 集群分析

旅行時間的集群分析中，透過 K-Means 方法將六種旅行資料之旅行時間資料分為三群，針對不同集群間的差異以及集群內的特性作探討。

I. 南港-宜蘭(ETC_KA1)

表 4.13 為分群結果與平均旅行時間(時速)之表格，分群後發現資料多集中在集群 2，佔了 89.9%，集群 1、3 的比例分別為 7.8%與 2.3%。從平均旅行時間來看，集群 1 在南港-石碇-坪林與石碇-坪林-頭城路段的旅行時間偏高，行駛時速分別為 42.53 公里與 37.31 公里；集群 2 則是全線都順暢的狀態，平均時速有達到 70 以上，坪林-頭城-宜蘭路段更高達 85.57 公里。集群 3 在 A21 路段非常壅塞，時速僅 13.57 公里，但過壅塞路段後，A32、A43 均順暢，時速分別為 70.7 公里與 81.48 公里。

表 4.13 ETC_KA1 分群結果與平均旅行時間(時速)

集群		1	2	3	總計
資料筆數		395,769	4,587,453	118,223	5,101,445
百分比		7.8%	89.9%	2.3%	100%
南港-石碇-坪林	旅行時間(秒)	465.5	267.18	1459.32	310.2
	時速(km)	42.53	74.11	13.57	63.83
石碇-坪林-頭城	旅行時間(秒)	2232.3	1049.16	1181.27	1144.01
	時速(km)	37.41	79.61	70.7	73.01
坪林-頭城-宜蘭	旅行時間(秒)	99.05	92.56	97.2	93.17
	時速(km)	79.96	85.57	81.48	85

表 4.14 為各集群於不同星期之資料比例，集群 1 的資料有 61%集中在星期六，比例最少的星期二僅佔了 3.7%；集群 2、3 較為平均分布，但集群 2 的星期五資料比例較其他高，比例為 17.3%，其次為星期日的 15.2%，其他星期的資料大都佔 13%；集群 3 在星期五、六兩天的資料也較高，比例分別為 17.6%與 17.4%。

表 4.14 ETC_KA1 各集群於不同星期之資料比例

星期 \ 集群	1	2	3
Mon	7.3%	13.5%	12.9%
Tues	3.7%	13.3%	11.8%
Wed	4.6%	13.8%	12.4%
Thurs	8.0%	13.8%	14.1%
Fri	11.2%	17.3%	17.6%
Sat	61.0%	13.0%	17.4%
Sun	4.2%	15.2%	13.9%
總計	100%	100%	100%

表 4.15 為各集群於不同時段之資料比例，從表格中可以看出集群 1 之資料多集中在上午及下午，包含了 95% 以上的資料。集群 2 之資料較平均分布於上午、下午及晚間。集群 3 之資料也是較集中於上午及下午，分別佔了 45.4% 與 35.6%，再來為晚間之資料佔了 14.5%。

表 4.15 ETC_KA1 各集群於不同時段之資料比例

時段 \ 集群	1	2	3
上午	55.2%	33.0%	45.4%
下午	40.7%	33.9%	35.6%
凌晨	0.9%	6.3%	4.2%
晚間	3.2%	26.8%	14.8%
總計	100%	100%	100%

表 4.16 為集群 1 分平假日、車種、時段之平均旅行時間及車速，發現大型車在平日凌晨時段的 A21、A32 路段以及假日晚上的 A32 路段花費了較多的旅行時間，推測此段多為速限較低之大型車行駛時段，用路人於平日凌晨通行南港-頭城兩路段、假日晚間行經石碇-坪林-頭城路段時應注意與大車之間距，留意自身安全。南港-石碇-坪林路段多為壅塞狀態但在假日晚間時段的行駛時間較短，時速平均在 70 公里以上，建議可以挑選假日晚間行經此路段，避開壅塞車多的路況。又因為集群 1 有 61% 的資料集中於星期六，且有 90% 之資料集中在上午及下午，將集群 1 命名為假日白天出遊之車潮。

表 4.16(集群 1)假日白天出遊之車潮

平假日	車種	時段	南港-石碇-坪林		石碇-坪林-頭城		坪林-頭城-宜蘭	
			旅行時間 (秒)	時速 (km)	旅行時間 (秒)	時速 (km)	旅行時間 (秒)	時速 (km)
平日	大型車	上午	473.58	41.81	2133.87	39.14	101.94	77.69
		下午	364.31	54.35	2325.44	35.92	99.04	79.97
		凌晨	545.07	36.33	3626.89	23.03	103.41	76.59
		晚間	414.01	47.82	2313.34	36.1	102.98	76.91
	小客車	上午	485.36	40.79	2151.48	38.82	100.12	79.11
		下午	376.79	52.55	2487.94	33.57	97.63	81.12
		凌晨	346.49	57.14	2388.39	34.97	98.76	80.19
		晚間	449.16	44.08	2291.61	36.45	101.26	78.21
	小貨車	上午	487.38	40.63	2160.15	38.66	103.01	76.89
		下午	365.03	54.24	2634.8	31.7	100.14	79.09
		凌晨	301.69	65.63	2673.13	31.24	93.94	84.31

		晚間	439.17	45.09	2360.92	35.38	101.73	77.85
假日	大型車	上午	428.08	46.25	2087.66	40.01	107.67	73.56
		下午	403.91	49.02	2309.21	36.17	98.09	80.74
		凌晨	473.86	41.78	1998.34	41.79	98.98	80.02
		晚間	272.27	72.72	3765.47	22.18	100.19	79.05
	小客車	上午	496.96	39.84	2213.78	37.73	99.7	79.44
		下午	460.17	43.03	2195.71	38.04	96.75	81.86
		凌晨	396.56	49.93	2394.74	34.88	103.75	76.34
		晚間	274.99	72	2264.26	36.89	95.9	82.59
	小貨車	上午	489.3	40.47	2199.03	37.98	100.1	79.12
		下午	462.93	42.77	2201.5	37.94	97.35	81.36
		凌晨	345.05	57.38	2471.88	33.79	97.34	81.36
		晚間	263.89	75.03	2363.06	35.34	96.51	82.06

表 4.17 為集群 2 分平假日、車種、時段之平均旅行時間及車速，可以發現集群 2 的資料中，各路段的行車狀態都屬於順暢，南港-石碇-坪林路段的時速較其他兩路段低，上午時段的行駛時間也較其他時段低，用路人若趕時間可避開上午時段行駛，或選擇替代道路從石碇上交流道，避開車多路段。集群 2 之資料以星期五之資料為多，為此集群命名為下班後前往宜蘭地區之車潮

表 4. 17(集群 2)下班後前往宜蘭地區之車潮

平假日	車種	時段	南港-石碇-坪林		石碇-坪林-頭城		坪林-頭城-宜蘭	
			旅行時間 (秒)	時速 (km)	旅行時間 (秒)	時速 (km)	旅行時間 (秒)	時速 (km)
平日	大型車	上午	279.96	70.72	1123.03	74.37	102.12	77.56
		下午	259.53	76.29	1061.79	78.66	94.53	83.78
		凌晨	257.2	76.98	1007.98	82.86	95.33	83.08
		晚間	253.21	78.2	1037.13	80.53	93.7	84.53
	小客車	上午	283.96	69.73	1102.57	75.75	93.95	84.3
		下午	259.9	76.18	1026.19	81.39	91.72	86.35
		凌晨	249.18	79.46	969.81	86.12	90.27	87.74
		晚間	255.38	77.53	1001.44	83.4	89.61	88.38
	小貨車	上午	283.01	69.96	1097.53	76.1	96.73	81.88
		下午	261.78	75.64	1029.41	81.13	93.85	84.39
		凌晨	247.52	79.99	974.28	85.72	94.17	84.1
		晚間	256.83	77.09	1005.32	83.08	91.22	86.82

假日	大型車	上午	287.54	68.86	1186.94	70.37	103.87	76.25
		下午	273.35	72.43	1120.24	74.56	94.53	83.78
		凌晨	262.19	75.52	1035.63	80.65	99.01	79.99
		晚間	253.24	78.19	1028.78	81.18	93.75	84.48
	小客車	上午	287.53	68.86	1135.16	73.58	94.35	83.94
		下午	269.58	73.45	1070.3	78.03	91.47	86.59
		凌晨	253.53	78.1	990.04	84.36	91.02	87.01
		晚間	253.55	78.09	987.58	84.57	89.82	88.18
	小貨車	上午	288.28	68.68	1135.4	73.56	95.96	82.53
		下午	270.25	73.27	1071.62	77.94	92.3	85.81
		凌晨	252.25	78.49	991.73	84.22	93.59	84.62
		晚間	255.02	77.64	992.57	84.15	91.15	86.89

表 4.18 為集群 3 分平假日、車種、時段之平均旅行時間及車速，可以看出南港-石碇-坪林路段不論甚麼時段跟車種，都是時速非常低的壅塞狀態。石碇-坪林-頭城部分則是上午時速較低，越晚時速越高。坪林-頭城-宜蘭在清晨與上午的時速較下午晚間低，非必要的情況下可考慮避免上午時間行走南港-石碇-頭城路段，並多利用下午及晚間時段可避開車潮。由於上午時段之資料佔了集群 345% 的比例，因此集群 3 命名為上午時段上班之車潮。

表 4.18(集群 3)上午時段上班之車潮

平假日	車種	時段	南港-石碇-坪林		石碇-坪林-頭城		坪林-頭城-宜蘭	
			旅行時間 (秒)	時速 (km)	旅行時間 (秒)	時速 (km)	旅行時間 (秒)	時速 (km)
平日	大型車	上午	1376.85	14.38	1187.45	70.34	105.46	75.1
		下午	1320.18	15	1079.01	77.4	97.52	81.21
		凌晨	1497.76	13.22	1134.35	73.63	99.77	79.38
		晚間	1366.57	14.49	1102.16	75.78	106.55	74.33
	小客車	上午	1396.54	14.18	1226.02	68.12	98.13	80.71
		下午	1504.8	13.16	1053.03	79.31	93.33	84.86
		凌晨	1466.89	13.5	1022.71	81.67	93.77	84.46
		晚間	1422.56	13.92	1041.82	80.17	92.41	85.71
	小貨車	上午	1395.86	14.18	1217.76	68.58	108.25	73.16
		下午	1553.69	12.74	1056.55	79.05	97.3	81.4
		凌晨	1511.42	13.1	1028.03	81.24	94.41	83.89
		晚間	1479.83	13.38	1047.25	79.75	94.1	84.17

假日	大型車	上午	1567.99	12.63	1365.44	61.17	101.32	78.17
		下午	1359.1	14.57	1165.67	71.65	98.76	80.19
		凌晨	1452.4	13.63	1245.22	67.07	102.59	77.2
		晚間	1244.14	15.91	1035.98	80.62	99.15	79.88
	小客車	上午	1564.42	12.66	1488.32	56.12	98.27	80.59
		下午	1488.91	13.3	1263.54	66.1	93.88	84.36
		凌晨	1477.73	13.4	1081.22	77.25	94.61	83.71
		晚間	1391.83	14.23	995.42	83.9	90.13	87.87
	小貨車	上午	1570.16	12.61	1450.78	57.57	108.24	73.17
		下午	1488.71	13.3	1260.17	66.28	94.73	83.61
		凌晨	1455.45	13.6	1080.43	77.3	100.37	78.91
		晚間	1428	13.87	1007.21	82.92	93.5	84.71

表 4.19 為南港-宜蘭路徑資料之集群分析重點整理，將三個集群命名後為各個集群提出行駛建議。

表 4.19 ETC_KA1 各級群命名與用路人建議

集群	建議
1 假日白天出遊之車潮	1. 假日晚間可避免南港路段壅塞 2. 晚間行車多注意大型車輛
2 下班後前往宜蘭地區之車潮	1. 普遍順暢
3 上午時段上班之車潮	1. 南港端壅塞，可走替代道路在由石碇、坪林上交流道 2. 上午南下車流量大，非必要可考慮 12 點過後出發

II. 南港-羅東(ETC_KA2)

表 4.20 為分群結果與平均旅行時間(時速)之表格，分群後發現資料多集中在集群 3，佔了 89.5%，集群 1、3 的比例分別為 2.6%與 7.9%。從平均旅行時間來看，集群 1 在南港-石碇-坪林非常壅塞，行駛時速為 13.5 公里；集群 2 從南港到頭城兩路段時速分別為 42.33 公里與 37.79 公里，屬於車多狀態，坪林-頭城-宜蘭有好轉但仍低於平均時速 84.44 公里，時速為 79.14 公里。集群 3 於各路段都屬於順暢的狀態，除石碇-坪林-頭城路段略低於平均時速外，其他路段之車速皆大於平均。

表 4.20 ETC_KA2 分群結果與平均旅行時間(時速)

集群		1	2	3	總計
資料筆數		118011	352860	4005543	4476414
百分比		2.6%	7.9%	89.5%	100%
南港-石碇-坪林	旅行時間(秒)	1466.15	467.77	268.87	316.11
	時速(km)	13.5	42.33	73.64	62.64
石碇-坪林-頭城	旅行時間(秒)	1171.19	2209.97	1051.05	1145.57
	時速(km)	71.31	37.79	79.46	72.91
坪林-頭城-宜蘭	旅行時間(秒)	96.22	100.08	93.17	93.79
	時速(km)	82.31	79.14	85.01	84.44
頭城-宜蘭-羅東	旅行時間(秒)	501.41	501.92	492.68	493.64
	時速(km)	93.34	93.24	94.99	94.81

表 4.21 為各集群於不同星期之資料比例，集群 1、3 之資料較平均分散於各星期，其中比例最高為星期五，分別佔了 17.4%與 17.7%。集群 2 之資料以星期六佔的比例最多，比例為 60.3%，其次為星期五佔了 11.1%。

表 4.21 ETC_KA2 各集群於不同星期之資料比例

星期 \ 集群	1	2	3
Mon	12.9%	7.5%	13.3%
Tues	11.9%	3.8%	13.1%
Wed	12.2%	4.7%	13.4%
Thurs	13.2%	8.3%	13.5%
Fri	17.4%	11.1%	17.7%
Sat	17.6%	60.3%	13.5%
Sun	14.7%	4.3%	15.6%
總計	100%	100%	100%

表 4.21 為各級群於不同時段之資料比例，可以看出集群 1 之資料多集中在上午佔了 41.1%，其次為下午佔了 37.8%，晚間資料佔了 17.5%。集群 2 之資料多集中於上午及下午，佔了整體的 95% 以上。集群 3 之資料較分散於上午、下午及晚間，比例分別為 32.6%、34% 與 27.4%。各個集群中，凌晨之行使資料都只是少數。

表 4.22 ETC_KA2 各集群於不同時段之資料比例

時段 \ 集群	1	2	3
上午	41.1%	53.2%	32.6%
下午	37.8%	42.2%	34.0%
凌晨	3.7%	0.9%	6.0%
晚間	17.5%	3.7%	27.4%
總計	100%	100%	100%

表 4.23 為集群 1 分平假日、車種、時段之平均旅行時間及車速彙整表，可以看出各車種於各個時段中，南港-石碇-坪林的車速都非常低，平均時速僅 13.5 公里，過了南港-石碇-頭城路段後皆屬於順暢的狀態，由於資料分散於各星期，且集中在上午及下午，將此集群命名為白天之車潮。

表 4.23(集群 1)白天之車潮

平 假 日	車 種	時段	南港-石碇-坪林		石碇-坪林-頭城		坪林-頭城-宜蘭		頭城-宜蘭-羅東	
			旅行時間 (秒)	時速 (km)	旅行時間 (秒)	時速 (km)	旅行時間 (秒)	時速 (km)	旅行時間 (秒)	時速 (km)
平 日	大型車	上午	1378.76	14.36	1204.33	69.35	103.48	76.54	523.37	89.42
		下午	1326.99	14.92	1081.8	77.21	97.05	81.61	521.69	89.71
		凌晨	1573.32	12.58	1147.71	72.77	104.09	76.09	523.84	89.34
		晚間	1366.2	14.49	1113.06	75.04	99.26	79.79	538.89	86.85
	小客車	上午	1413.46	14.01	1222.84	68.3	98.52	80.39	501.5	93.32
		下午	1503.15	13.17	1054.28	79.22	93.89	84.36	494.21	94.7
		凌晨	1423.41	13.91	1040.32	80.28	94.56	83.76	499.16	93.76
		晚間	1415.88	13.98	1044.49	79.96	92.49	85.63	497.72	94.03
	小貨車	上午	1420.78	13.94	1217.41	68.6	98.75	80.2	509.74	91.81
		下午	1564.09	12.66	1048.86	79.63	96.32	82.23	506.87	92.33
		凌晨	1486.31	13.32	1057.71	78.96	93.08	85.08	509.05	91.94
		晚間	1460.37	13.56	1049.89	79.55	95.36	83.05	515.91	90.71

假日	大型車	上午	1584.15	12.5	1371.87	60.88	102.41	77.34	524.9	89.16
		下午	1385.01	14.3	1185.98	70.42	99.31	79.75	528.25	88.59
		凌晨	1604.37	12.34	1288.38	64.83	96.41	82.15	529.82	88.33
		晚間	1370.94	14.44	1037.65	80.49	101.78	77.81	523.18	89.45
	小客車	上午	1581.24	12.52	1485.68	56.22	98.8	80.16	495.55	94.44
		下午	1476.66	13.41	1241.68	67.26	93.03	85.13	496.04	94.35
		凌晨	1473.28	13.44	1105.44	75.55	98.57	80.35	490.96	95.32
		晚間	1400.03	14.14	993.41	84.07	92	86.09	493.89	94.76
	小貨車	上午	1576.59	12.56	1460.14	57.2	101.33	78.16	497.64	94.04
		下午	1463.5	13.53	1238.48	67.44	95.85	82.63	502.39	93.15
		凌晨	1433.01	13.82	1115.06	74.9	92.31	85.8	497.27	94.11
		晚間	1396.33	14.18	996.85	83.78	92.01	86.08	502.87	93.07

表 4.23 為集群 2 分平假日、車種、時段之平均旅行時間及車速彙整表，整體來看可以發現南港-石碇-坪林與石碇-坪林-頭城兩個路段之車速較慢，但其中可看出假日晚間的車速較快。假日凌晨之大型車在坪林-頭城-宜蘭之路段車速也較慢，時速僅 56.92 公里，於假日清晨行經此處時應注意大型車輛，確保自身行車安全。由於本資料多為星期六之旅行資料，又資料多集中於上午及下午，將此集群命名為週六白天出遊之車潮。

表 4.24(集群 2)週六白天出遊之車潮

平假日	車種	時段	南港-石碇-坪林		石碇-坪林-頭城		坪林-頭城-宜蘭		頭城-宜蘭-羅東	
			旅行時間 (秒)	時速 (km)	旅行時間 (秒)	時速 (km)	旅行時間 (秒)	時速 (km)	旅行時間 (秒)	時速 (km)
平日	大型車	上午	463.9	42.68	1996.82	41.83	100.69	78.66	530.25	88.26
		下午	345.41	57.32	2226.88	37.51	101.43	78.09	532.8	87.84
		凌晨	399.13	49.61	3037.37	27.5	120.87	65.53	569.97	82.11
		晚間	390.96	50.64	2296.09	36.37	108.15	73.23	546.53	85.63
	小客車	上午	491.65	40.27	2136.21	39.1	100.74	78.62	500.28	93.55
		下午	385.34	51.38	2436.09	34.28	97.71	81.06	501.85	93.26
		凌晨	349.35	56.68	2395.35	34.87	98.44	80.46	504.96	92.68
		晚間	466.16	42.48	2259.74	36.96	102.04	77.62	517.04	90.51
	小貨車	上午	485.36	40.79	2140.79	39.01	100.29	78.97	508.61	92.02
		下午	370.52	53.44	2520.74	33.13	101.6	77.95	509.16	91.92
		凌晨	357.9	55.32	2679.05	31.18	93.25	84.94	509.13	91.92
		晚間	456.6	43.36	2273.48	36.74	102.45	77.3	523.3	89.43

假日	大型車	上午	425.96	46.48	1993.17	41.9	102.7	77.12	527.67	88.69
		下午	397.2	49.85	2065.16	40.44	99.11	79.91	530.2	88.27
		凌晨	403.49	49.07	2954.89	28.27	139.13	56.92	529.11	88.45
		晚間	277.12	71.45	2897.05	28.83	103.64	76.42	545.22	85.84
	小客車	上午	500.52	39.56	2204.86	37.88	101.39	78.12	498.11	93.95
		下午	462.66	42.8	2188.48	38.16	97.85	80.94	497.54	94.06
		凌晨	427.64	46.3	2199.5	37.97	119.08	66.51	500.24	93.55
		晚間	271.31	72.98	2166.63	38.55	98.13	80.71	510.53	91.67
	小貨車	上午	494.93	40.01	2199.22	37.98	102.04	77.62	502.67	93.1
		下午	464.32	42.64	2186.94	38.19	99.58	79.54	503.8	92.89
		凌晨	428.47	46.21	2260.82	36.94	114.84	68.97	525.87	89
		晚間	275.07	71.98	2194.33	38.06	102.78	77.06	514.05	91.04

表 4.24 為集群 3 分平假日、車種、時段之平均旅行時間及車速彙整表，整理行車狀況多為正常狀態，車速沒有出現過慢的情況，為可看出在上午之車速較其他時段低，且頭城-宜蘭-羅東之車速較前面三個路段高。因此群之資料多集中在上午、下午與晚上，且平均分布於各星期中命名為清晨以外之車潮

表 4.25(集群 3)清晨以外之車潮

平 假 日	車 種	時段	南港-石碇-坪林		石碇-坪林-頭城		坪林-頭城-宜蘭		頭城-宜蘭-羅東	
			旅行時間 (秒)	時速 (km)	旅行時間 (秒)	時速 (km)	旅行時間 (秒)	時速 (km)	旅行時間 (秒)	時速 (km)
平 日	大型車	上午	281.43	70.36	1138.54	73.36	99.6	79.52	523.12	89.46
		下午	258.19	76.69	1075.55	77.65	99.77	79.38	523.45	89.41
		凌晨	252.09	78.54	1008.98	82.78	97.9	80.9	519.33	90.12
		晚間	251.82	78.63	1049.65	79.57	100.75	78.61	527.67	88.69
	小客車	上午	285.08	69.46	1101.8	75.8	93.87	84.38	489.04	95.7
		下午	262.07	75.55	1027.24	81.3	92.38	85.74	485.82	96.33
		凌晨	250.98	78.89	970.4	86.07	90.64	87.38	485.61	96.37
		晚間	258.63	76.56	1005.2	83.09	90.87	87.16	489.6	95.59
	小貨車	上午	284.51	69.59	1099.06	75.99	95.65	82.8	499	93.79
		下午	264.88	74.75	1029.7	81.11	94.02	84.23	496.43	94.27
		凌晨	250.18	79.14	975.07	85.66	92.88	85.27	498.87	93.81
		晚間	260.55	75.99	1007.62	82.89	92.25	85.86	496.38	94.28

假日	大型車	上午	284.76	69.53	1184.9	70.49	102.25	77.46	523.6	89.38
		下午	272.9	72.55	1129.39	73.95	95.82	82.65	524.55	89.22
		凌晨	257.35	76.94	1036.67	80.57	99.17	79.87	519.59	90.07
		晚間	250.05	79.19	1037.21	80.52	98.37	80.51	527.63	88.7
	小客車	上午	287.55	68.86	1131.04	73.84	94.73	83.61	489.48	95.61
		下午	271.05	73.05	1070.36	78.03	91.45	86.6	488.37	95.83
		凌晨	255.38	77.53	990.59	84.31	92.19	85.91	484.82	96.53
		晚間	255.92	77.37	987.59	84.57	90.44	87.57	489.91	95.53
	小貨車	上午	288.47	68.64	1131.43	73.82	95.34	83.07	494.49	94.64
		下午	272.46	72.67	1071.18	77.97	92	86.09	493.81	94.77
		凌晨	254.34	77.85	993.56	84.06	93.43	84.77	496.29	94.3
		晚間	257.72	76.83	992.03	84.19	91.31	86.74	496.46	94.27

表 4.26 為南港-羅東路徑資料之集群分析重點整理，將三個集群命名後為各個集群提出行駛建議。

表 4.26 ETC_KA2 各級群命名與用路人建議

集群	建議
1 白天之車潮	1. 南港端車多易塞，可考慮替代道路
2 週六白天出遊之車潮	1. 南港端假日晚間較順暢 2. 假日清晨於頭城路段需注意大型車輛 3. 出遊民眾可考慮 12 點過後出發
3 清晨以外之車潮	1. 普遍順暢

III. 南港-蘇澳(ETC_KA3)

表 4.27 為分群結果與平均旅行時間(時速)之表格，分群後發現資料多集中在集群 3，佔了 89.24%，集群 1、3 的比例分別為 9.1%與 1.02%。從平均旅行時間來看，集群 1 在南港-石碇-坪林與石碇-坪林-頭城路段較為車多，行駛時速分別為 33.18 公里與 38.5 公里；集群 2 於南港-石碇-坪林路段屬於車較多的路段，平均時速為 57.65 公里，後續路段皆順暢，直到宜蘭-羅東-蘇澳路段出現異常壅塞的狀況，時速僅 8.98 公里。集群 3 之行使狀況從南港的 64.05 公里到蘇澳的 94.29 公里有越來越快的趨勢。沿線屬於順暢的狀態。

表 4. 27 ETC_KA3 分群結果與平均旅行時間(時速)

集群		1	2	3	總計
資料筆數		200281	20945	1835337	2056563
百分比		9.74%	1.02%	89.24%	100%
南港-石碇-坪林	旅行時間(秒)	596.68	343.43	309.11	337.47
	時速(km)	33.18	57.65	64.05	58.67
石碇-坪林-頭城	旅行時間(秒)	2169.2	1172.9	1060.5	1169.58
	時速(km)	38.5	71.21	78.76	71.41
坪林-頭城-宜蘭	旅行時間(秒)	104.41	102.71	95.15	96.13
	時速(km)	75.85	77.11	83.24	82.39
頭城-宜蘭-羅東	旅行時間(秒)	496.18	494.01	487.35	488.28
	時速(km)	94.32	94.73	96.03	995.85
宜蘭-羅東-蘇澳	旅行時間(秒)	214.25	2205.63	210	230.74
	時速(km)	92.41	8.98	94.29	85.81

表 4.28 為各集群於不同星期之資料比例，集群 1 之資料有 53% 集中在星期六，其次為星期五佔了 14%，其他星期之資料佔不到 10%。集群 2 之資料較分散於各星期，其中以星期五與星期六為最高，分別佔了 17.6% 與 17.7%。集群 3 之資料有 18% 是星期五之資料，資料較平均分布於各個星期中。

表 4.28 ETC_KA3 各集群於不同星期之資料比例

星期 \ 集群	1	2	3
Mon	9.2%	13.1%	13.3%
Tues	4.4%	11.7%	12.3%
Wed	5.5%	12.8%	12.7%
Thurs	9.4%	13.0%	13.7%
Fri	14.0%	17.7%	18.4%
Sat	53.0%	17.4%	14.2%
Sun	4.6%	14.4%	15.5%
總計	100%	100%	100%

表 4.29 為各集群於不同時段之資料比例，可以看出集群 1 之資料多集中在上午，佔了 62.3%，其次為下午佔了 31.8%，晚間及清晨資料共不到 6% 資料。集群 2 之資料多集中於上午及下午，分別佔了 41.8% 與 35.7%，晚間資料佔了 15%。集群 3 之資料有 41.1% 在上午，其次為下午，佔了 26.9%，晚間及清晨資料各佔了 16%。

表 4.29 ETC_KA2 各集群於不同時段之資料比例

時段 \ 集群	1	2	3
上午	62.3%	41.8%	41.1%
下午	31.8%	35.7%	26.9%
凌晨	2.3%	7.5%	16.1%
晚間	3.6%	15.0%	16.0%
總計	100%	100%	100%

表 4.29 為集群 1 分平假日、車種、時段之平均旅行時間及車速彙整表，發現大型車不分時段與平假日在南港-石碇-坪林路段都呈現非常壅塞的狀態，尤其是晚間時段的時速更只有 14 公里與 11 公里，其他車種在此路段也都是車多擁擠狀態，平均車速在 30 到 40 之間；石碇-坪林-頭城路段也都是車多壅擠的狀態，越往後面的路段狀況越好，唯獨大型車於平日晚間時段在宜蘭-羅東-蘇澳時速僅 58.05 公里，用路人於南港-石碇-坪林路段行車時須留意大型車輛，抑或是選擇替代道路由石碇或坪林上交流道，避開大型車潮。此集群由於資料多集中於星期六上午及下午，命名為假日白天之車潮。

表 4.30(集群 1)假日白天之車潮

平 假 日	車 種	時段	南港-石碇-坪林		石碇-坪林-頭城		坪林-頭城-宜蘭		頭城-宜蘭-羅東		宜蘭-羅東-蘇澳	
			旅行 時間	時速 (km)	旅行 時間	時速 (km)	旅行 時間	時速 (km)	旅行 時間	時速 (km)	旅行 時間	時速 (km)
平 日	大 型 車	上午	931.99	21.24	1999.85	41.76	110.34	71.78	533.47	87.73	236.53	83.71
		下午	815.45	24.28	2144.95	38.94	98.78	80.18	533.32	87.75	244.34	81.03
		凌晨	886.94	22.32	2652.32	31.49	95.32	83.09	521.32	89.77	220	90
		晚間	1325.41	14.94	2015.08	41.45	99.73	79.41	545.29	85.83	341.08	58.05
	小 客 車	上午	654.52	30.25	2070.6	40.34	105.3	75.21	496.39	94.28	213.16	92.89
		下午	531.09	37.28	2323	35.95	102.23	77.47	494.42	94.66	214.54	92.29
		凌晨	532.95	37.15	2060.89	40.53	102.83	77.02	498.74	93.84	209.04	94.72
		晚間	638.93	30.99	2218.38	37.65	105.73	74.91	512.63	91.29	215.79	91.76
	小 貨 車	上午	650.64	30.43	2065.53	40.44	103.64	76.42	506.25	92.45	215.5	91.88
		下午	537.22	36.86	2350.41	35.53	102.6	77.19	501.1	93.39	216.84	91.31
		凌晨	515.33	38.42	2128.8	39.23	102.45	77.31	501.56	93.31	216.26	91.56
		晚間	598.33	33.09	2216.9	37.67	106.42	74.42	515.81	90.73	215.23	92
假 日	大 型 車	上午	1004.73	19.71	1876.91	44.5	128.43	61.67	514.46	90.97	233.79	84.69
		下午	741.4	26.71	1978.72	42.21	101.15	78.3	523.46	89.41	241.82	81.88
		凌晨	1411.56	14.03	1651.97	50.56	110.79	71.48	512.26	91.36	215.19	92.01
		晚間	1780.91	11.12	1399.09	59.7	100.09	79.13	541	86.51	225.18	87.93
	小 客 車	上午	586.41	33.77	2203.91	37.9	105.99	74.72	492.11	95.1	213.6	92.7
		下午	530.36	37.33	2191.12	38.12	100.65	78.69	491.35	95.25	213.01	92.95
		凌晨	876.06	22.6	2116.77	39.46	112.76	70.24	497.36	94.1	210.42	94.1
		晚間	806.1	24.56	1993.91	41.89	107.28	73.83	505.74	92.54	211.17	93.76
	小 貨 車	上午	579.68	34.16	2197.64	38	105.49	75.08	495.6	94.43	214.26	92.41
		下午	527.76	37.52	2191.91	38.1	102.19	77.5	498.63	93.86	214.73	92.21
		凌晨	954.88	20.74	2018.13	41.38	127.44	62.15	497.16	94.14	218.36	90.68
		晚間	859.15	23.05	2063.44	40.48	103.02	76.88	513.84	91.08	239.61	82.64

表 4.30 為集群 2 分平假日、車種、時段之平均旅行時間及車速彙整表，可以看出南港-石碇-坪林路段屬於車多速度慢的狀況，僅少數狀況如平日清晨的車速較為正常，時速大約 70 到 80 公里，石碇-坪林-頭城路段在上午時段多為車多狀況，時速多在 70 公里以下；宜蘭-羅東-蘇澳路段壅塞程度最為嚴重，平均車速不到 10 公里，用路人應避免通過此路段，考慮於宜蘭或羅東下高速公路改走替代道路。此集群之資料較為平均分散於各個星期，時段則主要集中在上午至晚間，命名為清晨以外之車潮。

表 4.31(集群 2)清晨以外之車潮

平 假 日	車 種	時段	南港-石碇-坪林		石碇-坪林-頭城		坪林-頭城-宜蘭		頭城-宜蘭-羅東		宜蘭-羅東-蘇澳	
			旅行 時間	時速 (km)	旅行 時間	時速 (km)	旅行 時間	時速 (km)	旅行 時間	時速 (km)	旅行 時間	時速 (km)
平 日	大 型 車	上午	437.08	45.3	1230.59	67.87	107.28	73.82	561.97	83.28	2158.23	9.17
		下午	408.12	48.51	1074.46	77.73	112.35	70.49	523.01	89.48	2014.42	9.83
		凌晨	245	80.82	987.67	84.56	91.56	86.5	564.67	82.88	1976.33	10.02
		晚間	297.69	66.51	1050.06	79.54	93.9	84.34	531.06	88.13	1679.4	11.79
	小 客 車	上午	367.01	53.95	1209.27	69.07	102.13	77.55	492.5	95.03	2280.59	8.68
		下午	311.25	63.61	1087.1	76.83	98.06	80.77	488.7	95.76	2194.54	9.02
		凌晨	277.96	71.23	1022.7	81.67	109.24	72.5	488.79	95.75	2063.33	9.6
		晚間	302.72	65.41	1029.74	81.11	101.31	78.18	493.81	94.77	2125.11	9.32
	小 貨 車	上午	354.06	55.92	1182.56	70.63	100.23	79.02	502.03	93.22	2189.74	9.04
		下午	303.17	65.31	1065.25	78.4	99.28	79.77	497.05	94.15	2142.16	9.24
		凌晨	251.92	78.6	971.09	86.01	207.1	38.24	486.3	96.24	1735.67	11.41
		晚間	286.16	69.19	1037.75	80.48	90.96	87.07	498.08	93.96	2083.67	9.5
假 日	大 型 車	上午	506.01	39.13	1288.76	64.81	118.33	66.93	520.14	89.97	2280.12	8.68
		下午	455.48	43.47	1251.99	66.71	94.1	84.17	516.89	90.54	1849.61	10.7
		凌晨	385.27	51.39	1143.64	73.03	94.64	83.69	537.09	87.14	2259.91	8.76
		晚間	296.06	66.88	1030.57	81.04	104.41	75.85	527.61	88.7	1770.74	11.18
	小 客 車	上午	408.24	48.5	1428.8	58.45	103.04	76.87	492.37	95.05	2366.75	8.37
		下午	364.83	54.27	1317.56	63.39	92.02	86.07	489.81	95.55	2266.77	8.73
		凌晨	322.2	61.45	1065.17	78.41	96.29	82.25	486.6	96.18	2185.78	9.06
		晚間	302.76	65.4	1006.73	82.96	93.7	84.53	492.02	95.12	2140.33	9.25
	小 貨 車	上午	418.71	47.29	1425.65	58.58	109.89	72.07	498.25	93.93	2354.54	8.41
		下午	378.56	52.3	1325.54	63.01	96.47	82.1	499.28	93.74	2250.7	8.8
		凌晨	274.18	72.22	1030.4	81.06	182.09	43.5	482.24	97.05	1979.3	10
		晚間	339.55	58.31	1009.63	82.72	97.11	81.56	501.21	93.37	2103.09	9.41

表 4.31 為集群 3 分平假日、車種、時段之平均旅行時間及車速彙整表，發現大型車在南港-石碇-坪林路段多為車多狀態，平均車速不到 60 公里，上午時段較其他時段來的嚴重，平均時速不到僅 43 公里，其他路段之行車狀況都為順暢，所需花費之旅行時間較少。此集群包含比例較高之清晨行車資料，又資料分散無各個星期中，將其命名為全天之車潮。

表 4.32(集群 3)全天之車潮

平 假 日	車 種	時段	南港-石碇-坪林		石碇-坪林-頭城		坪林-頭城-宜蘭		頭城-宜蘭-羅東		宜蘭-羅東-蘇澳	
			旅行 時間	時速 (km)	旅行 時間	時速 (km)	旅行 時間	時速 (km)	旅行 時間	時速 (km)	旅行 時間	時速 (km)
平 日	大 型 車	上午	459.52	43.09	1131.46	73.82	132.82	59.63	510.4	91.69	229.87	86.14
		下午	321	61.68	1077.78	77.49	94.81	83.54	510.27	91.72	233.42	84.83
		凌晨	383.46	51.63	1025.52	81.44	105.35	75.17	505.41	92.6	218.42	90.65
		晚間	338.58	58.48	1042.7	80.1	96.56	82.02	519.14	90.15	372.56	53.15
	小 客 車	上午	327.83	60.4	1106.13	75.51	96.09	82.42	486.27	96.24	207.73	95.32
		下午	302.32	65.49	1037.21	80.52	94.06	84.2	481.8	97.14	208.74	94.85
		凌晨	286.4	69.13	992.59	84.14	92.8	85.35	483.83	96.73	206.24	96
		晚間	292.41	67.71	1014.69	82.31	92.71	85.43	487.41	96.02	211.81	93.48
	小 貨 車	上午	319.92	61.89	1100.74	75.88	97.03	81.63	494.69	94.61	211.21	93.75
		下午	304.24	65.08	1039.11	80.38	95.09	83.29	491.54	95.21	212.74	93.07
		凌晨	276.34	71.65	988.33	84.51	93.46	84.74	494	94.74	210.25	94.18
		晚間	295.65	66.97	1017.33	82.1	94.05	84.21	496.07	94.34	213.77	92.62
假 日	大 型 車	上午	454.37	43.58	1218.53	68.54	148.01	53.51	517.8	90.38	228.05	86.82
		下午	362.07	54.69	1148.12	72.74	94.61	83.71	518.23	90.31	246.93	80.18
		凌晨	433.44	45.68	1080.15	77.32	120.87	65.52	507.18	92.27	215.31	91.96
		晚間	393.47	50.32	1029.97	81.09	96.72	81.89	522.57	89.56	436.25	45.39
	小 客 車	上午	320.73	61.73	1137.83	73.4	96.57	82.02	485.25	96.45	207.29	95.52
		下午	309.71	63.93	1072.65	77.86	92.08	86.01	482.18	97.06	209.25	94.62
		凌晨	286.17	69.19	999.57	83.56	94.75	83.59	481.15	97.27	205	96.58
		晚間	286.07	69.21	986.51	84.66	91.56	86.5	485.44	96.41	211.77	93.5
	小 貨 車	上午	316.9	62.48	1136.43	73.49	96.48	82.09	489.22	95.66	209.25	94.62
		下午	310.94	63.68	1073.88	77.77	92.51	85.61	489.37	95.63	210.67	93.99
		凌晨	278.54	71.09	997.03	83.77	94.87	83.48	488.25	95.85	207.33	95.5
		晚間	288.2	68.7	992.97	84.11	93	85.16	495	94.55	215.17	92.02

表 4.33 為南港-蘇澳路徑資料之集群分析重點整理，將三個集群命名後為各個集群提出行駛建議。

表 4.33 ETC_KA3 各級群命名與用路人建議

集群	建議
1 假日白天之車潮	1. 大型車於南港路段可考慮改走替代道路
2 清晨以外之車潮	1. 建議中午過後出發，可避開上午車潮 2. 建議避開羅東-蘇澳路段，另尋替代道路
3 全天之車潮	1. 大型車避免行經南港路段 2. 普遍順暢

IV. 石碇-羅東(ETC_KB1)

表 4.34 為分群結果與平均旅行時間(時速)，集群 1 之資料佔了整體的 86.2%，其次為集群 3 佔了 13.55%，集群 2 僅佔了 0.24%，但資料量仍有 1605 筆之多。其中，集群 1 之行駛狀況都在平均之上，行車狀況為順暢；集群 2 在石碇-坪林-頭城路段非常壅塞，平均時速僅 10.59 公里；集群 3 在石碇-坪林-頭城路段顯示車多速度較慢的狀態。

表 4.34 ETC_KB1 分群結果與平均旅行時間(時速)

集群		1	2	3	總計
資料筆數		565590	1605	88938	656133
百分比		86.2%	0.24%	13.55%	100%
石碇-坪林-頭城	旅行時間(秒)	1070.6	7887	2209.3	1241.64
	時速(km)	78.01	10.59	37.8	67.27
坪林-頭城-宜蘭	旅行時間(秒)	94.18	112.41	100	95.01
	時速(km)	84.1	70.46	79.2	83.36
頭城-宜蘭-羅東	旅行時間(秒)	494.92	531.49	504.31	496.28
	時速(km)	94.56	88.05	92.8	94.3

表 4.35 為各集群於不同星期之資料比例，集群 1 之資料較平均分布於各個星期中；集群 2 之資料最多為星期六，佔了 19.8%，其他星期之資料較為平均，比例大都 12%至 14%；集群 3 之資料有 66.7%集中在星期六，其他星期之資料接不到 10%。

表 4.35 ETC_KB1 各集群於不同星期之資料比例

星期 \ 集群	1	2	3
Mon	14.1%	12.6%	6.1%
Tues	13.8%	13.8%	3.2%
Wed	14.0%	14.8%	3.5%
Thurs	14.4%	13.4%	7.2%
Fri	17.1%	13.4%	9.5%
Sat	13.3%	19.8%	66.7%
Sun	13.2%	12.3%	3.8%
總計	100%	100%	100%

表 4.36 為各集群於不同時段之資料比例，集群 1 之資料最多為上午時段，比例為 42.4%、其次為下午，佔了 32.7%、晚間佔了 21%；集群 2 之資料多集中在上午及下午，佔了整體的 85% 以上；集群 3 之資料也在上午及下午時段佔了 95% 以上之資料，顯示使旅行資料多為上午、下午之資料。

表 4.36 ETC_KB2 各集群於不同時段之資料比例

時段 \ 集群	1	2	3
上午	42.4%	38.6%	57.2%
下午	32.7%	47.1%	39.9%
凌晨	3.9%	6.0%	0.3%
晚間	21.0%	8.2%	2.6%
總計	100%	100%	100%

表 4.37 為集群 1 分平假日、車種、時段之平均旅行時間及車速，看出大客車在假日上午時段於石碇-坪林-頭城及坪林-頭城-宜蘭路段車速較其他時段、車種慢，平均時速僅 60 出頭。下午時段在石碇-坪林-頭城路段亦只有 64 公里時速。其他平假日、時段之車種行駛狀況皆順暢正常。此集群之資料多集中在凌晨時段以外，命名為凌晨以外之車潮。

表 4.37(集群 1)凌晨以外之車潮

平 假 日	車 種	時段	石碇-坪林-頭城		坪林-頭城-宜蘭		頭城-宜蘭-羅東	
			旅行 時間	時速 (km)	旅行 時間	時速 (km)	旅行 時間	時速 (km)
平 日	大 型 車	上午	1170.84	71.33	111.45	71.06	517.74	90.39
		下午	1115.21	74.89	98.62	80.31	522.06	89.64
		凌晨	1069.38	78.1	105.73	74.91	542.25	86.31
		晚間	1062.18	78.63	101.73	77.86	539.58	86.73
	小 客 車	上午	1105.27	75.57	94.43	83.87	491.66	95.19
		下午	1036.9	80.55	92.79	85.35	489.79	95.55
		凌晨	973.47	85.8	90.73	87.29	491.71	95.18
		晚間	1010.08	82.69	91.54	86.52	493.18	94.9
	小 貨 車	上午	1100.82	75.87	95.52	82.92	500.85	93.44
		下午	1042.34	80.13	95.68	82.78	502.35	93.16
		凌晨	979.63	85.26	93.46	84.75	503.21	93
		晚間	1016.14	82.19	93.24	84.95	501.64	93.29

假日	大型車	上午	1373.04	60.83	127.55	62.09	524.84	89.17
		下午	1300.82	64.21	107.98	73.34	530.59	88.2
		晚間	1061.56	78.68	97.17	81.51	526.57	88.88
	小客車	上午	1156.99	72.19	95.88	82.6	492.96	94.94
		下午	1092.91	76.42	92.62	85.51	492.11	95.1
		凌晨	974.47	85.71	93.12	85.05	486.21	96.25
		晚間	1000.55	83.47	91.77	86.31	494.56	94.63
	小貨車	上午	1154.76	72.33	95.88	82.6	498.77	93.83
		下午	1095.34	76.25	93.93	84.32	500.12	93.58
		凌晨	985.27	84.77	96.09	82.42	498.04	93.97
		晚間	1009.59	82.73	93.83	84.41	504.62	92.74

表 4.38 為集群 2 分平假日、車種、時段之平均旅行時間及車速，資料中可以明顯看出石碇-坪林-頭城路段非常壅塞，車速較慢，平均車速不到 11 公里，建議改走替代道路以避開壅塞之時段。此集群之資料多位於上午及下午，平均分散於各個星期，因此命名為白天之車潮。

表 4.38(集群 2)白天之車潮

平 假 日	車 種	時段	石碇-坪林-頭城		坪林-頭城-宜蘭		頭城-宜蘭-羅東	
			旅行 時間	時速 (km)	旅行 時間	時速 (km)	旅行 時間	時速 (km)
平 日	大型車	上午	8820.1	9.47	95.4	83.02	607.9	76.99
		下午	8201.26	10.18	167.22	47.36	925.91	50.54
		凌晨	9937.81	8.4	127.08	62.32	580.35	80.64
		晚間	7883.18	10.59	226.64	34.95	538.73	86.87
	小客車	上午	7913.78	10.55	106.32	74.5	513.02	91.22
		下午	7795.38	10.71	106.21	74.57	516.13	90.67
		凌晨	9466.28	8.82	92.44	85.67	512.5	91.32
		晚間	6858.42	12.18	116.76	67.83	513.46	91.15
	小貨車	上午	8095.72	10.32	111.61	70.96	526.26	88.93
		下午	7957.9	10.5	158.81	49.87	620.43	75.43
		凌晨	9410.75	8.87	125	63.36	547.15	85.53
		晚間	8994.46	9.29	149.02	53.15	570.92	81.97

假日	大型車	上午	8534.5	9.79	97.67	81.09	519.5	90.09
		下午	9189.94	9.09	101.28	78.2	611.17	76.57
		晚間	11720.33	7.13	96	82.5	508	92.13
	小客車	上午	7708.87	10.83	103.97	76.18	502.02	93.22
		下午	7381.82	11.31	93.24	84.94	501.04	93.41
		凌晨	7302.75	11.44	92.88	85.28	515.88	90.72
		晚間	7370.57	11.33	89	88.99	493.86	94.76
	小貨車	上午	7316.93	11.41	126.46	62.63	509	91.94
		下午	7033.37	11.87	93.28	84.9	506.24	92.45
		凌晨	10003.6	8.35	95.6	82.85	537	87.15
		晚間	5480.25	15.24	88	90	480.5	97.4

表 4.39 為集群 3 分平假日、車種、時段之平均旅行時間及車速，資料顯示石碇-坪林-頭城路段為車多速度較慢的狀況，後續路段之旅行狀態較為正常，為大客車於平日下午行經坪林-頭城-宜蘭時的車速較慢，時速為 55.6 公里。此集群之資料多分布於星期六的上午及下午時段，命名為週六白天之車潮。

表 4.39(集群 3)週六白天之車潮

平假日	車種	時段	石碇-坪林-頭城		坪林-頭城-宜蘭		頭城-宜蘭-羅東	
			旅行時間	時速(km)	旅行時間	時速(km)	旅行時間	時速(km)
平日	大型車	上午	2165.36	38.57	102.17	77.52	520.76	89.87
		下午	2211.11	37.77	142.45	55.6	531.89	87.99
		凌晨	2940.52	28.4	109.95	72.03	568.95	82.26
		晚間	2187.69	38.18	117.35	67.49	608.48	76.91
	小客車	上午	2146.7	38.91	99.35	79.72	504.38	92.79
		下午	2365.04	35.31	101.26	78.22	505.33	92.61
		凌晨	2483.98	33.62	99.09	79.92	518.66	90.23
		晚間	2286.51	36.53	101.23	78.24	518.81	90.21
	小貨車	上午	2152.55	38.8	101.66	77.9	512.58	91.3
		下午	2367.89	35.27	101.74	77.84	512.37	91.34
		凌晨	2508.12	33.3	119.46	66.3	534.67	87.53
		晚間	2336.04	35.75	108.91	72.72	547.12	85.54

假日	大型車	上午	1994.81	41.87	108.83	72.77	523.87	89.34
		下午	2174.94	38.4	114.8	68.99	530.84	88.16
		凌晨	2067	40.41	90.33	87.68	515.33	90.82
		晚間	1913.33	43.65	105.67	74.95	572.67	81.72
	小客車	上午	2212.98	37.74	99.18	79.85	499.33	93.72
		下午	2201.16	37.94	97.92	80.88	498.93	93.8
		凌晨	2171.75	38.46	93.03	85.14	507.42	92.23
		晚間	2094.48	39.88	97.87	80.93	527.93	88.65
	小貨車	上午	2215.97	37.69	99.11	79.91	506.13	92.47
		下午	2210.04	37.79	99.24	79.81	506.14	92.47
		凌晨	2361.45	35.37	124.45	63.64	611.32	76.56
		晚間	1970.15	42.39	97.79	80.99	527.19	88.77

表 4.40 為石碇-羅東路徑資料之集群分析重點整理，將三個集群命名後為各個集群提出行駛建議。

表 4.40 ETC_KB1 各級群命名與用路人建議

集群	建議
1 清晨以外之車潮	1. 普遍順暢 2. 大型車於假日可避開上午時段通過頭城路段
2 白天之車潮	1. 石碇路段車多壅塞，建議改走替代道路
3 週六白天之車潮	1. 頭城路段車多易塞，建議避開假日白天時段 2. 大型車行經宜蘭考慮避開平日上午時段

V. 石碇-蘇澳(ETC_KB2)

表 4.41 為分群結果與平均旅行時間(時速)，集群 3 之資料佔了整體的 84.07%，其次為集群 1 佔了 15.65%，集群 2 僅佔了 0.28%。集群 1 在石碇-坪林-頭城路段車速較慢，平均車速為 37.49 公里，後續路段之狀況有較佳，略低於整體平均車速；集群 2 之資料中，顯示石碇-坪林-頭城路段非常壅塞，車速非常慢，後續路段之車速也都低於平均車速；集群 3 之旅行狀況較其他集群佳，各個路段之車速都在平均之上。

表 4.41 ETC_KB2 分群結果與平均旅行時間(時速)

集群		1	2	3	總計
資料筆數		43754	778	235044	279576
百分比		15.65%	0.28%	84.07%	100%
石碇-坪林-頭城	旅行時間(秒)	2228.1	6742.1	1087.4	1281.68
	時速(km)	37.49	12.39	76.81	65.16
坪林-頭城-宜蘭	旅行時間(秒)	104.54	120.96	97.43	98.61
	時速(km)	75.76	65.48	81.29	80.32
頭城-宜蘭-羅東	旅行時間(秒)	499.45	537.96	492.12	493.39
	時速(km)	93.7	87	95.1	94.85
宜蘭-羅東-蘇澳	旅行時間(秒)	241.15	252.78	231.38	232.97
	時速(km)	82.11	78.33	85.57	84.99

表 4.42 為各集群於不同星期之資料比例，集群 1 之資料有 62.3% 落在星期六，其次為星期五的 11.1%，其他星期之資料皆小於 10%；集群 2 之資料有 21% 落在星期六，其次為星期四，佔了 17.2%；集群 3 之資料有 17.9% 落在星期五，其他星期之資料較為平均。

表 4.42 ETC_KB2 各集群於不同星期之資料比例

星期 \ 集群	1	2	3
Mon	7.5%	13.6%	14.1%
Tues	3.8%	13.5%	12.7%
Wed	4.5%	12.9%	13.3%
Thurs	7.6%	17.2%	14.4%
Fri	11.1%	11.6%	17.9%
Sat	62.3%	21.0%	13.8%
Sun	3.3%	10.3%	13.7%
總計	100%	100%	100%

表 4.36 為各集群於不同時段之資料比例，集群各個集群之資料多落在上午及下午，集群 1 佔了 97%，集群 2 佔了 92%，集群 3 佔了 79%。集群 3 在晚間之資料佔了 11%，其他時段之資料比例皆小於 10%。

表 4.43 ETC_KB2 各集群於不同時段之資料比例

<div> <div>集群</div> <div>時段</div> </div>	1	2	3
上午	64.4%	59.8%	52.7%
下午	33.2%	33.5%	26.8%
凌晨	0.6%	2.2%	9.5%
晚間	1.8%	4.5%	11.0%
總計	100%	100%	100%

表 4.44 為集群 1 分平假日、車種、時段之平均旅行時間及車速，可以看出石碇-坪林-頭城路段之車速不分車種時段都較慢，平均僅 37 公里；大型車於平日下午行經坪林-頭城-宜蘭及晚間行經宜蘭-羅東-蘇澳路段之車速較其他車種低，建議可以提早至 12 點以前通過坪林路段，便可連帶避免宜蘭路段之車速較慢問題。此集群之資料多集中於星期六上午及下午，命名為週六白天之車潮。

表 4.44(集群 1)週六白天之車潮

平 假 日	車 種	時段	石碇-坪林-頭城		坪林-頭城-宜蘭		頭城-宜蘭-羅東		宜蘭-羅東-蘇澳	
			旅行 時間	時速 (km)	旅行 時間	時速 (km)	旅行 時間	時速 (km)	旅行 時間	時速 (km)
平 日	大 型 車	上午	2167.01	38.54	101.67	77.9	523.07	89.47	280.52	70.58
		下午	2284.56	36.56	132.07	59.97	614.51	76.16	281.69	70.29
		凌晨	2652.86	31.48	95.43	82.99	517	90.52	221	89.59
		晚間	2020	41.35	102.5	77.27	544.67	85.92	349.67	56.63
	小 客 車	上午	2188.92	38.16	102.37	77.36	499.41	93.71	235.41	84.11
		下午	2348.34	35.57	103.01	76.88	502.94	93.05	242.06	81.8
		凌晨	2293.82	36.41	93.23	84.95	505.26	92.62	216.87	91.3
		晚間	2184.23	38.24	105.97	74.74	518.81	90.21	240.97	82.17
	小 貨 車	上午	2155.64	38.74	103.34	76.64	502.68	93.1	242.64	81.6
		下午	2299.79	36.32	105.89	74.79	517.74	90.39	249.84	79.25
		凌晨	2230.33	37.45	94.69	83.64	507.77	92.17	218.66	90.55
		晚間	2120.16	39.39	120.1	65.95	510.91	91.6	241.7	81.92

假日	大型車	上午	1975.52	42.28	152.03	52.1	515.28	90.82	246.15	80.44
		下午	2084.29	40.07	103.45	76.56	516.1	90.68	221.58	89.36
		晚間	1966	42.48	94	84.26	503	93.04	208.5	94.96
	小客車	上午	2253.1	37.07	103.92	76.21	495.56	94.44	239.81	82.57
		下午	2227.51	37.49	101.83	77.78	493.05	94.92	244.2	81.08
		凌晨	2014.76	41.45	92.94	85.22	541.58	86.41	245.94	80.51
		晚間	1965.44	42.49	97.74	81.03	510.82	91.62	223.41	88.62
	小貨車	上午	2245.12	37.2	105.82	74.84	502.67	93.1	240.11	82.46
		下午	2221.74	37.59	101.43	78.08	498.02	93.97	243.69	81.25
		凌晨	2377.4	35.13	90	88	492.6	95.01	204.93	96.62
		晚間	2080.04	40.15	100.5	78.81	539.12	86.81	287.42	68.89

表 4.45 為集群 2 分平假日、車種、時段之平均旅行時間及車速，可以看出石碇-坪林-頭城路段之行駛車速較慢，平均不到 13 公里，建議用路人可考慮趕走替代道路，小貨車在平日凌晨與假日晚間通過坪林-頭城-宜蘭路段之車速也較低，建議可下午時段再通過，避開車潮。大型車於平日早上通過坪林-頭城-宜蘭之車速亦較低，車速平均 46.72 公里，建議可中午過後通過。小貨車於平日晚間通過宜蘭-羅東-蘇澳路段之車速也不高，平均車速僅 46.28 公里，建議可提早以避開此時段。此集群之資料多落在上午及下午，並平均分布於各個星期，命名為白天之車潮。

表 4.45(集群 2)白天之車潮

平假日	車種	時段	石碇-坪林-頭城		坪林-頭城-宜蘭		頭城-宜蘭-羅東		宜蘭-羅東-蘇澳	
			旅行時間	時速(km)	旅行時間	時速(km)	旅行時間	時速(km)	旅行時間	時速(km)
平日	大型車	上午	5472.08	15.26	169.54	46.72	531.15	88.11	221.08	89.56
		下午	6059.5	13.78	96	82.5	528.25	88.59	219	90.41
		凌晨	7711	10.83	96.67	81.93	490	95.51	207.33	95.5
		晚間	7841.88	10.65	103.5	76.52	544.88	85.89	229.62	86.23
	小客車	上午	6763.88	12.35	143.61	55.15	546.67	85.61	266.93	74.18
		下午	6053.2	13.8	126.15	62.78	521.58	89.73	250.28	79.11
		凌晨	5405.75	15.45	209.5	37.8	478	97.91	211.25	93.73
		晚間	6538.73	12.77	94.33	83.96	839.13	55.77	226.07	87.58
	小貨車	上午	8261.95	10.11	117.69	67.3	595.32	78.61	239.01	82.84
		下午	6475.97	12.9	94	84.26	508.37	92.06	283.17	69.92
		凌晨	5847	14.28	95.6	82.85	525.2	89.11	216.6	91.41
		晚間	7023.12	11.89	88.5	89.49	500.5	93.51	427.88	46.28

假日	大型車	上午	7942	10.52	102.71	77.11	522.43	89.58	294.57	67.22
		下午	5738	14.56	94	84.26	508	92.13	578	34.26
		凌晨	6657	12.55	108	73.33	617	75.85	255	77.65
	小客車	上午	6623.1	12.61	94.18	84.09	496.56	94.25	216.69	91.37
		下午	6386.45	13.08	102.03	77.62	507.78	92.17	231.02	85.71
		凌晨	13121.67	6.37	95	83.37	527.67	88.69	214.33	92.38
		晚間	5325.5	15.68	258.75	30.61	548.75	85.28	606	32.67
	小貨車	上午	6680.92	12.5	93.83	84.4	501.62	93.3	212.08	93.36
		下午	6511.85	12.83	95.45	82.98	509.2	91.91	294.9	67.14
		凌晨	7447	11.22	101	78.42	548	85.4	217	91.24

表 4.46 為集群 3 分平假日、車種、時段之平均旅行時間及車速，資料顯示各個路段皆為順暢狀態，唯獨大型車於上午時段行經坪林-頭城-宜蘭時之車速較低，平日平均為 58.66 公里，假日平均為 47.78 公里，建議可提早或晚一點上路通過此路段。此集群之資料較前兩集群相比，較平均分散於各個時段，命名為一般車潮。

表 4.46(集群 3)一般車潮

平假日	車種	時段	石碇-坪林-頭城		坪林-頭城-宜蘭		頭城-宜蘭-羅東		宜蘭-羅東-蘇澳	
			旅行時間	時速(km)	旅行時間	時速(km)	旅行時間	時速(km)	旅行時間	時速(km)
平日	大型車	上午	1183.01	70.6	135.02	58.66	522.3	89.6	234.31	84.5
		下午	1097.31	76.11	96.05	82.45	517.35	90.46	243.8	81.22
		凌晨	1046.86	79.78	92.9	85.25	512.47	91.32	234.54	84.42
		晚間	1074.65	77.72	98.41	80.48	531.67	88.02	269.05	73.59
	小客車	上午	1116.31	74.82	96.64	81.95	489.98	95.51	230.5	85.9
		下午	1048.62	79.65	95.81	82.66	485.68	96.36	236.35	83.77
		凌晨	995.58	83.89	96.85	81.78	489.28	95.65	216.72	91.36
		晚間	1023.93	81.57	93.59	84.62	491.43	95.23	235.25	84.17
	小貨車	上午	1103.9	75.66	97.68	81.08	498.83	93.82	234.59	84.4
		下午	1051.86	79.4	97.59	81.16	497.24	94.12	237.94	83.22
		凌晨	991.32	84.25	96.07	82.44	501.72	93.28	218.53	90.61
		晚間	1025.34	81.46	96.1	82.42	496.81	94.2	234.18	84.55

假日	大型車	上午	1368.31	61.04	165.75	47.78	519.38	90.11	245.13	80.77
		下午	1230.75	67.86	98.97	80.02	515.76	90.74	235.53	84.06
		凌晨	1043.1	80.07	108.57	72.95	510.31	91.71	244.97	80.82
		晚間	1063.07	78.57	95.14	83.25	521.13	89.81	271.36	72.97
	小客車	上午	1165.75	71.64	97.16	81.51	489.49	95.61	226.89	87.27
		下午	1101.23	75.84	93.93	84.32	488.7	95.76	234.95	84.27
		凌晨	986.25	84.68	95.53	82.91	480.06	97.49	212.12	93.34
		晚間	1000.65	83.47	92.73	85.41	488.55	95.79	233.11	84.94
	小貨車	上午	1162.65	71.84	98.63	80.3	492.3	95.06	230.78	85.79
		下午	1100.47	75.9	96.04	82.46	497.02	94.16	228.76	86.56
		凌晨	990.54	84.32	99.43	79.66	496.16	94.32	215.1	92.05
		晚間	1009.66	82.72	95.63	82.82	499.68	93.66	228.72	86.57

表 4.47 為石碇-蘇澳路徑資料之集群分析重點整理，將三個集群命名後為各個集群提出行駛建議。

表 4.47 ETC_KB2 各級群命名與用路人建議

集群	建議
1 週六白天之車潮	1. 石碇路段車速較低，可考慮走替代道路 2. 大型車可考慮提早行走頭城路段，避開此處與宜蘭之車潮
2 白天之車潮	1. 小貨車可考慮利用下午時段通過坪林路段 2. 大型車可於中午過後行駛坪林路段
3 一般車潮	1. 大型車建議避開上午通過坪林路段 2. 普遍順暢

VI. 坪林-蘇澳(ETC_KC1)

表 4.48 為分群結果與平均旅行時間(時速)，集群 2 之資料佔了 97.14%，集群 1 與集群 3 分別佔了 1.73%與 1.13%。其中，集群 1 再宜蘭-羅東-蘇澳路段之車速非常慢，平均僅 9.54 公里，其他路段也都低於平均車速；集群 2 之行駛狀況較佳，時速都在平均之上。集群 3 於各個路段之車速都不高，坪林-頭城-宜蘭路段僅 45.26 公里、頭城-宜蘭-羅東路段僅 13.64 公里、宜蘭-羅東-蘇澳路段僅 29.52 公里。

表 4.48 ETC_KC1 分群結果與平均旅行時間(時速)

集群		1	2	3	總計
資料筆數		456	25574	298	26328
百分比		1.73%	97.14%	1.13%	100%
坪林-頭城-宜蘭	旅行時間(秒)	130.94	103.52	174.97	104.8
	時速(km)	60.48	76.51	45.26	75.57
頭城-宜蘭-羅東	旅行時間(秒)	668.54	501.16	3431.11	537.23
	時速(km)	70	93.38	13.64	87.11
宜蘭-羅東-蘇澳	旅行時間(秒)	2075.31	218.28	670.7	255.57
	時速(km)	9.54	90.71	29.52	77.47

表 4.49 為各集群於不同集群之資料比例，集群 1 之資料有 27.4%落在星期六，其次為星期五與星期一，各佔了 14.3%；集群 2 之資料中有 34.8%在星期六，其次為星期五，佔了 14.2%；集群 3 之資料較平均落在各個星期中。

表 4.49 ETC_KC1 各集群於不同星期之資料比例

星期 \ 集群	1	2	3
Mon	14.3%	10.0%	17.8%
Tues	10.3%	8.4%	14.1%
Wed	9.9%	9.2%	11.4%
Thurs	11.6%	11.2%	15.1%
Fri	14.3%	14.2%	15.1%
Sat	27.4%	34.8%	14.4%
Sun	12.3%	12.2%	12.1%
總計	100%	100%	100%

表 4.50 為各集群於不同時段之資料比例，集群 1 之資料有 43% 落在下午時段，其次為上午佔了 36.8%，晚間佔了 13.6%；集群 2 之資料有 85% 以上之資料落在上以及下午時段，晚間及清晨之資料不到 10%；集群 3 之資料有 32.6% 在上午，其次為晚上，有 29.2%，下午跟凌晨時段各佔了 19.2%。

表 4.50 ETC_KC1 各集群於不同時段之資料比例

時段 \ 集群	1	2	3
上午	36.8%	47.1%	32.6%
下午	43.0%	39.7%	19.1%
凌晨	6.6%	4.9%	19.1%
晚間	13.6%	8.4%	29.2%
總計	100%	100%	100%

表 4.51 為集群 1 分平假日、車種、時段之平均旅行時間及車速，可以看出坪林-頭城-宜蘭路段的小客車於平日下午及凌晨時段之車速較慢，大約只有 40 公里，小貨車於下午也只有 52.17 公里，建議可以改於上午或晚間時段通行；頭城-宜蘭-羅東路段的小客車、小貨車於平日晚間之車速也較低，建議可以提早在下午六點以前通過；宜蘭-羅東-蘇澳路段之車輛多為僑速慢速的狀況，平均車速不到 10 公里，建議用路人改道，利用替代道路前往蘇澳。此集群之資料有接近一半屬於假日資料，且凌晨之資料量較少，因此命名為假日凌晨時段之車潮。

表 4.51(集群 1)假日凌晨時段之車潮

平 假 日	車 種	時段	坪林-頭城-宜蘭		頭城-宜蘭-羅東		宜蘭-羅東-蘇澳	
			旅行 時間	時速 (km)	旅行 時間	時速 (km)	旅行 時間	時速 (km)
平 日	大 型 車	上午	112.83	70.19	520	90	1509	13.12
		下午	107	74.02	525	89.14	1605	12.34
		晚間	111	71.35	590	79.32	1149	17.23
	小 客 車	上午	122.88	64.45	658.62	71.06	1952.07	10.14
		下午	179.65	44.09	676.45	69.18	2048.36	9.67
		凌晨	194.54	40.71	781.08	59.92	1901.92	10.41
		晚間	126.11	62.8	1007.59	46.45	1823.68	10.86
	小 貨 車	上午	93.71	84.51	615.81	76	2051.43	9.65
		下午	151.81	52.17	679.65	68.86	2184.29	9.06
		凌晨	96.17	82.36	775.5	60.35	2093.67	9.46
		晚間	125.18	63.27	859.45	54.45	1827.18	10.84

假日	大型車	上午	109.5	72.33	543.25	86.15	1908.5	10.37
		下午	100.75	78.61	541.25	86.47	1993	9.93
		凌晨	96	82.5	478	97.91	1303	15.2
		晚間	105	75.43	496	94.35	1329	14.9
	小客車	上午	115.45	68.6	609.18	76.82	2405.27	8.23
		下午	117.2	67.58	572.86	81.7	2245.02	8.82
		凌晨	81.12	97.63	437.75	106.91	1559.88	12.69
		晚間	96.82	81.8	846.27	55.3	1990.45	9.95
	小貨車	上午	131.5	60.23	564.64	82.88	2513.79	7.88
		下午	95.27	83.13	543.82	86.06	2132.64	9.28
		凌晨	77.5	102.19	380	123.16	1611.5	12.29
		晚間	93	85.16	499	93.79	1594	12.42

表 4.52 為集群 2 分平假日、車種、時段之平均旅行時間及車速，個路段之行駛狀況多為順暢，僅大型車於平日上午之車速較低，時速為 58.53 公里，建議可利用其他時段出發。此集群之資料多為假日白天之資料，因此命名為假日白天之車潮。

表 4.52(集群 2)假日白天之車潮

平假日	車種	時段	坪林-頭城-宜蘭		頭城-宜蘭-羅東		宜蘭-羅東-蘇澳	
			旅行時間	時速(km)	旅行時間	時速(km)	旅行時間	時速(km)
平日	大型車	上午	135.32	58.53	536.64	87.21	243.83	81.2
		下午	111.96	70.74	510.64	91.65	235.93	83.92
		凌晨	100.07	79.14	540.86	86.53	274.64	72.09
		晚間	108.62	72.92	543.95	86.04	258.22	76.68
	小客車	上午	103.72	76.36	497.68	94.04	212.35	93.24
		下午	104.35	75.9	499.25	93.74	220.69	89.72
		凌晨	112.17	70.6	535.27	87.43	241.1	82.12
		晚間	107.37	73.76	527.01	88.8	240.35	82.38
	小貨車	上午	109.36	72.42	505.76	92.53	218.32	90.69
		下午	106.7	74.23	505.48	92.58	220.27	89.89
		凌晨	108.52	72.98	527.21	88.77	228.66	86.59
		晚間	114.75	69.02	545.1	85.86	239.12	82.8

假日	大型車	上午	111.43	71.07	519.58	90.07	242.25	81.73
		下午	100.91	78.49	515.75	90.74	255.69	77.44
		凌晨	100.45	78.84	541.09	86.49	225.64	87.75
		晚間	117.68	67.3	563.82	83.01	264.71	74.8
	小客車	上午	100.82	78.55	491.69	95.18	210.11	94.24
		下午	96.03	82.47	489.89	95.53	213.28	92.84
		凌晨	99.96	79.23	512.06	91.4	226.85	87.28
		晚間	97.29	81.4	516.54	90.6	232.65	85.11
	小貨車	上午	108.23	73.18	498.31	93.92	213.57	92.71
		下午	100.6	78.73	494.9	94.56	214	92.52
		凌晨	90.85	87.18	504.49	92.77	210.16	94.21
		晚間	98.06	80.77	513.86	91.08	226.86	87.28

表 4.53 為集群 3 分平日、車種、時段之平均旅行時間及車速，坪林-頭城-宜蘭路段中，大型車平日建議避開上午時段，此時段平均車速僅 14.36 公里，小貨車建議避開平日下午及晚間時段，小客車建議避開上午及下午時段；頭城-宜蘭-羅東路段則是全線壅塞慢速狀態，建議可改走替代道路；宜蘭-羅東-蘇澳路段亦多為壅塞狀態，大型車建議行駛平日上午及假日凌晨以外之時段、小客車及小貨車建議改走替代道路前往蘇澳。此集群之資料較平均分布於各個時段，及各星期，但由於資料量較低，命名為少數壅塞車潮。

表 4.53(集群 3)少數壅塞車潮

平日	車種	時段	坪林-頭城-宜蘭		頭城-宜蘭-羅東		宜蘭-羅東-蘇澳	
			旅行時間	時速(km)	旅行時間	時速(km)	旅行時間	時速(km)
平日	大型車	上午	551.67	14.36	3091.17	15.14	1395	14.19
	小客車	上午	220.73	35.88	3416.57	13.7	554.93	35.68
		下午	206.07	38.43	3278	14.28	564.29	35.09
		凌晨	137.18	57.74	3637.88	12.86	545.82	36.28
		晚間	147.66	53.64	3690.59	12.68	828.53	23.9
	小貨車	上午	309.77	25.57	3515.69	13.31	842.46	23.5
		下午	314.55	25.18	2752	17.01	359.55	55.07
		凌晨	100	79.2	2960.67	15.81	600.83	32.95
		晚間	162.47	48.75	3193.47	14.65	786.79	25.17

假日	大型車	上午	92	86.09	3118.8	15.01	277.4	71.38
		凌晨	167	47.43	7993	5.86	2904	6.82
		晚間	101	78.42	3593	13.03	243	81.48
	小客車	上午	138.65	57.12	3548.08	13.19	595.88	33.23
		下午	93.87	84.38	2972.27	15.75	530.67	37.31
		凌晨	116.8	67.81	3710.47	12.61	670.6	29.53
		晚間	85.43	92.71	3188.14	14.68	957.86	20.67
	小貨車	上午	96	82.5	3384.33	13.83	1059.67	18.69
		下午	96	82.5	2586.67	18.09	354.67	55.83
		凌晨	91	87.03	2571	18.2	330	60
		晚間	95	83.37	3590	13.04	942	21.02

表 4.54 為頭城-蘇澳路徑資料之集群分析重點整理，將三個集群命名後為各個集群提出行駛建議。

表 4.54 ETC_KB2 各級群命名與用路人建議

集群	建議
1 假日凌晨之車潮	1. 前往蘇澳建議改走替代道路 2. 小貨車平日建議避開下午及凌晨時段 3. 小型車平日建議避開晚間時段通過宜蘭路段
2 假日白天之車潮	1. 普遍順暢 2. 大型車建議避開平日上午行經頭城路段
3 少數壅塞車潮	3.1.特殊狀況建議多聽廣播與路況資訊，避開不併時出現之回堵

圖 4.2 為各路徑集群分析時速圖，不同路徑中，有一個集群是沿線順暢的狀態。在南港-宜蘭、南港-羅東、南港-蘇澳、石碇-羅東與石碇-蘇澳等路徑中，各有一個集群在南港-坪林以及石碇-頭城呈現壅塞狀態，平均車速大約 35 公里至 40 公里，而該集群之資料又多集中在周六上午及下午，判斷為假日出遊之車潮，建議可走替代道路、或是改時間通行南港-頭城路段，以避開壅塞路段。在南港-宜蘭、南港-羅東、石碇-羅東與石碇-蘇澳各有一集群之資料集中在平日上午，在南港-頭城兩個路段之車速僅 10 公里至 15 公里，判斷為上班之車潮，建議一般用路人可避開此平日上午之上班車潮上高速公路，上班之用路人亦可考慮替代道路於下一個交流道上高速公路。在南港-蘇澳與坪林-蘇澳兩路徑資料中有集群呈現在南港-蘇澳路段以及坪林-蘇澳路段異常壅塞，依不同變數分類後亦無法找出特別集中的範圍，建議用路人行駛前多關注路況資訊，以避開壅塞的問題。另外從各個集群中可以發現在晚間與凌晨的大型車輛會有行車速度較慢的狀況，晚間及凌晨時段通過南港路段時應多注意大型車輛，保障自身行車安全。以上行車狀況整理於表 4.55。

圖 4.2 各路徑集群分析時速圖

	南港-石碇	石碇-坪林	坪林-頭城	頭城-宜蘭	宜蘭-羅東	羅東-蘇澳
KA1_1	42.53	37.41	79.96			
KA1_2	74.11	79.61	85.57			
KA1_3	13.57	70.7	81.48			
KA2_1	13.5	71.31	82.31	93.34		
KA2_2	42.33	37.79	79.14	93.24		
KA2_3	73.64	79.46	85.01	94.99		
KA3_1	33.18	38.5	75.85	94.32	92.41	
KA3_2	57.65	71.21	77.11	94.73	8.98	
KA3_3	64.05	78.76	83.24	96.03	94.29	
KB1_1		78.01	84.1	94.56		
KB1_2		10.59	70.46	88.05		
KB1_3		37.8	79.2	92.8		
KB2_1		37.49	75.76	93.7	82.11	
KB2_2		12.39	65.48	87	78.33	
KB2_3		76.81	81.29	95.1	85.57	
KC1_1			60.48	70	9.54	
KC1_2			76.51	93.38	90.71	
KC1_3			45.26	13.64	29.52	

問題狀態	建議
假日白天車潮	南港端兩路段於上午時段車多壅塞，建議可於 12 點後通過此路段。
上班時間車潮	非上班之用路人可避免上午時段行駛南港路段。
大型車慢速行駛之狀況	清晨及晚間多注意大型車行駛狀況，保障行車安全
特殊壅塞路段(如：蘇澳)	建議多參考路況資訊，避開壅塞狀態。

表 4.55 集群分析彙整表

第三節、RStudio Server 之多核心計算

將 K-means 的運算過程以巨集包裝，將此巨集透過 Parallel 與 foreach 套件中的函數 mclapply、parLapply 與 foreach 進行不同核心的運算，最後將多核心運算的程式碼以 system.time() 函數包覆，用以計算所花費的時間，詳細的程式碼請參考 P.99 附錄(二)，從運算時間比較中發現以下結果：

3.1.lapply 方法之程式碼與運算結果(以四核心運算之程式碼為例)

透過 lapply 函數以及 parallel 套件中的 mclapply 與 parLapply 函數進行比較，程式碼如表 4.56，將集群分析 K-means 方法的程式碼以巨集包裝後，分別使用 lapply、mclapply 與 parLapply 進行運算，並以 system.time() 函數包覆運算過程以計算花費時間。

表 4. 56 lapply 方法與 parallel 套件之程式碼

名稱	程式碼
K-Means 巨集	1. ETC.cluster <- function(data, N) { set.seed(278613) K3=kmeans(x=data,centers=N) return(c('iter','totss')) }
lapply	2. system.time(lapply(FUN = function(x) { ETC.cluster(etc1, 3)}))
mclapply	3. system.time(mclapply(FUN = function(x) { ETC.cluster(etc1, 3)},mc.cores =4))
parLapply	4. cores <- detectCores(logical = FALSE)
	5. cl <- makeCluster(cores)
	6. xx<-3
	7. clusterExport(cl, c('ETC.cluster', 'etc1', 'xx'))
	8. system.time(parLapply(cl, function(x) {ETC.cluster(etc1, 3) }))

表 4.57 為 lapply 方法之多核心計算效能比較表，從運算時間結果可以發現不論哪一種多核心計算的方法都比單核心計算省時，使用 mclapply 函數在雙核心的環境下可以省下 48.8%至 52.4%的時間、三核心的環境下可節省 64.4%至 67.8%的運算時間、四核心的環境可節省 70.8%至 72.8%的運算時間；從不同資料量的方面可以發現在 20G 以上的資料量所節省的比例較 10G 多。

表 4.57 lapply 方法之多核心計算效能比較表(單位：秒(省時比例))

函數	核心數	10G 資料	20G 資料	25G 資料	30G 資料
lapply	1	412.88	874.75	1094.9	1318.95
mclapply	2	211.39 (48.8%)	422.4 (51.7%)	520.85 (52.4%)	631.46 (52.1%)
	3	147.18 (64.4%)	282.2 (67.7%)	360.54 (67.1%)	424.35 (67.8%)
	4	120.65 (70.8%)	241.54 (72.4%)	297.73 (72.8%)	362.28 (72.5%)
parLapply	2	240.24 (41.8%)	435.33 (50.2%)	625.44 (42.9%)	712.46 (46%)
	3	165.7 (59.9%)	312.78 (64.2%)	426.41 (61.1%)	505.77 (61.7%)
	4	154.58 (62.6%)	256.58 (70.7%)	365.03 (66.7%)	410.2 (68.9%)

3.2.for 迴圈之程式碼與運算結果

透過 for 迴圈以及 foreach 套件的函數進行比較，程式碼如表 4.58，將集群分析 K-means 方法的程式碼以巨集包裝後，使用 for 迴圈與 foreach 迴圈進行運算，並以 system.time()函數包覆運算過程以計算花費時間。

表 4. 58 lapply 方法與 parallel 套件之程式碼

名稱	程式碼
K-Means 巨集	<pre> 1. ETC.cluster <- function(data, N) { set.seed(278613) K3=kmeans(x=data,centers=N) return(c('iter','totss')) }</pre>
for 迴圈	<pre> 2. res2.s <- matrix(0, nrow=len10, ncol = 2) 3. system.time(for(i in 1:len10) {res2.s[i,] <- ETC.cluster(etc1, 3)})</pre>
foreach 迴圈	4. cores <- detectCores(logical=F)
	5. cl <- makeCluster(cores)
	6. registerDoParallel(cl, cores=cores)
	7. chunk.size <- len10/cores
	<pre> 8. system.time(foreach(i=1:cores, .combine='rbind') %dopar%{ res <- matrix(0, nrow=chunk.size, ncol=2) for(x in ((i-1)*chunk.size+1):(i*chunk.size)) {res[x - (i-1)*chunk.size,] <-ETC.cluster(etc1, 3)} res})</pre>
	9. stopImplicitCluster()
	10. stopCluster(cl)

表 4.59 為 for 迴圈方法之多核心計算效能比較表，從運算時間結果可以發現不論哪一種多核心計算的方法都比單核心計算省時，雙核心的環境可節省 45.7 至 46.6% 的運算時間，不同資料量間所節省的運算時間比例差距不超過 1%；三核心的環境可節省 45.8% 至 48% 的運算時間，20G 的資料量在此環境所節省的時間較少，25G 及 30G 的資料量在此環境所節省的時間最多；四核心的環境可節省 64.2 至 69% 的運算時間，10G 的資料量在此環境所節省的時間較少，20G 以上的環境所節省的運算時間較多。另外也可以發現雙核心與三核心的環境所節省的運算時間差異較小。

表 4.59 for 迴圈方法之多核心計算效能比較表(單位：秒(省時比例))

函數	核心數	10G 資料	20G 資料	25G 資料	30G 資料
For 迴圈	1	406.79	811.8	1007	1200.85
Foreach 迴圈	2	220.99 (45.7%)	433.28 (46.6%)	542.01 (46.2%)	645.28 (46.3%)
	3	214.13 (47.4%)	439.7 (45.8%)	524.43 (47.9%)	623.96 (48%)
	4	145.8 (64.2%)	251.97 (69%)	322.67 (68%)	362.19 (69.8%)

透過使用不同的資料量進行分析時間的比較後，單核心運算的兩個方法中 (lapply、for 迴圈)，運用 for 迴圈所花費的運算時間較 lapply 的低，但沒有明顯的差異；若使用多核心計算功能則明顯省下很多的運算時間。雙核心運算下使用 mclapply 與 Foreach 函數可以節省 45% 左右的運算時間、使用 parLapply 函數可以節省 40% 左右的運算時間。三核心運算的部分在使用 mclapply 與 parLapply 函數可以節省 60% 以上的運算時間、使用 Foreach 函數可以節省 45% 以上的運算時間。四核心運算使用 mclapply 函數可以節省 70% 到 75% 的運算時間、使用 parLapply 與 Foreach 可以節省 65% 到 70% 的運算時間。

綜合以上結果，使用 mclapply 函數所節省的運算時間最多，但由於 mclapply 函數僅限於 Linux 系統中使用，在 Windows 系統或 Mac OS 系統中若要使用多核心計算，可依照電腦硬體規格做選取，雙核心使用 Foreach 函數、三核心使用 parLapply 函數、四核心使用 parLapply 或 Foreach 函數均可。

圖 4.3 與圖 4.4 為 mclapply 函數及 parLapply 函數的運算時間比較折線圖，可以明顯看出多核心運算所需花費的運算時間明顯較單核心低，雙核心運算可節省 50% 左右的時間；三核心與四核心的運算時間差異較小，平均可節省 60% 到 70% 的時間。若礙於硬體限制，可選擇使用三核心運算，其效能接近四核心運算。

圖 4.3 mclapply 運算時間比較折線圖

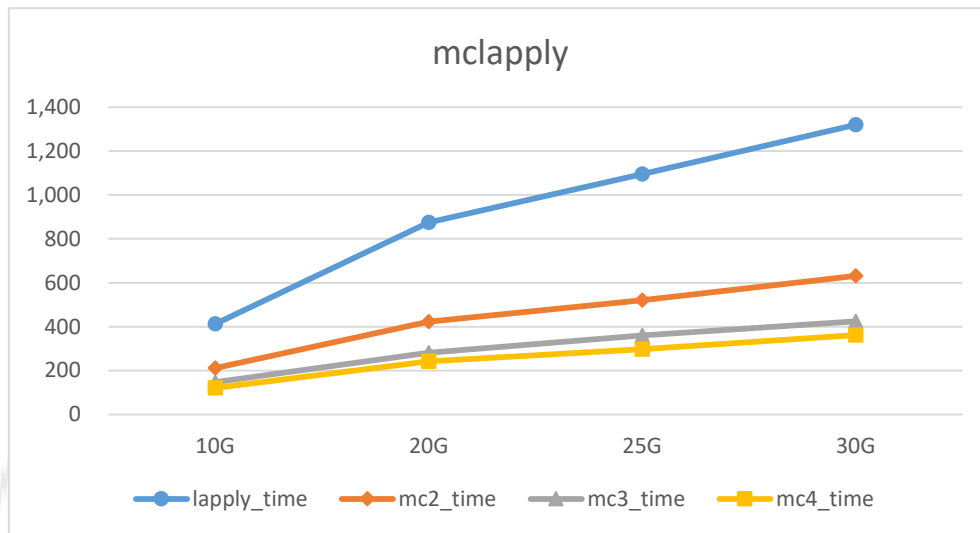


圖 4.4 parLapply 運算時間比較折線圖

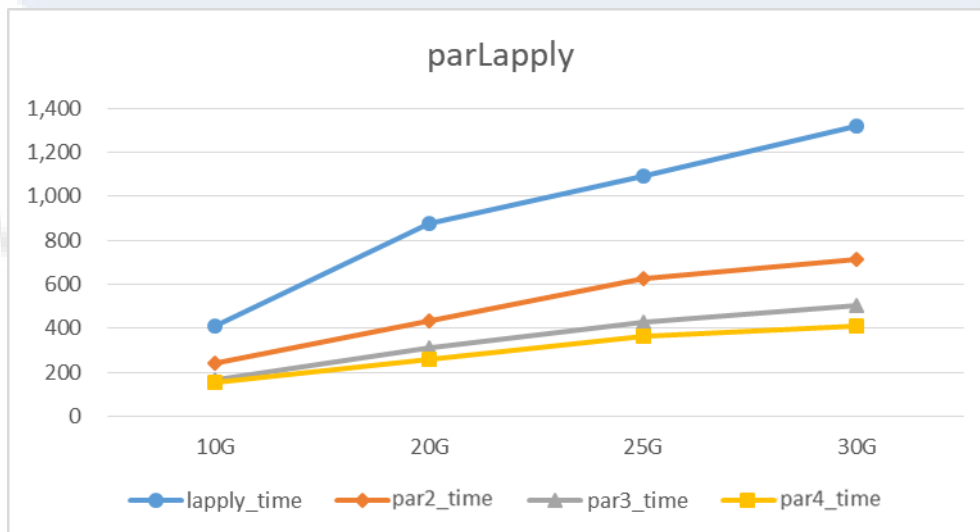
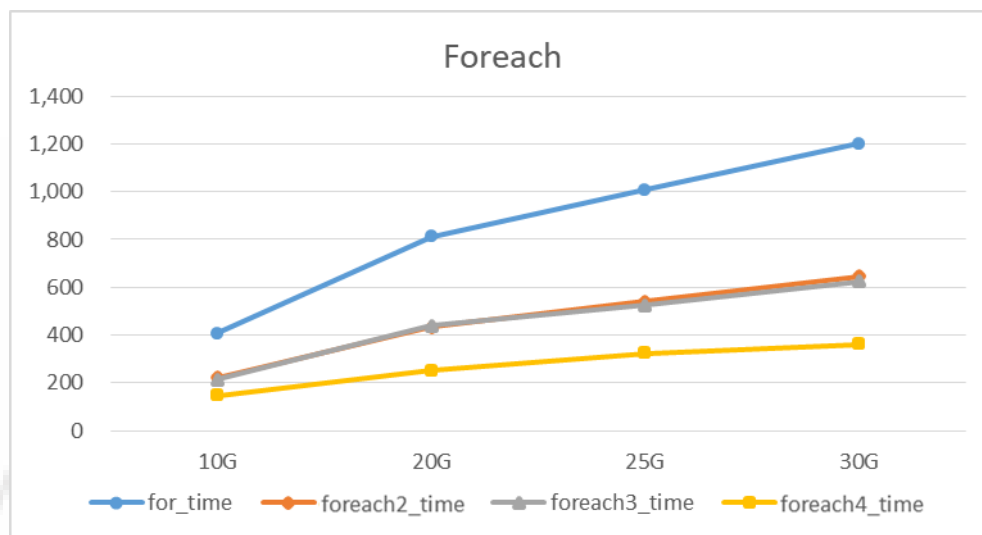


圖 4.5 為 Foreach 函數的運算時間比較折線圖，顯示多核心運算所需花費的運算時間明顯較單核心低，雙核心及三核心運算沒有太大的差異，均可節省 45% 左右的時間，硬體若無法提供到四核心，使用 Foreach 函數可以用雙核心運算，可達到與三核心一樣的效能；四核心則可節省近 70% 的運算時間。

圖 4.5 Foreach 運算時間比較折線圖



第五章、結論與建議

綜合前四章之成果，將結論分為針對系統架構與效能、針對資料分析兩個部分，最後是研究限制與未來研究之建議。

第一節、系統架構與效能

本研究於 Hortonworks 資料平台上進行資料的篩選過濾，並將資料匯入與平台結合之統計分析軟體 RStudio-Server 進行進階統計分析與資料探勘，透過伺服器版本之 R 語言多核心計算功能提升了資料的運算效能，不同的函數在不同環境下節省的效能不同，使用 mclapply 函數是最有效率的方法，但僅限於 Linux 系統使用，雙核心的環境可節省 50% 的運算時間，三核心及四核心可節省 65-75% 的運算時間，而不同資料量所節省的時間沒有太大差異；parLapply 函數的優勢在於可於 Windows 系統下使用，但所節省之效能較 mclapply 少雙核心可節省 45% 左右之效能，三核心可節省 60% 左右之效能、四核心可節省 65 到 70% 的時間，且在 20G 的資料運算中節省的時間最多；Foreach 迴圈的部分亦可作用於 Windows 系統，雙核心與三核心所節省的時間相近，硬體如有限制可使用雙核心運算即可，四核心的效能與 parLapply 函數在四核心環境的差異不大，不同資料量的運算時間亦沒有明顯差異。

第二節、資料分析

透過 K-Means 方法將六個旅行資料個別分為三群進行比較後，各個行駛資料中皆發現以下幾個共同點：

1. 星期六上午是車多壅塞的主要時段，這個時段應避免行駛南港-頭城路段，透過替代道路避開車潮前往宜蘭地區。
2. 平日上午於南港路段多為塞車的狀況，上班的用路人可考慮提早出，避開上班尖峰時間。
3. 部分大客車於晚間或清晨行駛車速較慢，行經南港-坪林路段時應多注意大車，保障行駛安全。
4. 雖資料可歸納出大致的行駛狀態，探高速公路上之變數較大，上路前應多留意行車資訊及廣播，避免遭遇突發的壅塞情況。

第三節、 研究限制

本研究中所遇到之限制包含平台架構與資料分析兩部分：

I. 平台架構

1. 操作雲端分析平台須具備資訊工程之相關技術，對 Java、Linux 等語言較不熟悉者，會花費較多時間在熟悉系統環境的操作上。
2. Hortonworks Sandbox 為一整合完全之作業系統，欲進行軟體的修改及套件的架設會有較多限制。
3. RStudio server 連接 Hadoop 之相關套件在微軟買向 R 語言之後暫時停止更新，部分功能與新版 R 語言會有衝突。

II. 資料分析

1. ETC 自 2014 年初啟用後之資料有部分遺漏，2015 年 2 月起才收集得到完整資料。
2. 行駛資料易受到受氣候、政策或車輛故障之影響，造成分群上的困難。

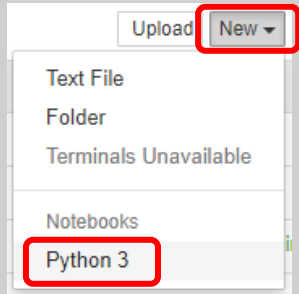
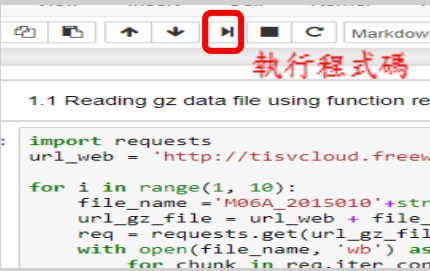
第四節、 未來研究之建議

在平台架構前需多熟悉 Java、Linux 等程式語言，遇到系統問題時較容易解決，另外可考慮利用乾淨之 Linux 作業系統架設 Hadoop 套件與 RStudio Server，對系統路徑及安裝步驟會更清楚。

對 ETC 資料的部分可考慮加入更多變數如氣候、颱風假、政府政策等變數，並探討故障車輛與離群值(飆車或龜速車)的處理方式，可將行駛資料作更精確之分類。

附錄(一)-Anaconda 安裝與 Python 程式碼解說

	<p>網址：</p> <p>https://www.continuum.io/</p> <p>進入官方網站後，點選右上角 Download 進入下載網頁</p>
	<p>依照自身作業系統版本選擇下載連結，本研究所使用的是 Windows 作業系統下之軟體</p>
	<p>選擇下載所需要的版本之 Anaconda 並安裝於電腦中，本研究使用之 Python 版本為 2.7</p>
	<p>開啟命令提示字元，輸入 ipython notebook，稍等一段時間便啟動成功</p>
	<p>啟動成功後，系統會自動開啟瀏覽器便連線至 Anaconda 的操作介面，亦可自行輸入網址 http://localhost:8888</p>

	<p>進入指定目錄後，點選畫面右上角的 NEW 按鈕，選擇 notebooks 的 Python3 即可開啟新的編輯簿</p>
	<p>在開啟之編輯簿中輸入 Python 程式碼，按下執行按鈕便可執行所輸入之內容，如有進行資料下載便會將資料存放至原先編輯簿所存放之路徑，本附錄以 2015 年 1 月為例。</p>
<pre> 1. import requests, 2. url_web = 'http://tisvcloud.freeway.gov.tw/history/TDCS/M06A/' 3. for i in range(1, 10):, 4. file_name = 'M06A_2015010'+str(i)+'.tar.gz', 5. url_gz_file = url_web + file_name, 6. req = requests.get(url_gz_file, stream=True), 7. with open(file_name, 'wb') as f:, 8. for chunk in req.iter_content(chunk_size=1024):, 9. if chunk:, 10. f.write(chunk), 11. f.flush(), 12. for i in range(10, 32):, 13. file_name = 'M06A_201501'+str(i)+'.tar.gz', 14. url_gz_file = url_web + file_name, 15. req = requests.get(url_gz_file, stream=True), 16. with open(file_name, 'wb') as f:, 17. for chunk in req.iter_content(chunk_size=1024):, 18. if chunk:, 19. f.write(chunk), 20. f.flush(), </pre>	<ol style="list-style-type: none"> 載入 requests 套件 將網址指派給 url_web 設定迴圈範圍(1 日-9 日) 檔案名稱設定 合併網址與檔案名稱 透過 requests 函數下載檔案 開啟檔案並指定名稱 f 設定寫入格式 寫入資料 寫入資料 刷新暫存空間 設定迴圈範圍(10 日-31 日) 檔案名稱設定 合併網址與檔案名稱 透過 requests 函數下載檔案 開啟檔案並指定名稱 f 設定寫入格式 寫入資料 寫入資料 刷新暫存空間 <p>[註] range(a,b) 第 3 行與第 12 行之迴圈表示抓取日期為 a 日至 b-1 日，其中 a<b。</p>

附錄(二)-多核心計算時間比較之程式碼解說

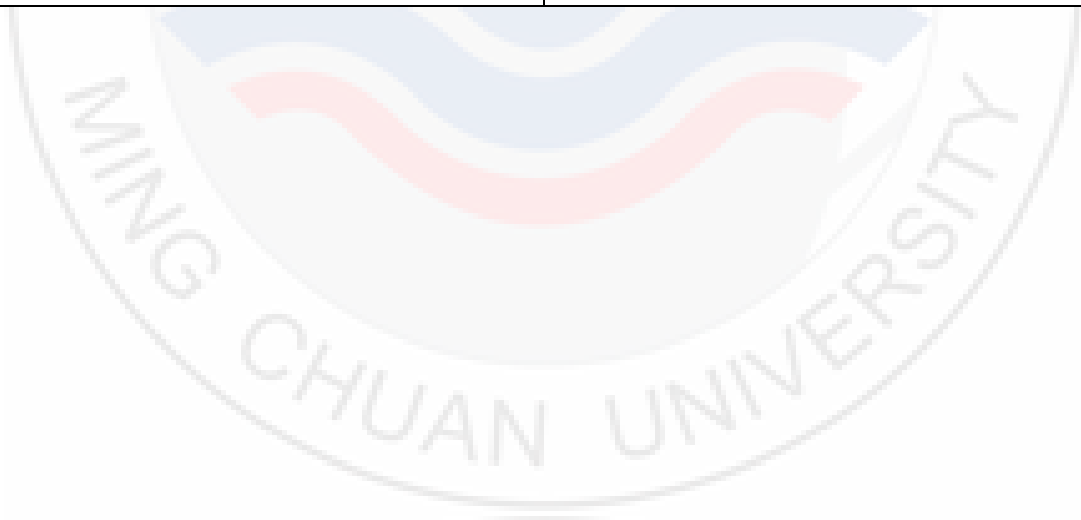
載入多核心計算所需之相關套件	11. library(parallel) 12. library(foreach) 13. library(doParallel)	載入多核心計算所需之相關套件
lapply 方法之單核心計算	14. system.time(lapply(FUN = function(x) { ETC.cluster(etc1, 3)}))	lapply 方法之單核心計算
mclapply 方法之多核心計算	15. system.time(mclapply(FUN = function(x) { ETC.cluster(etc1, 3)},mc.cores =4))	將集群分析之函數插入平行計算之函數中，mc.cores 為核心數量的設定參數。並計算所花費之時間
parLapply 方法之多核心計算	16. cores <- detectCores(logical = FALSE)	用 detectCores 確定當前硬體的內核數目
	17. cl <- makeCluster(cores)	建立多核心計算用之核心數量
	18. xx<-3	設定集群分群數 K
	19. clusterExport(cl, c('ETC.cluster', 'etc1', 'xx'))	利用 clusterExport 把計算所需的資料和函數分配給各個計算組
	20. system.time(parLapply(cl, function(x) {ETC.cluster(etc1, 3) })))	parLapply 將計算平均分配給計算組，收集合併結果並計算所花費之時間
for 迴圈之單核心計算	21. res2.s <- matrix(0, nrow=len10, ncol = 2) 22. system.time(for(i in 1:len10) {res2.s[i,] <- ETC.cluster(etc1, 3)}))	for 迴圈知單核心計算
foreach 迴圈之多核心計算	23. cores <- detectCores(logical=F)	用 detectCores 確定當前硬體的內核數目
	24. cl <- makeCluster(cores)	建立多核心計算用之核心數量
	25. registerDoParallel(cl, cores=cores)	使用 registerDoParallel 來設定 foreach 後端的計算方式
	26. chunk.size <- len10/cores	將向量均勻分配到各個計算組
	27. system.time(foreach(i=1:cores, .combine='rbind') %dopar%{ res <- matrix(0, nrow=chunk.size, ncol=2) for(x in ((i-1)*chunk.size+1):(i*chunk.size)) {res[x - (i-1)*chunk.size,] <-ETC.cluster(etc1, 3)} res})	透過 foreach 函數進行多核心計算，並計算所花費之時間
	28. stopImplicitCluster() 29. stopCluster(cl)	停用多核心模式

附錄(三)-RStudio Server 集群分析程式碼

<pre> 1. date<-ETC\$starttime 2. date<-as.POSIXct(date, format="%Y/%m/%d %H:%M") 3. year<-year(date) 4. month<-month(date) 5. day<-day(date) 6. weekday<-wday(date,label=T) 7. timestamp<-hour(date) 8. working<-hour(date) 9. vacation<-wday(date,label=T) 10. vacation<-gsub('Sun',"假日",vacation) 11. vacation<-gsub("Mon","平日",vacation) 12. timestamp<-gsub(23,"晚間",timestamp) 13. timestamp<-gsub(12,"下午",timestamp) 14. timestamp<-gsub(11,"上午",timestamp) 15. timestamp<-gsub(5,"凌晨",timestamp) 16. working<-gsub(23,"離峰",working) 17. working<-gsub(18,"尖峰",working) 18. cartype2<-ETC\$cartype 19. cartype2<-gsub('31',"小客車",cartype2) 20. cartype2<-gsub('32',"小貨車",cartype2) 21. cartype2<-gsub('41',"大型車",cartype2) 22. cartype2<-gsub('42',"大型車",cartype2) 23. cartype2<-gsub('5',"大型車",cartype2) 24. ETC<- cbind(date,year,month,day,weekday,hour,working,vac ation,timestamp,cartype2,ETC) </pre>	<pre> 1. ETC 資料表為本研究所使用之原始整理後，挑選 上路時間 starttime 進行計算，並存於 date 2. 定義 date 資料格式為日期時間格式 3. 定義年資料 4. 定義月資料 5. 定義日資料 6. 定義星期資料 7. 定義時段用之小時資料 8. 定義尖峰非尖峰用之小時資料 9. 定義平假日用之星期資料 10. 定義假日資料 11. 定義平日資料 12. 定義晚間時段資料 13. 定義下午時段資料 14. 定義上午時段資料 15. 定義凌晨時段資料 16. 定義離峰時段資料 17. 定義尖峰時段資料 18. 挑選車種 cartype 進行中文定義，並存於 cartype2 19. 定義各種車種之中文名稱 20. 定義各種車種之中文名稱 21. 定義各種車種之中文名稱 22. 定義各種車種之中文名稱 23. 定義各種車種之中文名稱 24. 將上述變數與原始資料合併 </pre>
<pre> 25. aa<-as.numeric(difftime(strptime(ETC\$S0055, "%Y- %m-%d %H:%M:%S"),strptime(ETC\$S0000, "%Y- %m-%d %H:%M:%S"),units='secs')) 26. bb<-as.numeric(difftime(strptime(ETC\$S0287, "%Y- %m-%d %H:%M:%S"),strptime(ETC\$S0055, "%Y- %m-%d %H:%M:%S"),units='secs')) 27. cc<-as.numeric(difftime(strptime(ETC\$S0309, "%Y- %m-%d %H:%M:%S"),strptime(ETC\$S0287, "%Y- %m-%d %H:%M:%S"),units='secs')) </pre>	<pre> 計算五個路段之行駛時間 </pre>

28. <code>dd<-as.numeric(difftime(strptime(ETC\$S0439, "%Y-%m-%d %H:%M:%S"),strptime(ETC\$S0309, "%Y-%m-%d %H:%M:%S"),units='secs'))</code> 29. <code>ee<-as.numeric(difftime(strptime(ETC\$S0494, "%Y-%m-%d %H:%M:%S"),strptime(ETC\$S0439, "%Y-%m-%d %H:%M:%S"),units='secs'))</code>	
30. <code>ETC_time<-cbind(ETC,aa,bb,cc,dd,ee)</code>	將行駛時間與原始資料合併
31. <code>ETC_A1<-subset(ETC_time, startgate=="05F0000S" & endgate=="05F0309S",select=-c(16,17,21,22))</code> 32. <code>ETC_A2<-subset(ETC_time, startgate=="05F0000S" & endgate=="05F0439S",select=-c(17,22))</code> 33. <code>ETC_A3<-subset(ETC_time, startgate=="05F0000S" & endgate=="05F0494S")</code> 34. <code>ETC_B1<-subset(ETC_time, startgate=="05F0055S" & endgate=="05F0439S",select=-c(12,17,18,22))</code> 35. <code>ETC_B2<-subset(ETC_time, startgate=="05F0055S" & endgate=="05F0494S",select=-c(12,18))</code> 36. <code>ETC_C1<-subset(ETC_time, startgate=="05F0287S" & endgate=="05F0494S",select=-c(12,13,18,19))</code>	將資料分為六大路徑資料，並刪除多餘變數
37. <code>ETC_A1<-ETC_A1[complete.cases(ETC_A1),]</code> 38. <code>ETC_A2<-ETC_A2[complete.cases(ETC_A2),]</code> 39. <code>ETC_A3<-ETC_A3[complete.cases(ETC_A3),]</code> 40. <code>ETC_B1<-ETC_B1[complete.cases(ETC_B1),]</code> 41. <code>ETC_B2<-ETC_B2[complete.cases(ETC_B2),]</code> 42. <code>ETC_C1<-ETC_C1[complete.cases(ETC_C1),]</code>	剔除有遺漏值之資料
43. <code>ETC_A11<-ETC_A1[complete.cases(ETC_A1),16:18]</code> 44. <code>ETC_A22<-ETC_A2[complete.cases(ETC_A2),17:20]</code> 45. <code>ETC_A33<-ETC_A3[complete.cases(ETC_A3),18:22]</code> 46. <code>ETC_B11<-ETC_B1[complete.cases(ETC_B1),16:18]</code> 47. <code>ETC_B22<-ETC_B2[complete.cases(ETC_B2),17:20]</code> 48. <code>ETC_C11<-ETC_C1[complete.cases(ETC_C1),16:18]</code>	篩選出旅行時間資料，以便於進行集群分析
49. <code>K=3</code> 50. <code>A11_K3 = kmeans(x = ETC_A11, centers = K)</code> 51. <code>A22_K3 = kmeans(x = ETC_A22, centers = K)</code> 52. <code>A33_K3 = kmeans(x = ETC_A33, centers = K)</code> 53. <code>B11_K3 = kmeans(x = ETC_B11, centers = K)</code> 54. <code>B22_K3 = kmeans(x = ETC_B22, centers = K)</code> 55. <code>C11_K3 = kmeans(x = ETC_C11, centers = K)</code>	將六個路徑資料分別進行 K-means 集群分析分為三群

<pre> 56. cluster<-A11_K3\$cluster 57. ETC_KA1<-cbind(ETC_A1,cluster) 58. cluster<-A22_K3\$cluster 59. ETC_KA2<-cbind(ETC_A2,cluster) 60. cluster<-A33_K3\$cluster 61. ETC_KA3<-cbind(ETC_A3,cluster) 62. cluster<-B11_K3\$cluster 63. ETC_KB1<-cbind(ETC_B1,cluster) 64. cluster<-B22_K3\$cluster 65. ETC_KB2<-cbind(ETC_B2,cluster) 66. cluster<-C11_K3\$cluster 67. ETC_KC1<-cbind(ETC_C1,cluster) 68. write.csv(ETC_KA1,file="ETC_KA1.csv") 69. write.csv(ETC_KA2,file="ETC_KA2.csv") 70. write.csv(ETC_KA3,file="ETC_KA3.csv") 71. write.csv(ETC_KB1,file="ETC_KB1.csv") 72. write.csv(ETC_KB2,file="ETC_KB2.csv") 73. write.csv(ETC_KC1,file="ETC_KC1.csv") 74. write.csv(ETC_time,file="ETC_time.csv") 75. write.csv(ETC,file="ETC.csv") </pre>	<p>將集群分析之分群結果 cluster 挑選出來，並與六大路徑資料合併</p>
--	---



參考文獻

- [1] 交通部國道高速公路局(民 104)。交通部臺灣區國道高速公路局交通資料庫(原始數據)。取自 <http://tisvcloud.freeway.gov.tw/>
- [2] 吳木富(民 104)。**ETC 資料分析應用於高速公路管理之案例分享**。大數據(Big Data)分析在交通管理與服務應用研討會，交通部運輸研究所。
- [3] 汪進財、邱孟佑(民 99)。以車流狀態為基礎之高速公路旅行時間預測模式。**運輸學刊**，3(22)，261-284。
- [4] 邱孟佑(民 99)。**以交通狀態為基礎之旅行時間預測**(博士論文)。國立交通大學，新竹市。
- [5] 李穎(民 96)。**考慮事故延續時間動態更新預測之高速公路旅行時間預測模式建立與資料簡化方法比較**(博士論文)。國立成功大學，台南市。
- [6] 邱逸彥(民 103)。**利用貝氏理論於旅行時間推估之研究**(碩士論文)。國立交通大學，新竹市。
- [7] 林志弘、翁玉麟、曾淑峰(民 101)。資料探勘應用：以 SAS Enterprise Miner 為工具。臺北市：梅霖文化事業股份有限公司。
- [8] 林士傑、李穎、魏建宏(民 92)。高速公路客運車輛旅行時間預測之實證評析。**運輸計劃季刊**，4(32)，651-679。
- [9] 梁衍忠(民 101)。資料探勘技術在學生國語學習成就之分析運用。**教育與多元文化研究**，7，65-110。
- [10] 張效娟(2008)。k-means 算法的改進和在保險業 CRM 中的應用。**大連交通大學學報**。29(1)，76 - 79
- [11] 高諒亨(民 102)。**雲端心電圖大量資料的分析與顯示**(碩士論文)。義守大學，高雄市。
- [12] 蔡宗諺、楊志平、彭克仲、陳仕偉(民 98)。應用資料探勘技術於農會超市之研究—以恆春鎮農會超市為例。**台灣農學會報**，10(4)，272-285。
- [13] 陳垂呈、利益多(民 91)。應用資料探勘技術發掘最適性之線上拍賣競標者。**運籌研究集刊**，1，1-14。
- [14] 廖知航(民 104)。**RHadoop 技術探討與實作**(碩士論文)。靜宜大學，台中市。
- [15] 趙建東、王浩、劉文輝、白繼根(2013)。基于收费数据的高速公路站间旅行时间预测。**同濟大學學報(自然學版)**。41(12)，1859-1854。
- [16] 顏嘉惠(民 91)。資料探勘於圖書館行銷及顧客關係管理之應用。**圖書與資訊學刊**，42，58-68。
- [17] Smith, B L & Demetsky, M J(1994) SHORT-TERM TRAFFIC FLOW PREDICTION: NEURAL NETWORK APPROACH. *Transportation research record*, 1453, 98-104.
- [18] Rakesh Kumar, Neha Gupta, Shilpi Charu, Somya Bansal & Kusum Yadav(2014) Comparison of SQL with HiveQL. *International Journal for Research in Technological Studies*| Vol. 1, Issue 9, August 2014 | ISSN : 2348-1439