TWCC 臺灣AI雲
TAIWAN COMPUTING CLOUD

全功能且自動化的人工智慧整合平台

NCHC

分析大師 (DAS: Data Analysis Service)

全功能且自動化的人工智慧整合平台結合資料收集、治理、分析、部署一系列功能

1. Business Value

2. Solution Overview

Ladder to AI

部署 **Infuse** – Deploy trusted AI-driven business processes

- Manage models across different stages
- Deploy models and scale for different use cases
- Monitor model and automatically retraining and redeployment
- **Infuse trusted and governed insights**

分析 **Analyze** - Scale insights with ML everywhere

- Understand and prepare data for analytics
- Build descriptive, predictive & prescriptive models
- Train at scale with support for distributed compute and GPUs
- **Analyze insights on demand**

治理 **Organize** – Create a trusted analytics foundation

- Find, Catalog, mask data
- Establish and enforce business policies on data
- Advanced transformation capabilities
- **Organize data so that it can be trusted**

收集 **Collect** – Make data simple & accessible

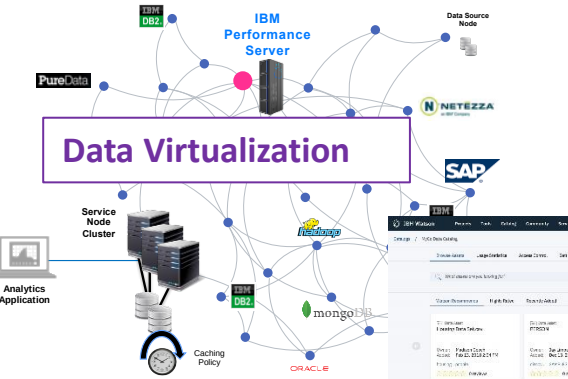
- Collect all Sources of Data
- Connect and discover content from data sources
- Provision new repositories as needed
- **Make it simple & accessible**

Strong Foundation – Built on “OpenShift”

DAS資料科學部署流程

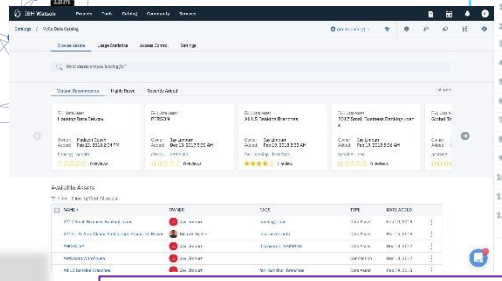
COLLECT ORGANIZE ANALYZE INFUSE

Collect, Connect and Access Data	Govern, Search, and Find Data	Understand and Prepare Data for Analysis	Build Descriptive, Predictive, and Prescriptive Models	Model Management and Deployment	Create Analytics Applications



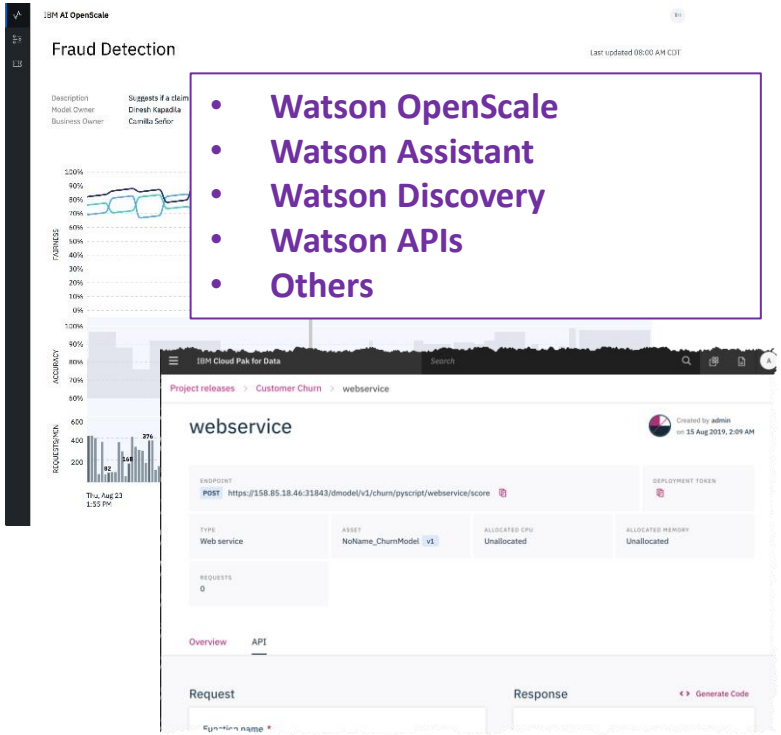
Data Virtualization

- Next Gen Appliances
- Hadoop (Cloudera)
- Db2
- Others



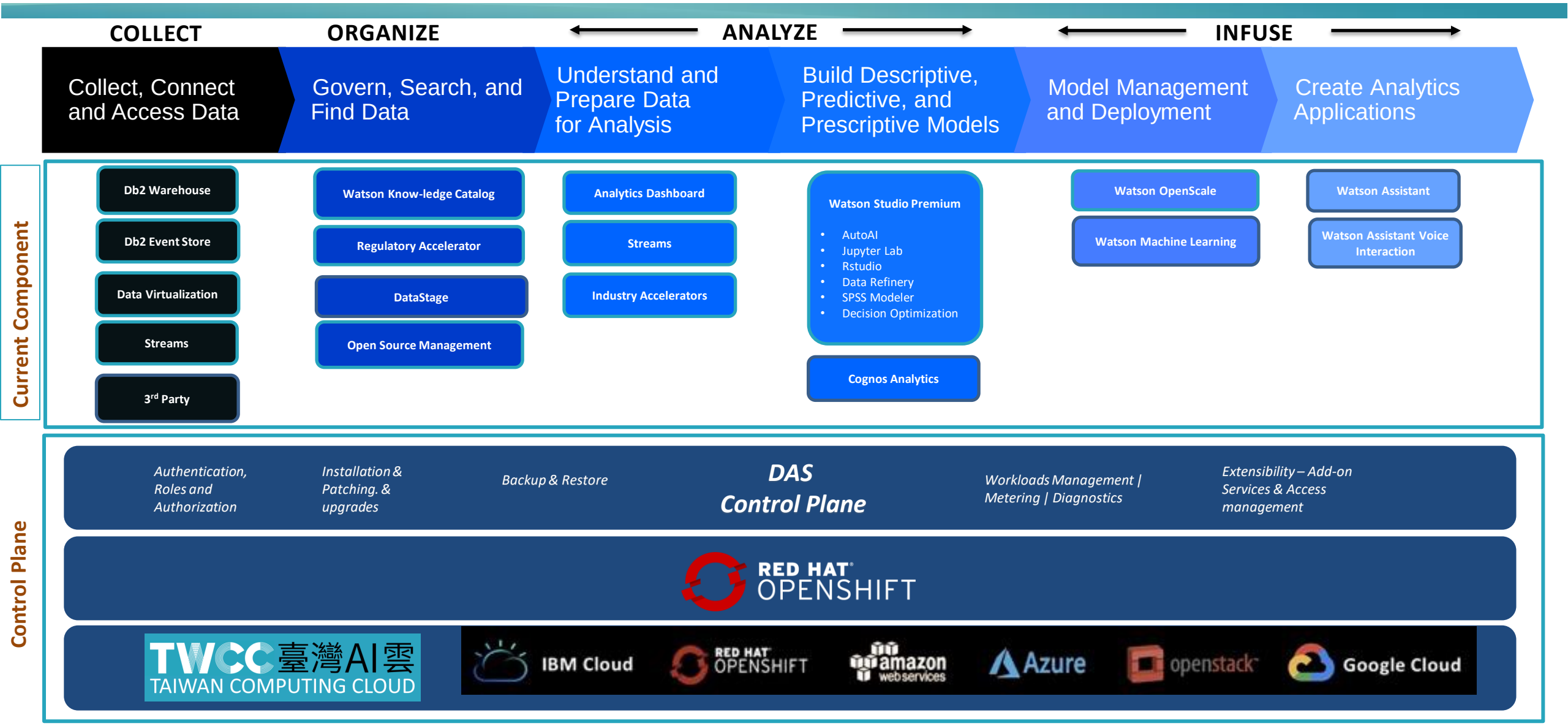
- Watson Knowledge Catalog
- Data Profiling

- Watson Studio + AutoAI
- Watson Machine Learning
- Cognos Analytics
- SPSS Modeler Flow



- Watson OpenScale
- Watson Assistant
- Watson Discovery
- Watson APIs
- Others

DAS底層設計與功能架構



DAS在資料科學上的循環架構與價值

資料推論與監控

在DAS的環境裡，可以透過部署空間的環境來取用API的方式使用模型做資料的推論，另外可以使用分析儀表板來監控資料的狀態和模型推論的成果

資料分析

DAS提供了notebook、RStudio、modeler flow、AutoAI等工具可以依需求進行資料分析建立模型，並可設定排程自動化建模

資料清理

通常在資料分析的專案中，資料清理大概佔了80%的工作時間，因此DAS提供Data Refinery這個工具，減少程式的撰寫並且提供pipeline的模式進行清理流程的編輯

資料收集

提供豐富的資料串接工具與介面，包含MySQL、Oracle、DB2、S3、Dropbox等等資料庫與雲端儲存都可以彙整於DAS的環境中

資料治理

DAS除了提供良好的資料上傳介面外，同時也提供了高效率的資料管理介面，包含資料的使用權限、資料的健康度、資料的類別屬性都可以進行統籌管理。

DAS在國網中心不同層面上的架構與角色

產業應用

可應用於金融業異常交易偵測、工業IOT製造異常分析、智慧交通流量預測等等AI與大資料應用場域，提升產業競爭力

分析大師

整合IBM與其他軟體工具，在雲端發揮資料收集、治理、分析、部署的價值，提供全功能且自動化的人工智慧整合工具

資料科學工具組

從資料收集、資料治理、資料清理、資料分析、模型分析部署都有相對應的軟體設置，可整合內外部的資料與相關人員，提高資料整合的價值

硬體層

完整串接 IaaS、PaaS，整合基礎計算設施、使用平台，自助設定資源；支援 Docker、Singularity 等容器技術，並透過 Kubernetes、Slurm 進行資源管理

DAS在推廣上的目標與預期成果

目標	關鍵成果
贏得客戶滿意	透過建置之DAS服務，客戶可以透過這個平台建立自己的資料科學團隊，有效的利用與分享資料和模型，同時完成模型建置與部署，加速公司內部的AI自動化能量
提昇產業效益	AI產業化與產業AI化： 建立產業內部資料、模型的彙整環境，加強產業應用資料開發模型與部署的能量及效率
鼓勵創新研發	藉由這個平台，中大型的公司可以彙整其內部的資料與AI量能，創造其內部的資料活化與加值應用
創造社會效益	中心藉由提供這個環境的過程，可以參與並且協助規劃相關AI與大數據產出，讓企業更關注資料分析與團隊合作
增加團隊合作	這項服務需要大量內部跨團隊的合作，同時提供服務後，企業本身也可以整合其內部資料科學的能量，企業與企業間也能夠有一個共有與共同合作的平台

組成內外部的DAS團隊

COLLECT ORGANIZE ANALYZE INFUSE

Collect, Connect and Access Data	Govern, Search, and Find Data	Understand and Prepare Data for Analysis	Build Descriptive, Predictive, Prescriptive and Planning Models	Model Management and Deployment	Create Analytics Applications

Amy Collect

- I know where the data is and can securely expose it for others to use
- I can make a single view of the data
- I can deploy a data sandbox if needed

Joe Organize

- I can catalog the data definitions
- I think a data dictionary would help
- Shop for Data… I can make that happen

Suzie Analyze

- I'm just going to whip up the best model ever ☺.
- Wish I could shop for my data?
- How do I connect to data sources I don't know exist?
- I have NO idea how to deploy an AI micro-service!!

Frank Infuse

- Just give me a model, I'll deploy it for you!
- I can create create real-time and batch interfaces
- I can even schedule it for you

Tina Evaluate

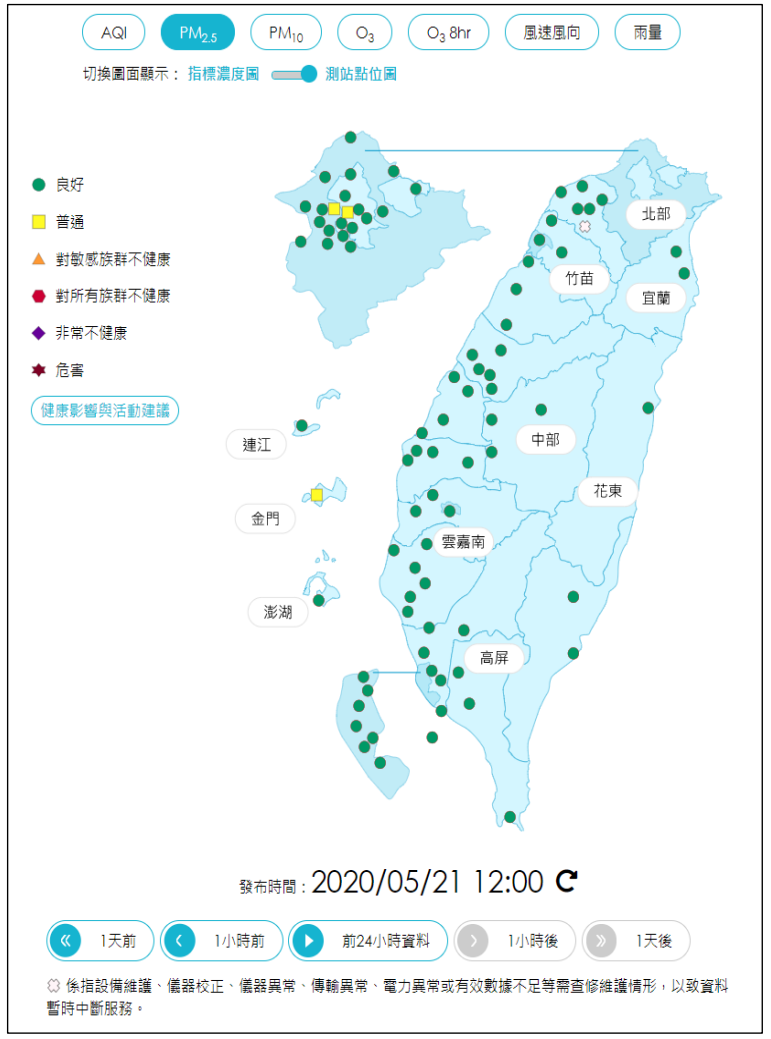
- Need to make sure the there is no bias in the model
- I need to make sure drift is being controlled?
- Who all is using the model now?

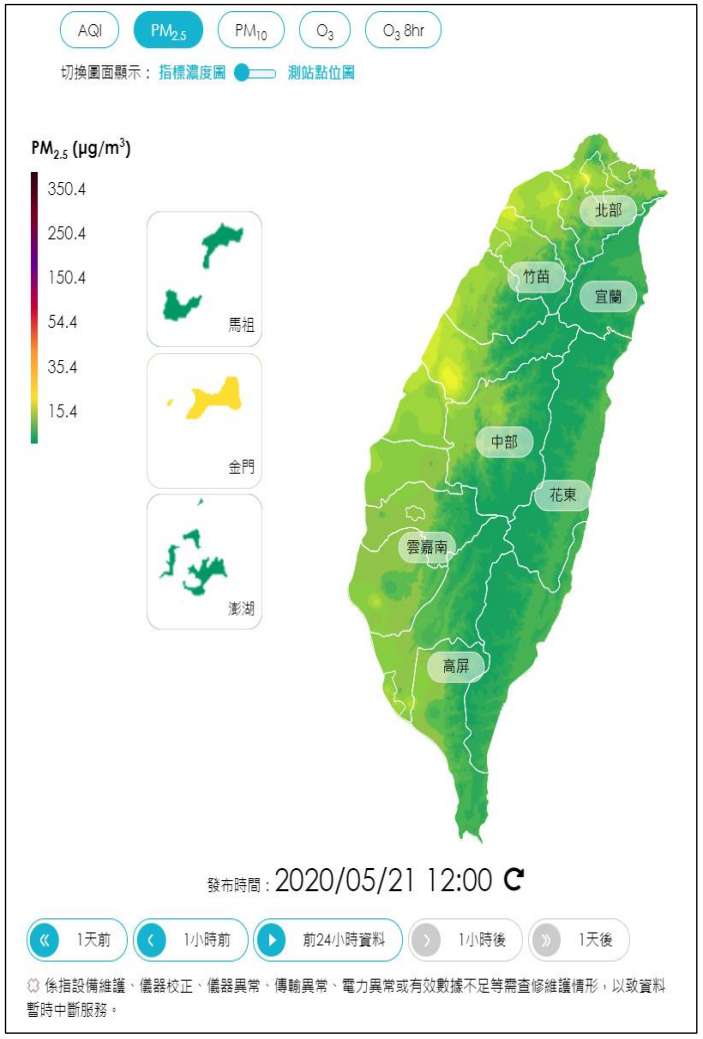
案例與功能介紹
PM2.5分析預測

問題背景與分析需求

- **展示背景：**
  1. 利用公開資料，對市民與政府機關提供PM2.5數值未來的預測值，提供查詢與決策。
  2. 並對於潛在可能造成汙染的地點與時段(如：工廠、特定路段等)，進行監控並做即時告警。
  3. PM2.5數值實時數值及未來五分鐘的預測值。

- **需求：**
  依公開資料(氣象資料、雨量資料、PM2.5資料)一預測PM2.5數值在未來5分鐘的變化。





空汙與氣象資料來源

- **氣象資料：**
  - 局屬氣象站 (共 45 站點)
  - 自動氣象站 (共 432 站點)
- **雨量資料：**
  - 自動雨量站 (共 998 站點)
- **PM2.5 資料：**
  - lass 社群 (共約 3000 點)

分析大師(DAS)Pipeline

COLLECT ⟵ ANALYZE ⟶ ⟵ INFUSE ⟶

Collect, Connect and Access Data	Understand and Prepare Data for Analysis	Build Descriptive, Predictive, Prescriptive and Planning Models	Model Management and Deployment	Create Analytics Applications

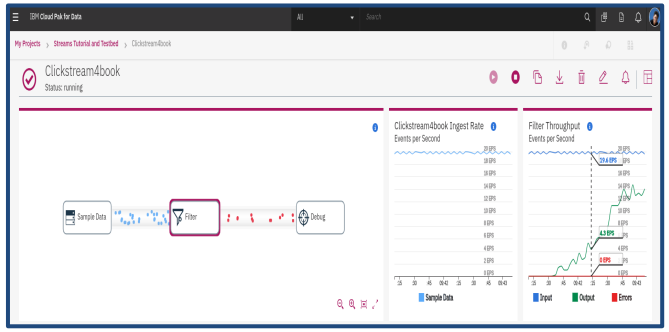
- Data Stream:
  氣象資料
  雨量資料
  PM2.5資料

氣象資料：
局屬自動氣象站(1h)
氣象站(1h)

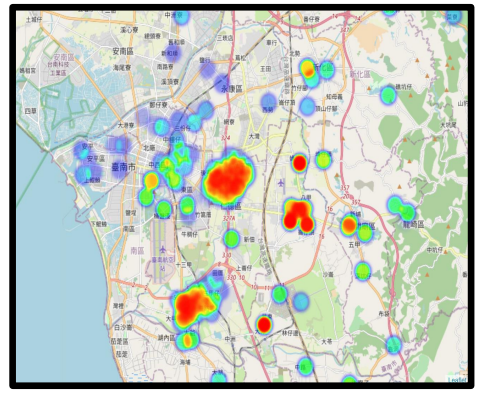
雨量資料：
自動雨量站(10min)

PM2.5 資料：
lass 社群(5min)

- Data Stream:
  資料清理與融合
- Notebook:
  開發資料處理與週期性任務的建立

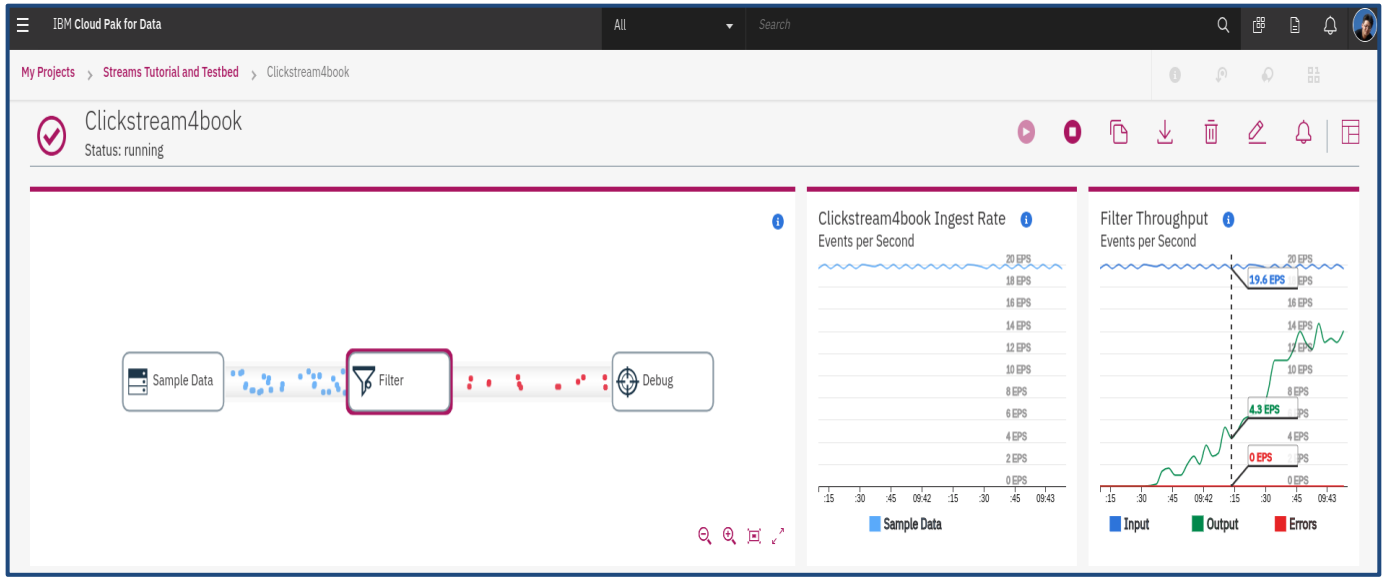


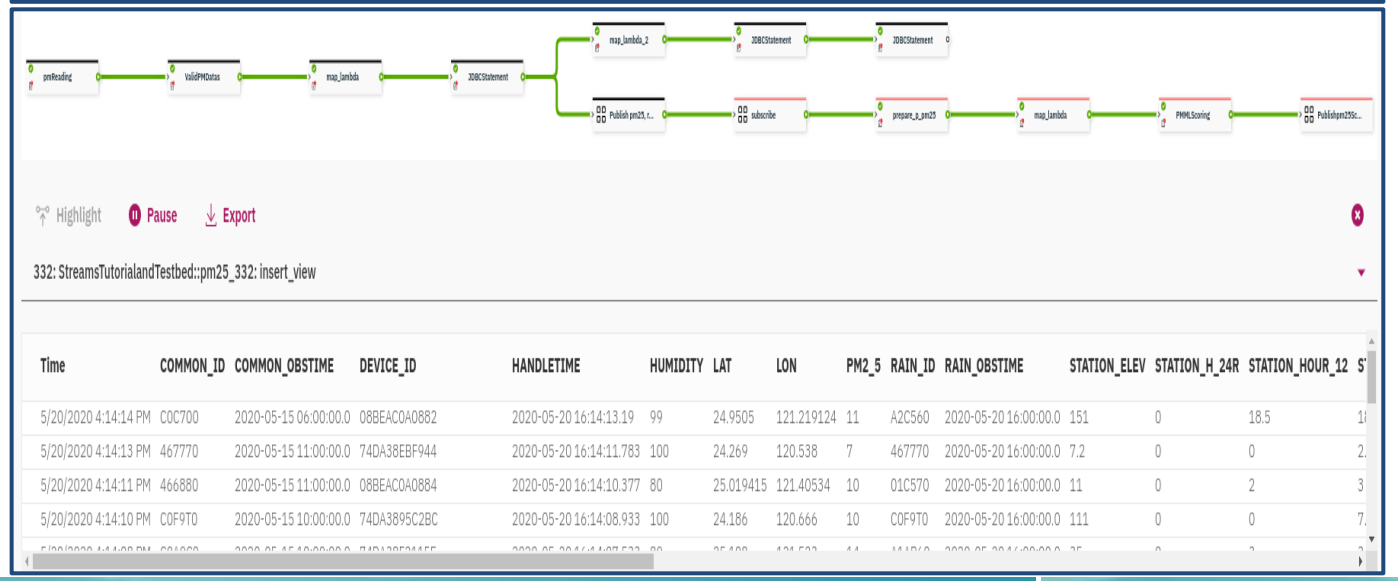
- Deployment:
  部署模型進行預測未來變化

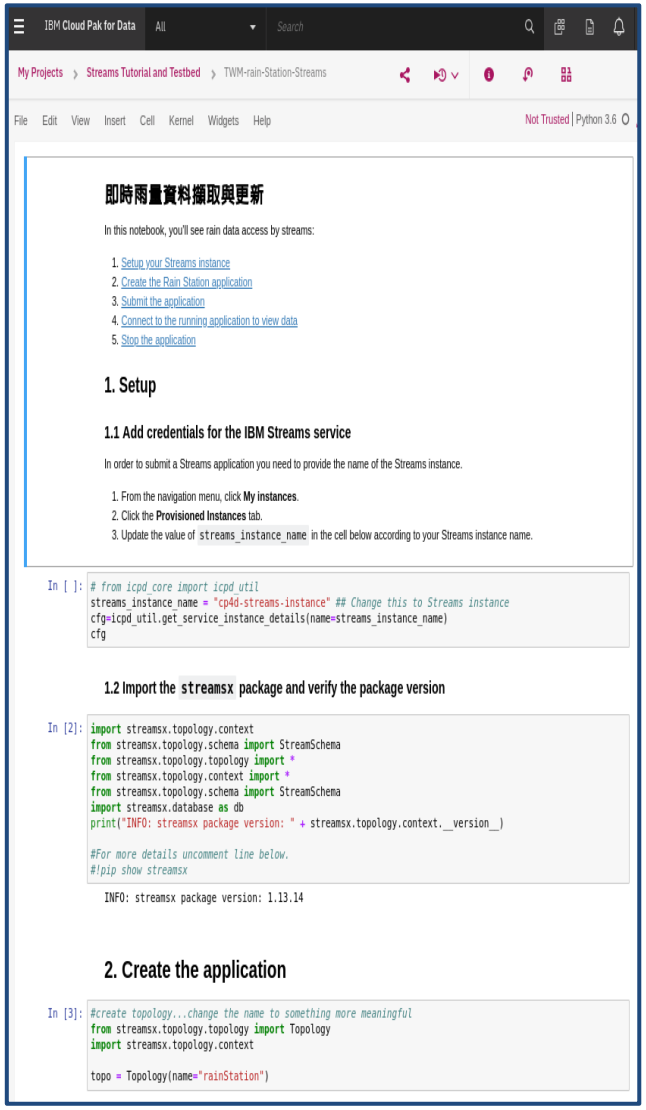


- Notebook:
  透過程式繪製合適的空間資料視覺化呈現

以Streams功能進行即時資料分析

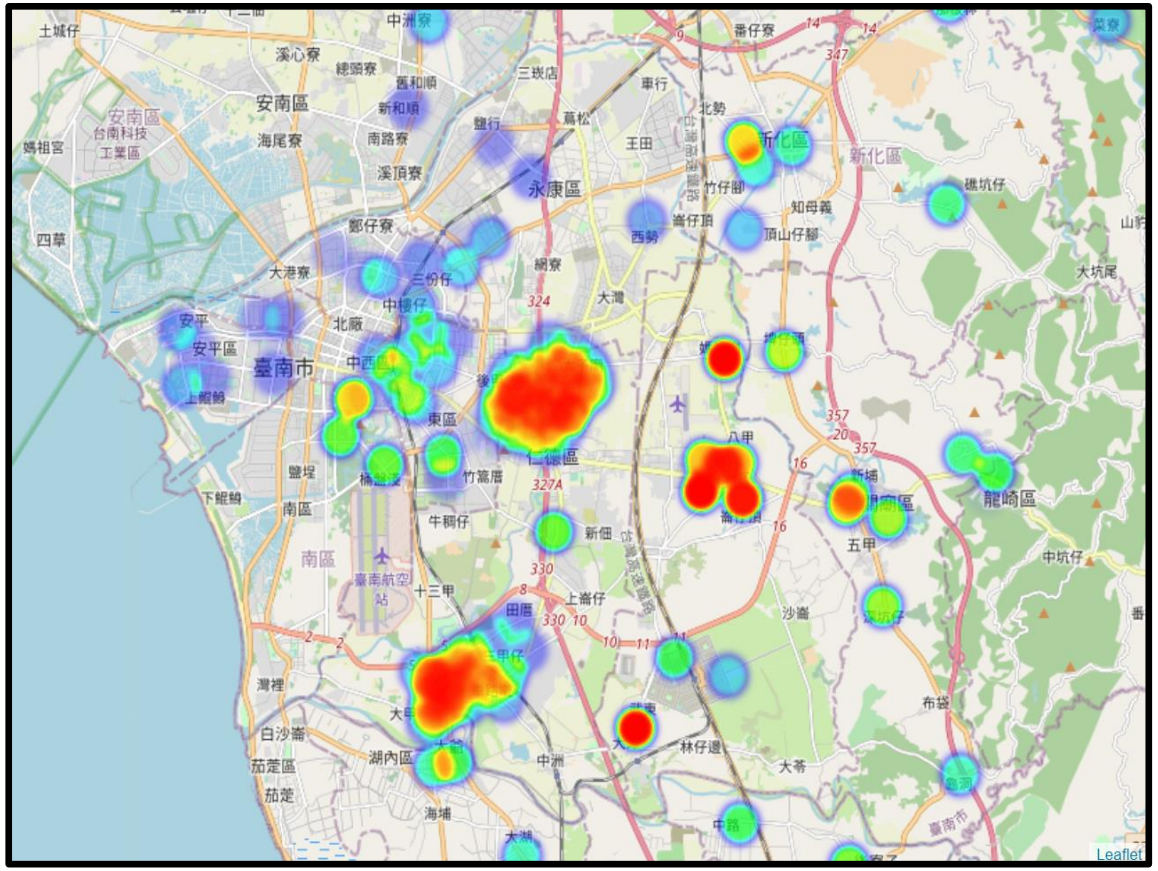


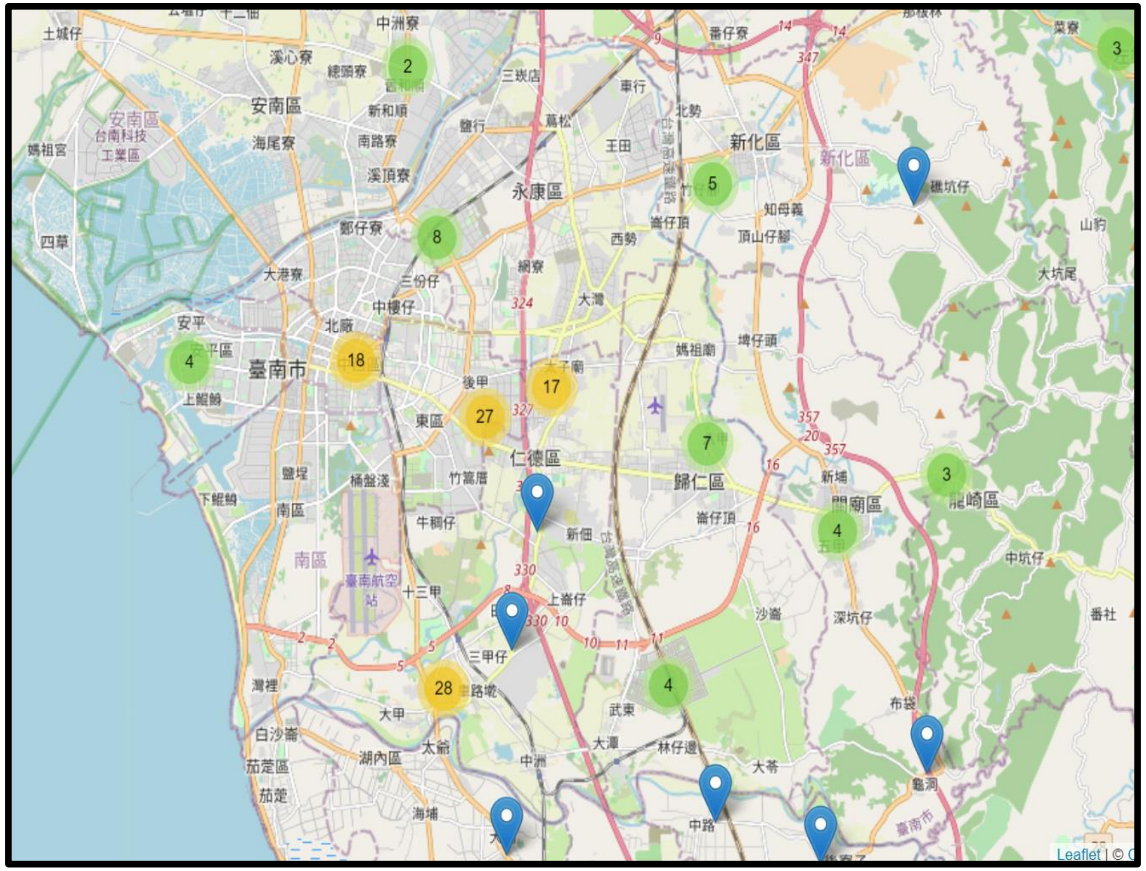




PM2.5 資料預測結果:以台南市為例

- 將分析結果以RStudio繪圖並部署到資料市集，可針對資料坐即時的監控





案例與功能介紹
Youbike借用應用

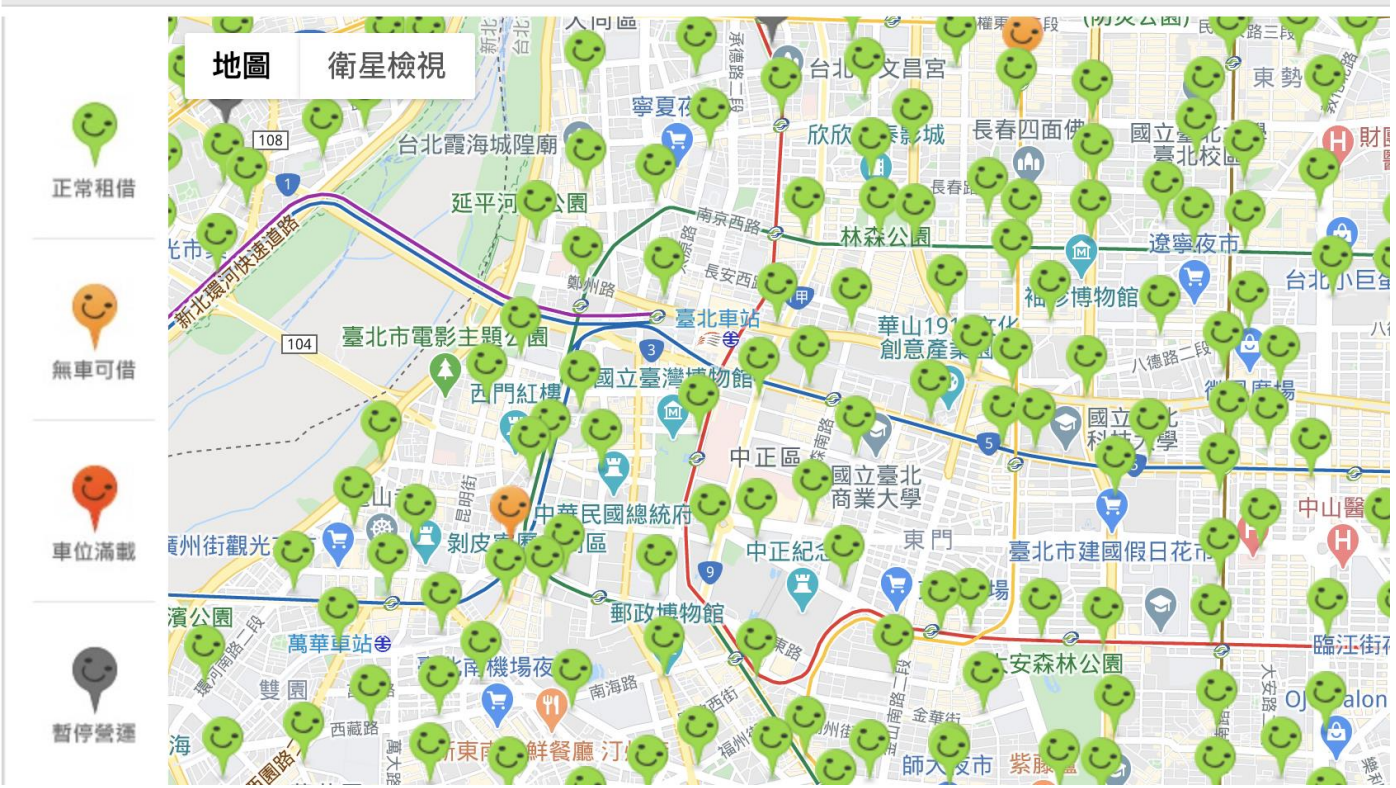
問題分析與需求

- Youbike公司背景：

1. Youbike公司在借用服務對市民提供即時各站目前剩餘車輛數及可還車空位數供市民查詢。
2. 對於熱門站點於特殊時段(如：公館捷運站出口放學時段需求量大增)需求只能依經驗作調度。
3. 調度量及各站間調度面臨困難。

- 需求：

  依各站借用歷史資料，結合氣象資料(氣溫、降雨)發展一預測模式作為車輛調度使用。



資料內容與結構

- Youbike:
  - sno：站點代號
  - tot：場站總停車格
  - sbi：場站目前車輛數量
  - bemp：空位數量
  - act：全站禁用狀態
  - utime: 系統時間

https://tcgbusfs.blob.core.windows.net/blobyoubike/YouBikeTP.json

- 天氣:

elementName	中文說明
ELE	高度，單位 公尺
WDIR	風向，單位 度，一般風向 0 表示無風
WDSD	風速，單位 公尺/秒
TEMP	溫度，單位 攝氏
HUMD	相對濕度，單位 百分比率，此處以實數 0-1.0 記錄
PRES	測站氣壓，單位 百帕
H_24R	日累積雨量，單位 毫米
H_FX	小時最大陣風風速，單位 公尺/秒
H_XD	小時最大陣風風向，單位 度
H_FXT	小時最大陣風時間，yyyy-MM-ddThh:mm:ss+08:00
D_TX	本日最高溫，單位 攝氏
D_TXT	本日最高溫發生時間，hhmm (小時分鐘)
D_TN	本日最低溫，單位 攝氏
D_TNT	本日最低溫發生時間，hhmm (小時分鐘)
CITY	縣市
CITY_SN	縣市編號
TOWN	鄉鎮
TOWN_SN	鄉鎮編號
補充說明	-99 皆表示 該時刻因故無資料。

https://opendata.cwb.gov.tw/api/v1/rest/datastore/O-A0001-001?Authorization=rdec-key-123-45678-011121314

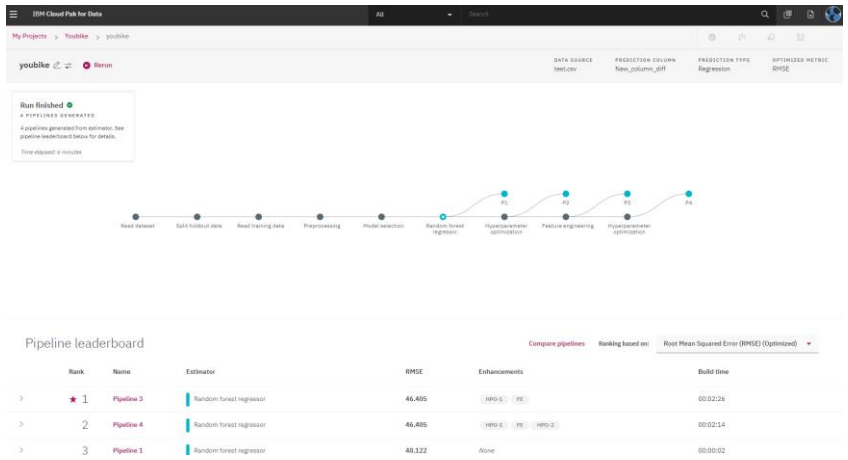
分析大師(DAS)Pipeline

COLLECT ← ANALYZE → ← INFUSE →

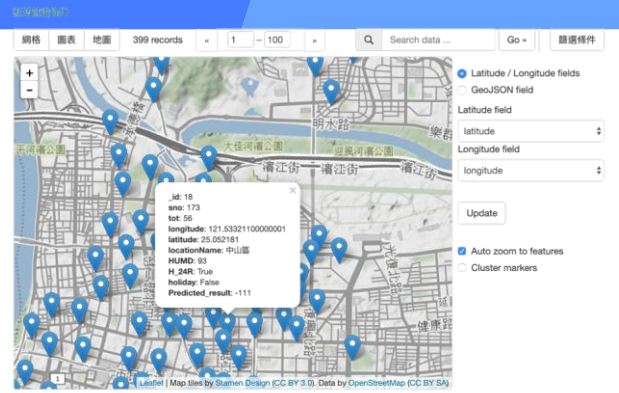
Collect, Connect and Access Data	Understand and Prepare Data for Analysis	Build Descriptive, Predictive, Prescriptive and Planning Models	Model Management and Deployment	Create Analytics Applications

- Data Connection:
  Youbike資料
  氣象觀測資料
  氣象預報資料

- Data Refinery:
  計算每天車輛借用情況，清理氣象資料



- AutoAI:
  利用現有借用資料及氣象觀測資料透過AutoAI訓練預測模型



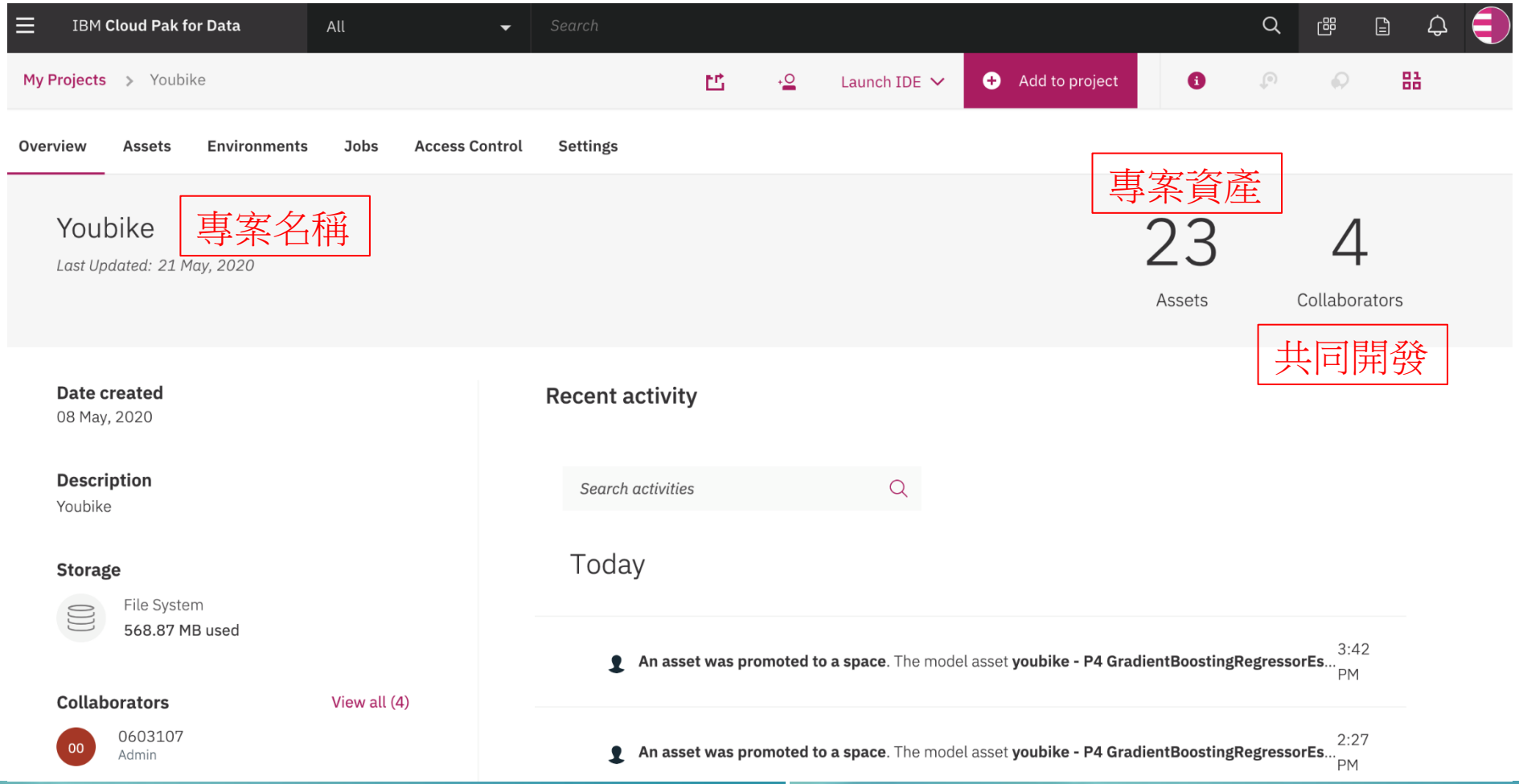
- Deployment:
  部署預測模型進行預測

- Dashboard:
  互動式監控面板

- Data sharing:
  資料市集、open data

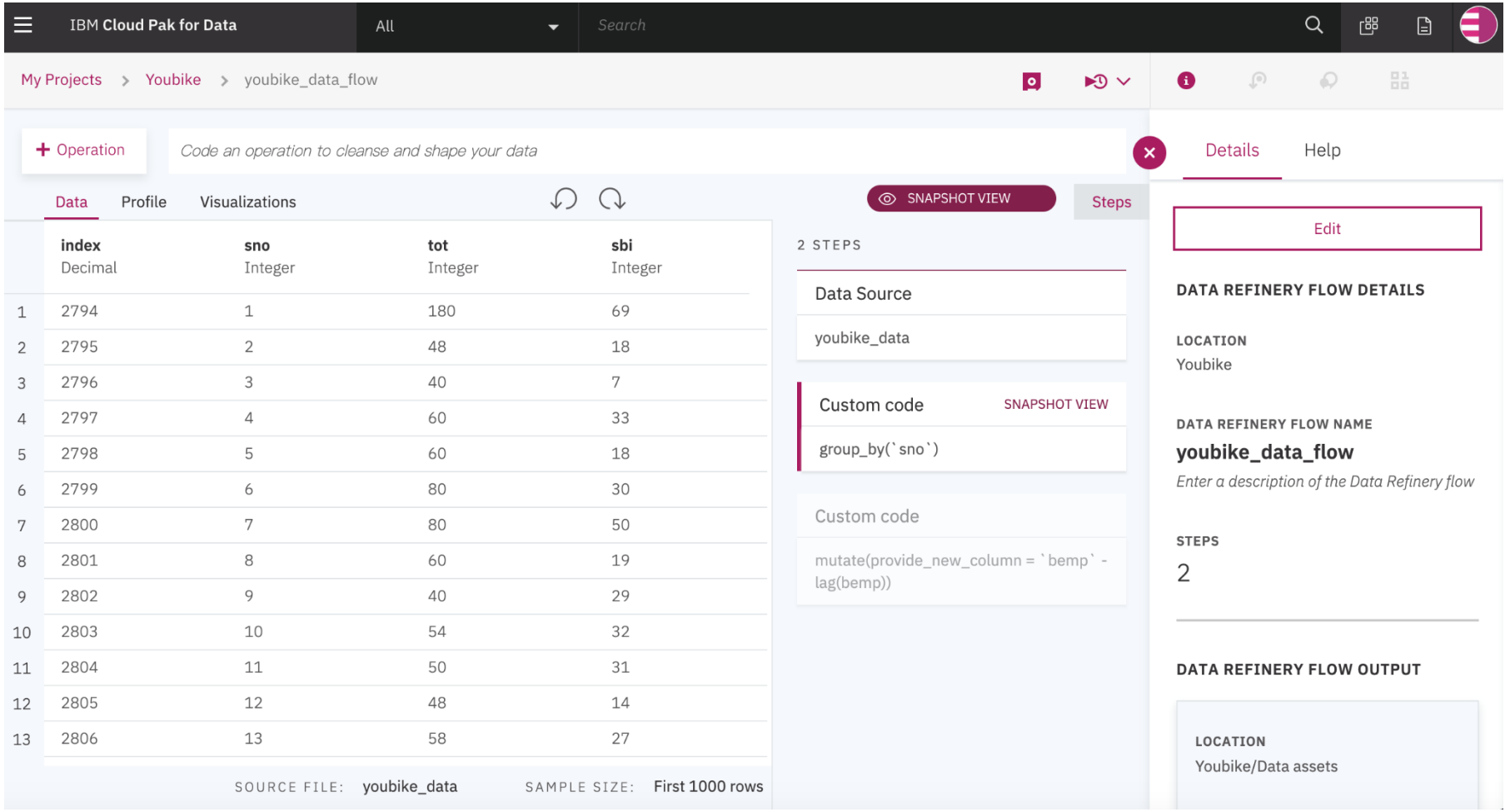
專案管理清單

- 建立專案並加入共同開發者



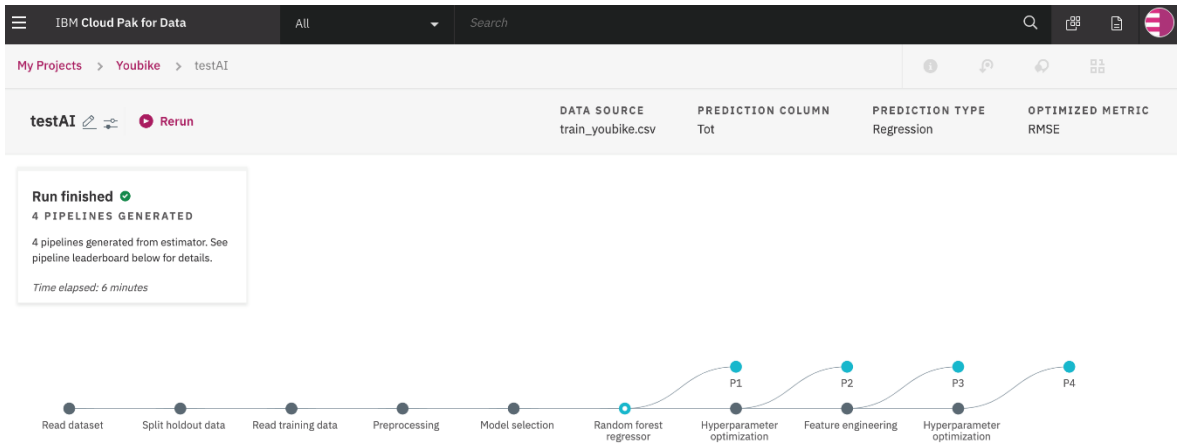
Data Refinery

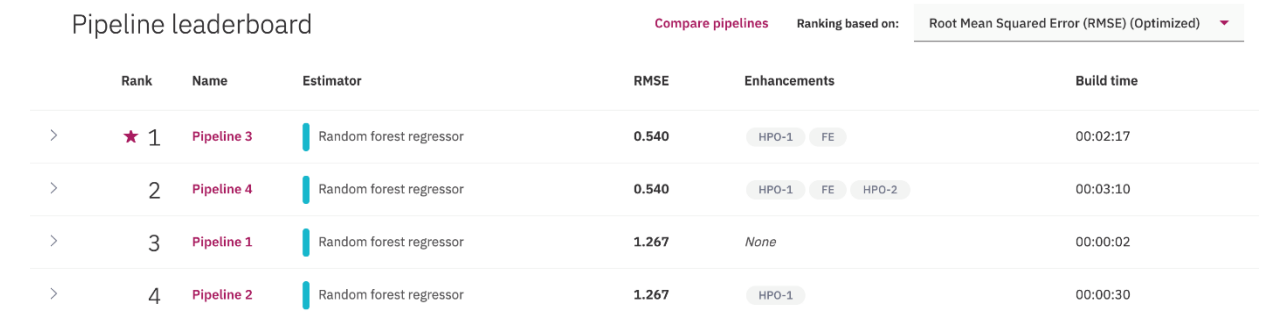
- 將資料導入並使用data refinery進行資料清理

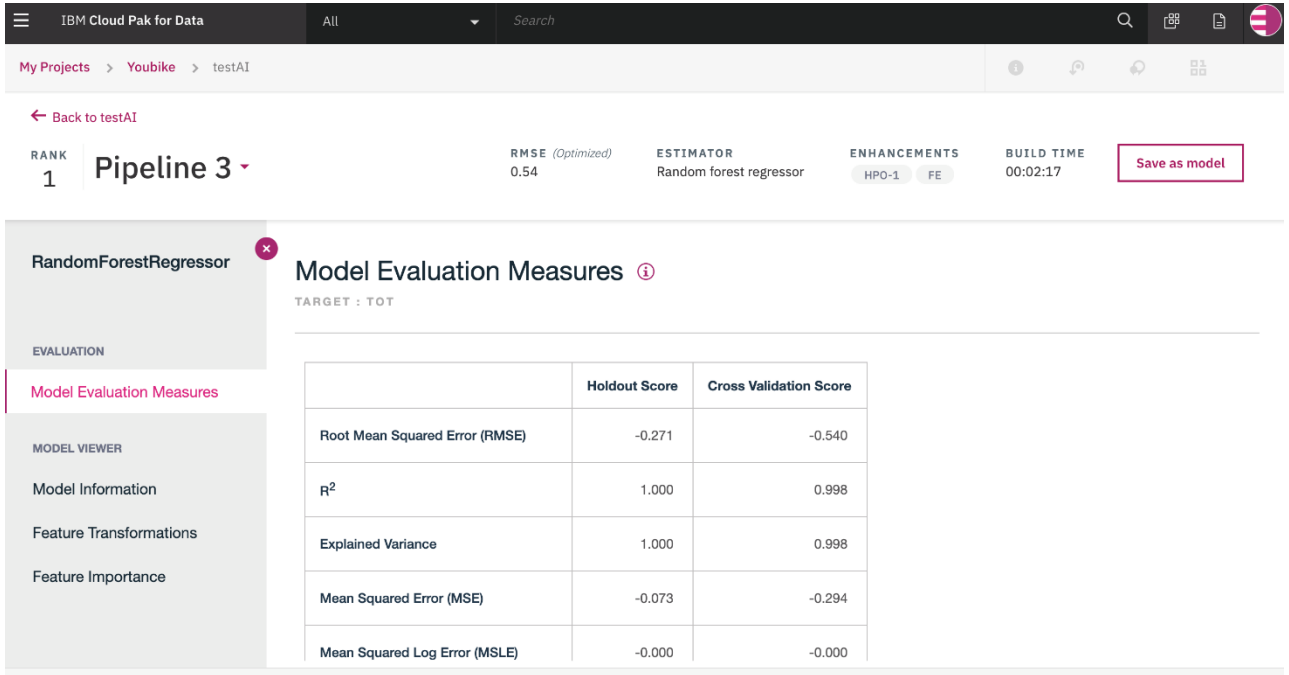


AutoAI

- 將清理好的資料導入AutoAI
- 自動選取回歸模式
- 由系統自動選取最佳的模型，並可觀察模型分析細節







Thank you for your attention.

分析大師 DAS 試用申請: https://event.nchc.org.tw/2020/dasapply/
聯絡我們: das@narlabs.org.tw