

應用於 MiinStore 平台之資料湖實作

MiinStore DataLake Implementation

指導教授：謝孫源教授

專題成員：陳廷笙、王昱翔

開發工具：Python、JavaScript

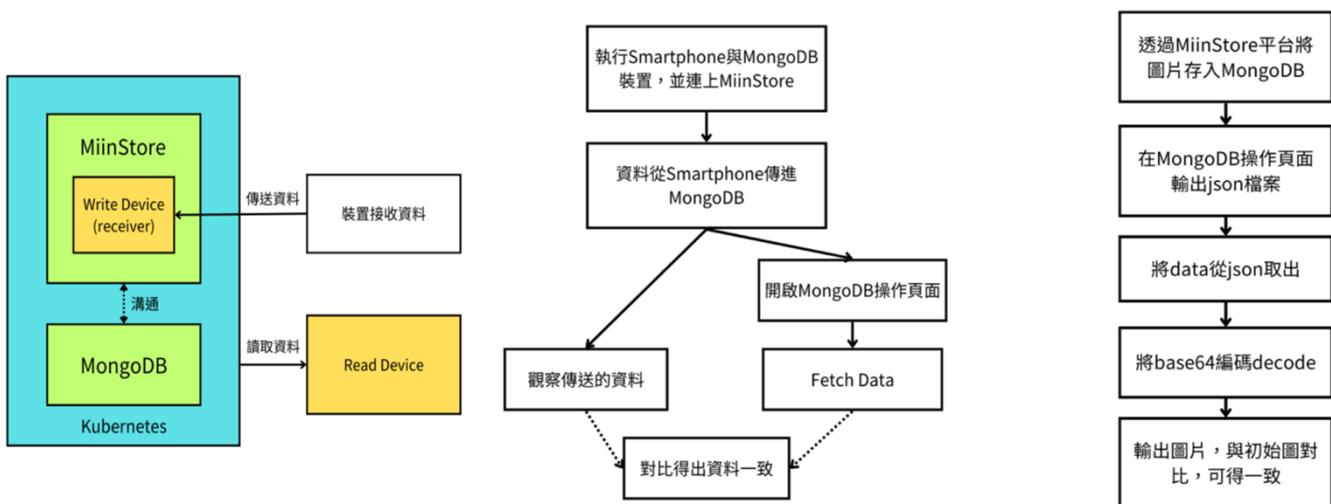
測試環境：Windows、Mac OS、Ubuntu 22.04

一、簡介：

本作品基於 MiinStore IoT 平台，加上先進的資料湖(Data Lake)架構，以利高度靈活的數據存儲與分析。資料湖的特點為：「在無需提前建立資料結構的情況下，即可儲存**結構化與非結構化**的資料。」而且，由於資料湖不需提前定義資料的類型，將可大幅增加實務上使用的**彈性與便利性**，並根據需求在不同的情況下，將收集到的資料進行最適合的分析。

再來介紹本次專題的實驗方法，先將由一台**行動裝置(Smartphone)產生資料**，透過我們在 MiinStore 內部設計的一個名為 **Receiver 的裝置(Device)**來接收後，把收集到的資料傳入我們用 MongoDB 實作的資料湖中，並使用在 MiinStore 中設計的 **MongoDB 讀取裝置**，從資料湖中讀取資料，最終下載為 JSON 或 CSV 格式。

作品未來將可被應用於各種不同場域，選擇使用建立於本地端的資料庫，而非外部現成的資料庫，可安全存放個人資訊、檢驗數據、醫學影像等高度隱私的資料，將可**提升相關資料的安全性**，與有效的防止資料外洩的情況發生；另外，使用資料湖與一般的資料庫相比，以醫學上為例，將可因應醫學上經常出現新的醫療技術，新型的技术經常會產生不同的資料類型，此時使用資料湖，就可利於我們將**各種不同型態的資料儲存在同一個地方**，不需額外花費成本與時間再建立另外的資料庫。



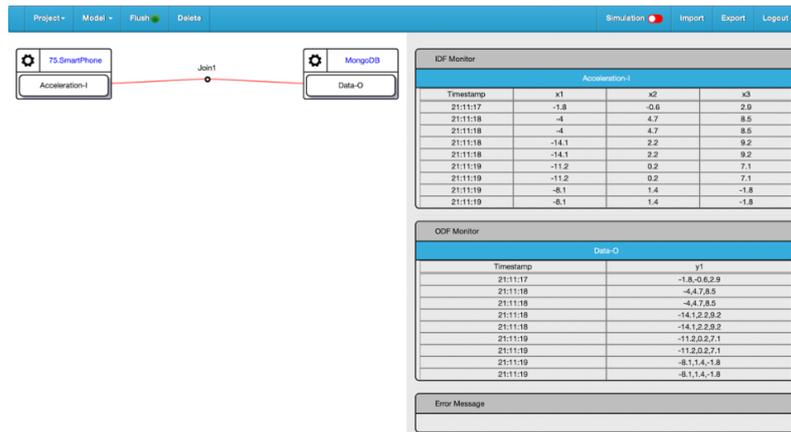
圖一、系統架構圖

圖二、結構化資料處理流程圖

圖三、非結構化資料處理流程圖

二、測試結果：

本次實驗會在 MiinStore 網頁平台上進行。將 MongoDB 變成 MiinStore 平台上的 Device，可供使用者直接使用，並將 input 與 output 給 join 起來，傳送給 MongoDB。右側的 x1,x2,x3代表手機所傳出的三軸加速度，如圖四(下)所示。



圖四、MiinStore 操作畫面

在 MongoDB 的操作畫面內，可以看到圖四的 x1~x3會被存為 Data，Timestamp 會被存為 Date 並存進 MongoDB，並且和圖四的資料一致，如圖五(下)所示。

No.	_id	Data	Date
1	6649fa77cde45ee72bbb2c69	-8.1,1.4,-1.8	2024/5/19 下午9:11:19
2	6649fa77cde45ee72bbb2c68	-8.1,1.4,-1.8	2024/5/19 下午9:11:19
3	6649fa77cde45ee72bbb2c67	-11.2,0.2,7.1	2024/5/19 下午9:11:19
4	6649fa77cde45ee72bbb2c66	-11.2,0.2,7.1	2024/5/19 下午9:11:19
5	6649fa76cde45ee72bbb2c65	-14.1,2.2,9.2	2024/5/19 下午9:11:18
6	6649fa76cde45ee72bbb2c64	-14.1,2.2,9.2	2024/5/19 下午9:11:18
7	6649fa76cde45ee72bbb2c63	-4,4.7,8.5	2024/5/19 下午9:11:18
8	6649fa76cde45ee72bbb2c62	-4,4.7,8.5	2024/5/19 下午9:11:18
9	6649fa75cde45ee72bbb2c61	-1.8,-0.6,2.9	2024/5/19 下午9:11:17

圖五、MiinStore 平台內，MongoDB 的操作畫面

本次實驗 Data Lake 的好處是除了結構化資料外，非結構化資料如圖片也可以處理。以下是照片的處理，讀取 Data Lake 中的照片檔案資料由 png 檔轉成 byte string 格式後，再以 base 64編碼，如圖三(上)及圖六(下)所示。

No.	_id	Data
1	66489fed202c6f91773edc5d	{"data": "iVBORw0KGgoAAAANSUhEUgAAC3wAA"}
2	66489fec202c6f91773edc5c	{"data": "iVBORw0KGgoAAAANSUhEUgAAC3wAA"}
3	66489feb202c6f91773edc5b	{"data": "iVBORw0KGgoAAAANSUhEUgAAC3wAA"}
4	66489fea202c6f91773edc5a	{"data": "iVBORw0KGgoAAAANSUhEUgAAC3wAA"}

圖六、MongoDB 操作非結構化資料