

# 地端到雲端 串聯數據 效能升級

Rock Chen  
is-land Systems Inc

# 亦思科技 Data Engineering



## ● 公司

- ◆ 新竹科學園區 成立超過十年
- ◆ 員工人數90
- ◆ 大數據平台服務團隊：R&D、顧問、專案與維運團隊



## ● 平台

- ◆ 唯一通過Cloudera **產品認證**之台灣廠商
- ◆ 數十座叢集的建置、維運經驗
- ◆ 平台維運經驗囊括各不同叢集大小及使用情境

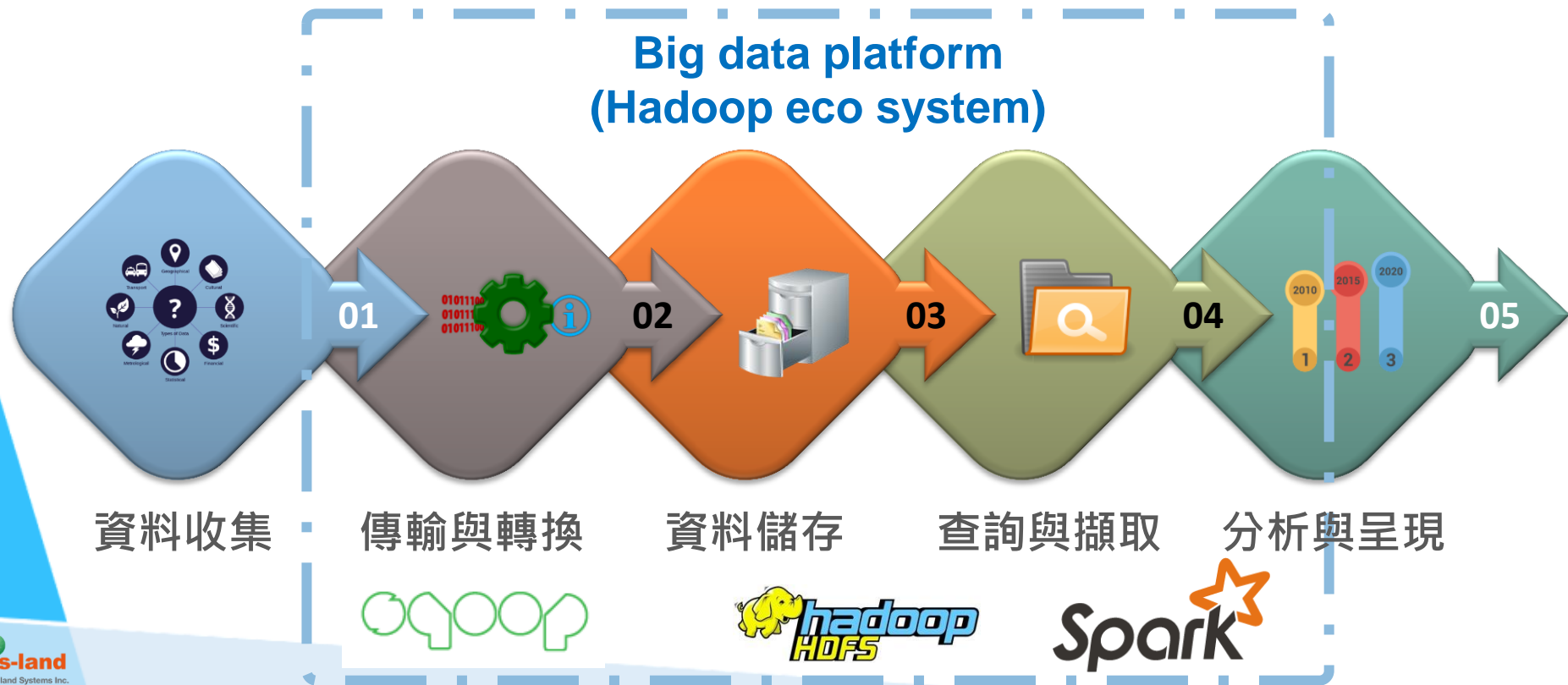


## ● 應用

- ◆ 所有台灣半導體廠皆有使用
- ◆ 熟知大數據系統開發流程與架構
- ◆ 橫跨製造、金融、電信、傳產
- ◆ 長期投入開源專案研究與開發



# 批次處理時代的大數據平台架構



# 傳統架構面臨的痛點

## 難以預估硬體資源

日常資料處理已有固定分配的硬體資源，on-demand的分析需求更耗用大量的Computing power，且分析人員常常無法根據需求即sizing出本次分析所需要的硬體設備。

## 大量細碎的即時資料

傳統Hadoop eco system無法承載即時資料，若大量細碎資料全部進入Hadoop，將影響Hadoop執行效率，因而需要額外的資料處理時間。

## IT/AP間資料的管理

原始資料與分析資料皆位於同一座cluster，對於資料安全保護的要求需更加嚴格。



# 用微服務概念打造大數據平台

## Data Engineering

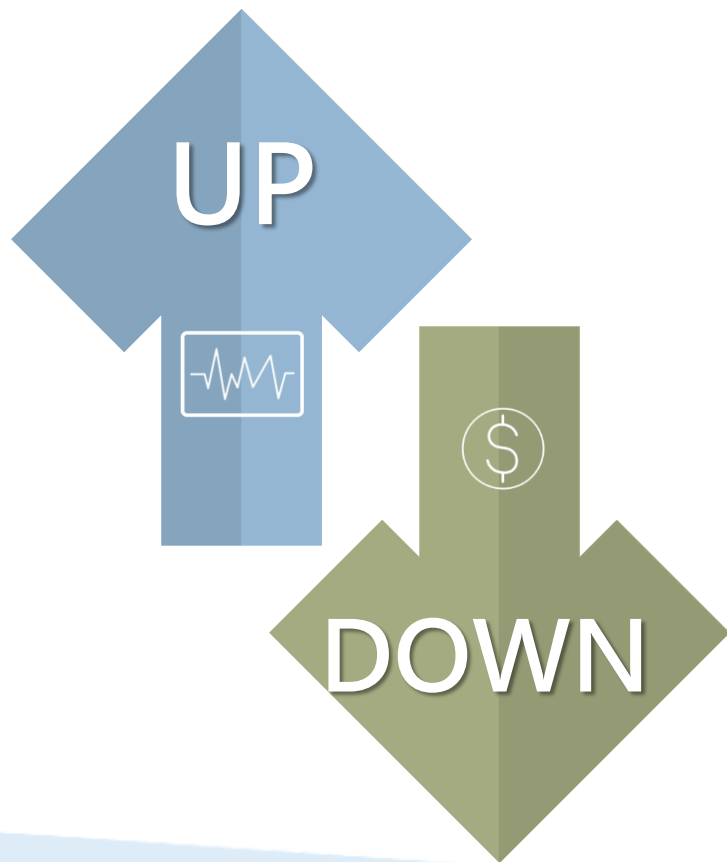
- 兼容Batch與Real-time
- 可乘載大量的through put
- 快速整合與開發

## Data Lake

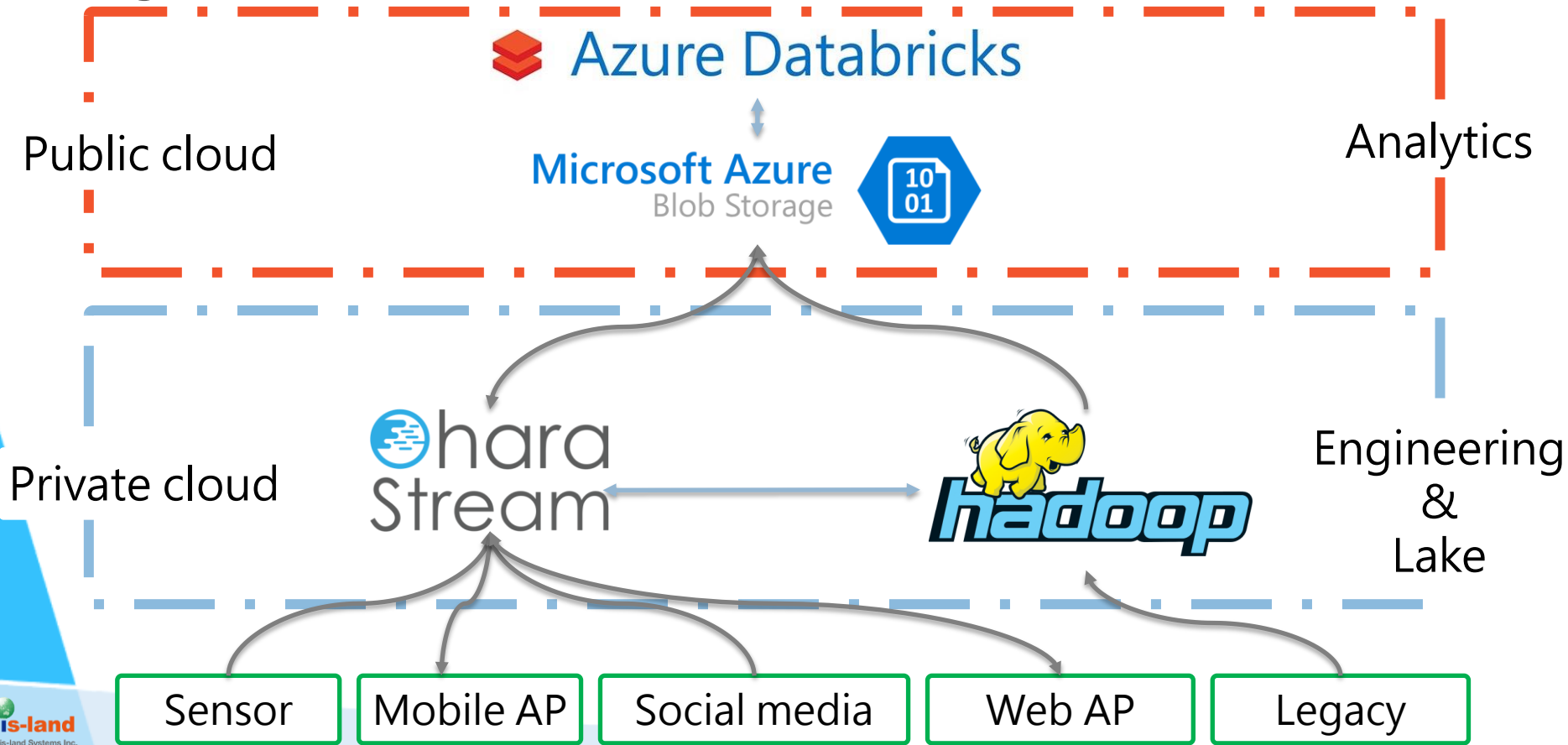
- 具備擴充彈性與可靠性
- 負責資料儲存與治理
- 資料前處理與資料提供

## Data analytics

- 快速建立分析環境
- 內建常用分析語言與演算法
- 具備有效率的分析框架



# Hybrid cloud 的大數據平台架構



# Ohara Stream 串流數據整合的利器

## 整合元件

- Apache Zookeeper / Kafka
- Docker & Kubernetes

## 資料整合

- Connector可介接FTP、RDB、HDFS、Storage Blob等外部系統
- 提供可客製化Connector API

## 分散式運算

- Kafka、Zookeeper 等分散式元件組成
- StreamApp可部署多節點，進行分散運算

## 動態部署

- 各服務封裝成獨立的Docker Image
- 動態部署叢集服務
- 動態上傳與部署StreamApp
- 動態運行資料流

## GUI視覺化

- 以視覺化呈現資料流狀態
- 用拖拉完成資料介接

## 狀態監控

- Components狀態監控。
- 資料流動狀態監控。

# Use Cases - ATM詐騙偵測

◆ 即時ATM詐騙警告

◆ 加速追蹤盜領軌跡

Ohara Stream

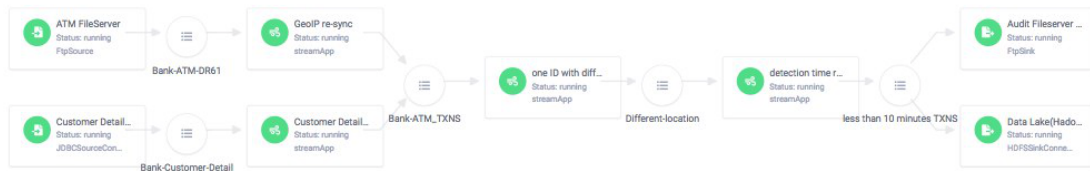
Pipelines

Nodes

Services



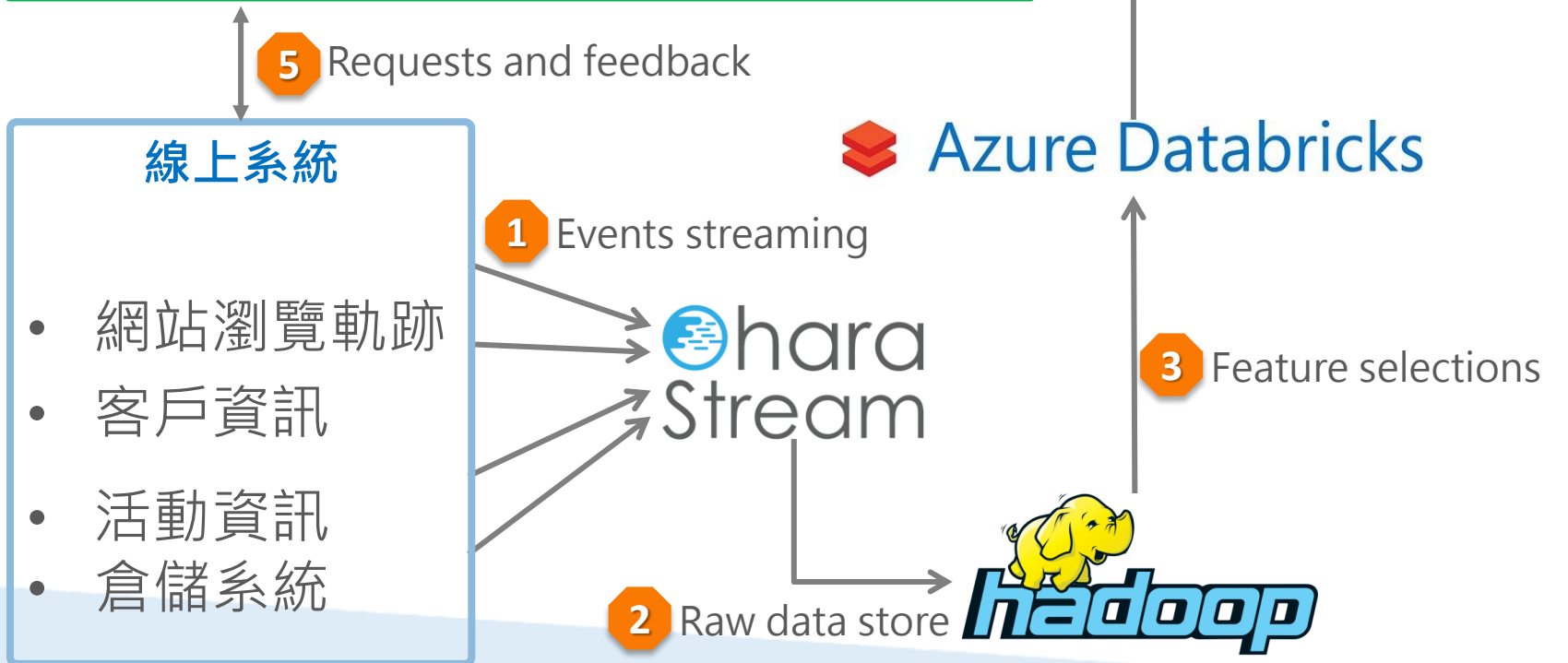
## Pipeline graph





# Use Cases – 電子商務

- Consumer Behavior Prediction Services
- Logistics Automation Services



# Summary

- Ohara Stream讓資料整合及傳輸更有效率，且降低資料科學家等待及整理資料的effort
- 大數據平台的效率分工，可讓資料快速轉換為業務行動
- 善用雲端的彈性，Azure Databricks快速佈建分析環境，提升分析效率  
降低硬體設備的成本

# Thank you for your attentions