



食話實說

餐廳評論數據分析

王承宇、王鈺婷、姚伯翰、楊鍇鎧、劉冠宏

目錄

- 一. 組員介紹
- 二. 專題介紹
- 三. 專題使用工具
- 四. 資料介紹
- 五. 系統建置
- 六. 資料預處理
- 七. 模型建置
- 八. 資料視覺化

組員介紹



王承宇

資料蒐集
資料視覺化
網頁建置



王鈺婷

資料蒐集
資料清洗
簡報製作



姚伯翰

資料蒐集
系統建置
網頁建置



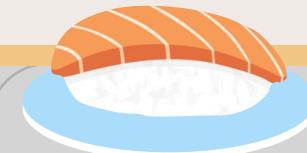
楊鍇鎧

資料蒐集
系統建置
模型建置



劉冠宏

資料蒐集
資料清洗
模型建置



資料蒐集

系統建置

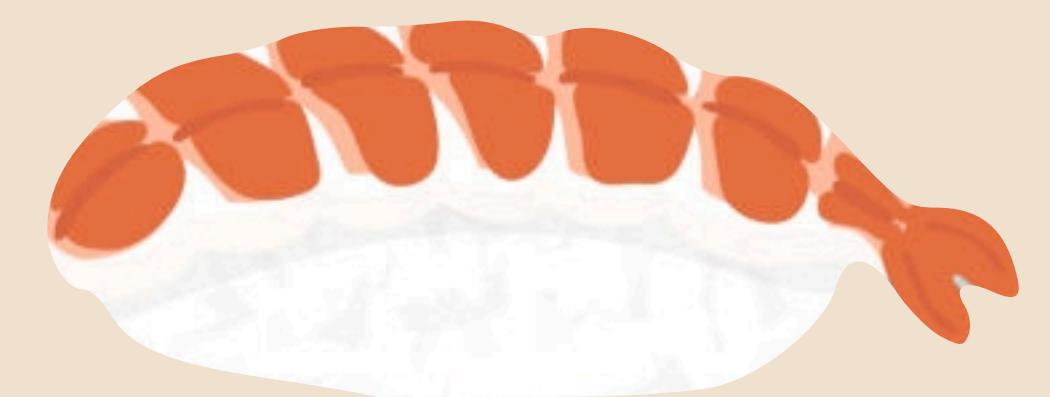
資料預處理

模型建置

資料視覺化

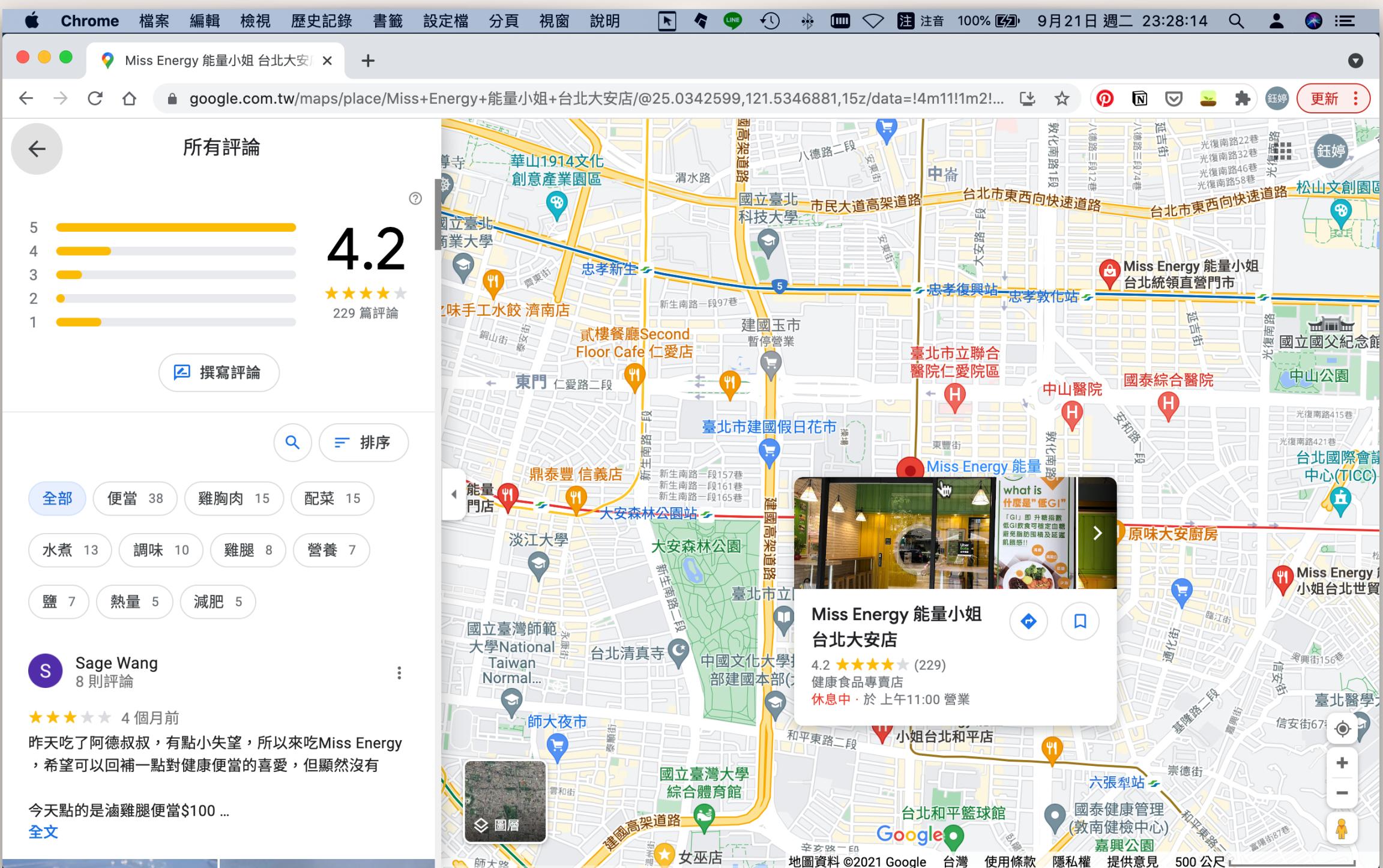
總結

專題介紹



緣起

- Google Maps的高星等美食卻不美味？
- 為何近滿星等卻仍會踩雷呢？



資料蒐集

系統建置

資料預處理

模型建置

資料視覺化

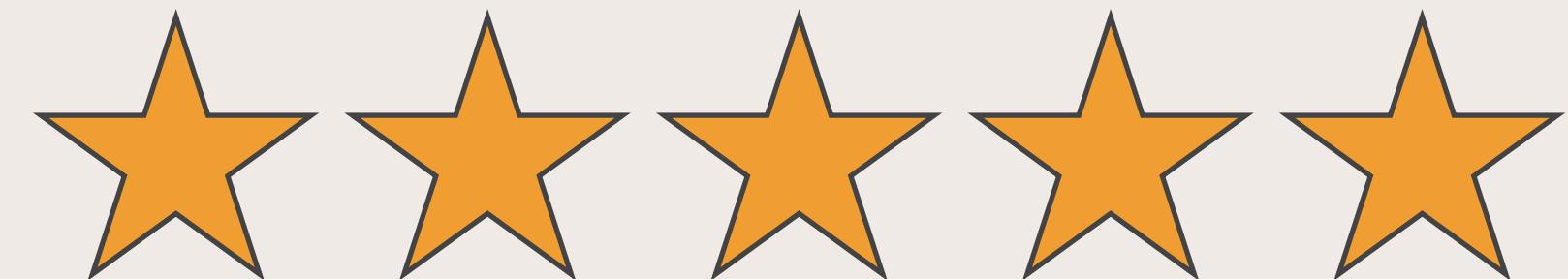
總結

緣起

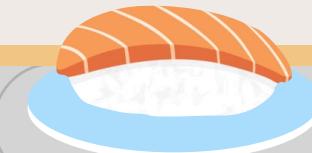
- 同樣的感受轉化成分數的標的不同



好吃！！



好吃！！



資料蒐集

系統建置

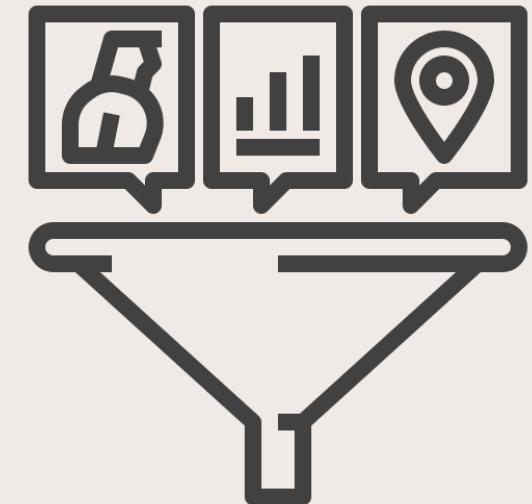
資料預處理

模型建置

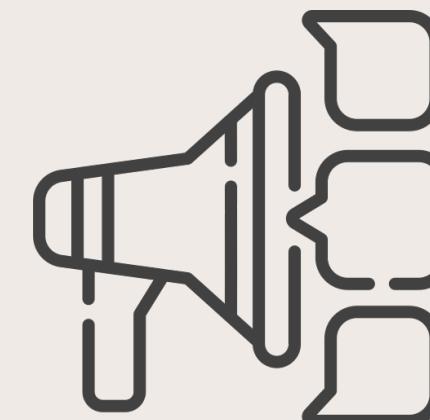
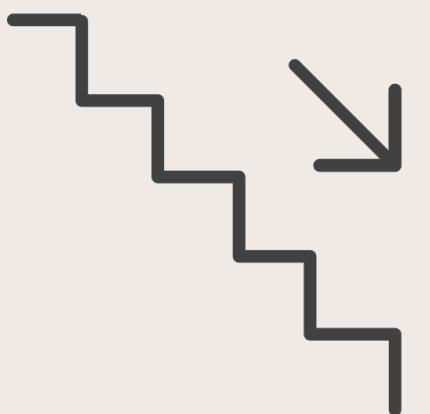
資料視覺化

總結

目標

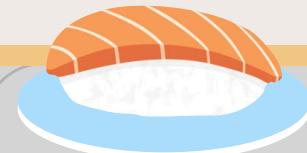


整合各大美食評論網站，降低
使用者屬性不同造成的差異



利用「文字描述」作為評分依據，預期降低不同評判標準造成的偏差

找出各餐廳被提及的高頻字眼，挖掘餐廳特色



資料蒐集

系統建置

資料預處理

模型建置

資料視覺化

總結

工具

系統建置



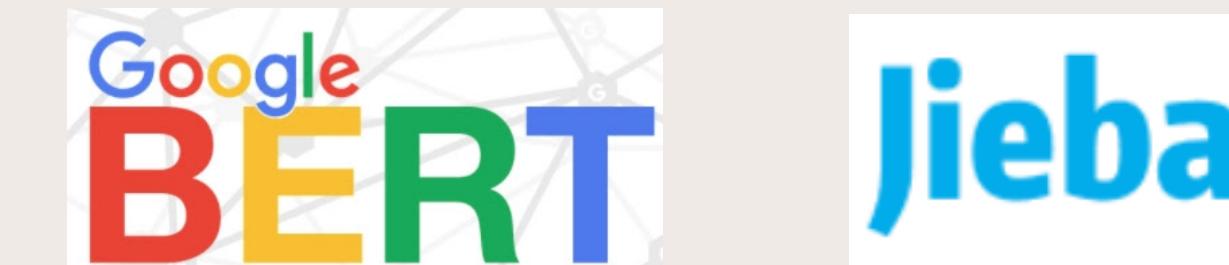
資料蒐集



資料清洗



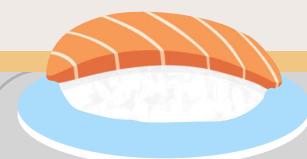
模型建置



資料視覺化



網頁建置



資料蒐集

系統建置

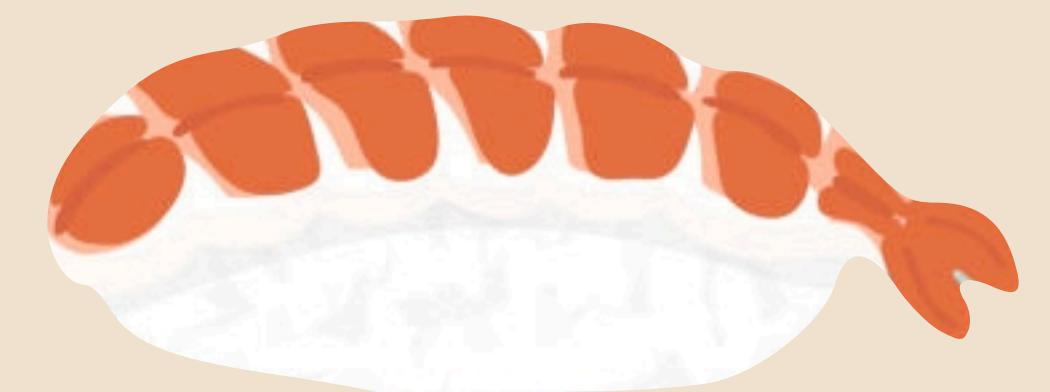
資料預處理

模型建置

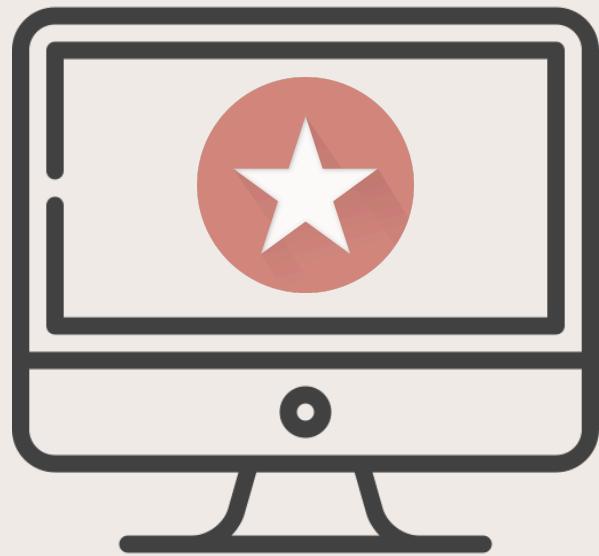
資料視覺化

總結

資料蒐集



資料蒐集



EZ Table



GoogleMaps



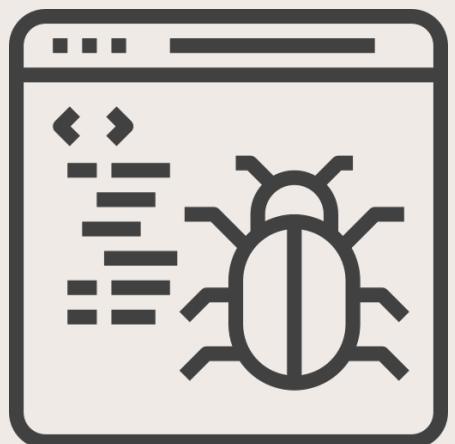
iFoodie



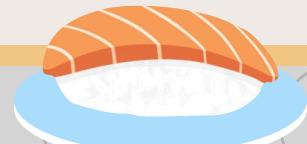
MenuTaiwan



Tripadvisor



- 利用爬蟲爬取五大美食評論平台：EZtable簡單桌、GoogleMaps、iFoodie愛食記、Menutaiwan美食誌、Tripadvisor
- 使用工具：**Python 及 Request, Selenium 套件**



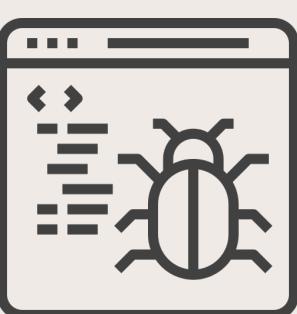
資料蒐集

The screenshot shows the homepage of the MENU website for Taipei. It features a header with a search bar and navigation links. Below the header, a section titled "台北市 排行榜" displays five restaurant cards with images, names, and average ratings. The cards are:

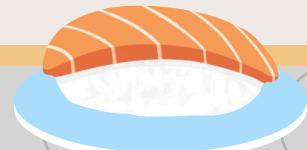
- 天廚菜館 (4.3)
- 葫蘆小郭鐵板燒 (4.6)
- GB鮮釀餐廳 (4.2)
- hot 7 新鐵板料理 (4.3)
- 全家便利商店 (4.2)

Each card includes a small image of the restaurant's food or interior.

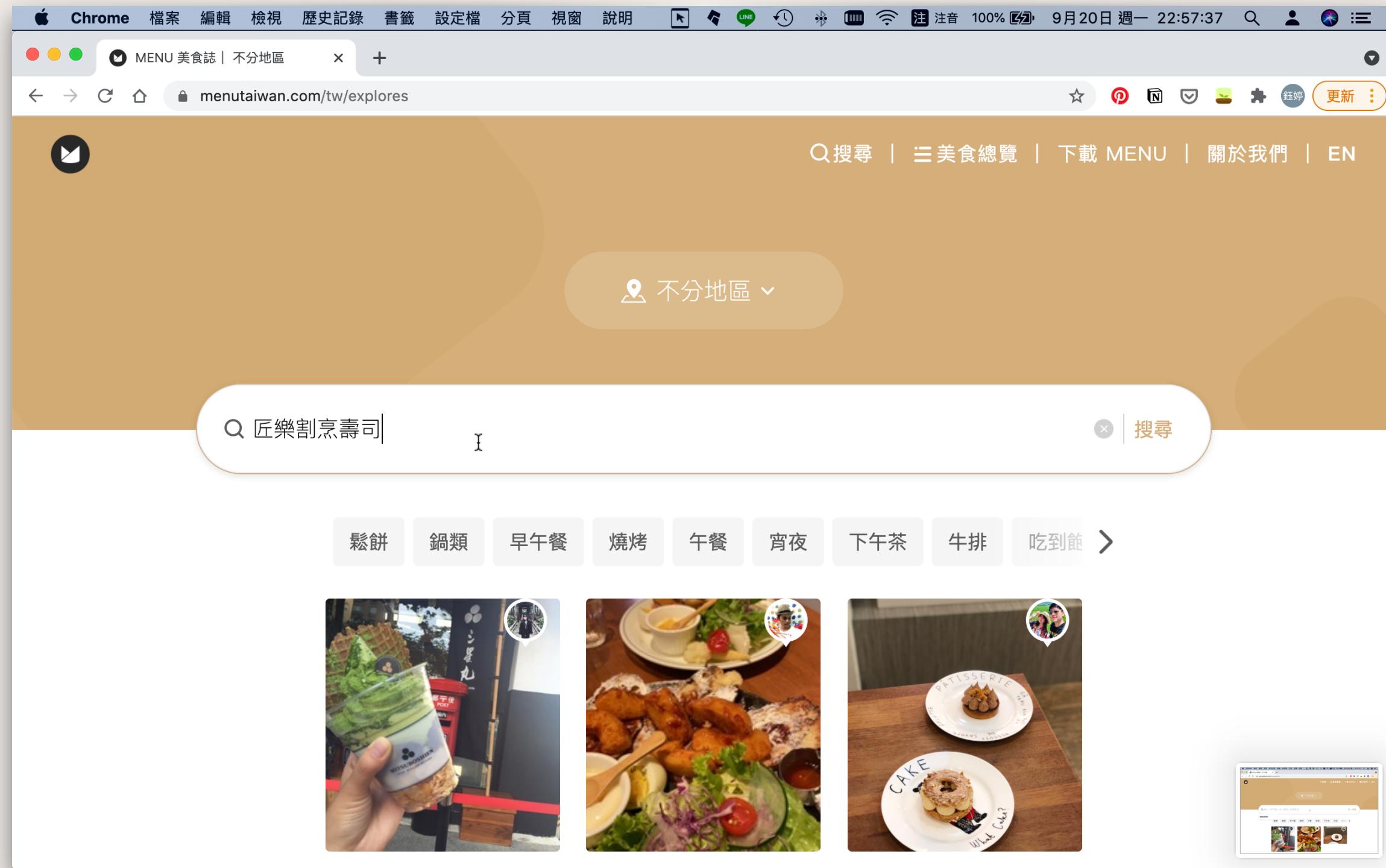
The screenshot shows a detailed view of a restaurant profile for "三槐堂" in Taipei, Zhongzheng District. The page includes a large image of the restaurant's exterior, a smaller image of a dessert, and a summary box with the restaurant's name, location, average rating (4.0), and a "closed" status. A "more" button is visible at the bottom right.



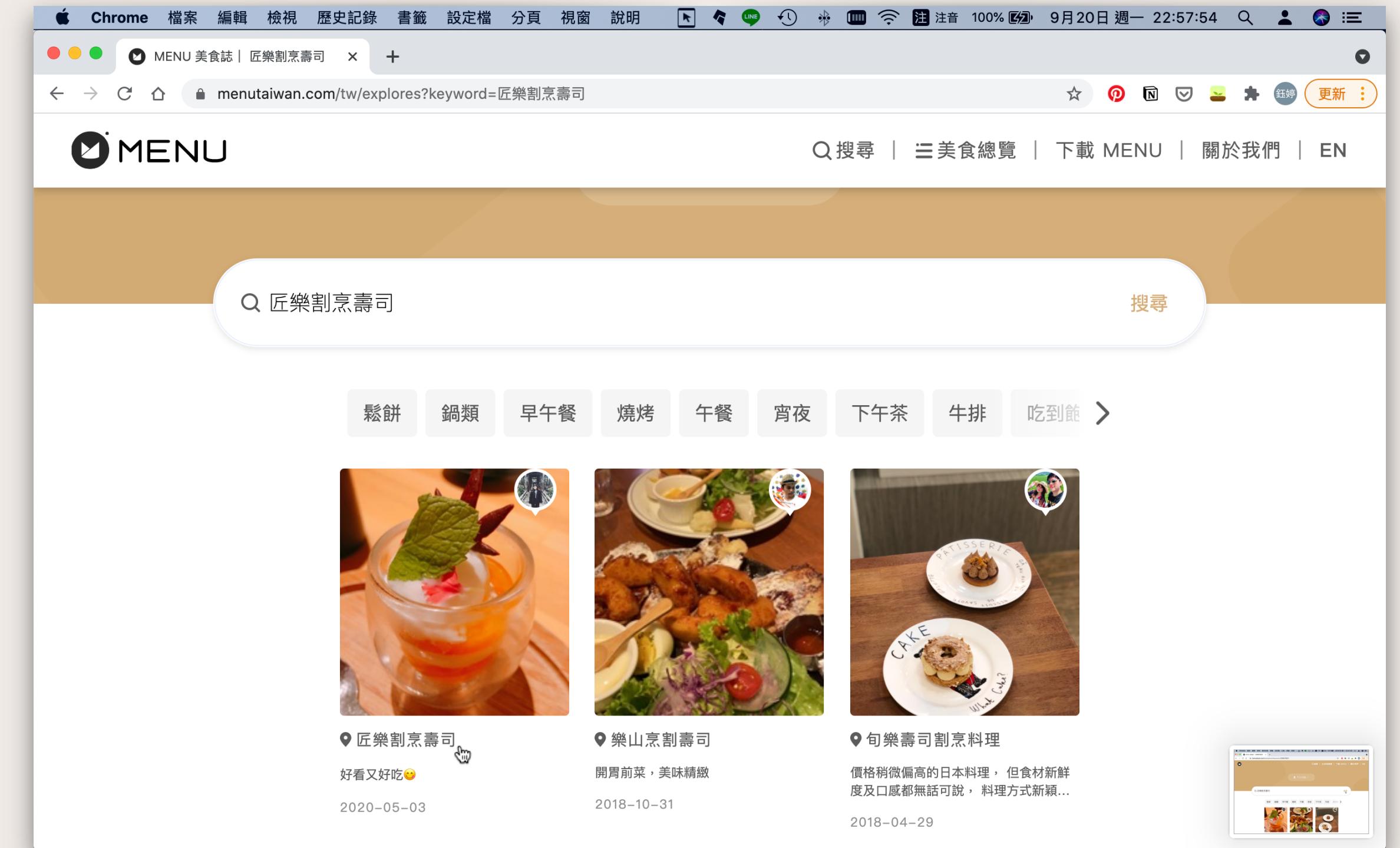
- 利用Python 的 Selenium 套件爬取資料
- 模擬使用者進入網頁後，點擊「more」按鍵，直至不可再點擊後，
獲取餐廳名稱、餐廳座落區域、餐廳總平均分數，儲存為 json 檔



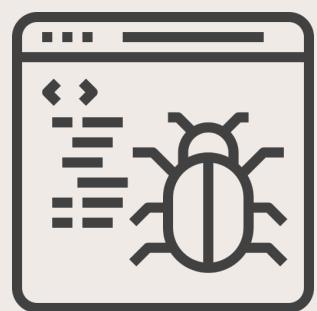
資料蒐集



A screenshot of the MENU website homepage. At the top, there's a navigation bar with links like '搜尋', '美食總覽', '下載 MENU', '關於我們', and 'EN'. Below the navigation is a search bar with the placeholder '不分地區' (Unspecified Region) and a dropdown arrow. A main search bar contains the text '匠樂割烹壽司'. Below the search bar is a horizontal menu with categories: 鬆餅, 鍋類, 早午餐, 燒烤, 午餐, 宵夜, 下午茶, 牛排, 吃到飽. Underneath this menu is a grid of three food images: a green smoothie, a plate of fried chicken, and a dessert. To the right of the grid is a small thumbnail of a computer screen displaying the same page.



A screenshot of the MENU website search results for '匠樂割烹壽司'. The search bar at the top contains the query '匠樂割烹壽司'. Below the search bar is a horizontal menu with categories: 鬆餅, 鍋類, 早午餐, 燒烤, 午餐, 宵夜, 下午茶, 牛排, 吃到飽. Three search results are displayed, each with a photo, a location pin, a restaurant name, a comment, and a date. The first result is '匠樂割烹壽司', with the comment '好看又好吃😊' and the date '2020-05-03'. The second result is '樂山烹割壽司', with the comment '開胃前菜，美味精緻' and the date '2018-10-31'. The third result is '旬樂壽司割烹料理', with the comment '價格稍微偏高的日本料理，但食材新鮮度及口感都無話可說，料理方式新穎...' and the date '2018-04-29'. A small thumbnail of a computer screen is visible in the bottom right corner.



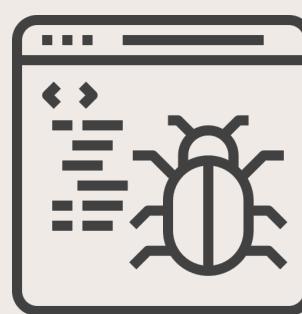
- 利用Python 的 Selenium 套件爬取資料
- 利用前頁取得的 json 檔，依序搜尋餐廳名稱，並點擊該餐廳，進入該餐廳資訊頁面



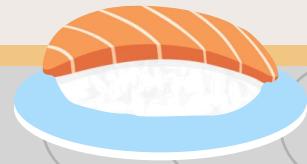
資料蒐集

This screenshot shows the homepage of the MENU website. At the top, there's a large banner featuring a dish. Below it, the restaurant's name '匠樂割烹壽司' is displayed in a prominent box along with a 5.0 rating. The main content area includes sections for '附近熱門' (Nearby Hotspots) and '餐廳服務' (Restaurant Services). A '叫車服務' (Car Hailing Service) button is highlighted in orange.

This screenshot shows the same restaurant page after scrolling down. It displays detailed information about the restaurant, including price range (\$1001 - \$2000), phone number (02 7709 1188), and address (台北市大安區四維路52巷31號). There's also a '帮我Google' (Help me Google) button. The '餐廳服務' (Restaurant Services) section is visible at the bottom, featuring a '叫車服務' (Car Hailing Service) button.



- 進入餐廳頁面，擷取餐廳名稱、評論者名稱、評論日期、評論內容
- 滾動至頁面底部，擷取餐廳價位、餐廳電話、餐廳地址
- 將上述資料，儲存為 json 檔



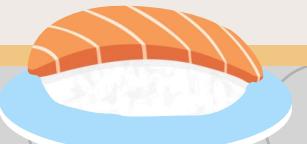
資料蒐集 - Google 限制爬取速度



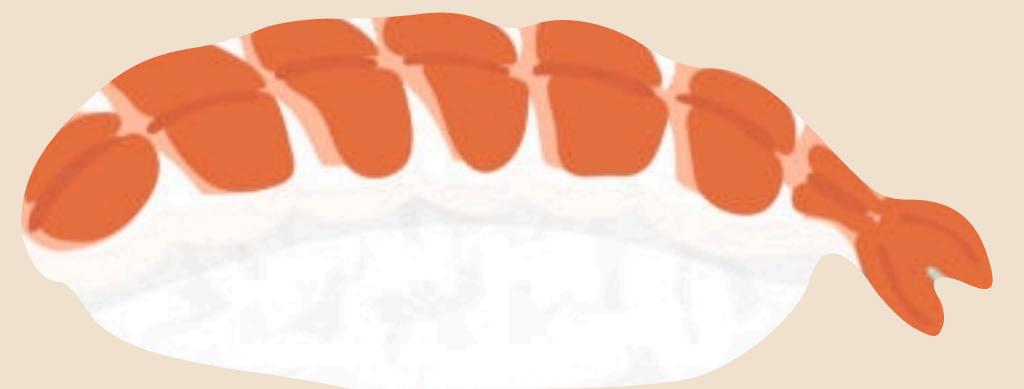
- 爬取 GoogleMaps 網站時，一旦被偵測到大量造訪頁面，便會跳出驗證機器人的頁面、甚至完全無法造訪網頁、被限制取得資訊速度



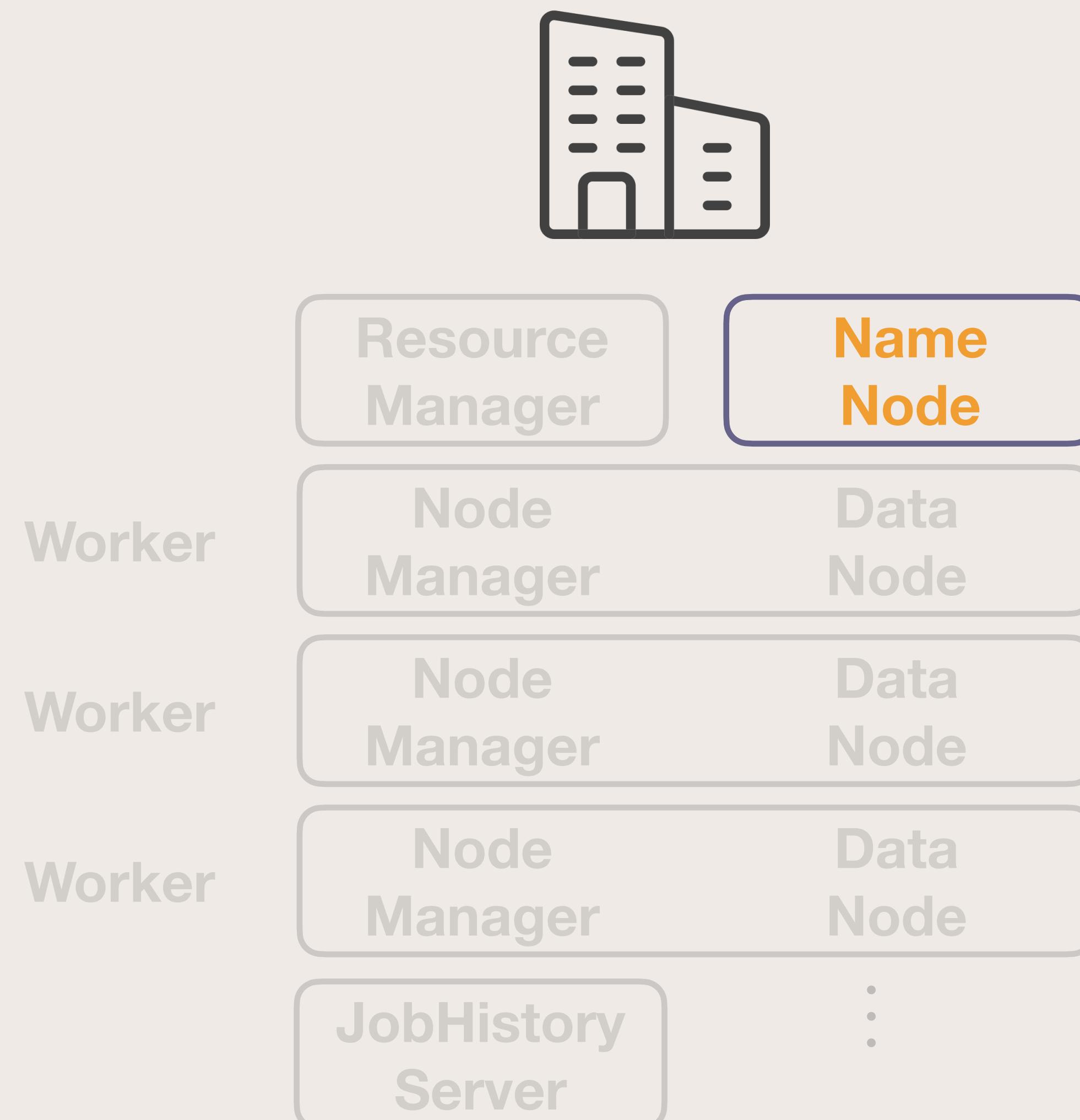
- 問題一：跳出驗證機器人**
- 問題二：JS模擬滾動面臨out of memory**
- 問題三：被限制資料爬取速度**
對策 1：透過 Google 釋出的 API
→ 每間餐廳被限制只可取得 5 則評論
對策 2：使用多台電腦同步爬取，分散工作量給多台電腦，降低被檔的風險



系統建置



系統架構

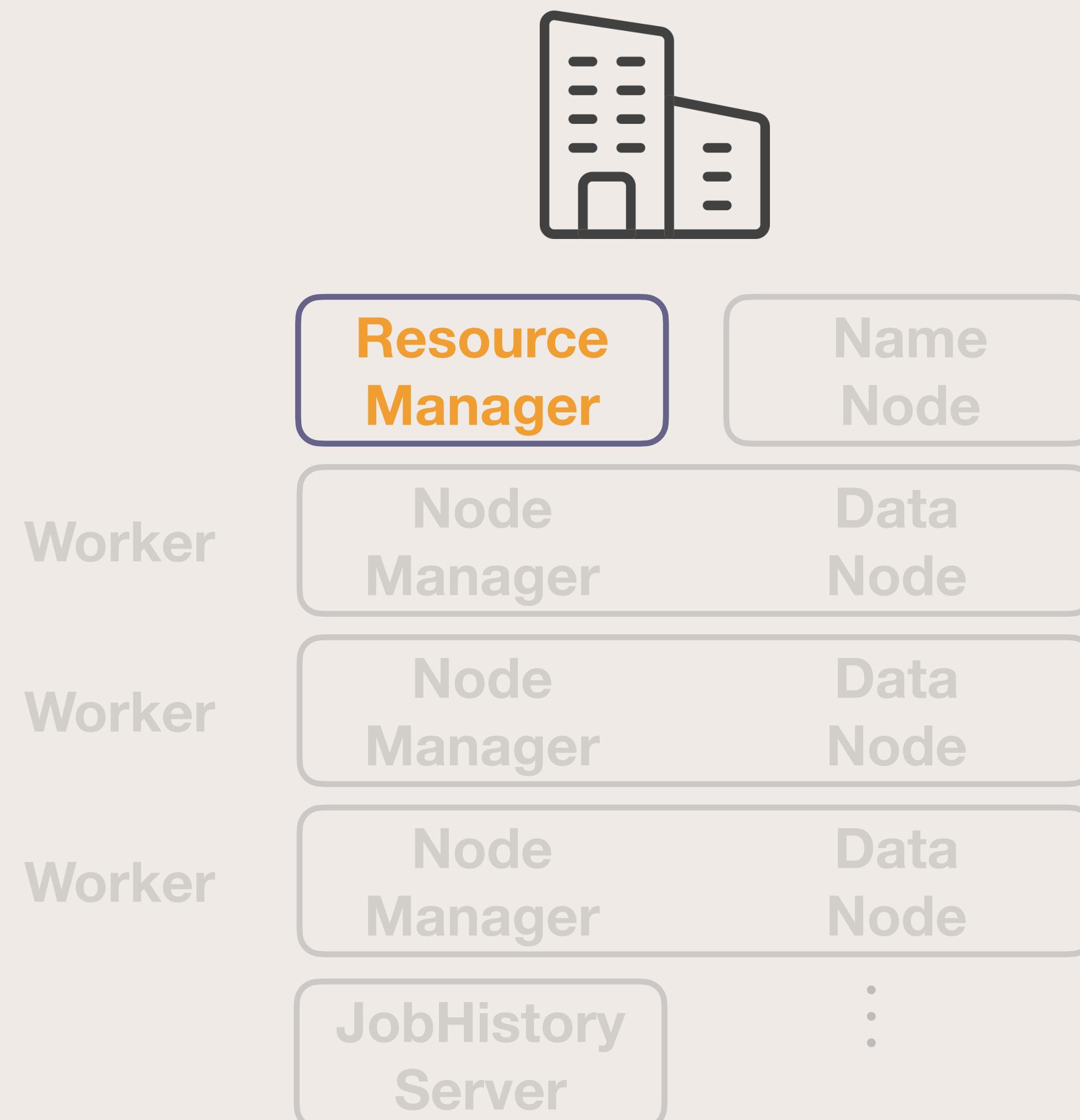


Name Node

- 負責管理和維護 HDFS 的 MetaData
- 控制檔案所有讀寫動作
- 包括檔案 block 存取及讀寫的 DataNode 位置



系統架構

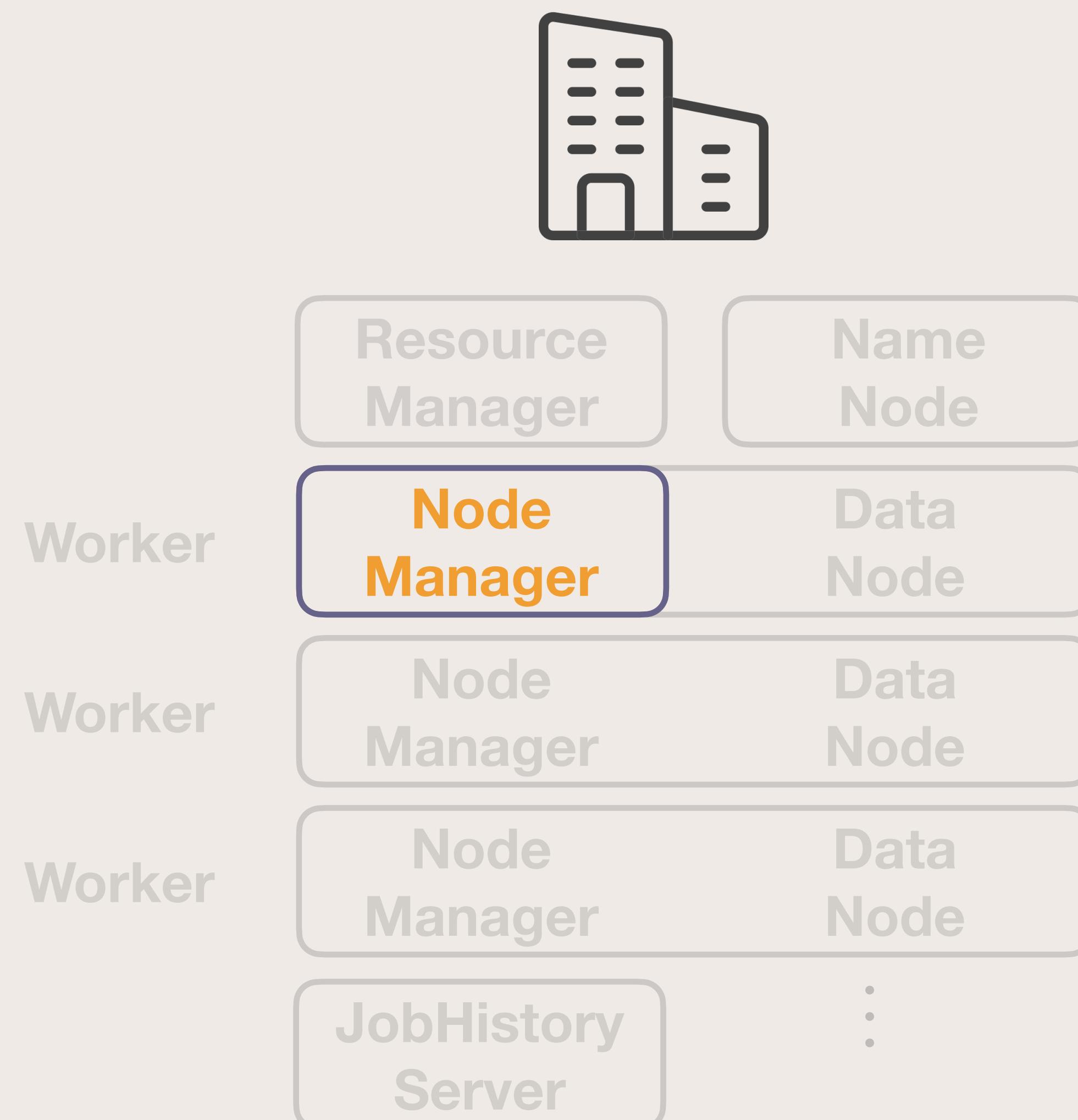


Resource Manager

- 負責管理和分配叢集的運算資源
- 接受 Client 端請求後，進行資源分配



系統架構

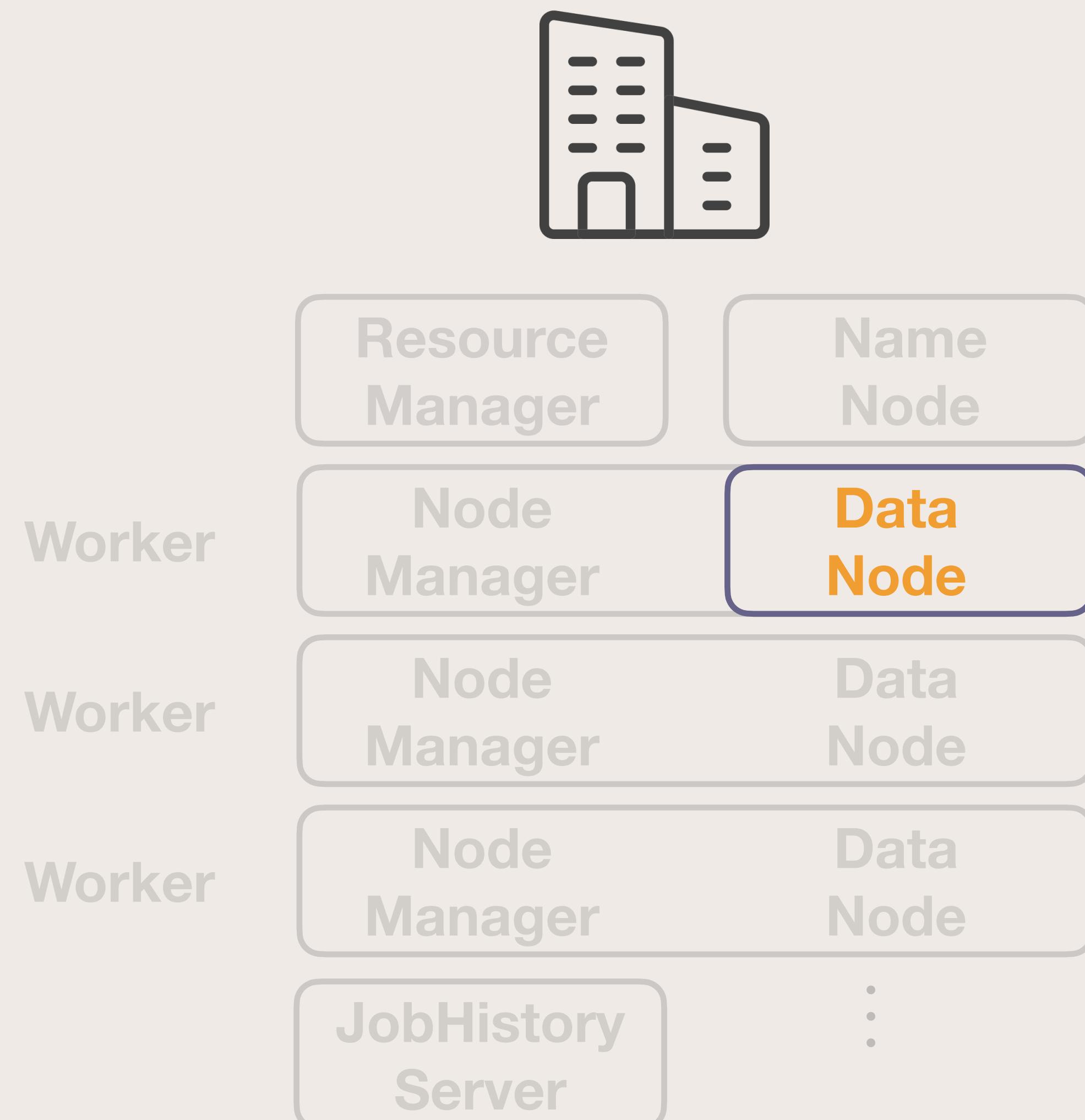


Node Manager

- 負責傳送 block 列表給 NameNode
- 讀取 NameNode 指定的 block
- 負責監控叢集內資源使用情況（如：memory, cpu, disk, network等）
- 回報資源使用情形給 ResourceManager

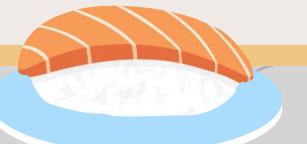


系統架構

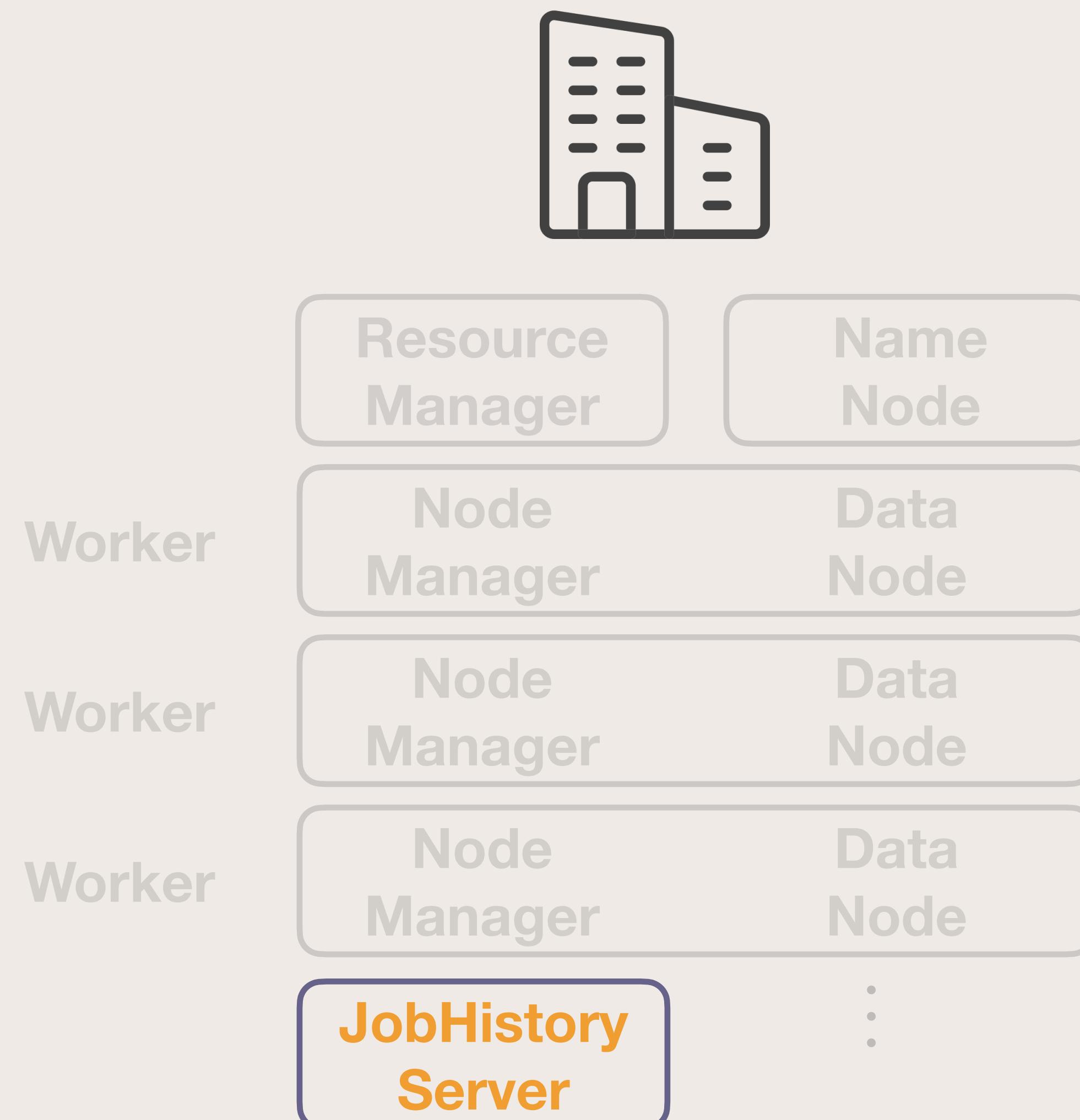


Data Node

- 管理 Client 端傳送的檔案資料
- Client 端切分完的每個檔案資料
皆於多個 DataNode 上存儲副本

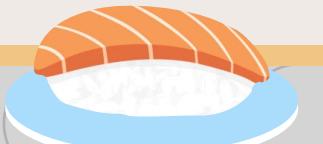


系統架構

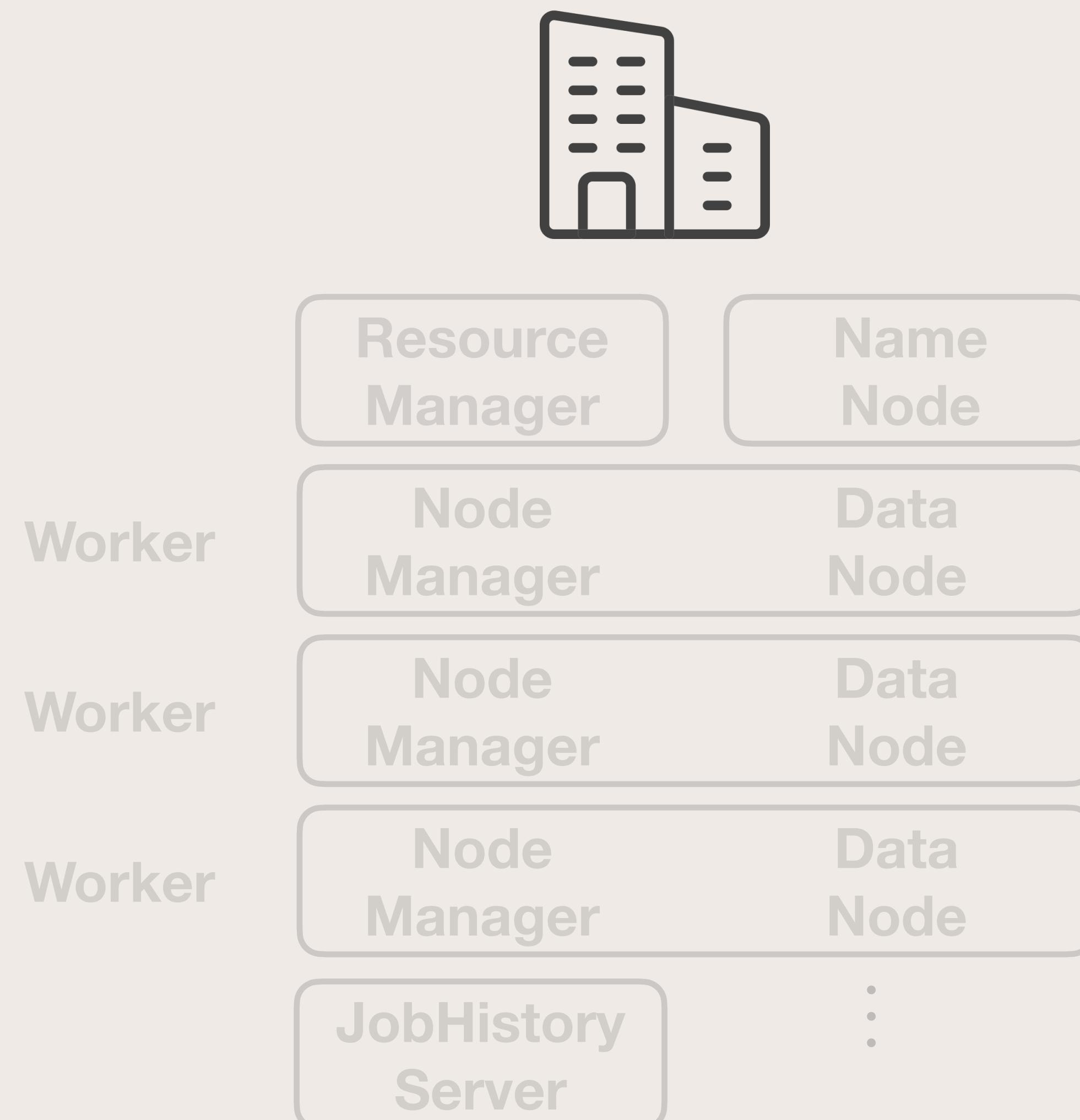


Job History Server

- 紀錄 DataNode 程序作業

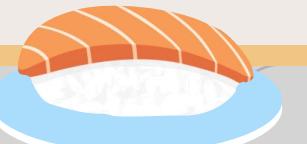


系統架構

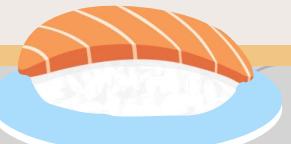
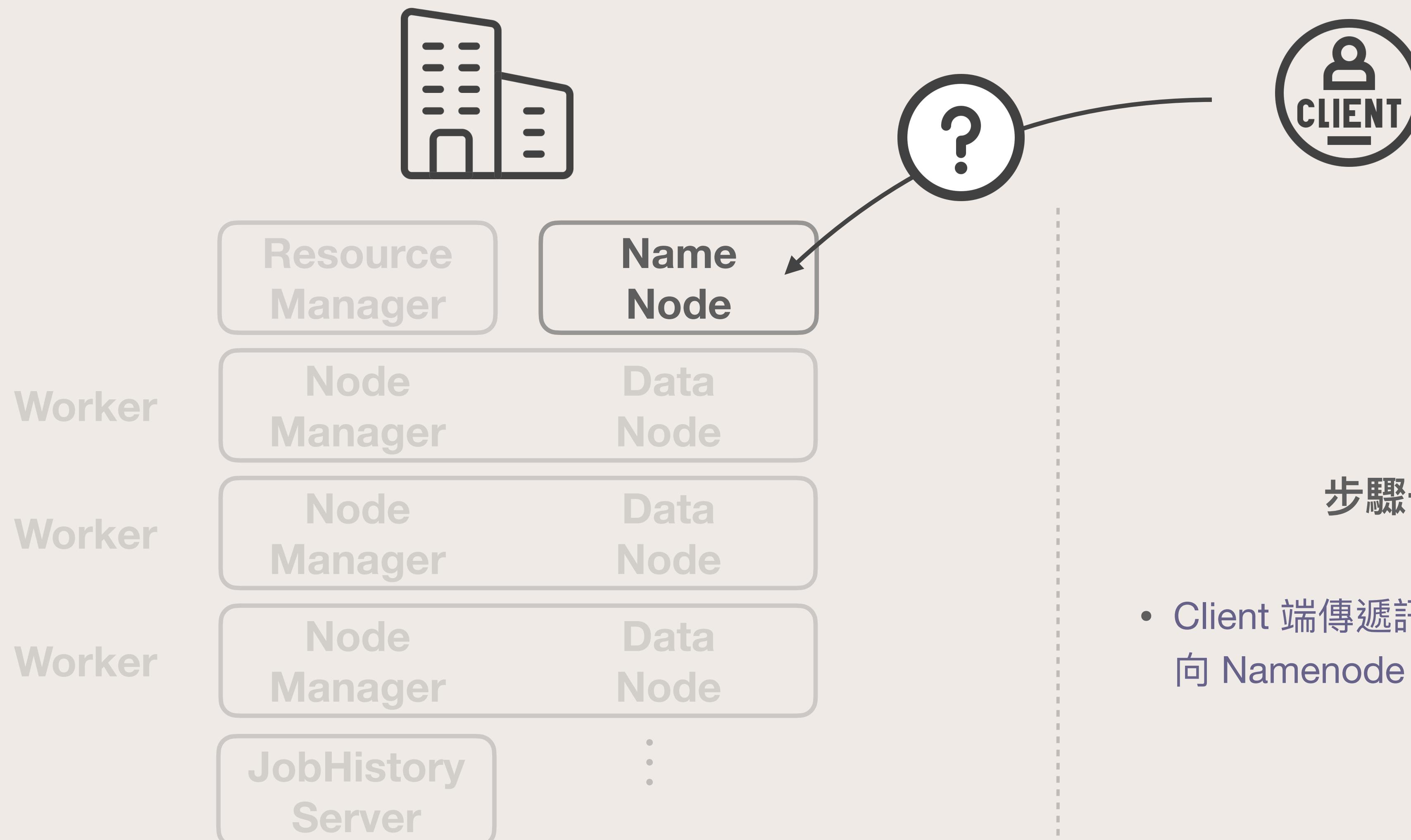


Client

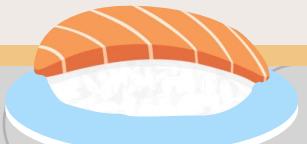
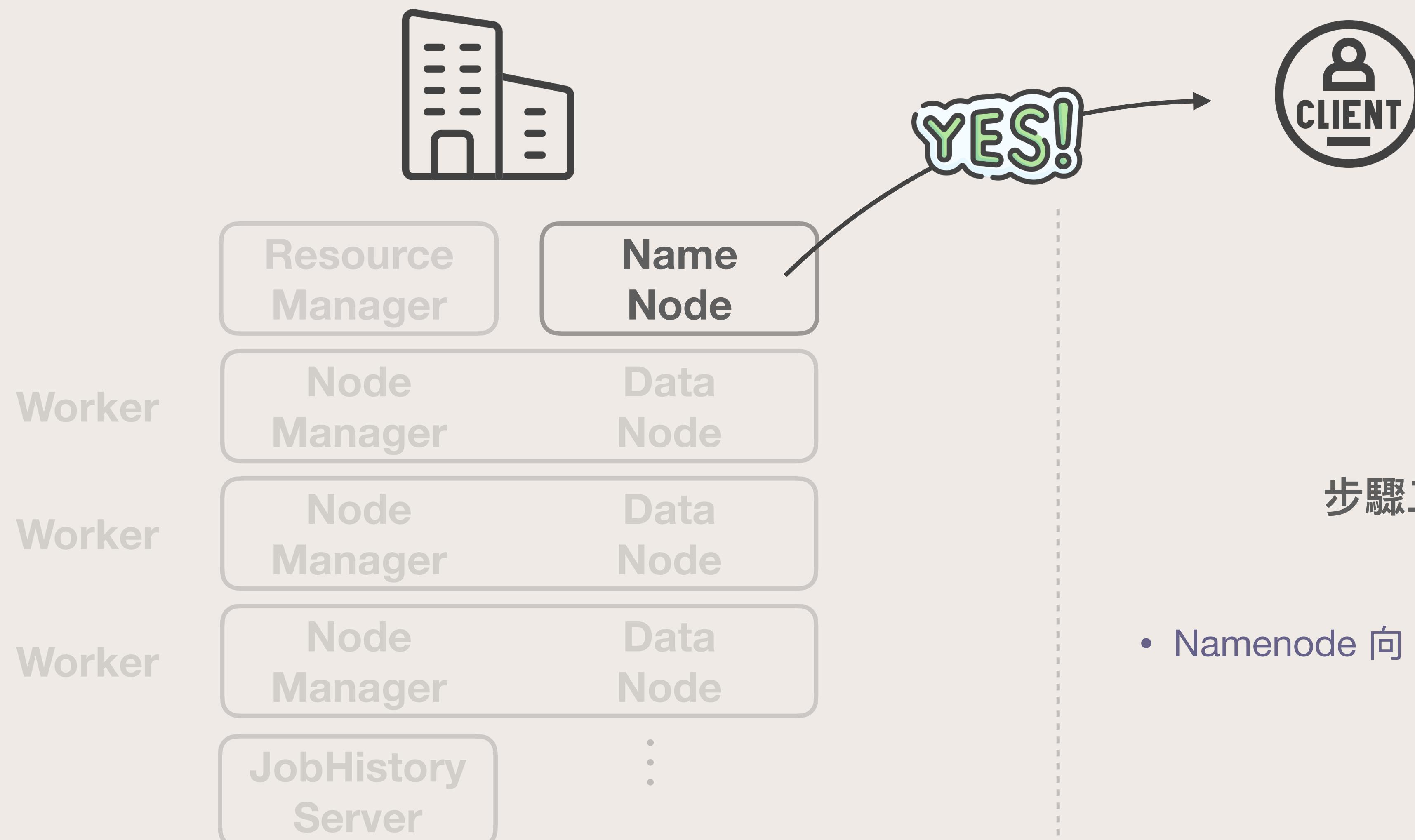
- 亦稱 Edge Node、Gateway Node
- 叢集內部、外部的邊緣節點都可為 Client 端
- 向叢集發送程式、進行請求



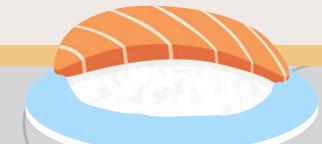
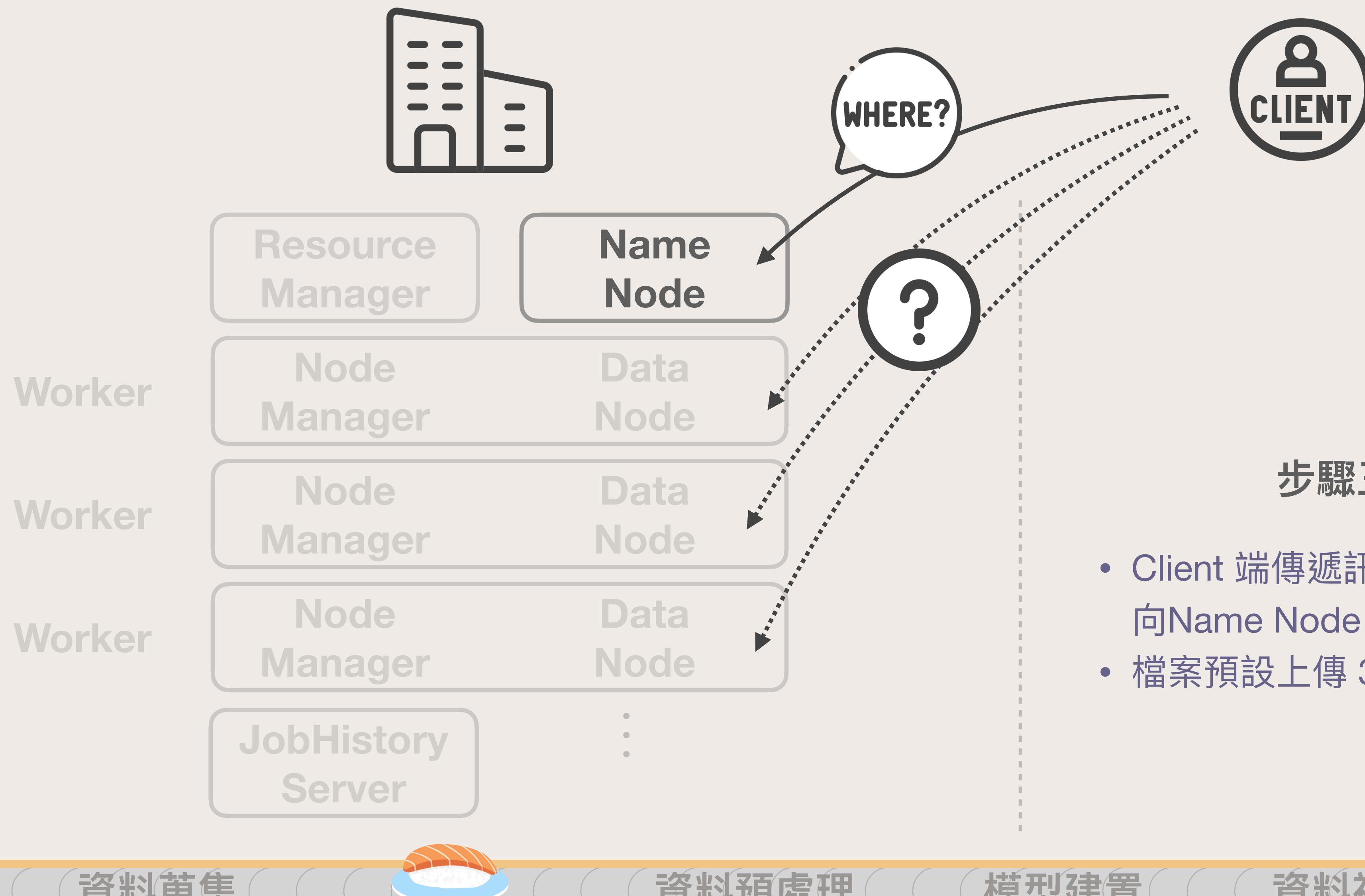
系統運作



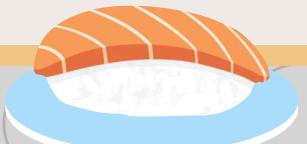
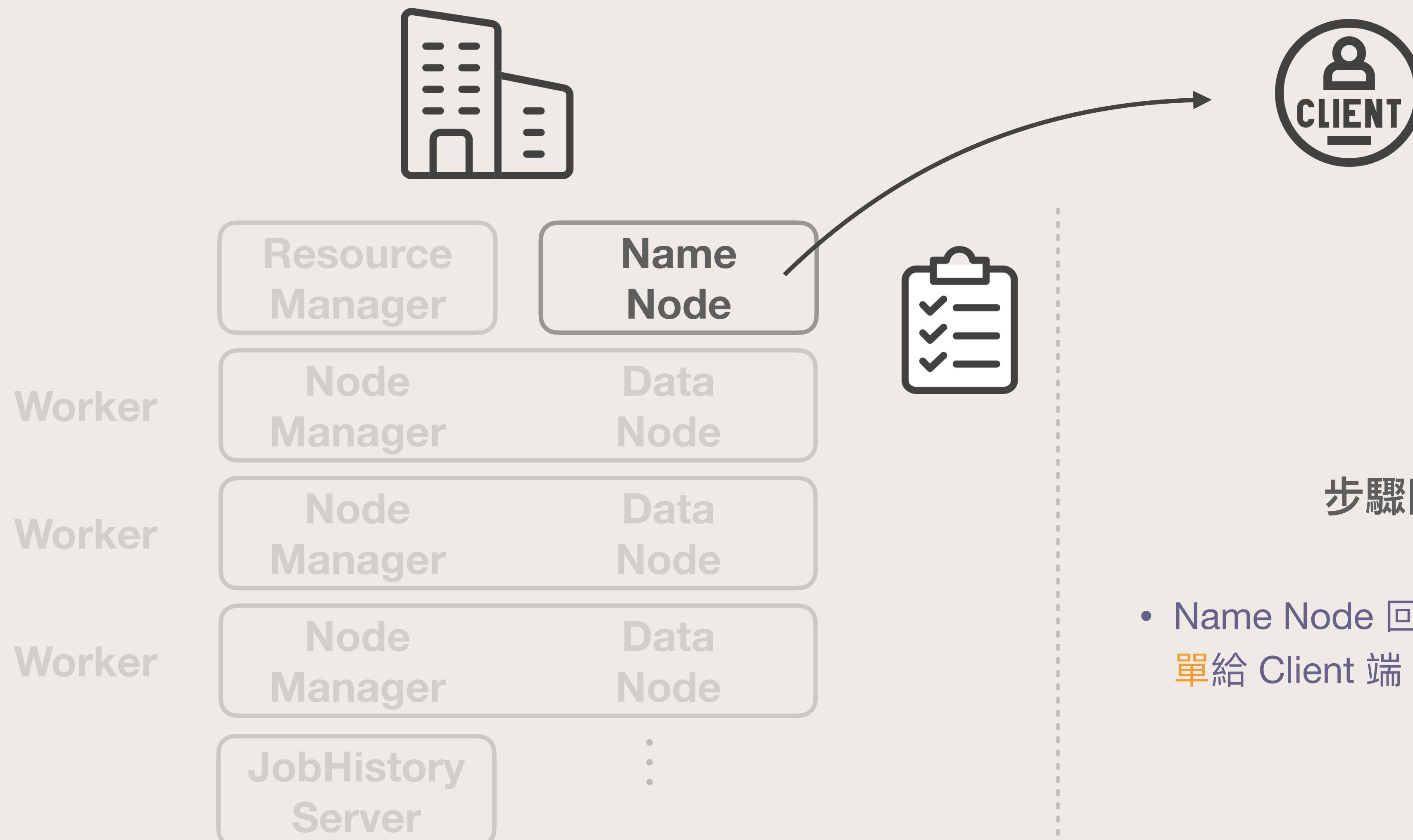
系統運作



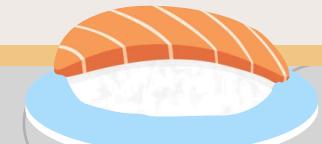
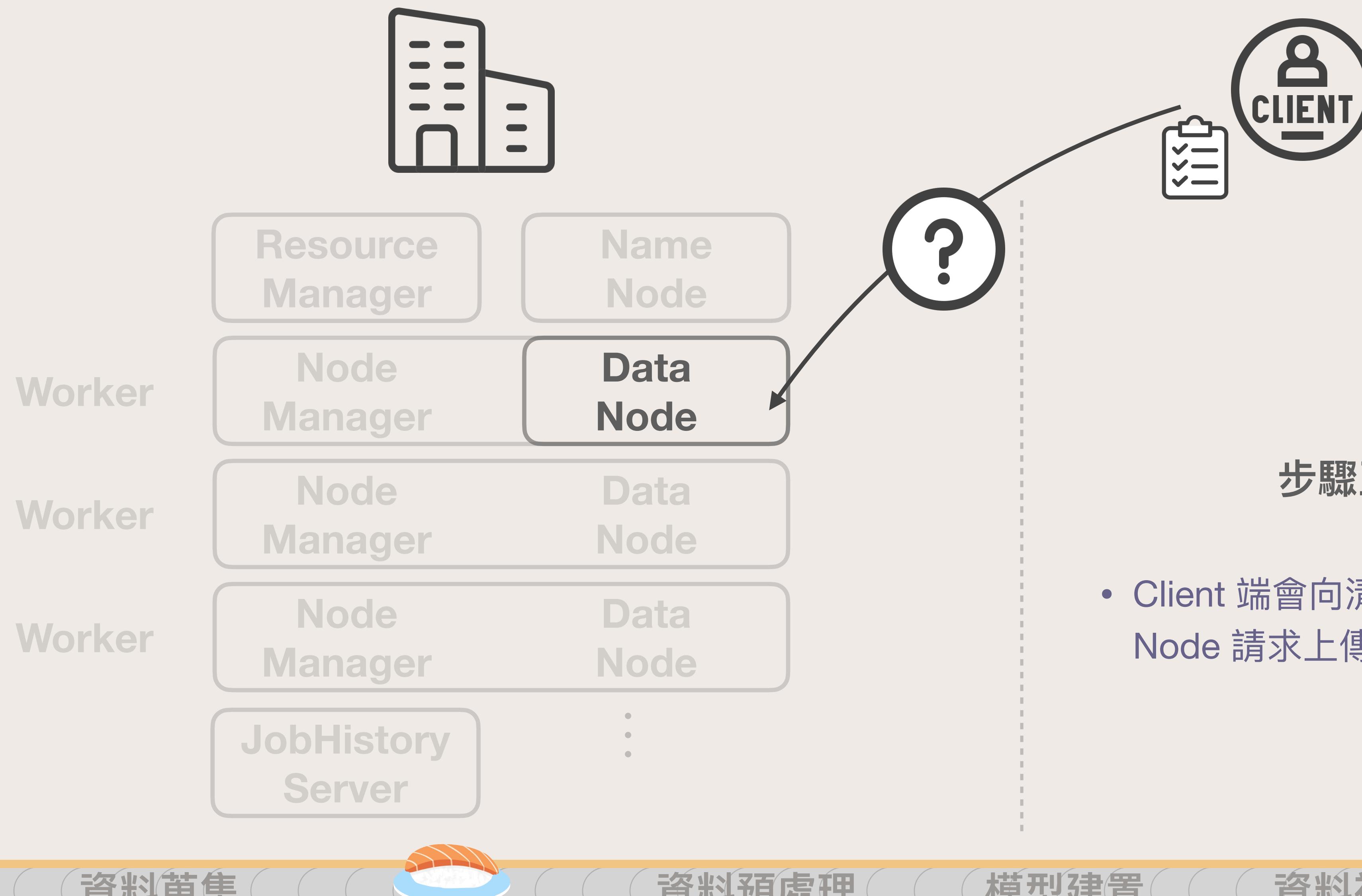
系統運作



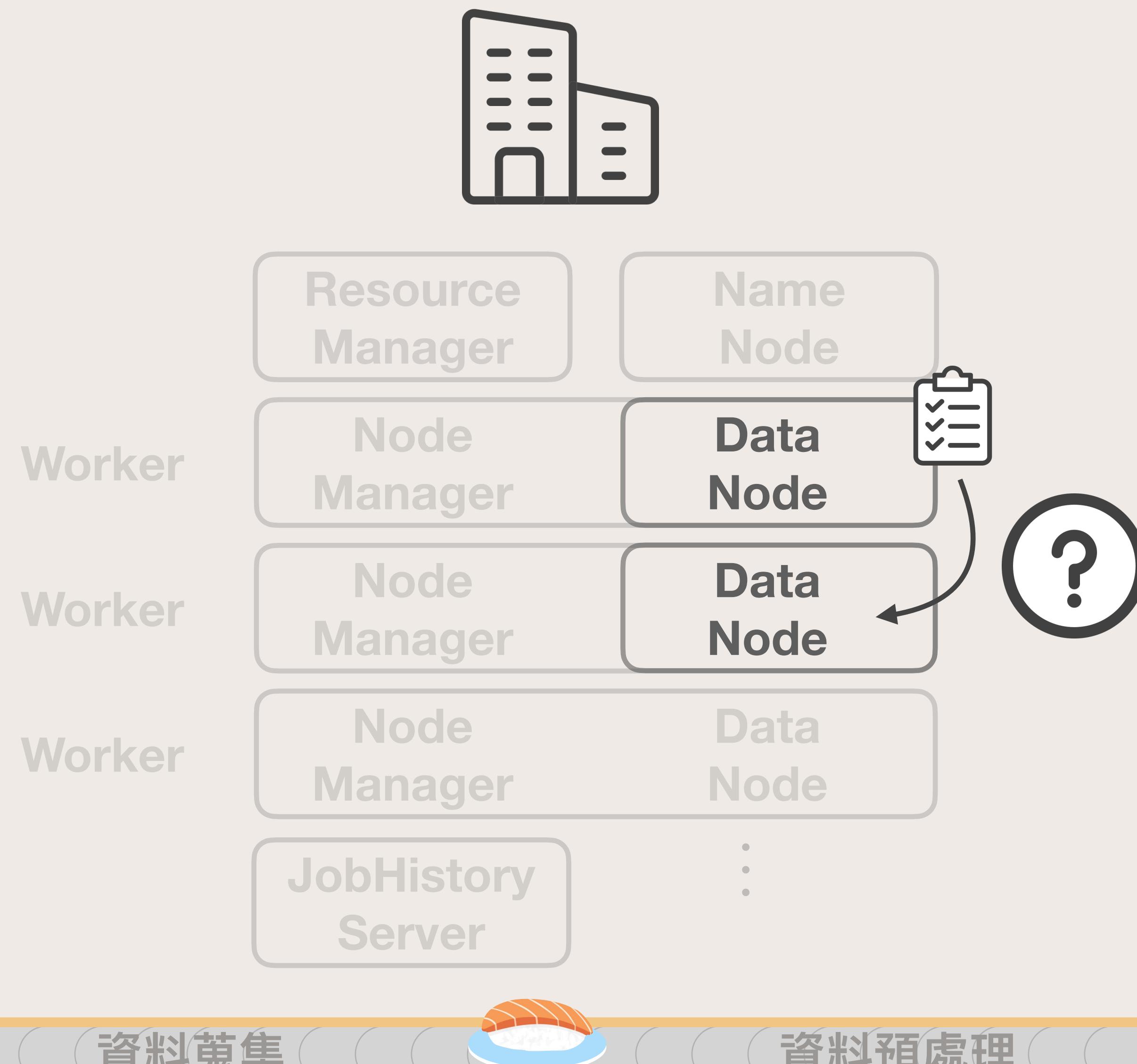
系統運作



系統運作

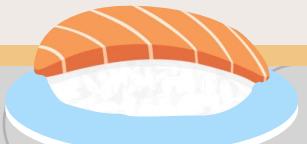


系統運作

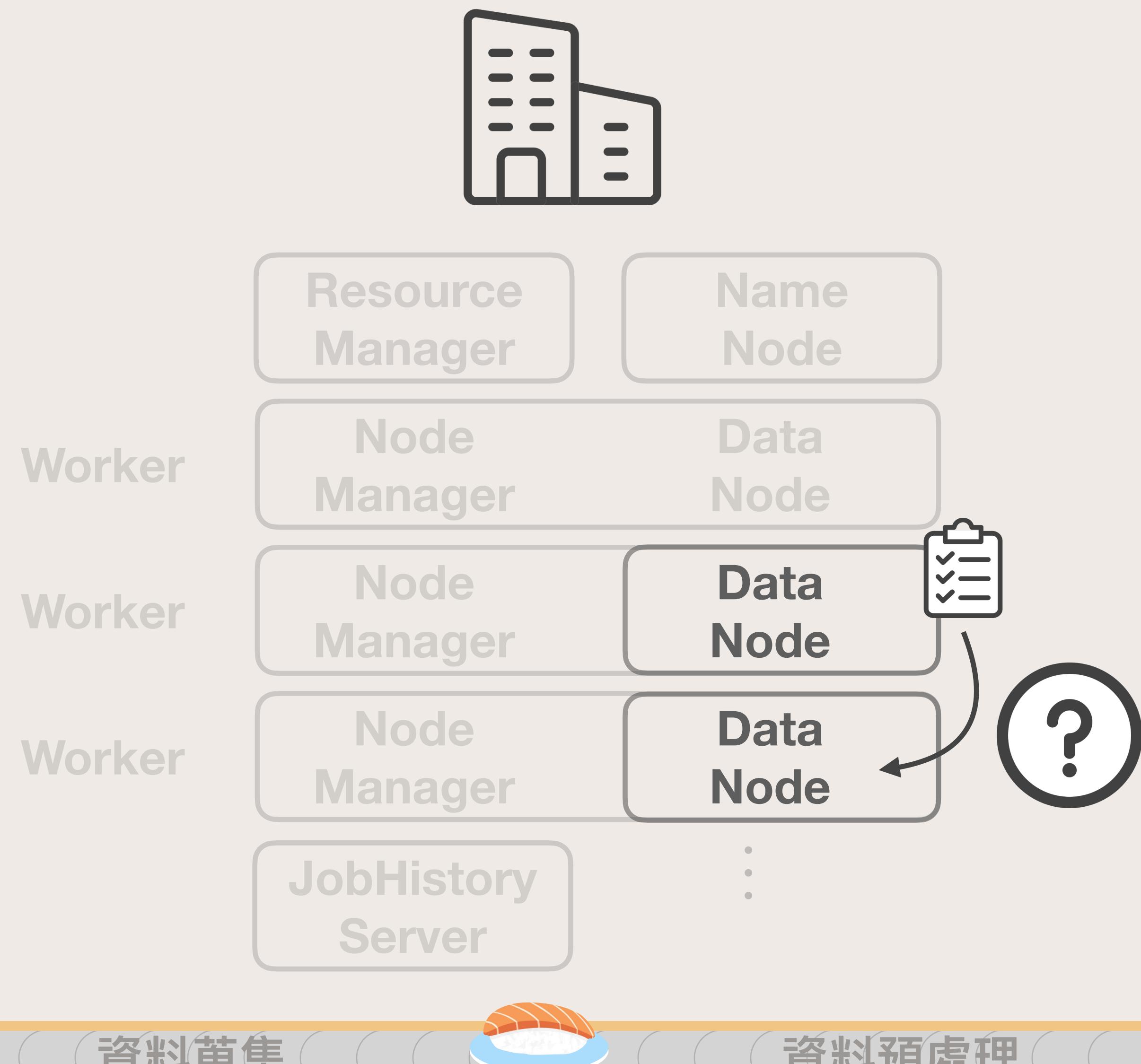


步驟六

- Data Node 會依序向清單上的其他 Data Node 請求上傳檔案

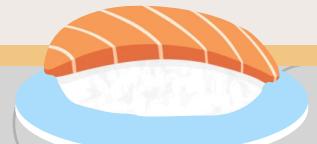


系統運作

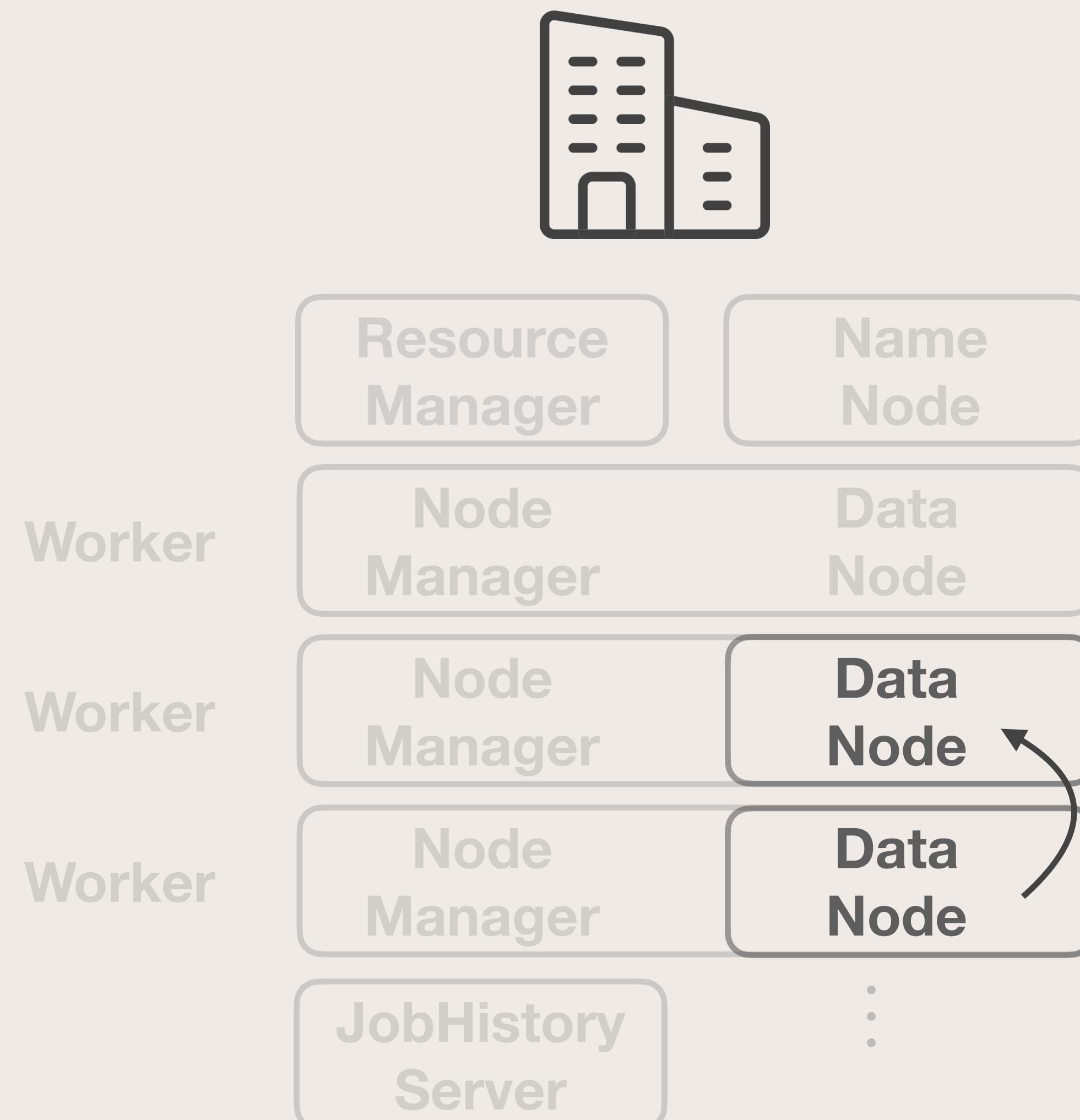


步驟七

- Data Node 會依序向清單上的其他 Data Node 請求上傳檔案
- 依序傳送請求，完成通信管道建立

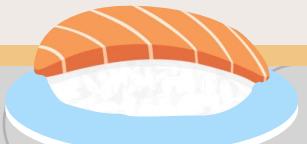


系統運作

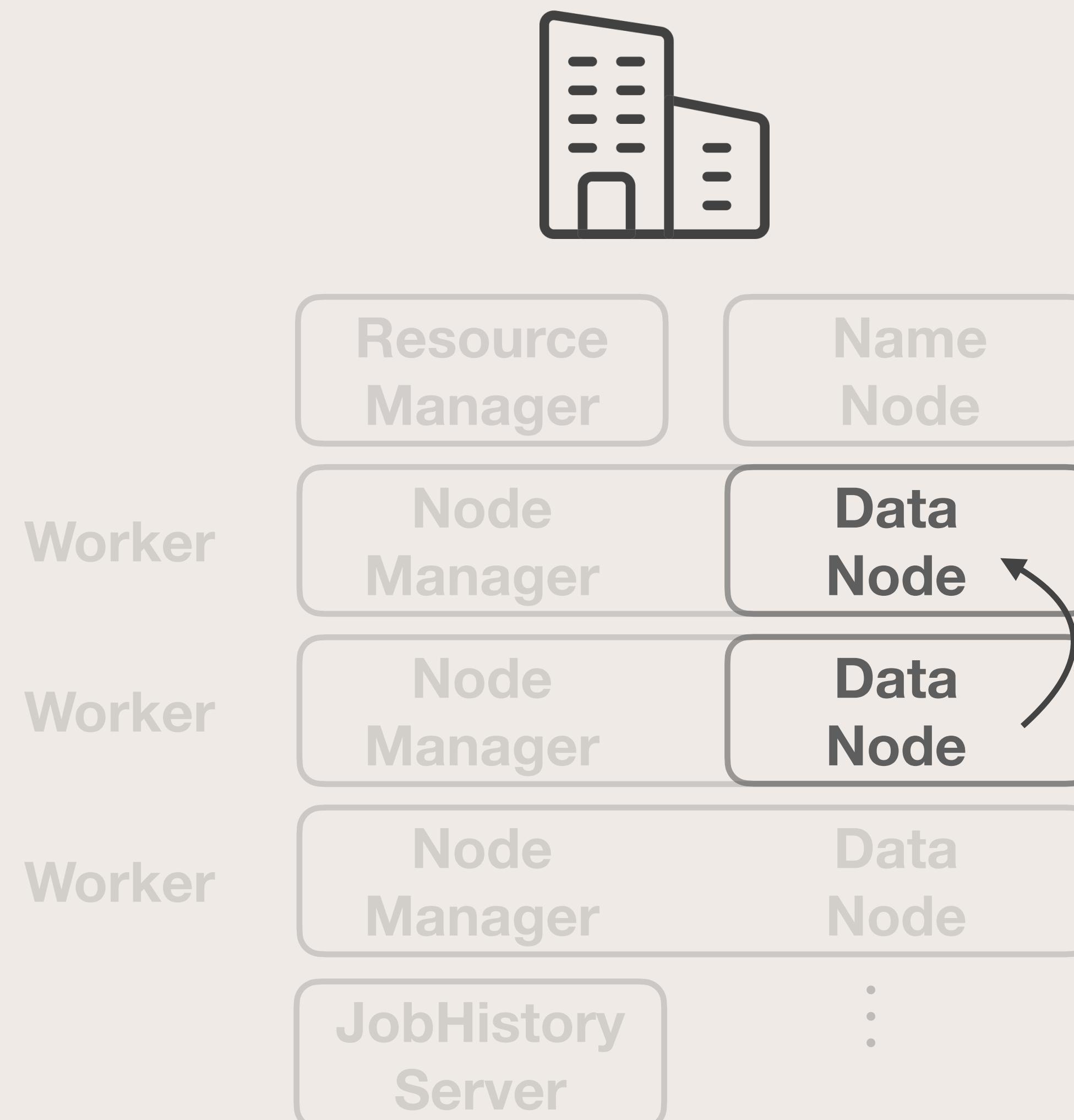


步驟八

- Data Node 依序回應 Client 端的請求



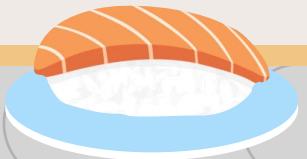
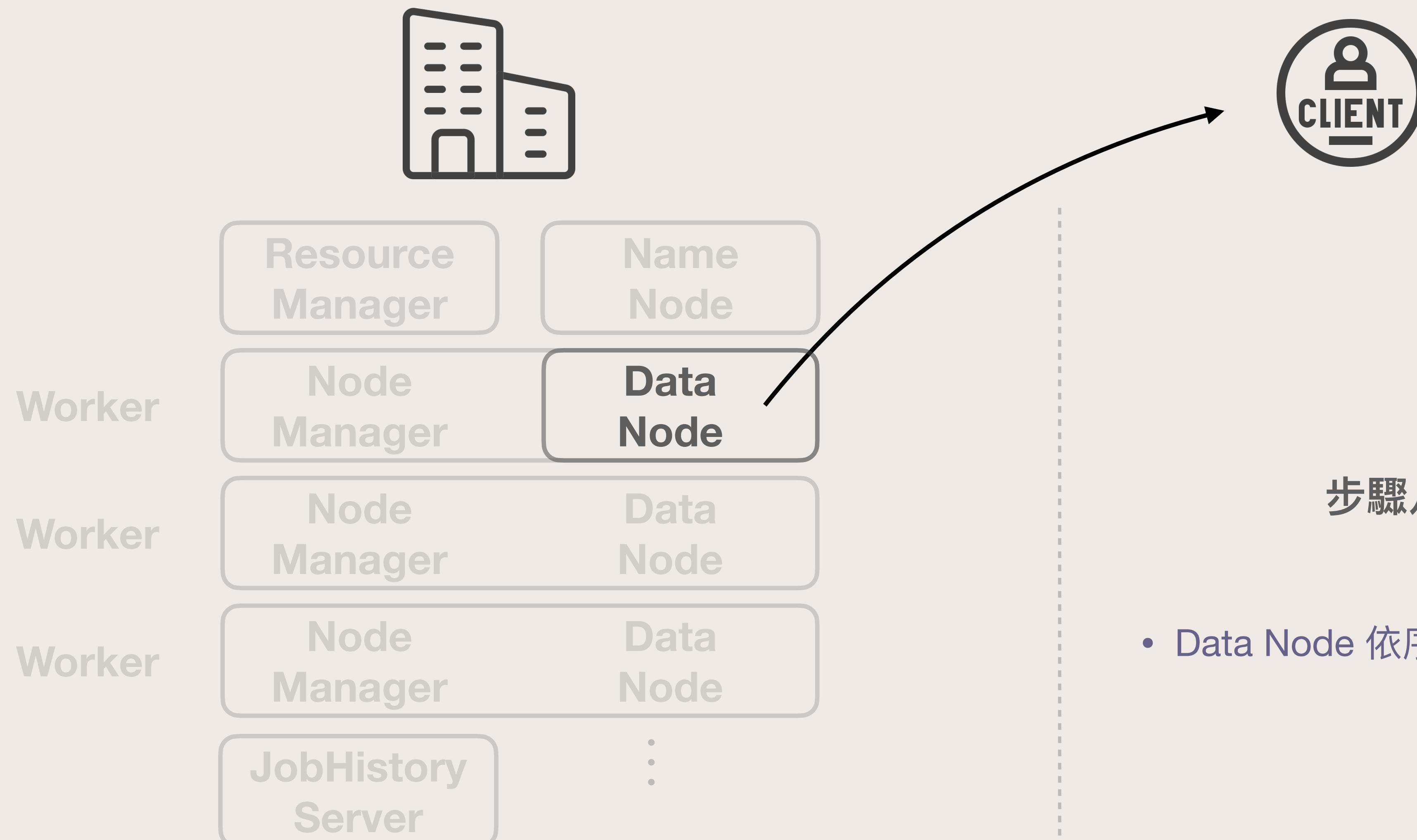
系統運作



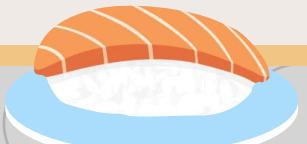
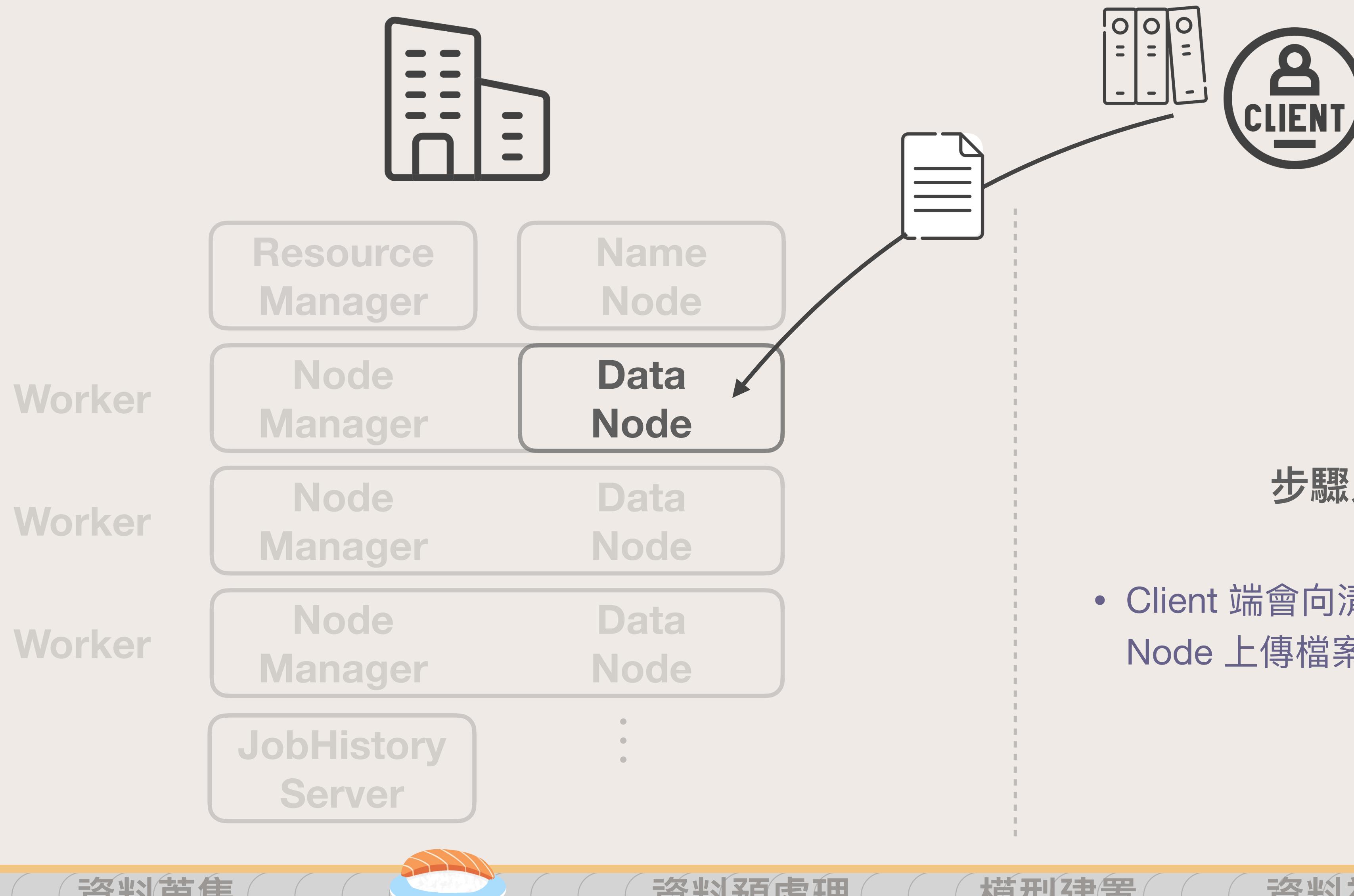
步驟八

- Data Node 依序回應 Client 端的請求

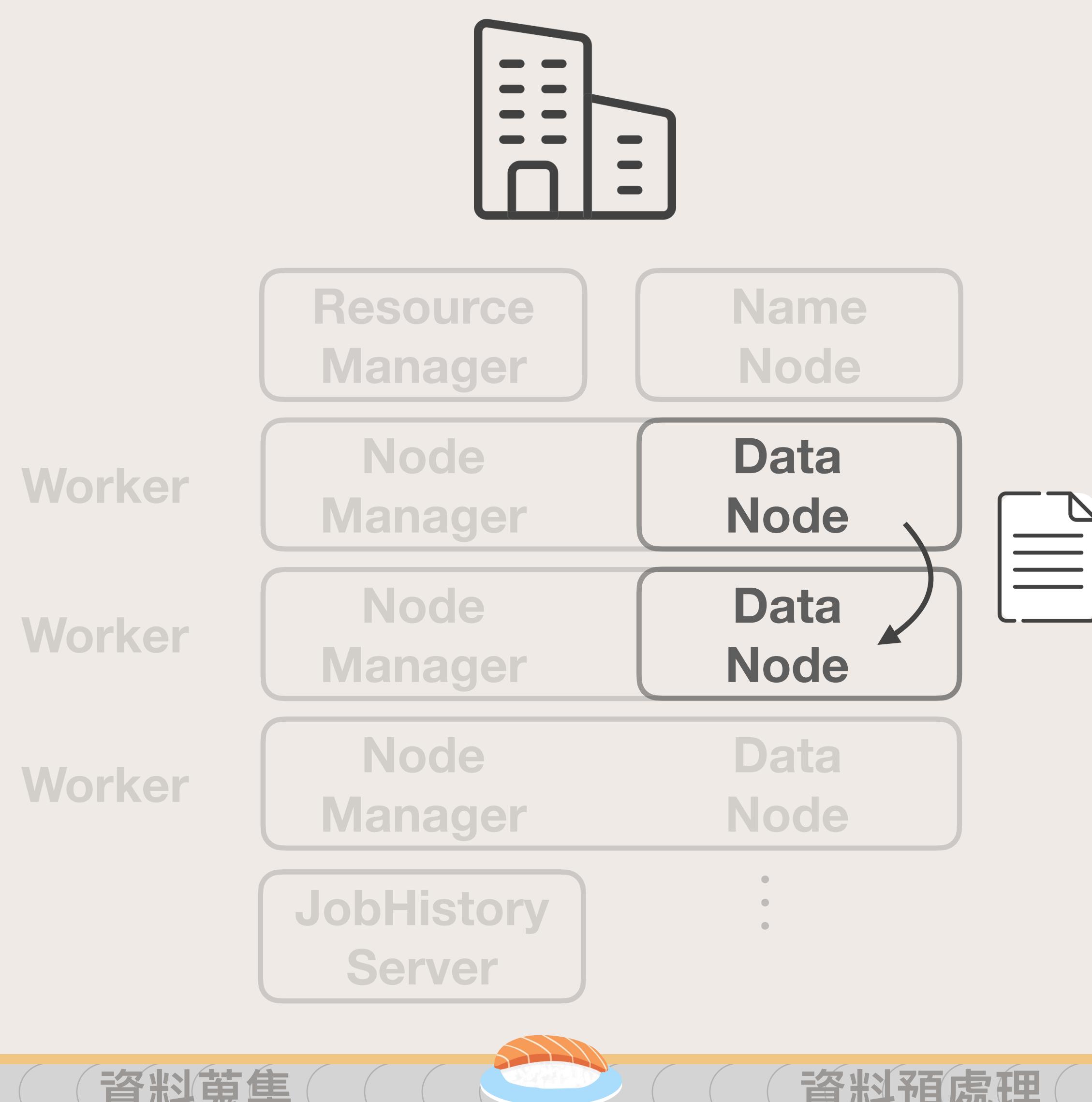
系統運作



系統運作

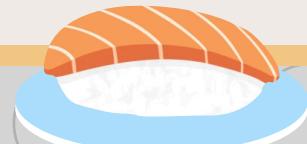


系統運作

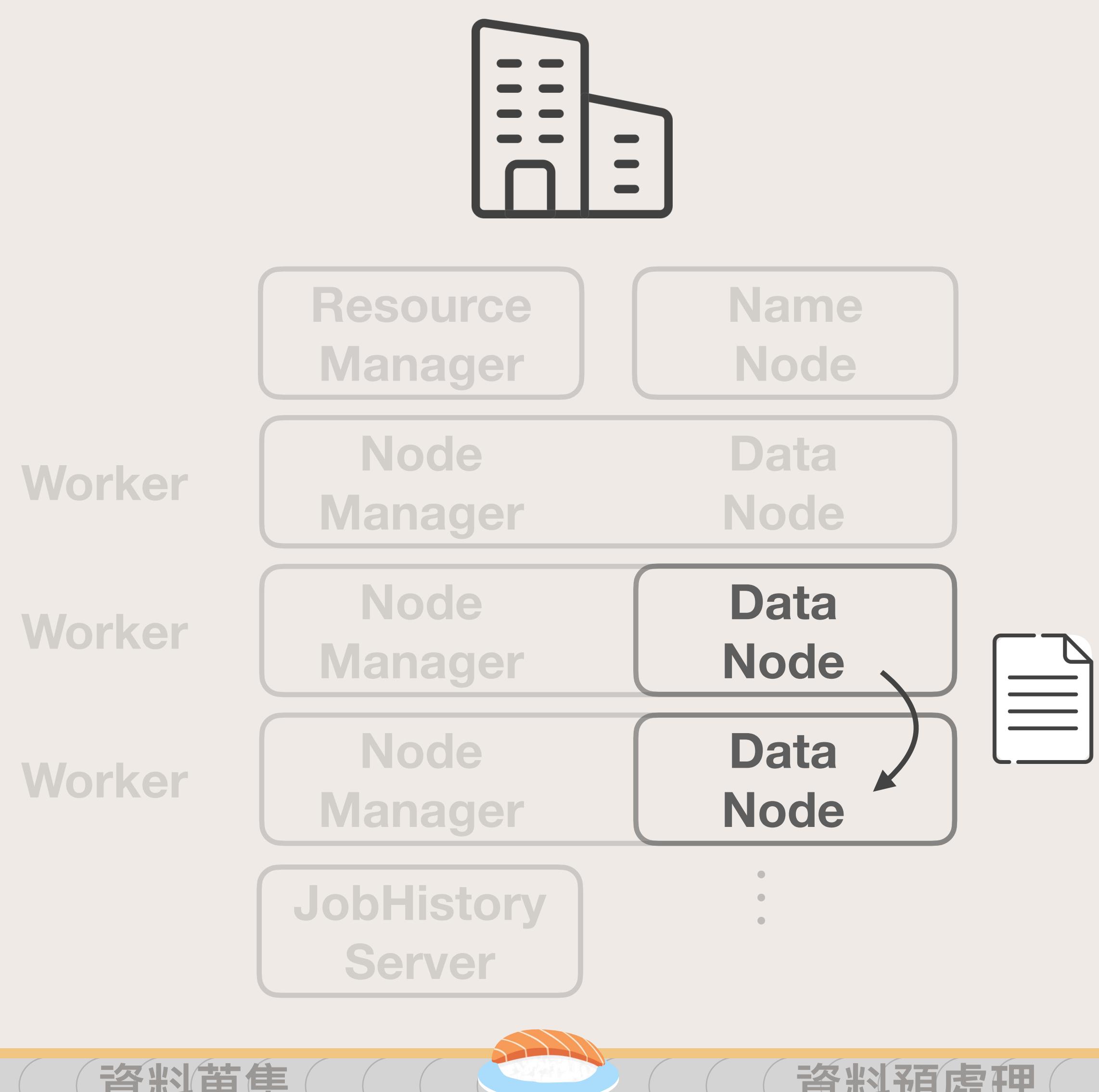


步驟十

- Data Node 依序上傳檔案至清單上的其他 Data Node

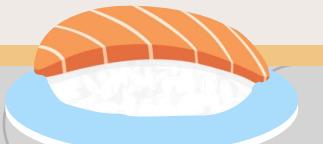


系統運作

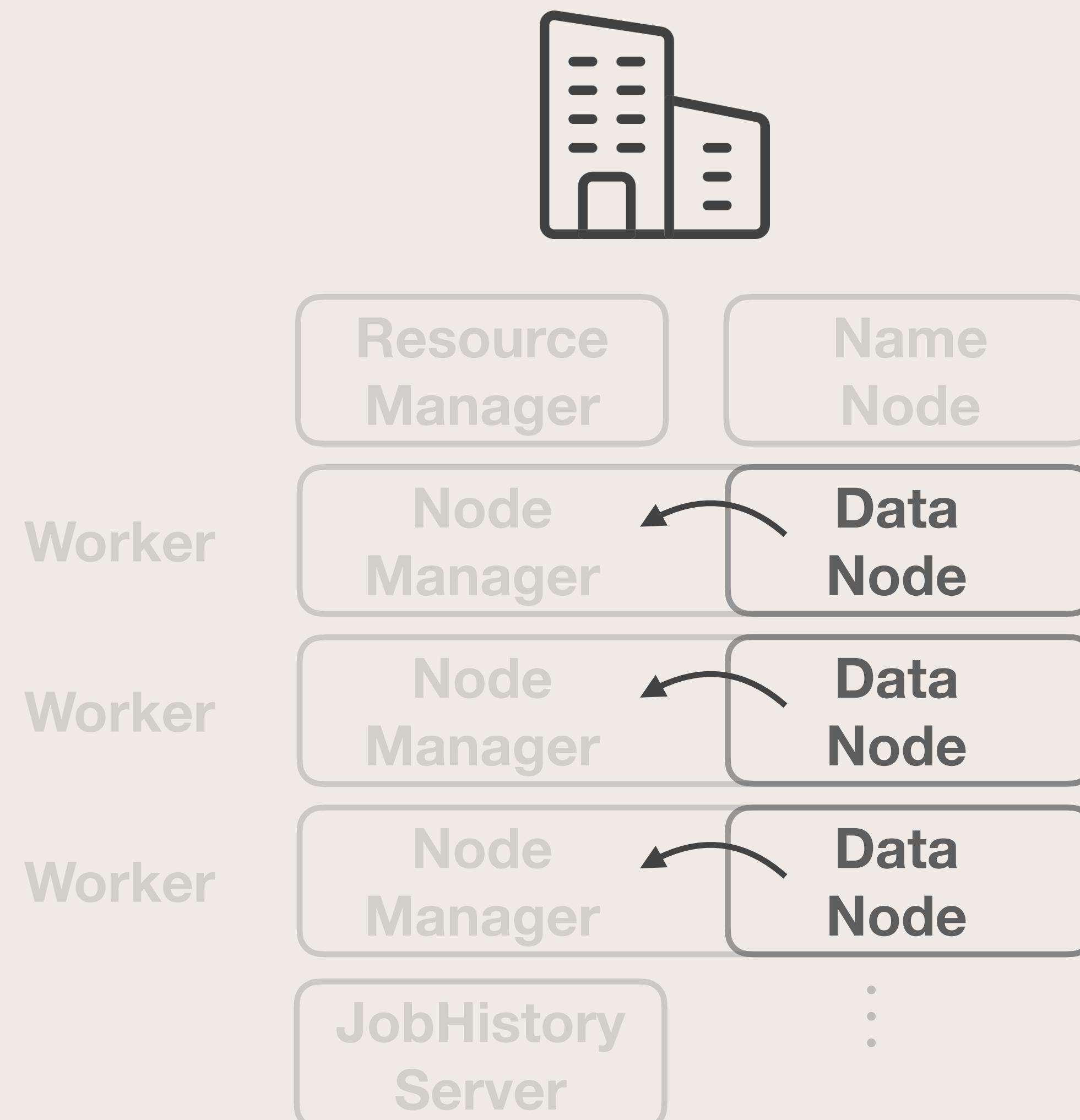


步驟十

- Data Node 依序上傳檔案至清單上的其他 Data Node

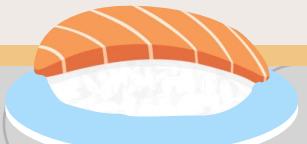


系統運作

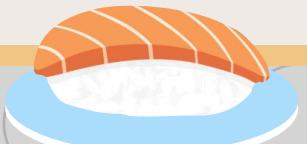
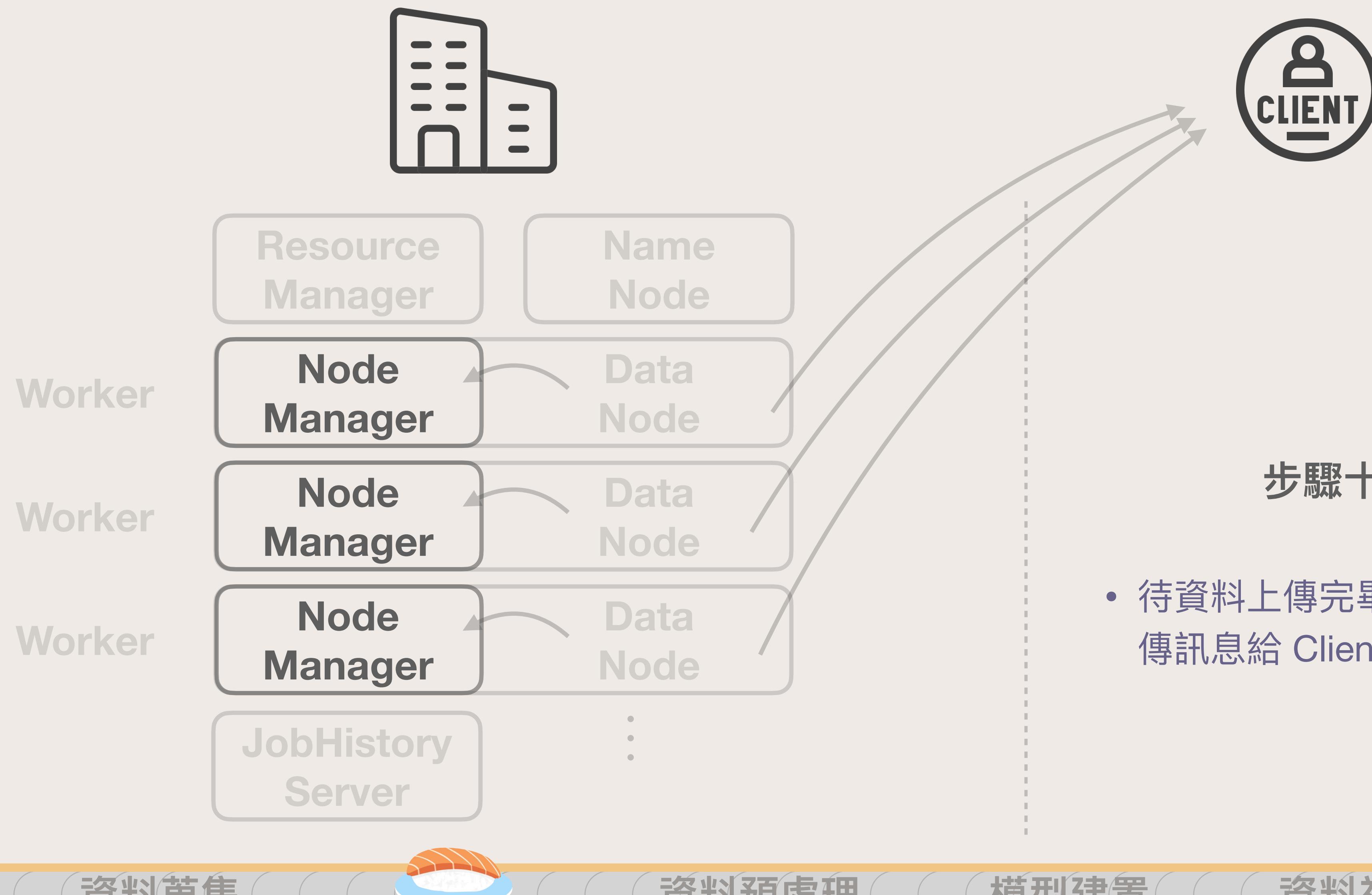


步驟十一

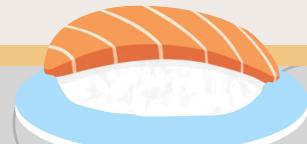
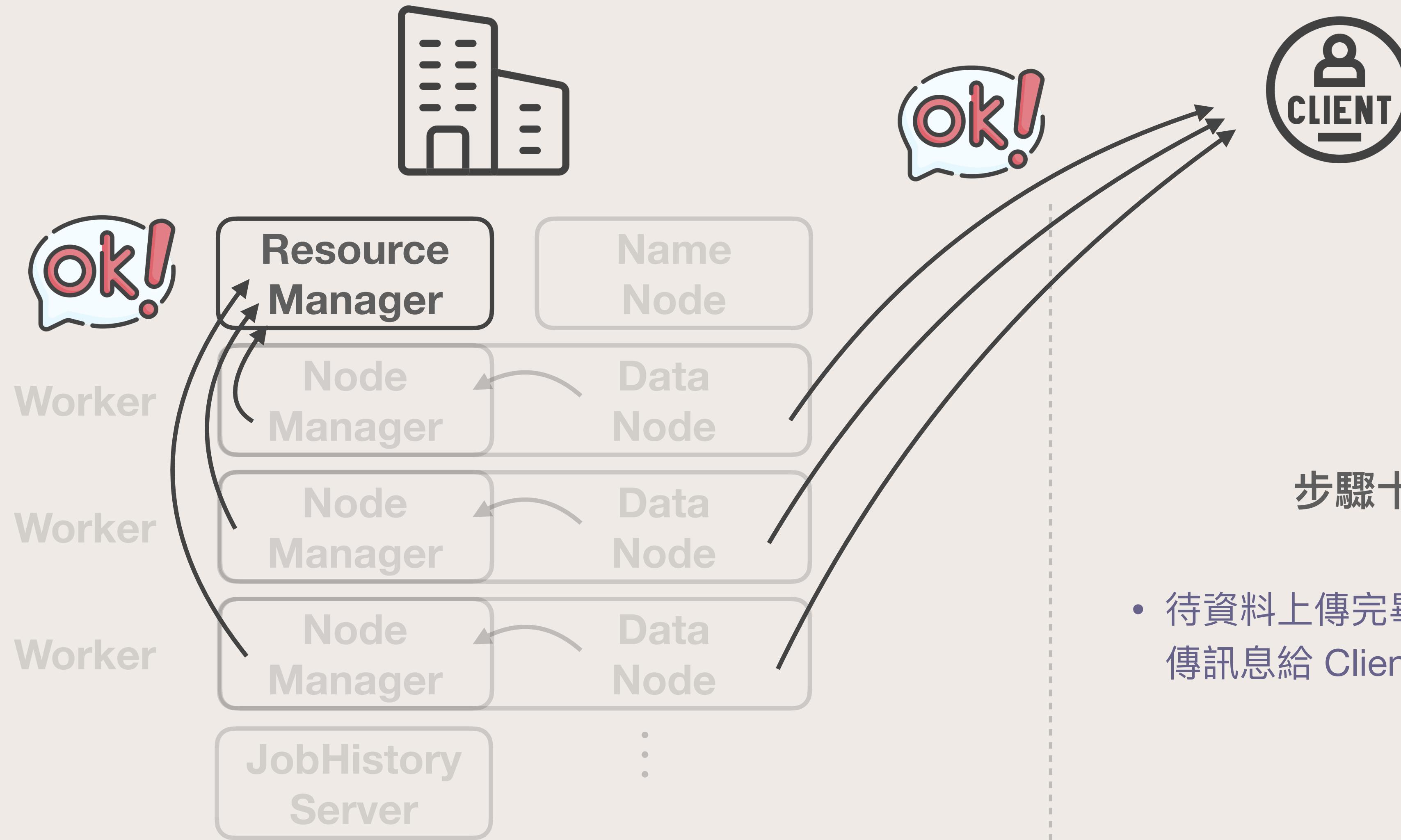
- 待資料上傳完畢，Data Node 會回傳訊息給 Client 端和 NameNode



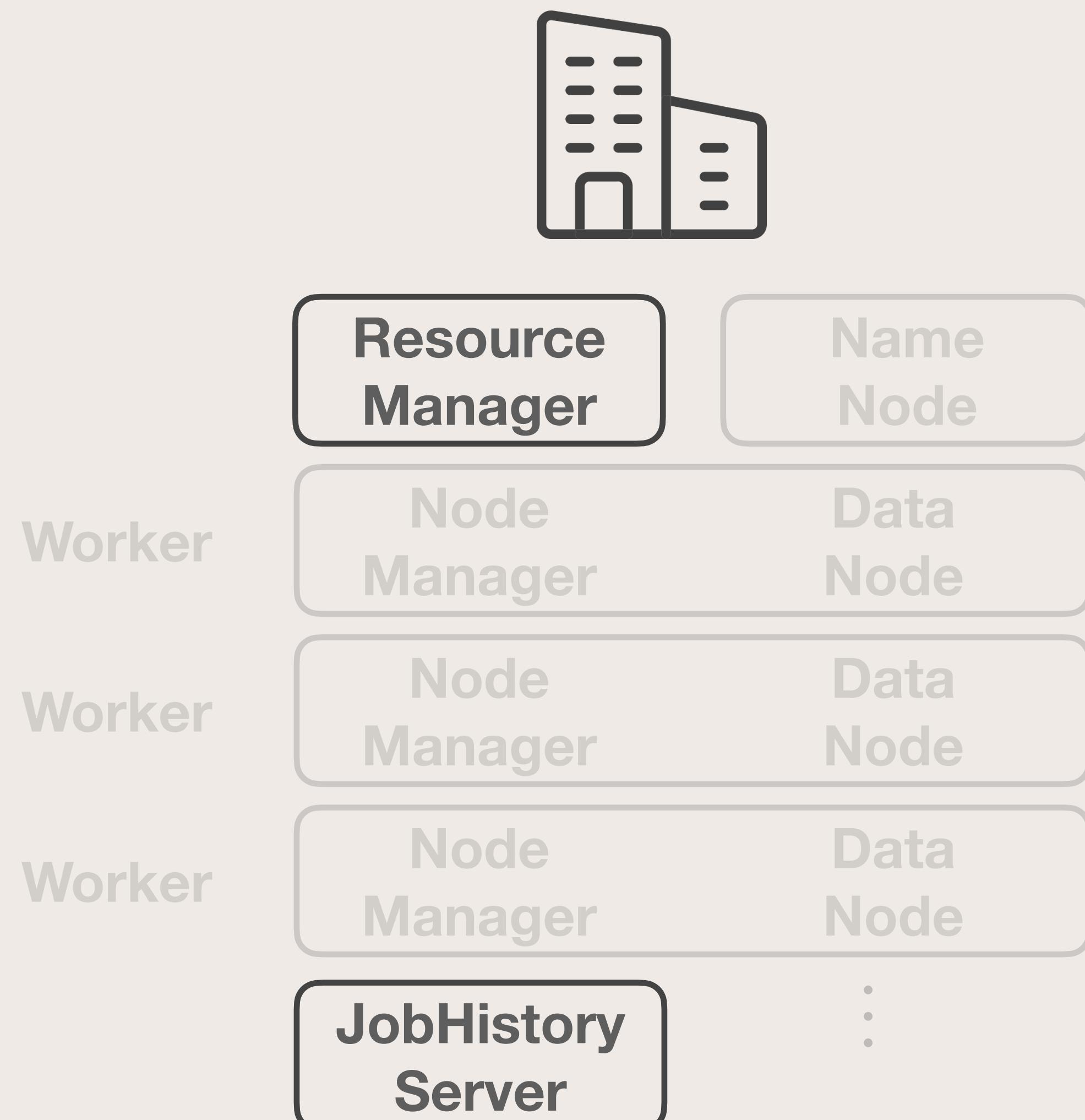
系統運作



系統運作

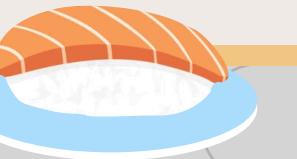


系統運作

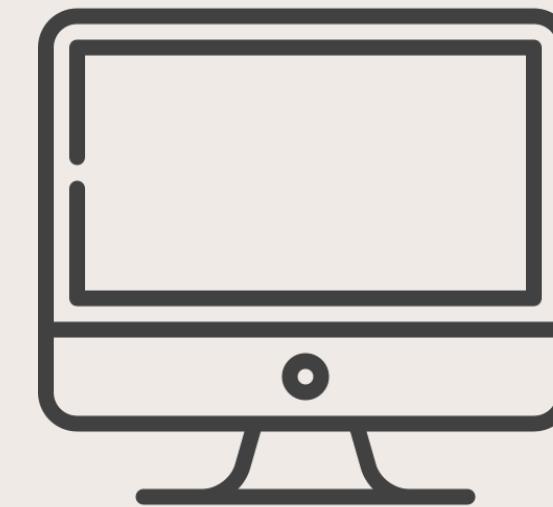
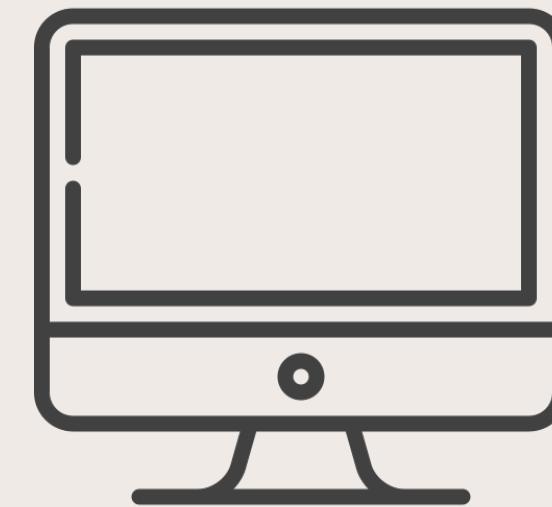
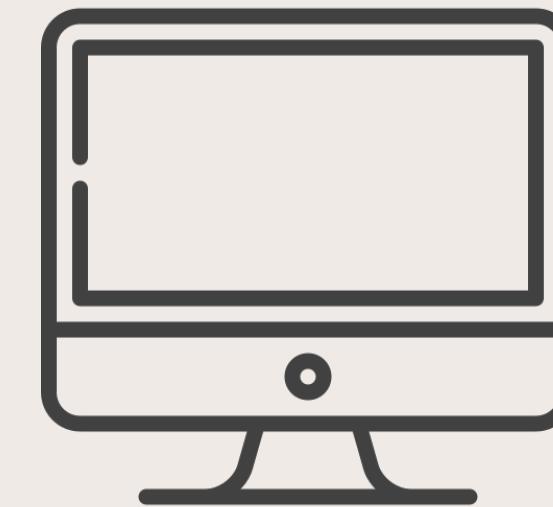
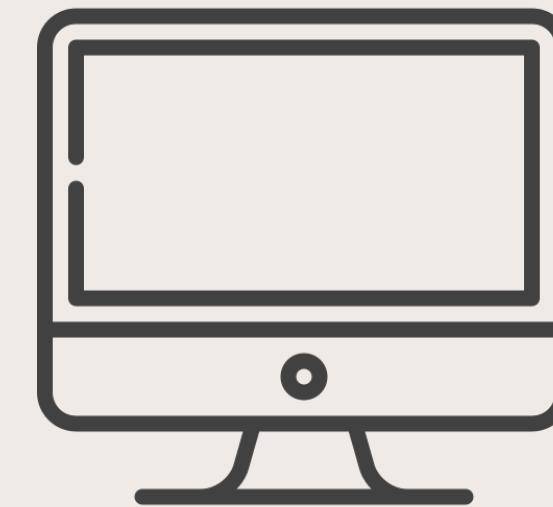


步驟一～十一

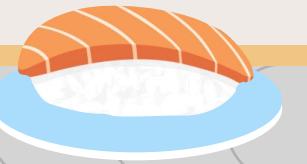
- ResourceManager 會負責虛擬機資源的調配
 - Job History Server 會紀錄 Datanode 運行的作業內容



系統建置



5 台實體主機

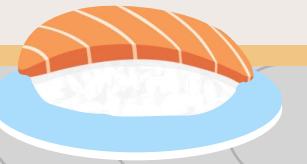


系統建置

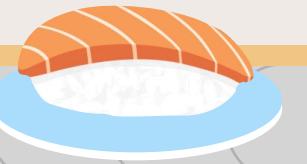
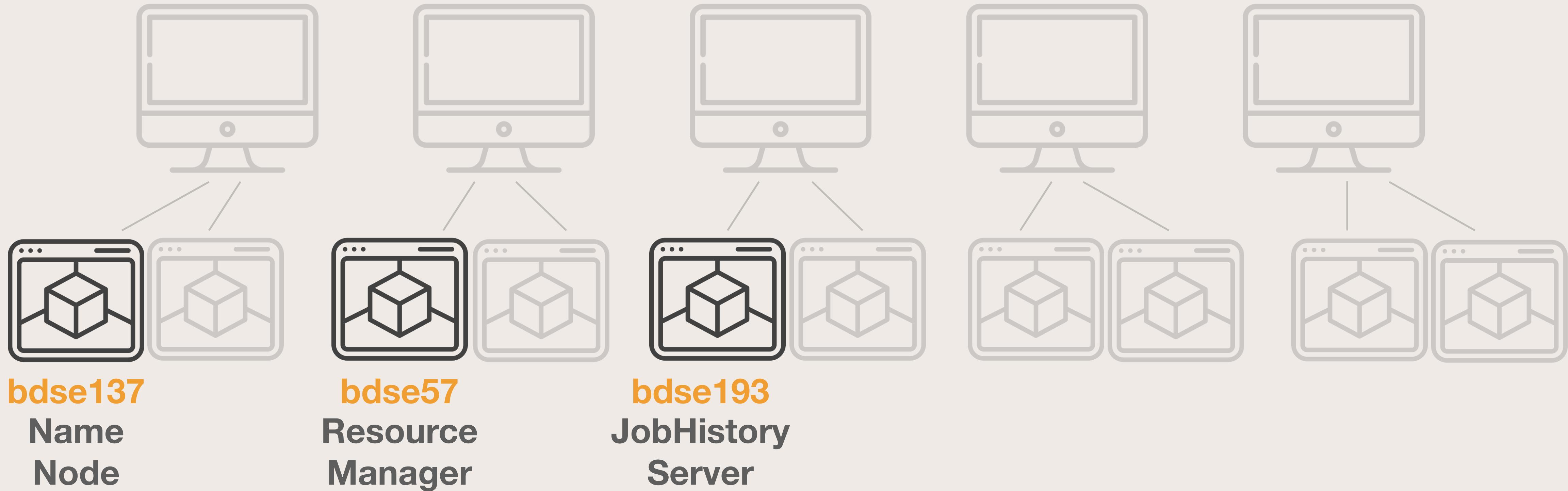


10 台虛擬主機

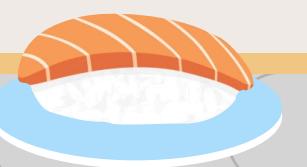
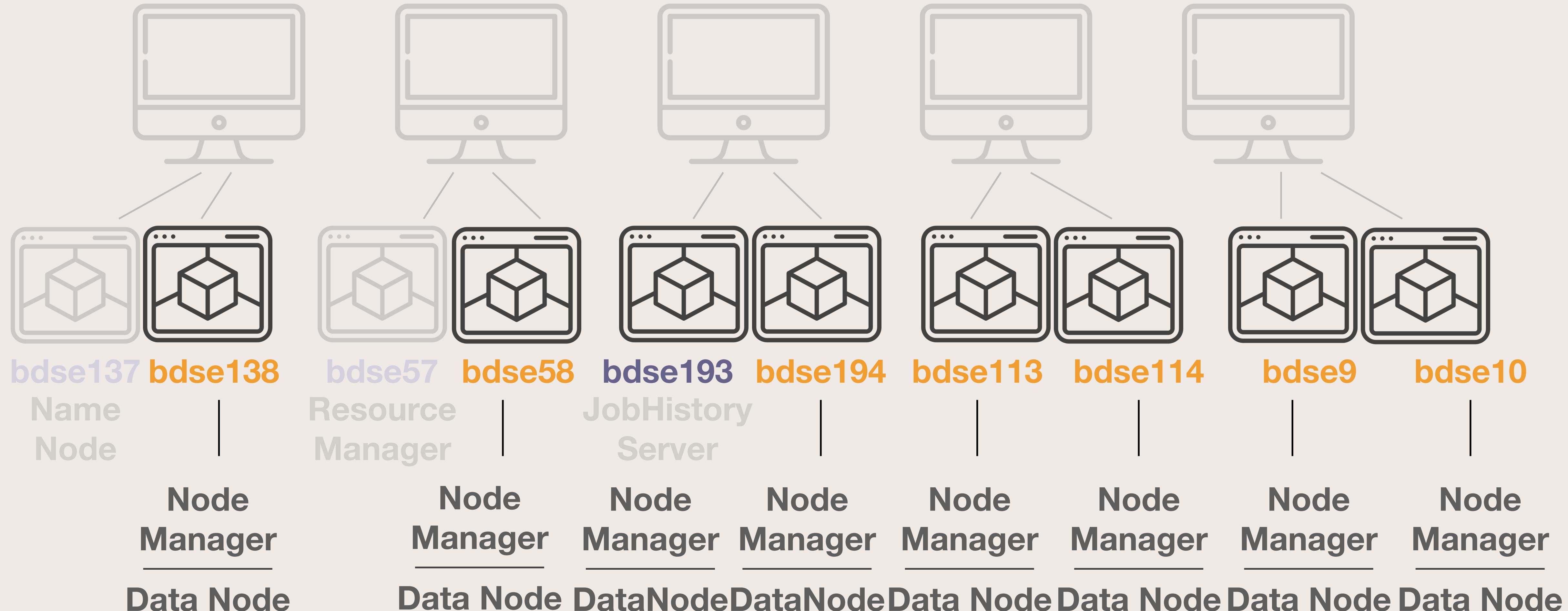
{ 記憶體：56 GB
CPU：16 核心



系統建置



系統建置



資料預處理





資料合併

Code 檔案 編輯 選取項目 檢視 移至 執行 終端機 視窗 說明 9月21日 週二 19:52:06 menutaiwan_all_110.json

```
限制模式適用於安全程式碼瀏覽。信任此視窗以啟用所有功能。 管理 深入了解
{} menutaiwan_all_110.json x
Users > wangyuting > Coding > projectDelicacy > {} menutaiwan_all_110.json > ...
{"allInfo": {"props": {"initialI18nStore": {"tw": {"restaurant": {}, "common": {"banner_title_1": "\u2019", "banner_title_2": "\u201c你錯過了什麼？\u201d", "banner_subtitle": "\u201d年度排行榜大回顧，全台最紅美食隆重加！\u201d", "search_btn": "\u201c搜尋\u201d", "search_placeHolder": "\u201d輸入\u201e早午餐\u201d或\u201e甜點\u201d找尋美食！\u201d", "more_btn": "\u201c看更多\u201d", "section_hot": "\u201d熱門餐廳\u201d", "hottest_posts": "\u201d熱門分享文\u201d", "section_collection": "\u201dMENU精選特輯\u201d", "section_newOpen": "\u201d美食客探索新餐廳\u201d", "update_time": "\u201d更新時間\u201d", "at": "\u201d在\u201d", "Cancel": "\u201d取消\u201d", "Copy": "\u201d複製嵌入框\u201d", "search_empty_1": "\u201d這題太難了！可以跳過嗎？\u201d", "search_empty_2": "\u201d麻煩請換下一個關鍵字\u201d"}, "initialLanguage": "\u201ctw\u201d", "i18nServerInstance": null, "pageProps": {"namespacesRequired": ["restaurant"], "common": {"cityId": "\u201d5d63c42937ca71b646626be5\u201d", "city": "\u201d台北市\u201d", "category": "\u201d \u201dlang": "\u201ctw\u201d", "relativeRecommend": [{"updatedAt": "\u201d2021-08-10T11:34:18.073Z\u201d, "createdAt": "\u201d2016-07-23T21:40:41.979Z\u201d, "name": "\u201d匠樂割烹壽司\u201d", "hasServiceFee": true, "areaLevelThree": "\u201d大安區\u201d", "areaLevelOne": "\u201d台北市\u201d, "webPhoto": {"userId": "\u201dnJmfGX3HTV\u201d", "userPictureUrl": "\u201dhttps://d1v47iv392tpm6.cloudfront.net/04b6a7a0-9e85-44cd-84c5-19e390070064.file.bin\u201d, "photoId": "\u201dW8J1wFarIT7\u201d, "createTime": "\u201dDate\u201d, "iso": "\u201d2020-05-03T07:43:46.389Z\u201d, "photoDescription": "\u201d好看又好吃\u201d", "mealId": "\u201dPogLRAJQCe\u201d, "photoUrl": "\u201dhttps://d1v47iv392tpm6.cloudfront.net/42aed2fc-b765-48b3-91b8-ce6b9373187d_519c1bb0-56cc-413f-bf89-2b55d541109f_i_image_resized.jpg\u201d, "secondPhotoUrl": "\u201dhttps://d1v47iv392tpm6.cloudfront.net/3b6c1a05-3b8c-44da-a930-13a45d7b047f_mfp_af7465df87031982c710d140ce33e90f_i_image_resized.jpg\u201d, "thirdPhotoUrl": "\u201dhttps://d1v47iv392tpm6.cloudfront.net/b0e03fce-6682-b456-ea062c48f038_3a5e8699-c9cd-4295-99f3-9847e5ff1496_i_image_resized.jpg\u201d}, "lowerRange": "1001", "upperRange": "2000", "averageRating": 4, "objectId": "\u201d04xK01CuquH\u201d, "type": "Object", "className": "Restaurant", "updatedAt": "\u201d2021-08-20T04:42:24.070Z\u201d, "createdAt": "\u201d2015-07-31T19:01:57.729Z\u201d, "name": "\u201d下酒菜 串燒 居酒屋\u201d, "areaLevelOne": "\u201d台北市\u201d, "areaLevelThree": "\u201d中正區\u201d, "lowerRange": "1505", "upperRange": "1505", "hasServiceFee": false, "periods": [{"close": {"day": 1, "time": "\u201d0100\u201d}, {"open": {"day": 0, "time": "\u201d1800\u201d}, {"close": {"day": 2, "time": "\u201d1800\u201d}, {"open": {"day": 1, "time": "\u201d0100\u201d}, {"close": {"day": 3, "time": "\u201d1800\u201d}, {"open": {"day": 2, "time": "\u201d0100\u201d}, {"close": {"day": 4, "time": "\u201d0100\u201d}, {"open": {"day": 3, "time": "\u201d1800\u201d}, {"close": {"day": 5, "time": "\u201d0100\u201d}, {"open": {"day": 4, "time": "\u201d1800\u201d}, {"close": {"day": 6, "time": "\u201d0100\u201d}, {"open": {"day": 5, "time": "\u201d1800\u201d}, {"close": {"day": 0, "time": "\u201d0100\u201d}, {"open": {"day": 1, "time": "\u201d1800\u201d}, {"close": {"day": 2, "time": "\u201d1800\u201d}, {"open": {"day": 3, "time": "\u201d0100\u201d}, {"close": {"day": 4, "time": "\u201d1800\u201d}, {"open": {"day": 5, "time": "\u201d0100\u201d}, {"close": {"day": 6, "time": "\u201d1800\u201d}, {"open": {"day": 0, "time": "\u201d1800\u201d}, {"averageRating": 4, "webPhoto": {"userId": "\u201dR7rvmTh0E5\u201d, "userPictureUrl": "\u201dhttps://d1v47iv392tpm6.cloudfront.net/efe79558-c635-4368-ad05-8f6792fa157d_file.bin\u201d, "photoId": "\u201dD0Bh1b3dY7\u201d, "createTime": "\u201dDate\u201d, "iso": "\u201d2020-12-05T13:24:10.324Z\u201d, "photoDescription": "\u201d非常的棒舒適\u201d", "mealId": "\u201dEz1LdaVLh4\u201d, "photoUrl": "\u201dhttps://d1v47iv392tpm6.cloudfront.net/7d72f71a-5282-4b2b-9e57-c2e8d23be66d_9725d62f-9851-4991-8e8f-eee336f2ac60_i_image_resized.jpg\u201d, "secondPhotoUrl": "\u201dhttps://d1v47iv392tpm6.cloudfront.net/8141a9dc-0ccf-4f91-9230-e751908b0aa4_13da4d1858d1fa3a769cb7aff800a3de_a_image_resized.jpg\u201d, "thirdPhotoUrl": "\u201dhttps://d1v47iv392tpm6.cloudfront.net/a16fd5fd-14be-42a9-9c47-8590667b7bc0_85dc871c-7b49-4501-9ee0-0081d2a29c8f_i_image_resized.jpg\u201d, "objectId": "\u201d0KRbfPALY6\u201d, "type": "Object", "className": "Restaurant", "name": "\u201d湖光市場現做肉圓\u201d, "branch": "\u201d \u201d, "updatedAt": "\u201d2021-08-18T00:14:57.574Z\u201d, "createdAt": "\u201d2015-06-05T12:30:37.418Z\u201d, "areaLevelOne": "\u201d台北市\u201d, "areaLevelThree": "\u201d內湖區\u201d, "hasServiceFee": false, "lowerRange": 0, "upperRange": 0, "averageRating": 4, "webPhoto": {"userId": "\u201dwMjwwmXR0C\u201d, "userPictureUrl": "\u201dhttps://d1v47iv392tpm6.cloudfront.net/5ea0107f-f5b1-4f38-80d7-149b8d1f624d_file.bin\u201d, "photoId": "\u201dC7reREBA5\u201d, "createTime": "\u201dDate\u201d, "iso": "\u201d2019-02-04T06:13:56.540Z\u201d, "photoDescription": "\u201d內湖人都知道的好味道www\u201d", "mealId": "\u201dikWyozm0HE\u201d, "photoUrl": "\u201dhttps://d1v47iv392tpm6.cloudfront.net/5c9affcc-56c0-4c82-aa8c-b18457d401c3_1df15ab4-e034-4c60-abc2-f80723eb2294_i_image_resized.jpg\u201d, "secondPhotoUrl": "\u201dhttps://d1v47iv392tpm6.cloudfront.net/60fd7fd6-b561-4f3c-8837-0015a8facd2_1369c8d-377c-4085-a7691d94776e9_i_image_resized.jpg\u201d, "thirdPhotoUrl": "\u201dhttps://d1v47iv392tpm6.cloudfront.net/0b419a8c-dbf7-4b96-9d3d-fe6c402aecfc_e90f31cd-a4c6-46fa-87a6-919cf1764612_i_image_resized.jpg\u201d}, "objectId": "\u201d0Pu8w9xpC\u201d, "type": "Object", "className": "Restaurant"}], "review": "非常的棒 舒適", "resource": "menutaiwan", "res_name": "\u201d下酒菜 串燒 居酒屋\u201d, "address": "\u201d100台灣台北市中正區金山南路一段81-1號\u201d, "phone": "02 3322 5557", "rank": "NaN"}]
```

受限模式 ① 0 △ 0 第1行, 第1欄 空格: 4 UTF-8 LF JSON ⌂

Code 檔案 編輯 檢視 歷史記錄 書籤 設定檔 分頁 視窗 說明 9月21日 週二 19:48:01 Desktop/clean3/ combine_ppt - Jupyter Notebook dropna - Jupyter Notebook * 流程紀錄.txt - Jupyter Text Edi +

Jupyter combine_ppt Last Checkpoint: 9 分鐘前 (unsaved changes)

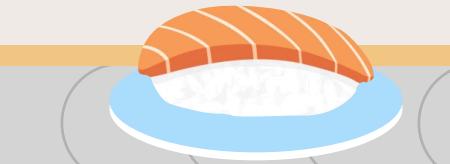
	res_name	address	phone	rank	review	resource
0	下酒菜 串燒 居酒屋	100台灣台北市中正區金山南路一段81-1號	02 3322 5557	NaN	非常的棒 舒適	menutaiwan
1	下酒菜 串燒 居酒屋	100台灣台北市中正區金山南路一段81-1號	02 3322 5557	NaN	居酒屋口袋名單 這個牛只能給坐坐 必須成為新歡！牛肉炒飯超級好吃 明太子山藥也是愛再來...	menutaiwan
2	下酒菜 串燒 居酒屋	100台灣台北市中正區金山南路一段81-1號	02 3322 5557	NaN	下酒菜的串燒最推的是這個 明太子雞肉串 明太子跟美奶滋的組合 波波的在嘴裡融化 雞肉也烤的...	menutaiwan
3	下酒菜 串燒 居酒屋	100台灣台北市中正區金山南路一段81-1號	02 3322 5557	NaN	去幾次都不會膩的下酒菜 牛肉串烤的很嫩 鹹香口味搭配啤酒一級棒	menutaiwan
4	湖光市場現做肉圓	台北市內湖區內湖區成功路四段26號 (湖光市場門口)	02 3322 5557	NaN	內湖人都知道的好味道♥♥♥	menutaiwan
...
25074	老晴光張媽切仔麵		NaN	NaN	# 切仔米粉 這家從小吃到大的切仔麵 這個滋味真的是令人難忘！已經有50年歷史的古早味♥ ...	menutaiwan
25075	罵子蛋炸物		NaN	NaN	剛炸起來罵子蛋的酥脆感超好吃的 加上泰式酸酸甜甜的 想一口接著一口♥♥	menutaiwan
25076	芳琳魯肉飯		NaN	NaN	在西湖站看到一個很吸引我的招牌 就這樣走進去了 結果發現 根本就是鬍鬚張呀！😍😍😍 不誇張...	menutaiwan
25077	三良品茶-永康店		NaN	NaN	# 新開店 每口都喝到濃郁咖啡香跟巧克力碎片，讓我再悠閒的下午茶時光，忘記減肥的痛苦😊 #台北...	menutaiwan
25078	SC粉紅雲朵		NaN	NaN	拿在手上份量很重，吃起來就像熟悉的小美淇淋。獨角獸的角是一根長長的糖霜麻花捲。兩片耳朵是軟...	menutaiwan

25079 rows x 6 columns

merge_menutaiwan.csv



- 合併統整自各網站爬取下的餐廳資訊、評分、評論
- 統一檔案格式為 csv 檔與欄位名稱：餐廳名稱、餐廳地址、餐廳電話、評論分數、評論內容





資料整理

Chrome 檔案 編輯 檢視 歷史記錄 書籤 設定檔 分頁 視窗 說明 9月21日 週二 19:45:55

localhost:8888/notebooks/Desktop/clean3/combine_ppt.ipynb

jupyter combine_ppt Last Checkpoint: 7分鐘前 (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Out[3]:

	res_name	address	phone	rank	review	resource
0	下酒菜 串燒 居酒屋	100台灣台北市中正區金山南路一段81-1號	02 3322 5557	Nan	非常的棒 舒適	menutaiwan
1	下酒菜 串燒 居酒屋	100台灣台北市中正區金山南路一段81-1號	02 3322 5557	Nan	居酒屋口袋名單 這個牛只能給四星半 必須成為新歡！牛肉炒飯超級好吃 明太子山藥也是愛再來...	menutaiwan
2	下酒菜 串燒 居酒屋	100台灣台北市中正區金山南路一段81-1號	02 3322 5557	Nan	下酒菜的串燒最推的是這個 明太子雞肉串 明太子跟美奶滋的組合 波波波的在嘴裡融化 雞肉也烤的...	menutaiwan
3	下酒菜 串燒 居酒屋	100台灣台北市中正區金山南路一段81-1號	02 3322 5557	Nan	去幾次都不會膩的下酒菜 牛肉串烤的很嫩 鹹香口味搭配啤酒一級棒	menutaiwan
4	湖光市場現做肉圓	台北市內湖區內湖區成功路四段26號 (湖光市場門口)	02 3322 5557	Nan	內湖人都知道的好味道❤️❤️❤️	menutaiwan
...
25074	老晴光張媽媽切仔麵		Nan	Nan	# 切仔米粉 這家從小吃到大的切仔麵 這個滋味真的是令人難忘！已經有50年歷史的古早味...	menutaiwan
25075	鴉子蛋炸物		Nan	Nan	剛炸起來鴉子蛋的酥脆感超好吃的 加上泰式酸酸甜甜的 想一口接著一口❤️❤️ 起司雞排搭配...	menutaiwan
25076	芳琳魯肉飯		Nan	Nan	在西湖站 看到一個很吸引我的招牌 就這樣走進去了 結果發現 根本就是爛頭張 呀！😂😂 不誇張...	menutaiwan
25077	三良品茶-永康店		Nan	Nan	#新開店 每口都喝到濃郁咖啡香跟巧克力碎片，讓我再悠閒的下午茶時光，忘記減肥的痛苦😊 #台北...	menutaiwan
25078	SC粉紅雲朵		Nan	Nan	拿在手上份量很重，吃起來就像熟悉的小美冰淇淋。獨角獸的角是一根長長的糖霜麻花捲。兩片耳朵是軟...	menutaiwan

25079 rows × 6 columns

addresource_menutaiwan.csv



- 利用Python 的 Pandas 彙整資料
- 在每個餐廳評論檔案加上欄位「資料來源」
- 將 5 間餐廳的評論檔案合併成 1 份資料

Chrome 檔案 編輯 檢視 歷史記錄 書籤 設定檔 分頁 視窗 說明 9月21日 週二 19:43:04

localhost:8888/notebooks/Desktop/clean3/combine_ppt.ipynb

jupyter combine_ppt Last Checkpoint: 5分鐘前 (autosaved)

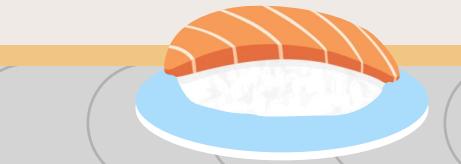
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Out[2]:

	res_name	address	phone	rank	review	resource
0	長園廳-長榮桂冠酒店(台北)	台北市中山區松江路63號	02 2518 8380	5.0	食物很好吃，調味都很好，服務也很棒，疫情時間客人比較，上菜速度也很快！結帳時剛好有使用信用卡...	eztable
1	長園廳-長榮桂冠酒店(台北)	台北市中山區松江路63號	02 2518 8380	4.0	芋泥鴨失望了，芋泥不香鴨肉材，好像放了三天呢！烏參太軟混了些牛筋 應該不會再點。蘿蔴糕竟然也不...	eztable
2	長園廳-長榮桂冠酒店(台北)	台北市中山區松江路63號	02 2518 8380	4.0	這次去長園廳點的是龍蝦套餐，從前菜到甜點每道味道都很不錯，唯一美中不足的是當天不知是不是服務...	eztable
3	長園廳-長榮桂冠酒店(台北)	台北市中山區松江路63號	02 2518 8380	4.0	這次透過EZTABLE的平台，預約到了長榮桂冠酒店，來進行一個非凡的父親節活動，不得不說他們...	eztable
4	長園廳-長榮桂冠酒店(台北)	台北市中山區松江路63號	02 2518 8380	5.0	這次剛好碰到朋友推薦這家餐廳、說氣氛佳食物美味，本來還覺得這家酒店地方不大，服務可能打點折扣...	eztable
...
679363	龙门客棧饺子馆	100台灣 中正區台北 中正區林森南路61巷19號	tel:+86 2 2351 0729	4.0	為什麼不？可能真的應該被命名為Dragon Gate Inn Dumpling House。這...	tripadvisor
679364	龜山島阿興現撈海產店(大同)	台灣 大同台北 大同區台北市承德路二段78號		Nan	新鮮美味的海產餐廳原為寧夏夜市老字號的店家，現搬至承德路與民生西路口，從下午4點開到凌晨3點...	tripadvisor
679365	龜山島阿興現撈海產店(大同)	台灣 大同台北 大同區台北市承德路二段78號		Nan	新鮮好吃上菜速度滿快，價格也還ok，食材新鮮...個人覺得要點白飯，可能我吃的清淡，覺得東西都有...	tripadvisor
679366	龜山島阿興現撈海產店(大同)	台灣 大同台北 大同區台北市承德路二段78號		Nan	上菜速度快，美味新鮮位於承德路上的龜山島阿興現撈海產店，每到夜晚總是聚集享受快炒美味的老饕們...	tripadvisor
679367	龜山島阿興現撈海產店(大同)	台灣 大同台北 大同區台北市承德路二段78號		Nan	好吃價格還算合理 選擇也很多樣 食材也都用得很新鮮 用餐環境還算可以不會太不舒服 有機會可...	tripadvisor

679368 rows × 6 columns

combine.csv



資料清洗

phone

(02)6631-6666

tel:0266316666

+8862-6631-6666

市話	手機
02 6631 6666	0911 296 699
2碼 4碼 3 or 4 碼	4碼 3碼 3碼

address

台北市大安區忠孝東路四段216巷19弄11號大安區

台灣 大同台北 大同區台北市承德路二段78號

100 台灣 中正區台北 中正區林森南路61巷19號

縣市	區域	路名	門牌號碼	樓層
台北市大同區承德路二段78號				
台北市大安區忠孝東路四段166號2樓				



- 進行資料清洗，整理欄位 phone 和 address 的格式





資料清洗

clean.csv

	res_name	address	phone	rank	review	district	resource
0	長園廳-長榮桂冠酒店 (台北)	台北市中山區松江路63號	02 2518 8380	5.0	食物很好吃，調味都很好，服務也很棒，疫情時間客人比較，上菜速度也很快！結帳時剛好有使用信用卡...	中山區	eztable
1	長園廳-長榮桂冠酒店 (台北)	台北市中山區松江路63號	02 2518 8380	4.0	芋泥鴨失望了，芋泥不香鴨肉材，好像放了三天呢！烏參太軟混了些牛筋應該不會再點。蘿蔔糕竟然也不...	中山區	eztable
2	長園廳-長榮桂冠酒店 (台北)	台北市中山區松江路63號	02 2518 8380	4.0	這次去長園廳點的是龍蝦套餐，從前菜到甜點每道味道都很不錯，唯一美中不足的是當天不知是不是服務...	中山區	eztable
3	長園廳-長榮桂冠酒店 (台北)	台北市中山區松江路63號	02 2518 8380	4.0	這次透過EZTABLE的平台，預約到了長榮桂冠酒店，來進行一個非凡的父親節活動，不得不說他們...	中山區	eztable
4	長園廳-長榮桂冠酒店 (台北)	台北市中山區松江路63號	02 2518 8380	5.0	這次剛好碰到朋友推薦這家餐廳、說氣氛佳食物美味，本來還覺得這家酒店地方不大，服務可能打點折扣...	中山區	eztable
...
679363	龙门客棧饺子馆	台北市中正區林森南路61巷19號	NaN	4.0	為什麼不？可能真的應該被命名為DragonGateInnDumpingHouse。這個地方曾...	中正區	tripadvisor
679364	龜山島阿興現撈海產店(大同)	台北市大同區承德路二段78號	NaN	4.0	新鮮美味的海產餐廳原為寧夏夜市老字號的店家，現搬至承德路與民生西路口，從下午4點開到凌晨3點...	大同區	tripadvisor
679365	龜山島阿興現撈海產店(大同)	台北市大同區承德路二段78號	NaN	4.0	新鮮好吃上菜速度滿快，價格也還ok，食材新鮮...個人覺得要點白飯，可能我吃的清淡，覺得東西都有...	大同區	tripadvisor
679366	龜山島阿興現撈海產店(大同)	台北市大同區承德路二段78號	NaN	4.0	上菜速度快，美味新鮮位於承德路上的龜山島阿興現撈海產店，每到夜晚總是聚集享受快炒美味的老饕們...	大同區	tripadvisor
679367	龜山島阿興現撈海產店(大同)	台北市大同區承德路二段78號	NaN	5.0	好吃價格還算合理選擇也很多樣食材也都用得很新鮮用餐環境還算可以不會太不舒服有機會可以去吃吃看...	大同區	tripadvisor

679368 rows × 7 columns

clean.csv

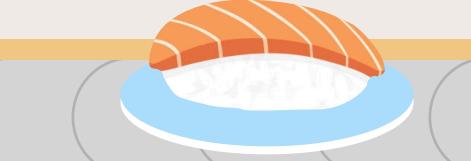
dropna.csv

	res_name	address	phone	rank	review	district	resource
0	長園廳-長榮桂冠酒店 (台北)	台北市中山區松江路63號	02 2518 8380	5.0	食物很好吃，調味都很好，服務也很棒，疫情時間客人比較，上菜速度也很快！結帳時剛好有使用信用卡...	中山區	eztable
1	長園廳-長榮桂冠酒店 (台北)	台北市中山區松江路63號	02 2518 8380	4.0	芋泥鴨失望了，芋泥不香鴨肉材，好像放了三天呢！烏參太軟混了些牛筋應該不會再點。蘿蔔糕竟然也不...	中山區	eztable
2	長園廳-長榮桂冠酒店 (台北)	台北市中山區松江路63號	02 2518 8380	4.0	這次去長園廳點的是龍蝦套餐，從前菜到甜點每道味道都很不錯，唯一美中不足的是當天不知是不是服務...	中山區	eztable
3	長園廳-長榮桂冠酒店 (台北)	台北市中山區松江路63號	02 2518 8380	4.0	這次透過EZTABLE的平台，預約到了長榮桂冠酒店，來進行一個非凡的父親節活動，不得不說他們...	中山區	eztable
4	長園廳-長榮桂冠酒店 (台北)	台北市中山區松江路63號	02 2518 8380	5.0	這次剛好碰到朋友推薦這家餐廳、說氣氛佳食物美味，本來還覺得這家酒店地方不大，服務可能打點折扣...	中山區	eztable
...
679363	龙门客棧饺子馆	台北市中正區林森南路61巷19號	NaN	4.0	為什麼不？可能真的應該被命名為DragonGateInnDumpingHouse。這個地方曾...	中正區	tripadvisor
679364	龜山島阿興現撈海產店(大同)	台北市大同區承德路二段78號	NaN	4.0	新鮮美味的海產餐廳原為寧夏夜市老字號的店家，現搬至承德路與民生西路口，從下午4點開到凌晨3點...	大同區	tripadvisor
679365	龜山島阿興現撈海產店(大同)	台北市大同區承德路二段78號	NaN	4.0	新鮮好吃上菜速度滿快，價格也還ok，食材新鮮...個人覺得要點白飯，可能我吃的清淡，覺得東西都有...	大同區	tripadvisor
679366	龜山島阿興現撈海產店(大同)	台北市大同區承德路二段78號	NaN	4.0	上菜速度快，美味新鮮位於承德路上的龜山島阿興現撈海產店，每到夜晚總是聚集享受快炒美味的老饕們...	大同區	tripadvisor
679367	龜山島阿興現撈海產店(大同)	台北市大同區承德路二段78號	NaN	5.0	好吃價格還算合理選擇也很多樣食材也都用得很新鮮用餐環境還算可以不會太不舒服有機會可以去吃吃看...	大同區	tripadvisor

541346 rows × 7 columns



- 將沒有評論的欄位刪除



關鍵字提取

Python 套件
Jieba (結巴)

以 TF-IDF (關鍵詞提取算法)
提取餐廳評論中的關鍵詞



將單一餐廳所有評論
讀入套件內

關鍵字提取

Python 套件
Jieba (結巴)

以 TF-IDF (關鍵詞提取算法)
提取餐廳評論中的關鍵詞



將單一餐廳所有評論
讀入套件內

設置「停止詞」
忽略贅字詞彙 (如：
啊、標點符號等贅字)

關鍵字提取

Python 套件
Jieba (結巴)

以 TF-IDF (關鍵詞提取算法)
提取餐廳評論中的關鍵詞



將單一餐廳所有評論
讀入套件內

設置「停止詞」
忽略贅字詞彙 (如：
啊、標點符號等贅字)

$$\text{詞頻(TF)} = \frac{\text{某個詞在文章中出現次數}}{\text{文章總詞數}}$$
$$\text{逆向檔案頻率(IDF)} = \log\left(\frac{\text{語料庫文檔總數}}{\text{包含該詞的文檔數} + 1}\right)$$
$$\text{TF - IDF} = \text{詞頻(TF)} \times \text{逆向檔案頻率(IDF)}$$

使用 `jieba.analyse.extract_tags`
函式設定目標詞彙 (形容詞)
進行關鍵字提取

關鍵字提取

Python 套件
Jieba (結巴)

以 TF-IDF (關鍵詞提取算法)
提取餐廳評論中的關鍵詞



將單一餐廳所有評論
讀入套件內

設置「停止詞」
忽略贅字詞彙 (如：
啊、標點符號等贅字)

$$\text{詞頻(TF)} = \frac{\text{某個詞在文章中出現次數}}{\text{文章總詞數}}$$
$$\text{逆向檔案頻率(IDF)} = \log\left(\frac{\text{語料庫文檔總數}}{\text{包含該詞的文檔數} + 1}\right)$$
$$\text{TF - IDF} = \text{詞頻(TF)} \times \text{逆向檔案頻率(IDF)}$$

```
[('簡單', 11.9547675029),  
 ('貼心', 11.9547675029),  
 ('愉快', 7.51548325298),  
 ======  
 "重修舊好"  
 這間餐廳的關鍵詞: ['簡單', '貼心', '愉快']]
```

使用 `jieba.analyse.extract_tags`
函式設定目標詞彙 (形容詞)
進行關鍵字提取

依照 TF-IDF 值進行排序
列出比重前N名的關鍵字
進行視覺化

關鍵字提取

陶板屋



Jieba

→
經過分析

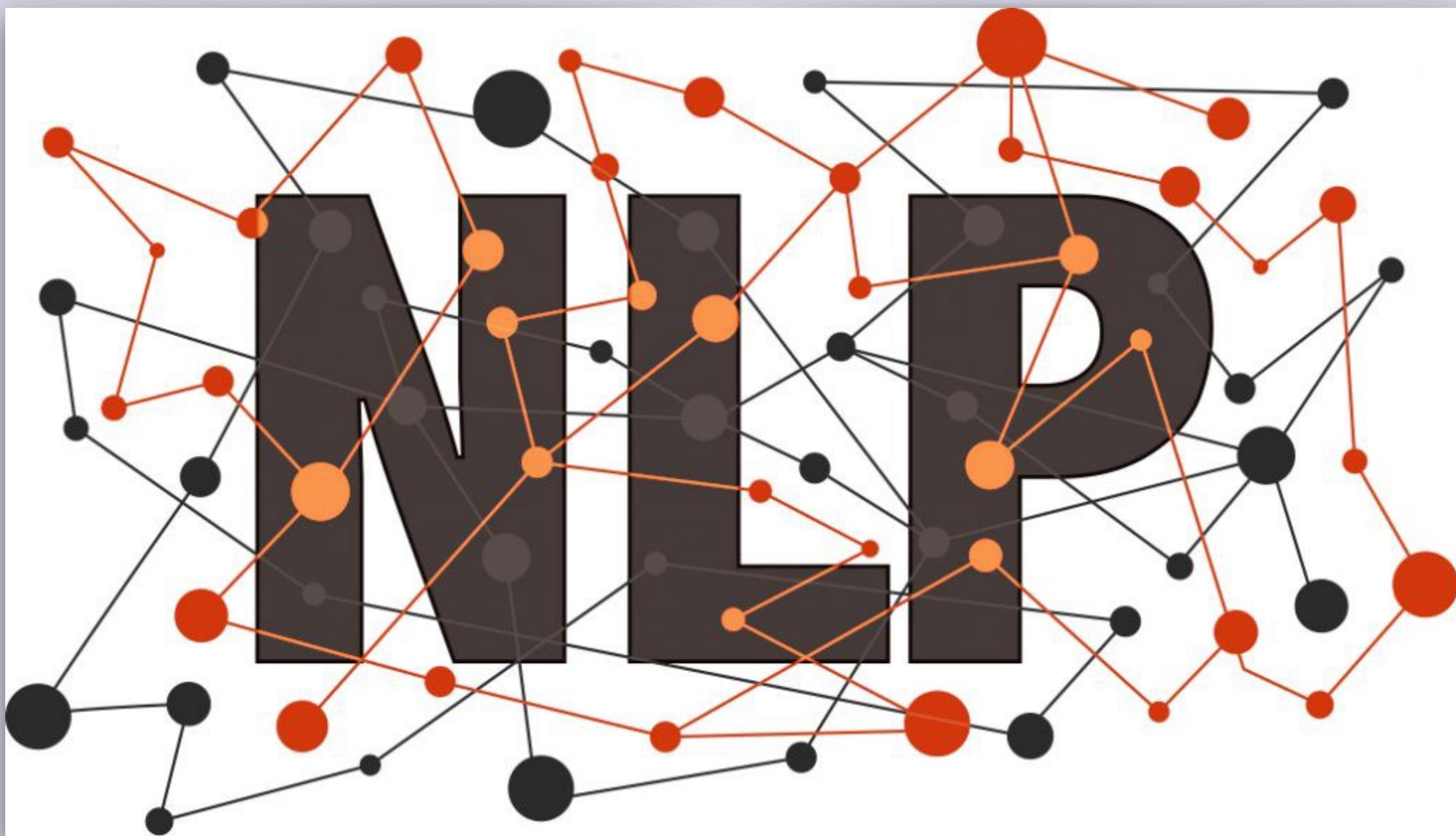


乾淨 出現頻率 25 %

美味 出現頻率 18 %

難吃 出現頻率 11 %

⋮



模型建置



機器學習

正向評論數 **70** 條

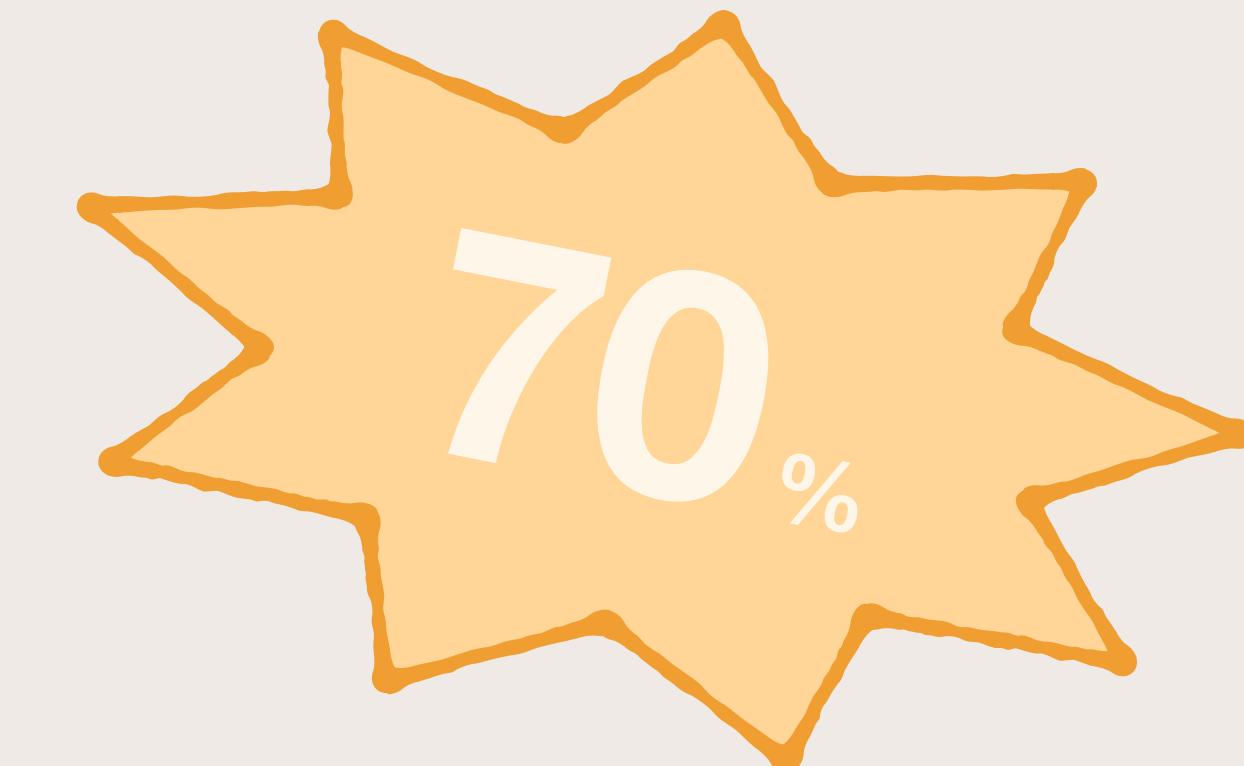


好評程度

$$= \frac{\text{正向評論數}}{\text{全部評論數}} = \frac{70}{100}$$



經過分析



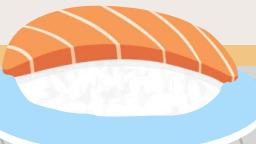
機器學習

450,000+

筆評論資料

NLP 自然語言處理

文本情感分析



機器學習

- 訓練資料集：

北京大學
餐飲評論數據集

資料特色

- 資料量：47萬筆
- 資料自帶標籤
- 均為餐飲相關評論

- 訓練資料預處理：

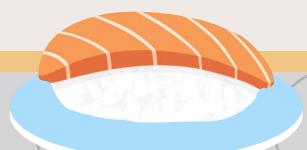
- 刪除停用詞
- 移除中空欄位、亂碼、表情符號等其他贅詞
- 篩選情緒詞數量：
正評論、負評論、中立評論各 **2.3萬** 筆

- 特徵處理

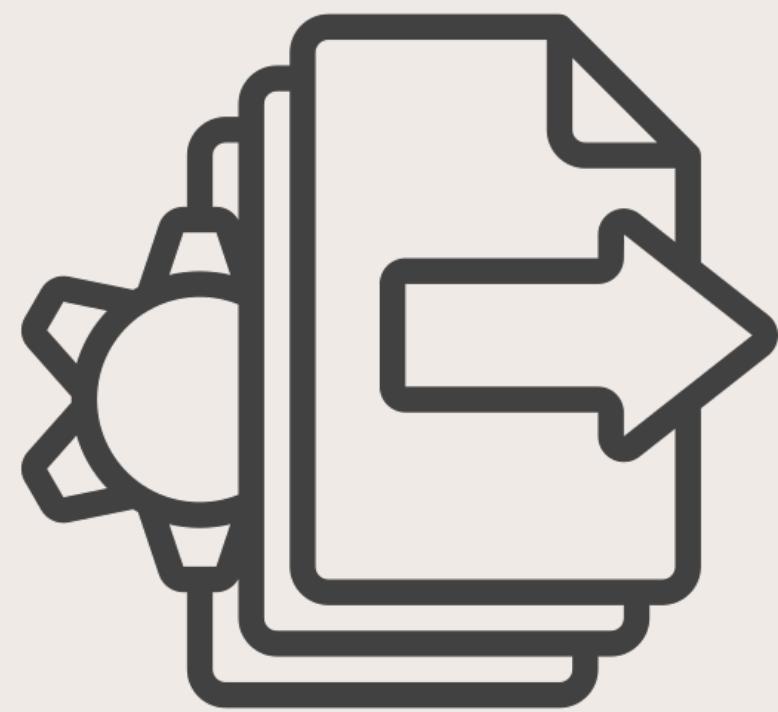
- 刪除無用處欄位(人名、時間等.....)
- 評分量化
- 劃分驗證集及測試集



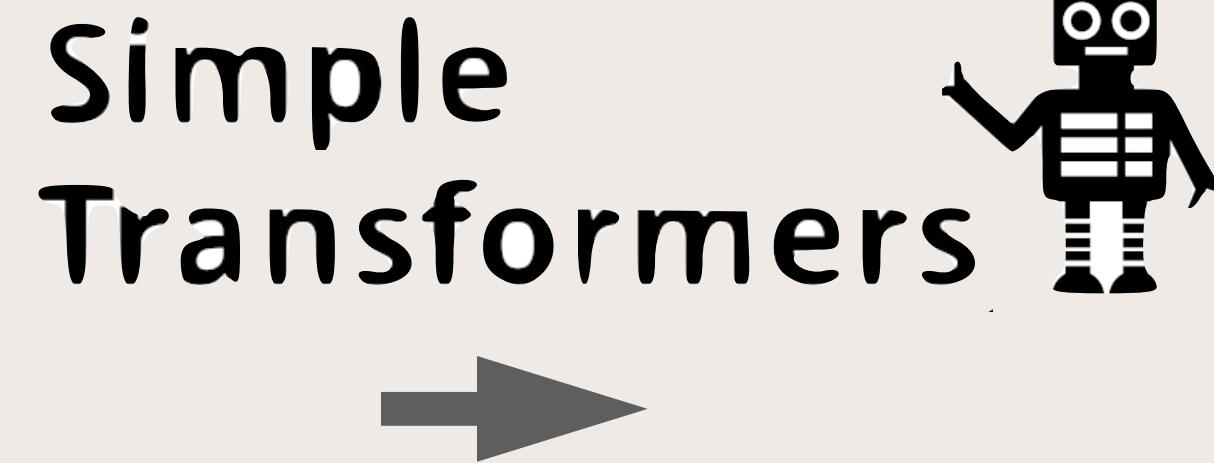
標籤化資料



機器學習



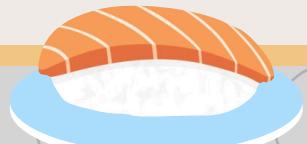
已標籤化資料



送入資料
載入預訓練模型



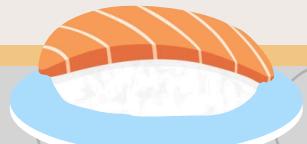
開始訓練



機器學習

	Neg	
True	兩邊標籤都預測正確 其中一項預測錯誤	其中一項預測錯誤
Pos	其中一項預測錯誤	兩邊標籤都預測正確
	Neg	Pos
	Predict	

- **混淆矩陣：**
 - 可簡易表示模型好壞的可視化圖表
 - X 軸：模型預測的標籤
Y 軸：真實的標籤
 - 可呈現訓練後模型的準確率



機器學習

- 中立標籤數 大於 正負面標籤數

- 測試資料集：

北京大學 餐廳評論數據集

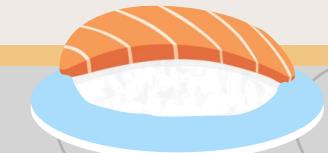
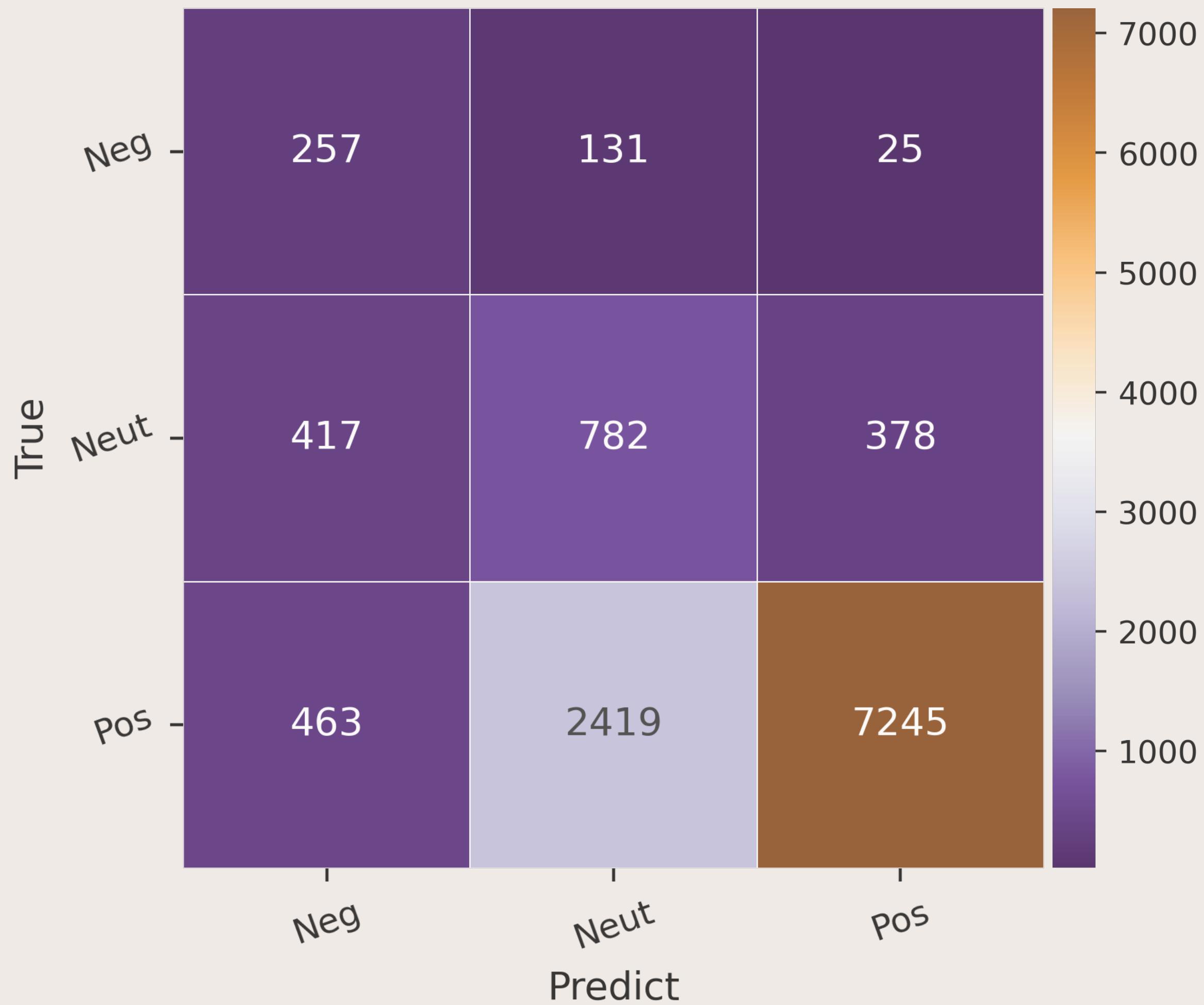
- 資料量：1 萬 2,117 筆

- 測試模型：

- 訓練標籤：
正評論、負評論、中立評論

- 預測準確率：

68.36 %



機器學習

- 測試資料集：

北京大學 餐廳評論數據集

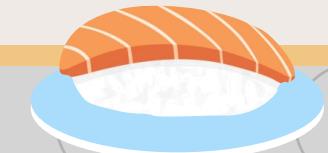
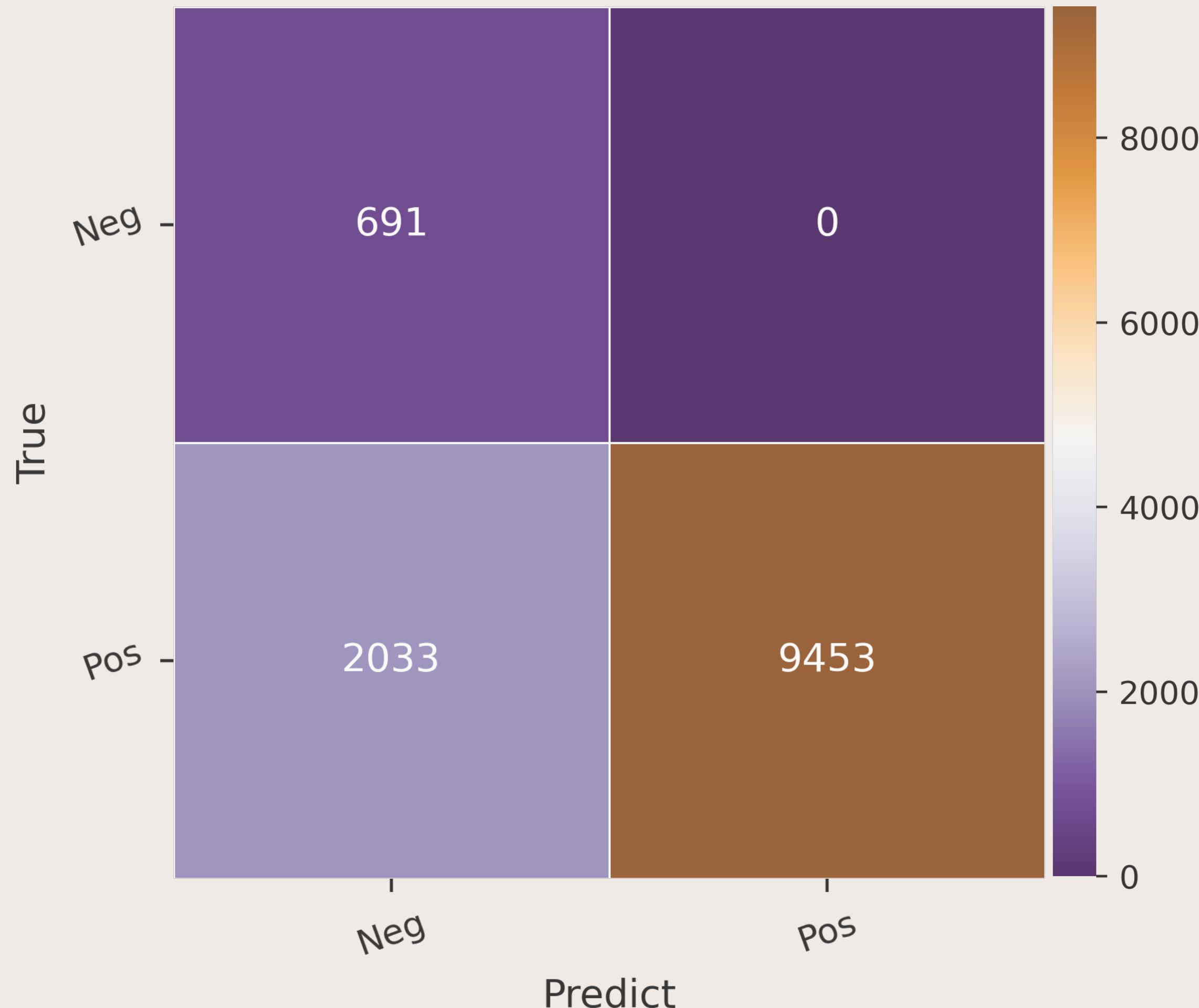
- 資料量：1 萬 2,117 筆

- 測試模型：

- 訓練標籤：
正評論、負評論
- 預測準確率：

83.30 %

- 刪除中立語意標籤，僅
留下正、負標籤



機器學習

- 測試資料集：

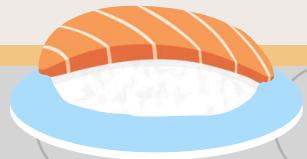
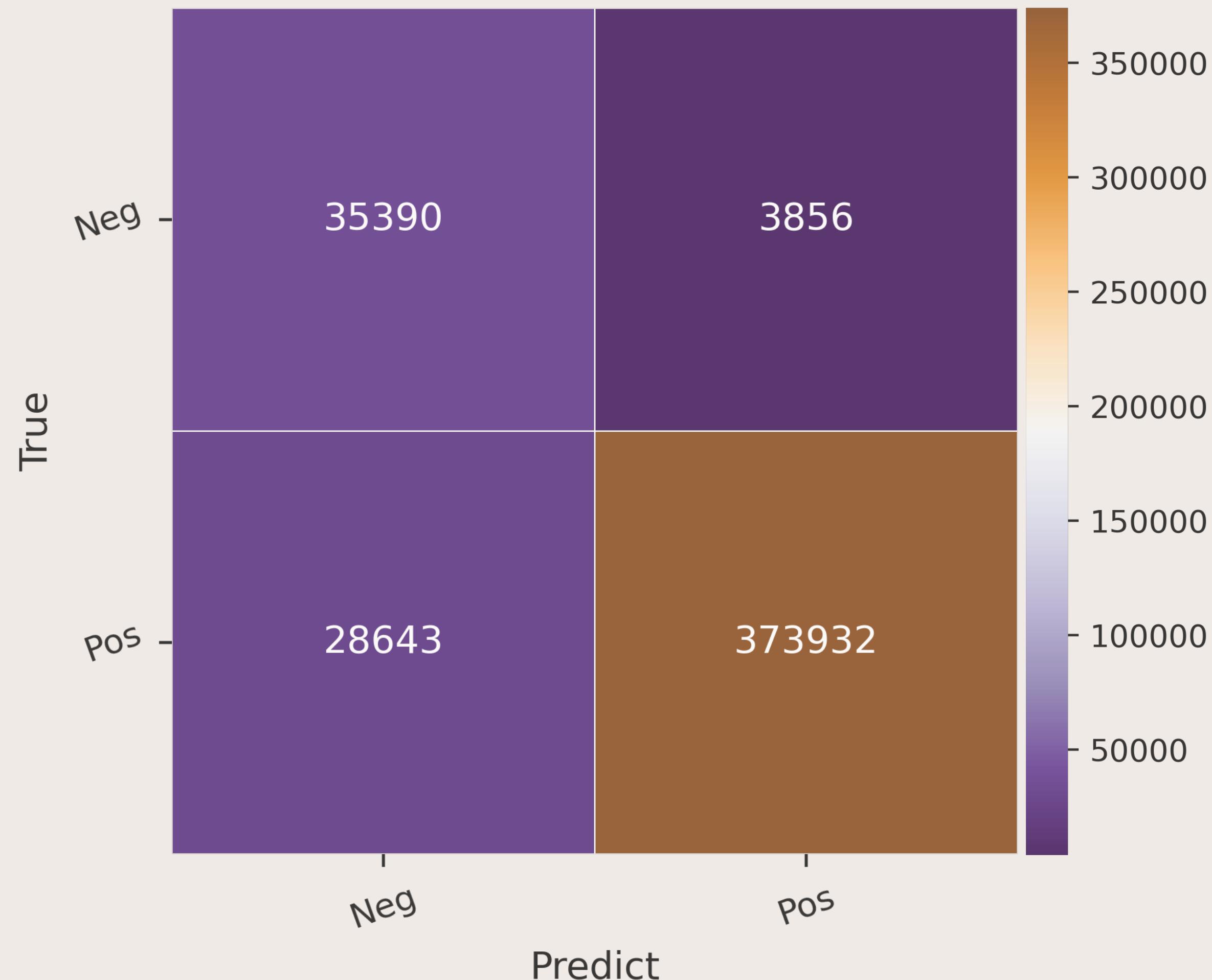
五大評論網站
餐廳評論文字集

- 資料量：約 45 萬

- 模型訓練成果：

預測準確率：

92.64 %





資料視覺化



資料視覺化

平安京茶事

平安京茶事

樂山烹割壽司

旬樂壽司割烹料理

吉兆割烹壽司

京壽司手作割烹

平安京茶事

地址：台北市中正區師大路
165號

電話：02 2368 2277

4.4

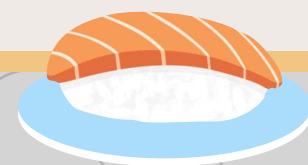
86%

- 道地
- 小貴
- 濃郁



竟然座無虛席，而店員告知需等候一小時才能有座位（我們前面沒有其他顧客在等候）

抹茶味濃蛋糕捲的視覺感受普通，但實際吃起來味道蠻濃郁的！



資料視覺化

餐廳名稱	餐廳地址	餐廳電話	總評論數	總正評數	正評比率	平均星數	關鍵字1	關鍵字2	關鍵字3
Milano Pizzeria	台北市中山區四平街55號2樓	02 2502 0902	67	61	91%	4.2	不錯	驚艷	鮮甜
滙海鮮蒸氣鍋	台北市中山區松江路76號1樓	02 2523 9963	54	49	91%	4.5	不錯	舒適	親切

餐廳基本資料

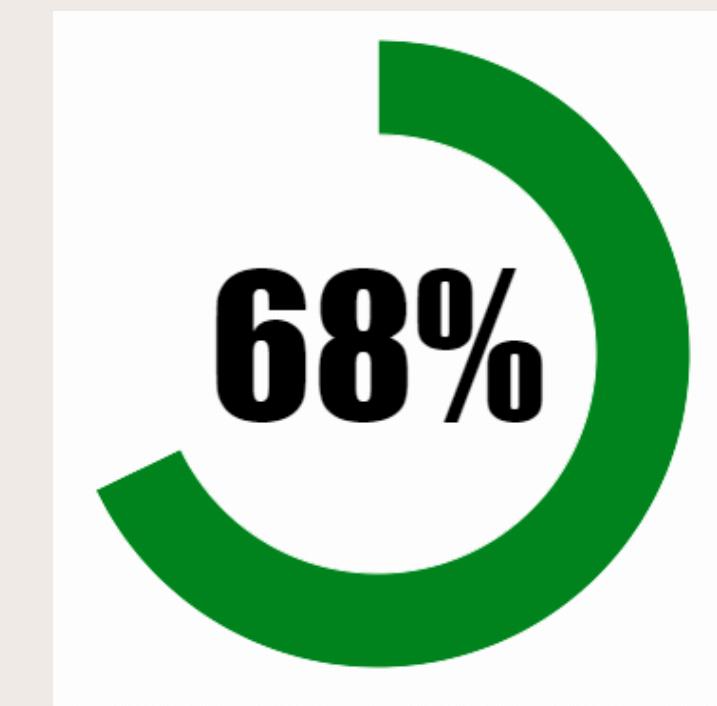


資料視覺化

餐廳名稱	餐廳地址	餐廳電話	總評論數	總正評數	正評比率	平均星數	關鍵字1	關鍵字2	關鍵字3
Milano Pizzeria	台北市中山區四平街55號2樓	02 2502 0902	67	61	91%	4.2	不錯	驚艷	鮮甜
灑海鮮蒸氣鍋	台北市中山區松江路76號1樓	02 2523 9963	54	49	91%	4.5	不錯	舒適	親切



Bert 模型正面評論百分比



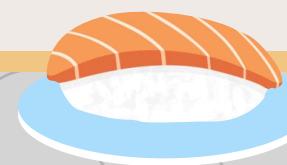
資料視覺化

餐廳名稱	餐廳地址	餐廳電話	總評論數	總正評數	正評比率	平均星數	關鍵字1	關鍵字2	關鍵字3
Milano Pizzeria	台北市中山區四平街55號2樓	02 2502 0902	67	61	91%	4.2	不錯	驚艷	鮮甜
灑海鮮蒸氣鍋	台北市中山區松江路76號1樓	02 2523 9963	54	49	91%	4.5	不錯	舒適	親切



各評論網站評分星星數

平均星星數



資料視覺化

餐廳名稱	餐廳地址	餐廳電話	總評論數	總正評數	正評比率	平均星數	關鍵字1	關鍵字2	關鍵字3
Milano Pizzeria	台北市中山區四平街55號2樓	02 2502 0902	67	61	91%	4.2	不錯	驚艷	鮮甜
灑海鮮蒸氣鍋	台北市中山區松江路76號1樓	02 2523 9963	54	49	91%	4.5	不錯	舒適	親切

Jieba 漲取三個關鍵字





資料視覺化

後端



資料庫



雲端佈署



前端





資料視覺化

後端



資料庫



雲端佈署



前端





資料視覺化

後端



資料庫



雲端佈署



前端



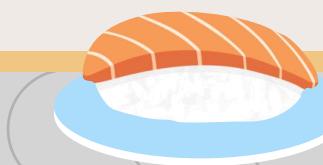
專題介紹

資料蒐集

系統建置

資料預處理

模型建置



總結



資料視覺化

後端



資料庫



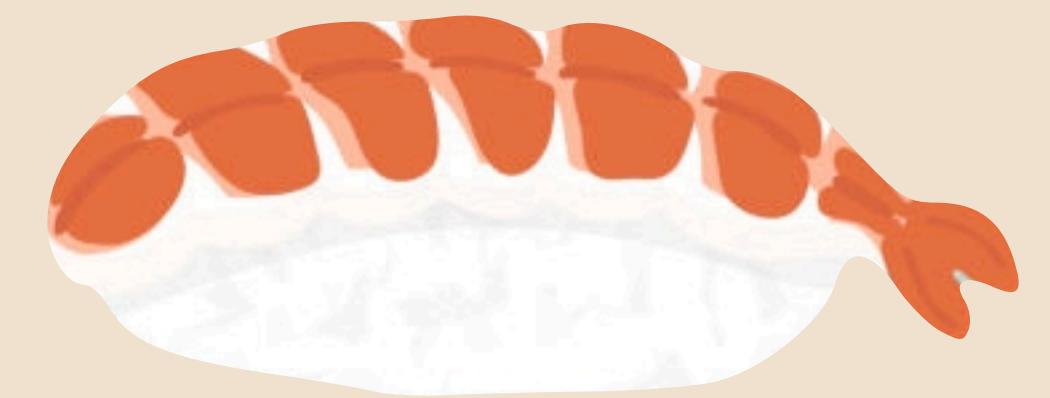
雲端佈署



前端



總結



總結

優化方向

- 調整模型參數，提升模型預測準確度
- 將語意標籤依正負面區分

未來展望

- 提供使用者對店家評論
- 串接經濟部經濟地理資訊系統的「商業選址」功能
- 與店家合作，提供食材產銷履歷等資訊



金魚

金魚

金魚

金魚