

Midterm Project Report

資管碩一 黃彥哲 R13725045

1. System components

- 透過 user data 在每個 cloud node 啟動時寫入 add_chord_node.py、chord_flask_api.py，功能分別如下：

➤ add_chord_node.py

功能：

- **初始化 Chord 環**：如果當前節點是環中的第一個節點，則創建一個新的 Chord 環。
- **加入已有的 Chord 環**：如果當前節點是新加入的，它將查找已存在的節點，並加入到已有的 Chord 環中。
- **Data Migration**：當新節點加入後，會根據節點的 hash 範圍從後繼節點遷移檔案。
- **查詢節點的前驅和後繼節點**：節點之間需要互相查詢彼此的前驅和後繼節點，以確保資料能夠正確地分佈並保持一致性。

實作方法：

- **使用現有工具**：程式使用了 boto3（AWS SDK for Python）來與 AWS 服務（如 EC2 和 Auto Scaling）進行互動，並獲取當前 EC2 實例的公共 IP 地址和其他實例的 IP 地址。
- **自定義程式碼**：
 - 使用 Python 編寫程式邏輯，根據 Chord 協定來實現節點的加入、查詢、數據遷移等功能。

- 使用 msgpackrpc 來實現節點之間的 RPC 通訊，實現節點之間的資料共享和同步。
- 使用 subprocess 運行 chord 執行檔，來初始化或加入 Chord 環。

➤ **chord_flask_api.py**

功能：

提供一個 HTTP 接口，查詢當前節點上的檔案及其 hash 值。

實作方法：

- **使用現有工具：**使用 Flask 框架來實現 Web API。
- **自定義代碼：**
 - 程式會列出當前節點的 /files 目錄中的檔案，並計算每個檔案的 hash 值（使用 hashlib 庫）。
 - 提供 /get_files 路由，將檔案名稱及其哈希值返回給用戶，這些檔案被存儲在當前節點的文件系統中。

2. System functionalities

2.1 Data Migration

1. 找到 new_node 的 predecessor 和 successor，透過 new_node 的 hash 值、predecessor 的 hash 值確認須遷移的 hash 值範圍。
2. 透過 flask api 查詢 successor 有哪些檔案。
3. 判斷 successor 內每個檔案的 hash 值是否在 new_node 的 hash 值、predecessor 的 hash 值之間（需考慮環狀邊界）。
4. 如果是的話在 new_node 透過 download.py 將檔案下載到 new_node。

2.2 File Chunks

透過 upload_chunk.py、download_chunk.py 來執行上傳和下載。

- upload_chunk.py
 1. 根據指定的大小將檔案分割為小 chunk，每個 chunk 會寫入一個單獨的檔案並以 .chunkN 為名稱。
 2. 在第一個 chunk 中，加入關於原始檔案的 metadata（檔案名稱和 chunk 總數）。
 3. 使用 upload.py 腳本將每個 chunk 上傳到 chord node。
- download_chunk.py
 1. 根據每個 chunk 的名稱（chunk_name），從指定的 Chord 節點（node_ip）下載該 chunk。
 2. 在下載的第一個 chunk 中提取 metadata，這些 metadata 包含了原始檔案名稱和總 chunk 數量。這些資訊用來確保檔案能夠被正確地組合成完整的檔案。
 3. 將所有下載的 chunk 按順序合併，重建原始檔案，這些 chunk 會依次寫入到一個新的檔案中，第一個檔案塊的元數據會被跳過，不會被寫入到最終檔案中。

3. Experiment

3.1 Data Migration

1. 初始 2 個 chord node，並各自存放數個檔案。
2. 手動新增 1 個 node。
3. ssh 進入新的查看 chord.log，確認遷移範圍、遷移的檔案。
4. 至/files 查看是否有確實遷移檔案。

3.2 File Chunks

1. 透過 upload_chunk.py 上傳檔案(chunk_test.pdf 檔案大小 3801 KB)、並設定 chunk 大小 1MB。
2. ssh 進各個 node 查看是否有成功上傳 chunks。
3. 透過 download_chunk.py 下載檔案(chunk_test.pdf)。

4. 至資料夾查看是否成功下載檔案。

4. References

1. ChatGPT

2.

https://boto3.amazonaws.com/v1/documentation/api/latest/reference/services/autoscaling/client/describe_auto_scaling_groups.html

3.

https://boto3.amazonaws.com/v1/documentation/api/latest/reference/services/ec2/client/describe_instances.html