

國立台南大學資訊工程學系106級專題

Hadoop實作-雲端股票推薦系統

成員：張辰安

1. 摘要

本系統是以「Hadoop為基底建構Hive資料倉儲的股票推薦系統」，透過HDFS儲存資料，使用Hive對分散式資料做統計分析後，再將分析資料傳至推薦系統，給投資者股票上的推薦。

現在的上市公司有上千萬家，投資者取得的資料十分大量，如何處理資料變成有用資訊是一個難題。本系統利用Hadoop儲存、運算特性與Hive的資料處理能力，實行股票推薦的功能，讓投資者面對股票龐大數據不再感到憂心。

3. 使用者介面

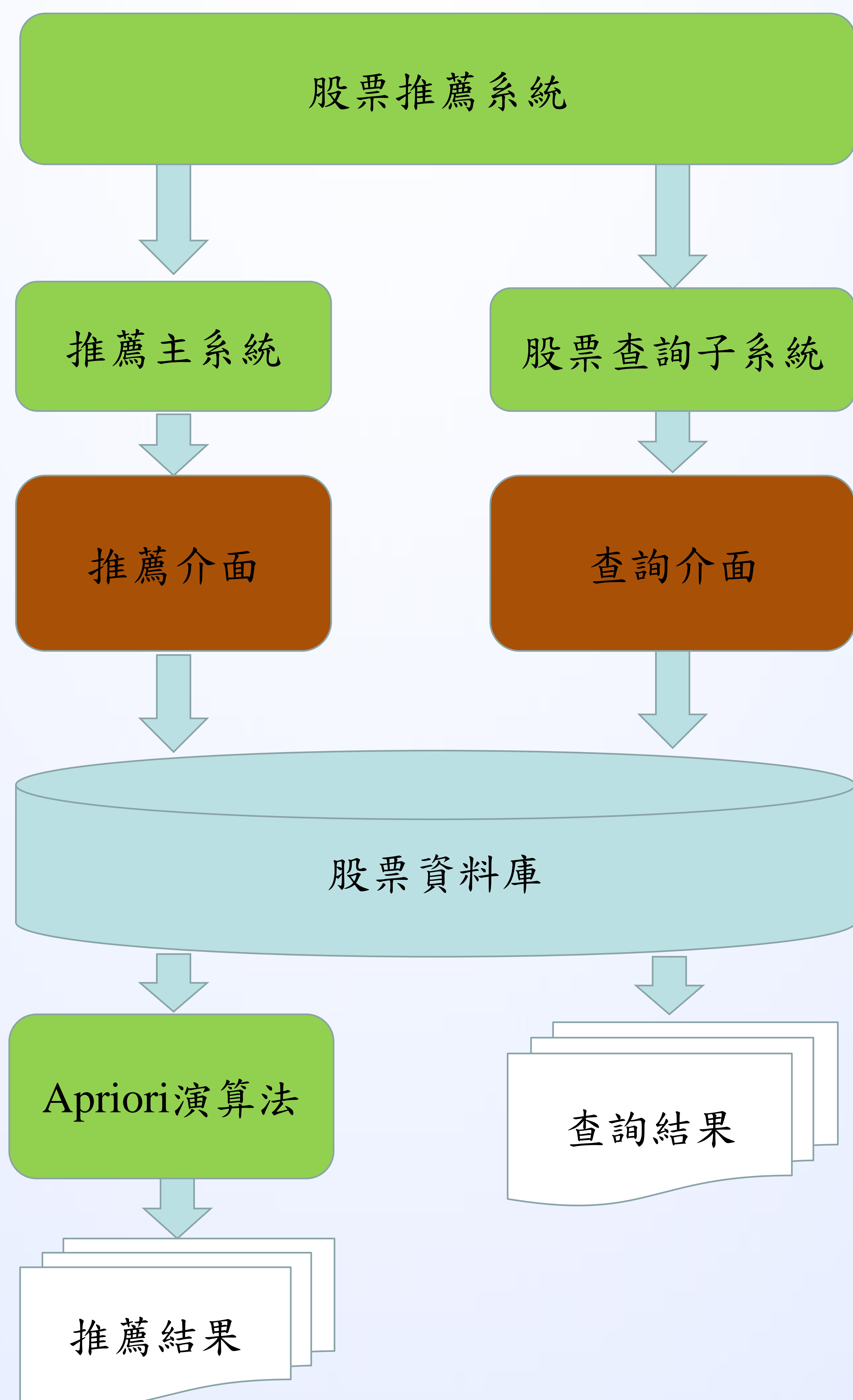
選擇查詢的股票: ☒ 台泥 ☐ 日月光 ☐ 台機電 ☐ 第一金 ☐ 聯電

選擇查詢的年份: ☒ 105

選擇查詢的月份: ☒ Jan. ☐ Feb. ☐ Mar. ☐ Apr. ☐ May. ☐ Jun. ☐ Jul. ☐ Aug. ☐ Sep. ☐ Oct. ☐ Nov. ☐ Dec.

日期	成交筆數	成交金額	成交股數	開盤價	最高價	最低價	收盤價	漲跌價差
1050801	2365	144365898	4262763	34	34	33.6	33.9	0
1050802	2227	146370667	4293001	33.9	34.4	33.85	34.4	0.5
1050803	3317	221607455	6428564	33.7	34.95	33.3	34.9	0.5
1050804	2399	152776627	4434113	34.9	34.9	34.1	34.5	-0.4
1050805	3453	197560841	5666704	34.5	35.15	34.35	34.9	0.4
1050808	1768	141276456	4043087	35	35.1	34.65	35	0.1
1050809	1929	165037436	4720704	34.85	35.05	34.8	35	0

2. 系統架構



4. 系統說明及優點

1. 本系統使用Apriori演算法與關聯法則從股票漲跌資訊中找出高度相關性的股票，並提高推薦的可信度。
2. 在使用介面中，使用者輸入欲查詢的股票、推薦日期、漲跌與支持度的參數，系統便會由信賴度的高低列出符合參數的推薦股票，信賴度越高則推薦信心則越大，給投資者當作一個參考的指標。
3. 本系統使用分散式的儲存方式，相較於一般單機下的資料庫，資料上限可大大提升。
4. 本系統所採用的Hive可將結構化的數據文件轉成資料表並提供查詢語言，是十分適合數據倉儲的統計分析。
5. Hive在本系統會將資料處理方式，轉換成MapReduce處理的形式，MapReduce在運算中使用Divide&Conquer來解決問題，效率及時間都叫一般方法提升許多。

5. 結論

本系統可以依據過去的股票資訊，以信賴度來推薦投資者想要的股票，而支持度的變化也會影響顯示的股票數量，當支持度太過龐大，顯示的數量也會有過少的情況。

現實生活中影響股票波動的因素不只本系統的參數而已，然而本系統的參數基於理論上較有量化的可能，其他如天災、金融政策等較不可量化之因素是比較難以參考的，希望本系統未來能將更多因素納入考量，以提升更高的推薦準確度。