



(19) 中華民國智慧財產局

(12) 發明說明書公告本

(11) 證書號數：TW I626548 B

(45) 公告日：中華民國 107 (2018) 年 06 月 11 日

(21) 申請案號：106110943

(22) 申請日：中華民國 106 (2017) 年 03 月 31 日

(51) Int. Cl. : G06F17/30 (2006.01)

(71) 申請人：東森信息科技股份有限公司 (中華民國) EASTERN INFORMATION AND TECHNOLOGY CO., LTD (TW)

新北市中和區景平路 258 號 3 樓

(72) 發明人：粘仲仁 NIEN, CHUNG JEN (TW)

(74) 代理人：廖嫻涵

(56) 參考文獻：

TW I333380

TW 201640352A

TW 201643707A

US 6675133B2

US 9047717B2

審查人員：林文琦

申請專利範圍項數：15 項 圖式數：5 共 24 頁

(54) 名稱

資料收集與儲存系統及其方法

DATA COLLECTION AND STORAGE SYSTEM AND METHOD THEREOF

(57) 摘要

一種資料收集與儲存系統，用於收集並儲存來自至少一使用者裝置之複數訊息資料，該資料收集與儲存系統係主要包含一資料收集子系統，該資料收集子系統包括至少一資料中介模組、至少一事件處理模組，及至少一事件載入模組，該資料中介模組包括一第一資料區及一第二資料區，該等訊息資料係儲存於該第一資料區，該事件處理模組自該資料中介模組之第一資料區讀取該等訊息資料，再根據該等訊息資料及一預設條件得到至少一資料段，並將該資料段寫入該資料中介模組之第二資料區，該事件載入模組定時地自該資料中介模組之該第二資料區讀取該資料段。

A data collection and storage system for collecting and storing a plurality of information data from at least one user device, the data collection and storage system mainly comprising a data collection subsystem, the data collection subsystem including at least one data middleware module, at least one event processing module, and at least one event loading module, the data middleware module including a first data area and a second data area, the information data are stored in the first data area, the event processing module reading the information data from the first data area of the data middleware module and obtaining at least one data segment based on the information data and a predetermined condition, and writing the data segment into the second data area of the data middleware module, and the event loading module reading the data segment from the second data area of the data middleware module at regular intervals.

指定代表圖：

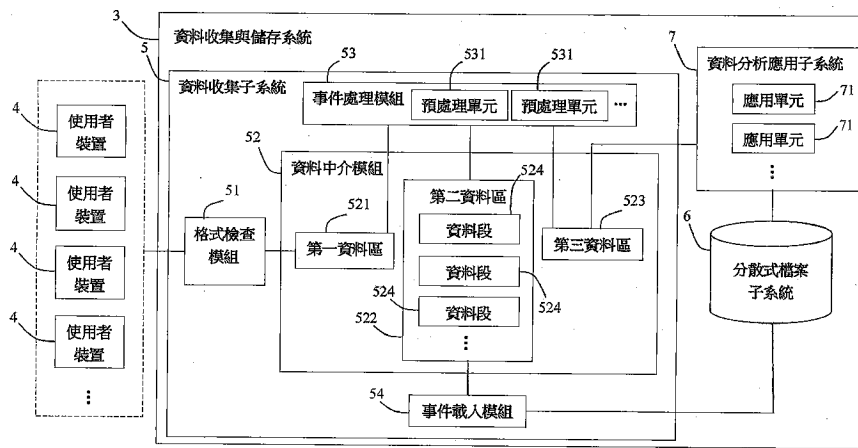


圖 2

符號簡單說明：

3 . . . 資料收集與儲存系統

4 . . . 使用者裝置

5 . . . 資料收集子系統

51 . . . 格式檢查模組

52 . . . 資料中介模組

521 . . . 第一資料區

522 . . . 第二資料區

523 . . . 第三資料區

524 . . . 資料段

53 . . . 事件處理模組

531 . . . 預處理單元

54 . . . 事件載入模組

6 . . . 分散式檔案子系統

7 . . . 資料分析應用子系統

71 . . . 應用單元

圖 1

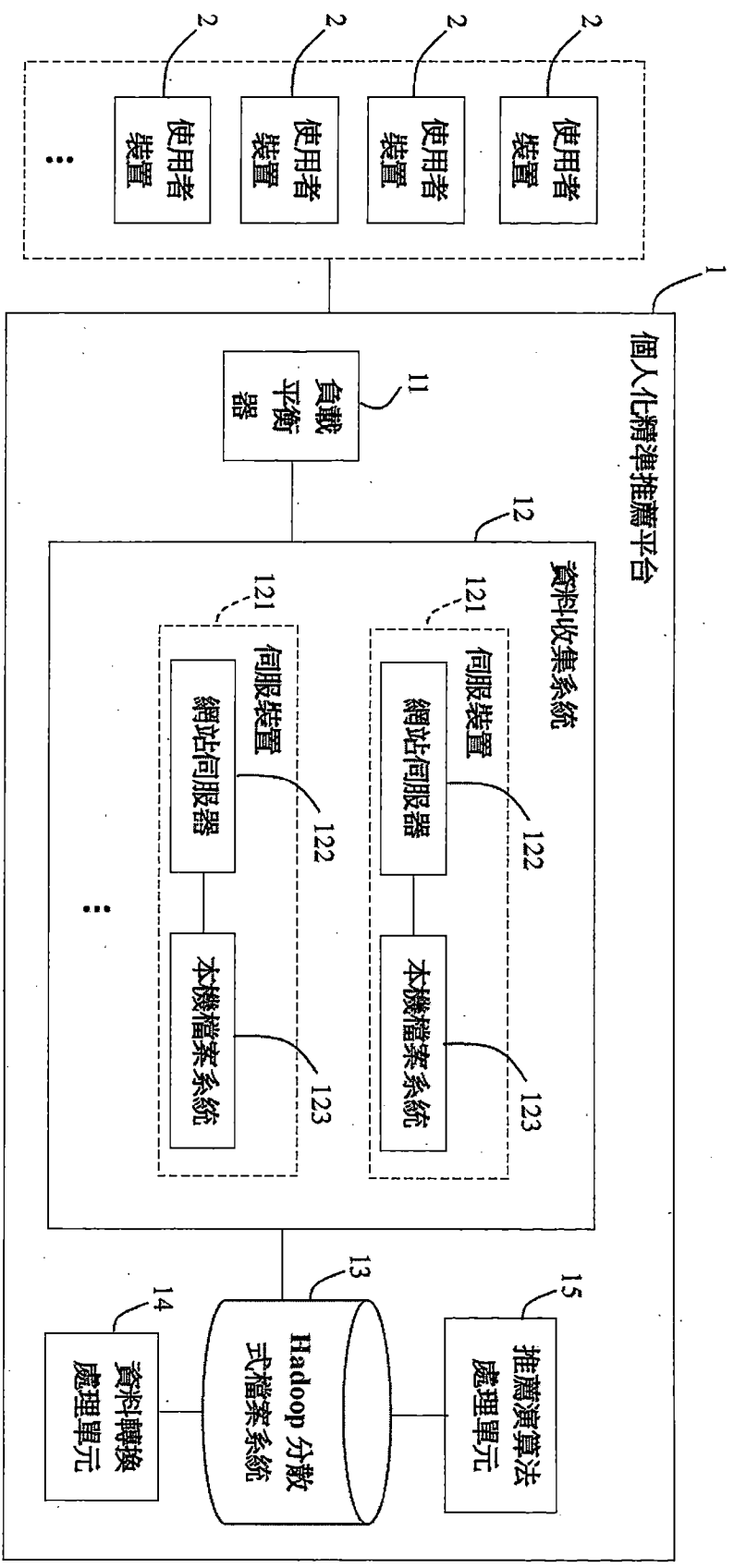


圖 1

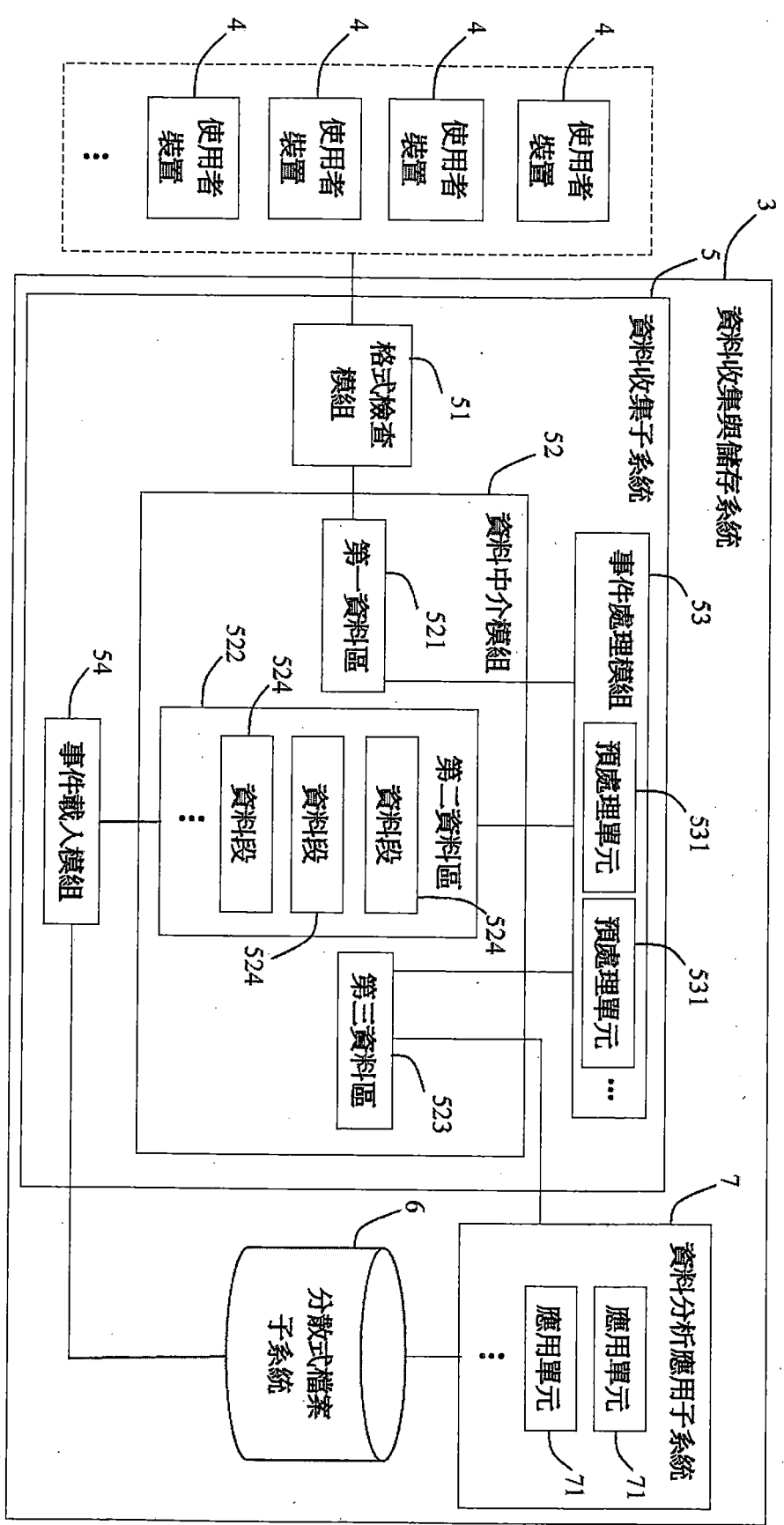


圖 2

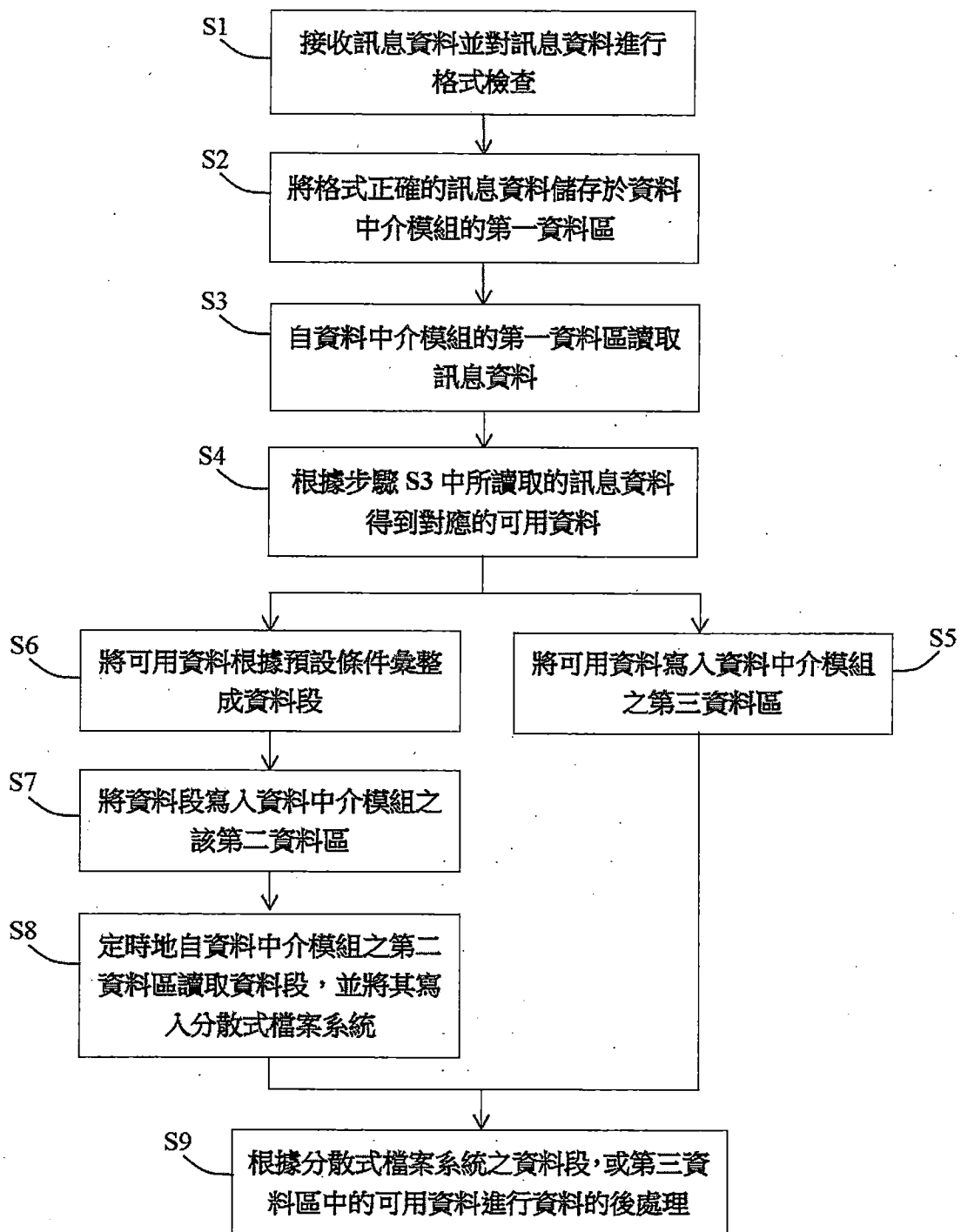


圖 3

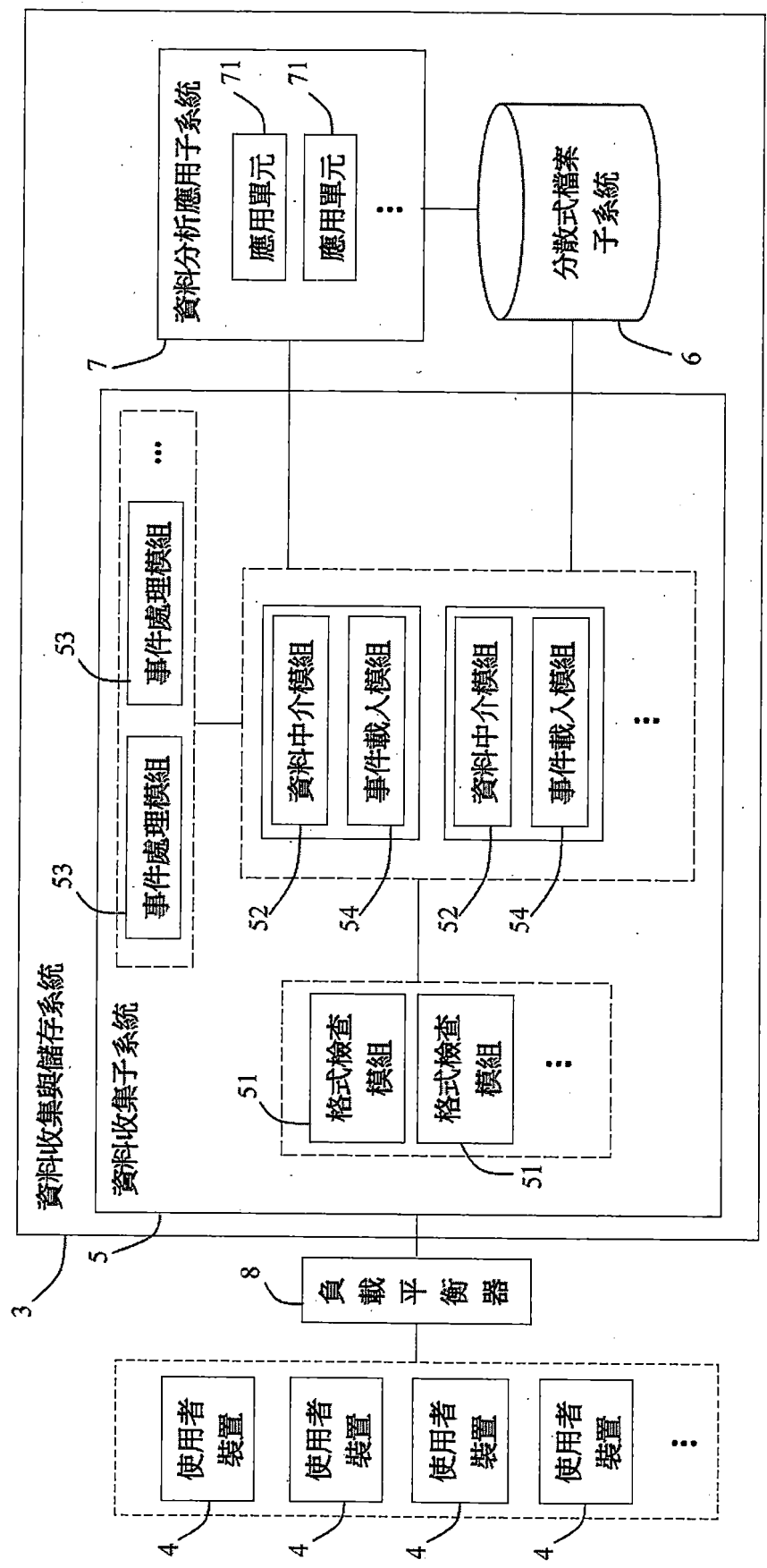


圖 4

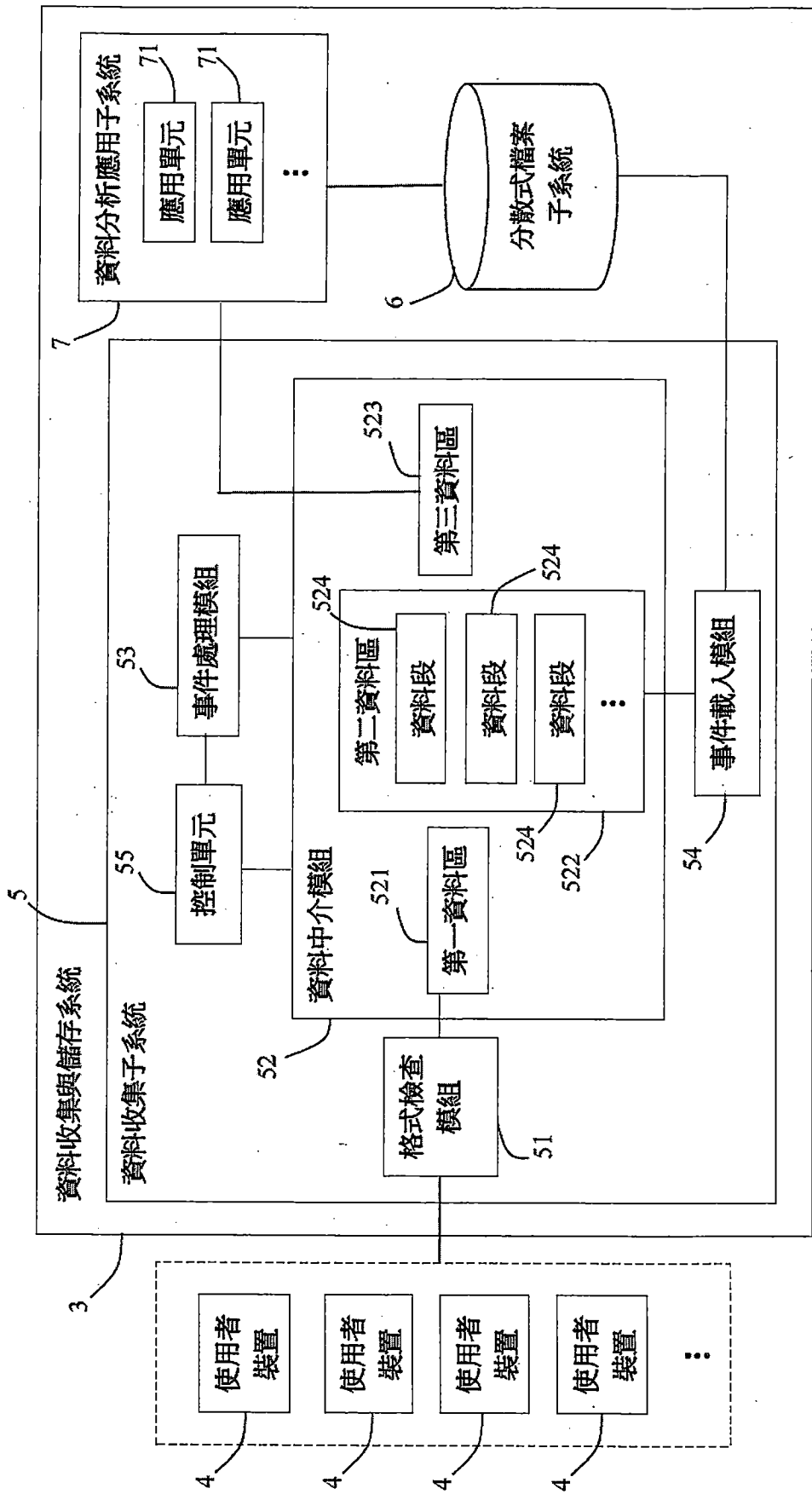


圖 5

發明專利說明書

(本說明書格式、順序，請勿任意更動)

【發明名稱】(中文/英文)

資料收集與儲存系統及其方法/ DATA COLLECTION AND STORAGE
SYSTEM AND METHOD THEREOF

【技術領域】

【0001】 本發明是有關於一種資料收集與儲存系統及其方法，特別是指一種應用於大數據(Big Data)的資料收集與儲存系統及其方法。

【先前技術】

【0002】 近年來，隨著大數據之應用日益廣泛，與大數據處理有關的技術及運算平台，例如，Hadoop，也隨之蓬勃發展。Hadoop為Apache軟體基金會底下的開放原始碼計劃(Open Source Project)，它可以提供大量資料的分散式運算環境及儲存空間。

【0003】 請參閱圖1，習知的一個人化精準推薦平台1，舉例來說，個人化精準推薦平台1係為一種精準推薦和消費者行為分析平台(Etu Recommender)，架構在Hadoop平台上，該個人化精準推薦平台1用於接收來自於複數使用者裝置2的訊息資料，並基於該等訊息資料給出至少一推薦清單。該個人化精準推薦平台1包含一負載平衡器(Load Balancer)11、一資料收集系統12、一Hadoop分散式檔案系統(Hadoop Distributed File System，簡稱HDFS)13、一資料轉換處理單元14，及一推薦演算法處理單元15；其中，該資料收集系統12包括複數伺服器裝置121，每一伺服器裝置121包括一網站伺服器(Web Server)122，及一本機檔案系統(Local File System)123，該網站伺服器122係使用NGINX來實現。

【0004】 其中，來自於該等使用者裝置2的訊息資料是以GET方法(GET Method)經由該負載平衡器11傳送到該資料收集系統12的該等伺服裝置121，該等訊息資料包括複數資料鍵/資料值對(Key/Value Pairs)；每一伺服裝置121的該網站伺服器122接收該等訊息資料，並將其寫入該本機檔案系統123；該資料收集系統12定期地(例如，每小時)將收集到的訊息資料寫入該Hadoop分散式檔案系統13；該資料轉換處理單元14係對該Hadoop分散式檔案系統13的內容進行Hadoop擷取-轉換-載入(Extract-Transform-Load，簡稱ETL)處理，以將資料轉換成可以匯入資料庫的格式；該推薦演算法處理單元15係根據該Hadoop分散式檔案系統13的內容進行運算分析(例如，資料相關性分析、資料相似度分析)，以給出該推薦清單，其中，該推薦清單包括商品推薦資訊、內容推薦資訊、站內廣告推薦資訊、、、等。

【0005】 習知的該個人化精準推薦平台1的資料收集系統12僅是收集訊息資料，然後定期地將收集到的訊息資料批次寫入該Hadoop分散式檔案系統13，並未對訊息資料進行任何的轉換或加工，使得後續對於訊息資料的應用較為受限且不夠便利。

【發明內容】

【0006】 因此，本發明之目的，即在提供一種資料收集與儲存系統，適用於收集並儲存來自於至少一使用者裝置之複數訊息資料，該資料收集與儲存系統係主要包括一資料收集子系統，該資料收集子系統包括至少一資料中介模組、至少一事件處理模組，及至少一事件載入模組，其中該資料中介模組包括一第一資料區、一第二資料區及一第三資料區，該等訊息

資料係儲存於該第一資料區，其中，每一筆訊息資料對應一事件發生時間；該事件處理模組電性連接於該資料中介模組，用以自該資料中介模組的該第一資料區讀取該等訊息資料，再根據該等訊息資料及一預設條件得到至少一資料段，並將該資料段寫入該資料中介模組的該第二資料區，其中該事件處理模組包括至少一預處理單元，該預處理單元用以根據該等訊息資料進行資料預處理以得到對應之複數可用資料，該事件處理模組還將該等可用資料寫入該資料中介模組之該第三資料區；該事件載入模組電性連接於該資料中介模組，用以定時地自該資料中介模組的該第二資料區讀取該資料段。

【0007】 本發明之另一目的，即在提供一種資料收集與儲存方法，適用於收集並儲存來自於至少一使用者裝置的複數訊息資料，該資料收集與儲存方法係以一資料收集與儲存系統來執行，該資料收集與儲存系統包含一資料收集子系統及一分散式檔案子系統，該資料收集子系統包括一資料中介模組，該資料中介模組包括一第一資料區、一第二資料區，及一第三資料區，該資料收集與儲存方法包含下列步驟：**(a)**接收該等訊息資料並將該等訊息資料儲存於該資料中介模組之第一資料區；**(b)**自該資料中介模組之第一資料區讀取該等訊息資料；**(c)**根據該等訊息資料及一預設條件得到至少一資料段，其中該步驟**(c)**包括下列子步驟：**(c-1)**根據該等訊息資料進行資料預處理以得到對應之複數可用資料，及**(c-2)**將對應於該等訊息資料的可用資料依該等事件發生時間進行排序，並以一時間區間為單位彙整成該資料段；**(d)**將該資料段寫入該資料中介模組之第二資料區；及**(e)**定時地自該資料中介模組之第二資料區讀取該資料段，並將該資料段寫入該分

散式檔案子系統。

【0008】 本發明之功效在於：該事件處理模組根據該等訊息資料及該預設條件得到該資料段，再由該事件載入模組定時地將該資料段寫入該分散式檔案子系統，增加了該等訊息資料在後續的應用範圍及便利性。

【圖式簡單說明】

【0009】 本發明之其他的特徵及功效，將於參照圖式之實施方式中清楚地呈現，其中：

【0010】 圖1是一方塊圖，說明習知之一個人化精準推薦平台；

【0011】 圖2是一方塊圖，說明本發明資料收集與儲存系統之一第一較佳實施例；

【0012】 圖3是一流程圖，說明對應該第一較佳實施例的一資料收集與儲存方法；

【0013】 圖4是一方塊圖，說明本發明資料收集與儲存系統之一第二較佳實施例；及

【0014】 圖5是一方塊圖，說明本發明資料收集與儲存系統之一第三較佳實施例。

【實施方式】

【0015】 有關本發明之前述及其他技術內容、特點與功效，在以下配合參考圖式之三個較佳實施例之詳細說明中，將可清楚的呈現。

【0016】 在本發明被詳細描述之前，應當注意在以下的說明內容中，類似或等同之元件是以相同之編號來表示。

【0017】 請參閱圖2，本發明資料收集與儲存系統3之一第一較佳實

施例，適用於收集並儲存來自於至少一使用者裝置4的複數訊息資料，該資料收集與儲存系統3電性連接於該使用者裝置4；該資料收集與儲存系統3包括一資料收集子系統5、電性連接於該資料收集子系統5之一分散式檔案子系統6，及電性連接於該資料收集子系統5及該分散式檔案子系統6之一資料分析應用子系統7。

【0018】 在本第一較佳實施例中，該分散式檔案子系統6為一Hadoop分散式檔案子系統；該資料分析應用子系統7包括至少一應用單元71，其中，該資料分析應用子系統7可以為具有各種特定應用或功能的一伺服器，該應用單元71可以軟體的方式來實施，其實施態樣為建置於該資料分析應用子系統7內的應用程式，然，該應用單元71亦可以韌體或硬體的方式來實施，並不限於本第一較佳實施例所揭露。

【0019】 該資料收集子系統5包含一格式檢查模組51、電性連接於該格式檢查模組51之一資料中介模組52、電性連接於該資料中介模組52之一事件(Event)處理模組53、及電性連接於該資料中介模組52之一事件載入模組54。其中，該資料中介模組52包括一第一資料區521、一第二資料區522，及一第三資料區523。

【0020】 該資料收集子系統5之該格式檢查模組51接收來自於該使用者裝置4之該等訊息資料，並對該等訊息資料進行格式檢查，格式正確之該等訊息資料會傳送至該資料收集子系統5之該資料中介模組52，並儲存於該第一資料區521。在本第一較佳實施例中，該等訊息資料是以POST方法(POST Method)由使用者裝置4傳送至資料處理系統5，更進一步來說，該等訊息資料是採用JSON格式、放在訊息體(message-body)內傳送到該資料收集

子系統5。

【0021】 該事件處理模組53自該資料中介模組52之該第一資料區521讀取該等訊息資料，每一筆訊息資料對應一事件發生時間，該事件處理模組53係根據該等訊息資料及一預設條件得到至少一個資料段524，並將該資料段524寫入該資料中介模組52的該第二資料區522。

【0022】 在本第一較佳實施例中，該事件處理模組53包括至少一預處理(Pre-Processing)單元531，該事件處理模組53之該預處理單元531用以根據該等訊息資料進行資料預處理以得到對應之複數可用資料。舉例來說，該預處理單元531可由該等訊息資料中之IP位址(IP Adress)得到這個IP所屬國別、來源等可用資料；其中，該預處理單元531可視後端之該資料分析應用子系統7之需求來進行開發或透過實體介面介接擴充。更進一步來說，在本第一較佳實施例中，該預設條件係相關於該等事件發生時間，換言之，該事件處理模組53係將對應於該等訊息資料之可用資料依事件發生時間進行排序，然後以一時間區間(例如，每小時)為單位彙整成該資料段524後，寫入該資料中介模組52之該第二資料區522；此外，該事件處理模組53還可將未依事件發生時間進行排序之該等可用資料直接寫入該資料中介模組52之該第三資料區523；而該事件載入模組54定時地(例如，每小時)自該資料中介模52之該第二資料區522讀取該資料段524，並將該資料段524寫入該分散式檔案子系統6。

【0023】 該資料分析應用子系統7用以根據該分散式檔案子系統6之該資料段524，或該資料中介模組52之該第三資料區523中的可用資料進行資料的後處理；其中，資料的後處理包括由該應用單元71進行的資料統計、

分析、歸納等處裡。在本第一較佳實施例中，當在該資料分析應用子系統7之應用單元71中，有需要根據即時更新之可用資料來進行資料的後處理時，該資料分析應用子系統7可以直接自該資料中介模組52之該第三資料區523進行讀取；當在該資料分析應用子系統7之應用單元71中，有需要根據每一時間區間所彙整而成之該資料段524來進行統計、分析時，該資料分析應用子系統7即可自該分散式檔案子系統6讀取其所需之該資料段524。更進一步來說，在本第一較佳實施例中，該分散式檔案子系統6為Hadoop分散式檔案子系統，由於Hadoop分散式檔案子系統不支援隨機存取(Random Access)的方式，只支援批次存取方式；再加上當每一訊息資料產生(對應事件發生時間)後，可能會有臨時斷訊情況發生，若不將該等可用資料彙整成該資料段524，無法保證該資料中介模組52中該等可用資料是依其等對應之事件發生時間儲存，當該資料分析應用子系統7之應用單元71需要某一時間區間內的可用資料時，將無法得知需要自該分散式檔案子系統6連續讀取多少位移量(Offset)的資料；因此，將該等可用資料彙整成該資料段524可使得該資料分析應用子系統7之應用單元71在資料上讀取更為便利。

【0024】 此外，在本第一較佳實施例中，各元件之間於網路上資料傳輸皆是經過壓縮，以降低整體之網路流量負載。

【0025】 請參閱圖2與圖3，對應上述第一較佳實施例之一資料收集方法包含下列步驟：

【0026】 如步驟S1所示，接收來自於該使用者裝置4之該等訊息資料，並對該等訊息資料進行格式檢查；

【0027】 如步驟S2所示，將格式正確之該等訊息資料儲存於該資料

中介模組52之該第一資料區521；

【0028】 如步驟S3所示，自該資料中介模組52之該第一資料區521讀取該等訊息資料；

【0029】 如步驟S4所示，根據步驟S3中所讀取之該等訊息資料得到對應之該等可用資料；

【0030】 如步驟S5所示，將該等可用資料寫入該資料中介模組52之該第三資料區523；

【0031】 如步驟S6所示，根據該預設條件將該等可用資料彙整成該資料段524；

【0032】 如步驟S7所示，將該資料段524寫入該資料中介模組52之該第二資料區522；及

【0033】 如步驟S8所示，定時地自該資料中介模組52之該第二資料區522讀取該資料段524，並將該資料段524寫入該分散式檔案子系統6。

【0034】 如步驟S9所示，根據該分散式檔案子系統6之該資料段524，或該資料中介模組52之該第三資料區523中的可用資料進行資料的後處理。

【0035】 請參閱圖4，本發明資料收集與儲存系統3之一第二較佳實施例，電性連接於一負載平衡器8；該負載平衡器8電性連接於至少一使用者裝置4。該資料收集與儲存系統3包含一資料收集子系統5、一分散式檔案子系統6，及一資料分析應用子系統7；其中，該資料收集子系統5包括複數格式檢查模組51、電性連接於該等格式檢查模組51的複數資料中介模組52、電性連接於該等資料中介模組52的複數事件處理模組53，及電性連接

於該等資料中介模組52之複數事件載入模組54。

【0036】 由於該等格式檢查模組51、該等資料中介模組52、該等事件處理模組53，及該等事件載入模組54之運作，類似於該第一較佳實施例，故以下不再贅述其等元件的實施細節，僅就其配置上的差異進行描述。

【0037】 該負載平衡器8用於接收來自於該使用者裝置4之複數訊息資料，並將該等訊息資料分配至該等格式檢查模組51進行格式檢查；較佳地，每一資料中介模組52係與一事件載入模組54整合在同一裝置內，以更進一步地降低整體網路流量負載。在本第二較佳實施例中，該等資料中介模組52及該等事件載入模組54係以Kafka佇列(Kafka Queues)方式來實現，每一資料中介模組52為一Kafka服務器(Kafka-Broker)。

【0038】 請參閱圖5，本發明資料收集與儲存系統3之一第三較佳實施例，電性連接於至少一使用者裝置4，該資料收集與儲存系統3接收來自於該使用者裝置4的複數訊息資料，該資料收集與儲存系統3包含一資料收集子系統5、一分散式檔案子系統6，及一資料分析應用子系統7；其中，該資料收集子系統5包括一格式檢查模組51、電性連接於該格式檢查模組51的一資料中介模組52、電性連接於該資料中介模組52的一控制單元55、電性連接於該控制單元55及該資料中介模組52的一事件處理模組53，及電性連接於該資料中介模組52之一事件載入模組54。其中，該資料中介模組52包括一第一資料區521、一第二資料區522，及一第三資料區523。與上述第一、第二較佳實施例類似之處不再贅述，以下僅就差異處進行描述。

【0039】 在本第三較佳實施例中，該控制單元55可供進行條件設定，並將設定好的條件提供給該事件處理模組53以進行相關處理。舉例來

說，可於該控制單元55設定一篩選條件，以供該事件處理模組53篩選來自於該資料中介模組52之第一資料區521之該等訊息資料；也可以於該控制單元55設定一預設條件，以供該事件處理模組53根據該等訊息資料及該預設條件得到至少一個資料段524，並將該資料段524寫入該資料中介模組52的該第二資料區522；廣義言之，該事件處理模組53根據該控制單元55設定好的條件進行相關處理，可得到不同分群之資料段524，使得該等訊息資料之應用範圍更為廣泛而便利。

【0040】 歸納上述，請參閱圖2、圖4及圖5，本發明資料收集與儲存系統3係將大數據之資料收集與儲存進行最佳化，其功效如下：其一、該等訊息資料以POST方法傳送，不用擔心資料量大小之限制，該等訊息資料內容可以更為豐富且多樣，且以POST方法傳送之資料安全性也較高；其二、藉由該(等)事件處理模組53根據該等訊息資料進行資料預處理以得到該等可用資料，並根據該預設條件將該等可用資料彙整成該資料段524後，由該(等)事件載入模組54定時地寫入該分散式檔案子系統6，而該資料分析應用子系統7可視其應用單元71之需求，直接自該資料中介模組52的該第三資料區523讀取該等可用資料，或讀取該分散式檔案子系統6內的該資料段524，大大地增加了該等訊息資料之應用範圍及便利性；其三、各元件之間於網路上的資料傳輸皆是經過壓縮，可降低整體的網路流量負載；故確實能達成本發明之目的。

【0041】 惟以上所述者，僅為本發明之實施例而已，當不能以此限定本發明實施之範圍，凡是依本發明申請專利範圍及專利說明書內容所作之簡單的等效變化與修飾，皆仍屬本發明專利涵蓋之範圍內。

【符號說明】**【0042】**

- 3 資料收集與儲存系統
- 4 使用者裝置
- 5 資料收集子系統
 - 51 格式檢查模組
 - 52 資料中介模組
 - 521 第一資料區
 - 522 第二資料區
 - 523 第三資料區
 - 524 資料段
 - 53 事件處理模組
 - 531 預處理單元
 - 54 事件載入模組
 - 55 控制單元
- 6 分散式檔案子系統
- 7 資料分析應用子系統
 - 71 應用單元
- S1~9 步驟

發明摘要

※ 申請案號：106110943

※ 申請日：106/03/31

※IPC 分類：G06F 17/30 (2006.01)

【發明名稱】(中文/英文)

資料收集與儲存系統及其方法/ DATA COLLECTION AND STORAGE
SYSTEM AND METHOD THEREOF

【中文】

一種資料收集與儲存系統，用於收集並儲存來自至少一使用者裝置之複數訊息資料，該資料收集與儲存系統係主要包含一資料收集子系統，該資料收集子系統包括至少一資料中介模組、至少一事件處理模組，及至少一事件載入模組，該資料中介模組包括一第一資料區及一第二資料區，該等訊息資料係儲存於該第一資料區，該事件處理模組自該資料中介模組之第一資料區讀取該等訊息資料，再根據該等訊息資料及一預設條件得到至少一資料段，並將該資料段寫入該資料中介模組之第二資料區，該事件載入模組定時地自該資料中介模組之該第二資料區讀取該資料段。

【英文】

A data collection and storage system for collecting and storing a plurality of information data from at least one user device, the data collection and storage system mainly comprising a data collection subsystem, the data collection subsystem including at least one data middleware module, at least one event processing module, and at least one event loading module, the data middleware module including a first data area and a second data area, the information data

are stored in the first data area, the event processing module reading the information data from the first data area of the data middleware module and obtaining at least one data segment based on the information data and a predetermined condition, and writing the data segment into the second data area of the data middleware module, and the event loading module reading the data segment from the second data area of the data middleware module at regular intervals.

【代表圖】

【本案指定代表圖】：第（2）圖。

【本代表圖之符號簡單說明】：

- 3 資料收集與儲存系統
- 4 使用者裝置
- 5 資料收集子系統
- 51 格式檢查模組
- 52 資料中介模組
- 521 第一資料區
- 522 第二資料區
- 523 第三資料區
- 524 資料段
- 53 事件處理模組
- 531 預處理單元
- 54 事件載入模組

申請專利範圍

1. 一種資料收集與儲存系統，適用於收集並儲存來自於至少一使用者裝置之複數訊息資料，該資料收集與儲存系統包含：
 - 一資料收集子系統，包括：
 - 至少一資料中介模組，包括一第一資料區、一第二資料區，及一第三資料區，該等訊息資料係儲存於該第一資料區，其中，每一筆訊息資料對應一事件發生時間；
 - 至少一事件處理模組，電性連接於該資料中介模組，該事件處理模組自該資料中介模組之該第一資料區讀取該等訊息資料，再根據該等訊息資料及一預設條件得到至少一資料段，並將該資料段寫入該資料中介模組之該第二資料區，其中該事件處理模組包括至少一預處理單元，該預處理單元用以根據該等訊息資料進行資料預處理以得到對應之複數可用資料，該事件處理模組還將該等可用資料寫入該資料中介模組之該第三資料區；
 - 及
 - 至少一事件載入模組，電性連接於該資料中介模組，該事件載入模組定時地自該資料中介模組之第二資料區讀取該資料段。
2. 如請求項1所述之資料收集與儲存系統，其中該資料收集子系統更包括電性連接於該資料中介模組之至少一格式檢查模組，該格式檢查模組用以對該等訊息資料進行格式檢查，並將格式正確的該等訊息資料儲存於該資料中介模組之該第一資料區。

3. 如請求項1所述之資料收集與儲存系統，其中該資料收集子系統更包括電性連接於該資料中介模組及該事件處理模組之一控制單元，用以設定該預設條件。
4. 如請求項1所述之資料收集與儲存系統，更包括一分散式檔案子系統，該分散式檔案子系統係電性連接該資料收集子系統之事件載入模組，用以接收並儲存自該事件載入模組所傳送之資料段。
5. 如請求項4所述之資料收集與儲存系統，更包括電性連接於該資料收集子系統及該分散式檔案子系統之一資料分析應用子系統，該資料分析應用子系統用以根據該分散式檔案子系統之該資料段，或該資料中介模組之該第三資料區中之可用資料進行資料之後處理。
6. 如請求項5所述之資料收集與儲存系統，其中該資料分析應用子系統包括至少一應用單元。
7. 如請求項1所述之資料收集與儲存系統，其中該資料中介模組與該事件載入模組係整合在同一裝置內。
8. 如請求項1所述之資料收集與儲存系統，其中每一資料中介模組為一Kafka服務器。
9. 一種資料收集與儲存方法，適用於收集並儲存來自於至少一使用者裝置之複數訊息資料，該資料收集與儲存方法係以一資料收集與儲存系統來執行，該資料收集與儲存系統包含一分散式檔案子系統及一資料收集子系統，該資料收集子系統包括一資料中介模組，該資料中介模組包括一第一資料區、一第二資料區，及一第三資料區，該資料收集與儲存方法包

含下列步驟：

(a)接收該等訊息資料並將該等訊息資料儲存於該資料中介模組之該第一資料區；

(b)自該資料中介模組之第一資料區讀取該等訊息資料；

(c)根據該等訊息資料及一預設條件得到至少一資料段，其中該步驟(c)包括下列子步驟：

(c-1)根據該等訊息資料進行資料預處理以得到對應之複數可用資料；及

(c-2)將對應於該等訊息資料的可用資料依該等事件發生時間進行排序，並以一時間區間為單位彙整成該資料段；

(d)將該資料段寫入該資料中介模組之第二資料區；及

(e)定時地自該資料中介模組之第二資料區讀取該資料段，並將該資料段寫入該分散式檔案子系統。

10.如請求項9所述之資料收集與儲存方法，更包含步驟(a)之前的下列步驟：

(f)對該等訊息資料進行格式檢查；及

(g)將格式正確該等訊息資料儲存於該資料中介模之該第一資料區。

11.如請求項9所述之資料收集與儲存方法，其中該步驟(c)所述之該預設條件係相關於對應該等訊息資料的複數事件發生時間。

12.如請求項9所述之資料收集與儲存方法，還包括子步驟(c-1)之後的下列步驟：

(h)將該等可用資料寫入該資料中介模組之第三資料區。

13.如請求項12所述之資料收集與儲存方法，還包括步驟(e)之後的下列步驟：

(i)根據該分散式檔案子系統之該資料段，或該資料中介模組之該第三資料區中的可用資料進行資料的後處理。

14.如請求項9所述之資料收集與儲存方法，其中步驟(a)之訊息資料是以POST方法(POST Method)由使用者裝置傳送至資料處理子系統。

15.如請求項9所述之資料收集與儲存方法，其中步驟(a)之訊息資料是採用JSON格式、放在訊息體(message-body)內傳送到該資料收集子系統。

are stored in the first data area, the event processing module reading the information data from the first data area of the data middleware module and obtaining at least one data segment based on the information data and a predetermined condition, and writing the data segment into the second data area of the data middleware module, and the event loading module reading the data segment from the second data area of the data middleware module at regular intervals.

【代表圖】

【本案指定代表圖】：第（2）圖。

【本代表圖之符號簡單說明】：

- 3 資料收集與儲存系統
- 4 使用者裝置
- 5 資料收集子系統
- 51 格式檢查模組
- 52 資料中介模組
- 521 第一資料區
- 522 第二資料區
- 523 第三資料區
- 524 資料段
- 53 事件處理模組
- 531 預處理單元
- 54 事件載入模組

- 6 分散式檔案子系統
- 7 資料分析應用子系統
- 71 應用單元

【本案若有化學式時，請揭示最能顯示發明特徵的化學式】：