

# **USER BEHAVIOR DATA PROCESSING METHOD AND DEVICE, AND COMPUTER-READABLE STORAGE MEDIUM**

## **CROSS-REFERENCE TO RELATED APPLICATIONS**

**[0001]** The present application is the National Stage of International Application  
5 with No. PCT/CN2017/104101, filed September 28, 2017, which claims the benefit of  
China Patent Application No. 201710636447.1, filed July 28, 2017 with the State  
Intellectual Property Office and entitled “User Behavior Data Processing Method,  
Device, and Computer-readable Storage Medium”, the entirety of which is hereby  
incorporated herein by reference.

10 **[0002]** FIELD

**[0003]** The present disclosure relates to the field of data processing, and more  
particularly relates to a user behavior data processing method and a user behavior data  
processing device, as well as a computer-readable storage medium.

**[0004]** BACKGROUND

15 **[0005]** With the on-going increasing of internet users, user behavior analysis has  
become one of the important analysis methods in the internet field. Recording and  
analysis of the users’ operation behavior are crucial to the development of product.

**[0006]** Currently, the analysis of user behavior is carried out as follows: the  
client-side captures data and uploads the captured data to a WEB server in real time;  
20 when the WEB server receives the data uploaded by the client-side, the WEB server  
uploads the data to a database server; when the database server receives the data  
uploaded by the WEB server, the database server processes the received data and  
analyzes the user behavior based on the data.

**[0007]** However, for the database server needs to process the received data in real  
25 time, as a result, the database server needs to process data for lots of times, which

**[0031]** based on the database table, performing data statistics on the user behavior information according to a preset rule, to obtain statistical information of the user behavior information.

**[0032]** In an exemplary embodiment, after the step of performing data statistics on the user behavior information according to a preset rule, the user behavior data processing method further includes:

**[0033]** obtaining attribute information of the user behavior information;

**[0034]** displaying the attribute information and the statistical information.

**[0035]** In addition, in order to realize the above aim, the present disclosure also provides a user behavior data processing device. The user behavior data processing device includes: a memory, a processor and a user behavior data processing program which stored in the memory and executable by the processor, the user behavior data processing program when being executed by the processor performing steps of the user behavior data processing method.

**[0036]** In addition, in order to realize the above aim, the present disclosure also provides a computer-readable storage medium. The computer-readable storage medium stores a user behavior data processing program when being executed by the processor performing steps of the user behavior data processing method.

**[0037]** In accordance with the present disclosure, log information stored in a preset sever is obtained at a fixed time, wherein, the log information includes user behavior data which is captured and sent to the preset server by client. Next, whether row data satisfying a preset format exists in each row data of the log information is determined, wherein, the row data is the data corresponding to each row of the log information. Then, the row data satisfying the preset format is set as user behavior information. Finally, data statistics is performed on the user behavior information according to a preset rule to obtain statistical information of the user behavior information. Obtaining log information at a fixed time by a HADOOP large data platform avoids the real-time data processing of the HADOOP large data platform, and reduces the amount of processing as well as in turn reduces the operating pressure of the HADOOP large data platform. At the same time, the cleaning operations for the log information by the HADOOP large data platform reduces the amount of data

processing and improves data processing efficiency of the HADOOP large data platform.

**[0038] BRIEF DESCRIPTION OF THE DRAWINGS**

**[0039]** Fig.1 is an schematic diagram of a terminal to which a user behavior data processing device belongs in a hardware operating environment according to an embodiment of this disclosure;

**[0040]** Fig.2 is an illustrative flowchart of a first embodiment of a user behavior data processing method according to this disclosure;

**[0041]** Fig.3 is a detailed illustrative flowchart of the step of setting the row data satisfying the preset format as user behavior information in a second embodiment of a user behavior data processing method according to this disclosure;

**[0042]** Fig.4 is a detailed illustrative flowchart of the step of determining that whether the row data satisfying the preset format includes row data satisfying a preset permission in a third embodiment of a user behavior data processing method according to this disclosure;

**[0043]** Fig.5 is a detailed illustrative flowchart of the step of setting the row data satisfying the preset format as user behavior information in a fourth embodiment of a user behavior data processing method according to this disclosure;

**[0044]** Fig.6 is a detailed illustrative flowchart of the step of setting the row data satisfying the preset format as user behavior information in a fifth embodiment of a user behavior data processing method according to this disclosure;

**[0045]** Fig.7 is a detailed illustrative flowchart of the step of performing data statistics on the user behavior information according to a preset rule to obtain statistical information of the user behavior information in a sixth embodiment of a user behavior data processing method according to this disclosure;

**[0046]** Fig.8 is an illustrative flowchart of a seventh embodiment of a user behavior data processing method according to this disclosure.

**[0047]** The realizing of the aim, functional characteristics and advantages of the

character string dpmp\_log;, row data ending with or beginning with other character or character string, and so on. For example, the preset format to be row data starting with the character string dpmp\_log;, when a user clicks an interface of an application, the client captures the user's behavior data according to a customized script and stores the captured behavior data in log information of the preset server. Then the HADOOP large data platform obtains the log information at a fixed time, and in the HADOOP large data platform, whether the log information is the row data satisfying the preset format is determined, namely whether the row data corresponding to the log information is the row data starting with the preset character or the preset character string, namely the row data starting with the character string dpmp\_log;. When the row data is satisfied with the start of the character string dpmp\_log;, the row data corresponding to the log information is the row data satisfying the preset format.

**[0092]** In other embodiments, it may take various ways to clean the obtained log information in order to get data related to the user behavior in the log information.

For example, it's first to filter out information satisfying the preset format in the log information obtained in the HADOOP large data platform, then, according to a preset condition, to filter layer by layer in order to obtain information data satisfying the preset condition, wherein the preset condition including limitation requirements such as format, symbol, satisfaction formula, number of words after characters and so on in the information data, finally, through filtering of these preset conditions, to obtain the information data satisfying the condition clearly and quickly. Specifically, it's first to obtain row data satisfying a preset format in each row data of the log information, next to obtain row data satisfying a preset permission in the row data satisfying the preset format, to obtain row data including a preset character or a preset character string from the row data satisfying the preset format, and to perform validation on the first word after the preset character or the preset character string in the row data including the preset character or the preset character string, then to determine whether the second word is operators' information, finally to determine in the row data including the preset character or the preset character string, whether exists row data that the number of words after the preset character or the preset character string is less than a preset threshold. The obtained row data less than the preset threshold is user

behavior information. The order of each step in the above cleaning operations can be arbitrarily ordered and combined.

**[0093]** S300, setting the row data satisfying the preset format as user behavior information.

5 **[0094]** In the exemplary embodiment, when row data satisfying the preset format exists in each row data of the log information, the row data satisfying the preset format is set as user behavior information. User behavior data often involves multiple aspects, and user behavior information is user behavior data of a specific aspect that needs to be analyzed or counted. For example, the preset format starting with the  
10 character string dpmp\_log:, when a user clicks to access an interface of an application, the client captures the user's behavior data according to a customized script and stores the captured behavior data into log information in the preset server. The HADOOP large data platform obtains the log information at a fixed time, and in the HADOOP large data platform, it's to determine whether row data starting with the character  
15 string dpmp\_log: exists in the row data corresponding to the log information. When the row data starting with the character string dpmp\_log: exists in the row data corresponding to the log information, the row data is set as user behavior information. Setting the row data as user behavior information includes reserving a storage space in the database named user behavior information library, then saving the row data in the  
20 user behavior information library, or directly naming the row data satisfying the preset format as user behavior information.

**[0095]** S400, performing data statistics on the user behavior information according to a preset rule to obtain statistical information of the user behavior information.

25 **[0096]** The data statistics include: statistics on the number of occurrences of a certain fixed information in the data after cleaning, the number of times the function point appears, the data input by the user, the data corresponding to the time when the cursor staying time exceeds a certain time, and the data corresponding to the button of the user clicking the interface. When a user accesses a preset web page, the client runs  
30 the capture script of the user behavior data embedded in the preset web page corresponding to the preset server and sends the user behavior data captured to the

preset server, and the preset server stores the user behavior data obtained in log information, next the HADOOP large data platform obtains the log information stored in the preset server at a fixed time, then the log information obtained at the fixed time is cleaned, finally statistics and analysis are performed on the user behavior data. For example, log information of user behavior of the investment and finance to be needed, when a user accesses a web page related to the investment and finance, the client captures the access behavior data left in the web page related to the investment and finance according to a customized script, and stores the captured behavior data into log information in the preset server, next the HADOOP large data platform obtains the log information stored in the preset server at a fixed time, then in the HADOOP large data platform, the obtained log information is to be cleaned, finally statistics and analysis are performed on the cleaned log information data. The number of all users accessing the web page can be counted, and the number of users is a manifestation of the number of occurrences of the certain fixed information.

**[0097]** User behavior information is counted according to a certain preset rule. Statistics according to a preset rule can be done according to a statistical identification. For example, it can be counted that the number of times a fixed message appears in the cleaned data or the number of times the function point appears, time sorting of certain data input by users, the specific staying time of data corresponding to cursor staying time exceeds a certain time, and the specific content of data corresponding to a button of a user clicking interface.

**[0098]** In the exemplary embodiment, the purpose of data statistics is to arrange and display the data information obtained after cleaning according to a certain preset rule, therefore it's needed to perform data statistics on the cleaned data. Statistical information refers to information obtained by statistics according to a preset rule and attribute information of the information. In the exemplary embodiment, the statistical information includes: number of users, page views, APP user registrations and rankings, and so on, the attribute information includes: information name, page URL,

both  $\alpha$  and \$\$, the corresponding row data meets the requirements of the preset character and the preset character string.

**[0116]** S340, determining that in the row data including the preset character or preset character string, whether exists row data that the second word after the preset character or the preset character string is a preset information.

**[0117]** In the exemplary embodiment, it's first to filter out row data including a preset character or a preset character string, then to determine whether the second word after the preset character or the preset character string in the row data corresponding to the log information is a preset information, or the third word or the word at another position after or before the preset character or the preset character string. The position of the word is not limited in the exemplary embodiment.

**[0118]** The preset information can be reasonably set in accordance with the storage format of user behavior data. For example, the preset information can be information of operators. If this is the case, it's to determine that in the row data including the preset character or preset character string, whether exists row data that the second word after the preset character or the preset character string is the information of operators. For example, when the preset character string is \$\$, it's to determine whether the second word after \$\$ in the log file row is the information of operators. Specifically, regular expression: `^[\w\d_-\+\.\.]*$` can be used to determine whether the second word after \$\$ is the information of operators. When in the row data including the preset character or the preset character string, row data satisfying the regular expression exists, it's to determine that in the row data including the preset character or the preset character string, row data that the second word after the preset character or the preset character string is the preset information exists.

**[0119]** S350, setting the row data that the second word after the preset character or the preset character string is the preset information as user behavior information.

**[0120]** In the exemplary embodiment, it's to determine whether the second word after the preset character or the preset character string in the row data is a preset information exists. If the second word after the preset character or the preset character string in the row data is the preset information exists, the row data is set as user behavior information.

Referring to Fig.8, in the exemplary embodiment, after the step S300, the user behavior data processing method further includes:

**[0141]** S500, obtaining attribute information of the user behavior information.

**[0142]** In the exemplary embodiment, the log information has its corresponding attribute information. The attribute information of the log information obtained after cleaning can be directly obtained on the HADOOP large data platform, wherein the attribute information includes page information URL, information name, time, information type and so on.

**[0143]** S600, displaying the attribute information and the statistical information.

**[0144]** In the exemplary embodiment, statistics on the user behavior data is performed in the database according to a preset rule, and then attribute information of the information data is obtained directly. The attribute information includes identification information and source information of the user behavior information, wherein, the identification information includes information type, information name and the like of the user behavior information, the source information includes source URL information, time and the like of the user behavior information.

**[0145]** The attribute information and the statistical information of behavior log data are directly displayed on the HADOOP large data platform. The statistical information includes data obtained by statistics such as the number of times a user clicking on the APP interface, the length of time to access to the APP interface, and the like. The direct display of the log information obtained by statistics on the HADOOP large data platform is clear.

**[0146]** In accordance with the user behavior data processing method proposed in the exemplary embodiment, attribute information of the user behavior information is obtained, then the attribute information and the statistical information are displayed. These allow the administrator of the HADOOP large data platform to view related information of the log information at a glance, so that the administrator can easily perform subsequent processing on the log information.

**[0147]** The present disclosure also provides a computer-readable storage medium. In the exemplary embodiment, the computer-readable storage medium stores a user

10. The method of claim 9, wherein the attribute information comprises identification information and source information of the user behavior information.

11. The method of claim 3, wherein, after the step of performing data statistics on the user behavior information according to a preset rule, the user behavior data processing method, the method further comprises:

- obtaining attribute information of the user behavior information;
- displaying the attribute information and the statistical information.

12. The method of claim 11, wherein the attribute information comprises identification information and source information of the user behavior information.

13. The method of claim 4, wherein, after the step of performing data statistics on the user behavior information according to a preset rule, the user behavior data processing method, the method further comprises:

- obtaining attribute information of the user behavior information;
- displaying the attribute information and the statistical information.

14. The method of claim 13, wherein the attribute information comprises identification information and source information of the user behavior information.

15. The method of claim 5, wherein, after the step of performing data statistics on the user behavior information according to a preset rule, the user behavior data processing method, the method further comprises:

- obtaining attribute information of the user behavior information;
- displaying the attribute information and the statistical information.

16. The method of claim 15, wherein the attribute information comprises identification information and source information of the user behavior information.

17. The method of claim 6, wherein, after the step of performing data statistics on the user behavior information according to a preset rule, the user behavior data processing method, the method further comprises:

- obtaining attribute information of the user behavior information;
- displaying the attribute information and the statistical information.

18. The method of claim 17, wherein the attribute information comprises identification information and source information of the user behavior information.

19. A user behavior data processing device, comprising a memory, a processor, and a user behavior data processing program executable by the processor and stored in the

memory, the user behavior data processing program when being executed by the processor performing the following steps:

- obtaining log information stored in a preset server at a fixed time, wherein, the log information comprises user behavior data which is captured and sent to the preset server by client;

- determining that whether row data satisfying a preset format exists in each row data of the log information, wherein, the row data is the data corresponding to each row of the log information;

- setting the row data satisfying the preset format as user behavior information;

- performing data statistics on the user behavior information according to a preset rule to obtain statistical information of the user behavior information.

20. A computer-readable storage medium storing a user behavior data processing program, the user behavior data processing program when being executed by a processor performing the following steps:

- obtaining log information stored in a preset server at a fixed time, wherein, the log information comprises user behavior data which is captured and sent to the preset server by client;

- determining that whether row data satisfying a preset format exists in each row data of the log information, wherein, the row data is the data corresponding to each row of the log information;

- setting the row data satisfying the preset format as user behavior information;

- performing data statistics on the user behavior information according to a preset rule to obtain statistical information of the user behavior information.