DATA QUERY APPARATUS, METHOD, AND NON-TRANSITORY TANGIBLE MACHINE-READABLE MEDIUM THEREOF

This application claims priority to China Patent Application No. 201910514605.5 filed on June 14, 2019, which is hereby incorporated by reference in its entirety.

CROSS-REFERENCES TO RELATED APPLICATIONS

Not applicable.

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to a data query apparatus, method, and non-transitory tangible machine-readable medium thereof. Specifically, the present invention relates to an apparatus and method capable of querying time series data and a non-transitory tangible machine-readable medium thereof.

Descriptions of the Related Art

Industrial data usually has time information (e.g., each sales data in the retail industry corresponds to the sales time of a product). If these data and the corresponding time information can be stored and further analyzed, the development tendency of the industry in the past can be learned and, moreover, the future development of the industry can be predicted.

According to the prior art, data along with their time information are stored in one or more tables of a database and a field is set in one or more of these tables to record the time information. Specifically, in the prior art, each time a piece of data is received by the database, that piece of data is appended behind the existing pieces of data in the database. Since the data often comes from different locations, the chronological order regarding the pieces of data received by the database (i.e., the order that the pieces of data are stored in the database) is locally inconsistent with the chronological order regarding the events corresponding to the pieces of data under the influence of various factors (e.g., network speed, system performance). That is, some events occur later than other events, but their pieces of data are received by the database earlier. As a consequence, the time information in the database is not ordered perfectly/completely (i.e. the time information of these data are arranged in order in terms of a coarser time precision, but the time information of some of these data are not arranged in order in terms of a finer time precision).

In order to let users query data of a certain period of time in the past and even to aggregate the data within that period of time, conventionally technology creates indexes for the time information of these data (e.g., creates the indexes by using the binary tree). However, this type of conventional technology has drawbacks. First, creation of indexes becomes extremely time-consuming when the amount of data is quite huge. Second, the indexes have to be re-created each time a piece of data is added or deleted, which is extremely inconvenient and, moreover, makes the indexes generated each time have poor benefit. Third, since the time information in the database is arranged in order completely, creating the indexes by the

technique of a binary tree will derive an extremely unbalanced binary tree. As a result, queries become inefficient.

When it comes to data with a high data injection rate, the aforesaid drawbacks become more obvious and make a system crush. Accordingly, a technology that can query time series data efficiently and that will not crush a system is in an urgent need.

SUMMARY OF THE INVENTION

To solve the problems described in the related art, the present invention provides a data query apparatus, method, and non-transitory tangible machine-readable medium thereof that are capable of querying a time series fact dataset.

The data query apparatus provided by the present invention comprises a storage, a processor, and a transceiving interface, wherein the processor is electrically connected to the storage and the transceiving interface. The storage is configured to store a plurality of first historical data, wherein each of the first historical data corresponds to a first transmission time length. The processor is configured to calculate a first extension time length according to the first transmission time lengths. The transceiving interface is configured to receive a query command, wherein the query command records a preset query time interval. The processor is further configured to extend the preset query time interval to an extended query time interval according to the first extension time length and query a fact dataset according to the extended query time interval.

Preferably, the processor extends the preset query time interval to the extended query time

interval by setting a start time of the extended query time interval to be the first extension time length earlier than a start time of the preset query time interval and setting an end time of the extended query time interval to be the first extension time length later than an end time of the preset query time interval.

Preferably, the processor calculates an expected time length by inputting the first transmission time lengths to a Poisson process and calculates the first extension time length according to the expected time length and an expected precision.

Preferably, the storage is further configured to store a plurality of second historical data, wherein each of the second historical data corresponds to a second transmission time length. The processor further calculates a second extension time length according to the second transmission time lengths. The processor further determines that the preset query time interval corresponds to the first extension time length. The preset query time interval is extended to the extended query time interval by the processor according to the first extension time length after determining that the preset query time interval corresponds to the first extension time length.

The data query method provided by the present invention is adapted for use in an electronic computing apparatus. The electronic computing apparatus stores a plurality of historical data, wherein each of the historical data corresponds to a transmission time length. The data query method comprises steps (a) to (d). The step (a) calculates an extension time length according to the transmission time lengths. The step (b) receives a query command, wherein the query command records a preset query time interval. The step (c) extends the preset query time

interval to an extended query time interval according to the extension time length. The step (d) queries a fact dataset according to the extended query time interval.

The non-transitory tangible machine-readable medium provided by the present invention stores a computer program, wherein the computer program comprises a plurality of codes. When the codes of the computer program are loaded into an electronic computing apparatus, the codes executed by the electronic computing apparatus performs the data query method described in the above paragraph. The electronic computing apparatus stores a plurality of pieces of historical data, wherein each of the historical data corresponds to a transmission time length.

The data query technology provided by the present invention (including the apparatus, the method, and the non-transitory tangible machine-readable medium at least) broadens a preset query time interval to an extended query time interval based on the transmission time length of past historical data. That is, the data query technology provided by the present invention evaluates the network transmission time of the historical data. Therefore, even if the historical data in the fact dataset is not arranged in order completely (i.e. although some historical data corresponds to an earlier event time than some of the others, these historical data is arranged in the later row(s) in the fact dataset comparing to them (i.e. the aforesaid some of the others)), most of the disordered historical data can be retrieved by broadening the preset query time interval to the extended query time interval. By the aforesaid operations, the data query technology provided by the present invention can efficiently and accurately respond to the query of the user and can be applied to an industry with a high data injection rate.

The detailed technology and preferred embodiments implemented for the subject invention are described in the following paragraphs accompanying the appended drawings for people skilled in this field to well appreciate the features of the claimed invention.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A depicts a schematic view of a data query apparatus **1** according to a first embodiment;

FIG. 1B depicts a specific example of a fact dataset **12**; and

FIG. 2 depicts a flowchart of a data query method according to a second embodiment.

DESCRIPTION OF THE PREFERRED EMBODIMENT

In the following description, a data query apparatus, method, and non-transitory tangible machine-readable medium thereof provided by the present invention will be explained with reference to embodiments thereof. However, these embodiments of the present invention are not intended to limit the present invention to any environment, applications, or implementations described in these embodiments. Therefore, description of these embodiments is only for purpose of illustration rather than to limit the scope of the present invention. It shall be appreciated that, in the following embodiments and the attached drawings, elements unrelated to the present invention are omitted from depiction. In addition, dimensions of elements and dimensional proportions among individual elements in the attached drawings are provided only for illustration, but not to limit the scope of the present invention.

A first embodiment of the present invention is a data query apparatus **1** and a schematic view of which is depicted in **FIG. 1A**. The data management apparatus **1** comprises a storage **11**, a processor **13**, and a transceiving interface **15**, wherein the processor **13** is electrically connected to the storage **11** and the transceiving interface **15**. The storage **11** may be a hard disk (e.g., a solid-state hard disk) or other non-transitory storage medium or apparatus that can be installed with various kinds of data storage systems and well known to those skilled in the art. The processor **13** may be one of various processors, central processing units (CPUs), microprocessors, digital signal processors (DSPs), or any other computing apparatuses that have the same function and well known to those skilled in the art. The transceiving interface **15** may be any interface (e.g., a universal serial bus (USB) interface, a network interface, without being limited thereto) that is capable of receiving data inputted by the user from a near-end (e.g., a keyboard) of the data query apparatus **1** or capable of receiving the data remotely (e.g., over the network).

The storage **11** stores a fact dataset **12**, wherein the fact dataset **12** comprises a plurality of historical data and each of the historical data has an event time. For example, if the historical data comprised in the fact dataset **12** is sales data, the event time of each historical data can be the sales time of a certain product. In this embodiment, the fact dataset **12** is continuously updated by a stream of data (not shown). For example, the transceiving interface **15** or another transceiving interface receives the stream of data continuously and then the processor **13** updates the fact dataset **12** with the fact data in the stream of data (i.e., adding new fact records into the fact dataset **12** to cover the new fact data). The aforesaid stream of data may be data

generated by an Internet of Things (IoT) system, data generated by a mobile application, or data generated by a large-scale Web service, but it is not limited thereto.

In this embodiment, the processor **13** sequentially stores the historical data in the fact dataset **12** according to the time at which the historical data is received by the data query apparatus **1**. Since the sources of the historical data are not all the same (e.g., the historical data comes from different points of sale), some historical data will be arranged in the later row(s) in the fact dataset **12** even their corresponding event times are earlier than the event time of some other historical data under the influence of various factors (e.g., network speed, system performance). For better understanding, please refer to a specific example shown in **FIG. 1B**. Please note that this specific example is not intended to limit the scope of the present invention. In this specific example, the event time corresponding to the historical data with a row number of "150003" is earlier than the event time corresponding to the historical data with a row number of "150002." but the historical data with the row number of "150003" is arranged after the historical data with the row number of "150002" in the fact dataset **12**.

In order to let users query the historical data of a certain time interval in the past in the fact dataset **12**, the data query apparatus **1** does not create index for the event time of the historical data. Instead, the data query apparatus **1** slightly broadens the time interval that is queried by the user based on the transmission time length of the historical data over the network and then sequentially searches whether the historical data in the fact dataset **12** falls into the query time interval that is broadened. In the following description, the specific operations and the principle of the data query apparatus **1** will be described in detail.

For each historical data in the fact dataset **12**, the processor **13** calculates a transmission time length according to its event time and the time that the data query apparatus **1** receives the historical data (for example, the data query apparatus **1** may derive the transmission time length by subtracting the event time of the historical data from the time that the data query apparatus **1** receives the historical data). It means that the historical data is received by the data query apparatus **1** after travelling over the network for the transmission time length. In this embodiment, the processor **13** further records the transmission time length corresponding to each historical data in the fact dataset **12** as shown in **FIG. 1B**. In some other embodiments, the transmission time length corresponding to each historical data is not required to be recorded in the fact data set **12**.

In this embodiment, the processor **13** divides a day into two periods, including a transmission peak period (for example, from 9 a.m. to 5 p.m.) and a transmission off-peak period (that is, a non-transmission peak period). The processor **13** calculates an extension time length (not shown) for each of the two periods individually. Specifically, the processor **13** divides the historical data comprised in the fact dataset **12** into two groups according to the two periods, wherein the first group comprises the historical data whose corresponding event time falls into the transmission peak period and the second group comprises the historical data whose corresponding event time falls into the transmission off-peak period. For the first group, the processor **13** calculates an extension time length (not shown) according to the transmission time length corresponding to the historical data comprised in the first group. Similarly, for the second group, the processor **13** calculates an extension time length (not shown) according to

the transmission time length corresponding to the historical data comprised in the second group.

Furthermore, the transmission time lengths corresponding to the historical data comprised in each group conforms to the Poisson Process. Now, the parameter N(t) is used to represent the number of historical data received after 0 to t seconds has elapsed (that is, the number of historical data whose transmission time length is of 0 to t seconds). If N(t) conforms to the Poisson process, the event of receiving the historical data at the time point X_i (where the variable *i* may be any positive integer greater than 1) is independent and the variable X_i is of the exponential distribution with a parameter of λ. In this way, the probability of not receiving the historical data after 0 to t seconds is $\mathrm{P}(X_i > t) = 1 - e^{-\lambda t}$.

Thus, in some embodiments, the processor **13** may calculate the extension time length of each group by using the formula of the Poisson process. Specifically, for the first group, the processor **13** calculates an expected time length (not shown) by inputting the transmission time lengths corresponding to the historical data comprised in the first group to the formula of the Poisson process, that is, calculates an expected value of the transmission time length (i.e., the aforesaid parameter λ). Next, the processor **13** calculates the extension time length (not shown) of the first group based on the expected time length and an expected precision. Specifically, the processor **13** calculates the extension time length by using the following formula (1):

$$1 - \mathrm{P}(X_i > t) = 1 - e^{-\lambda t} = N\% \qquad (1)$$

In the above formula (1), *N*% represents the expected precision and the parameter t represents the transmission time length under the expected precision of N%, that is, the extension time length to be calculated. Similarly, for the second group, the processor **13**

calculates an expected time length (not shown) by inputting the transmission time lengths corresponding to the historical data comprised in the second group to the formula of the Poisson process, that is, calculates an expected value of the transmission time length. Next, the processor **13** uses the above formula (1) to calculate the extension time length (not shown) of the second group according to the expected time length and an expected precision.

It shall be noted that the expected precision used by the processor **13** for calculating the extension time length of each group may be a default value or may be a value inputted by a user that is received through the transceiving interface **15** or another transceiving interface.

Thereafter, the user may input a query command **Q** to query the historical data of a previous time interval, wherein the query command **Q** records the date range to be queried (for example, March of last year) and a preset query time interval (for example, 9 a.m. to 12 a.m.). The query command **Q** inputted by the user is received by the transceiving interface **15**. Next, the processor **13** determines which extension time length the preset query time interval recorded in the query command corresponds to. Specifically, the processor **13** determines the preset query time interval recorded in the query command **Q** corresponds to which group and then takes the extension time interval of that group as the extension time length corresponding to the preset query time interval.

Next, the processor **13** extends the preset query time interval to an extended query time interval according to the extension time length corresponding to the preset query time interval. In some embodiments, the processor **13** extends the preset query time interval to the extended query time interval by setting a start time of the extended query time interval to be the first

extension time length earlier than a start time of the preset query time interval and setting an end time of the extended query time interval to be the first extension time length later than an end time of the preset query time interval. Thereafter, the processor **13** queries the fact dataset **12** according to the extended query time interval.

Please note that, in this embodiment, the processor **13** calculates the extension time lengths of different groups according to the transmission time lengths of the historical data stored in the fact dataset **12** and the object queried by the user later is still the fact dataset **12** (however, the fact dataset **12** may have been updated by the new fact data when it is queried by the user). In other embodiments, the processor **13** may calculate the extension time lengths of different groups according to the transmission time lengths of the historical data stored in other fact datasets, but the object queried by the user later is the fact dataset **12**.

It shall be additionally noted that, in this embodiment, the processor **13** divides a day into two periods, divides the historical data stored in the fact dataset **12** into two groups according to the two periods, and then calculates the extension time lengths for different groups. In other embodiments, the processor **13** may divide a day into multiple periods, divide the historical data stored in the fact dataset **12** into multiple groups according to the multiple periods, and then calculate an extension time length for each group. Regarding the additional operations that the processor **13** needs to perform when a day is divided into multiple periods shall be appreciated by those skilled in the art based on the aforesaid content. Thus, the details will not be further described.

In addition, in some embodiments, the processor **13** does not divide a day into two or more

periods. Instead, the processor **13** directly calculate an extension time length according to the transmission time lengths of the historical data stored in the fact dataset **12**. For those embodiments, when the user queries the fact dataset **12**, the processor **13** will not determine which extension time length does the preset query time interval recorded in the query command inputted by the user corresponds to. The processor **13** directly extends the preset query time interval by the extension time length.

According to the above description, the data query apparatus **1** stores the historical data in the fact dataset **12** sequentially based on the time when the historical data is received. In order to let users query the historical data of a certain time interval in the past in the fact dataset **12**, the data query apparatus **1** broadens the preset query time interval inputted by the user to an extended query time interval according to the transmission time lengths of the historical data over the network and then sequentially searches the historical data in fact dataset **12** for those fall into the extended query time interval.

The data query apparatus **1** broadens the preset query time interval to the extended query time interval based on the transmission time lengths of the historical data, which means that the data query apparatus **1** has evaluated the transmission time of the historical data over the network. Therefore, even if the historical data in the fact dataset **12** is not arranged in order completely (that is, although some historical data corresponds to an earlier event time comparing to some of the others, these historical data is arranged in the later row(s) in the fact dataset **12** comparing to them (i.e. the aforesaid some of the others)), most of the disordered historical data can be retrieved by broadening the preset query time interval inputted by the user

to the extended query time interval.

If the data query apparatus **1** calculates the extension time length based on the Poisson process, the precision of the derived historical data by query is the aforesaid expected precision (i.e., the aforesaid N%). By the aforesaid operation, the data query apparatus **1** can efficiently and accurately respond to the query of the user and can be applied to an industry with a high data injection rate.

A second embodiment of the present invention is a data query method and a flowchart of which is depicted in **FIG. 2**. The data query method is adapted for use in an electronic computing apparatus, e.g., the data query apparatus **1** of the first embodiment. The electronic computing apparatus stores a plurality of historical data, wherein each of the historical data corresponds to a transmission time length. The data query method executes steps **S201** to **S207**.

In the step **S201**, the electronic computing apparatus calculates an extension time length according to the transmission time lengths. In some embodiments, the step **S201** calculates an expected time length by inputting the transmission time lengths to a Poisson process and calculate the extension time length according to the expected time length and an expected precision.

In the step **S203**, the electronic computing apparatus receives a query command, wherein the query command records a preset query time interval.

In the step **S205**, the electronic computing apparatus extends the preset query time interval to an extended query time interval according to the extension time length. In some

embodiments, the step **S205** extends the preset query time interval to the extended query time interval by setting a start time of the extended query time interval to be the extension time length earlier than a start time of the preset query time interval and setting an end time of the extended query time interval to be the extension time length later than an end time of the preset query time interval.

In step **S207**, the electronic computing apparatus queries a fact dataset according to the extended query time interval.

In some embodiments, the data query method may divide a day into a plurality of periods, divide the historical data into a plurality of groups according to the periods, and execute the step **S201** for each of the groups. For those embodiments, the data query method further executes a step after executing the step **S203** to determine to which group does the preset query time interval recorded in the query command corresponds to and then the step **S205** extends the preset query time interval to the extended query time interval according to the extension time length of that corresponding group.

In addition to the aforesaid steps, the second embodiment can execute all the operations and steps of the data query apparatus **1** set forth in the first embodiment, have the same functions, and deliver the same technical effects as the first embodiment. How the second embodiment executes these operations and steps, has the same functions, and delivers the same technical effects as the first embodiment will be readily appreciated by those skilled in the art based on the explanation of the first embodiment and, thus, will not be further described herein.

The data query method described in the second embodiment may be implemented as a

computer program having a plurality of codes. The computer program is stored in a non-transitory tangible machine-readable medium, which may be a read only memory (ROM), a flash memory, a floppy disk, a hard disk, a compact disk (CD), a digital versatile disc (DVD), a mobile disk, a database accessible to networks, or any other storage media with the same function and well known to those of ordinary skill in the art. After the computer program is loaded into an electronic computing apparatus, the electronic computing apparatus executes the codes to perform the data query method described in the second embodiment.

According to the above embodiments, the data query technology provided by the present invention (including the apparatus, the method, and the non-transitory tangible machine-readable medium at least) broadens the preset query time interval to an extended query time interval based on the transmission time lengths of past historical data. That is, the data query technology provided by the present invention evaluates the transmission times of the historical data over the network. Therefore, even if the historical data in the fact dataset is not arranged in order completely (that is, although some historical corresponds to an earlier event time comparing to some of the others, these historical data is arranged in the later rows in the fact dataset comparing to them (i.e. the aforesaid some of the others)), most of the disordered historical data can be retrieved by broadening the preset query time interval inputted by the user to the extended query time interval. By the aforesaid operation, the data query technology provided by the present invention can efficiently and accurately respond to the query of the user and can be applied to an industry with a high data injection rate.

The above disclosure is related to the detailed technical contents and inventive features

thereof. People skilled in this field may proceed with a variety of modifications and replacements based on the disclosures and suggestions of the invention as described without departing from the characteristics thereof. Nevertheless, although such modifications and replacements are not fully disclosed in the above descriptions, they have substantially been covered in the following claims as appended.

What is claimed is:

1. A data query apparatus, comprising:

a storage, being configured to store a plurality of first historical data, wherein each of the first historical data corresponds to a first transmission time length;

a processor, being electrically connected with the storage and configured to calculate a first extension time length according to the first transmission time lengths; and

a transceiving interface, being electrically connected to the processor and configured to receive a query command, wherein the query command records a preset query time interval,

wherein the processor is further configured to extend the preset query time interval to an extended query time interval according to the first extension time length and query a fact dataset according to the extended query time interval.

2. The data query apparatus of Claim 1, wherein the processor extends the preset query time interval to the extended query time interval by setting a start time of the extended query time interval to be the first extension time length earlier than a start time of the preset query time interval and setting an end time of the extended query time interval to be the first extension time length later than an end time of the preset query time interval.

3. The data query apparatus of Claim 1, wherein the processor calculates an expected time length by inputting the first transmission time lengths to a Poisson process and calculates

the first extension time length according to the expected time length and an expected precision.

4. The data query apparatus of Claim 1, wherein the storage is further configured to store a plurality of second historical data, each of the second historical data corresponds to a second transmission time length, the processor further calculates a second extension time length according to the second transmission time lengths, the processor further determines that the preset query time interval corresponds to the first extension time length, wherein the preset query time interval is extended to the extended query time interval by the processor according to the first extension time length after determining that the preset query time interval corresponds to the first extension time length.

5. A data query method, being adapted for use in an electronic computing apparatus, the electronic computing apparatus storing a plurality of first historical data, each of the first historical data corresponding to a first transmission time length, the data query method comprising the following steps of:

(a) calculating a first extension time length according to the first transmission time lengths;

(b) receiving a query command, wherein the query command records a preset query time interval;

(c) extending the preset query time interval to an extended query time interval according to the first extension time length; and

(d) querying a fact dataset according to the extended query time interval.

6. The data query method of Claim 5, wherein the step (c) extends the preset query time interval to the extended query time interval according to the first extension time length by setting a start time of the extended query time interval to be the first extension time length earlier than a start time of the preset query time interval and setting an end time of the extended query time interval to be the first extension time length later than an end time of the preset query time interval.

7. The data query method of Claim 5, wherein the step (a) calculates an expected time length by inputting the first transmission time lengths to a Poisson process and calculates the first extension time length according to the expected time length and an expected precision.

8. The data query method of Claim 5, wherein the electronic calculating apparatus further stores a plurality of second historical data, each of the second historical data corresponds to a second transmission time length, and the data query method further comprises the following steps of:

(e) calculating a second extension time length according to the second transmission time lengths; and

(f) determining that the preset query time interval corresponds to the first extension time length;

wherein the step (c) is executed after the step (f).

9. A non-transitory tangible machine-readable medium, storing a computer program

comprising a plurality of codes, the codes being able to perform a data query method when the computer program is loaded into an electronic computing apparatus, the electronic computing apparatus storing a plurality of first historical data, each of the first historical data corresponding to a first transmission time length, the data query method comprising the following steps:

(a) calculating a first extension time length according to the first transmission time lengths;

(b) receiving a query command, wherein the query command records a preset query time interval;

(c) extending the preset query time interval to an extended query time interval according to the first extension time length; and

(d) querying a fact dataset according to the extended query time interval.

10. The non-transitory tangible machine-readable medium of Claim 9, wherein the step (c) extends the preset query time interval to the extended query time interval according to the first extension time length by setting a start time of the extended query time interval to be the first extension time length earlier than a start time of the preset query time interval and setting an end time of the extended query time interval to be the first extension time length later than an end time of the preset query time interval.

11. The non-transitory tangible machine-readable medium of Claim 9, wherein the step (a) calculates an expected time length by inputting the first transmission time lengths into a Poisson process and calculates the first extension time length according to the expected time

length and an expected precision.

12. The non-transitory tangible machine-readable medium of Claim 9, wherein the electronic calculating apparatus further stores a plurality of second historical data, each of the second historical data corresponds to a second transmission time length, and the data query method further comprises the following steps of:

(e) calculating a second extension time length according to the second transmission time lengths; and

(f) determining that the preset query time interval corresponds to the first extension time length;

wherein the step (c) is executed after the step (f).