SYSTEM AND METHOD FOR CUSTOMER-SPECIFIC DATA CLEANSING

FIELD OF THE INVENTION

[0001]

The present invention relates to a system and method for customer-specific data cleansing.

[0002]

Specifically, the present invention relates to a customer-specific data cleansing system and method which are suitably applied to a data providing system which can gather various types of data provided by multiple data providers, such as multiple service providing companies, perform compilation, analysis, data mining, or the like on the gathered data, and provide the resulting data to data receivers.

BACKGROUND ART

[0003]

With the development of the Internet, there have been increased the number of various types of service providing companies that provide services useful for the categories or business types of companies. Examples of such service providing companies include service providing companies that provide merchandise information, such as popular menu items, to the restaurant industry, service providing companies and

retail service providing companies that provide food material or the like to such service providing companies, and producers of food materials or the like.

[0004]

For example, there have existed service providing systems in recent years where through the internet or Web, service providing companies that provide particular business services to service receiving companies, such as restaurants, are provided with an infrastructure of the particular business services and then support the business of the service receiving companies by providing business services based on the service infrastructure.

[0005]

Such a service providing system includes a service infrastructure company device, service providing company devices, service receiving company devices, and user terminal devices.

[0006]

The service infrastructure company device is a system or device that provides functions (services) based on the service infrastructure to be provided, for example, a SaaS (Software as a Service)-type operation system, to the service providing company devices.

[0007]

A service providing company device is a system or device

that receives an operation system (business SaaS) based on the service infrastructure provided by the service infrastructure company device. For example, the service providing company device customizes the operation system in accordance with a request from a service receiving company (or directly from a user) and provides the customized operation system to the service receiving company.

[0008]

A service receiving company device is a system or device which receives services provided by service providing companies. For example, a service receiving company device is a device that provides user information provided by users to service providing company devices.

[0009]

A user terminal device is, for example, a terminal which itself acquires user information and provides the user information to a service receiving company device. Further, a user terminal device may be a device which can automatically gather information from the mobile terminal of the user when the mobile terminal is held over the device.

Original user information possessed by a service receiving company or service providing company includes useful information. If the service infrastructure company can receive and analyze such useful original user information, it can provide more suitable business support.

[0010]

Related technologies in this technical field include Japanese Patent Application Laid-Open Publication No. 2009-93569 (Patent Document 1), Japanese Patent Application Laid-Open Publication No. H09-153094 (Patent Document 2), and Japanese Patent Application Laid-Open Publication No. 2012-141847 (Patent Document 3).

[0011]

When changing an SaaS environment to an SI environment, the technology of Patent Document 1 reduces occurrence of sunk cost or opportunity loss in a business service whose use has been started in SaaS form. Specifically, when transferring business services from an SaaS provider environment (company A) to an SI-form environment (city B), the technology provides the city B with a service address list for recording the addresses of the business services and, each time a business service is transferred, updates an address in the service address list. Further, when transferring business data stored in the company A, the technology provides the company A and the city B with data address lists for recording the addresses of pieces of business data and, each time a piece of business data is transferred, updates addresses in both data address lists. With respect to the update of the addresses, there is disclosed a technology which changes the access method so that the addresses logically become the same.

[0012]

Patent Document 2 discloses a method for analyzing the effects of merchandise sales promotion from information about customers, merchandise information that the customers have referred to when considering the purchase of merchandise, and information about merchandise purchased by the customers; and a system for performing the method.

[0013]

Patent Document 3 describes a data transfer system. The data transfer system includes a definition file generation unit configured to generate a definition file for defining processing conditions used in subsequent processes and to store the definition file in a storage region, a checking unit configured to check the validity of data in a source DB using the definition file read from the storage region before or after data cleansing and to output the check result, a verification unit configured to verify the validity of data loaded into a destination DB using the definition file read from the storage region and to output the verification result, and a screen processing unit configured to display the check result and the verification result on a display device.

CITATION LIST

Patent Document

[0014]

Patent Document 1: Japanese Patent Application Laid-Open Publication No. 2009-93569

Patent Document 2: Japanese Patent Application Laid-Open Publication No. H09-153094

Patent Document 3: Japanese Patent Application Laid-Open Publication No. 2012-141847

SUMMARY OF THE INVENTION

PROBLEMS TO BE SOLVED BY THE INVENTION

[0015]

According to a service providing system as described in Patent Document 1, a service infrastructure company can support the business of service providing companies. Each service providing company can support the business of service receiving companies. Such business support can increase the business efficiency of the entire system.

[0016]

The service providing companies and service infrastructure company, which provide such services, typically hold (store) and manage data about users, for example, data such as preferences or behavior histories of the users.

[0017]

Accordingly, if there is a system where a service infrastructure company can analyze the data about users and then service providing companies can provide useful

information or original data based on the analysis result to service receiving companies (stores, general companies, or the like), the business efficiency can be further increased.

[0018]

In the present invention, “a user” refers to an individual, organization, or corporation that uses the present system and includes a specified user (a user which can be identified as a single individual, single organization, or single corporation) and an unspecified user (a user which cannot be identified as any individual, organization, or corporation or is difficult to identify as such an entity). A terminal or the like used by a user in the present system will be referred to as “a user terminal.” In the present invention, the term “individual” described with respect to a user includes an organization and corporation serving as users.

[0019]

“A service receiving company" refers to a company that receives an original business support service according to the present system (first service). A system component device disposed on the side of the service receiving company will be referred to as “a service receiving company device.” As used herein, “disposed on the side of” does not mean only being disposed inside the service receiving company. The term “disposed on the side of” will be used in a similar manner.

[0020]

“A service providing company" refers to a company that provides the original business support service according to the present system (first service). A system component device disposed on the side of the service providing company will be referred to as “a service providing company device.”

[0021]

“A service infrastructure company” refers to a company that organizes multiple service providing companies, and provides a service infrastructure, such as a business function (business service) or reserve function (reserve service), to the service providing companies and/or service receiving companies. The service infrastructure company gathers information from the service receiving company devices and/or service providing company devices on the basis of contracts as necessary, performs, for example, an analysis on the gathered information, and provides the analyzed information (second service) to the service receiving companies and/or service providing companies. At this time, the analyzed information need not necessarily be provided. For example, in accordance with a request from the data providing company, the original data, which has not been analyzed, may be provided. A system component device disposed on the side of the service infrastructure company will be referred to as “a service infrastructure company device.”

[0022]

"A data providing company" or "data receiver" refers to a service providing company or service receiving company. A service providing company or service receiving company transmits, for example, original user information Ia and Ib as raw data or as anonymized data XIa and XIb satisfying k-anonymity to the service infrastructure company and receives data resulting from analysis of the original user information, or the raw data from the service infrastructure company. A system component device disposed on the side of the data providing company will be referred to as a "data providing company device," and a system component device disposed on the side of the data receiving company as a "data receiver device."

While a data provider is expressed as a "data providing company and a data receiver expressed as a "data receiving company," they are not limited to "companies" and may be "individuals" serving as users. Accordingly, in an embodiment to be discussed later, they may be referred to as a "data provider" and "data receiver," respectively.

[0023]

If the service infrastructure company, which is a company organizing multiple service providing companies, can fuse, integrate, and analyze various types of data stored in the service providing companies, it can perform higher-value added analysis. Thus, the service infrastructure company can provide,

as a service, the higher-value added analysis result to users including the service providing companies and/or service receiving companies.

[0024]

Under circumstances where an IT environment which allows analysis of large-volume data is being formed, it is technically possible to perform the above high-value added analysis on large-volume data.

[0025]

"A high-value added analysis" refers to an analysis for providing basic data which allows personalization (transmission of information corresponding to preferences or attributes of each client or prospective user through emails or the like), recommendation (a service where preferences of a user are analyzed and information which the user may be interested in is selected and displayed), or the like.

[0026]

When the service infrastructure company gathers various types of data from each service providing company and performs, for example, compilation and analysis on the gathered data, it needs to cleanse the data so that the data can be subjected to compilation, analysis, or the like. The reason is that part of the data may become invalid values due to data entry errors, a failure of data processing, or the like.

[0027]

There are two main methods for performing data cleansing: One of the methods is as follows: a server provider (hereafter referred to as data provider) device 10 cleanses data (original data) which the data provider device 10 is to provide and then provides the cleansed data to a data receiver (hereafter referred to as data receiver) device 20. In this method, the data provider and the data receiver are required to previously make an agreement on according to which of the data cleansing policies of the data provider and the data receiver the data should be cleansed.

[0028]

The other method is as follows: the data receiver device 20 receives original data from the data provider device 10 and then cleanses the original data. In this method, (1) the data receiver device 20 has difficulty in determining which of a normal value and an invalid value the original data is. Further, (2) it is necessary to make a determination on which of the data cleansing policies of the data provider device 10 and the data receiver device 20 should be complied with. The data provider and the data receiver are required to previously make agreements on (1) and (2).

[0029]

With regard to (1), the data provider, which is the owner of the original data, has the best knowledge about which of a normal value and an invalid value (a value to be subjected to

data cleansing) the original data is.

[0030]

With regard to (2), it is the data receiver that uses the original data provided by the data provider. Accordingly, it is preferred to cleanse the original data according to the data cleansing policy of the data receiver, which uses the data.

[0031]

As seen above, when the data provider and the data receiver are one-on-one, the two parties can easily make an agreement on the method by which the data provider or data receiver performs data cleansing.

[0032]

However, no consideration is given to a method for applying data cleansing to a system including multiple data providers having different data cleansing policies and multiple data receivers having different data cleansing policies, for example, a system including N number of data providers and M number of data receivers.

[0033]

That is, when data cleansing is applied to a system including increased numbers of data providers and data receivers, such as a system including N number of data providers and M number of data receivers, the system configuration becomes complicated, causing the following

problems.

[0034]

When the N number of data providers process values to be subjected to data cleansing, each data provider cleanses the values according to the data cleansing policies of the M number of data receivers. Accordingly, the N number of data providers must each perform data cleansing M times.

[0035]

In contrast, when the M number of data receivers perform data cleansing, each data receiver acquires original data from the N number of data providers and simultaneously receives values to be subjected to data cleansing (invalid value) from the N number of data providers. Accordingly, the M number of data receivers must each perform data cleansing N times.

[0036]

That is, there is a problem that when the system including the N number of data providers and the M number of data receivers requires a device for performing data cleansing efficiently.

[0037]

The above Patent Document 1 describes the data transfer system which includes a data cleansing unit for performing data cleansing on a source DB on the basis of an instruction from a client and which allows the user to determine the validity of data in a destination system. However, Patent

Document 1 does not consider performing data cleansing systematically in a system which receives and gathers data from multiple data providers and provides the cleansed data to multiple data receivers. The same applies to Patent Documents 2 and 3.

[0038]

In view of the foregoing, the present invention provides a customer-specific data cleansing system and method which can perform data cleansing systematically even in systems including increased numbers of data providers and data receivers, such as a system including N number of data providers and M number of data receivers.

[0039]

To address the above problem, the present invention provides a configuration where an infrastructure company device includes customer-specific data cleansing means (data cleansing server) configured to perform data cleansing on the basis of a data cleansing policy master and data cleansing policies corresponding to the types of invalid values of original data provided by a data provider device of the master and where a data receiver device can set the data cleansing policy.

[0040]

For example, the present invention provides a customer-specific data cleansing system serving as a data providing

system. The system includes: at least one service infrastructure company device; multiple data provider devices; and multiple data receiver devices. The service infrastructure company device includes: input/output means; storage means; and control means. The input/output means is configured to provide or receive data to or from the data provider devices and the data receiver devices. The storage means includes a data cleansing policy master with respect to data provided by the data provider devices. The control means has a function of cleansing the data in accordance with a data cleansing policy of the data cleansing policy master. The control means includes: a receiving unit configured to receive data provided by the data provider devices and field attribute information of the data; a data cleansing policy receiving unit configured to receive the data cleansing policy; data cleansing means configured to cleanse data provided by the data provider devices in accordance with the data cleansing policy; and means configured to transmit the data cleansed by the data cleansing means to the data receiver devices.

[0041]

The control means further includes means configured to receive default data cleansing policies of the data receiver devices, data receiver data cleansing policies for data provided by the data provider devices, default data cleansing policies of the data provider devices, and data provider data

cleansing policies for data provided by the data provider devices and to incorporate the received policies into the data cleansing policy master.

[0042]

The control means further includes means including a field attribute master, an invalid value master, and a processing method master and configured to generate a data cleansing policy of the service infrastructure company device from the field attribute master, the invalid value master, and the processing method master. The means configured to incorporate into the data cleansing policy master incorporates the data cleansing policy of the service infrastructure company device into the data cleansing policy master.

[0043]

The data cleansing policy master includes original data field names of the data provided by the data provider devices and processing methods corresponding to invalid values which may occur in field attributes of the field names. The invalid values include one or more of null, double-byte/single-byte inconsistency, invalid character, abnormal value, hiragana character inconsistency, katakana character inconsistency, alphabetical character inconsistency, value inconsistency, symbol inconsistency, kanji character inconsistency, phone number denormalization, zip code denormalization, denormalized postal address, and denormalized amount. The field attributes

includes one or more of primary key, name, zip code, postal address, quantity, amount, and date. The processing methods include one or more of doing nothing, deleting all records, complying with a data provider format, complying with a normalized data provider, making a null, and replacing with an average value.

[0044]

The present invention also provides a customer-specific data cleansing system for cleansing data shared by multiple companies, the system including an infrastructure company device configured to receive original data provided by a device for a data provider serving as a company and to provide the original data to a data receiver device. The infrastructure company device includes: a data cleansing control device configured to cleanse the original data provided by the device for the data provider; an invalid value master defining invalid values which may occur in fields of the original data; and a data cleansing policy master including processing methods corresponding to the invalid values which may occur in the fields of the original data, the device for the data provider includes means configured to, when a value in the original data is an invalid value, determine to which of items in the invalid value master the invalid value corresponds. The data receiver device includes means configured to set a processing policy corresponding to

invalid values with respect to field attributes which are comprehensively prepared for the original data. The data cleansing control device includes: reading means configured to read the processing policy set by the data receiver device from the data cleansing policy master with respect to field attribute information of the original data set by the device for the data provider; and control means configured to cleanse the original data in accordance with the read data cleansing policy.

[0045]

The present invention also provides a customer-specific data cleansing method, the method being performed by a data providing system including at least one service infrastructure company device; multiple data provider devices; and multiple data receiver devices. The service infrastructure company device includes: input/output means; storage means; and control means, the input/output means being configured to provide or receive data to or from the data provider devices and the data receiver devices, the storage means including a data cleansing policy master with respect to data provided by the data provider devices, the control means having a function of cleansing the data in accordance with a data cleansing policy of the data cleansing policy master. The method includes: receiving, by the control means, data provided by the data provider devices and field attribute information of

the data; receiving, by the control means, a data cleansing policy of the data cleansing policy master; cleansing, by the control means, data provided by the data provider devices in accordance with the data cleansing policy; and transmitting, by the control means, the data cleansed by the data cleansing means to the data receiver devices.

[0046]

The method further includes receiving, by the control means, default data cleansing policies of the data receiver devices, data receiver data cleansing policies for data provided by the data provider devices, default data cleansing policies of the data provider devices, and data provider data cleansing policies for data provided by the data provider devices and incorporating the received policies into the data cleansing policy master.

[0047]

The control means includes a field attribute master, an invalid value master, and a processing method master, and the method further includes generating, by the control means, a data cleansing policy of the service infrastructure company device from the field attribute master, the invalid value master, and the processing method master; and incorporating, by the means configured to incorporate into the data cleansing policy master, the data cleansing policy of the service infrastructure company device into the data cleansing policy

master.

[0048]

The data cleansing policy master includes original data field names of the data provided by the data provider devices, field attributes corresponding to invalid values which may occur in field attributes of the field names, and processing methods corresponding to invalid values which may occur in field attributes of the field names. The invalid values include one or more of null, double-byte/single-byte inconsistency, invalid character, abnormal value, hiragana character inconsistency, katakana character inconsistency, alphabetical character inconsistency, value inconsistency, symbol inconsistency, kanji character inconsistency, phone number denormalization, zip code denormalization, denormalized postal address, and denormalized amount. The field attributes include one or more of primary key, name, zip code, postal address, quantity, amount, and date. The processing methods include one or more of doing nothing, deleting all records, complying with a data provider format, complying with a normalized data provider, making a null, and replacing with an average value.

EFFECT OF THE INVENTION

[0049]

According to the present invention, it is possible to

provide a customer-specific data cleansing system and method which can perform data cleansing systematically even in systems including increased numbers of data providers and data receivers, such as a system including N number of data providers and M number of data receivers.

[0050]

Further, since the infrastructure company device holds the "data cleansing policy master" and centrally performs data cleansing, the data provider only has to specify an "invalid value" in the "original data" once, and the data receiver only has to set a "data cleansing policy" with respect to the "invalid value" once. Furthermore, it is possible to make a change or addition to the data cleansing policy specified by the data provider or infrastructure company device.

[0051]

Problems, configurations, and effects other than those described above will be clarified in the description of the following embodiment.

BRIEF DESCRIPTION OF THE DRAWINGS

[0052]

Fig. 1 is a diagram showing an outline of a service providing system which employs a customer-specific data cleansing system of the present invention;

Fig. 2 is a block diagram showing the configuration of a

service infrastructure providing company server of Fig. 1 and the relationship between the service infrastructure providing company server, and a data providing company server and a data receiver server;

Fig. 3 is a block diagram showing an example configuration of a computation device (control means) of the service infrastructure company server of Fig. 2;

Fig. 4 is a diagram showing an example of data (original data) provided by the data provider server;

Fig. 5 is an example of a master table of original data (Ia) provided by the data provider server, the master table being prepared in the service infrastructure company server;

Fig. 6 is an example of a master table showing information of a data provider (company A) in the service infrastructure company server;

Fig. 7 is an example of a master table showing field attribute information of original data provided by the data provider server, the master table being prepared in the service infrastructure company server;

Fig. 8 is an example of a master table showing field attributes of data provided by the data provider server; Fig. 9 is an example of a master table showing invalid values which are comprehensively prepared in the service infrastructure company server;

Fig. 10 is an example of a master table showing

processing methods which are comprehensively prepared in the service infrastructure company server;

Fig. 11 is a diagram showing an example of a data cleansing policy setting screen with which the service receiver server sets a data cleansing policy;

Fig. 12 is a diagram showing an example of a default data cleansing policy for data (Ia) of a data provider (company A);

Fig. 13 is a diagram showing an example of a data receiver (company P) master in the service infrastructure company server;

Fig. 14 is a diagram showing an example of a data cleansing policy master in the service infrastructure company server;

Fig. 15 is a diagram showing the flowchart of a process where the service infrastructure company device cleanses data (Ia and Ib) provided by data providers (companies A and B) and transmits the cleansed data to the data receiver server;

Fig. 16 is a diagram schematically showing an aspect where the data provider (company A) transmits data (Ia), a default data cleansing policy (CaIa) of the data, and a data provider default data cleansing policy (Ca) to the infrastructure company;

Fig. 17 is a diagram schematically showing an aspect where the data receiver (company P) transmits a default data cleansing policy (Cp) of the data receiver (company P) and a

data receiver (company P) data cleansing policy (CpIa) for data (Ia) of the data provider to the infrastructure provider and where the infrastructure company performs data cleansing using at least one of a default data cleansing policy Cg of the infrastructure company, the data provider data cleansing policy (CaIa) of the data, and the default data cleansing policy (Ca) of the data provider;

Fig. 18 is a diagram showing a flowchart for generating a display of a data cleansing policy setting screen in the data receiver server;

Fig. 19 is a diagram showing the flowchart of the process of Fig. 16;

Fig. 20 is a diagram showing the flowchart of the process of Fig. 17;

Fig. 21 is a diagram showing the flowchart of a process related to data, records, and fields in Fig. 20;

Fig. 22 is a diagram showing the flowchart of one process according to a processing method;

Fig. 23 is a diagram showing the flowchart of another process according to a processing method;

Fig. 24 is a diagram showing the flowchart of another process according to a processing method;

Fig. 25 is a diagram showing the flowchart of another process according to a processing method; and

Fig. 26 is a diagram showing the flowchart of another

process according to a processing method.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0053]

Hereinafter, the present embodiment will be described with reference to the drawings. First, a service providing system for supporting the business of service receiving companies will be outlined.

The service providing system includes a service infrastructure company system or device configured to provide application services or SI, that is, services such as a reserve function or business SaaS (first services S1) to service providing companies (hereafter referred to as “the service infrastructure company device”); and service providing company systems or devices configured to customize first services in accordance with requests from service receiving companies (restaurants, or the like) and to provide the customized first services to the service receiving companies (hereafter referred to as “the service providing company devices”).

[0054]

A service receiving company (restaurant, or the like) provides users (consumers) with food and drink, merchandise, or the like and receives user information (original information) from users.

[0055]

The service infrastructure company device may be a device for providing PaaS (Platform as a Service) services for lending a platform including an OS, ASP (Application Service provider) services, or the like rather than SaaS.

[0056]

Hereinafter, the restaurant industry (restaurants), service providing companies therefor, and a service infrastructure company will be mainly described.

First, a service providing system will be outlined. The service infrastructure company device of the service providing system includes an application server configured to provide a reserve management service (application service) or the like to the service providing company devices (hereafter referred to as "the service infrastructure company server).

[0057]

Each service providing company device includes an application server configured to customize this reserve management service in accordance with requests from the service receiving companies and to receive and manage user (consumer) reserve information from the service receiving companies (hereafter referred to as "the service providing company server). That is, the service providing system has a mechanism where the service infrastructure company device and the service providing company devices collaborate with each

other to support the business of the service receiving companies.

[0058]

The system having such a mechanism typically addresses an individual case corresponding to each service receiving company. Accordingly, the service infrastructure company device only has to provide the first service S1 to each service providing company device, while each service providing company device only has to provide the customized first service (start of service) to each service receiving company and to manage pieces of user information (original information) Ia and Ib based on the provided first service, or the like.

[0059]

Examples of the method for providing this service include SaaS. SaaS is an ICT utilization environment which allows a service providing company or service receiving company to use a function provided by an information system (software) (this function will be referred to as “the first service”) via a network without having to construct or introduce a new system in particular.

[0060]

SaaS typically has an advantage that it allows a service providing company or service receiving company to immediately use a needed service when needed without having to construct a

special system. Further, since SaaS does not need to prepare a special environment in the company, it is possible to advance the time when the function (service) of SaaS starts operating and to reduce cost. Owing to the technological innovation of networks or browsers, a service providing company or service receiving company can operate SaaS, which uses the Internet or WAN, just as the service providing company or service receiving company uses an intra-company system.

[0061]

Furthermore, the service providing company or service receiving company can reduce cost or increase sales by operating or maintaining SaaS or SI provided by the service infrastructure company.

[0062]

The present invention can cleanse, for each of the data receivers, a result (data) obtained by performing data mining on data provided by such data providing company devices (service providing company devices A and B, service receiving company devices a1, a2, b1, and b2, or the like) and provide the cleansed data to the devices of the data receivers.

[0063]

Hereinafter, the embodiment of the present invention will be described with reference to the drawings. In the present embodiment, it is possible to accept even unevenly cleansed data, to, for example, fuse or integrate the accepted multiple

pieces of data, and to efficiently analyze the resulting data. Further, the analysis data or original data is cleansed according to the setting of the data receiver device and then transmitted according to a distribution request from the data receiver.

[0064]

A system according to the present embodiment is as follows:

(1) If the system including N number of data providers and M number of data receivers, the data providers and the data receivers receive data through an infrastructure provider.

(2) The server of the infrastructure provider includes “data cleansing means” configured to cleanse the original data of a data provider. That is, the system of the present invention is configured centering on the infrastructure provider.

(3) The server of the infrastructure provider holds a “data cleansing policy master" (Fig. 14).

(4) In the “data cleansing policy master” (Fig. 14), values in a “default data cleansing master” (Fig. 12) set by each data provider are set as default values.

(5) Each data provider registers the “attribute” of each field of “original data” (Fig. 4) in an “original data field attribute information master” (Fig. 7). Each “original data" is managed in an “original data source master (Fig. 2).

(6) Each data provider selects a “field ID” of the “original data field attribute master" (Fig. 7) from a “field attribute master (Fig. 8).

(7) In the “field attribute master” (Fig. 8), field types are comprehensively prepared. If the “field attribute master” does not include a corresponding type,” the data provider adds the corresponding field type in the field attribute master.”

(8) Invalid values” are managed in an “invalid value master (Fig. 5).” In the “invalid value master (Fig. 5),” invalid values are previously comprehensively prepared. If the “invalid value master (Fig. 5)” does not include a corresponding invalid value, the data provider adds the corresponding invalid value to the “invalid value master.”

(9) “Processing methods” of an invalid value” are managed in a “processing method master” (Fig. 10). In the "processing method master,” processing methods are previously comprehensively prepared. If the “processing method master” does not include a corresponding processing method, the data provider adds the corresponding processing method to the “processing method master” (Fig. 10).

(10) The data provider specifies a “processing method” corresponding to a “field attribute” on a “data cleansing policy setting screen” (Fig. 11) and sets each ID (value) in a “default data cleansing master” (Fig. 12). A data providing company ID defaults to “0.”

(11) In the “data cleansing policy setting screen” (Fig. 11), “processing IDs” are set as the values of field attributes (“primary key,” “name,” “zip code,” "postal address,” "quantity,” “amount,” and the like) corresponding to “invalid values (“null,” “double-byte/single-byte inconsistency,” “invalid character,” “unformatted,” and the like).

(12) Each data receiver can set a “data cleansing policy” using the “data cleansing policy setting screen” (Fig. 11). The set “data cleansing processing policy” is incorporated into the “data cleansing policy master” (Fig. 14). The data receivers are managed in a “data receiver master” (Fig. 13).

(13) Each data receiver sets a processing policy, that is, a “processing method” corresponding to an “invalid value” of data in the “data cleansing policy master” once (Fig. 14). As seen above, each data receiver sets a processing policy corresponding to an “invalid value” of the original data of the data provider. As a result, each data receiver can set data cleansing policies corresponding to the respective original data of the N number of data providers by setting a processing policy once.

(14) Each data provider and each data receiver can set a data cleansing policy corresponding to specific data of the data provider. If a data cleansing policy corresponding to the specific data is already set, the data cleansing policy is

used with higher priority than the default data cleansing policy.

(15) A data cleansing server of the infrastructure provider reads, from the data cleansing policy master, a “processing method” which corresponds to an “invalid value” (an invalid value ID of Fig. 12) specified by a data provider and which is specified by a data receiver, and then executes a processing program.

(16) The server of the infrastructure provider includes a program for performing a process corresponding to the “invalid value master” (Fig. 9) and the “processing method master" (Fig. 10).

Hereinafter, a detailed embodiment will be described.

First Embodiment

[0065]

Fig. 1 is a diagram showing the overall configuration of a service providing system where a customer-specific data cleansing system of the present invention is applied to a service infrastructure company device.

[0066]

The service providing system includes a data providing company device 12 of a data providing company and a service infrastructure company device 3 of a service infrastructure company.

[0067]

The data providing company device 12 includes multiple service providing company devices 2 including multiple service providing company devices (A, B, and the like) and service receiving company devices 1 including multiple service receiving company devices (a, b, and the like).

[0068]

The data providing company device 12 includes a data providing company server 121. The service receiving company devices 1 are, for example, devices of multiple stores (a1 and a2) or multiple companies (b1 and b2) and include service receiving company servers 11. Each service receiving company server 11 outputs original user information, that is, original data (Ia or Ib) to the corresponding service providing company device 2.

[0069]

The original user information Ia or Ib is sent from a user in response to the first service S1 (business/reserve function, or the like) provided by the service infrastructure company device 3 through a network 4.

[0070]

Each service providing company device 2 includes a service providing company server 21. The service providing company server 21 receives each original data I (Ia, Ib) and transmits the original data directly to a service

infrastructure company device 3 through a network 4 or transmits data which is anonymized by using a well known anonymization technology to the extent that the source of the original data, that is, the user cannot be identified (including the provided data = the original data I(Ia), field attribute information F (FIa) of the data, and a data cleansing policy C (Ca) to the service infrastructure company device 3 through the network 4.

[0071]

The service infrastructure company device 3 includes a service infrastructure company server 31. The service infrastructure company server 31 receives the data I(Ia), the field attribute information F (FIa), and the data cleansing policy C (Ca) from the data providing company (company A) server 121.

[0072]

The service infrastructure company device 3 receives a data cleansing policy C (Cp) from a data receiving company (company P) server (see 1216 of Fig. 2). The service infrastructure company device 3 has a function of cleansing the original data provided by the data providing company server in accordance with the received data cleansing policy and outputting cleansed data S.

[0073]

Fig. 2 is a block diagram showing the configurations of

the service infrastructure company server 31, a data providing company server 121, and a data receiving company server 1216.

[0074]

This example will be described assuming that the service providing company server 21 of Fig. 1 is the data providing company server 121, that the service receiving company server 11 is the data receiving company server 1216, and that the data provided to the service infrastructure company server 31 by the data providing company server 121 is original data I (Ia, Ib, and the like). This example will be also described assuming that the data provided by the data providing company server 121 is cleansed and that the cleansed data S is transmitted to the data receiving company server 1216. Note that the original data may be subjected to desired analysis as necessary. While this analysis may be conducted by the data receiving company on the basis of the cleansed data, the analysis is not the essential of the present invention and therefore will not be described.

[0075]

While the data providing company server 121 is a server for providing the original data I, the information F indicating the field attribute of the data, the data cleansing policy C, and the like to the service infrastructure company server 31, it may be used as a data receiver server for receiving the cleansed data S from the service infrastructure

company server 31. While the data receiver server 1216 is a server that provides the data cleansing policy Cp of the data receiver in receiving the cleansed data S, it may be used as a data providing company server configured to provide data to the service infrastructure company server.

[0076]

To receive the cleansed data S from the service infrastructure company server 31, the data receiving company server 1216 requests the service infrastructure company server 31 to receive the cleansed data S (data reception request signal), as well as transmits the data cleansing policy C (Cp) of the data to the service infrastructure company server 31. The data cleansing policy C (Cp) may be previously set and registered in the service infrastructure company server 31.

[0077]

The service infrastructure company server 31 includes an input/output device (input/output interface) 311, a storage device (storage means) 312, and a computation device (computation means) 313. These devices are connected together through a LAN.

[0078]

The storage device 312 includes a data provider and data receiver management master DB/data and field attribute information management DB 3121, a data provider, data receiver, and data cleansing policy information management DB 3122, a

field attribute master DB 3123, an invalid value master DB 3124, and a processing method master DB 3125.

[0079]

The data provider and data receiver management master DB/data and field attribute information management DB 3121 is a DB for managing data or field attribute information of a data provider and data receiver. In this DB, each data provided by the data provider server 121 and the field attribute information of the data are stored and managed in an associated manner.

[0080]

The data provider, data receiver, and data cleansing policy information management DB 3122 stores a data cleansing policy ("data provider ID/data receiver ID," "original data source ID," "original data field ID," "original data field name," "invalid value," "processing method ID," and the like) set by a data provider or data receiver.

[0081]

In other words, the data provider, data receiver, and data cleansing policy information management DB 3122 manages the data cleansing policy master, which the service infrastructure company server 31 uses when cleansing the data I provided by the data provider server 121.

[0082]

The field attribute master DB 3123 stores "field

attribute ID" and "attribute," that is, "data number," "purchase data," "store name," "customer name," "trade name," "amount," "the number of items purchased," "date," and the like.

[0083]

The invalid value master DB 3124 stores "invalid value ID" and "invalid value type," that is, "null," "double-byte/single-byte inconsistency," "invalid character," "unformatted," and the like.

[0084]

The processing method master DB 3125 stores "processing IDs" and "processing," that is, "do nothing," "delete all records," "make null," "replace with an average value," "comply with a data receiver data cleansing policy," "comply with a data receiver default data cleansing policy," "comply with a data provider data cleansing policy," "comply with an infrastructure company data cleansing policy, "comply with a data provider default data cleansing policy," and the like. Details of these DBs will be described later.

[0085]

The computation device 313 includes control means 3131 configured to cleanse the original data I. The control means 3131 includes a data provider and data receiver recognition unit 31311, a data and field attribute information receiving unit 31312, a data cleansing unit 31313, a cleansed data

distribution unit 31314, a data cleansing policy receiving unit 31315, a data write/read unit 31316, and the like.

[0086]

The data provider and data receiver recognition unit 31311 recognizes a data provider and a data receiver.

[0087]

The data and field attribute information receiving unit 31312 receives the original data I and the field attribute information F from the data provider server 121.

[0088]

The data cleansing policy receiving unit 31315 receives the data cleansing policy information C (Ca, Cp) from the data provider server 121 and the data receiver server 1216.

[0089]

The data cleansing unit 31313 cleanses the data provided by the data provider server 121 in accordance with the data cleansing policy in the data cleansing policy information management DB 3122.

[0090]

For example, this data cleansing process is intended to allow the service infrastructure company server 31 to analyze the data of the data provider server 121. Thus, the data receiver server 1216 can receive cleansed data from the service infrastructure company server 31 by once only setting a data cleansing policy that the data receiver server 1216

hopes will be used to perform data cleansing.

[0091]

The data write/read unit 31316 writes each data I (Ia, Ib, and the like) of the data provider server or the data cleansing policy C (Ca, Cp) for each data to the DB 3121 or reads such data from the DB.

[0092]

The cleansed data distribution unit 31315 transmits the data cleansed by the data cleansing unit 31313 to the data receiver server 1216.

[0093]

Fig. 3 is a function block diagram showing an example of the service infrastructure company server 31. In Fig. 3, the server 31 receives the provided data I and the field attribute information F of the data at the data and field attribute information receiving unit 31312 (see Fig. 2) and stores the data and the field attribute information in the data and field attribute information management DB 3121 in an associated manner.

[0094]

The provided data I is stored in the data and field attribute information management DB 3121 in such a manner to remain in the format provided by the data provider server 121 and then cleansed by the data cleansing unit 31313.

[0095]

The data cleansing unit 31313 may store the cleansed data S in the data and field attribute information management DB 3121.

[0096]

The field attribute information of the provided data I is stored in the data and field attribute information management DB 3121 in such a manner to be associated with the data I.

[0097]

Hereinafter, details of each DB will be described. Fig. 4 is a diagram showing an example of data (original data I) provided by a data provider server and is a table showing an original data name "sales journal table" of a data provider name "kakaka store, Inc." in an original data master of Fig. 5.

[0098]

In Fig. 4, the original data includes information 401 to 407, including original data primary keys ("1," "2," "3," "4," and the like), purchase dates ("2013.3.2," "2013.3.2," "2013.3.3," "2013.3.4," and the like), store names ("Shibuya central," "Osaki higashi," "Osaki higashi," "Shinagawa main store," and the like), customer names ("ota □rou," "oki □rou," "otatsu□rou," "J Ferry," and the like), trade names ("optical mouse," "four-color ball-point pen," "100-sheet name card case," "pencil case," and the like), the numbers of items purchased ("5," "100," "10," "5," and the like), and amounts ("10000," "20, 000," "30, 000," "40000," and the like).

[0099]

Fig. 5 is a table showing an original data master which is a master of data (original data Ia) provided by the data provider servers. The original data master is prepared in the service infrastructure company server.

[0100]

In Fig. 5, the original data source master includes information 501 to 504, including original data IDs (“1,” “2,” “3,” “4,” and the like), original data names (“sales journal table,” “reserved journal table,” “time zone-specific sales table,” “reserved journal table,” and the like), a data provider IDs (“1,” “2,” “3,” “3,” and the like), and data provider names (“kakaka store, Inc.,” “kikiki hotel, Inc.,” “kekeke beauty salon, Inc.,” “kekeke beauty salon, Inc.,” and the like).

[0101]

Fig. 6 is a table showing a data provider (company A) master in the service infrastructure company server.

[0102]

In Fig. 6, the data provider master includes information 601 to 602, including data provider IDs (“1,” “2,” “3,” and the like), data provider names (“kakaka store, Inc.,” “kikiki hotel, Inc.,” “kekeke beauty salon, Inc.,” and the like).

[0103]

Fig. 7 is a table showing an original data field

attribute master of data provided by the data provider servers. The original data field attribute master is prepared in the service infrastructure company server.

[0104]

In Fig. 7, the original data field attribute master includes information 701 to 704, including field IDs (“1,” “2,” “3,” “4,” “5,” “6,” “7,” and the like), original data IDs (“1,” “1,” “1,” “1,” “1,” “1,” “1,” and the like) field names (“data number,” “purchase date,” “store name,” “customer name,” “trade name,” “the number of items purchased,” “amount,” and the like), and field attribute IDs (“1,” “8,” “3,” “3,” “3,” “7,” “6,” and the like).

[0105]

Fig. 8 is a table showing a field attribute master of data provided by a data provider server.

[0106]

In Fig. 8, the field attribute master includes information 801 and 802, including field type attribute IDs (“1,” “2,” “3,” “4,” “5,” “6,” “7,” “8,” and the like) and attributes (“data number,” “purchase date,” “name,” “postal address,” “phone number,” “amount,” “quantity,” “date,” and the like).

[0107]

Fig. 9 is a table showing an invalid value master which is comprehensively prepared by the service infrastructure

company server.

In Fig. 9, the invalid value master includes information 901 and 902, including invalid value IDs (“1,” “2,” “3,” “4,” and the like), invalid value types (“null,” “double-byte/single-byte inconsistency,” “invalid character,” “unformatted,” and the like).

[0108]

Fig. 10 is a table showing the processing method master, which is comprehensively prepared by the service infrastructure company server.

[0109]

In Fig. 10, the processing method master includes information 1001 and 1002, including processing method IDs (“1,” “2,” “3,” “4,” “5,” “6,” “7,” “8,” “9,” and the like), processes (“do nothing,” “delete all records,” “make a null,” “replace with an average value,” “make an alphanumeric character a double-byte character,” “comply with the default data cleansing policy of a data receiver=>company P default data cleansing policy Cp,” “comply with the data cleansing policy of a data provider => company A data cleansing policy CaIa for original data Ia of company A,” “comply with infrastructure company default data cleansing policy => infrastructure company default data cleansing policy Cg,” “comply with the default data cleansing policy of a data provider => default data cleansing policy Ca of company A,”

and the like).

[0110]

Fig. 11 is a diagram showing a data cleansing policy setting screen for setting a data cleansing policy in a server receiver server.

[0111]

In Fig. 11, the data cleansing policy setting screen horizontally displays field names 703 (see Fig. 7), that is, "data number" 7031, "purchase date" 7032, "store name" 7033, "customer name" 7034, "trade name" 7035, "the number of items purchased" 7036, and "amount" 7037, and the field attributes of the field names," that is, "primary key" 7041,' "date" 7042,' "name" 7043,' "name" 7044,' "name" 7045,' "quantity" 7046,' "amount" 7047,' and the like. The data cleansing policy setting screen also vertically displays "invalid values" (see Fig. 9), that is, "null," "double-byte/single-byte inconsistency," "invalid character," "unformatted," and the like. The processing method IDs (1 to 9, and the like) in the processing method master (see Fig. 10) are set in the "field attributes" corresponding to the "invalid values" using select boxes.

[0112]

Fig. 12 is a diagram showing an example of a default data cleansing policy for data (Ia) of the data provider (company A).

[0113]

In Fig. 12, the default data cleansing policy master includes information 1201 to 1206, including data provider IDs (“1,” “1,” “1,” “1,” and the like), original data source IDs (“1,” “1,” “1,” “1,” and the like), original data field IDs (“1,” …, “2,” …, “3,” , “4,” …, and the like), original data field names (“original data primary key,” …, “purchase date,” …, “store name,” …, “customer name,”…, and the like), invalid value IDs (“1,” “2,” “3,” “4,” “1,” “2,” “3,” “4,” “1,” “2,” “3,” “4,” and the like), and processing method IDs (“2,” “2,” “2,” “2,” “1,” “1,” “3,” “3,” “1,” “5,” “3,” “3,” “1,” “5,” “5,” “3”).

[0114]

Fig. 13 is a diagram showing an example of a data receiver (company P) master in the service infrastructure company server.

[0115]

In Fig. 13, the data receiver master includes information 1301 and 1302, including data receiver IDs (“1,” “2,” “3,” and the like) and data receiver names (“sasasa factory, Inc.,” “shishishi noodle factory, Ltd.,” “sususu trading, Inc.,” and the like).

[0116]

Fig. 14 is a diagram showing an example of a data cleansing policy master in the service infrastructure company

server.

[0117]

In Fig. 14, the data cleansing policy master includes information 1401 to 1406, including data receiver IDs (“1,” “1,” “1,” “1,” and the like), original data source IDs (“1,” “1,” “1,” “1,” and the like), original data field IDs (“1,” …, “2,” …, “3,” …, “4,” …, and the like), original data field names (“original data primary key,” , “purchase date,” …, “store name,” …, “customer name,” …, and the like), invalid value IDs (“1,” “2,” “3,” “4,” “1,” “2,” “3,” “4,” “1,” “2,” “3,” “4” and the like), and processing method IDs (“2,” “2,” “2,” “2,” “1,” “1,” “3,” “3,” “1,” “5,” “3,” “3,” “1,” “5,” “5,” “3”).

[0118]

Next, the process steps of the customer-specific data cleansing method will be described. Fig. 15 is a diagram showing the process flow of the customer-specific data cleansing method of the present invention. This example shows a case where there are two data provider servers 121 and where there is one data receiver server 1216. It is assumed that a default data cleansing policy is already set in each of these servers and that a field attribute master, an invalid value master, and a processing method master are shared by the data provider servers and the data receiver server. In Fig. 15, the process is performed as follows:

(1) A data provider (company A) generates field attribute information FIa for data Ia and a company A data cleansing policy CaIa for the data Ia on the basis of the above three masters.

(2) The data provider (company A) transmits the data Ia, the field attribute information FIa, and the company A data cleansing policy CaIa for the data Ia to the infrastructure providing company.

(3) A data provider (company B) also performs steps similar to the above (2) and (3).

(4) A data receiver (company P) requests the infrastructure company to provide the field attribute information FIa of the data Ia and the company A data cleansing policy CaIa for the data Ia.

(5) In response to this request, the infrastructure company transmits, to the data receiver (company P), the field attribute information FIa of the data Ia of the data provider (company A) and the company A data cleansing policy CaIa for the data Ia.

(6) The data receiver (company P) receives and refers to FIa and CaIa, generates a data cleansing policy CpIa for the data Ia, and transmits CaIa to the infrastructure company.

(7) The data receiver (company P) also requests the infrastructure company to provide the cleansed data.

(8) Based on the request, the infrastructure company performs

data cleansing and transmits the cleansed data to the server receiver (company P).

(9) The service receiving company (company P) receives the cleansed data.

[0119]

In step S1211, the data provider (company A) server 121 transmits the data Ia and the field attribute information FIa of the data to the service infrastructure company server 31.

[0120]

Similarly, in step S1212, the server 121 of the other the data provider (company B) transmits data Ib and the field attribute FIb of the data to the service infrastructure company server 31.

[0121]

On the other hand, in step S12161, the data receiver (company P) server 1216 requests the service infrastructure company server 31 to transmit the field attribute information FIa of the data Ia and the data cleansing policy CaIa. In step S12163, the server 1216 requests the server 31 to receive the cleansed data, as well as transmits the data cleansing policy CpIa.

[0122]

In step S311, the service infrastructure company server 31 receives the data Ia and Ib and the field attribute information FIa and FIb and writes and stores the data and

information to the data and field attribute information management DB 3121. Further, in step S315, in response to the request from the data receiver (company P) server 1216, the service infrastructure company server 31 transmits the field attribute information FIa of the data Ia of the data provider (company A) and the company A data cleansing policy CaIa for the data Ia.

[0123]

In step S312, the service infrastructure company server 31 receives the data request and the data cleansing policy Cp and writes and stores these to the data provider, data receiver, and data cleansing policy information management DB 3122.

[0124]

Next, in step S313, the service infrastructure company server 31 performs data cleansing in accordance with a processing method compliant with the data cleansing policy Cp in the data cleansing policy information management DB 3122. Each original data is cleansed in compliance with the data cleansing policy Cp.

[0125]

When necessary, this data cleansing process may be performed according to a receiver data cleansing policy CpIa which the data receiver obtains by partially changing the default data cleansing policy Ca set by the data provider, the

data cleansing policy CaIa for the data of the data provider, the infrastructure providing company data cleansing policy Cg set by the infrastructure company, or the default data cleansing policy CaIa for the data provider set by the data provider.

[0126]

Subsequently, in step S314, the service infrastructure company server 31 transmits, to the data receiver server 121, cleansed data obtained by cleansing the data (original data) Ia provided by the data provider server 121.

[0127]

At this time, although not shown, the service infrastructure company server 31 converts the data from the data provider server 121 into a data format compliant with the data cleansing policy of the server 31 and analyzes the data in the unified infrastructure data format using a desired analysis technique (data mining). Alternatively, the data may remain in the infrastructure data format or in the data format of the data provider server 121. These processes are performed in accordance with a data format request from the data receiver server 1216.

[0128]

In step S12162, the data receiver server 1216 receives the cleansed data.

[0129]

In this way, the data receiver can receive the desired cleansed data on which the data cleansing policy of the data receiver is reflected.

[0130]

Fig. 16 is a diagram schematically showing the flow of a process where the data provider (company A) transmits original data, a field attribute, and a data cleansing policy to the infrastructure company.

[0131]

In Fig. 16, the data provider (company A) first captures the field attribute master (see Fig. 8), the invalid value master (see Fig. 9), and the processing method master (see Fig. 10) from the storage device of the infrastructure company. Based on these masters, the data provider generates a default data cleansing policy Ca of the company A and transmits it to the infrastructure company.

[0132]

The data provider also transmits original data Ia of the company A to the infrastructure company. At this time, based on the original data Ia of the company A and the field attribute master (see Fig. 8), the invalid value master (see Fig. 9), and the processing method master (see Fig. 10) captured from the infrastructure company, the data provider generates a field attribute FIa of the original data Ia of the company A and a data cleansing policy CaIa for the original

data Ia of the company A and transmits these to the infrastructure company.

[0133]

The infrastructure company receives the default data cleansing policy Ca of the company A, the original data Ia of the company A, the field attribute FIa of the original data Ia of the company A (see Fig. 8), and the default data cleansing policy CaIa for the data Ia of the company A and stores these in the data cleansing policy information management DB 3122 of the storage device thereof (see Fig. 12).

[0134]

Fig. 17 is a diagram schematically showing the flow of a process where the data receiver (company P) transmits a data cleansing policy to the infrastructure company and where the infrastructure company transmits cleansed data to the data receiver.

[0135]

In Fig. 17, the data receiver (company P) first captures the field attribute master (see Fig. 8), the invalid value master (see Fig. 9), and the processing method master (see Fig. 10) from the storage device of the infrastructure company. Based on these masters, the data receiver generates a default data cleansing policy Cp of the company P and transmits it to the infrastructure company.

[0136]

The data receiver also captures the original data Ia of the company A generated by the data provider (company A), the field attribute FIa of the original data Ia of the company A, and the data cleansing policy CaIa for the original data Ia of the company A from the storage device of the infrastructure company.

[0137]

If the data receiver wants to make a change to the default data cleansing policy CaIa for the original data Ia of the company A, that is, if the data receiver wants to set a processing policy based on the will of the data receiver, the data receiver (company P) sets a desired data cleansing policy CpIa and transmits it to the infrastructure company. For example, the data receiver refers to the field attribute information FIa of the original data Ia of the company A and the company A data cleansing policy for the original data of the company A captured from the infrastructure company, generates a company P data cleansing policy for the original data Ia of the company A and incorporates the generated policy into the default data cleansing processing master of the infrastructure company so that data cleansing can be performed in compliance with the policy.

[0138]

Such data cleansing policy setting is performed, for example, using a data cleansing setting screen as shown in Fig.

10.

[0139]

The infrastructure company receives the company P default data cleansing policy Cp and the company P data cleansing policy CpIa for the original data Ia of the company A from the data receiver (company P). Based on these processing policies, or based on the company A data cleansing policy CaIa for the original data Ia of the company A, the infrastructure company default data cleansing policy Cg, or the company A default data cleansing policy Ca, the infrastructure company cleanses the original data Ia of the company A.

[0140]

Subsequently, the infrastructure company performs desired format conversion or desired analysis on the cleansed data and transmits the resulting data to the data receiver (company P).

[0141]

Fig. 18 is a flowchart for generating a display of a data cleansing policy setting screen between the data receiver and the infrastructure company. In step S1111 of Fig. 18, the data receiver acquires a list of invalid values from the "invalid value master" of the infrastructure company. In step S1112, the data receiver displays the invalid values on the vertical axis of a "data cleansing policy setting screen" (see Fig. 11).

[0142]

In step S1114, the data receiver acquires a list of field

attributes from the “field attribute master” (see Fig. 8) and displays the values of the field attributes on the horizontal axis of the “data cleansing policy setting screen.”

[0143]

In step S1116, the data receiver acquires processing method IDs from the “processing method master” and displays the values of the processing method IDs in the value fields of the “data cleansing policy setting screen” in such a manner that the values are selectable.

[0144]

In step S1117, the data receiver sets “processing methods” corresponding to the “invalid values” of the “field attributes” on the “data cleansing policy setting screen.” In step S1118, the data receiver transmits the set “processing methods” to the infrastructure company (server).

The infrastructure company receives the “processing methods” from the infrastructure company and, in step S1119, generates a “data cleansing policy master” on the basis of the processing methods and stores the policy master in the storage device.

[0145]

Fig. 19 is a flowchart for generating a display of a data cleansing policy setting screen between the data provider and the infrastructure company.

[0146]

In step S1121 of Fig. 19, the data provider (the data provider, company A) captures the “field attribute master,” the “invalid value master,” and the “processing method master” from the infrastructure company.

[0147]

In step S1122, using the “data cleansing policy setting screen, the infrastructure company sets an “infrastructure company data cleansing policy Cg” on the basis of the “field attribute master,” the “invalid value master,” and the “processing method master.”

[0148]

In step S1123, using the “data cleansing policy setting screen,” the data provider, the company A, sets a “company A default data cleansing policy Ca” on the basis of the “field attribute master,” the "invalid value master,” and the “processing method master.”

[0149]

In step S1124, the data provider, the company A, transmits the “original data Ia of company A” to the infrastructure company.

[0150]

In step S1125, the data provider, the company A, transmits the “field attribute information FIa” of the “original data Ia of the company A” to the infrastructure company.

[0151]

In step S1126, the data provider, the company A, refers to the "default data cleansing policy Ca of the company A" to set a "data cleansing policy CaIa for the original data Ia of the company A."

[0152]

In step S1127, the data provider, the company A, transmits the "default data cleansing policy CaIa for the original data Ia of the company A" to the infrastructure company.

[0153]

Figs. 20 to 26 are flowcharts showing an example of a series of processes from a data cleansing policy generation process to a data cleansing process.

[0154]

In step S1411 of Fig. 20, the data receiver, the company P, captures the "field attribute master," "invalid value master," and the "processing method maser" from the infrastructure company, generates a "default data cleansing policy Cp of the company P" using the "data cleansing policy setting screen," and transmits the policy Cp to the infrastructure company.

[0155]

In step S1412, the data receiver, the company P, captures the "field attribute information FIa of the data Ia of the

company A" and the "data cleansing policy CaIa for the data Ia of the company A" from the infrastructure company.

[0156]

In step S1413, using the "data cleansing policy setting screen," the data receiver, the company P, generates a "company P data cleansing policy CpIa for the original data Ia of the company A" with reference to the "field attribute information FIa of the data Ia of the company A" and the "data cleansing policy CaIa for the data Ia of the company A" and transmits the policy CpIa to the infrastructure company.

[0157]

Subsequently, the process proceeds to *1 of Fig. 21. In step S1421 of Fig. 21, the infrastructure company performs processes until a record number i of the data Ia reaches from 1 to the number of records (i ≤ the number of records). In step S1422, the infrastructure company performs processes until a field number j of the record i of the data Ia reaches from 1 to the number of fields (j ≤ the number of fields). The infrastructure company then proceeds to step S1423 to determine whether the field (i, j) is an invalid value.

[0158]

If the field (i, j) is an invalid value (Y), the infrastructure company proceeds to step S1424 to determine whether there is a company P data cleansing policy CpIa for the original data Ia of the company A.

If the field (i, j) is not an invalid value (N), the infrastructure company proceeds to step S1430 to perform a process to be described later.

[0159]

If there is a CpIa in step S1424, the infrastructure company proceeds to *0 of Fig. 25 to perform a process according to a processing method of Fig. 10 to be discussed later. If there is no CpIa, the infrastructure company proceeds to step S1425 to determine whether there is a data cleansing policy CaIa for the original data Ia of the company A.

[0160]

If there is a CaIa in step S1425, the infrastructure company proceeds to *7out of Fig. 23 to perform a process according to a processing method of Fig. 10 to be discussed later. If there is no CaIa, the infrastructure company proceeds to step S1426 to determine whether there is a company P default data cleansing policy Cp.

[0161]

If there is a Cp in step S1426, the infrastructure company proceeds to 6out of Fig. 24 to perform a process according to a processing method of Fig. 10 to be discussed later. If there is no Cp, the infrastructure company proceeds to step S1427 to determine whether there is a company A default data cleansing policy Ca.

[0162]

If there is a Ca in step S1427, the infrastructure company proceeds to *9out of Fig. 25 and perform a process according to a processing method of Fig. 10 to be discussed later. If there is no Ca, the infrastructure company proceeds to step S1428 to determine whether there is an infrastructure company default data cleansing policy Cg.

[0163]

If there is a Cg in step S1428, the infrastructure company proceeds to *8out of Fig. 26 to perform a process according to a processing method of Fig. 10 to be discussed later. If there is no Cg, the infrastructure company returns to *100, increments j by 1 (j = j + 1) in step S1430, and increments i by 1 (i + 1) in step S1431, ending the process.

[0164]

Next, referring to Figs. 22 to 26, there will be described the processes according to the processing methods of Fig. 10.

If there is a CpIa in step S1424 of Fig. 21, the infrastructure company refers to the processing policy of Fig. 14 in step S14241 of Fig. 22 to acquire a processing method ID.

[0165]

Next, in step S14242, the infrastructure company determines whether the processing method ID is one of 1, 2, 3, 4, and 5. If the ID is one of 1, 2, 3, 4, and 5 (Y), the

infrastructure company returns to *4 of Fig. 21 and performs a process according to the processing method ID in step S1429. The infrastructure company then performs the above steps S1430 and S1431, ending the process.

[0166]

If the ID is not any of 1, 2, 3, 4, and 5 in step S14243 of Fig. 22 (N), the infrastructure company proceeds to step S14243 to determine whether the processing method ID is 7.

[0167]

If the processing method ID is 7 (Y), the infrastructure company returns to *7in of Fig. 21 to perform the above steps. If the processing method ID is not 7 (N), the infrastructure company proceeds to step S14244 to determine whether the processing method ID is 6.

[0168]

If the processing method ID is 6 (Y), the infrastructure company returns to *6in of Fig. 21 to perform the above steps. If the processing method ID is not 6 (N), the infrastructure company proceeds to step S14245 to determine whether the processing method ID is 9.

[0169]

If the processing method ID is 9 (Y), the infrastructure company returns to *9in of Fig. 21 to perform the above steps. If the processing method ID is not 9 (N), the infrastructure company proceeds to step S14246 to determine whether the

processing method ID is 8.

[0170]

If the processing method ID is 8 (Y), the infrastructure company returns to *8in of Fig. 21 to perform the above steps. If the processing method ID is not 8 (N), the infrastructure company end the process.

[0171]

If there is a CaIa in step S1425 of Fig. 21, the infrastructure company acquires a processing method ID with reference to the processing policy of Fig. 14 in step S14251 of Fig. 23.

[0172]

Next, in step S14252, the infrastructure company determines whether the processing method ID is one of 1, 2, 3, 4, and 5. If the ID is one of 1, 2, 3, 4, and 5 (Y), the infrastructure company returns to *4 of Fig. 21 and performs a process according to the processing method ID in step S1429. The infrastructure company then performs the above steps S1430 and S1431, ending the process.

[0173]

If the ID is not any of 1, 2, 3, 4, and 5 (N), the infrastructure company proceeds to step S14254 to determine whether the processing method ID is 6.

[0174]

If the processing method ID is 6 (Y), the infrastructure

company returns to *6in of Fig. 21 to perform the above steps. If the processing method ID is not 6 (N), the infrastructure company proceeds to step S14255 to determine whether the processing method ID is 9.

[0175]

If the processing method ID is 9 (Y), the infrastructure company returns to *9in of Fig. 21 to perform the above steps. If the processing method ID is not 9 (N), the infrastructure company proceeds to step S14256 to determine whether the processing method ID is 8.

[0176]

If the processing method ID is 8 (Y), the infrastructure company returns to *8in of Fig. 21 to perform the above steps. If the processing method ID is not 8 (N), the infrastructure company ends the process.

[0177]

If there is a Cp in step S1426 of Fig. 21, the infrastructure company acquires a processing method ID with reference to the policy of Fig. 14 in step S14261 of Fig. 24.

[0178]

Next, in step S14262, the infrastructure company determines whether the processing method ID is one of 1, 2, 3, 4, and 5. If the ID is one of 1, 2, 3, 4, and 5 (Y), the infrastructure company returns to *4 of Fig. 21 and performs a process according to the processing method ID in step S1429.

The infrastructure company then performs the above steps S1430 and S1431, ending the process.

[0179]

If the ID is not any of 1, 2, 3, 4, and 5 (N), the infrastructure company proceeds to step S14265 to determine whether the processing method ID is 9.

[0180]

If the processing method ID is 9 (Y), the infrastructure company returns to *9in of Fig. 21 to perform the above steps. If the processing method ID is not 9 (N), the infrastructure company proceeds to step S14266 to determine whether the processing method ID is 8.

[0181]

If the processing method ID is 8 (Y), the infrastructure company returns to *8in of Fig. 21 to perform the above steps. If the processing method ID is not 8 (N), the infrastructure company ends the process.

[0182]

If there is a Ca in step S1427 of Fig. 21, the infrastructure company acquires a processing method ID with reference to the policy of Fig. 14 in step S14271 of Fig. 25.

[0183]

Next, in step S14272, the infrastructure company determines whether the processing method ID is one of 1, 2, 3, 4, and 5. If the ID is one of 1, 2, 3, 4, and 5 (Y), the

infrastructure company returns to *4 of Fig. 21 and performs a process according to the processing method ID in step S1429. The infrastructure company then performs the above steps S1430 and S1431, ending the process.

[0184]

If the ID is not any of 1, 2, 3, 4, and 5 (N), the infrastructure company proceeds to step S14276 to determine whether the processing method ID is 8.

[0185]

If the processing method ID is 8 (Y), the infrastructure company returns to *8in of Fig. 21 to perform the above steps. If the processing method ID is not 8 (N), the infrastructure company ends the process.

[0186]

If there is a Cg in step S1428 of Fig. 21, the infrastructure company acquires a processing method ID with reference to the policy of Fig. 14 in step S14281 of Fig. 26.

[0187]

Next, in step S14282, the infrastructure company determines whether the processing method ID is one of 1, 2, 3, 4, and 5. If the ID is one of 1, 2, 3, 4, and 5 (Y), the infrastructure company returns to *4 of Fig. 21 and performs a process according to the processing method ID in step S1429. The infrastructure company then performs the above steps S1430 and S1431, ending the process.

[0188]

If the ID is not any of 1, 2, 3, 4, and 5 (N), the infrastructure company returns to *100 of Fig. 21 and performs the above steps, ending the process.

[0189]

The present invention is not limited to the above embodiment and includes various modifications. For example, while the embodiment has been described in detail to clarify the present invention, the invention need not always include all the described components. Further, some components of an embodiment can be replaced with components of another embodiment, and an embodiment can additionally include components of another embodiment. Further, for some components of each embodiment, other components can be added, deleted, or replaced.

[0190]

Some or all of the above components, functions, processing units, processing means, and the like may be realized as hardware, for example, by designing these using integrated circuits. Alternatively, the above components, functions, and the like may be realized as software when a processor interprets and executes a program for realizing the functions. Information for achieving the functions, such as a program, a table, and a file, can be stored in a storage device, such as a memory or hard disk.

[0191]

The drawings show only control lines and information lines considered necessary for description and do not necessarily show all of the control lines and information lines of the product. It can be considered that almost all the components are actually coupled to each other.

REFERENCE SIGNS LIST

[0192]

1: service receiving company device

2: service providing company device

3: service infrastructure company device

31: infrastructure company server

311: input/output device (input/output interface means)

312: storage device (storage means)

313: computation device (control means)

121: data provider server

31: service infrastructure company server

121: data provider server

1216: data receiver server

3121: data provider and data receiver management master DB/data and field attribute information management DB

3122: data provider, data receiver, and data cleansing policy information management DB

3123: field attribute master DB

3124: invalid value master DB

3125: processing method master DB

31311: data provider and data receiver recognition unit

31312: data and field attribute information receiving unit

31313: data cleansing unit

31314: cleansed data distribution unit

31315: data cleansing policy receiving unit

31316: data write/read unit

Claims

1. A customer-specific data cleansing system serving as a data providing system, the system comprising:

at least one service infrastructure company device;

a plurality of data provider devices; and

a plurality of data receiver devices,

wherein the service infrastructure company device includes:

an input/output means;

a storage means; and

a control means,

the input/output means is configured to provide or receive data to or from the data provider devices and the data receiver devices,

the storage means includes a data cleansing policy master with respect to data provided by the data provider devices, and

the control means has a function of cleansing the data in accordance with a data cleansing policy of the data cleansing policy master,

wherein the control means further includes:

means configured to receive default data cleansing policies of the data receiver devices, data receiver data cleansing policies for data provided by the data provider devices, default data cleansing policies of the data provider devices, and data provider data cleansing policies for data provided by the data provider devices and to incorporate the

received policies into the data cleansing policy master;

a receiving unit configured to receive data provided by the data provider devices and field attribute information of the data;

a data cleansing policy receiving unit configured to receive the data cleansing policy;

a data cleansing means configured to cleanse data provided by the data provider devices in accordance with the data cleansing policy; and

means configured to transmit the data cleansed by the data cleansing means to the data receiver devices.

2. The customer-specific data cleansing system according to claim 1,

wherein the control means further transmit the data provider data cleansing policies to the data receiver devices in response to request of the data receiver devices.

3. The customer-specific data cleansing system according to claim 1,

wherein the control means further includes means including a field attribute master, an invalid value master, and a processing method master and configured to generate a data cleansing policy of the service infrastructure company device

from the field attribute master, the invalid value master, and the processing method master, and

the means configured to incorporate into the data cleansing policy master incorporates the data cleansing policy of the service infrastructure company device into the data cleansing policy master.

4. The customer-specific data cleansing system according to claim 1,

wherein the data cleansing policy master includes original data field names of the data provided by the data provider devices and processing methods corresponding to invalid values which may occur in field attributes of the field names, and

the invalid values include one or more of null, double-byte/single-byte inconsistency, invalid character, abnormal value, hiragana character inconsistency, katakana character inconsistency, alphabetical character inconsistency, value inconsistency, symbol inconsistency, kanji character inconsistency, phone number denormalization, zip code denormalization, denormalized postal address, and denormalized amount, and

the field attributes includes one or more of primary key, name, zip code, postal address, quantity, amount, and

date, and

the processing methods includes one or more of doing nothing, deleting all records, complying with a data provider format, complying with a normalized data provider, making a null, and replacing with an average value.

5. A customer-specific data cleansing system for cleansing data shared by a plurality of companies, the system including an infrastructure company device configured to receive original data provided by a device for a data provider serving as a company and to provide the original data to a data receiver device,

wherein the infrastructure company device includes:

a data cleansing control device configured to cleanse the original data provided by the device for the data provider;

an invalid value master defining invalid values which may occur in fields of the original data; and

a data cleansing policy master including processing methods corresponding to the invalid values which may occur in the fields of the original data,

the device for the data provider includes means configured to, when a value in the original data is an invalid value, determine to which of items in the invalid value master the invalid value corresponds,

the data receiver device includes means configured to set a processing policy corresponding to invalid values with respect to field attributes comprehensively prepared for the original data, and

the data cleansing control device includes:

a reading means configured to read the processing policy set by the data receiver device from the data cleansing policy master with respect to field attribute information of the original data set by the device for the data provider; and

a control means configured to cleanse the original data in accordance with the read data cleansing policy.

6. A customer-specific data cleansing method, the method being performed by a data providing system comprising:

at least one service infrastructure company device;

a plurality of data provider devices; and

a plurality of data receiver devices,

the service infrastructure company device including:

an input/output means;

a storage means;and

a control means,

the input/output means being configured to provide or receive data to or from the data provider devices and the data receiver devices,

the storage means including a data cleansing policy master with respect to data provided by the data provider devices,

the control means having a function of cleansing the data in accordance with a data cleansing policy of the data cleansing policy master, and

the control means further including means configured to receive default data cleansing policies of the data receiver devices, data receiver data cleansing policies for data provided by the data provider devices, default data cleansing policies of the data provider devices, and data provider data cleansing policies for data provided by the data provider devices and to incorporate the received policies into the data cleansing policy master;

the method comprising the steps of:

receiving data provided by the data provider devices and field attribute information of the data;

receiving a data cleansing policy of the data cleansing policy master;

cleansing data provided by the data provider devices in accordance with the data cleansing policy; and

transmitting the data cleansed to the data receiver devices.

7. The customer-specific data cleansing method according to claim 6, further comprising the step of:

transmitting, by the control means, the data provider data cleansing policies to the data receiver devices in response to request of the data receiver devices.

8. The customer-specific data cleansing system according to claim 6,

wherein the control means includes a field attribute master, an invalid value master, and a processing method master,

the method further comprising:

generating, by the control means, a data cleansing policy of the service infrastructure company device from the field attribute master, the invalid value master, and the processing method master, and

incorporating, by the means configured to incorporate into the data cleansing policy master, the data cleansing policy of the service infrastructure company device into the data cleansing policy master.

9. The customer-specific data cleansing method according to claim 6,

wherein the data cleansing policy master includes original data field names of the data provided by the data provider devices, field attributes corresponding to invalid values which may occur in field attributes of the field names, and processing methods corresponding to invalid values which may occur in field attributes of the field names, and

the invalid values include one or more of null, double-byte/single-byte inconsistency, invalid character, abnormal value, hiragana character inconsistency, katakana character inconsistency, alphabetical character inconsistency, value inconsistency, symbol inconsistency, kanji character inconsistency, phone number denormalization, zip code denormalization, denormalized postal address, and denormalized amount, and

the field attributes include one or more of primary key, name, zip code, postal address, quantity, amount, and date, and

the processing methods include one or more of doing nothing, deleting all records, complying with a data provider format, complying with a normalized data provider, making a null, and replacing with an average value.