(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 April 2021 (22.04.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/076324 A1

(51) International Patent Classification:
*G06F 3/048* (2013.01) *G06Q 30/02* (2012.01)
*G06F 17/30* (2006.01)

(21) International Application Number:
PCT/US2020/053419

(22) International Filing Date:
30 September 2020 (30.09.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/923,377 18 October 2019 (18.10.2019) US
16/795,570 19 February 2020 (19.02.2020) US

(71) Applicant: **ASG TECHNOLOGIES GROUP, INC. DBA ASG TECHNOLOGIES** [US/US]; 708 Goodlette Road North, Naples, Florida 34102 (US).

(72) Inventors: **MACNEILL, Marcus**; c/o ASG Technologies, 708 Goodlette Road North, Naples, Florida 34102 (US). **MORESMAU, Jean-Philippe**; c/o ASG Technologies, 708 Goodlette Road North, Naples, Florida 34102 (US).

(74) Agent: **GOEL, Sonia** et al.; Carr & Ferrell LLP, 120 Constitution Drive, Menlo Park, California 94025 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Published:**
— *with international search report (Art. 21(3))*

(54) Title: MULTI-FACETED TRUST SYSTEM

(57) Abstract: Provided are methods and systems for determining multi-faceted trust scores for data. A method may commence with receiving data and determining a plurality of metadata items associated with the data. The method may continue with determining one or more facets associated with each of the plurality of metadata items. The method may further include determining a parameter and a weight associated with each of the one or more facets. Upon determining the parameter and the weight, a trust score associated with each of the plurality of metadata items may be calculated based on the parameter and the weight associated with each of the one or more facets. The method may further include calculating a multi-faceted trust score of the data based on the trust score of each of the plurality of metadata items.

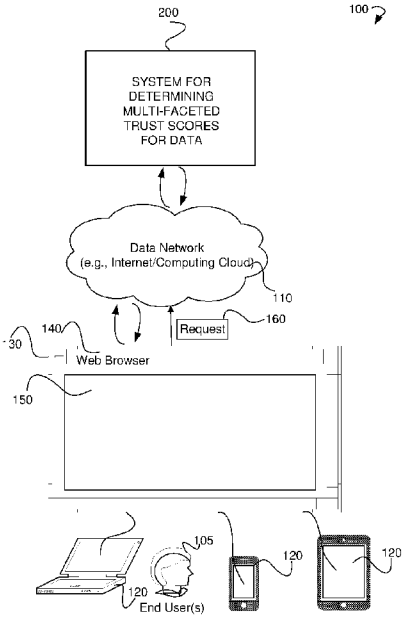


FIG. 1

WO 2021/076324 A1

# MULTI-FACETED TRUST SYSTEM

## CROSS-REFERENCE TO RELATED APPLICATIONS

**[0001]** This application claims priority of U.S. Provisional Patent Application No. 62/923,377 filed on October 18, 2019, entitled "MULTI-FACETED TRUST SYSTEM," which is incorporated herein by reference in its entirety.

## FIELD

**[0002]** This application relates generally to data processing and, more specifically, to systems and methods for determining multi-faceted trust scores for data.

## BACKGROUND

**[0003]** Defensive data strategies, such as regulatory compliance, require a truth-based approach to understanding data that is both highly automated and accurate. Instituting the same approach for offensive data strategies, such as self-service data analytics, would be costly, time consuming, and even impossible in some situations (for example, when combining internal and external data).

## SUMMARY

**[0004]** This summary is provided to introduce a selection of concepts in a simplified form that are further described in the Detailed Description below. This summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used as an aid in determining the scope of the claimed subject matter.

**[0005]** Provided are methods and systems for determining multi-faceted trust scores for data. In some embodiments, a method for determining multi-faceted trust scores for data may commence with receiving data and determining a plurality of

metadata items associated with the data. The method may continue with determining one or more facets associated with each of the plurality of metadata items. The method may further include determining a parameter and a weight associated with each of the one or more facets. Upon determining the parameter and the weight, a trust score associated with each of the plurality of metadata items may be calculated based on the parameter and the weight associated with each of the one or more facets. The method may further include calculating a multi-faceted trust score of the data based on the trust score of each of the plurality of metadata items.

**[0006]** In some example embodiments, a system for determining multi-faceted trust scores for data may include a data collection unit, a data analyzing unit, and a score calculation unit. The data collection unit may be configured to receive data. The data analyzing unit may be configured to determine a plurality of metadata items associated with the data. The data analyzing unit may be further configured to determine one or more facets associated with each of the plurality of metadata items. The data analyzing unit may determine a parameter and a weight associated with each of the one or more facets. The score calculation unit may be configured to calculate a trust score associated with each of the plurality of metadata items based on the parameter and the weight associated with each of the one or more facets. The score calculation unit may be further configured to calculate a multi-faceted trust score of the data based on the trust score of each of the plurality of metadata items.

**[0007]** Additional objects, advantages, and novel features will be set forth in part in the detailed description section of this disclosure, which follows, and in part will become apparent to those skilled in the art upon examination of this specification and the accompanying drawings or may be learned by production or operation of the example embodiments. The objects and advantages of the concepts may be realized and attained by means of the methodologies, instrumentalities, and combinations particularly pointed out in the appended claims.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0008]** Exemplary embodiments are illustrated by way of example and not limitation in the figures of the accompanying drawings, in which like references indicate similar elements.

**[0009]** FIG. 1 illustrates an environment within which methods and systems for determining multi-faceted trust scores for data can be implemented, according to an example embodiment.

**[0010]** FIG. 2 is a block diagram illustrating a system for determining multi-faceted trust scores for data, according to an example embodiment.

**[0011]** FIG. 3 is a flow diagram illustrating a method for determining multi-faceted trust scores for data, according to an example embodiment.

**[0012]** FIG. 4 is a schematic diagram showing a widget for calculating a trust score, according to an example embodiment.

**[0013]** FIG. 5 is a computing system that can be used to implement a method for determining multi-faceted trust scores for data, according to an example embodiment.

## DETAILED DESCRIPTION

**[0014]** The following detailed description includes references to the accompanying drawings, which form a part of the detailed description. The drawings show illustrations in accordance with exemplary embodiments. These exemplary embodiments, which are also referred to herein as "examples," are described in enough detail to enable those skilled in the art to practice the present subject matter. The embodiments can be combined, and other embodiments can be formed, by introducing structural and logical changes without departing from the scope of what is claimed. The following detailed description is, therefore, not to be taken in a limiting sense and the scope is defined by the appended claims and their equivalents.

**[0015]** In this document, the terms "a" or "an" are used, as is common in patent documents, to include one or more than one. In this document, the term "or" is used to refer to a nonexclusive "or," such that "A or B" includes "A but not B," "B but not A," and "A and B," unless otherwise indicated. Furthermore, all publications, patents, and patent documents referred to in this document are incorporated by reference herein in their entirety, as though individually incorporated by reference. In the event of inconsistent usages between this document and those documents so incorporated by reference, the usage in the incorporated reference(s) should be considered supplementary to that of this document; for irreconcilable inconsistencies, the usage in this document controls.

**[0016]** The present disclosure relates to systems and methods for determining multi-faceted trust scores for data. The system of the present disclosure provides an approach for understanding data in support of offensive data strategies that include augmenting truth (where available) with a dynamic trust model comprised of multiple facets. The system of the present disclosure can be used to analyze critical data components or other data items and determine a degree to which the data can be trusted by the users. The trust score can be calculated based on several factors, which

are referred to herein as facets. More specifically, each of the facets includes a characteristic of a metadata item of data. The trust score is also referred to as a "score" herein.

**[0017]** The method for determining multi-faceted trust scores for data may commence with receiving data and determining a plurality of metadata items associated with the data. The method may continue with determining one or more facets associated with each of the plurality of metadata items. The method may further include determining a parameter and a weight associated with each of the one or more facets. In an example embodiment, the weight of a facet can be entered by an operator based on empirical observations or established using a predetermined technique based on historical trustworthiness associated of the facet. Upon determining facets and corresponding parameters and weights, a trust score for each of the plurality of metadata items can be calculated based on the parameter and the weight associated with each of the facets. The method can further include calculating a multi-faceted trust score of the data based on the trust score of each of the plurality of metadata items.

**[0018]** Thus, the trust score for a specific metadata item can be calculated based on different facets, which may include objective facets (metadata, data, business information), subjective facets (for example, based on ratings of a user), and synthetic facets (calculated from heuristics on lineage). All facets can be evaluated and scored. The facets contribute to the final score based on their relative weights.

**[0019]** A decision table may be used to determine which facets contribute to the trust score of a particular metadata item (i.e., to determine which facets to use and which weights to assign to each facet). The decision table can use an item type (for example, tables and columns are evaluated differently) and other criteria. The result of the decision is a list of selected facets with parameters and weights. The data intelligence (DI) repository may then calculate each facet based on the information the facet contains, using the parameters provided. For example, the DI repository may

check if a business term was associated with the metadata item by determining whether the metadata item is tagged by the tag of type "Business Glossary." Other types of tags may also be referenced in the trust facet configuration. All scores can be then aggregated, taking into account the weights of each facet (some facets may contribute more than others to the final trust score). The final score and each individual facet score can be stored alongside the metadata item.

**[0020]** The score may be expressed as a percentage; for example, 100% means "complete trust" while 0% means "no trust at all." The user can visualize the score and how each facet contributed to the score in order to evaluate why the score is what it is and what caused the score to change over time. Thresholds can be associated with different levels of trust for more intuitive visualization; for example, a trust score of over 90% can be shown in green, under 50% can be shown in red, and so forth. Facet visualization may also be indicative of whether a facet is objective (such as a metadata field value), subjective (such as user ratings), or synthetic (for example, a facet calculated from heuristics on lineage).

**[0021]** The trust score calculation can be performed automatically when changes occur to the metadata item (for example, re-analysis of the metadata item, or manual tagging) or be triggered by the user. The DI repository may provide some trust configurations that can be readjusted based on specific needs. Other facets can be added to the facets provided by default.

**[0022]** Referring now to the drawings, FIG. 1 illustrates an environment 100 within which systems and methods for determining multi-faceted trust scores for data can be implemented. The environment 100 may include a data network 110 (e.g., an Internet or a computing cloud), end user(s) 105, client device(s) 120 associated with the end user 105, and a system 200 for determining multi-faceted trust scores for data. Client device(s) 120 may comprise a personal computer (PC), a desktop computer, a laptop, a smartphone, a tablet, or so forth.

**[0023]** The client device 120 may have a user interface 130. Furthermore, a web browser 140 may be running on the client device 120 and displayed using the user interface 130. The web browser 140 may communicate with the system 200 via the data network 110.

**[0024]** The data network 110 may include the Internet or any other network capable of communicating data between devices. Suitable networks may include or interface with any one or more of, for instance, a local intranet, a corporate data network, a data center network, a home data network, a Personal Area Network, a Local Area Network (LAN), a Wide Area Network (WAN), a Metropolitan Area Network, a virtual private network, a storage area network, a frame relay connection, an Advanced Intelligent Network connection, a synchronous optical network connection, a digital T1, T3, E1 or E3 line, Digital Data Service connection, Digital Subscriber Line connection, an Ethernet connection, an Integrated Services Digital Network line, a dial-up port such as a V.90, V.34 or V.34bis analog modem connection, a cable modem, an Asynchronous Transfer Mode connection, or a Fiber Distributed Data Interface or Copper Distributed Data Interface connection. Furthermore, communications may also include links to any of a variety of wireless networks, including Wireless Application Protocol, General Packet Radio Service, Global System for Mobile Communication, Code Division Multiple Access or Time Division Multiple Access, cellular phone networks, Global Positioning System, cellular digital packet data, Research in Motion, Limited duplex paging network, Bluetooth radio, or an IEEE 802.11-based radio frequency network. The data network can further include or interface with any one or more of a Recommended Standard 232 (RS-232) serial connection, an IEEE-1394 (FireWire) connection, a Fiber Channel connection, an IrDA (infrared) port, a Small Computer Systems Interface connection, a Universal Serial Bus (USB) connection or other wired or wireless, digital or analog interface or connection, mesh or Digi® networking.

**[0025]** The web browser 140 can render a web page associated with the system 200 which end user(s) 105 can use to calculate a trust score. The web browser 140 can establish a communication channel with the system 200 and generate and render virtual screens based on data received from the system 200.

**[0026]** The end user 105 may send a request 160 to the system 200 using the client device 120. The request 160 may include data for which a trust score needs to be calculated. In response to the request 160, the system 200 may calculate the trust score and render the results by the web browser 140.

**[0027]** FIG. 2 is a block diagram illustrating a system 200 for determining multi-faceted trust scores for data, according to an example embodiment. The system 200 may include a data collection unit 210, a data analyzing unit 220, and a score calculation unit 230. In an example embodiment, the operations performed by each of the data collection unit 210, the data analyzing unit 220, and the score calculation unit 230 may be performed by a processor and a memory for storing instructions executable by the processor. Example one or more processors 510 are shown in FIG. 5.

**[0028]** The operations performed by each of the data collection unit 210, the data analyzing unit 220, and the score calculation unit 230 of the system 200 are described in detail below with reference to FIG. 3.

**[0029]** FIG. 3 shows a process flow diagram of a method 300 for determining multi-faceted trust scores for data, according to an example embodiment. In some embodiments, the operations may be combined, performed in parallel, or performed in a different order. The method 300 may also include additional or fewer operations than those illustrated. The method 300 may be performed by processing logic that may comprise hardware (e.g., decision making logic, dedicated logic, programmable logic, and microcode), software (such as software run on a general-purpose computer system or a dedicated machine), or a combination of both.

**[0030]** The method 300 may commence with the data collection unit receiving data at operation 310. The data may include any type of data that needs to be analyzed and for which a trust score is required to be determined. The data may be received from any source (for example, a database, a remote source, an online source, a cloud database, the system 200 itself, and so forth). At step 320, a plurality of metadata items associated with the data may be determined by the data analyzing unit. The metadata may include information associated with the data. The method 300 may further include determining, by the data analyzing unit, one or more facets associated with each of the plurality of metadata items at operation 330. In an example embodiment, each of the one or more facets includes a characteristic of a metadata item. The facets may include one of the following: data quality dimensions, criticality of an item of the data, governance of the item of the data (governed/not governed), a rating, a review, an issue, a proximity of the item of the data to a source (for example, how close the data item is to the source), an existence of a data lineage, a fact of scanning the item of the data from an active source or a spreadsheet, a tag associated with the data item, a frequency of update, a frequency of use, a null value, usefulness of the item of the data for an intended purpose, and so forth. The facets may include objective facets (metadata, data, business information), subjective facets (for example, based on ratings of a user), and synthetic facets (calculated from heuristics on lineage).

**[0031]** At operation 340, the data analyzing unit may determine a parameter and a weight associated with each of the one or more facets. The parameter may include, for example, a value of the facet, a 'yes/no' parameter, a percentage, a range, and so forth. The method 300 may continue at operation 350 with calculating, by the score calculation unit, a trust score for each of the plurality of metadata items based on the parameter and the weight associated with each of the one or more facets. In an example embodiment, the calculation may include summing all parameters of facets (related to the metadata item) multiplied by corresponding weights and dividing the summation

by the number of facets related to the metadata item. The result can be a number between 0 and 1 or a percentage between 0% and 100%. Thereafter, at operation 360, the method 300 may proceed with calculating, by the score calculation unit, a multi-faceted trust score of the data based on the trust score of each of the plurality of metadata items. In an example embodiment, the calculation may include summing all trust scores of each of the plurality of metadata items and dividing the summation by the number of metadata items related to the data.

**[0032]** In an example embodiment, the method 300 may further include selecting, from the one or more facets, a plurality of facets that contribute to the trust score for each of the plurality of metadata items. In this embodiment, the calculation of the trust score for each of the plurality of metadata items may be based on the parameter and the weight associated with the selected plurality of facets.

**[0033]** Optionally, a trust score associated with each of the one or more facets may be calculated to obtain a plurality of trust scores for the one or more facets. The calculation may be based on the parameter and the weight associated with each of the one or more facets. In this embodiment, the trust score of each of the plurality of metadata items may be a sum of the plurality of trust scores of the one or more facets associated with each of the plurality of metadata items.

**[0034]** The method 300 may further include determining whether the multi-faceted trust score exceeds one or more predetermined thresholds. For example, the trust score that reached 100% threshold means "complete trust," while the trust score that reached 0% means "no trust at all." Thresholds can be associated with different levels of trust for more intuitive visualization; for example, a trust score that exceeded 90% threshold can be shown in green, a trust score that is under 50% threshold can be shown in red, and so forth.

**[0035]** Once the trust score is calculated, the trust score can be used as an input for a variety of use cases both inside a data intelligence product (e.g., a machine

learning an artificial intelligence node) itself and also other products. The trust score can trigger predetermined actions when the trust score reaches specific thresholds for a metadata item and/or can be used in conjunction with other data intelligence information to provide additional information associated with the data. Trust scores can be also combined with (or constructed entirely from) sources outside of the data intelligence in order to drive additional information into data and decisions beyond the data intelligence.

**[0036]** In an example embodiment, the method 300 may further include determining that the data have been changed. The change may include at least one of the following: re-analysis of at least one item of the data, manual tagging of the data, a trigger initiated by a user, and so forth. Based on the determined change, the multi-faceted trust score may be recalculated for the data.

**[0037]** In an example embodiment, factors that contribute to the trust score can be defined in policies. Specifically, a policy type of "enrichment" can be defined and used for any data that can contribute additional data. The factors may include a type of the data item, a catalog in which the item is placed, and so forth. More factors can be added if additional fine-grained calculations are needed. The result of setting the policy is a list of "Trust Factors." A factor can define what data item needs to be considered and the relative weight of the factor in the overall calculation. The calculation for the given item may be done on the DI side; the policy only sets how to calculate the trust score.

**[0038]** In an example embodiment, the setting of a policy may be performed as follows. For items of type 'Table', quality with a weight of 20%, the average rating with a weight of 30%, the number of ratings (if more than two ratings) with a weight of 20%, and the fact that the item is tagged with a Glossary tag (a tag called Glossary or any sub tag) may add 30% to the score. In this example, all weights sum up to 1, but this is not

required. Once the policy is set, the policy can be uploaded to a policy service upon running a configuration step (e.g., ConfigureDev script in a development environment).

**[0039]** Upon setting the policies, the trust score can be calculated by a backend service that reacts to events. The trust score and an explanation as to what factors contributed to the trust score can be displayed.

**[0040]** The trust factors may include the following facets: quality, ratings average, ratings count, tags, linked items, and so forth. The weight of each trust factor may be determined or set to get a range between 0 and 1.

**[0041]** **Quality**. The quality percentage score as calculated by the quality process (this is driven by configuration) may be taken. In particular, the percent divided by 100 may be taken.

**[0042]** **Ratings average**. When ratings go between 1 and 5, 3 is average. Thus, a rating of 1 results in a score of 0, a rating of 5 results in a score of 1, and a rating of 3 results in a score of 0.5.

**[0043]** **Ratings count**. There is a minimum value for the count of ratings to count (2 in the above example). If there are more, the score weighted by the count for each rating is calculated. Since ratings with low counts are filtered out, this can give a different result than the average.

**[0044]** **Tags**. The tag can include determining whether the item is tagged by any tag in the hierarchy of the given tag name. For example, any tag created from a glossary term may increase the trust factor, meaning that if an item is linked to a glossary term, it has been analyzed somewhat and is trusted more.

**[0045]** **Linked items**. The linked items may be taken and the average of their trust score can be calculated. For example, a dataset trust score may include the trust score of the items referenced by the dataset, the score of a table may include the score of each column, and so forth. The filter value is the name of the link to follow. Several

links can be specified, delimited by commas. By default, links can be followed in both directions, but the link name may be prefixed, or indicate in, out, or both directions.

**[0046]** **Trust calculation**. The trust score can be calculated based on data stored in the DI repository and by using an application programming interface (API), depending on a facet. The trust score can be calculated when a specific event, such as a TrustUpdateRequestEvent, is received. The event may carry explicit item identifiers (IDs) to calculate the trust score only on these items, explicit query to calculate the trust score only on the items returned by the query, or nothing to calculate the trust score on all items in the catalog that have trust score policies linked to them.

**[0047]** Other events can be added to recalculate the trust score automatically when an item is changed, rated, tagged, and so forth.

**[0048]** When the calculation is done, another event, such as a TrustUpdatedEvent, can be sent. The TrustUpdatedEvent is a subclass of ItemUpdatedEvent, but carries the old and new trust score, such that a user interface (UI) can properly indicate that the item has been updated if the item has been viewed.

**[0049]** Based on the calculation, the trust score and a JSON Array containing information on how much each factor contributed in the trustFactors attribute can be stored in the trust score attribute.

**[0050]** Trust API. A trust factors service can be provided to allow triggering a calculation of the trust score by sending the proper event (so the calculation is asynchronous).

**[0051]** FIG. 4 is a schematic diagram showing a widget 400 for calculating a trust score 405, according to an example embodiment. The UI may be using a chart 410, such as an ngx-chart (Gauge), to show the trust score 405 graphically. Red/yellow/green colors 415 may be shown based on the trust score 405 (green if over 90%, yellow if over 60%, and red otherwise). The colors can be configurable. Upon hovering over the chart 410 by the user, corresponding explanations 420 can be shown to the user. In an

example embodiment, a widget, such as the widget 400 showing a trust score associated with the quality, can be shown in search results and the like.

**[0052]** FIG. 5 illustrates an exemplary computing system 500 that may be used to implement embodiments described herein. The exemplary computing system 500 of FIG. 5 may include one or more processors 510 and memory 520. Memory 520 may store, in part, instructions and data for execution by the one or more processors 510. Memory 520 can store the executable code when the exemplary computing system 500 is in operation. The exemplary computing system 500 of FIG. 5 may further include a mass storage 530, portable storage 540, one or more output devices 550, one or more input devices 560, a network interface 570, and one or more peripheral devices 580.

**[0053]** The components shown in FIG. 5 are depicted as being connected via a single bus 590. The components may be connected through one or more data transport means. The one or more processors 510 and memory 520 may be connected via a local microprocessor bus, and the mass storage 530, one or more peripheral devices 580, portable storage 540, and network interface 570 may be connected via one or more input/output buses.

**[0054]** Mass storage 530, which may be implemented with a magnetic disk drive or an optical disk drive, is a non-volatile storage device for storing data and instructions for use by a magnetic disk or an optical disk drive, which in turn may be used by one or more processors 510. Mass storage 530 can store the system software for implementing embodiments described herein for purposes of loading that software into memory 520.

**[0055]** Portable storage 540 may operate in conjunction with a portable non-volatile storage medium, such as a compact disk (CD) or digital video disc (DVD), to input and output data and code to and from the computing system 500 of FIG. 5. The system software for implementing embodiments described herein may be stored on such a portable medium and input to the computing system 500 via the portable storage 540.

**[0056]** One or more input devices 560 provide a portion of a user interface. The one or more input devices 560 may include an alphanumeric keypad, such as a keyboard, for inputting alphanumeric and other information, or a pointing device, such as a mouse, a trackball, a stylus, or cursor direction keys. Additionally, the computing system 500 as shown in FIG. 5 includes one or more output devices 550. Suitable one or more output devices 550 include speakers, printers, network interfaces, and monitors.

**[0057]** Network interface 570 can be utilized to communicate with external devices, external computing devices, servers, and networked systems via one or more communications networks such as one or more wired, wireless, or optical networks including, for example, the Internet, an intranet, LAN, WAN, cellular phone networks (e.g., Global System for Mobile communications network, packet switching communications network, circuit switching communications network), Bluetooth radio, and an IEEE 802.11-based radio frequency network, among others. Network interface 570 may be a network interface card, such as an Ethernet card, optical transceiver, radio frequency transceiver, or any other type of device that can send and receive information. Other examples of such network interfaces may include Bluetooth®, 3G, 4G, and WiFi® radios in mobile computing devices as well as a USB.

**[0058]** One or more peripheral devices 580 may include any type of computer support device to add additional functionality to the computing system. The one or more peripheral devices 580 may include a modem or a router.

**[0059]** The components contained in the exemplary computing system 500 of FIG. 5 are those typically found in computing systems that may be suitable for use with embodiments described herein and are intended to represent a broad category of such computer components that are well known in the art. Thus, the exemplary computing system 500 of FIG. 5 can be a PC, handheld computing device, telephone, mobile computing device, workstation, server, minicomputer, mainframe computer, or any other computing device. The computer can also include different bus configurations,

networked platforms, multi-processor platforms, and so forth. Various operating systems (OS) can be used including UNIX, Linux, Windows, Macintosh OS, Palm OS, and other suitable operating systems.

**[0060]** Some of the above-described functions may be composed of instructions that are stored on storage media (e.g., computer-readable medium). The instructions may be retrieved and executed by the processor. Some examples of storage media are memory devices, tapes, disks, and the like. The instructions are operational when executed by the processor to direct the processor to operate in accord with the example embodiments. Those skilled in the art are familiar with instructions, processor(s), and storage media.

**[0061]** It is noteworthy that any hardware platform suitable for performing the processing described herein is suitable for use with the example embodiments. The terms "computer-readable storage medium" and "computer-readable storage media" as used herein refer to any medium or media that participate in providing instructions to a central processing unit (CPU) for execution. Such media can take many forms, including, but not limited to, non-volatile media, volatile media, and transmission media. Non-volatile media include, for example, optical or magnetic disks, such as a fixed disk. Volatile media include dynamic memory, such as random access memory (RAM). Transmission media include coaxial cables, copper wire, and fiber optics, among others, including the wires that include one embodiment of a bus. Transmission media can also take the form of acoustic or light waves, such as those generated during radio frequency and infrared data communications. Common forms of computer-readable media include, for example, a floppy disk, a flexible disk, a hard disk, magnetic tape, any other magnetic medium, a CD-read-only memory (ROM) disk, DVD, any other optical medium, any other physical medium with patterns of marks or holes, a RAM, a PROM, an EPROM, an EEPROM, a FLASHEPROM, any other memory chip or cartridge, a carrier wave, or any other medium from which a computer can read.

**[0062]** Various forms of computer-readable media may be involved in carrying one or more sequences of one or more instructions to a CPU for execution. A bus carries the data to system RAM, from which a CPU retrieves and executes the instructions. The instructions received by system RAM can optionally be stored on a fixed disk either before or after execution by a CPU.

**[0063]** Thus, various embodiments of methods and systems for determining multi-faceted trust scores for data have been described. Although embodiments have been described with reference to specific example embodiments, it will be evident that various modifications and changes can be made to these example embodiments without departing from the broader spirit and scope of the present application. Accordingly, the specification and drawings are to be regarded in an illustrative rather than a restrictive sense. There are many alternative ways of implementing the present technology. The disclosed examples are illustrative and not restrictive.

# CLAIMS

What is claimed is:

1. A method for determining multi-faceted trust scores for data, the method comprising:

receiving data;

determining a plurality of metadata items associated with the data;

determining one or more facets associated with each of the plurality of metadata items;

determining a parameter and a weight associated with each of the one or more facets;

calculating a trust score associated with each of the plurality of metadata items based on the parameter and the weight associated with each of the one or more facets; and

calculating a multi-faceted trust score of the data based on the trust score of each of the plurality of metadata items.

2. The method of claim 1, wherein each of the one or more facets includes a characteristic of a metadata item of the plurality of metadata items.

3. The method of claim 1, wherein the one or more facets includes one of the following: data quality dimensions, criticality of an item of the data, governance of the item of the data, a rating, a review, an issue, a proximity of the item of the data to a source, an existence of a data lineage, a fact of scanning the item of the data from an active source or a spreadsheet, a tag associated with the data item, a frequency of

update, a frequency of use, a null value, and usefulness of the item of the data for an intended purpose.

4. The method of claim 1, wherein the one or more facets include objective facets, subjective facets, and synthetic facets.

5. The method of claim 1, further comprising determining whether the multi-faceted trust score exceeds one or more predetermined thresholds.

6. The method of claim 1, further comprising:

determining that the data have been changed; and

based on the determination, recalculating the multi-faceted trust score for the data.

7. The method of claim 6, wherein the change includes at least one of the following: re-analysis of at least one item of the data, manual tagging of the data, and a trigger initiated by a user.

8. The method of claim 1, further comprising selecting, from the one or more facets, a plurality of facets that contribute to the trust score for each of the plurality of metadata items, wherein the calculating the trust score for each of the plurality of metadata items is based on the parameter and the weight associated with the selected plurality of facets.

9. The method of claim 1, further comprising calculating a trust score associated with each of the one or more facets to obtain a plurality of trust scores for the

one or more facets, the calculation being based on the parameter and the weight associated with each of the one or more facets.

10. The method of claim 9, wherein the trust score of each of the plurality of metadata items is a sum of the plurality of trust scores of the one or more facets associated with each of the plurality of metadata items.

11. A system for determining multi-faceted trust scores for data, the system comprising:

a data collection unit configured to receive data;

a data analyzing unit configured to:

determine a plurality of metadata items associated with the data;

determine one or more facets associated with each of the plurality of metadata items; and

determine a parameter and a weight associated with each of the one or more facets;

a score calculation unit configured to:

calculate a trust score associated with each of the plurality of metadata items based on the parameter and the weight associated with each of the one or more facets; and

calculate a multi-faceted trust score of the data based on the trust score of each of the plurality of metadata items.

12. The system of claim 11, wherein the one or more facets includes one of the following: a data quality dimension, a criticality of an item of the data, governance of the item of the data, a rating, a review, an issue, a proximity of the item of the data to a source, an existence of a data lineage, a fact of scanning the item of the data from an

active source or a spreadsheet, a tag associated with the data item, a frequency of update, a frequency of use, a null value, and usefulness of the item of the data for an intended purpose.

13. The system of claim 11, wherein the one or more facets include objective facets, subjective facets, and synthetic facets.

14. The system of claim 11, wherein the data analyzing unit is further configured to determine whether the multi-faceted trust score exceeds one or more predetermined thresholds.

15. The system of claim 11, wherein the data analyzing unit is configured to determine that the data has been changed; and

wherein the score calculation unit is further configured to recalculate, based on the determination, the multi-faceted trust score for the data.

16. The system of claim 15, wherein the change includes at least one of the following: re-analysis of at least one item of the data, a manual tagging of the data, and a trigger initiated by a user.

17. The system of claim 11, wherein the data analyzing unit is further configured to select, from the one or more facets, a plurality of facets that contribute to the trust score for each of the plurality of metadata items, wherein the calculating of the trust score for each of the plurality of metadata items is based on the parameter and the weight associated with the selected plurality of facets.

18. The system of claim 11, wherein the score calculation unit is further configured to calculate a trust score associated with each of the one or more facets to obtain a plurality of trust scores for the one or more facets, the calculation being based on the parameter and the weight associated with each of the one or more facets.

19. The system of claim 18, wherein the trust score of each of the plurality of metadata items is a sum of the plurality of trust scores of the one or more facets associated with each of the plurality of metadata items.

20. A system for determining multi-faceted trust scores for data, the system comprising:

a data collection unit configured to receive data;

a data analyzing unit configured to:

determine a plurality of metadata items associated with the data;

determine one or more facets associated with each of the plurality of metadata items;

determine a parameter and a weight associated with each of the one or more facets; and

select, from the one or more facets, a plurality of facets that contribute to the trust score for each of the plurality of metadata items; and

a score calculation unit configured to:

calculate a trust score associated with each of the plurality of metadata items based on the parameter and the weight associated with each of the one or more facets, wherein the calculating of the trust score for each of the plurality of metadata items is based on the parameter and the weight associated with the selected plurality of metadata items;

calculate a trust score associated with each of the one or more facets to obtain a plurality of trust scores of the one or more facets, the calculation being based on the parameter and the weight associated with each of the one or more facets; and

calculate a multi-faceted trust score of the data based on the trust score of each of the plurality of metadata items.

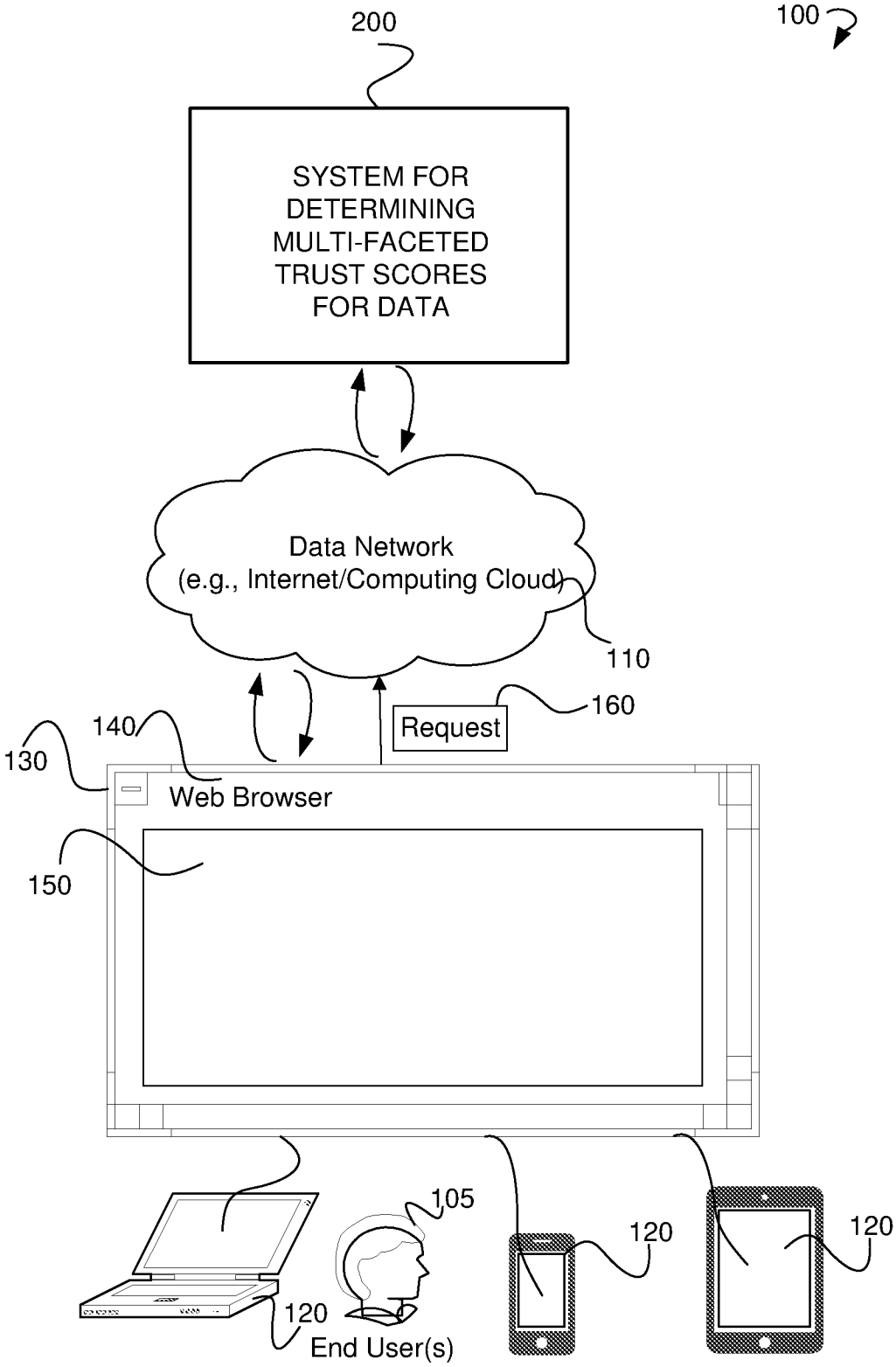
100
200
SYSTEM FOR DETERMINING MULTI-FACETED TRUST SCORES FOR DATA
Data Network
(e.g., Internet/Computing Cloud)
110
Request
160
140
130
Web Browser
150
105
120
120
120
End User(s)

*FIG. 1*

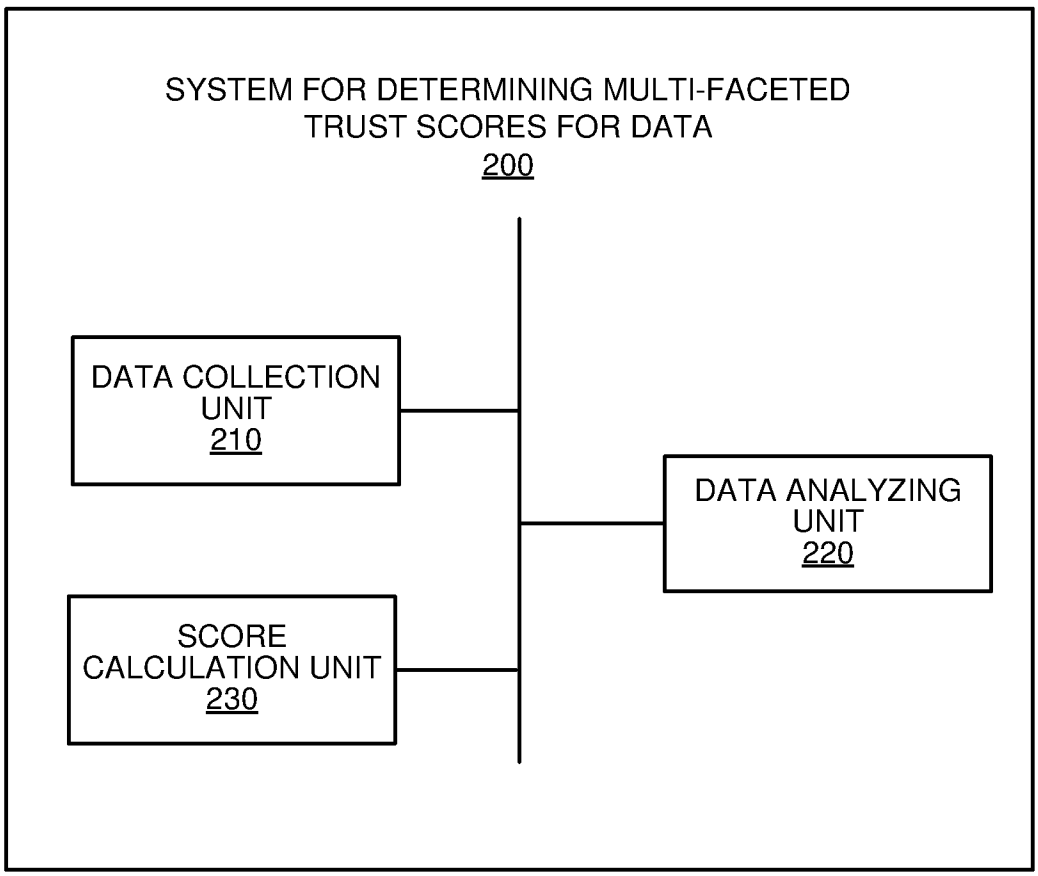
SYSTEM FOR DETERMINING MULTI-FACETED TRUST SCORES FOR DATA
200
DATA COLLECTION UNIT
210
DATA ANALYZING UNIT
220
SCORE CALCULATION UNIT
230

*FIG. 2*

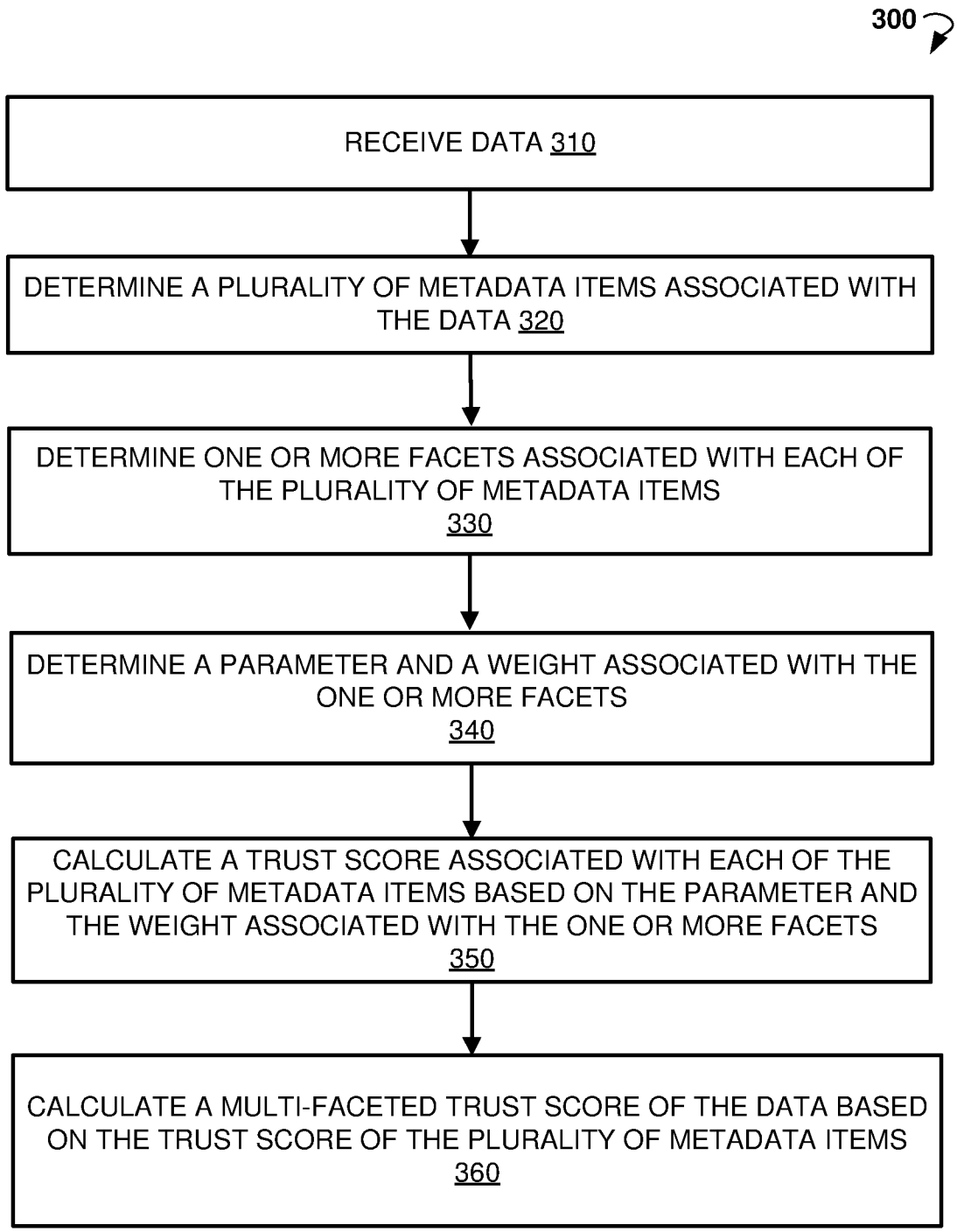
300
RECEIVE DATA 310
DETERMINE A PLURALITY OF METADATA ITEMS ASSOCIATED WITH THE DATA 320
DETERMINE ONE OR MORE FACETS ASSOCIATED WITH EACH OF THE PLURALITY OF METADATA ITEMS 330
DETERMINE A PARAMETER AND A WEIGHT ASSOCIATED WITH THE ONE OR MORE FACETS 340
CALCULATE A TRUST SCORE ASSOCIATED WITH EACH OF THE PLURALITY OF METADATA ITEMS BASED ON THE PARAMETER AND THE WEIGHT ASSOCIATED WITH THE ONE OR MORE FACETS 350
CALCULATE A MULTI-FACETED TRUST SCORE OF THE DATA BASED ON THE TRUST SCORE OF THE PLURALITY OF METADATA ITEMS 360

*FIG. 3*

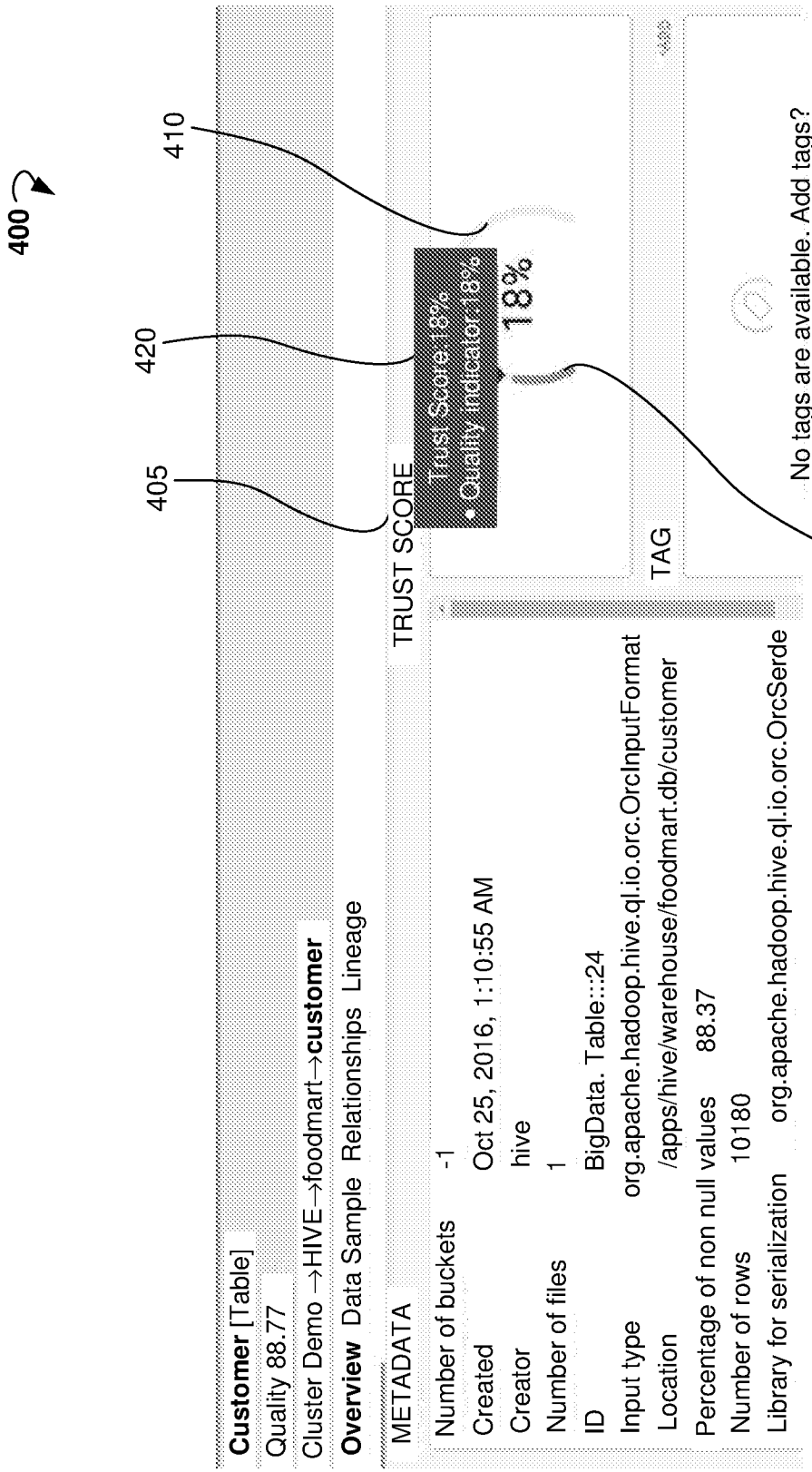
400
410
420
405
415
Customer [Table]
Quality 88.77
Cluster Demo →HIVE→foodmart→customer
Overview Data Sample Relationships Lineage
METADATA
TRUST SCORE
Trust Score 18%
• Quality indicator 18%
18%
TAG
No tags are available. Add tags?
Number of buckets -1
Created Oct 25, 2016, 1:10:55 AM
Creator hive
Number of files 1
ID BigData. Table:::24
Input type org.apache.hadoop.hive.ql.io.orc.OrcInputFormat
Location /apps/hive/warehouse/foodmart.db/customer
Percentage of non null values 88.37
Number of rows 10180
Library for serialization org.apache.hadoop.hive.ql.io.orc.OrcSerde

*FIG. 4*

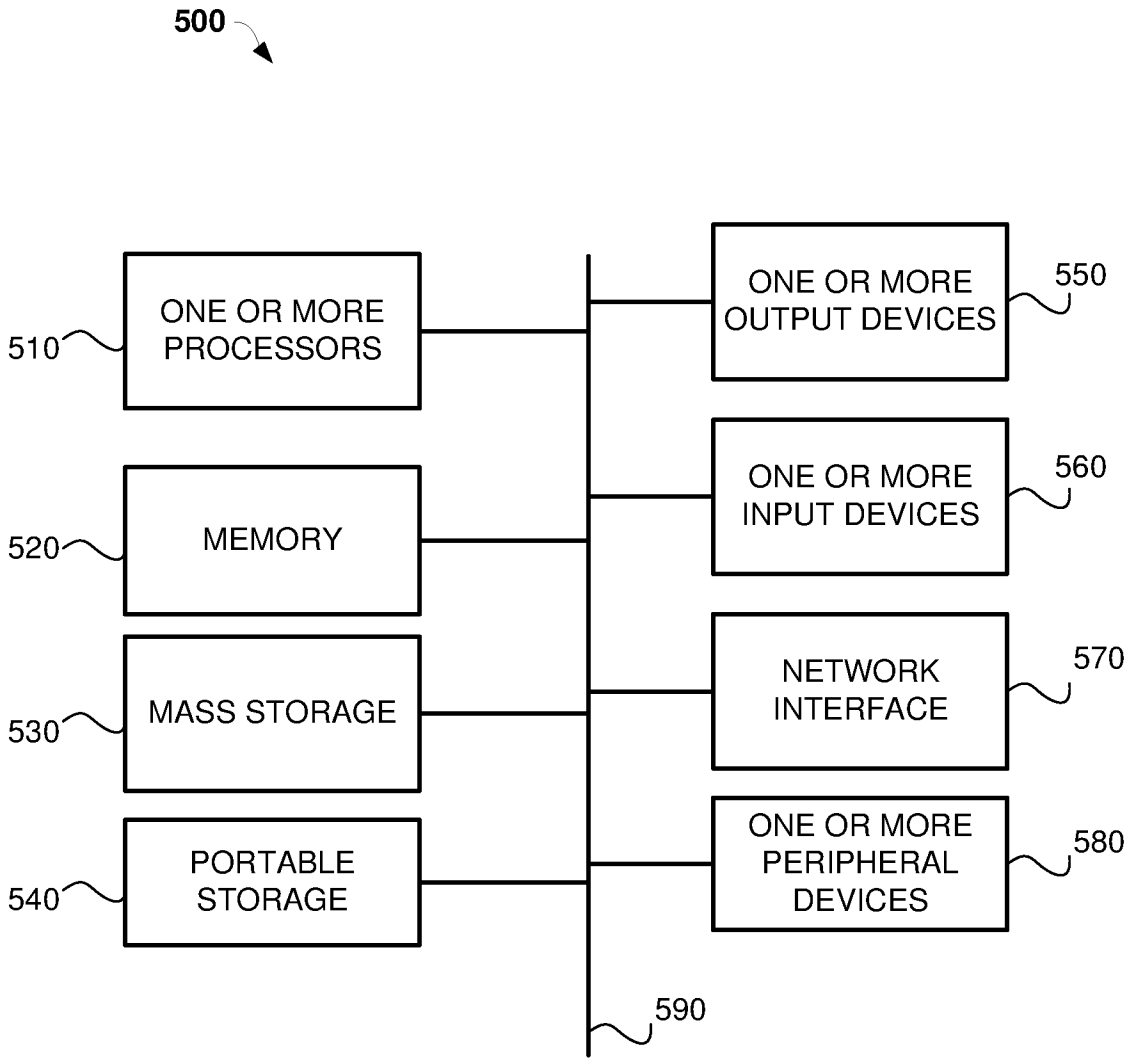
500
ONE OR MORE PROCESSORS
510
MEMORY
520
MASS STORAGE
530
PORTABLE STORAGE
540
ONE OR MORE OUTPUT DEVICES
550
ONE OR MORE INPUT DEVICES
560
NETWORK INTERFACE
570
ONE OR MORE PERIPHERAL DEVICES
580
590

FIG. 5

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2020/053419

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06F 3/048; G06F 17/30; G06Q 30/02 (2020.01)
CPC - G06Q 10/0635; G06Q 10/107; G06F 16/90348; G06F 16/9535; G06Q 30/018; G06Q 30/02 (2020.08)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
see Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
see Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
see Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | US 2010/0106560 A1 (LI et al) 29 April 2010 (29.04.2010) entire document | 1-20 |
| A | US 2016/0359711 A1 (CISCO TECHNOLOGY, INC.) 08 December 2016 (08.12.2016) entire document | 1-20 |
| A | US 2015/0127660 A1 (HERE GLOBAL B.V.) 07 May 2015 (07.05.2015) entire document | 1-20 |
| A | US 2014/0114962 A1 (LEXISNEXIS, A DIVISION OF REED ELSEVIER INC et al) 24 April 2014 (24.04.2014) entire document | 1-20 |
| A | WO 2017/147694 A1 (WWW.TRUSTSCIENCE.COM INC.) 08 September 2017 (08.09.2017) entire document | 1-20 |
| A | US 2010/0274815 A1 (VANASCO) 28 October 2010 (28.10.2010) entire document | 1-20 |
| A | WO 2015/139119 A1 (VEROSOURCE SOLUTIONS INC.) 24 September 2015 (24.09.2015) entire document | 1-20 |

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 09 December 2020 | 14 JAN 2021 |
| Name and mailing address of the ISA/US<br>Mail Stop PCT, Attn: ISA/US, Commissioner for Patents<br>P.O. Box 1450, Alexandria, VA 22313-1450<br>Facsimile No. 571-273-8300 | Authorized officer<br>Blaine R. Copenheaver<br>Telephone No. PCT Helpdesk: 571-272-4300 |