

FIG. 1

S. U. S.
CHITRA SUNDAR
IN PA/3274
BERUAR AND BERUAR LLP
AGENT OF THE APPLICANT

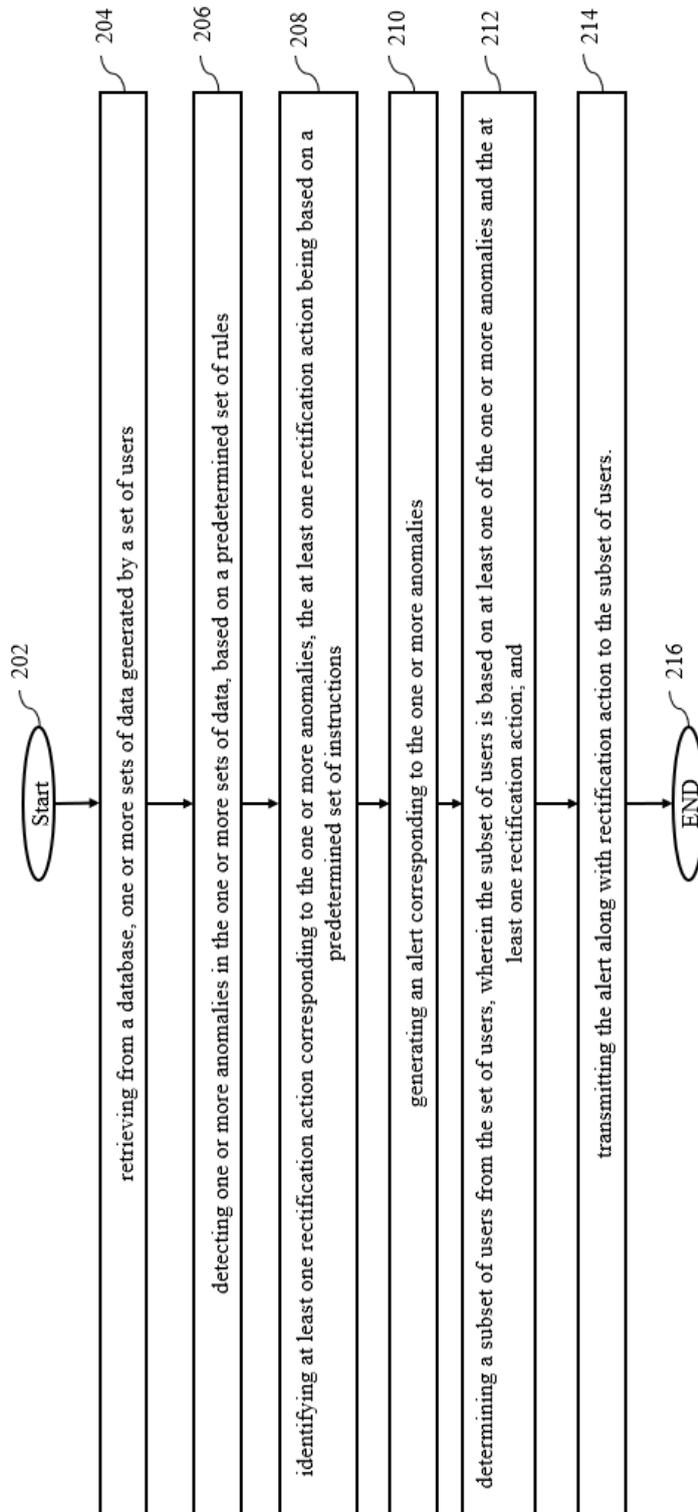


FIG. 2

S. U. S.

CHITRA SUNDAR
IN PA/3274
BERUAR AND BERUAR LLP
AGENT OF THE APPLICANT

	A	B
1	Company Name	Industry
2	家用电子制造有限公司	消费电子产品
3	奇妙玩具企业	玩具
4	快乐客户设施管理有限公司	设施管理
5	美味食品公司	食品及饮料
6	和谐音乐产品	娱乐
7		

	Company Name	Industry	Company Name_translated	Industry_translated
0	家用电子制造有限公司	消费电子产品	Home Electronics Manufacturing Co., Ltd.	Consumer Electronics
1	奇妙玩具企业	玩具	Wonderful Toys Enterprises	Toy
2	快乐客户设施管理有限公司	设施管理	Happy Customers Facility Management Ltd.	Facilities Management
3	美味食品公司	食品及饮料	Delicious Food Company	Food and Beverages
4	和谐音乐产品	娱乐	Harmony Music Products	entertainment

FIG. 3

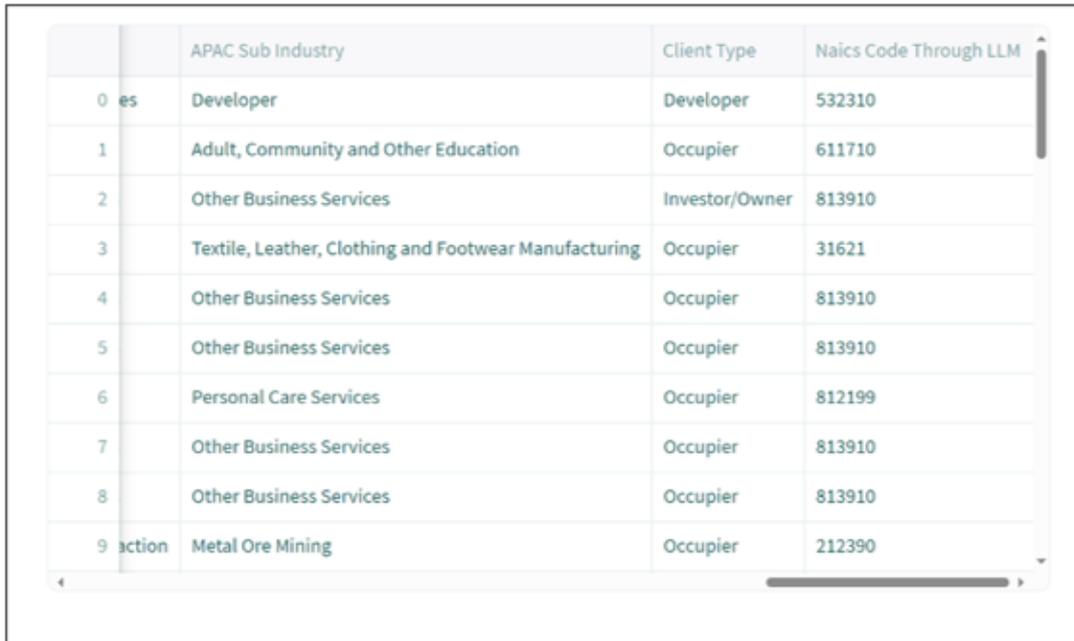
S. U. S.
CHITRA SUNDAR
IN PA/3274
BERUAR AND BERUAR LLP
AGENT OF THE APPLICANT

	A	B	C
1	Company	Name	Checked by LLM
2	Hogwarts Magic Limited	Arun Python Sharma	Male
3	Hogwarts Magic Limited	Arnav Sochbandhu	Male
4	Hogwarts Magic Limited	Dhruv Prernasarathi	Male
5	Hogwarts Magic Limited	Prithak Balwan	Male
6	Hogwarts Magic Limited	Jyestha Mitra	Female
7			

	A	B	C	D
1	Salutation	Company	Name	Checked by LLM
2	Mr.	Hogwarts Magic Limited	Arun Python Sharma	Male
3	Mr.	Hogwarts Magic Limited	Arnav Sochbandhu	Male
4	Mr.	Hogwarts Magic Limited	Dhruv Prernasarathi	Male
5	Mr.	Hogwarts Magic Limited	Prithak Balwan	Male
6	Ms.	Hogwarts Magic Limited	Jyestha Mitra	Female

FIG. 4

S. U. S.
CHITRA SUNDAR
IN PA/3274
BERUAR AND BERUAR LLP
AGENT OF THE APPLICANT



	APAC Sub Industry	Client Type	Naics Code Through LLM
0 es	Developer	Developer	532310
1	Adult, Community and Other Education	Occupier	611710
2	Other Business Services	Investor/Owner	813910
3	Textile, Leather, Clothing and Footwear Manufacturing	Occupier	31621
4	Other Business Services	Occupier	813910
5	Other Business Services	Occupier	813910
6	Personal Care Services	Occupier	812199
7	Other Business Services	Occupier	813910
8	Other Business Services	Occupier	813910
9 action	Metal Ore Mining	Occupier	212390

FIG. 5

S. U. S.
CHITRA SUNDAR
IN PA/3274
BERUAR AND BERUAR LLP
AGENT OF THE APPLICANT

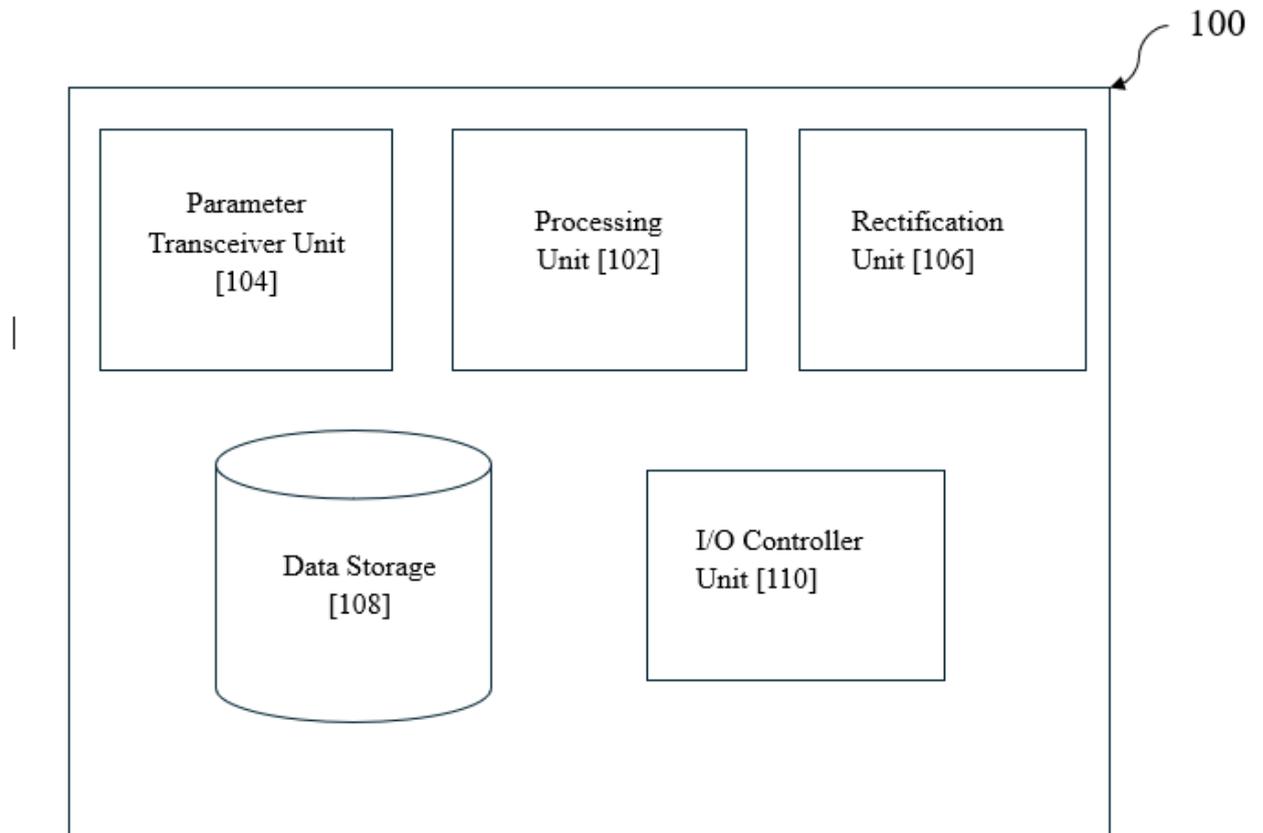


FIG. 1

S. Chitra

CHITRA SUNDAR
IN PA/3274
BERUAR AND BERUAR LLP
AGENT OF THE APPLICANT

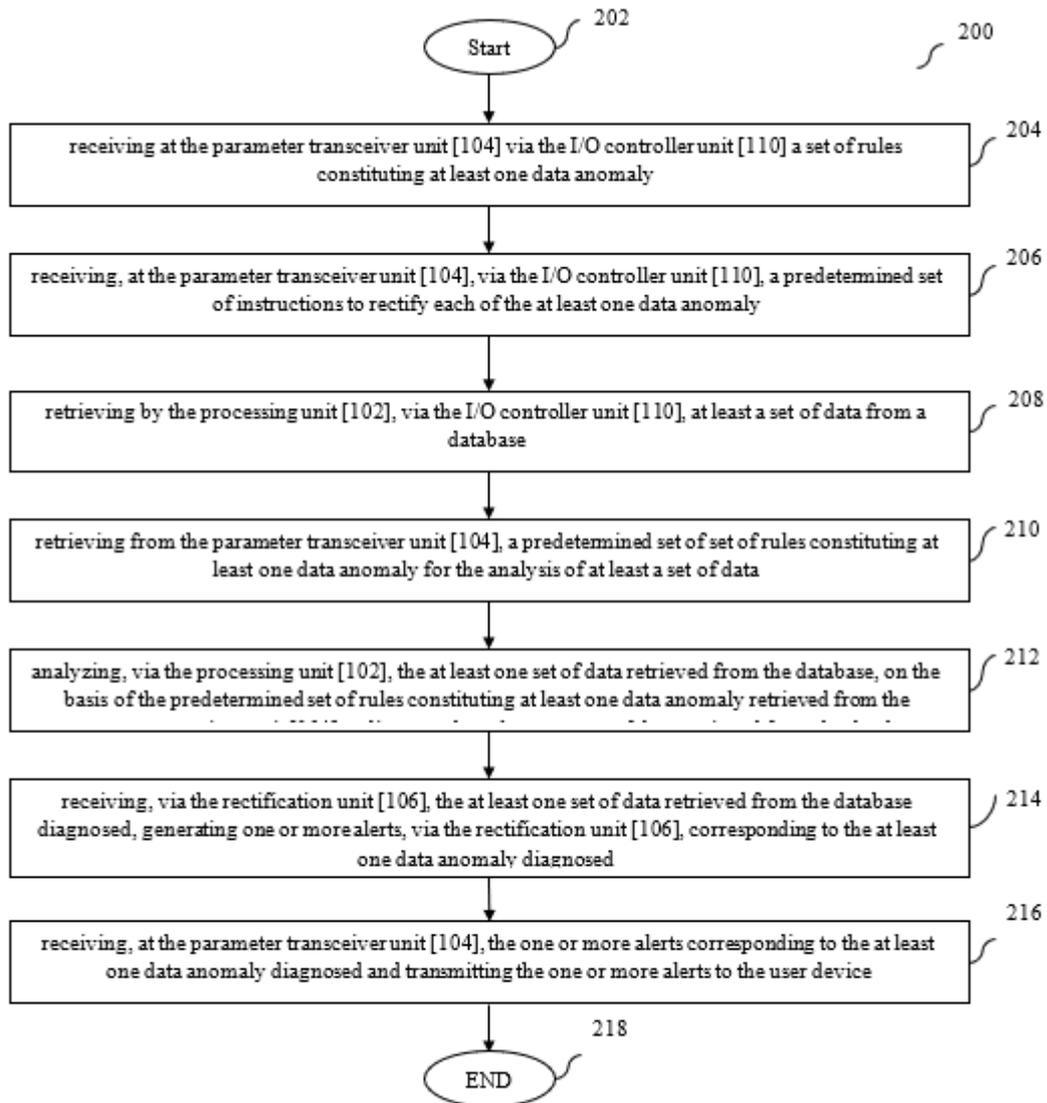


FIG. 2

g. chow

CHITRA SUNDAR
IN PA/3274
BERUAR AND BERUAR LLP
AGENT OF THE APPLICANT

300

Opportunity Name	Opportunity Name_Terminated	Sales Stage	Opportunity Priority	Location	Location_Terminated	Opportunity Investment Size	Investment
MT-Smartmore-Qinshai Kerry Center 5,500 sq.	MT-Smartmore-Qinshai Kerry Center 5,500	Cancelled-Closed	Cancelled-Closed	重庆	Qinshai	¥1,130,812	
MT-中国国际贸易中心新加坡住宅公寓 1000sqm 群	MT-China Construction International new	Cancelled-Closed	Cancelled-Closed	重庆	Qinshai	¥1,125,812	
MT-广筑设计-北科创业大厦1F 482sqm 群	MT - Regal Design - Bafu Enterprise	Cancelled-Closed	Cancelled-Closed	佛山	Nanshan	¥1,117,900	
MT -Wonderlab- 搬迁 - 2500平 - 848125 群	MT - Wonderlab - Relocation - 2500 square	Cancelled-Closed	Cancelled-Closed	佛山高新区	Heshan High-tech Park	¥1,104,741	
MT - 复星国际 商务楼 20层 9450sq - 012	MT - Fosun Yang@Qinshai Metrow 20th Fl.	Cancelled-Closed	Cancelled-Closed	重庆	Qinshai	¥1,102,500	
MT - 中铁商厦 - 4F 200sqm - 2楼中心 - 群	MT - Zhongyuan Trading 4th 200sqm - 2nd	Cancelled-Closed	Cancelled-Closed	福州亭前北	Fuzhou Muzheng North	¥1,117,900	
MT - 中国农业银行深圳龙岗大厦支行 - 2楼	MT - Agricultural Bank of China Shenzhen I	Cancelled-Closed	Cancelled-Closed	福州台江	Fuzhou TAI	¥1,100,171	
MT - 招商银行深圳康乐支行五楼 - 楼群 - 400	MT - China MinSheng Bank Shenzhen Tai	Cancelled-Closed	Cancelled-Closed	福州东台路	Fuzhou Chengde Temple	¥1,100,179	
MT - 精英文化 - 搬迁 - 175,81 648 20sqm - 群	MT - Qia Culture - Relocation - 175,81 648	Cancelled-Closed	Cancelled-Closed	佛山禅口	Nanshan Station	¥1,107,100	
MT - Dymon - 搬迁 - 3000 - 800 1200 - 群(天利)	MT - Dymon - New Best 3000 - 800 1200 -	Cancelled-Closed	Cancelled-Closed	佛山石角	Heshan Houtou	¥1,100,000	

FIG. 3

400

Country	Contact ID	Status	Unique Contact ID	Created Date	Checked by LLM
2	0034v00002vpzKZ	Correct Salutation	0034v00002vpzKZAAAY	1/1/2021	mostly_male
8	0034v00002vpzjy	Correct Salutation	0034v00002vpzjyAAA	1/1/2021	male
10	0034v00002vpznC	Correct Salutation	0034v00002vpznCAAQ	1/1/2021	mostly_male
11	0034v00002vpznD	Correct Salutation	0034v00002vpznDAAQ	1/1/2021	mostly_female
13	0034v00002vq0r5	Correct Salutation	0034v00002vq0r5AAA	1/1/2021	male
19	0034v00002vq30o	Correct Salutation	0034v00002vq30oAAA	2/1/2021	male
20	0034v00002vq31D	Correct Salutation	0034v00002vq31DAAQ	2/1/2021	male
23	0034v00002vq32p	Correct Salutation	0034v00002vq32pAAA	2/1/2021	male
28	0034v00002vq342	Correct Salutation	0034v00002vq342AAA	2/1/2021	male
29	0034v00002vq3CL	Correct Salutation	0034v00002vq3CLAAY	2/1/2021	female

FIG. 4

g. cfto
CHITRA SUNDAR
IN PA/3274
BERUAR AND BERUAR LLP
AGENT OF THE APPLICANT

	APAC Sub Industry	Client Type	Naics Code Through LLM
0 es	Developer	Developer	522310
1	Adult, Community and Other Education	Occupier	611710
2	Other Business Services	Investor/Owner	813910
3	Textile, Leather, Clothing and Footwear Manufacturing	Occupier	31621
4	Other Business Services	Occupier	813910
5	Other Business Services	Occupier	813910
6	Personal Care Services	Occupier	812199
7	Other Business Services	Occupier	813910
8	Other Business Services	Occupier	813910
9 action	Metal Ore Mining	Occupier	212390

FIG. 5

R. Chitra

CHITRA SUNDAR
IN PA/3274
BERUAR AND BERUAR LLP
AGENT OF THE APPLICANT

FORM – 2

THE PATENTS ACT, 1970

(39 OF 1970)

AND

5

THE PATENT RULES, 2003

(AS AMENDED)

COMPLETE SPECIFICATION

(SEE SECTION 10; RULE 13)

10

**“SYSTEM AND METHOD FOR ANOMALY MANAGEMENT IN A
DATABASE”**

15 **CBRE South Asia Private Limited, 6th & 7th Floor DLF Square M Block,
Jacaranda Marg DLF City Phase II, Gurgaon, Haryana, India 122002.**

*The following specification particularly describes the invention and the manner in
which it is to be performed.*

SYSTEM AND METHOD FOR ANOMALY MANAGEMENT IN A DATABASE

TECHNICAL FIELD

[0001] The invention of the present disclosure relates to the field of data diagnostics. More specifically, embodiments of the present disclosure relate to a system and method for the management of a data anomaly associated with the data
5 in a database.

BACKGROUND

[0002] This section is intended to provide information relating to the general state of the art and thus any approach/functionality described herein below should not be assumed to be qualified as prior art merely by its inclusion in this section.

10 [0003] Digital databases are used to record, store and organize large sets of data. A large number of databases are based on manual feeding of data into a database by one or more users within an organization. Naturally, when dealing with recording and organizing of large data sets in a database, there is a likelihood of certain amount of qualitative and/or quantitative errors in the process.

15 [0004] This leads to a degradation in the quality of data and use of faulty data may be counterproductive to the use-case of such data. Similarly, the presence of anomalies in a data set may lead to a misrepresentation of the data, and thereby negatively impact the outcomes of data-driven decisions in any organization.

[0005] The conventional solutions for the diagnosis and management of one or
20 more anomalies that may exist in a database involve hiring dedicated data analysts. Thereafter, a dedicated data analyst may use the tools and skills at their disposal to review all of the data sets in a database and determine one or more anomalies and/or the requisite course of action to fix the anomalies in the data. However, there are certain drawbacks to this approach.

[0006] Firstly, hiring a dedicated data analyst may not be cost-efficient or feasible for small-scale organizations. Further, an organization may separate databases for different branches or verticals at separate geographical locations, thereby prompting the need to hire multiple data analysts.

5 [0007] Secondly, it is difficult to maintain standardization in data quality across different verticals and branches due to the personalized approaches that may be followed by each analyst. Therefore, setting common standards, and ensuring coordination and implementation of such standards may incur excessive managerial burden on organizations. Differences in standards or rules for the management of
10 data across verticals and branches of an organization may also make it difficult to assess and compare the performance at a higher organizational level.

[0008] Thirdly, there is a need in the state of the art to reduce the time taken for management of anomalies in data sets. Depending on the size and nature of data sets in a database, individual data analysts may require a significant amount of time
15 to review the data, detect anomalies and a devise plan of action for the management of such anomalies.

[0009] Furthermore, with an increase in the size and complexity in data, it becomes harder to deliver alerts and instructions personally to the users that may be required to perform one or more corrective actions. Hence, there is a need for a technically
20 advanced solution to at least solve some of the above-mentioned drawbacks in the state of the art.

[0010] Moreover, with multiple sources of data, such as diversified data sets, or a plurality of databases, the use of extract, transform and load (ETL) tools becomes a necessity to collect and consolidate large sets of data from these diverse sources
25 into a single point of analysis. However, this can be problematic for data analysts, as well as to the organizations that may want to make use of the data because in some cases, some data sets may be in different language from others and therefore may not be understandable to the data analyst, or to the organization.

5 [0011] It may be noted that this section is not intended to identify an exhaustive list of drawbacks present in the state of the art, but to give a general picture of some of the major limitations associated with the state of the art, and as such, the scope of the present disclosure is not limited to the solution of the aforementioned limitations only.

OBJECTS OF THE INVENTION

[0012] This section is provided to introduce certain objects and aspects of the present disclosure in a simplified form that are further described below in the description:

10 [0013] In order to overcome at least a few problems associated with the known solutions as provided in the previous section, an object of the invention is to significantly reduce the limitations and drawbacks of the prior arts are described hereinabove.

15 [0014] An object of the invention is to provide a system and method for the management of anomalies in one or more data sets associated with a database.

[0015] Another object of the present invention is to provide a solution that enables centralised monitoring and analysis of databases spread across different verticals, branches and or geographical locations of an organization.

20 [0016] Another object of the invention is to provide a resource sparing approach to standardisation of rules and instructions for the management of anomalies in a database.

[0017] Yet another object of the present invention is to enhance the quality and accuracy of one or more data sets in a database.

25 [0018] Yet another object of the present invention is to increase cost-efficiency of data analysis and management of anomalies in a database, while reducing the time-duration of operations.

[0019] Yet another object of the present disclosure is to enable users of database to receive personalised alerts regarding one or more anomalies and requisite corrective measures in data associated with the users.

5 [0020] Yet another object of the present disclosure is to provide a solution to process and translate data sets in real time for anomaly management in a database.

SUMMARY

[0021] An aspect of the present disclosure relates to a method for anomaly management in a database. The said method comprises, retrieving from a database, one or more sets of data generated by a set of users. The method further comprises
10 the detection of one or more anomalies in the one or more sets of data on the basis of a predetermined set of rules. Thereafter, the method further comprises identifying at least one rectification action corresponding to the one or more anomalies, the at least one rectification action being based on a predetermined set of instructions and generating an alert corresponding to the one or more anomalies. The method also
15 comprises determining a subset of users from the set of users, wherein the subset of users is based on at least one of the one or more anomalies and the at least one rectification action, and then transmitting the alert to the subset of users.

[0022] In an exemplary aspect of the present disclosure, the one or more anomalies comprise at least one of one or more gaps, errors, inconsistencies, irregularities,
20 incongruities and deviations leading to a degradation in quality of data.

[0023] In an exemplary aspect of the present disclosure, the method further comprises dividing the one or more sets of data into a valid data set and an anomalous data set, based on the detection of one or more anomalies in the one or more sets of data.

25 [0024] In an exemplary aspect of the present disclosure, the predetermined set of rules comprise at least one of a minimum threshold for a set of data to qualify as a valid data and one or more parameters for detecting the one or more anomalies.

[0025] In an exemplary aspect of the present disclosure, the predetermined set of instructions comprise one or more instructions to convert an anomalous data into a valid data.

5 **[0026]** In an exemplary aspect of the present disclosure, prior to retrieving one or more sets of data from a database, the database may be connected to an extract, transform, and load (ETL) tool.

[0027] In an exemplary aspect of the present disclosure, the method further comprises converting the predetermined set of rules to one or more data flows in an extract, transform, and load (ETL) tool.

10 **[0028]** In an exemplary aspect of the present disclosure, the detection of one or more anomalies in the one or more sets of data and identification of at least one rectification action corresponding to the one or more anomalies are performed using one or more LLM models.

15 **[0029]** In an exemplary aspect of the present disclosure, the alert corresponding to the one or more anomalies comprises an individual alert for each of the users in the subset of users.

[0030] In an exemplary aspect of the present disclosure, the transmission of the alert to a subset of users is performed based on the performance of one or more actions by an admin user.

20 **[0031]** In an exemplary aspect of the present disclosure, the method further comprises displaying, via one or more user interfaces, at least one of the valid data set, the anomalous data set and the at least one rectification action.

25 **[0032]** Another aspect of the present disclosure may relate to a system for anomaly management in a database. The said system comprises a processor configured to retrieve from a database, one or more sets of data generated by a set of users and detect one or more anomalies in the one or more sets of data, based on a predetermined set of rules. The processor may be further configured identify at least

one rectification action corresponding to the one or more anomalies, the at least one rectification action being based on a predetermined set of instructions, and thereafter, generate an alert corresponding to the one or more anomalies. The processor may be further configured to determine a subset of users from the set of users, wherein the subset of users is based on at least one of the one or more anomalies and the at least one rectification action, and then transmit the alert to the subset of users.

BRIEF DESCRIPTION OF DRAWINGS

[0033] The accompanying drawings, which are incorporated herein, constitute a part of this disclosure. Components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Some drawings may indicate the components using block diagrams and may not represent the internal circuitry of each component. It will be appreciated by those skilled in the art that disclosure of such drawings includes disclosure of electrical components or circuitry commonly used to implement such components. Although exemplary connections between sub-components have been shown in the accompanying drawings, it will be appreciated by those skilled in the art, that other connections may also be possible, without departing from the scope of the invention. All sub-components within a component may be connected to each other, unless otherwise indicated.

[0034] FIG. 1 illustrates an exemplary system block diagram for anomaly management in a database, in accordance with an exemplary implementation of the present disclosure.

[0035] FIG. 2 illustrates an exemplary method flow diagram for anomaly management in a database, in accordance with an exemplary implementation of the present disclosure.

[0036] FIG. 3 illustrates an exemplary representation of an interface depicting translation done via the one or more large language models (LLMs), in accordance with exemplary implementations of the present disclosure.

5 [0037] FIG. 4 illustrates an exemplary representation of an interface depicting of a gender identification via the LLMs, in accordance with exemplary implementations of the present disclosure.

[0038] FIG. 5 illustrates an exemplary representation of an interface depicting identification of a standard industry code, in accordance with exemplary implementations of the present disclosure.

10 **DETAILED DESCRIPTION**

[0039] In the following description, for the purposes of explanation, various specific details are set forth in order to provide a thorough understanding of the embodiments of the present invention. It will be apparent, however, that embodiments of the present disclosure may be practiced without these specific
15 details. Several features described hereafter can each be used independently of one another or with any combination of other features. An individual feature may not address any of the problems discussed above or might address only some of the problems discussed above. Some of the problems discussed above might not be fully addressed by any of the features described herein. Example embodiments of
20 the present disclosure are described below, as illustrated in various drawings.

[0040] The ensuing description provides exemplary embodiments only, and is not intended to limit the scope, applicability, or configuration of the disclosure. Rather, the ensuing description of the exemplary embodiments will provide those skilled in the art with an enabling description for implementing an exemplary embodiment.
25 It should be understood that various changes may be made in the function and arrangement of elements without departing from the spirit and scope of the disclosure as set forth.

[0041] Specific details are given in the following description to provide a thorough understanding of the embodiments. However, it will be understood by one of ordinary skill in the art that the embodiments may be practiced without these specific details. For example, circuits, systems, networks, processes, and other components may be shown as components in block diagram form in order not to obscure the embodiments in unnecessary detail. In other instances, well-known circuits, processes, algorithms, structures, and techniques may be shown without unnecessary detail in order to avoid obscuring the embodiments.

[0042] Also, it is noted that individual embodiments may be described as a process which is depicted as a flowchart, a flow diagram, a data flow diagram, a structure diagram, or a block diagram. Although a flowchart may describe the operations as a sequential process, many of the operations can be performed in parallel or concurrently. In addition, the order of the operations may be re-arranged. A process is terminated when its operations are completed but could have additional steps not included in a figure.

[0043] The word “exemplary” and/or “demonstrative” is used herein to mean serving as an example, instance, or illustration. For the avoidance of doubt, the subject matter disclosed herein is not limited by such examples. In addition, any aspect or design described herein as “exemplary” and/or “demonstrative” is not necessarily to be construed as preferred or advantageous over other aspects or designs, nor is it meant to preclude equivalent exemplary structures and techniques known to those of ordinary skill in the art. Furthermore, to the extent that the terms “includes,” “has,” “contains,” and other similar words are used in either the detailed description or the claims, such terms are intended to be inclusive—in a manner similar to the term “comprising” as an open transition word—without precluding any additional or other elements.

[0044] As used herein, a “processing unit” or a “processor” includes processing unit, wherein processor refers to any logic circuitry for processing instructions. The processor may be a general-purpose processor, a special purpose processor, a

conventional processor, a digital signal processor (DSP), a plurality of microprocessors, one or more microprocessors in association with a DSP core, a controller, a microcontroller, Application Specific Integrated Circuits (ASIC), Field Programmable Gate Array circuits (FPGA), any other type of integrated circuits, etc. The processor may perform signal coding, data processing, input/output processing, and/or any other functionality that enables the working of the system according to the present disclosure. More specifically, the processor is a hardware processor.

[0045] Further, as used herein, “Storage” refers to a machine or computer-readable medium including any mechanism for storing information including but not limited to text, images, audio, and video files in a form readable by a computer or similar machine. For example, a computer-readable medium includes read-only memory (“ROM”), random access memory (“RAM”), magnetic disk storage media, optical storage media, flash memory devices or other types of machine-accessible storage media. The storage unit stores at least the data that may be required by one or more units of the server/system/user device to perform their respective functions.

[0046] Further, as used herein, “database” refers to a collection of a plurality of organized sets of data, stored electronically on a computing device, such as but not limited to, a server, wherein each of the data set may be derived from the same source, or a different source.

[0047] As discussed in the background section, the current known solutions have shortcomings. The present disclosure aims to overcome the shortcomings discussed above and other existing problems in the field of data diagnostics, specifically, for the management of data anomalies associated with the data in a database. The present disclosure provides a technically advanced system and method for anomaly management in a database that enables analysis of large data sets in a database simultaneously in significantly shorter time span as compared to the hiring of one or more data analysts. Furthermore, the present disclosure provides a technically advanced solution to streamline and standardise rules for data analysis, anomaly

detection and management across verticals and branches of an organisations while maintaining cost and time efficiency. Further, the present disclosure provides a technically advanced solution that increase data quality and data accuracy for better data-driven decision making. In part, the present disclosure enables this by providing real-time translation of data sets for consistent and accurate analysis and anomaly management. Furthermore, the present disclosure provides an advanced solution for providing generating and providing personalised alerts for data anomalies and rectification actions in response to detected anomalies.

[0048] Hereinafter, exemplary embodiments of the present disclosure will be described in detail with reference to the accompanying drawings so that those skilled in the art can easily carry out the solution provided by the current disclosure.

[0049] The invention of the present disclosure relates to anomaly management in a database, wherein the present disclosure provides a technically advanced solution for the diagnosis of anomalies in a database, and to

[0050] Referring now to FIG. 1, it illustrates an exemplary system [100] for anomaly management in a database. As illustrated, the system [100] comprises at least one of the following components, namely, a processor [102], a rectification unit [104] and a storage [106]. As used herein, all components/units of the system [100] shall be assumed to be inter-connected and working in conjunction with each other to perform anomaly management in a database, unless explicitly stated otherwise. Further, the scope of the present disclosure encompasses that the system [100] may be configured with more than one of each component/unit/module, however, only one instance of each is shown in FIG.1 for clarity and brevity.

[0051] It may be understood by a person ordinarily skilled in the art that the system [100] may comprise any other configuration of one or more components, units and/or modules as may be required to implement the features of the present disclosure, and the present description is not intended to limit the scope of the present disclosure. In one example, the system [100] may comprise only one or

more of processor [102] that may be configured to implement the features of the present disclosure independently.

[0052] Further, to perform anomaly management in a database, the system [100] may be implemented in a wide variety of computer electronic devices that may be configured to host one or more databases, or may be connected and configured to access one or more databases, such as a hosting server, a laptop, a desktop, a tablet etc. In one example, the system [100] may be directly implemented on a server on which the database exists, and therefore, the system may be configured to natively perform anomaly management in a database on the server. In another example, the system [100] may be implemented on a computing device that may be remotely connected to a hosting server/device. In one example, the computing device on which the system [100] is implemented may be connected via wired connection, such as ethernet, USB, fiber optic and the like. In another example, the connection may be a wireless connection using any wireless communication technology as may be known by a person skilled in the art, e.g., wide area networks (WAN), such as internet, 4G or 5G networks etc, or a local area network (LAN), such as Wi-Fi, Li-Fi, Bluetooth, etc.

[0053] It may be understood that the aforementioned description is only to illustrate the features of system [100] and is not intended to limit the scope of the present disclosure, and as such, the system [100] may be implemented on any computing device as may be known to a person ordinarily skilled in the art, and using any mode of communication as may be known by a person ordinarily skilled in the art.

[0054] In operation, the processor [102] of the system [100] may be configured to perform anomaly management in a database by retrieving one or more data sets from a database. The one or more data sets may comprise data generated by a set of users that are associated with the database. For example, an organization may utilize a customer relationship management (CRM) database to maintain all client data records, and their interactions with the organization. Herein, the client related information may be collected by a plurality of employees during their interactions

with one or more clients at different stages of operations of the organization. Accordingly, the employees that record and input the data sets, i.e., the customer related information into the database, i.e., the CRM database, then such employees may be the set of users that generate the one or more data sets in the database.

5 **[0055]** In an exemplary implementation, the processor [102] may be configured to retrieve the one or more data sets using an extract, transform and load (ETL) tool. In another exemplary implementation, the processor [102] may be further configured to retrieve the one or more data sets from across a plurality of databases, via the ETL tool.

10 **[0056]** For example, the one or more data sets may be related to different types and/or sources of data in a database, e.g., a database may comprise a data set relating to customer relationship management, as well as a data set related to sales operations. Alternatively, there may be multiple databases in an organization, wherein each database may correspond to a different type of data set obtained from
15 a different source. Moreover, an organization may maintain a vast database spanning across different verticals and/or branches of the organization, wherein the database contains an independent data set for each vertical/branch of the organization. Additionally, the organization may employ a separate database for any of its verticals and/or branches, each of which may further comprise one or more
20 data sets. In any of the aforementioned scenarios, it may helpful that the one or more data sets may be retrieved using an ETL tool. This enables efficient collection of the data sets from multiple sources and allows organization of the data sets for anomaly management.

[0057] In yet another exemplary implementation, wherein the one or more data sets
25 retrieved by the processor [102], are in a language other than the preferred language of operations, the processor [102] may be further configured to translate the one or more data sets into the preferred language of operation using a Large Language Model (LLM). For example, the one or more sets of data may be retrieved from five different countries to be assessed at a centralised level, and therefore the data may

exist in five different languages. Therefore, upon detecting that the one or more data sets are in a language other than the configured/preferred language of operation within the system [100], the processor [102] may translate the same using one or more LLMs.

5 **[0058]** Returning to the operation of the system [100], the processor [102] may be further configured to detect one or more anomalies in the one or more sets of data, wherein the detection action may be performed on the basis of a predetermined set of rules [108].

[0059] As used herein, the one or more anomalies may be anomalies relating to the
10 data in the one or more data sets. Accordingly, the one or more anomalies may indicate an occurrence that may be attributed to a degradation or a loss in the quality of data present in the one or more data sets. As such, one or more occurrences of at least one of the following may constitute an anomaly, namely, a gap, an error, an inconsistency, an irregularity or an incongruity within the data.

15 **[0060]** In an exemplary implementation, the processor [102] may be further configured to divide the one or more sets of data into two sets of data, namely, a valid data set and an anomalous data set. Herein, an anomalous data set may comprise a data set having one or more anomalies and a valid data set may comprise a data set with no anomalies.

20 **[0061]** Further, as used herein, the predetermined set of rules [108] may comprise one or more parameters required to detect the one or more anomalies and the minimum threshold/standards required by a data object to qualify as a valid data. Herein, the parameters may include the necessary elements required in a data object for it to be valid, and in absence of which it may be categorized as an invalid data.

25 Additionally, the parameters may include prescription on appropriate qualitative and quantitative ranges for the data objects in a data set, and/or one or more standardized formats for the entry of data objects into a set of data. Additionally, the parameter values may include any other metrics that may be known to a person

ordinarily skilled in the art for the management of anomalies in one or more data sets and setting thresholds for the validity of data.

[0062] For example, in a data set containing client names and contact information, a set of rules may direct that in a data object wherein the designation of the client is missing, or wherein the appropriate country code is missing for the associated phone number, then such a data object within the data set may be categorized as an anomaly. Accordingly, for the data object to be classified as a valid data object, it is not sufficient that the data object contains only the correct name and phone number of a client, but that it must also contain the correct designation and country code associated with the phone number of the client.

[0063] Similarly, in another example, an organization may have multiple branches across the globe, and hence users of different nationalities and backgrounds may be generating the one or more data sets in the organization's database. Let us assume that the organization is involved in financial management and the database comprises different numerical results associated with the testing of various financial models and their outcomes. When entering large numbers into the database, the certain set of users may be inclined to enter the numerical values using the Indian place-value system for commas, whereas, a different set of users may be inclined to enter the numerical values using the international place-value system for commas. Now, here the data may still represent the same numeric value, however, inconsistent standardization of the data may pose a problem in the usage and interpretation of data, especially in databases comprising a large number of complex data sets. Hence, the predetermined set of rules [108] may comprise parameters prescribing that for a data object to be considered valid, it must be entered using the Indian-place value system. Therefore, an entry that may be made using the international-place value system may technically represent the same value, however, it may be detected as an anomaly within the data set. This implementation aids in the standardization of data quality across different domains, verticals and branches, hence improving the overall data-driven decision making process.

[0064] In another exemplary implementation, the predetermined set of rules [108] may be defined by an admin user, and thereafter stored in the storage [106] of the system. Further, the processor [102] may be configured to retrieve the predetermined set of rules [108] as and when required for the detection of one or more anomalies and dividing the one or more data sets into the valid data set and the anomalous data set.

[0065] In yet another exemplary implementation, wherein the processor [102] retrieves the one or more data sets with the use of an ETL tool, the processor [102] may be further configured to convert the predetermined set of rules [108] into one or more data flows using the ETL tool, wherein the converted data flows may then be used as the basis for the detection of one or more anomalies in the retrieved data sets within the ETL tool.

[0066] Returning to the operation of the system [100], thereafter, the processor [102] may be configured to cause the rectification unit [104] to analyze the one or more anomalies detected by the processor [102]. Thereafter, the rectification unit [104] may then identify at least one rectification action required to be performed by a user to fix the one or more anomalies. Here, the at least one rectification action identified by the rectification unit [104] may be based on a predetermined set of instructions [110] that may be stored in the storage [106].

[0067] As used herein, the predetermined set of instructions [110] may comprise one or more instructions required to convert an anomalous data into a valid data, or to correct one or more anomalies present in one or more data sets. The predetermined set of instructions [110] may include a number of outcomes to be achieved in order to turn anomalous data objects within an anomalous data set, into valid data objects.

[0068] Further, a rectification action, based on the predetermined set of instructions [110] may comprise at least one step required to be performed by a user in order to achieve an outcome for converting an anomalous data object, in an anomalous data set, into a valid data object.

[0069] For the ease of understanding, continuing with one of our previous examples, let us assume that a set of users generated a set of data using international place-value system, wherein the predetermined set of rules [108] prescribed the usage of the Indian place-value system. Here, the processor [102] may detect the anomalies in the set of data and thereafter cause the rectification unit [104] to analyze the anomalies and identify at least one rectification action on the basis of the predetermined set of instructions. [110] Now, here the predetermined set of instructions [110], may comprise a corresponding outcome to turn the anomalous data set into a valid data set, and as such the outcomes may prescribe, that the anomalous data objects be changed into numbers based on the Indian place-value system. Further, here the at least one rectification action may comprise steps to achieve the outcome, and hence may contain steps for a user to access and navigate to the specific anomalous data objects within a data set, the specific corrections to be performed with respect to each of the data objects, saving any changes, unflagging the anomaly or notifying the admin user and/or the system [100] of the rectification actions performed to facilitate review etc. Therefore, it may be understood that the at least one rectification action may comprise additional steps to be performed, which may be identified on the basis of the desired outcomes laid down in the predetermined set of instructions [110].

[0070] In an exemplary implementation, the predetermined set of instructions [110] may be defined by an admin user, and thereafter, stored in the storage [106]. In another exemplary implementation, in an event wherein the predetermined set of instructions [110] do not contain an instruction to convert one or more anomalies in an anomalous data set into valid data, then rectification unit [104] may be configured to generate one or more set of instructions to convert the one or more anomalies into valid data. Thereafter, the generated set of instructions may be stored in the storage [106], in addition to the predetermined set of instructions defined by the admin user.

[0071] In another exemplary implementation, the predetermined set of instructions [110] may be wholly generated by the rectification unit [104]. In yet another

exemplary implementation, the processor [102] may perform the detection of the one or more anomalies, and cause the identification of the at least one rectification action, using at least one Large Language Models (LLMs).

5 [0072] Returning to the operation of the system [100], upon detecting the one or more anomalies and identifying the at least one rectification action, the processor [102] may be further configured to cause the rectification unit [104] to generate an alert corresponding to the one or more anomalies detected. As used herein, an alert may include a notification or an indication of the one or more anomalies detected, and at least one rectification action to rectify the one or more anomalies.

10 [0073] In an exemplary implementation, the action of generating an alert corresponding to the one or more anomalies may comprise generating an alert for each of the corresponding one or more anomalies to which the at least one rectification action may be associated to. This will be explained in further detail in subsequent paragraphs.

15 [0074] Returning to the operation of the system [100], thereafter, the processor [102] may determine a subset of users for the transmission of the generated alert, from within the set of users responsible for generating the one or more data sets. The processor [102] may determine the subset of users on the basis the one or more anomalies detected and/or the at least one rectification action.

20 [0075] It may be understood that as used herein, the subset of users may be the users responsible for causing the one or more data anomalies. In another exemplary implementation, the subset of users may also comprise a set of users other than those responsible for causing the one or more anomalies.

25 [0076] For example, the processor [102] may determine that based on the type and magnitude of anomalies within the one or more sets of data, another set of users may be better equipped to perform the at least one rectification action identified for the said one or more anomalies. Alternatively, it may also be the case that the processor [102] may determine that the subset of users responsible for causing the

one or more anomalies are occupied in the rectification of prior anomalies, or another task, the processor [102] may determine that another subset of users may have a better availability to perform the rectification action. As used herein, the determined subset of users may comprise at least one user.

5 [0077] Thereafter, the generated alert may be transmitted to the subset of users. In an exemplary implementation, the generated alert comprising individual alerts corresponding to each of the one or more anomalies may be transmitted as a single broadcast to each of the subset of users. Thereafter, the subset of users may view the alert, identify the at least one rectification action allocated to them, and the
10 corresponding anomaly, and perform the rectification of anomaly.

[0078] Alternatively, in another exemplary implementation, an individual alert corresponding to the one or more anomalies may be transmitted to each of the subset of users, wherein, each of the alerts may be personalized for the recipient, and therefore each alert may only comprise the corresponding one or more anomalies
15 and the associated rectification action(s) for the specific user from the subset of users.

[0079] In yet another exemplary implementation, the alert may be transmitted to the subset of users in response to the performance of an action by the admin user. For example, the performance of an action may comprise the admin user manually
20 selecting when to transmit the alert to each of the subset of users.

[0080] In yet another exemplary implementation, the system [100] may also transmit the at least one or more alerts to the user device in the form of one or more time-bound email notifications, wherein the user receives personalised email highlighting various information metrics such as, but not limited to, count of data
25 anomalies that the user is required to rectify, user login credentials, and comments related to data anomalies.

[0081] Additionally, the scope of the present disclosure also encompasses that that the system [100] may provide one or more graphical representations of the

outcomes of the anomaly management in a database to a admin user, user or any other person, via one or more user interfaces. The outcomes may comprise, but are not limited to, one or more details of the users of the database, one or more data anomalies detected during diagnosis, one or more corresponding instructions or requirements for the user to rectify the data anomalies, one or more comments describing the data anomaly, one or more links to data anomalies enabling easy navigation for rectification, amount of pending actions, one or more support information, and one or more summary data.

[0082] For instance, in an exemplary implementation, the one or more graphical representations may take the form of a landing page, summary page, comparative page, recognition page, data preparation tool-based flowchart page, or email notification page. Also, the one or more graphical representations may be controlled by the administrator (such as per the hierarchy in an organization, the user may be enabled to view one or more gaps on multiple levels. A junior level user may be enabled to view only his/her data gaps. Whereas a team lead user may be able to view the one or more gaps of their whole team and themselves). The landing page and summary page may have one or more key performance indicators which enable the user to quickly know the count of gaps (i.e. anomalies) in each data object (i.e. data). The one or more graphical representations may have one or more time stamps, one or more user name (such as data owner name), one or more record type (such as account, contact, lease, meeting), one or more records of anomalies, one or more list of fields such as phone number), one or more values, one or more comments (such as brief description about what exactly is wrong in a data point), one or more links and one or more support contacts (such as email address of the administrator). The comparative page may showcase one or more current data accuracy level which may be compared with one or more past data accuracy levels. The recognition page may showcase a list of users who have corrected the maximum number of data gaps (i.e. anomalies) in a past one week. The email notification page may have one or more feature such as personalization, data champions and link. Every user may receive a personalized email which

mentioned one or more counts of gaps (i.e. anomalies) and a username that the user may require to input to log into the system. A list of weekly data champions may be shared with all the users to recognize one or more top performers. Also, a link may be shared to the user and the user may directly open the link from the email.

- 5 **[0083]** The system [100] may provide one or more graphical representation of outcomes of data diagnosis relevant to a user (such as data administrator/ admin user, a business administrator, an analyst, a business manager or any other such person, to evaluate one or more metrics such as, but not limited to, one or more statistics pertaining to the quality of data, one or more statistics pertaining to the data accuracy, one or more performance metrics assessed across various verticals and locations of the business and comparative analysis of the metrics.

[0084] Additionally, the system [100] may further provide one or more graphical representation of rankings of data creators based on one or more metrics such as, but not limited to, number of data anomalies corrected within a fixed timespan.

- 15 **[0085]** Similarly, in another exemplary implementation, the alert corresponding to the one or more anomalies may direct the subset of users to one or more user interfaces, wherein, each of the user interfaces may provide a graphical representation of the contents of the alert, including information relating to the valid data set, the anomalous data set and the at least one rectification action.

- 20 **[0086]** Referring now to FIG. 2, it illustrates an exemplary flow diagram for a method [200] for anomaly management in a database. As illustrated, the method [200] begins at step [202].

[0087] At step [204], one or more sets of data that are generated by a set of users are retrieved from a database.

- 25 **[0088]** In an exemplary implementation, the method [200] may further comprise retrieving the one or more data sets using an extract, transform and load (ETL) tool. In another exemplary implementation, the method [200] may further comprise

retrieving the one or more data sets from across a plurality of databases, via the ETL tool.

[0089] At step [206], one or more anomalies may be detected in the one or more data sets on the basis of a predetermined set of rules [108].

5 **[0090]** As used herein, the one or more anomalies may be anomalies relating to the data in the one or more data sets. Accordingly, the one or more anomalies may indicate an occurrence that may be attributed to a degradation or a loss in the quality of data present in the one or more data sets. As such, one or more occurrences of at least one of the following may constitute an anomaly, namely, a gap, an error, an
10 inconsistency, an irregularity or an incongruity within the data.

[0091] In an exemplary implementation, the one or more sets of data may be divided into two sets of data, namely, a valid data set and an anomalous data set. Herein, an anomalous data set may comprise a data set having one or more anomalies and a valid data set may comprise a data set with no anomalies.

15 **[0092]** Further, as used herein, the predetermined set of rules [108] may comprise one or more parameters required to detect the one or more anomalies and the minimum threshold/standards required by a data object to qualify as a valid data. Herein, the parameters may include the necessary elements required in a data object for it to be valid, and in absence of which it may be categorized as an invalid data.
20 Additionally, the parameters may include prescription on appropriate qualitative and quantitative ranges for the data objects in a data set, and/or one or more standardized formats for the entry of data objects into a set of data. Additionally, the parameter values may include any other metrics that may be known to a person ordinarily skilled in the art for the management of anomalies in one or more data
25 sets and setting thresholds for the validity of data.

[0093] In yet another exemplary implementation, prior to retrieving the one or more data sets with the use of an ETL tool, the predetermined set of rules [108] may be converted into one or more data flows using the ETL tool, wherein the converted

data flows may then be used as the basis for the detection of one or more anomalies in the retrieved data sets within the ETL tool.

[0094] At step [208], the method [200] may comprise identifying at least one rectification action corresponding to the one or more anomalies. Here, the at least
5 one rectification action may be an action based on a predetermined set of instructions [110].

[0095] As used herein, the predetermined set of instructions [110] may comprise one or more instructions required to convert an anomalous data into a valid data, or to correct one or more anomalies present in one or more data sets. The
10 predetermined set of instructions [110] may include a number of outcomes to be achieved in order to turn anomalous data objects within an anomalous data set, into valid data objects.

[0096] Further, a rectification action, based on the predetermined set of instructions [110] may comprise at least one step required to be performed by a user in order to
15 achieve an outcome for converting an anomalous data object, in an anomalous data set, into a valid data object.

[0097] In yet another exemplary implementation, the method [200] may comprise detecting the one or more anomalies, and identifying the at least one rectification action, using at least one Large Language Models (LLMs).

[0098] At step [210], an alert corresponding to the detected one or more anomalies
20 may generated.

[0099] At step [212], the method [200] may comprise determining a subset of users from the set of users responsible for generating the one or more sets of data. Here, the determination of the subset of users maybe based upon the one or more
25 anomalies detected and the at least one rectification action identified.

[00100] At step [214], the alert may be transmitted to the determined subset of users.

[00101] In an exemplary implementation, an individual alert corresponding to the one or more anomalies may be transmitted to each of the subset of users, wherein, each of the alerts may be personalized for the recipient, and therefore each alert may only comprise the corresponding one or more anomalies and the associated rectification action(s) for the specific user from the subset of users.

[00102] In yet another exemplary implementation, the alert may be transmitted to the subset of users in response to the performance of an action by the admin user. For example, the performance of an action may comprise the admin user manually selecting when to transmit the alert to each of the subset of users.

[00103] In another exemplary implementation, the alert corresponding to the one or more anomalies may direct the subset of users to one or more user interfaces, wherein, each of the user interfaces may provide a graphical representation of the contents of the alert, including information relating to the valid data set, the anomalous data set and the at least one rectification action.

[00104] Thereafter, at step [216], the method [200] terminates.

[00105] Referring now to FIG. 3, an exemplary representation of an interface [300] depicting translation done via the one or more large language models (LLMs), in accordance with exemplary implementations of the present disclosure is depicted. The processing unit [102] may translate in real time, the set of data stored in the database in different languages via utilizing one or more large language models (LLMs). Thereafter, a translated version of the one or more sets of data may be retrieved for anomaly management.

[00106] Referring now to FIG. 4, an exemplary representation of an interface [400] depicting gender identification via the LLMs, in accordance with exemplary implementations of the present disclosure is depicted. The processing unit [102] may, via the use of one or more large language models (LLMs) identify the gender of a person in the at least a set of data, using the name of such person, wherein, the identification of gender may be used as a parameter in the segregation of the

analyzed data. For example, the identification of gender may be used to check for the use of correct salutations within the database.

[00107] Referring now to FIG. 5, an exemplary representation of an interface [500] depicting identification of a standard industry code, in accordance with exemplary implementations of the present disclosure is depicted. The system [100] may, via the processing unit [102], use at least one or more large language models (LLMs) to identify one or more standard industry codes for one or more companies present in the database, wherein, a user may use the one or more standard industry code to identify companies within the relevant industry that are present in the database.

[00108] The said system also features the capability to ingest additional variables from external systems such as Revenue Management Systems and Invoicing Applications, thereby enhancing data integration. It establishes a direct relationship between data quality and the accuracy of the revenue pipeline, demonstrating that improved data quality leads to better visibility of the revenue pipeline. Continuous data monitoring and anomaly detection further refine data trends, enabling more precise revenue relationship analysis and fostering stronger, data-driven insights.

[00109] As is evident from the paragraphs above, the present disclosure provides a technically advanced solution for anomaly management in a database. Thus, in view of the above disclosure, the method and system are cost-effective, easy to scale, easy to track, also the method and system assists the user to stay updated about one or more existing data anomalies present in the current data. Moreover, one or more personalized notification about the one or more data anomalies are shared with the users or one or more data creators. Hence, the present solution reduces the high requirement of resource associated with large databases and data diagnostics, enhances a centralized monitoring of the databases, efficiently and effectively identifies the one or more anomalies in the data.

[00110] While considerable emphasis has been placed herein on the disclosed implementations, it will be appreciated that many implementations can be made and that many changes can be made to the implementations without departing from the principles of the present disclosure. These and other changes in
5 the implementations of the present disclosure will be apparent to those skilled in the art, whereby it is to be understood that the foregoing descriptive matter to be implemented is illustrative and non-limiting.

We Claim:

1. A method [200] for anomaly management in a database, the method [200] comprising:

retrieving from a database, one or more sets of data generated by a set of users;

detecting one or more anomalies in the one or more sets of data, based on a predetermined set of rules [108];

identifying at least one rectification action corresponding to the one or more anomalies, the at least one rectification action being based on a predetermined set of instructions [110];

generating an alert corresponding to the one or more anomalies;

determining a subset of users from the set of users, wherein the subset of users is based on at least one of the one or more anomalies and the at least one rectification action; and

transmitting the alert to the subset of users.

2. The method as claimed in claim 1, wherein the one or more anomalies comprise at least one of one or more gaps, errors, inconsistencies, irregularities, incongruities and deviations leading to a degradation in quality of data.
3. The method as claimed in claim 1, wherein based on detecting one or more anomalies in the one or more sets of data, the one or more sets of data is divided into a valid data set and an anomalous data set.

4. The method as claimed in claim 1, wherein the predetermined set of rules [108] comprise at least one of a minimum threshold for a set of data to qualify as a valid data and one or more parameters for detecting the one or more anomalies.
5. The method as claimed in claim 1, wherein the predetermined set of instructions [110] comprise one or more instructions to convert an anomalous data into a valid data.
6. The method as claimed in claim 1, wherein prior to retrieving one or more sets of data from a database, the database may be connected to an extract, transform, and load (ETL) tool.
7. The method as claimed in claims 1 and 6, wherein the predetermined set of rules are converted to one or more data flows in an extract, transform, and load (ETL) tool.
8. The method as claimed in claim 1, wherein the detection of one or more anomalies in the one or more sets of data and identification of at least one rectification action corresponding to the one or more anomalies are performed using one or more LLM models.
9. The method as claimed in claim 1, wherein the alert corresponding to the one or more anomalies comprises an individual alert for each of the users in the subset of users.
10. The method as claimed in claim 1, wherein the transmission of the alert to a subset of users is performed based on the performance of one or more actions by an admin user.

11. The method as claimed in claim 1 and 3, wherein at least one of the valid data set, the anomalous data set and the at least one rectification action may be displayed via one or more user interfaces.
12. A system for anomaly management in a database, the system comprising:

a processor [102] configured to,

retrieve from a database, one or more sets of data generated by a set of users;

detect one or more anomalies in the one or more sets of data, based on a predetermined set of rules [108];

identify at least one rectification action corresponding to the one or more anomalies, the at least one rectification action being based on a predetermined set of instructions [110];

generate an alert corresponding to the one or more anomalies;

determine a subset of users from the set of users, wherein the subset of users is based on at least one of the one or more anomalies and the at least one rectification action; and

transmit the alert to the subset of users.
13. The system as claimed in claim 12, wherein the one or more anomalies comprise at least one of one or more gaps, errors, inconsistencies, irregularities, incongruities and deviations leading to a degradation in quality of data.
14. The system as claimed in claim 1, wherein based on detection of one or more anomalies in the one or more sets of data, the processor is further configured to

divide the one or more sets of data into a valid data set and an anomalous data set.

15. The system as claimed in claim 1, wherein the predetermined set of rules [108] comprise at least one of, at least one of a minimum threshold for a set of data to qualify as a valid data and one or more parameters to detect the one or more anomalies.
16. The system as claimed in claim 1, wherein the predetermined set of instructions [110] comprise one or more instructions to convert an anomalous data into a valid data.
17. The system as claimed in claim 1, wherein prior to the retrieval of the one or more sets of data from a database, the database may be connected to an extract, transform, and load (ETL) tool.
18. The system as claimed in claims 12 and 17, wherein the predetermined set of rules are converted to one or more data flows in an extract, transform, and load (ETL) tool.
19. The system as claimed in claim 1, wherein the processor is further configured to detect the one or more anomalies in the one or more sets of data and identify the at least one rectification action corresponding to the one or more anomalies using one or more LLM models.
20. The method as claimed in claim 1, wherein the alert corresponding to the one or more anomalies comprises an individual alert for each of the users in the subset of users.

21. The method as claimed in claim 1, wherein the transmission of the alert to a subset of users is performed based on the performance of one or more actions by an admin user.

22. The system as claimed in claim 12, wherein the processor is further configured to display, at least one of the valid data set, the anomalous data set and the at least one rectification action, via one or more user interfaces.

Dated this 31st day of May, 2024

S. U. S.
CHITRA SUNDAR

IN-PA/3274

**AGENT FOR THE APPLICANT(S)
OF BERUAR & BERUAR LLP**

ABSTRACT

SYSTEM AND METHOD FOR ANOMALY MANAGEMENT IN A DATABASE

A method for anomaly management in a database, the method comprising retrieving from a database, one or more sets of data generated by a set of users and detecting one or more anomalies in the one or more sets of data, based on a predetermined set of rules [108]. The method further comprises identifying at least one rectification action corresponding to the one or more anomalies, the at least one rectification action being based on a predetermined set of instructions [110]. The method further comprising generating an alert corresponding to the one or more anomalies, then determining a subset of users from the set of users, wherein the subset of users is based on at least one of the one or more anomalies and the at least one rectification action and transmitting the alert to the subset of users.

Reference: FIG. 2