



(51) International Patent Classification:

G06F 17/00 (2019.01)

(21) International Application Number:

PCT/US2020/034704

(22) International Filing Date:

27 May 2020 (27.05.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

16/423,456

28 May 2019 (28.05.2019)

US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:

US

16/423,456 (CON)

Filed on

28 May 2019 (28.05.2019)

(71) Applicant: JPMORGAN CHASE BANK, N.A. [US/US];

383 Madison Avenue, New York, New York 10179 (US).

(72) Inventors: YAZDI, Matthew; c/o JPMorgan Chase Bank,

N.A., 383 Madison Avenue, New York, New York 10179

(US). DOSHI, Manish; c/o JPMorgan Chase Bank, N.A.,

383 Madison Avenue, New York, New York 10179 (US).

POLADIAN, Grant; c/o JPMorgan Chase Bank, N.A.,

383 Madison Avenue, New York, New York 10179 (US).

KATAKDOUN, Sachin, Narhari; c/o JPMorgan Chase

Bank, N.A., 383 Madison Avenue, New York, New York

10179 (US).

(74) Agent: MUELLER, Robert, W. et al.; Greenblum & Bern-

stein, P.L.C., 1950 Roland Clarke Place, Reston, Virginia

20191 (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,

(54) Title: A MECHANISM FOR REPLICATION OF A DATA IN HADOOP FROM A PRIVATE NETWORK TO A PUBLIC CLOUD INSTANCE

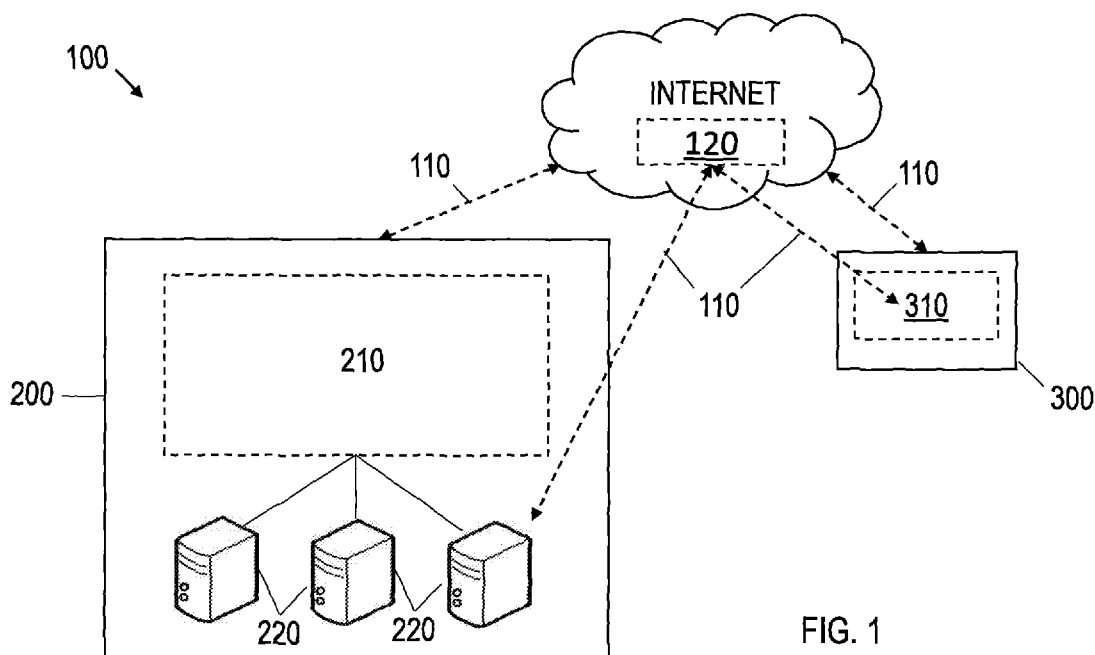


FIG. 1

(57) Abstract: Systems and methods for populating data in a client specific repository in a public cloud infrastructure is described. The system includes a private master data repository (MDR) implemented in a private network and having data stored therein. A client submits a data request to access the MDR and to perform an operation on data within the MDR. The data request may include information for identifying the client and information for determining whether the client is authorized to perform the requested operation. The system further includes a persistence mechanism configured to interface with the MDR in response to the data request. Upon verifying the client is authorized to access the MDR, the persistence mechanism may be configured to provision a directory in the public cloud infrastructure and to create and populate data in the client specific repository as a persistent state of the data in the MDR.

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, 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).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

- *with international search report (Art. 21(3))*

A MECHANISM FOR REPLICATION OF A DATA IN HADOOP FROM A PRIVATE
NETWORK TO A PUBLIC CLOUD INSTANCE

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority from U.S. Patent Application No. 16/423,456, filed May 28, 2019, which is herein incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] The present disclosure relates generally to a mechanism for populating data in a public cloud infrastructure from a private network and methods of using the same.

BACKGROUND

[0003] Large organizations are often tasked with the management of large amounts of data requiring secure and effective storage and retrieval. In some situations, organizations act as custodians of data for one or more clients. Certain data, such as confidential or highly confidential data, requires storage that is kept secure and isolated from networked connection so as to avoid any possibility of breach. Whenever a client requires access to this secure data, the organization must be able to effectively and securely provide access to this data.

[0004] Data repositories, for example, master data repositories such as a Hadoop data repository, may be used to hold large sets of data for multiple subscribers within an organization's private network. However, these solutions are typically only accessible to those within the organization's network, rendering it difficult and/or impossible for the organization to provide the secure data to a client without compromising the scalability and the security of its network and/or data. Additionally, traditional mechanisms for providing access to clients for client specific data is accomplished, for example, via File Transfer Protocols (FTP) for accessing the data files, and/or direct download by the client for the client specific

data. Unfortunately, each of these mechanisms have limitations related to scalability, and/or the ability to provide real-time and/or near real-time access to data.

SUMMARY

[0005] In an exemplary embodiment, a computer implemented method for populating data in a client specific repository in a public cloud infrastructure from a Hadoop master data repository in a private network is provided. The method includes the step of receiving a data request from a client to perform an operation on data in the Hadoop master data repository. The data request includes information for identifying the client and a data of the data in the Hadoop master data repository. The method further includes identifying the client from the data request and verifying that the client is authorized to access the data. The method also includes establishing a secure connection with the public cloud infrastructure upon verifying that the client is authorized to access the data. Additionally, the method includes provisioning the public cloud infrastructure to create the client specific repository in the public cloud infrastructure. The method further includes accessing the data identified in the data request. The method also includes executing a redundancy operation on the data to create a client instance of the data in a persistent state within the client specific repository, and providing access to the data in the persistent state to the client.

[0006] These and other objects, features, and advantages of the present disclosure will become apparent from the following detailed description of illustrative embodiments thereof, which is to be read in connection with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] These and other features of the present disclosure will become better understood with regard to the following description and accompanying drawings in which:

[0008] *Fig. 1* illustrates an exemplary embodiment of a system in accordance with the disclosure provided herein;

- [0009] *Fig. 2* illustrates an exemplary embodiment of a Hadoop data repository in a private network in accordance with the disclosure provided herein;
- [0010] *Fig. 3* illustrates an exemplary embodiment of a client specific repository in a public cloud infrastructure in accordance with the disclosure provided herein; and
- [0011] *Fig. 4* illustrates a flowchart for an exemplary embodiment of a method for implementing a persistence mechanism in accordance with the disclosure provided herein.

DETAILED DESCRIPTION

- [0012] Aspects and implementations of the present disclosure will be understood more fully from the detailed description given below and from the accompanying drawings of the various aspects and implementations of the disclosure. This should not be taken to limit the disclosure to the specific aspects or implementations, but explanation and understanding only.
- [0013] Referring now to the drawings, which are for purposes of illustrating exemplary embodiments of the subject matter herein only and not for limiting the same, *Fig. 1* shows an exemplary embodiment of a system 100 for populating data in a public cloud infrastructure.
- [0014] An embodiment of the present disclosure is directed to a persistence mechanism operable to separate the processing of data, for example, within a data repository (e.g., a Hadoop data repository), from the persistence of the data, such that creating an instance of the data may be routed through the persistence mechanism in order to replicate the data in the Hadoop data repository in a private network to a client specific repository implemented in a public cloud infrastructure. It should be appreciated that references to Hadoop data repositories herein are for exemplary purposes only, as additional data repositories and/or master data repositories may utilize the system 100 and/or be utilized within the system 100 in a manner consistent with the reference Hadoop data repository.
- [0015] As shown in *Fig. 1*, the system 100 may include at least a private network 200 in operable communication with a public cloud infrastructure 300 via a

communication link 110. The communication link 110 may be a direct communication link, a remote communications link, or any communication link known to persons of ordinary skill in the art and configurable to allow for communication and/or interfacing between the devices and/or components of the system 100. Examples of such communication links 110 may include Local Area Networks (LAN), Wide Area Networks (WAN), and Global Area Networks (GAN) having wire or wireless branches, including, for example, the Internet. In some embodiments, communications with the public cloud infrastructure 300 may include connections via a secured connection portal 120. The secured connection portal 120 may be similar to the communication link 110 in that it may be configured to allow for a secure communication and/or interfacing between the devices and/or components within the system 100.

[0016] The system 100 may include a Hadoop data repository 210 implemented in the private network 200 for providing access to data or a subset of data within the Hadoop data repository 210 to one or more data subscribers (also referred to herein as clients) 220 within the private network 200.

[0017] The client 220 may be an individual authorized to access the Hadoop data repository 210, or in yet a further embodiment, a computer or similar device under the control of an individual authorized to access the Hadoop data repository 210 within the private network 200. It should be appreciated that the client 220 may be remote from the Hadoop data repository 210, i.e., outside of the private network 200 (*Fig. 2*), and may be able to access the Hadoop data repository 210 within the private network 200 via the secure connection portal 120. *Fig. 1* illustrates an example of the client 220 connecting to the public cloud infrastructure 300 via the secure connection portal 120.

[0018] With continued reference to the figures, and now with additional reference to *Fig. 2*, an exemplary embodiment of the private network 200 with Hadoop data repository 210 is provided. As shown in *Fig. 2*, the private network 200 may include a client data manager 230 operably configured to interface with one or more clients 220 and the Hadoop data repository 210.

[0019] The client data manager 230 may be configured to monitor activity, for example, at the client device, within the private network 200 and between one or more clients 220 and the Hadoop data repository 210. In some embodiments, the client data manager 230 may be configured to determine whether the client is authorized to access data in the Hadoop data repository 210. The determination may be based on comparing client credentials, for example, provided by the client, to a client profile and/or an account database with authenticating information stored thereon. The client data manager 230 may access the client profile and/or the account database to compare the provided credentials to information within the client profile for determining whether the client is authorized to access data in Hadoop data repository 210.

[0020] Additionally, the client data manager 230 may be configured to deny the client access to data in the Hadoop data repository 210 upon determining that the client is not authorized, for example, based on the provided credentials. It should be appreciated that the client profile may be stored within the private network 200, and/or outside the private network 200, for example, via a data source in communication with the client data manager 230.

[0021] Additionally, or alternatively, the client data manager 230 may be configured to determine whether the client is permitted to create its own repository in the public cloud infrastructure 300. It should be appreciated that in some embodiments, determining whether the client is permitted to create its own specific repository in the public cloud infrastructure 300 may be established by the client data manager 230 when determining whether access to the Hadoop data repository 210 should be allowed and/or denied.

[0022] Additionally, or alternatively, the client data manager 230 may be configured to determine whether the client is authorized to access specific data sets within the Hadoop data repository 210. For example, the client data manager 230 may access one or more business rules to verify that the client is a holder of a specific data set and that the client is authorized to receive and/or otherwise access data for the specific data set. In some embodiments, in addition to verifying access to specific data sets, the client data manager 230 may be configured to determine whether the

client is authorized to access individual data points, for example, within the specific data sets. To determine whether the client is authorized to access these data points, the client data manager 230 may be configured to access one or more data point attributes for the data points, which may include information that uniquely identifies the client's ownership of the datapoints and/or associated data sets.

[0023] Additionally or alternatively, upon authenticating the client, the client data manager 230 may further be configured to initiate the provisioning process in the public cloud infrastructure 300 to ensure the health and performance of any client specific repositories or client instances in the public cloud infrastructure 300. In some embodiments, for example, to determine the health and/or performance of one or more repositories and/or instances in the public cloud infrastructure 300, the client data manager 230 may be configured to monitor the status of one or more components, for example, forming the public cloud infrastructure 300. For example, these components may include any hardware and/or software components inside the public cloud infrastructure 300.

[0024] Additionally, or alternatively, the client data manager 230 may be configured to dynamically structure a system for the repositories and/or instances, for example, via one or more tools provided by the public cloud infrastructure 300. In this embodiment, for example, the client data manager 230 may be configured to capture a snapshot of any components of the public cloud infrastructure 300, and should one or more components become unstable, for example, due to corruption and/or changes in configurations, restore the unstable component using the captured snapshot from, for example, a stable version of the unstable component and/or a component similarly configured to the unstable component.

[0025] In some embodiments, the client data manager 230 may further be configured to apply updates and/or fixes, for example, critical updates, to any components of the public cloud infrastructure, for example, during monitoring of the public cloud infrastructure 300 and/or provisioning of the repositories and/or instances.

[0026] Additionally or alternatively, the system 100 may include a client data router, for example, within and/or in operable communication with the private network 200. The client data router may be configured to implement one or more business rules,

configurations, and/or subscriptions to determine which client record is being requested in the data request. When a request to access data is identified, for example, the client data router may be configured to access and/or deliver any client specific data directly into public cloud infrastructure 300, for example, upon authenticating the client. It should be appreciated that the client data router may work with the client data manager 230, for example, upon identifying the client specific data, to confirm that the client is authorized to work in the public cloud infrastructure 300 environment.

[0027] With continued reference to the figures, and now with additional reference to *Fig. 3*, an exemplary embodiment of the client specific repository 310 in a public cloud infrastructure 300 is provided. In one exemplary embodiment, the public cloud infrastructure 300 may include a client data processor 330 (also referred to as a persistence mechanism) operably configured to receive data D1 from the Hadoop data repository 210, and to populate the received data D1 in the client specific repository 310 in the public cloud infrastructure 300. The client data processor 330 may access the data in real-time or near real-time. In some embodiments, updates to the data may be applied, via the client data processor 330, in an ongoing basis creating a persistent state 320 of the data from the Hadoop data repository 210 in the client specific repository 310 in real-time or near real-time. In other embodiments, updates may be applied in a batch operation, for example, via the client data processor 330.

[0028] With continued reference to the figures, and now with additional reference to *Fig. 4*, a method 400 for populating data in a client specific repository 310 in a public cloud infrastructure 300 from a Hadoop data repository is provided.

[0029] It should be appreciated that the method 400 may be performed in a different order, with illustrated steps omitted, with additional steps added, or with a combination of reordered, combined, omitted, or additional steps.

[0030] In step 410, the method 400 may include the step of receiving a data request from a client 220 to perform an operation on data in the Hadoop data repository 210. In this step, the client 220 may submit the data request, for example, via a user

interface, to perform an operation, for example, add, remove, or update, on data or a subset of data in the Hadoop data repository 210.

[0031] In step 420, the method 400 may include the step of identifying the client 220 from the data request and verifying that the client is authorized to access at least a subset of the data (also referred to as a dataset). In this step, the submitted data request may include client information for identifying the client and/or permissions for the client.

[0032] The permissions may be provided to determine if the client is authorized to access the Hadoop data repository 210, and/or authorized to perform the operation specified in the data request. Additionally or alternatively, the permission may be provided to determine whether the client may be authorized to establish the client specific repository 310 in the public cloud infrastructure 300.

[0033] Upon identifying the client 220 and verifying that the client 220 is authorized to access and/or perform the requested operation on data in the Hadoop data repository 210, the method 400 includes the step 430 of establishing a secure connection with a public cloud infrastructure 300. In this step, a connection to the public cloud infrastructure 300 may be established so that the client data processor 330 may create the client specific repository 310 in the public cloud infrastructure 300. In some embodiments, the client data manager 230 may be configured to determine whether a connection to the components of the public cloud infrastructure 300 is authorized, and if authorized, the client data manager 230 may assist in establishing the secured communication link.

[0034] In step 440, upon establishing a connection to the public cloud infrastructure 300, the public cloud infrastructure 300 may be provisioned to create the client specific repository/directory 310 in the public cloud infrastructure 300.

[0035] In some embodiments, for example, the client data processor 330 may access one or more seed markers, via the client data router, to create the client specific repository 310. The seed markers may include and/or otherwise define various parameters, for example, for establishing structures within the public cloud infrastructure 300. For example, the client data processor 330 may be configured to interpret the information provided via one or more of the seed markers, and to

establish data structures and/or granular data holders within the public cloud infrastructure 300, based on the accessed seed markers, and as desired or otherwise needed for the provisioning process and/or for creating the client specific repository 310.

[0036] Additionally, or alternatively, the information provided by the seed markers may also be utilized for updating and/or removing any existing data structures and/or granular data holders, for example, to accommodate revised data points, which may be submitted, for example, via client data manager 230. In yet a further exemplary embodiment, the information provided by the seed markers may be used for various administrative activities such as, for example, limiting or otherwise restricting the size of the client specific repository 310, optimizing the client specific repository 310 and/or miscellaneous clean-up activities.

[0037] Upon provisioning the public cloud infrastructure 300, in step 450, the data identified in the data request may be accessed in the Hadoop data repository 210 for performing the requested operation on the data in the Hadoop data repository 210. In some embodiments, the client data router may transmit any updates to the client specific data in the public cloud infrastructure 300. That is, the updates to the client specific data may occur in the public cloud infrastructure 300.

[0038] Additionally, or alternatively, any updates to the client specific data may occur in the Hadoop data repository 210. In this embodiment, for example, the client data router may provide information identifying that an update to the client specific data is available and/or has been made in the Hadoop data repository 210, and the client data processor 330 may be configured to interpret the provided information from the client data router, and update the client specific data in the client specific repository 310 based on the interpreted information.

[0039] Once the data has been updated in the Hadoop data repository 210 and/or the client specific repository 310, in step 460, the method includes the step of executing a redundancy operation on the updated data, for example, to create a client instance of the data in a persistent state 320.

[0040] In some embodiments, for example, the persistent state 320 of the data may be created in the Hadoop data repository 210 and/or the client specific repository 310

based upon where the requested operation is performed. For example, if the operation was performed in the Hadoop data repository 210, then the client instance may be created in the client specific repository 310 in the public cloud infrastructure 300.

[0041] Upon creation of the client instance, in step 470, the method includes the step of providing access to the data in the persistent state 320 to the client 220. It should be appreciated that upon creation of the client specific repository 310, the client 220 may access the client specific data directly from the client specific repository 310 instead of the Hadoop data repository 210, for example, for any future operations on the client specific data. Additionally, or alternately, once communication between the Hadoop data repository 210 and client specific repository 310 is formalized and/or initialized, all subsequent updates, for example, to the client data, may be recorded to both Hadoop data repository 210 and the client specific repository 310 in real-time or near real-time.

[0042] The term “module” or “engine” used herein will be appreciated as comprising various configurations of computer hardware and/or software implemented to perform operations. In some embodiments, modules, mechanisms, or engines as described may be represented as instructions operable to be executed by a processor and a memory. In other embodiments, modules or engines as described may be represented as instructions read or executed from a computer readable media. A module or engine may be generated according to application specific parameters or user settings. It will be appreciated by those of skill in the art that such configurations of hardware and software may vary, but remain operable in substantially similar ways.

[0043] It is to be understood that the detailed description is intended to be illustrative, and not limiting to the embodiments described. Other embodiments will be apparent to those of skill in the art upon reading and understanding the above description. Moreover, in some instances, elements described with one embodiment may be readily adapted for use with other embodiments. Therefore, the methods and systems described herein are not limited to the specific details, the representative embodiments, or the illustrative examples shown and described. Accordingly,

departures may be made from such details without departing from the spirit or scope of the general aspects of the present disclosure.

[0044] It should further be understood that the terms "comprising", "including", and "having", are open-ended and does not exclude other elements or steps; and the use of articles "a" or "an" does not exclude a plurality. Accordingly, the particular arrangements disclosed are meant to be illustrative only and should not be construed as limiting the scope of the claims or disclosure, which are to be given the full breadth of the appended claims, and any and all equivalents thereof.

CLAIMS

We claim:

1. A computer implemented method for populating data in a public cloud infrastructure comprising the steps of:
 - accessing a data request via a client device to perform an operation on data in a master data repository in a private network, wherein the data request includes information for identifying the data in the master data repository;
 - connecting to the public cloud infrastructure;
 - creating a client specific repository in the public cloud infrastructure;
 - creating a client instance of the data in a persistent state within the client specific repository; and
 - providing access to the client instance of the data.
2. The method of claim 1, further comprising:
 - prior to creating the client specific repository, identifying a client from the data request; and
 - verifying that the client is authorized to perform the requested operation by:
 - accessing a client profile corresponding to the client to identify one or more client permissions; and
 - identifying client permissions granting access to the data and authorizing the creation of the client specific repository in the public cloud infrastructure.
3. The method of claim 1, further comprising:
 - prior to creating the client specific repository, provisioning the public cloud infrastructure by:
 - determining a health of the public cloud infrastructure; and
 - upon determining that the public cloud infrastructure is healthy, creating the client specific repository.

4. The method of claim 3, wherein determining the health of the public cloud infrastructure comprises:
 - identifying one or more components forming the public cloud infrastructure;
 - and
 - monitoring the one or more components to identify a stable component for creating the client specific repository.
5. The method of claim 4, further comprising:
 - upon identifying a stable component, creating a snapshot of the stable component for subsequently restoring unstable versions of the stable component or similarly configured components based on the snapshot.
6. The method of claim 1, further comprising:
 - accessing an updated data request to perform an additional operation on the data;
 - accessing the client specific repository in the public cloud infrastructure; and
 - providing the client instance for performing the additional operation on the client instance.
7. The method of claim 6, further comprising:
 - upon the performance of the additional operation on the client instance, executing a redundancy operation on the client instance in the client specific repository to update the data in the master data repository based on updates to the client instance from the additional operation such that the data in the master data repository replicates the client instance in the client specific repository.
8. A system for populating data in a public cloud infrastructure comprising:
 - a memory including instructions for populating data stored thereon;
 - a processor in communication with the memory and configured to execute the instructions to perform the following operations:

accessing a data request to perform an operation on data in a master data repository, wherein the data request includes information for identifying the data in the master data repository;

connecting to the public cloud infrastructure;

creating a client specific repository in the public cloud infrastructure;

creating a client instance of the data in a persistent state within the client specific repository; and

providing access to the client instance of the data.

9. The system of claim 8, wherein the processor is further configured to execute the instructions to perform the following operation prior to creating the client specific repository:

identifying a client based on the data request; and

verifying that the client is authorized to perform the requested operation by:

accessing a client profile corresponding to the client to identify permissions for the client; and

identifying client permissions granting access to the data and authorizing the creation of the client specific repository in the public cloud infrastructure.

10. The system of claim 8, wherein the processor is further configured to execute the instructions to perform the following operation prior to creating the client specific repository:

provisioning the public cloud infrastructure by:

determining a health of the public cloud infrastructure; and

upon determining that the public cloud infrastructure is healthy, creating the client specific repository in the public cloud infrastructure.

11. The system of claim 10, wherein determining the health of the public cloud infrastructure comprises:

identifying one or more components of the public cloud infrastructure; and

monitoring the one or more components to identify a healthy component for creating the client specific repository.

12. The system of claim 11, wherein the processor is further configured to execute the instructions to perform the following operation upon identifying the healthy component:

creating a snapshot of the healthy component for subsequently restoring one or more unhealthy components based on the snapshot.

13. The system of claim 8, wherein the processor is further configured to execute the instructions to perform the following operation:

accessing an updated data request to perform an additional operation on the data;

accessing the client specific repository in the public cloud infrastructure; and

providing the client instance for performing the additional operation on the client instance.

14. The system of claim 13, wherein the processor is further configured to execute the instructions to perform the following operation upon the performance of the additional operation on the client instance:

executing a redundancy operation on the client instance in the client specific repository to update the data in the master data repository based on updates to the client instance during the additional operation such that the data in the master data repository replicates the client instance in the client specific repository.

15. A non-transitory computer readable medium comprising instructions for populating data in a public cloud infrastructure that, when executed by a processor, causes the processor to perform the following operation:

accessing a data request via a client device to perform an operation on data in a private master data repository, wherein the data request includes information for identifying the data in the private master data repository;

connecting to the public cloud infrastructure;

- creating a client specific repository in the public cloud infrastructure;
 - creating a persistent state of the data identified in the data request within the client specific repository; and
 - providing access to the persistent state of the data.
16. The non-transitory computer readable medium of claim 15, further comprising instructions, that, when executed by the processor, causes the processor to perform the following operation:
- prior to creating the client specific repository, identifying a client based on the data request; and
 - verifying that the client is authorized to access the data.
17. The non-transitory computer readable medium of claim 15, further comprising instructions, that, when executed by the processor, causes the processor to perform the following operation:
- provisioning the public cloud infrastructure prior to creating the client specific repository.
18. The non-transitory computer readable medium of claim 17, wherein provisioning the public cloud infrastructure comprises:
- determining a health of the public cloud infrastructure; and
 - upon determining that the public cloud infrastructure is healthy, creating the client specific repository.
19. The non-transitory computer readable medium of claim 15, further comprising instructions, that, when executed by the processor, causes the processor to perform the following operation:
- accessing an updated data request to perform an additional operation on the data;
 - accessing the client specific repository in the public cloud infrastructure; and
 - providing the persistent state of the data for performing the additional operation on the persistent state of the data.

20. The non-transitory computer readable medium of claim 19, further comprising instructions, that, when executed by the processor, causes the processor to perform the following operation:

upon performance of the additional on the persistent state of the data in the client specific repository, executing a redundancy operation on the persistent state of the data in the client specific repository to update the data in the private master data repository based on updates to the persistent state of the data such that the data in the private master data repository replicates the persistent state of the data in the client specific repository.

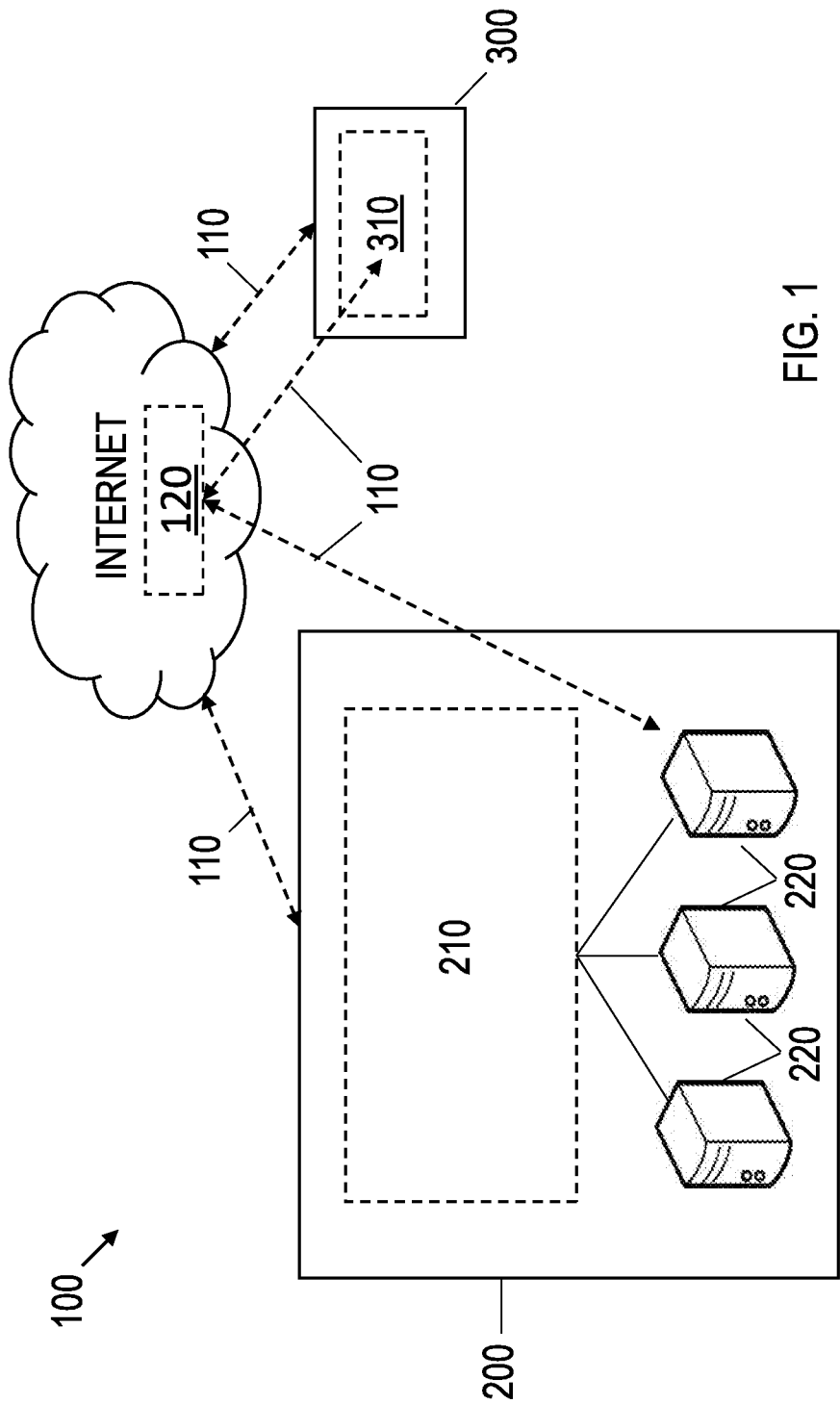


FIG. 1

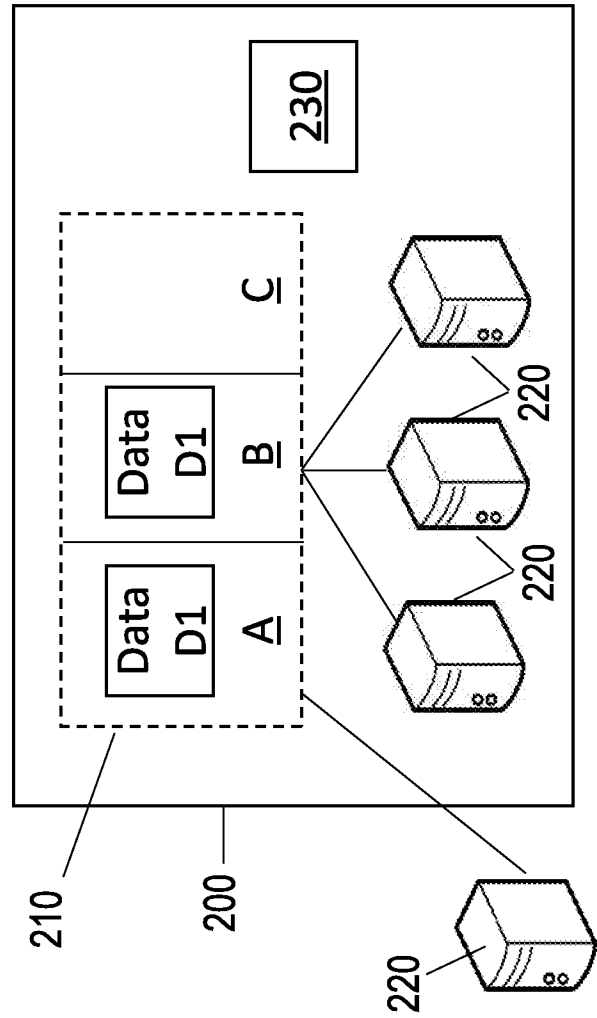


FIG. 2

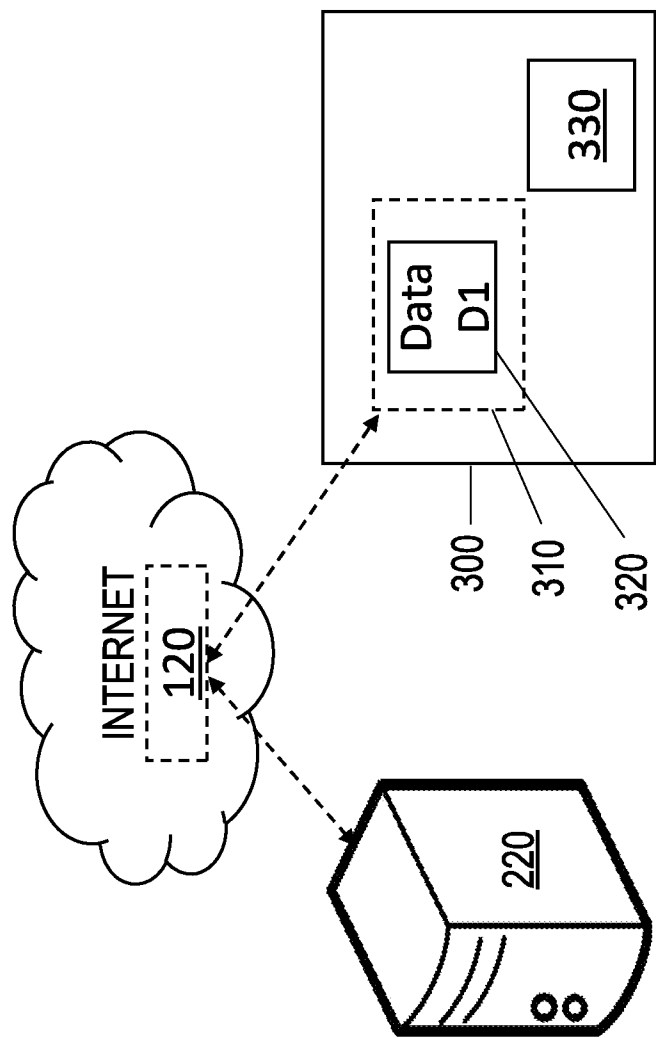


FIG. 3

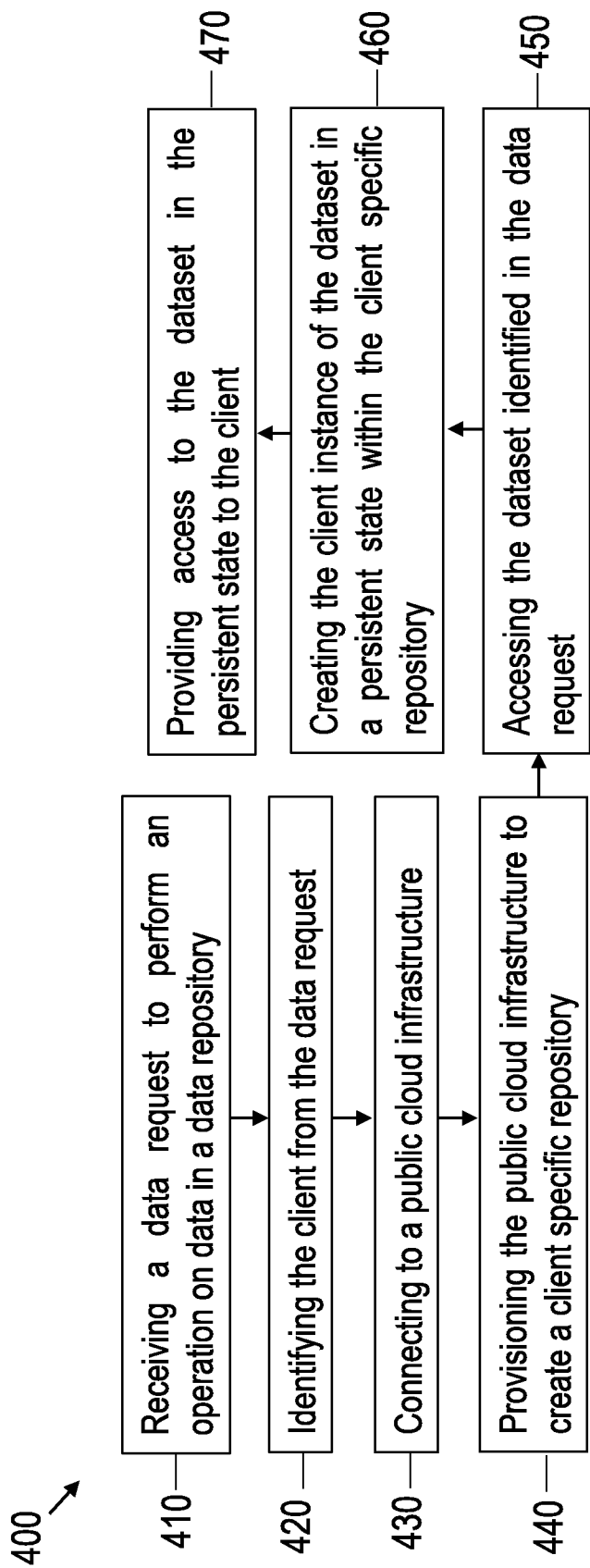


FIG. 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 20/34704

A. CLASSIFICATION OF SUBJECT MATTER

IPC - G06F 17/00 (2020.01)

CPC - G06F 16/164, G06F 16/13, G06F 16/174, Y10S 707/916, Y10S 707/99932

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 2017/0228248 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 10 August 2017 (10.08.2017), entire document, especially abstract.	1-20
A	US 2017/0322954 A1 (MONGODB, INC.) 09 November 2017 (09.11.2017), entire document, especially abstract.	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

17 July 2020 (17.07.2020)

Date of mailing of the international search report

10 AUG 2020

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Lee Young

Telephone No. PCT Helpdesk: 571-272-4300