



- (51) International Patent Classification:  
*G06F 21/60* (2013.01) *G06F 17/30* (2006.01)
- (21) International Application Number:  
PCT/US2014/067937
- (22) International Filing Date:  
1 December 2014 (01.12.2014)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
3764/CHE/2014 31 July 2014 (31.07.2014) IN
- (71) Applicant: **HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P.** [US/US]; 11445 Compaq Center Drive W., Houston, Texas 77070 (US).
- (72) Inventors: **SIVASHANMUGAM, Jothivelavan**; Sy.No.192, Whitefield Road, Mahadevapura Post, Karnataka, Bangalore 560048 (IN). **KANNAN, Rajkumar**; Sy.No.192, Whitefield Road, Mahadevapura Post, Karnataka, Bangalore 560048 (IN). **KOOMTHANAM, Annmary Justine**; Sy.No.192, Whitefield Road, Mahadevapura Post, Karnataka, Bangalore 560048 (IN).

(74) Agents: **ORTEGA, Arthur** et al.; Hewlett-Packard Company, Intellectual Property Administration, 3404 East Harmony Road, Mail Stop 35, Fort Collins, Colorado 80528 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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).

[Continued on next page]

(54) Title: POLICY-BASED DATA PROTECTION

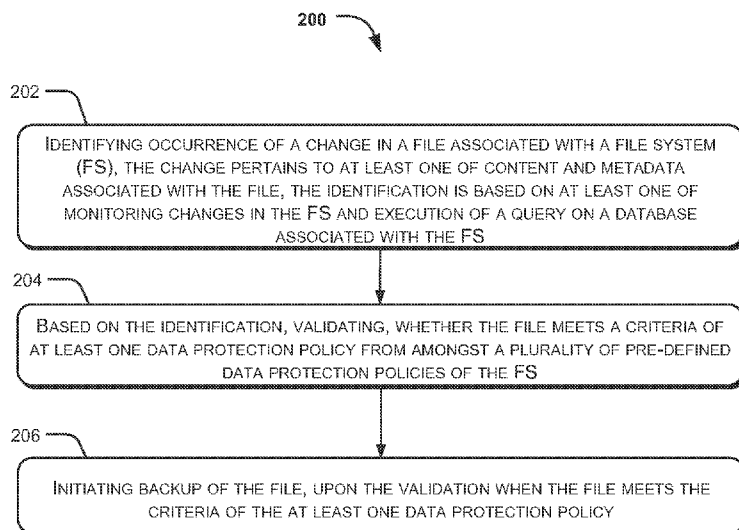


FIG. 2A

(57) Abstract: To protect files in a file system (FS), occurrence of a change in a file associated with the FS is identified. The change may pertain to at least one of content or metadata associated with the file. The identification is based on at least one of monitoring changes in the FS and execution of a query on a database associated with the FS. Based on the identification, it is validated whether the file meets criteria of at least one data protection policy from amongst a plurality of pre-defined data protection policies of the FS. Upon the validation, when the file meets the criteria of the at least one data protection policy, backup of the file is initiated.

**Declarations under Rule 4.17:**

- *as to the identity of the inventor (Rule 4.17(i))*
- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

**Published:**

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

## POLICY-BASED DATA PROTECTION

## BACKGROUND

**[0001]** Data is increasingly produced and stored digitally on computing  
5 systems and storage devices. The data may get corrupted or lost accidentally,  
for example, due to failure of a component, human error, software corruption,  
or system failure. Protection of data has thus become a priority for most  
organizations. Generally, organizations use data protection systems that  
backup data periodically and allow users to restore data in case original data  
10 is not available.

## BRIEF DESCRIPTION OF DRAWINGS

**[0002]** The detailed description is described with reference to the  
accompanying figures. In the figures, the left-most digit(s) of a reference  
number identifies the figure in which the reference number first appears. The  
15 same numbers are used throughout the figures to reference like features and  
components:

**[0003]** FIG. 1A illustrates components of a data protection system,  
according to an example of the present subject matter.

**[0004]** FIG. 1B illustrates a network implementation of the data  
20 protection system, according to another example of the present subject matter.

**[0005]** FIG. 2A illustrates a method for policy-based data protection in  
a file system, according to an example of the present subject matter.

**[0006]** FIG. 2B illustrate method for policy-based data protection in a file system, according to other examples of the present subject matter.

**[0007]** FIG. 3 illustrates a computer readable medium storing instructions for policy-based data protection in a file system, according to an  
5 example of the present subject matter.

#### DETAILED DESCRIPTION

**[0008]** Organizations often take backup of data to ensure availability in case data gets corrupt due to, for example, failure of a system component, human error, software corruption, or some other error. Generally, data  
10 protection systems rely on scheduled backups which involve taking backup of data at specific time intervals. The scheduled backups generally involve an agent that tracks changes occurring in a file system (FS). When a next backup is scheduled, the agent initiates the backup of the tracked data. Thus, the data protection systems are time-based and backup the data at specific time  
15 intervals. Further, generally, the data protection systems determine the schedule of the backups and initiate the backups based on Recovery Time Objective (RTO) values and Recovery Point Objective (RPO) values of an organization.

**[0009]** Even when periodicity of backup schedules is high, at a given  
20 time there exists some amount of unprotected data, which is not replicated or copied to a backup storage device. Such data may be protected upon intervention from a system administrator or when the next scheduled backup occurs. However, such unprotected data may frequently get corrupted or lost, for example, in environments with large amounts of rapidly changing data,  
25 such as large archive storage clusters. The scheduled data protection systems are inadequate for such environments. Further, the scheduled data protection

systems preserve any particularly relevant data of the organization as per the same backup schedule as other less relevant data in the FS. Any error in a computing system may cause loss of the particularly relevant data also, which may not be recoverable if not backed up.

- 5   **[0010]**       According to various examples, systems and methods for policy-based data protection are disclosed. A policy-based data protection system of the present subject matter automatically initiates a backup of data when criteria defined in one or more data protection policies of an organization are met.

- 10   **[0011]**       The policy based data protection system may be referred to as data protection system hereinafter. The data protection system may identify occurrence of a change in a file associated with a file system (FS), such as a journaling FS. The file may be understood as any computer generated document, content, and metadata of which are modifiable by users. For example, the file may be a medical record, employee details, a legal document,  
15   and the like. The metadata associated with the file may be stored in a database, such as a pipeline database, that may be associated with the FS. The database maintains indexed, ordered, and queryable tables which may be queried. In an example, the change may pertain to a change in the content or the metadata associated with the file. Further, based on the identification of  
20   the occurrence of the change in the file, the data protection system may validate whether the file meets a criterion pertaining to a data protection policy or not. The data protection policy defines terms, such as, influx of data in the FS, size of the file, type of the file, criticality of the file, and nature of the file (retained or WORM), for backing up the file. If the criterion of the data  
25   protection policy is met, the backup of the file may be initiated.

- [0012]**       In another example, the data protection system may receive a user input to retrieve information about the changes occurring in the FS. Based

on the user input, the data protection system may generate at least one query. For example, the user input may be parameters and criteria associated with more than one data protection policies. In this case, the data protection system may generate different queries for different data protection policies. Once  
5 generated, the data protection system may execute the at least one query on the database to obtain changes occurring in the FS. As mentioned above, the data protection system may validate these changes against the data protection policies that are pre-defined in the FS to determine whether the criterion of the data protection policy is met or not. If the criterion is met, the backup of the file  
10 may be initiated.

**[0013]** In order to initiate the backup, the data protection system may identify a backup destination, such as a backup system, as may be indicated in the data protection policy. The backup destination may include, but is not limited to, a disk storage, a tape storage system having tape drives and tape  
15 cartridges, an optical disk library, a disaster recovery (DR) site, and an independent software vendor (ISV) based backup. Once the backup destination is identified, the data protection system may invoke a backup application corresponding to the backup destination to backup the file.

**[0014]** In accordance with the examples of the disclosed systems and  
20 methods, the backup may be initiated when the criteria for any data protection policy is met. The data protection policy can be related to changes occurring in the FS in addition to or alternatively to recovery based objectives. Thus, the data protection system monitors the changes occurring in the FS and based on the changes, automatically initiates the backup, when the criteria pertaining  
25 to any of the data protection policy is met. Thus, the data protection system can be made content aware and does not have to wait for a next scheduled backup to protect the data. Moreover, the data protection policies can be

dynamically incorporated into the data protection system based on user inputs, thereby helping the data protection system to respond quickly to changes in business demands.

**[0015]** The various systems and the methods are further described in conjunction with the following figures. It should be noted that the description and figures merely illustrate the principles of the present subject matter. Further, various arrangements may be devised that, although not explicitly described or shown herein, embody the principles of the present subject matter and are included within its scope.

**[0016]** The manner in which the systems and the methods for policy-based data protection in a file system are implemented are explained in details with respect to FIG. 1A, FIG. 1B, FIG. 2A, FIG. 2B, and FIG. 3. While aspects of described systems and methods for policy-based data protection in a file system can be implemented in any number of different computing systems, environments, and/or implementations, the examples and implementations are described in the context of the following system(s).

**[0017]** FIG. 1A illustrates the components of a data protection system 100, according to an example of the present subject matter. In one example, the data protection system 100 may be implemented as any computing system, such as a desktop, a laptop, a server, and the like. In an example, the data protection system 100 can be implemented in a network environment comprising a variety of network devices, including routers, bridges, servers, computing devices, storage devices, etc.

**[0018]** In one implementation, the data protection system 100 includes a processor 102 and modules 104 coupled to the processor 102. The processor 102 may include microprocessors, microcomputers,

microcontrollers, digital signal processors, central processing units, state machines, logic circuitries, and/or any other devices that manipulate signals and data based on computer-readable instructions. Further, functions of the various elements shown in the figures, including any functional blocks labeled  
5 as “processor(s)”, may be provided through the use of dedicated hardware as well as hardware capable of executing computer-readable instructions.

**[0019]** The modules 104, amongst other things, include routines, programs, objects, components, and data structures, which perform particular tasks or implement particular abstract data types. The modules 104 may also  
10 be implemented as, signal processor(s), state machine(s), logic circuitries, and/or any other device or component that manipulates signals based on operational instructions. Further, the modules 104 can be implemented by hardware, by computer-readable instructions executed by a processing unit, or by a combination thereof. In one implementation, the modules 104 include  
15 an identification module 106 and an initiation module 108.

**[0020]** In an implementation, the identification module 106 may identify changes occurring in a file system (FS). For example, the identification module 106 may generate a query to retrieve information pertaining to changes occurring in the FS. The changes in the FS can include changes made to at  
20 least one file associated with the FS. In an example, the FS may be a journaling FS that maintains a log, also referred to as a journal, which includes a list of actions performed on the FS. An action may be understood as a modification in the content or metadata associated with the files of the FS. Further, a file may be understood as any computer generated document,  
25 content and metadata of which are modifiable by users. For example, the file may be a medical record, employee details, a legal document, and the like. Further, the metadata of the file may be stored in a database, such as a pipeline database, associated with the FS. In addition, the changes occurring

in the file may pertain to changes in either content or metadata associated with the file.

**[0021]** In an example, an administrator may define a plurality of data protection policies based on the objectives of the enterprise. A data protection policy defines terms for backing up the files when the criterion is satisfied. The criteria of the data protection policies may be based on a plurality of parameters associated with the file. For example, the parameters may include, but are not limited to size of a file, type of a file, criticality of a file, and nature of a file. Accordingly, the criteria may be defined as initiating the backup when the size of the file is more than or equal to 10MB, when the type of file is .DOC or .BIN, initiating the backup when the file is one of a retained or write once read many (WORM) file, and the like. Further, different data protection policies may be applied on different files based on the data included in the file. In an implementation, the users may provide input to the identification module 106, such as in the form of the parameter and the criteria of the data protection policies. Based on the input, the identification module 106 may generate at least one query to be executed on the database 158.

**[0022]** Considering an example where the user input may include the parameters as size of the file and type of the file. In addition, the user input may include the criteria as size of the file to be more than 5MB and type of the file to be a spreadsheet document. Based on the user input, the identification module 106 generates a query for being executed on the database. In the above example, as there are two parameters provided as input by the user, the identification module 106 may generate two queries for being executed on the database or a single query including both parameters. In an example, the user input may include multiple parameters, based on which the identification module 106 may generate complex queries. Further, the queries, when

executed on the database, may provide a list of changes that have occurred in the FS, with respect to the parameters and the criteria provided in the input. In an example, the users may also indicate a previous point in time with respect to which the changes have to be listed.

5     **[0023]**         Further, the identification module 106 may generate a report of the changes that are listed upon execution of the at least one query on the database. In an example, the identification module 106 may automatically generate the report by executing queries on the database at pre-defined time intervals. In another example, the initiation module 108 may request the  
10    identification module 106 to provide the report. Upon receiving the request, the identification module 106 may generate the report and share the report with the initiation module 108. In yet another example, the identification module 106 may execute the query on the database and store the report thus created. Subsequently, the identification module 106 may share the report with the  
15    initiation module 108 at the pre-defined time interval or when requested by the initiation module 108. Accordingly, the identification module 106 provides a fast query mechanism and saves time in responding to the requests of the initiation module 108.

20    **[0024]**         In addition, the list of changes obtained upon execution of the query may be validated to determine whether the files indicated in the list have to be backed up or not. In an example, the initiation module 108 may validate whether the file meets the criterion of a data protection policy or not. In another example, the identification module 106 may validate whether the file meets the criterion of the data protection policy or not. In case the criteria of any of the  
25    data protection policies is met, the initiation module 108 may initiate backup of the files.

**[0025]** When the file meets the criterion of a data protection policy, the initiation module 108 may identify a backup destination, such as a backup system, based on the data protection policy. Upon identification of the backup system, the initiation module 108 may invoke a backup application  
5 corresponding to the backup system for taking the backup of the file. The process of determination of changes in the FS and backup initiation by the data protection system 100 is described in greater detail in conjunction with Fig. 1B.

**[0026]** FIG. 1B illustrates a network environment 150 including the data  
10 protection system 100 according to another example of the present subject matter. As mentioned previously, the data protection system 100 may be implemented in various computing systems, such as personal computers, servers, etc. The data protection system 100 may be implemented on a network interfaced computing system. In one example, for the purpose of  
15 policy-based data protection in the network environment 150, the data protection system 100 may communicate with a plurality of user devices 152-1, 152-2, ... , 152-N over a network 154. The network 154 may be a wired network, a wireless network or a combination of a wired and wireless network. The network 154 can also be a collection of individual networks, which may  
20 use different protocols for communication, interconnected with each other. Further, the network 154 can include various network elements, such as gateways, modems, routers; however, such details have been omitted for ease of understanding. In one example, the network 154 may be a private network, such as an enterprise network, or a public network, such as a cloud network,  
25 or a hybrid network. The user devices 152-1, 152-2, ..., 152-N can be collectively referred to as user devices 152 and individually referred to as a user device 152 hereinafter. The user devices 152 can include, but are not restricted to, desktop computers, laptops, data servers, and the like. In an

implementation, the data protection system 100 may initiate the backup of the at least one file stored in the user devices 152.

**[0027]** In one example, the data protection system 100 may be coupled to one or more backup systems 156. The backup systems 156 may store a  
5 backup copy of the files associated with the FS, such as the journaling FS. In an example, the backup systems 156 may be used for protection of the files, such as in an event of meeting of a criterion of a data protection policy. The backup systems 156 may include any suitable secondary storage device for maintaining a backup copy of the files. The secondary storage devices may  
10 include, but are not limited to, a disk storage, a tape storage system comprised of one or more tape drives and tape cartridges, an optical disk library, a disaster recovery (DR) site, and an independent software vendor (ISV) based backup. Though not depicted in the illustrated example, in an example, the backup systems 156 may be a storage area network (SAN) or a cloud based  
15 storage.

**[0028]** In an implementation, the data protection system 100 may be coupled to a database 158 over the network 154 or any other network in the network environment 150. Although not shown in the figure, the database 158 may also be directly connected to the data protection system 100. In an  
20 example, the database 158 hosts data used by the data protection system 100 for initiating the backup of the data in the FS. The FS may be deployed in each of the user devices 152. Alternatively, the FS may be implemented in the data protection system 100. For example, the database 158 may be used for storing changes in the metadata associated with a file. Based on the changes, the  
25 files for being backed up by the backup system 156 may be identified. Further, the database 158 may store a data protection policy associated with the FS and the details about the backup system 156.

**[0029]** In an example, the database 158 may be provided as a relational database and may store data in various formats, such as relational tables, object oriented relational tables, and indexed tables. The database 158 may be a single database, multiple databases, or a distributed database. The database 158 may also be provided as other types of databases, such as operational databases, analytical databases, hierarchical databases, and distributed or network databases. In the present implementation, the database 158 may be an express query (EQ) database for storing information pertaining to the backup system 156, information pertaining to the files, details about the data protection policies that may be applicable on the files, and the like. In an alternative implementation, the data protection policies and the information pertaining to the backup system 156 may be stored on a database separate from the database storing the metadata of the FS.

**[0030]** In an implementation, the data protection system 100 includes the processor 102 and a memory 160 connected to the processor 102. The memory 160, communicatively coupled to the processor 102, can include any non-transitory computer-readable medium known in the art including, for example, volatile memory, such as static random access memory (SRAM) and dynamic random access memory (DRAM), and/or non-volatile memory, such as read only memory (ROM), erasable programmable ROM, flash memories, hard disks, optical disks, and magnetic tapes.

**[0031]** The data protection system 100 also includes interface(s) 162. The interfaces 162 may include a variety of interfaces, for example, interfaces 162 for user device(s), such as the user devices 152, the backup storage system 156, and network devices of the network 154. The interface(s) 162 may include data input and output devices, referred to as I/O devices. The interface(s) 162 facilitate the communication of the data protection system 100

with various communication and computing devices and various communication networks, such as networks that use a variety of protocols, for example, Hypertext Transfer Protocol (HTTP) and Transmission Control Protocol/Internet Protocol (TCP/IP).

5   **[0032]**       Further, in addition to the identification module 106 and the initiation module 108, the modules 104 may include a policy management module 164, and other module(s) 166. The other module(s) 166 may include programs or coded instructions that supplement the applications or functions performed by the data protection system 100. The modules 104 may be  
10   implemented as described in relation to FIGS. 1A and 1B.

**[0033]**       In an example, the data protection system 100 includes data 168. The data 168 may include input data 170, validation data 172, and other data 174. The other data 174 may include data generated and saved by the modules 104 for implementing various functionalities of the data protection  
15   system 100.

**[0034]**       In an example, the identification module 106 may identify the changes occurring in the FS, for example, from a journal associated with the FS. In an implementation, the identification module 106 may be implemented in a journal scanner of the FS. The identification module 106 may therefore,  
20   parse the data written on the journal to determine the changes occurring in the FS. The identification module 106 may store the parsed data of the journal in the database 158. For example, the identification module 106 may store the changes occurring in the FS in the database 158. The identification module 106 may further communicate with the policy management module 164 to  
25   access the data with respect to the data protection policies of the FS. For example, the identification module 106 may compare the changes with respect to the pre-defined data protection policies of the FS. The identification module

106 may monitor the changes occurring in the FS, such as what modifications are made in the files, how many times a particular type of file is modified, and changes in the files in terms of percentage since a last backup.

**[0035]** The identification module 106 may, based on the comparison of the changes with the data protection policy, determine if any file has crossed a threshold of backed up data. The threshold defines a criteria of the data protection policies, based on which the backup is initiated. In an implementation, the criteria may be associated with the content or metadata of the files. For example, a data protection policy may define that all files which are modified more than 10 times a day have to be backed up. The data protection policy may also indicate the type of backup source where the copy of the file may be stored if the criteria is met. In an example, the identification module 106 may parse the actions listed in the journal in a pre-defined manner to determine the influx of data in the FS. In the present example, the identification module 106 may maintain a counter to check influx of data in the FS. The counters may facilitate in detecting rate of influx of data in the FS when the data is being written in the FS. This detection of change in the data, facilitates in early initiation of the backup of files, i.e., as soon as the changes are made in the FS rather than waiting for a scheduled backup which may risk the data of the FS. Upon the detection of the changes, the data protection system 100 may trigger backup of the data.

**[0036]** The identification module 106 may, while monitoring the changes occurring in the FS, compare the changes with the criteria of each of the pre-defined data protection policies. The identification module 106 thus ensures that huge amount of data does not remain lying unprotected in the FS, waiting for a scheduled backup to happen.

**[0037]** The policy management module 164 may facilitate an administrator to define the data protection policies in the FS. In an implementation, the data protection policies may be stored in the database 158 or a database different from the database 158. In an example, the database 158 though shown outside the data protection system 100 may reside inside the data protection system 100. In an example, the administrator may define the data protection policies based on the objectives of the enterprise. Further, these data protection policies may be updated or modified by the administrator as per the demands of the enterprise. As mentioned above, the data protection policies define a criterion for initiating backup of the data of the FS. When the criterion is met, the data protection system 100 may push the data to be backed up in the backup systems 156 as may be defined in the data protection policies.

**[0038]** The policy management module 164 may define a schedule for validation of the plurality of data protection policies. The schedule may be indicative of a time interval after which the validity of the data protection policies is checked. For example, the administrator may schedule a query to be executed on the database 158 after every half an hour. Accordingly, the identification module 106 may execute the query at the scheduled time interval. The results obtained upon execution of the query may be validated against the data protection policies. For example, the identification module 106 may compare the result of the query with the parameter defined in each of the data protection policies, after the pre-defined time interval, in this case, half an hour. The identification module 106, upon validating the results with the data protection policy, may initiate the backup of the data. In case the results are not in accordance with the data protection policy, the backup will not be initiated by the initiation module 108.

**[0039]** In addition, the policy management module 164 may store priority information about each of the data protection policies. The priority information may be understood as priorities assigned by the administrator to each of the data protection policies. For example, the administrator may define that some of the data protection policies, if the criterion is met, have to be given priority over other operations of the data protection system 100. Accordingly, when the criterion of a data protection policy is met, the policy management module 164 may share the priority information corresponding to the data protection policy with the identification module 106. Consequently, the identification module 106 may request the initiation module 108 to trigger the backup of such files for which the data protection policy is applicable.

**[0040]** In an implementation, the identification module 106 may query the database 158 to retrieve information about the various changes occurring in the FS. As mentioned above, the database 158 may include queryable tables which may be queried. Therefore, the users may query the database 158 by providing input through the interface 162. For example, the input may include the parameter and the criteria associated with the data protection policies. Based on the parameter and the criteria, the identification module 106 may generate at least one query to be run on the database 158. For instance, the user may provide input as "number of WORM and retained files between time instance 1 and time instance 2". Based on the input, the identification module 106 may generate the query as provided below:

```
"select pathname from data_fileobjects_by_pathname where ondisktimesec  
>=$startTimestamp and ondisktimesec <$endTimestamp and  
retentionstate='3';",
```

where retentionstate of '3' specifies that the file is WORM and retained.

**[0041]** In an implementation, the database 158 may reside on a computing system other than the computing system implementing the FS. When the queries are executed on the database 158, resources, such as memory and processor, of the computing system having the database 158 may be consumed and usage of resources of the computing system in which the FS resides is reduced. In an example, the identification module 106 may store the input received from the users as the input data 170.

**[0042]** In another example, the users may provide the parameters and the criteria associated with multiple data protection policies as the input. In such cases, the identification module 106 may break or segment the input in multiple queries, each relating to one data protection policy. In an implementation, the identification module 106 may be configured to automatically query the database 158 without receiving input from the users. For example, the identification module 106 may execute pre-defined queries, as may be scheduled and defined by the administrator. The results of these queries may be stored along with the input data 170 in the data protection system 100. In an example, the identification module 106 may generate single output for multiple queries. Considering a scenario where the database 158 resides on multiple computing systems, which are different from the computing system on which the FS is residing. When a user selects the parameter and the criteria as the input of the query, the identification module 106 may execute the query on the database 158 of the multiple computing systems as multiple queries.

**[0043]** In an implementation, the FS, the identification module 106, and the database 158 of the present subject matter may be deployed in a cloud based environment. In such environment, the FS, the identification module 106, and the database 158 may be located across multiple computing

systems. In such a scenario, the policy management module 164 may also be stored across the multiple computing systems. When the user input is received by the identification module 106, such as in the form of a query, to obtain some information, as mentioned above, the identification module 106 retrieves the  
5 output from the multiple computing systems on which the database 158 is residing.

**[0044]** Based on an output of the queries run on the database 158, the identification module 106 may generate reports containing details about the changes that were identified in the FS since a last backup or last execution of  
10 a similar query. These reports facilitates in determining the files which have to be backed up. The identification module 106 may share the reports with the initiation module 108 to analyze and invoke the backup applications for storing a copy of the files for which the criteria pertaining to the data protection policy is met. In an implementation, the identification module 106 may generate the  
15 reports at pre-defined time intervals or upon receipt of the user input. In an alternative implementation, the identification module 106 may generate the reports upon being requested by the initiation module 108. In an example, the identification module 106 may generate the queries and store the results of the queries in the database 158. When the initiation module 108 sends a  
20 request to the identification module 106 to provide a list of files that have to be backed up, the identification module 106 shares the results of the pre-executed queries, in the form of reports, to the initiation module 108. The identification module 106 thereby facilitates in saving upon the time spent in executing queries upon receipt of the request from the initiation module 108.

25 **[0045]** In an example, upon receiving the reports from the identification module 106, the initiation module 108 may validate whether or not the criteria of a data protection policy is met. In an example, the initiation module 108 may

validate the file to determine whether the threshold as defined in the data protection policy has been crossed or not. In an implementation, the validation may be performed based on the scheduled time interval, as may be defined by the administrator. The initiation module 108 may store the validation details  
5 as the validation data 172. Once the file is validated, either by the identification module 106 or the initiation module 108, the initiation module 108 may identify the backup system 156 as may be defined in the data protection policy, for storing the copy of the files. The initiation module 108 may invoke the backup application corresponding to the backup system 156 to initiate the backup of  
10 the files. In an example, the initiation module 108 may invoke the backup system 156 through a representational state transfer (REST) interface. In another example, the initiation module 108 may invoke the backup system 156 by a command line argument.

**[0046]** The data protection system 100 facilitates the users to query the  
15 database 158 to determine changes occurring in the FS. Based on these changes, the backup may be initiated when the file meets the criteria of any data protection policy. The data protection system 100 is content aware and does not wait for next scheduled backup to protect the data. Further, as the data protection policies may be modified based on the objectives of the  
20 enterprise, the data protection system 100 may facilitate in saving resources and time spent in reconfiguration of all systems within the IT network of the enterprise to adhere to the data protection policies.

**[0047]** FIGS. 2A and 2B illustrate methods 200 and 220 for policy-based data protection in a file system (FS), according to an example of the present  
25 subject matter. The order in which the methods 200 and 220 are described is not intended to be construed as a limitation, and some of the described method blocks can be combined in a different order to implement the methods 200 and

220, or an alternative method. Additionally, individual blocks may be deleted from the methods 200 and 220 without departing from the spirit and scope of the subject matter described herein. Furthermore, the methods 200 and 220 may be implemented in any suitable hardware, computer-readable instructions, or combination thereof.

**[0048]** The steps of the methods 200 and 220 may be performed by either a computing device under the instruction of machine executable instructions stored on a computer readable medium or by dedicated hardware circuits, microcontrollers, or logic circuits. Herein, some examples are also intended to cover computer readable medium, for example, digital data storage media, which are machine or computer readable and encode machine-executable or computer-executable instructions, where said instructions perform some or all of the steps of the described methods 200 and 220.

**[0049]** With reference to method 200 as depicted in FIG. 2A, at block 202, the method 200 includes identifying occurrence of a change in a file associated with the FS. In an example, the change in the file may pertain to change in the content or the metadata of the file. The identification is based on at least one of monitoring changes in the FS and execution of a query on a database associated with the FS. In an implementation, the identification module 106 may identify the occurrence of the change in the FS. For instance, the identification module 106 may identify the change, such as size of a file, type of a file, criticality of a file, and nature of a file. In an example, in order to determine the criticality of the file, a custom metadata, associated with the file may be identified. For example, 'critical' may be the custom metadata that may be added to the file to indicate criticality of the file. The custom metadata may also be provided with values, such as 'normal', 'medium', and 'high'. When a user provides input as 'highly critical files', the identification module 106 may

generate a query for retrieving files for which criticality is 'high'. The custom metadata may be stored in the database 158.

**[0050]** At block 204, the method 200 includes validating, based on the identification, whether the file meets the criteria of at least one data protection policy from amongst a plurality of data protection policies of the FS. In an implementation, when the changes are identified based on parsing the data being stored in the FS, the identification module 106 may validate whether the file meets the criteria of the at least one data protection policy or not. In an alternative implementation, when the changes are identified based on execution of the query on the database 158, the initiation module 108 may validate whether the file meets the criteria of the at least one data protection policy or not.

**[0051]** At block 206, the method 200 includes initiating backup of the file, upon the validation, when the file meets the criteria of the at least one data protection policy. In an implementation, the initiation module 108 may initiate the backup of the file. Accordingly, the initiation module 108 may invoke a backup application associated with the backup destination, such as the backup system 156, to store the file.

**[0052]** Referring to FIG. 2B, at block 222, the method 220 may include identifying occurrence of a change in a file associated with a file system (FS). In an example, the change in the file may pertain to change in the content or the metadata of the file. In an implementation, the identification module 106 may identify occurrence of the change in the FS, based on the changes in the content or the metadata. In an example, the identification module 106 may parse data being stored in the FS and based on the parsing; the identification module 106 may identify the changes occurring in the FS. In another example, the identification module 106 may generate a query based on at least one

parameter associated with the data protection policies. The query may be executed on the database to determine the changes occurring in the FS.

**[0053]** In an example, the query may be generated based on the parameters and the criteria associated with a plurality of data protection policies that may be pre-defined in the FS. For example, the criteria may be defined as initiating the backup when the size of the file is more than or equal to 10MB, when the file is modified for more than 20 times in a week, when an extension of the file is changed, and the like. In another example, the criteria for initiating the backup may include influx of data in the FS. The identification module 106 may determine rate of influx of data in the FS and may initiate the backup of the data, when the data entering the FS crosses a pre-defined threshold. The data protection policy also include information about the backup system 156 where the file is to be stored when the criteria is met by the file.

**[0054]** As shown in block 224, the method 220 may include validating, based on the identification, whether the file meets a criterion of at least one data protection policy from amongst a plurality of pre-defined data protection policies of the FS. In an implementation, when the changes are identified based on parsing the data being stored in the FS, the identification module 106 may validate whether the file meets the criteria of the at least one data protection policy or not. In an alternative implementation, when the changes are identified based on execution of the query on the database 158, the initiation module 108 may validate whether the file meets the criteria of the at least one data protection policy or not. In this implementation, the initiation module 108 may validate the results of the query at scheduled time interval. To validate, the identification module 106 and the initiation module 108 communicate with the policy management module 164.

**[0055]** As depicted in block 226, the method 220 may include initiating backup of the file, upon the validation, when the file meets the criteria of the at least one data protection policy. In an implementation, the initiation module 108 may initiate the backup of the file. Accordingly, the initiation module 108  
5 may invoke a backup application associated with a backup destination, such as the backup system 156, to store the file. In an example, when the identification module 106 validates the file, the identification module 106 may send a message to the initiation module 108 to initiate the backup of the file.

**[0056]** At block 228, the method 220 may include identifying the backup  
10 system 156 based on the at least one data protection policy. In an implementation, the initiation module 108 may identify the backup system 156 of the file from the data protection policy. In an example, the data protection policies stored by the policy management module 164 may indicate the backup destination, such as the backup system 156 based on the changes associated  
15 with the file. The backup destination may include, a disk storage, a tape storage system comprised of one or more tape drives and tape cartridges, an optical disk library, a disaster recovery (DR) site, and an independent software vendor (ISV) based backup.

**[0057]** At block 230, the method 220 may include providing the file for  
20 backup to a backup application corresponding to the backup destination. In an implementation, the initiation module 108 may invoke the backup application to initiate the backup of the file.

**[0058]** FIG. 3 illustrates an example network environment 300  
implementing a non-transitory computer readable medium 302 for policy-  
25 based data protection in a file system (FS), according to an example of the present subject matter. The network environment 300 may be a public networking environment or a private networking environment. In one

implementation, the network environment 300 includes a processing resource 304 communicatively coupled to the non-transitory computer readable medium 302 through a communication link 306.

**[0059]** For example, the processing resource 304 can be a processor  
5 of a computing system, such as the data protection system 100. The non-transitory computer readable medium 302 can be, for example, an internal memory device or an external memory device. In one implementation, the communication link 306 may be a direct communication link, such as one formed through a memory read/write interface. In another implementation, the  
10 communication link 306 may be an indirect communication link, such as one formed through a network interface. In such a case, the processing resource 304 can access the non-transitory computer readable medium 302 through a network 308. The network 308 may be a single network or a combination of multiple networks and may use a variety of communication protocols.

**[0060]** The processing resource 304 and the non-transitory computer  
15 readable medium 302 may also be communicatively coupled to data sources 310 over the network 308. The data sources 310 can include, for example, databases and computing devices. The data sources 310 may be used by the database administrators and other users to communicate with the processing  
20 resource 304.

**[0061]** In one implementation, the non-transitory computer readable  
medium 302 includes a set of computer readable instructions, such as the identification module 106 and the initiation module 108. The set of computer  
25 readable instructions, referred to as instructions hereinafter, can be accessed by the processing resource 304 through the communication link 306 and subsequently executed to perform acts for network service insertion.

**[0062]** For discussion purposes, the execution of the instructions by the processing resource 304 has been described with reference to various components introduced earlier with reference to description of FIGS. 1A and 1B.

5 **[0063]** In an example, on execution by the processing resource 304, the identification module 106 may receive data to be written in a file associated with a file system (FS). The identification module 106 may parse the received data to identify a change occurring in the file. Based on the identification, the identification module 106 may compare the file with the criteria defined in each  
10 of the plurality of pre-defined data protection policies of the FS. The identification module 106 may therefore validate whether the file meets the criteria of the data protection policies or not. If the file meets the criteria, the identification module 106 may send a request to the initiation module 108 to initiate the backup of the file.

15 **[0064]** In an implementation, the identification module 106 may generate at least one query based on a user input to retrieve information about changes occurring in the FS. The identification module 106 may then execute the at least one query on the database 158 to obtain changes occurring in a plurality of files stored in the FS.

20 **[0065]** Although implementations of policy-based data protection in a file system have been described in language specific to structural features and/or methods, it is to be understood that the present subject matter is not necessarily limited to the specific features or methods described. Rather, the specific features and methods are disclosed and explained in the context of a  
25 few implementations for policy-based data protection in the file system.

I/We claim:

1. A method for policy-based data protection in a file system (FS), the method comprising:
  - 5 identifying, by a processor, occurrence of a change in a file associated with the FS, wherein the change pertains to at least one of content and metadata associated with the file, and wherein the identification is based on at least one of monitoring changes in the FS and execution of a query on a database associated with the FS;
  - 10 based on the identification, validating, by the processor, whether the file meets criteria of at least one data protection policy from amongst a plurality of pre-defined data protection policies of the FS; and
  - initiating, by the processor, backup of the file, upon the validation when the file meets the criteria of the at least one data protection policy.
- 15 2. The method as claimed in claim 1 further comprising:
  - receiving, by the processor, changes occurring in metadata associated with a file of the FS, wherein the changes are recorded in a journal associated with the FS;
  - parsing, by the processor, the received changes to identify
  - 20 changes occurring in the FS; and
  - storing, by the processor, the parsed metadata into a database.
3. The method as claimed in claim 1 further comprising:
  - receiving, by the processor, a user input to retrieve information about changes occurring in the FS;
  - 25 generating, by the processor, at least one query based on the user input; and

executing, by the processor, the at least one query on the database to obtain changes occurring in a plurality of files stored in the FS.

- 5     4.     The method as claimed in claim 3, wherein the user input is based on at least one parameter associated with the criteria of the at least one data protection policy.
5.     The method as claimed in claim 3 further comprising generating, by the processor, a report comprising changes obtained upon execution of the  
10     at least one query.
6.     The method as claimed in claim 1, wherein the initiating the backup comprises:  
         identifying, by the processor, a backup destination based on the at least one data protection policy; and  
15     providing, by the processor, the file for backup to a backup application corresponding to the backup destination.
7.     The method as claimed in claim 1, wherein the validation of the at least one data protection policy is performed at pre-defined time intervals.
- 20     8.     A data protection system comprising:  
         a processor;  
         an identification module, coupled to the processor, to,  
         generate at least one query to retrieve information  
         pertaining to changes occurring in a file system (FS); and  
25     execute the at least one query on the database to identify a change occurring in a plurality of files stored in the FS, wherein

- the change pertains to at least one of content and metadata associated with a file of the FS; and  
an initiation module, coupled to the processor, to initiate backup of a file, when the file meets a criteria of at least one data protection policy from amongst a plurality of pre-defined data protection policies of the FS.
- 5
9. The data protection system as claimed in claim 8, wherein the initiation module to further validate, based on the change, whether the file meets criteria of the at least one data protection policy.
- 10 10. The data protection system as claimed in claim 8, wherein the identification module to, further,  
receive changes occurring in metadata associated with a file of the FS, wherein the changes are recorded in a journal associated with the FS;  
15 parse the received changes to identify the change occurring in the FS; and  
validate, based on the change, whether the file meets criteria of the at least one data protection policy.
- 20 11. The data protection system as claimed in claim 8, wherein the identification module to generate the at least one query based on a user input, and wherein the user input comprises a parameter associated with the plurality of data protection policies.
- 25 12. The data protection system as claimed in claim 8, wherein the identification module to execute the at least one query at pre-defined time intervals.

13. The data protection system as claimed in claim 8 further comprises a policy management module, coupled to the processor, to,  
store a plurality of data protection policies, wherein the plurality  
of data protection policies are pre-defined in the FS; and  
5 define a schedule for validation of the plurality of data protection  
policies.
14. A non-transitory computer-readable medium having a set of computer  
readable instructions that, when executed, cause a data protection  
system to:  
10 receive changes occurring in metadata associated with a file of  
the FS, wherein the changes are recorded in a journal associated with  
the FS;  
parse the received metadata to identify a change occurring in a  
file of the FS;  
15 validate, based on the identification, whether the file meets  
criteria of at least one data protection policy from amongst a plurality of  
pre-defined data protection policies of the FS; and  
based on the determination, initiate backup of the file of the FS,  
when the file meets the criteria of the at least one data protection policy.
- 20 15. The non-transitory computer-readable medium as claimed in claim 14,  
wherein the set of computer readable instructions that, when executed,  
further cause the data protection system to:  
generate at least one query based on a user input to retrieve  
information about changes occurring in the FS; and  
25 execute the at least one query on the database to obtain  
changes occurring in a plurality of files stored in the FS.

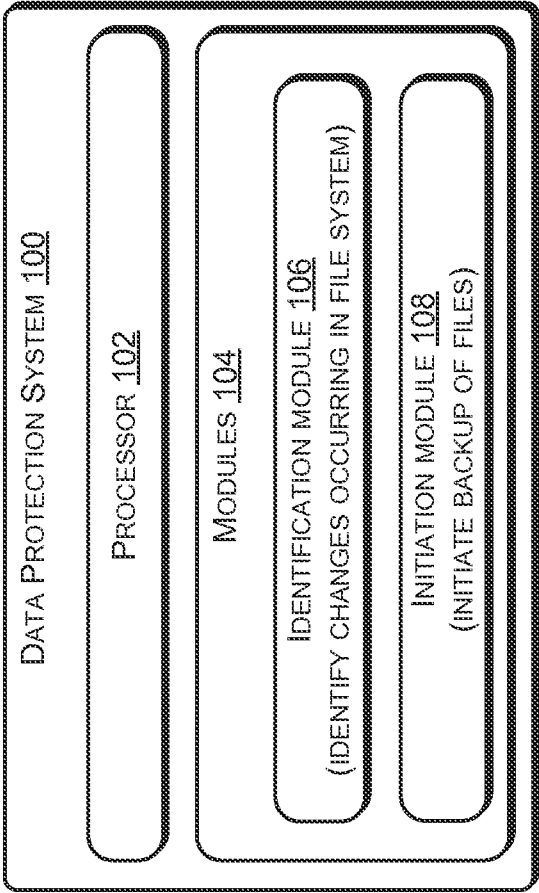


FIG. 1A

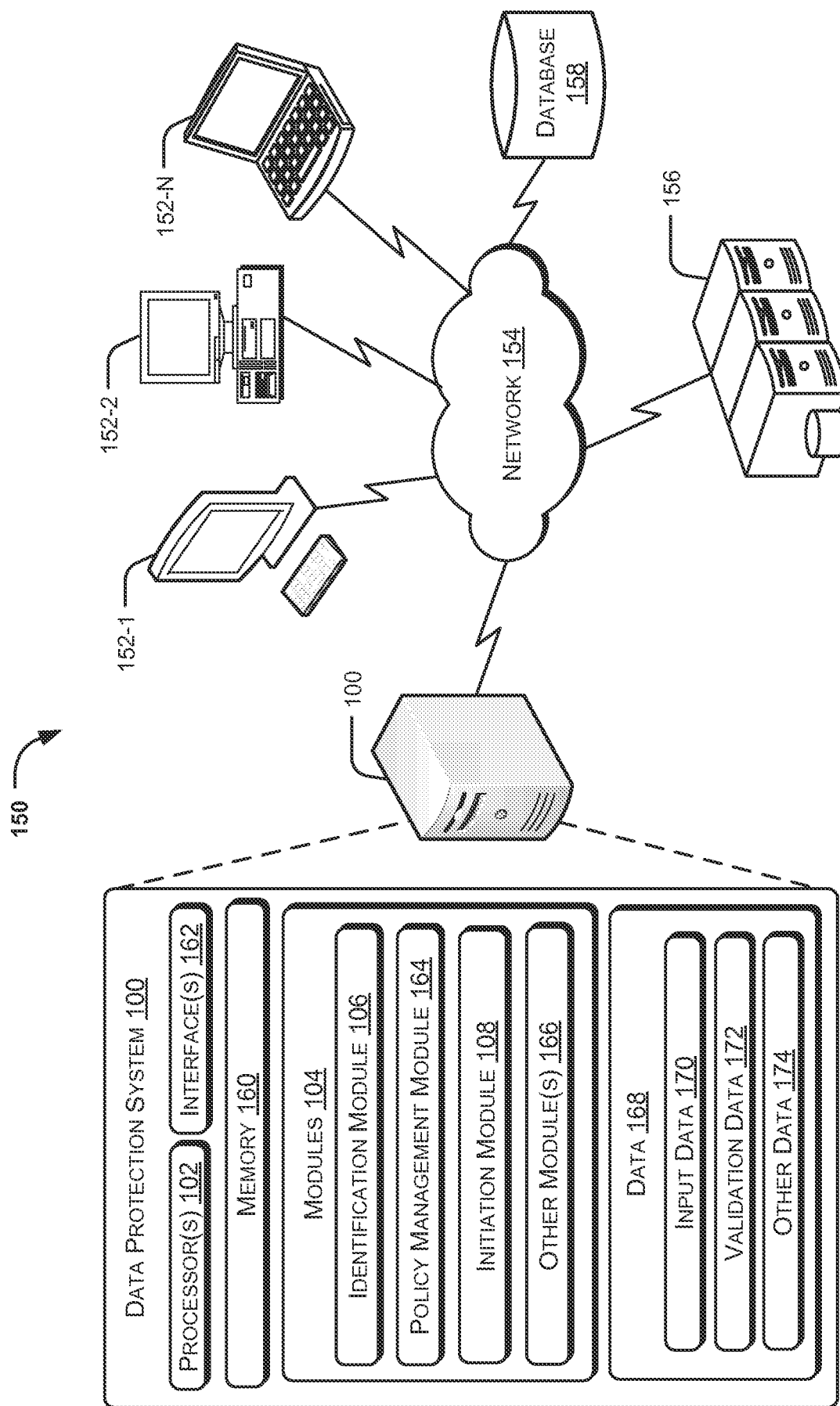


FIG. 1B

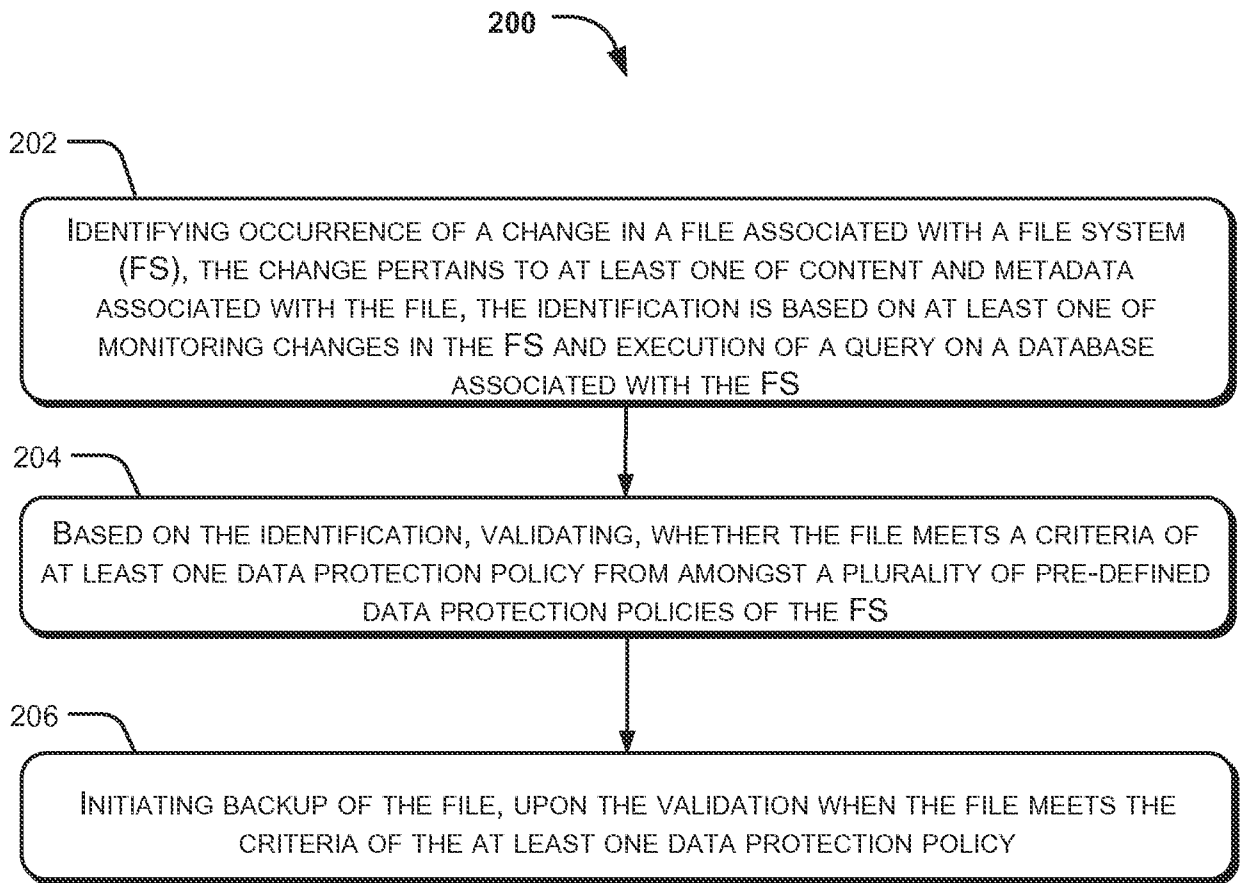


FIG. 2A

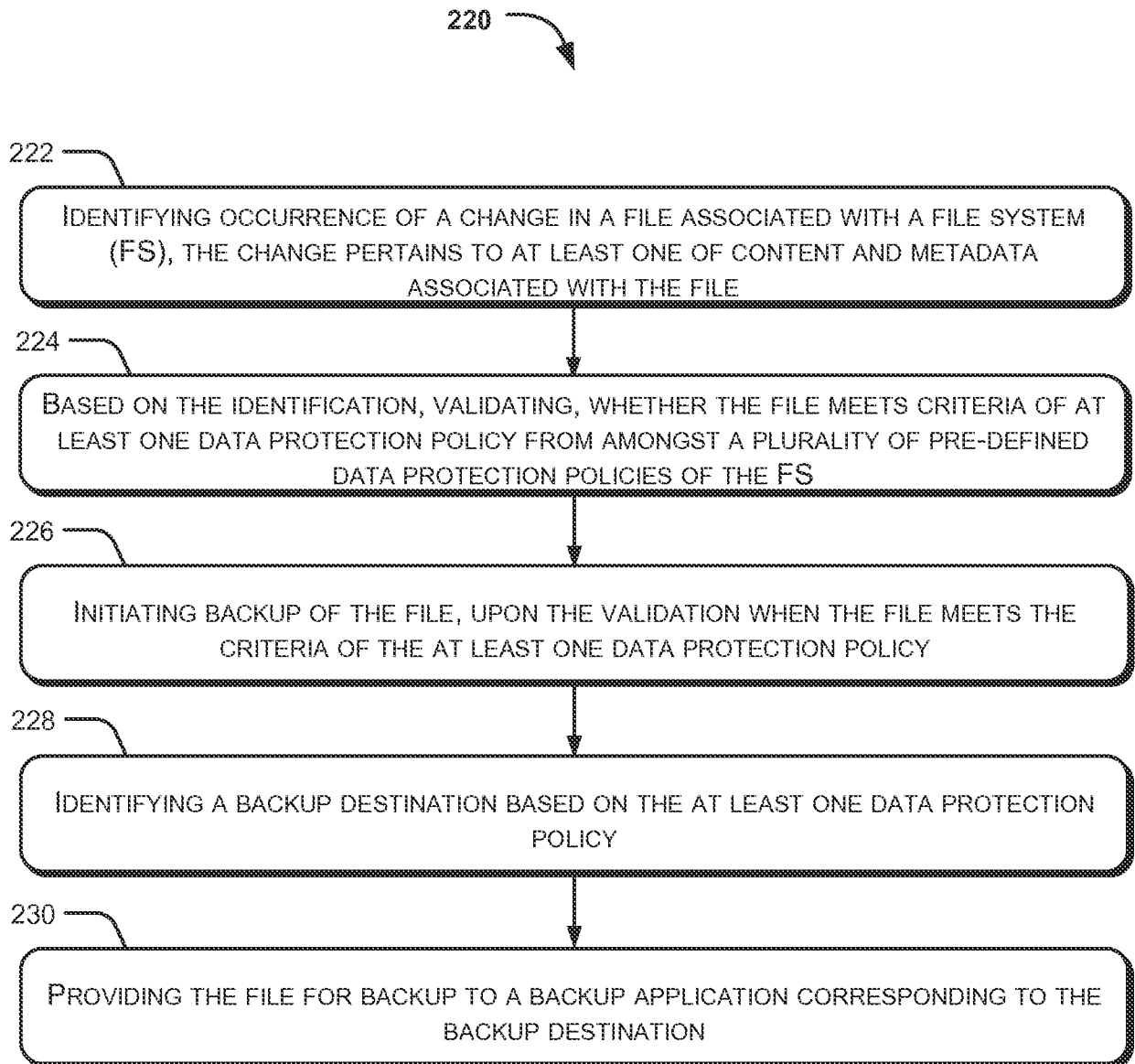


FIG. 2B

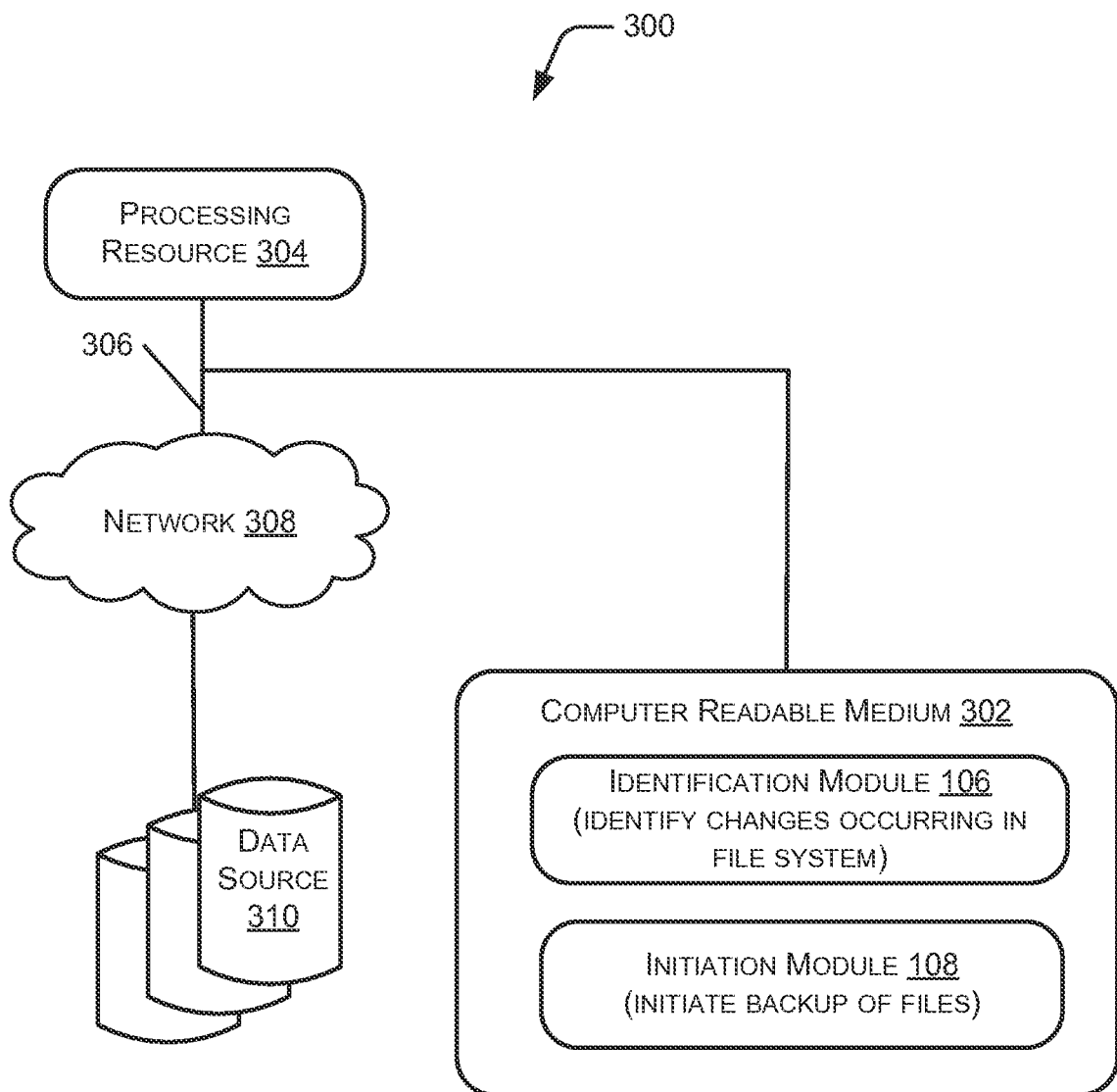


FIG. 3

## INTERNATIONAL SEARCH REPORT

International application No.  
**PCT/US2014/067937****A. CLASSIFICATION OF SUBJECT MATTER****G06F 21/60(2013.01)i, G06F 17/30(2006.01)i**

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)

G06F 21/60; G06F 9/54; G06F 3/00; G06F 17/30; G06F 7/00; G06F 11/14; G06F 11/00

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

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

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

eKOMPASS(KIPO internal) &amp; Keywords:journaling, backup, policy, criteria, change, file system(FS), data protection

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                  | Relevant to claim No. |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Y         | US 2009-0240744 A1 (STEVEN S. THOMSON et al.) 24 September 2009<br>See paragraphs [0013], [0043]-[0044]; claims 1, 10; and figure 1.                                                | 1-15                  |
| Y         | US 7690000 B2 (DONALD MUNRO FARMER) 30 March 2010<br>See column 2, lines 56-65; column 3, lines 46-49; column 7, lines 53-58;<br>claims 1, 24; and figure 4.                        | 1-15                  |
| Y         | WO 2012-045575 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION)<br>12 April 2012<br>See page 7, line 6 - page 8, line 5; page 10, lines 5-6;<br>page 12, lines 20-21; and figure 1. | 4, 7, 11, 13          |
| A         | US 8055613 B1 (PAUL YUEDONG MU et al.) 08 November 2011<br>See column 4, line 66 - column 5, line 34; and figure 1.                                                                 | 1-15                  |
| A         | US 8060889 B2 (SIEW YONG SIM-TANG) 15 November 2011<br>See column 13, line 37 - column 14, lines 59; and figure 5.                                                                  | 1-15                  |



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

"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

"&amp;" document member of the same patent family

Date of the actual completion of the international search

20 March 2015 (20.03.2015)

Date of mailing of the international search report

**20 March 2015 (20.03.2015)**

Name and mailing address of the ISA/KR

International Application Division  
Korean Intellectual Property Office  
189 Cheongsu-ro, Seo-gu, Daejeon Metropolitan City, 302-701,  
Republic of Korea

Facsimile No. ++82 42 472 7140

Authorized officer

AHN, Jeong Hwan

Telephone No. +82-42-481-8440



**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No.

**PCT/US2014/067937**

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s)                                                                                                            | Publication<br>date                                                                            |
|-------------------------------------------|---------------------|---------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|
| US 2009-0240744 A1                        | 24/09/2009          | US 8700574 B2<br>WO 2009-117202 A1                                                                                                    | 15/04/2014<br>24/09/2009                                                                       |
| US 7690000 B2                             | 30/03/2010          | US 2005-0165865 A1                                                                                                                    | 28/07/2005                                                                                     |
| WO 2012-045575 A1                         | 12/04/2012          | CN 103154901 A<br>JP 2013-542517 A<br>US 2012-0089572 A1                                                                              | 12/06/2013<br>21/11/2013<br>12/04/2012                                                         |
| US 8055613 B1                             | 08/11/2011          | None                                                                                                                                  |                                                                                                |
| US 8060889 B2                             | 15/11/2011          | EP 1745397 A2<br>EP 1745397 A4<br>US 2005-0251540 A1<br>US 2010-0031274 A1<br>US 7565661 B2<br>WO 2005-111870 A2<br>WO 2005-111870 A3 | 24/01/2007<br>09/09/2009<br>10/11/2005<br>04/02/2010<br>21/07/2009<br>24/11/2005<br>19/04/2007 |