(12) UK Patent (19) GB (11) 2588524 (13) B

(45) Date of B Publication 13.10.2021

(54) Title of the Invention: Managing big data in process control systems

(51) INT CL: ***G06F 16/22*** (2019.01) ***G05B 19/418*** (2006.01) ***G05B 23/02*** (2006.01) ***G06F 16/20*** (2019.01)

(21) Application No: **2017946.1**

(22) Date of Filing: **21.01.2015**

Date Lodged: **13.11.2020**

(30) Priority Data:
(31) **14169965** (32) **31.01.2014** (33) **US**

(62) Divided from Application No
1501042.4 under section 15(9) of the Patents Act 1977

(43) Date of A Publication **28.04.2021**

(72) Inventor(s):
Mark John Nixon
Paul R Muston
Deji Chen
Song Han

(73) Proprietor(s):
Fisher-Rosemount Systems, Inc
Bldg. 1, 1100 W. Louis Henna Blvd, Round Rock, Texas 78681, United States of America

(74) Agent and/or Address for Service:
Forresters IP LLP
Rutland House, 148 Edmund Street, BIRMINGHAM, B3 2JA, United Kingdom

(56) Documents Cited:
US 20080301123 A1

(58) Field of Search:
As for published application 2588524 A viz:
INT CL **G05B, G06F**
Other: **WPI, EPODOC**
updated as appropriate

Additional Fields
Other: **None**

GB 2588524 B

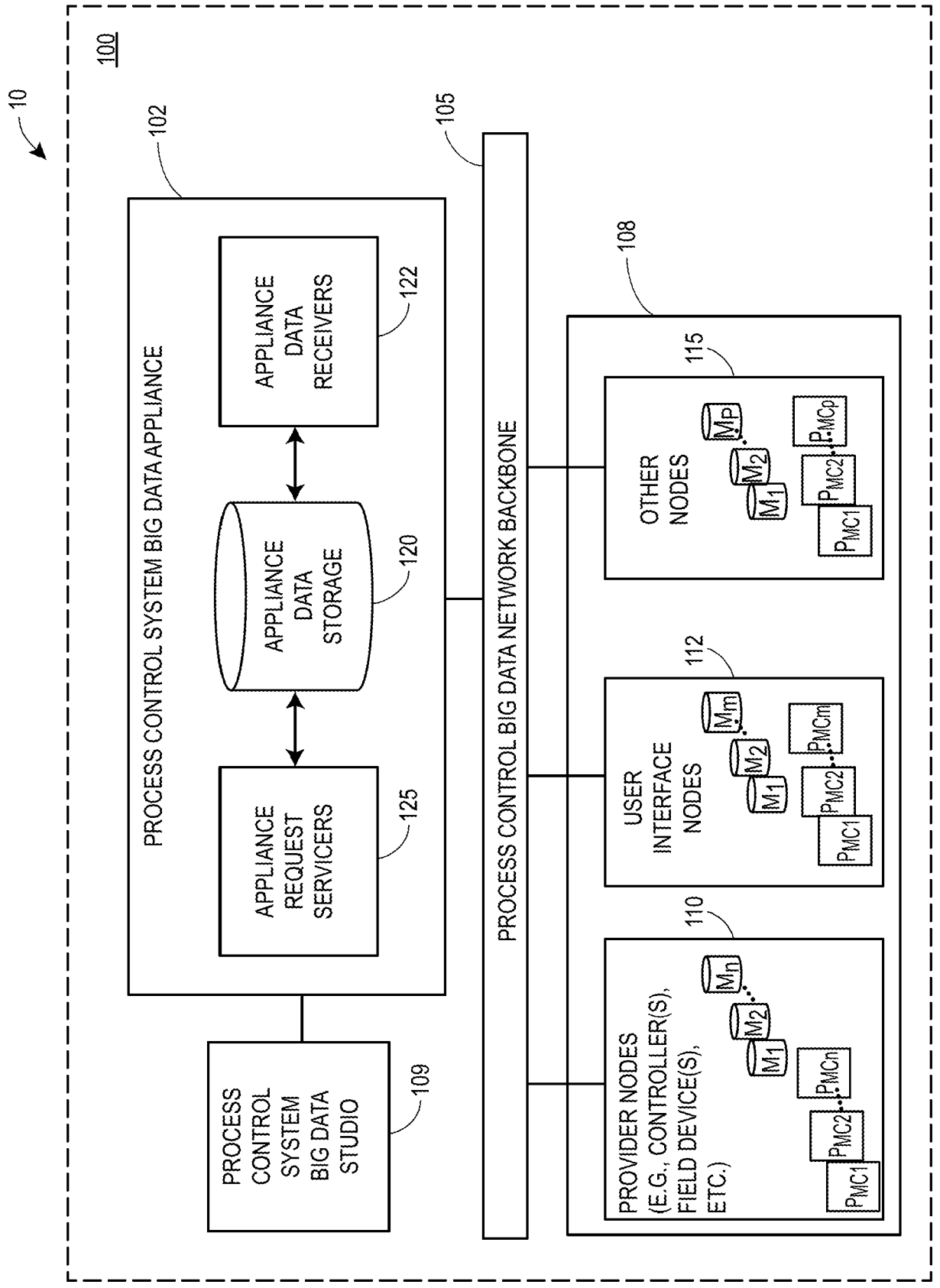
10
100
102
105
PROCESS CONTROL SYSTEM BIG DATA APPLIANCE
APPLIANCE DATA RECEIVERS
122
APPLIANCE DATA STORAGE
120
APPLIANCE REQUEST SERVICERS
125
PROCESS CONTROL SYSTEM BIG DATA STUDIO
109
PROCESS CONTROL BIG DATA NETWORK BACKBONE
108
OTHER NODES
115
USER INTERFACE NODES
112
PROVIDER NODES (E.G., CONTROLLER(S), FIELD DEVICE(S), ETC.)
110

FIG. 1

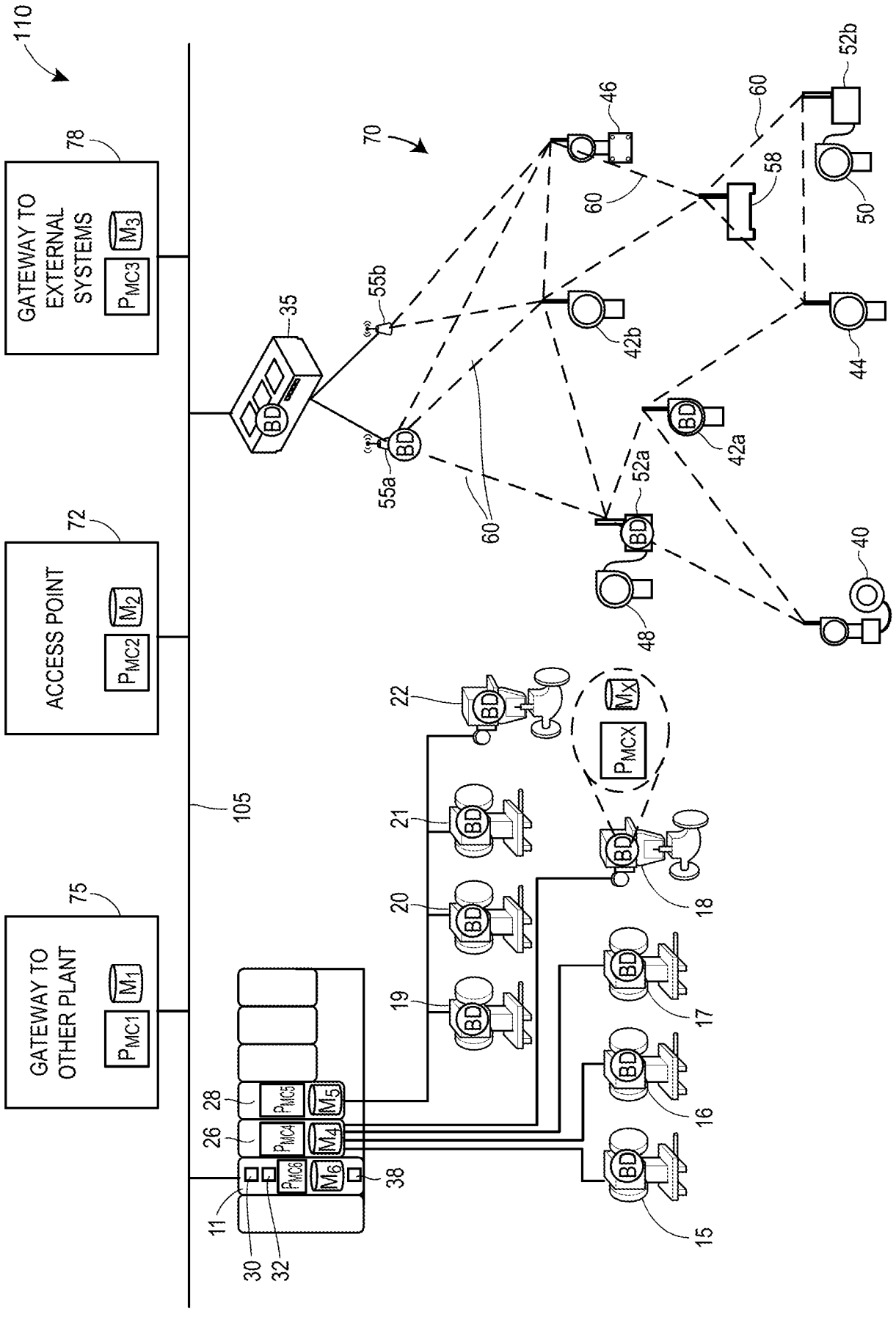
110
GATEWAY TO EXTERNAL SYSTEMS
78
P_{MC3}
M_3
ACCESS POINT
72
P_{MC2}
M_2
GATEWAY TO OTHER PLANT
75
P_{MC1}
M_1
105
70
35
BD
55a
55b
60
46
42a
42b
44
48
50
52a
52b
58
40
P_{MCX}
M_X
P_{MC5}
M_5
P_{MC4}
M_4
P_{MC6}
M_6
11
15
16
17
18
19
20
21
22
26
28
30
32
38

FIG. 2

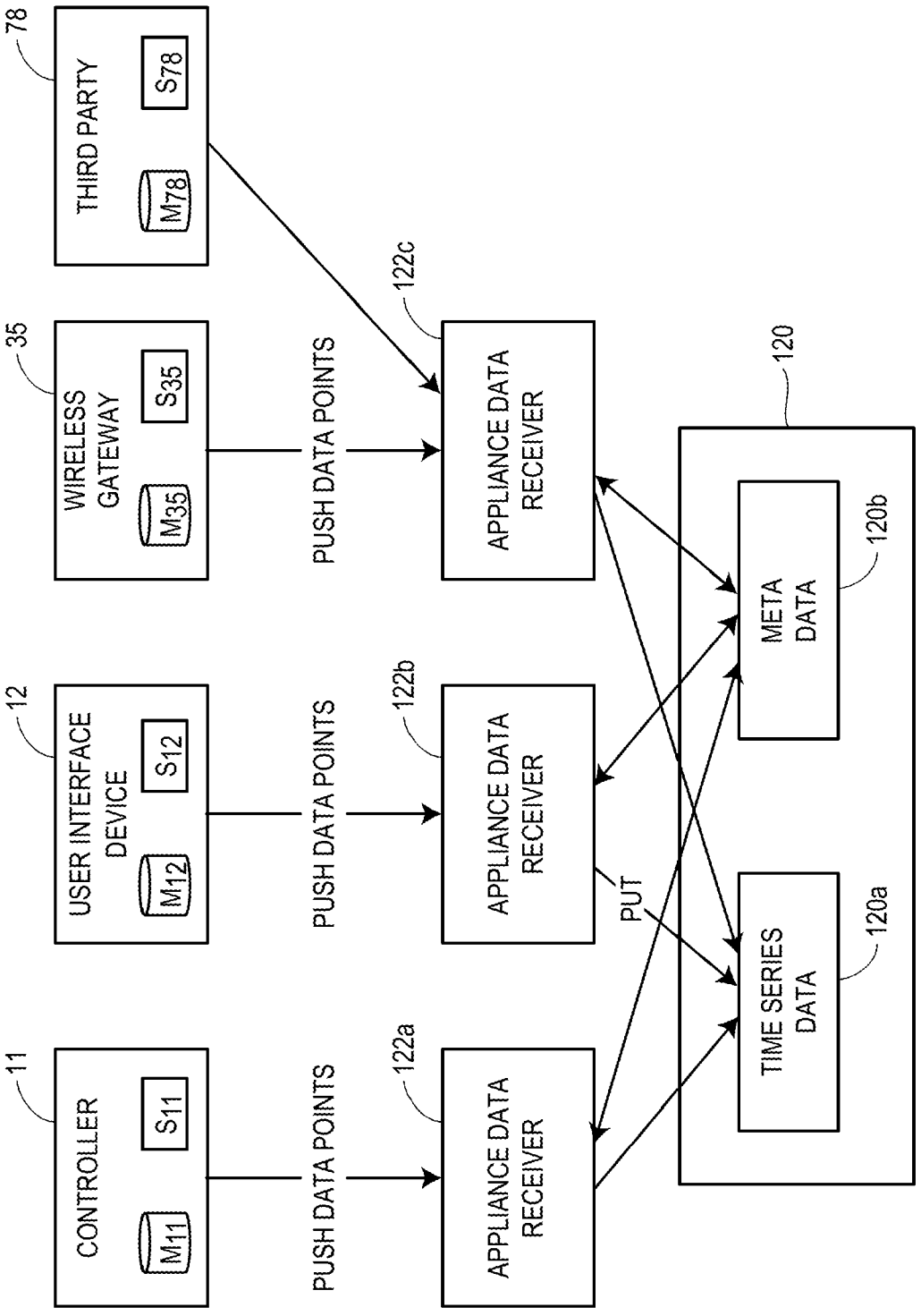
78
THIRD PARTY
S78
M78
35
WIRELESS GATEWAY
S35
M35
12
USER INTERFACE DEVICE
S12
M12
11
CONTROLLER
S11
M11
PUSH DATA POINTS
PUSH DATA POINTS
PUSH DATA POINTS
122c
APPLIANCE DATA RECEIVER
122b
APPLIANCE DATA RECEIVER
122a
APPLIANCE DATA RECEIVER
PUT
120
META DATA
120b
TIME SERIES DATA
120a

FIG. 3

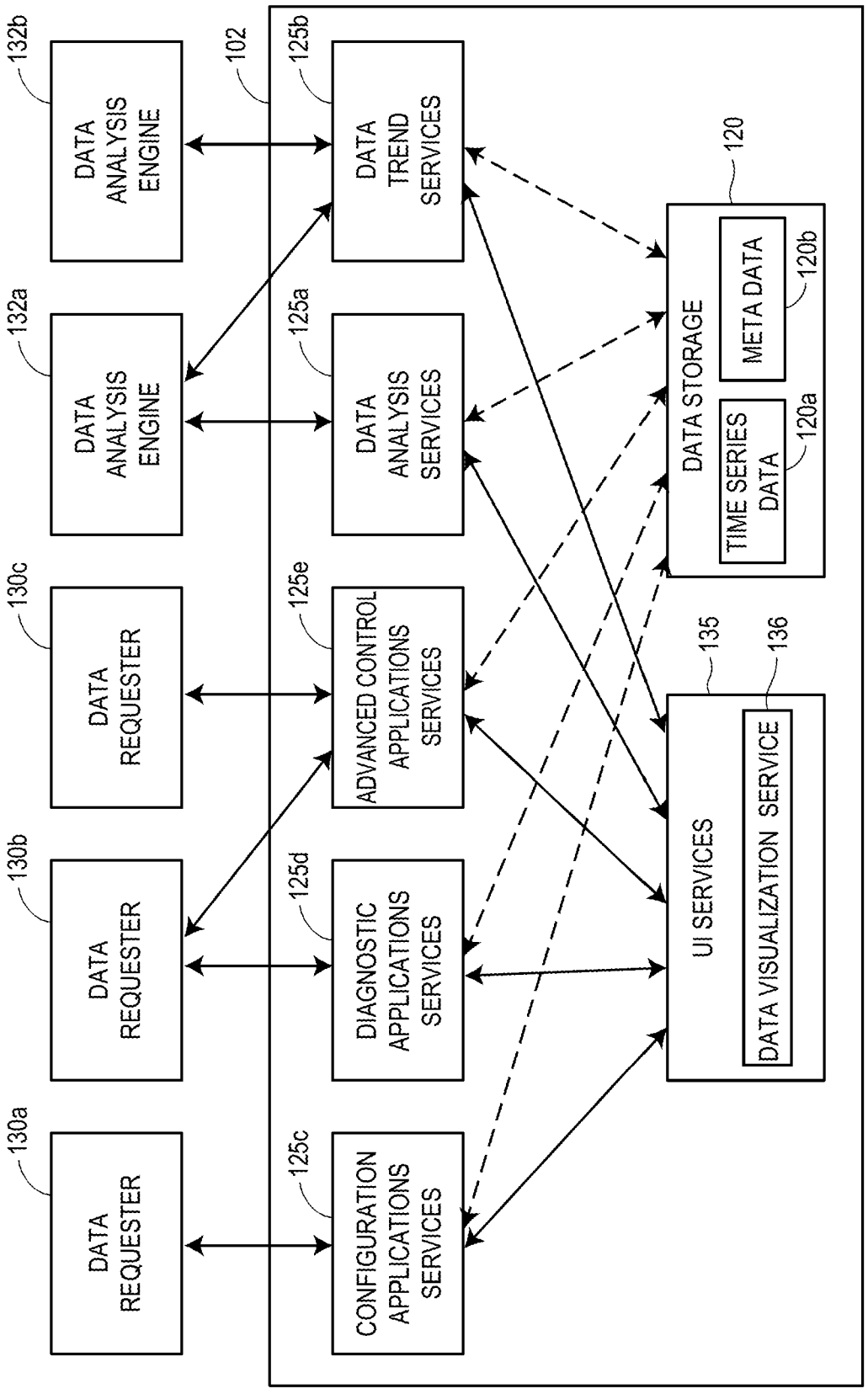
132b
DATA ANALYSIS ENGINE
102
125b
DATA TREND SERVICES
120
132a
DATA ANALYSIS ENGINE
125a
DATA ANALYSIS SERVICES
DATA STORAGE
META DATA
120b
TIME SERIES DATA
120a
130c
DATA REQUESTER
125e
ADVANCED CONTROL APPLICATIONS SERVICES
135
136
130b
DATA REQUESTER
125d
DIAGNOSTIC APPLICATIONS SERVICES
UI SERVICES
DATA VISUALIZATION SERVICE
130a
DATA REQUESTER
125c
CONFIGURATION APPLICATIONS SERVICES

FIG. 4

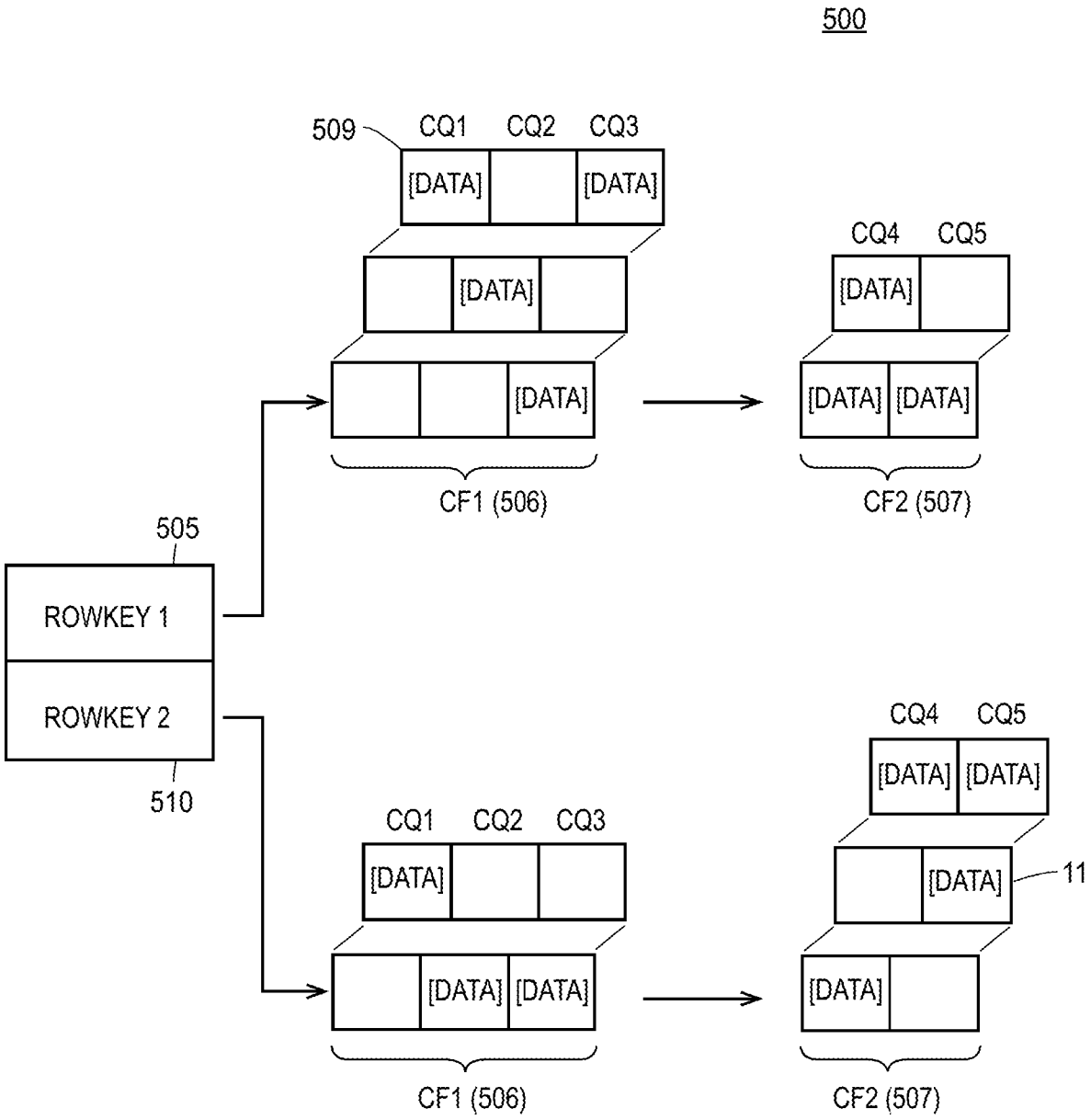
500
509
CQ1
CQ2
CQ3
[DATA]
[DATA]
[DATA]
[DATA]
[DATA]
CQ4
CQ5
[DATA]
[DATA]
[DATA]
CF1 (506)
CF2 (507)
505
ROWKEY 1
ROWKEY 2
510
CQ1
CQ2
CQ3
[DATA]
[DATA]
[DATA]
CQ4
CQ5
[DATA]
[DATA]
[DATA]
[DATA]
11
CF1 (506)
CF2 (507)

FIG. 5A

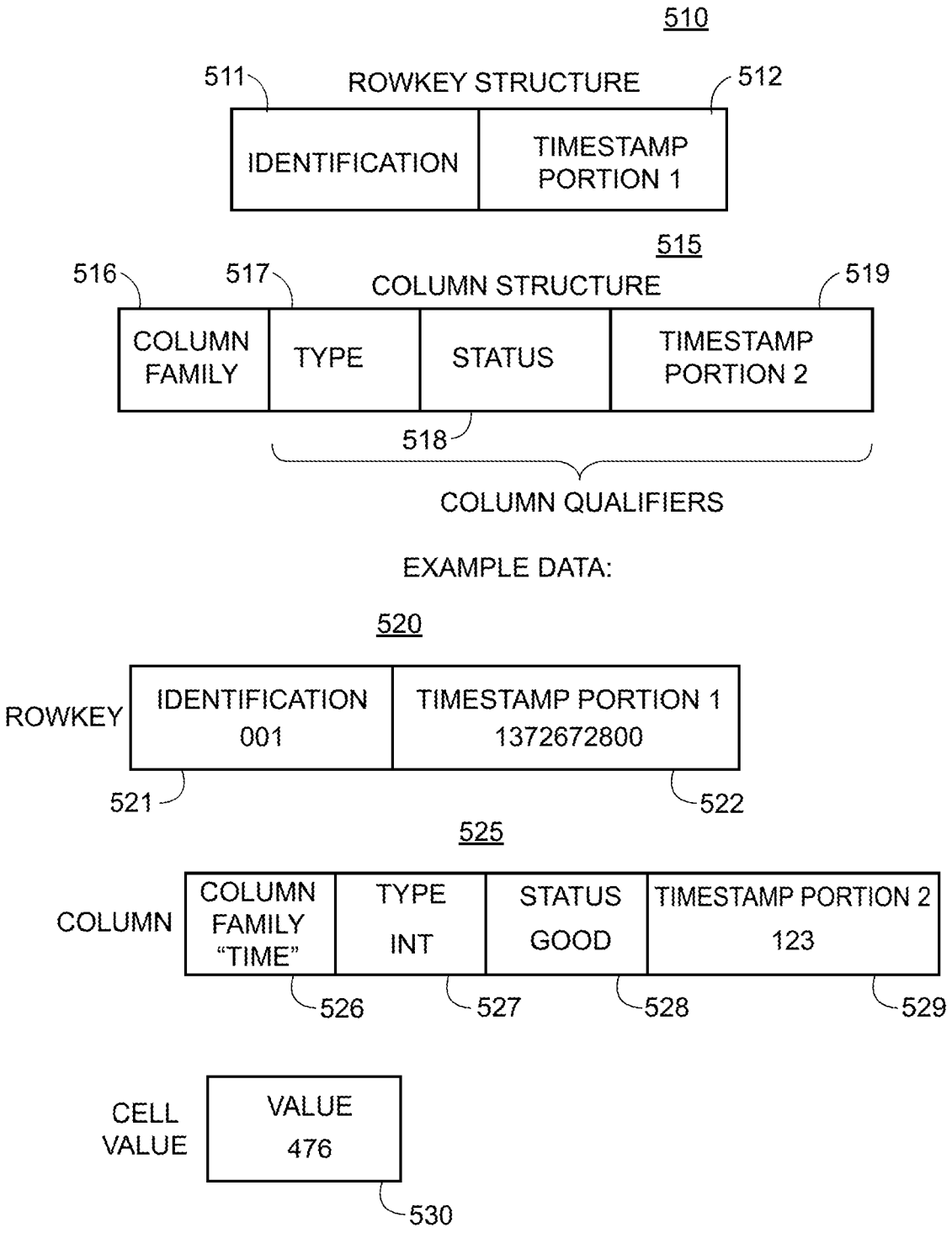
510
511
ROWKEY STRUCTURE
512
IDENTIFICATION
TIMESTAMP PORTION 1
515
516
517
COLUMN STRUCTURE
519
COLUMN FAMILY
TYPE
STATUS
TIMESTAMP PORTION 2
518
COLUMN QUALIFIERS
EXAMPLE DATA:
520
ROWKEY
IDENTIFICATION
001
TIMESTAMP PORTION 1
1372672800
521
522
525
COLUMN
COLUMN FAMILY "TIME"
TYPE
INT
STATUS
GOOD
TIMESTAMP PORTION 2
123
526
527
528
529
CELL VALUE
VALUE
476
530

FIG. 5B

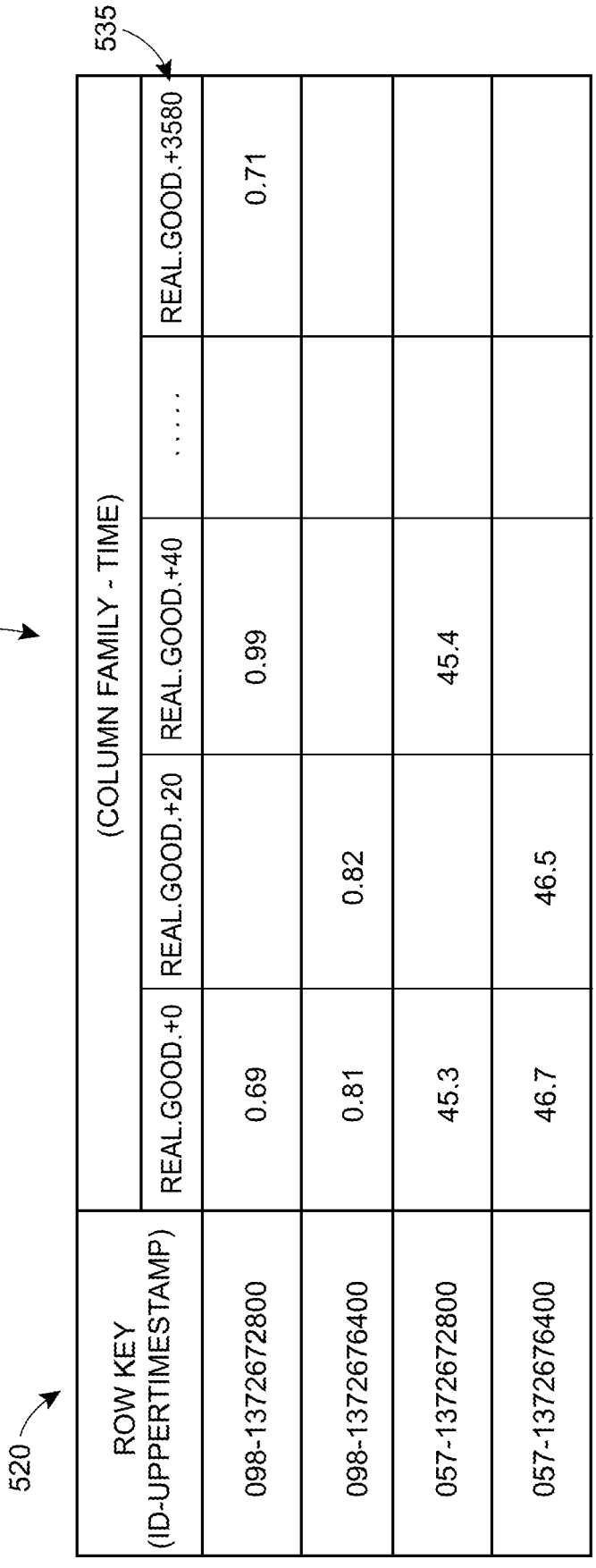
520
531
535
ROW KEY (ID-UPPERTIMESTAMP)
(COLUMN FAMILY - TIME)
REAL.GOOD.+0
REAL.GOOD.+20
REAL.GOOD.+40
.....
REAL.GOOD.+3580
098-1372672800
0.69
0.99
0.71
098-1372676400
0.81
0.82
057-1372672800
45.3
45.4
057-1372676400
46.7
46.5

FIG. 5C

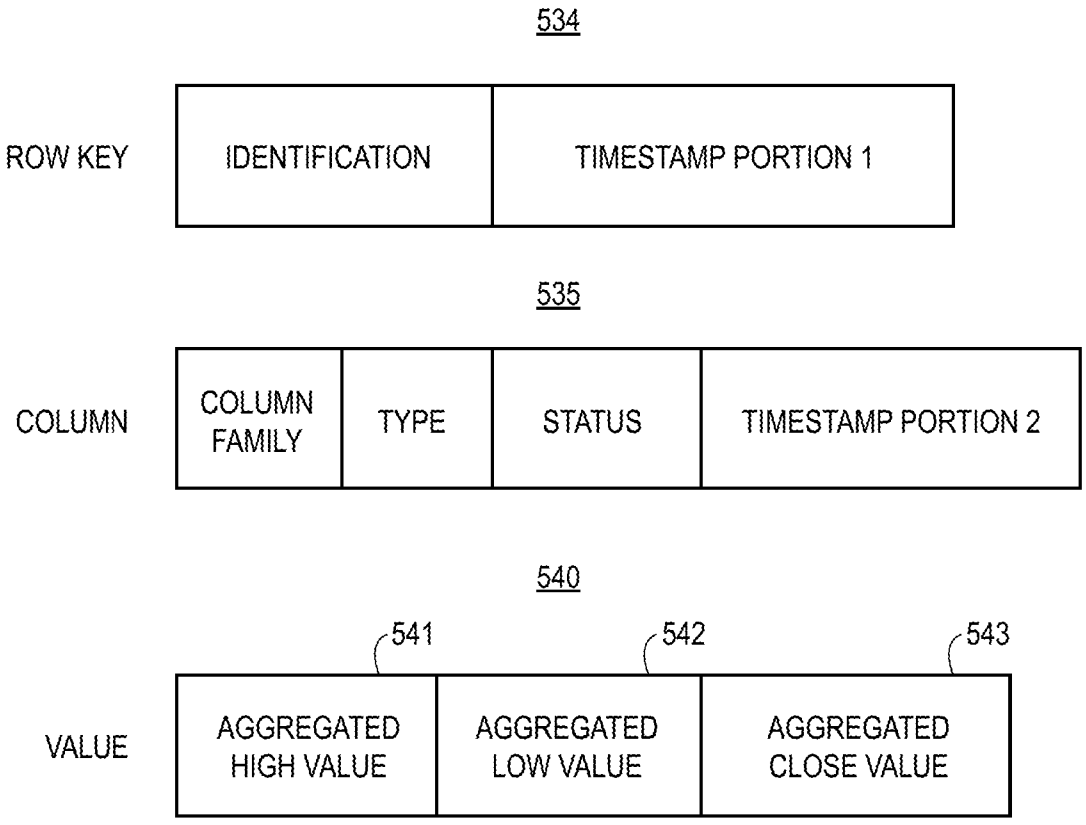
534
ROW KEY
IDENTIFICATION
TIMESTAMP PORTION 1
535
COLUMN
COLUMN FAMILY
TYPE
STATUS
TIMESTAMP PORTION 2
540
541
542
543
VALUE
AGGREGATED HIGH VALUE
AGGREGATED LOW VALUE
AGGREGATED CLOSE VALUE

FIG. 5D

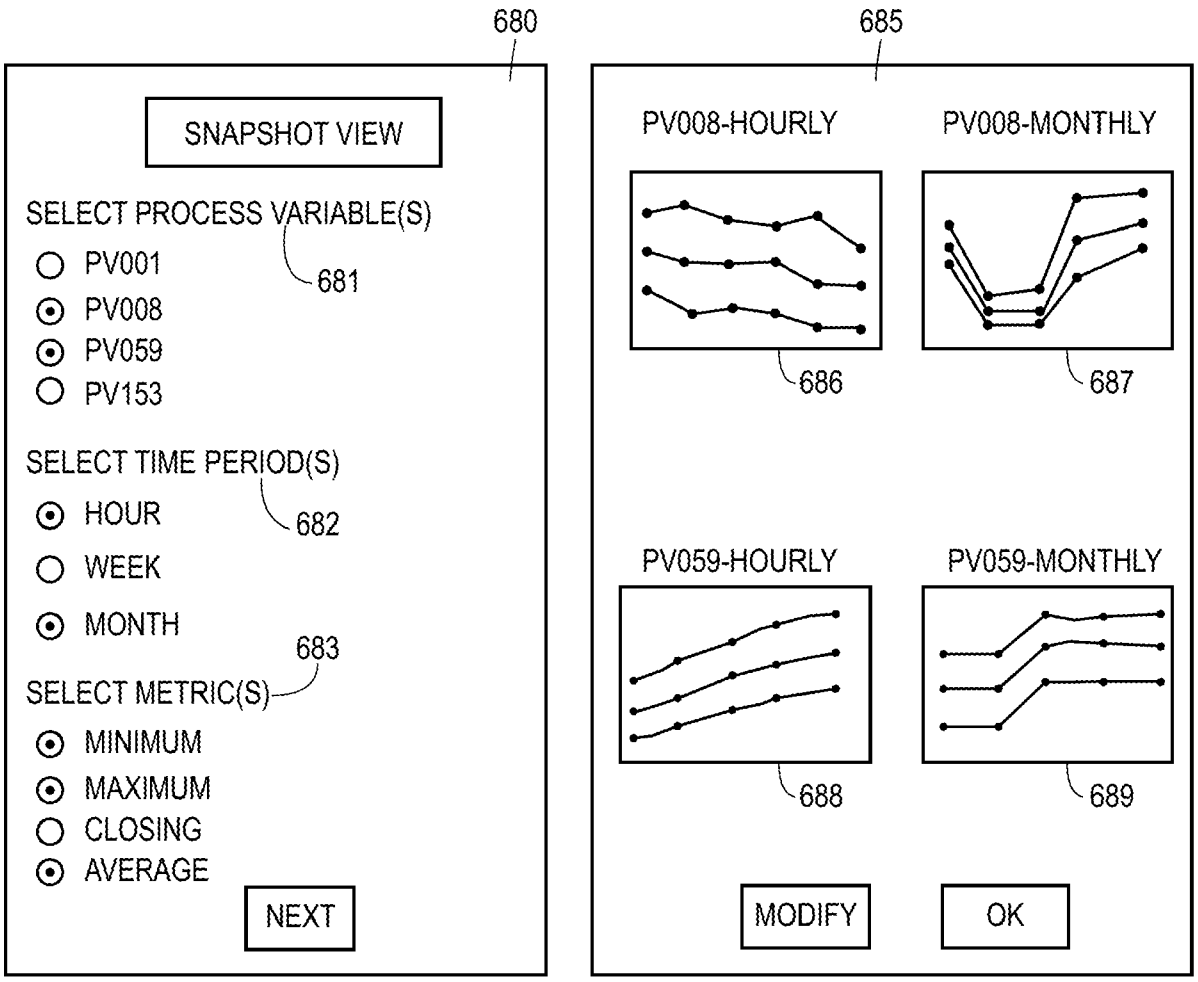
680
685
SNAPSHOT VIEW
SELECT PROCESS VARIABLE(S)
681
PV001
PV008
PV059
PV153
SELECT TIME PERIOD(S)
682
HOUR
WEEK
MONTH
683
SELECT METRIC(S)
MINIMUM
MAXIMUM
CLOSING
AVERAGE
NEXT
PV008-HOURLY
PV008-MONTHLY
686
687
PV059-HOURLY
PV059-MONTHLY
688
689
MODIFY
OK

FIG. 6

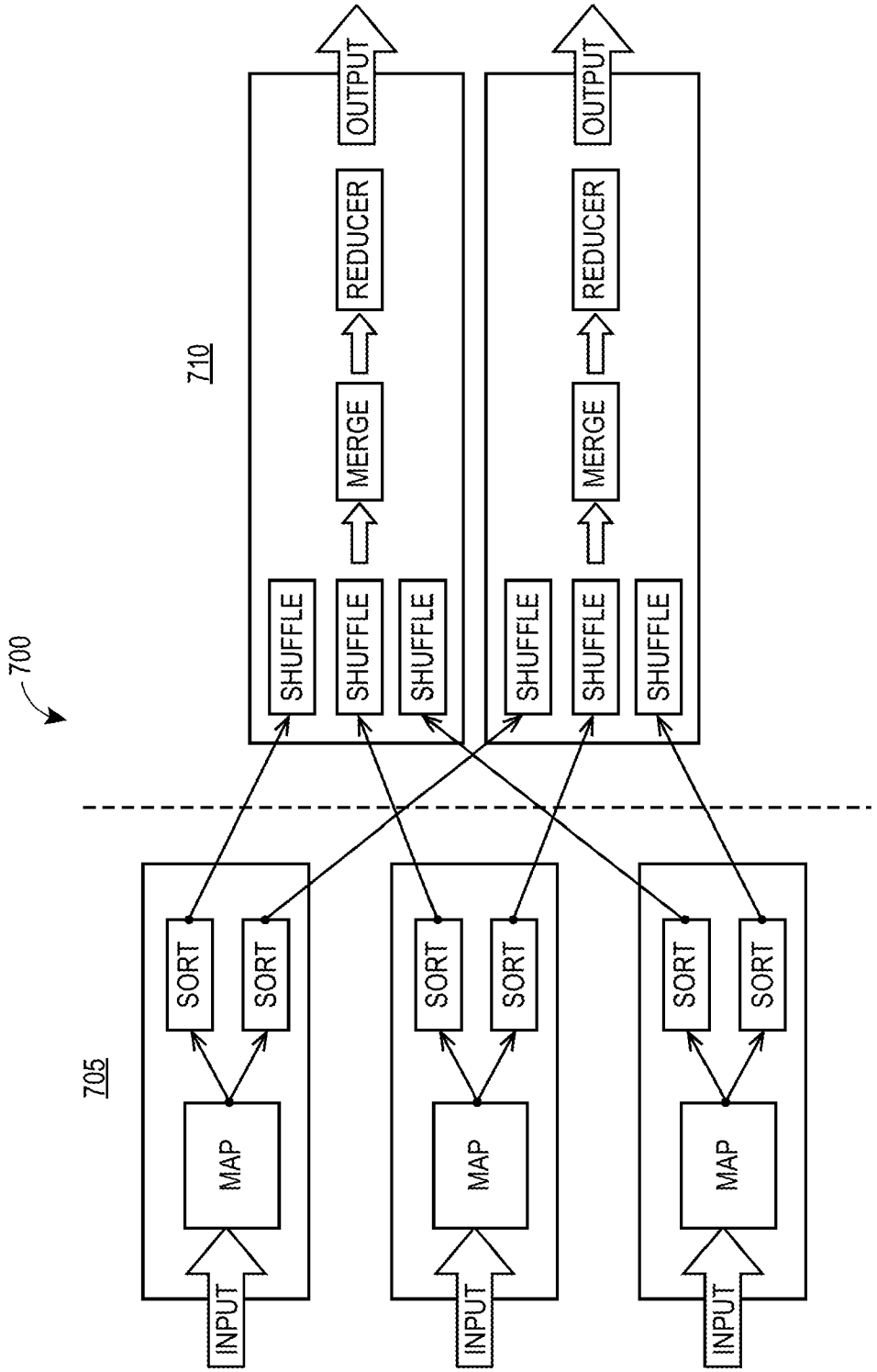
700
705
710
INPUT
MAP
SORT
SORT
SHUFFLE
SHUFFLE
SHUFFLE
MERGE
REDUCER
OUTPUT

FIG. 7

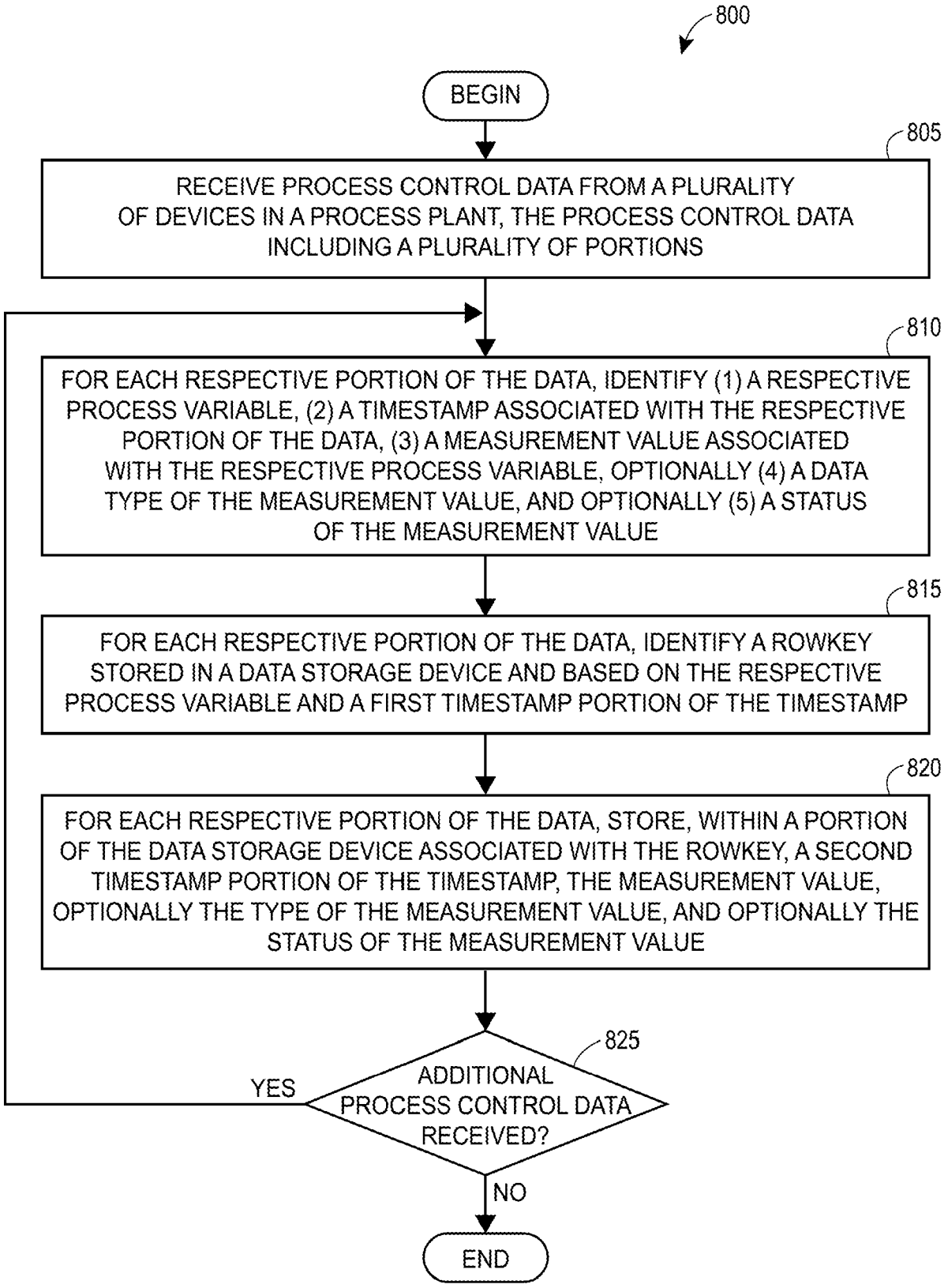
800
BEGIN
805
RECEIVE PROCESS CONTROL DATA FROM A PLURALITY OF DEVICES IN A PROCESS PLANT, THE PROCESS CONTROL DATA INCLUDING A PLURALITY OF PORTIONS
810
FOR EACH RESPECTIVE PORTION OF THE DATA, IDENTIFY (1) A RESPECTIVE PROCESS VARIABLE, (2) A TIMESTAMP ASSOCIATED WITH THE RESPECTIVE PORTION OF THE DATA, (3) A MEASUREMENT VALUE ASSOCIATED WITH THE RESPECTIVE PROCESS VARIABLE, OPTIONALLY (4) A DATA TYPE OF THE MEASUREMENT VALUE, AND OPTIONALLY (5) A STATUS OF THE MEASUREMENT VALUE
815
FOR EACH RESPECTIVE PORTION OF THE DATA, IDENTIFY A ROWKEY STORED IN A DATA STORAGE DEVICE AND BASED ON THE RESPECTIVE PROCESS VARIABLE AND A FIRST TIMESTAMP PORTION OF THE TIMESTAMP
820
FOR EACH RESPECTIVE PORTION OF THE DATA, STORE, WITHIN A PORTION OF THE DATA STORAGE DEVICE ASSOCIATED WITH THE ROWKEY, A SECOND TIMESTAMP PORTION OF THE TIMESTAMP, THE MEASUREMENT VALUE, OPTIONALLY THE TYPE OF THE MEASUREMENT VALUE, AND OPTIONALLY THE STATUS OF THE MEASUREMENT VALUE
825
ADDITIONAL PROCESS CONTROL DATA RECEIVED?
YES
NO
END

FIG. 8

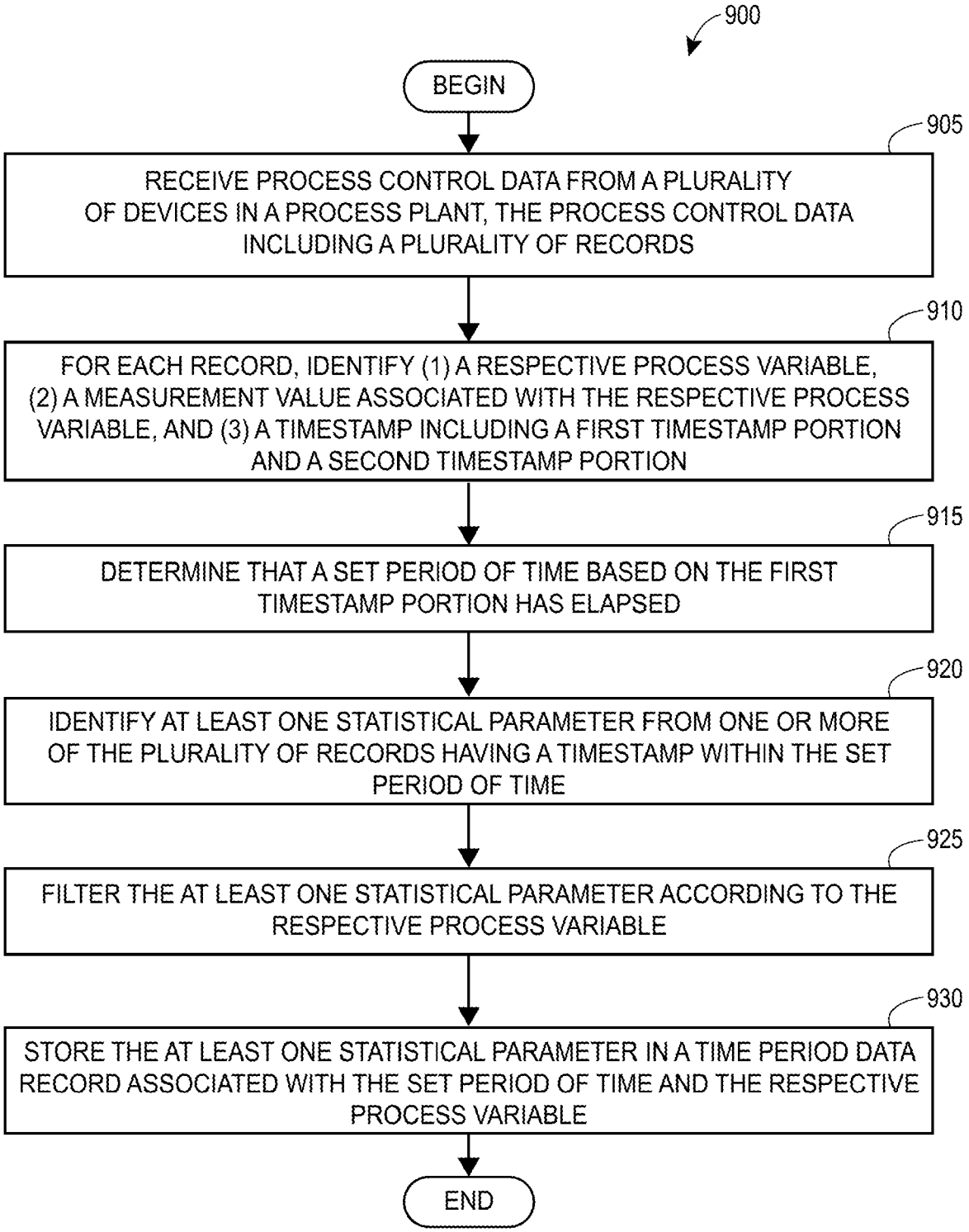
900
BEGIN
905
RECEIVE PROCESS CONTROL DATA FROM A PLURALITY OF DEVICES IN A PROCESS PLANT, THE PROCESS CONTROL DATA INCLUDING A PLURALITY OF RECORDS
910
FOR EACH RECORD, IDENTIFY (1) A RESPECTIVE PROCESS VARIABLE, (2) A MEASUREMENT VALUE ASSOCIATED WITH THE RESPECTIVE PROCESS VARIABLE, AND (3) A TIMESTAMP INCLUDING A FIRST TIMESTAMP PORTION AND A SECOND TIMESTAMP PORTION
915
DETERMINE THAT A SET PERIOD OF TIME BASED ON THE FIRST TIMESTAMP PORTION HAS ELAPSED
920
IDENTIFY AT LEAST ONE STATISTICAL PARAMETER FROM ONE OR MORE OF THE PLURALITY OF RECORDS HAVING A TIMESTAMP WITHIN THE SET PERIOD OF TIME
925
FILTER THE AT LEAST ONE STATISTICAL PARAMETER ACCORDING TO THE RESPECTIVE PROCESS VARIABLE
930
STORE THE AT LEAST ONE STATISTICAL PARAMETER IN A TIME PERIOD DATA RECORD ASSOCIATED WITH THE SET PERIOD OF TIME AND THE RESPECTIVE PROCESS VARIABLE
END

FIG. 9

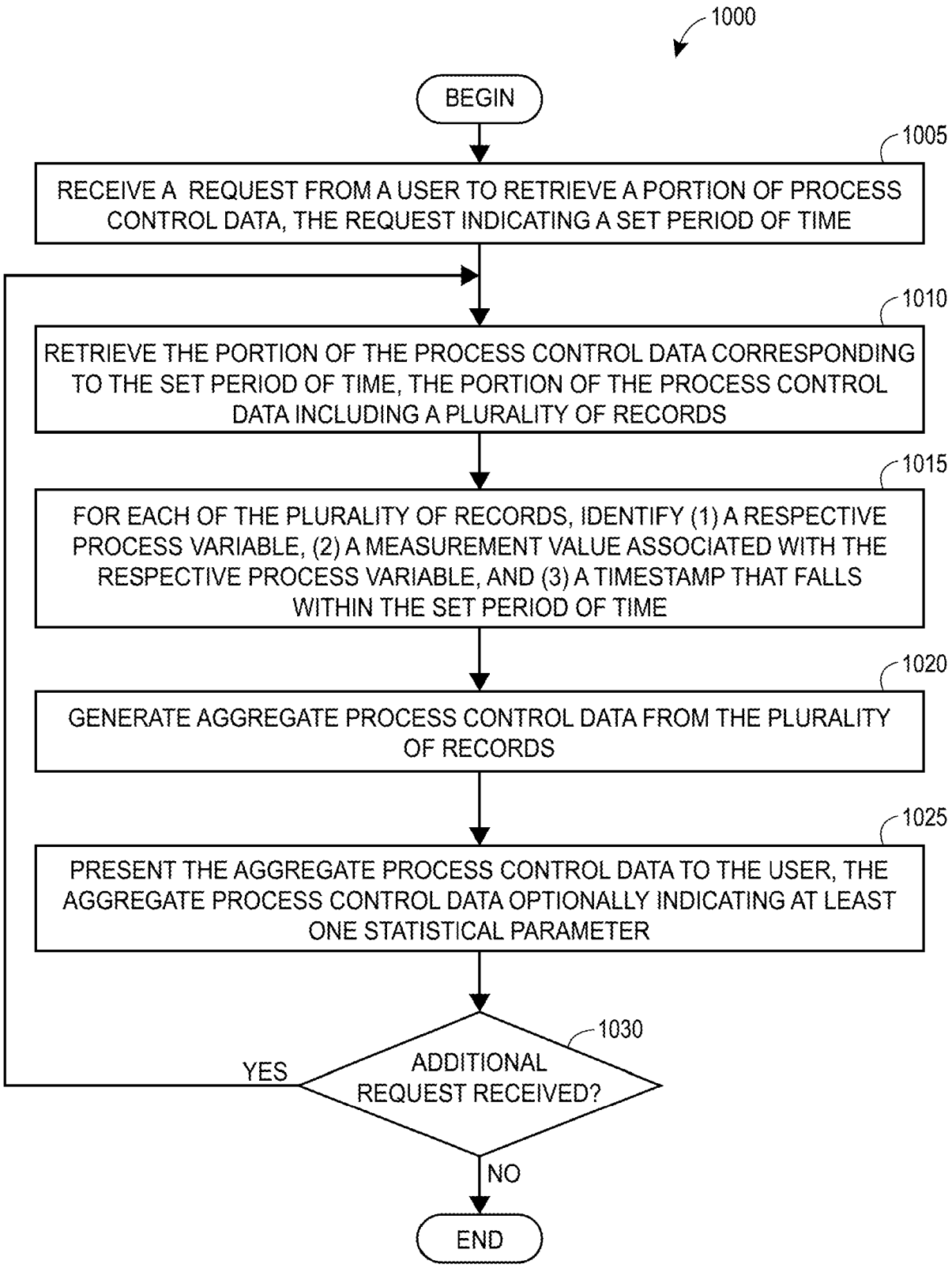
1000
BEGIN
1005
RECEIVE A REQUEST FROM A USER TO RETRIEVE A PORTION OF PROCESS CONTROL DATA, THE REQUEST INDICATING A SET PERIOD OF TIME
1010
RETRIEVE THE PORTION OF THE PROCESS CONTROL DATA CORRESPONDING TO THE SET PERIOD OF TIME, THE PORTION OF THE PROCESS CONTROL DATA INCLUDING A PLURALITY OF RECORDS
1015
FOR EACH OF THE PLURALITY OF RECORDS, IDENTIFY (1) A RESPECTIVE PROCESS VARIABLE, (2) A MEASUREMENT VALUE ASSOCIATED WITH THE RESPECTIVE PROCESS VARIABLE, AND (3) A TIMESTAMP THAT FALLS WITHIN THE SET PERIOD OF TIME
1020
GENERATE AGGREGATE PROCESS CONTROL DATA FROM THE PLURALITY OF RECORDS
1025
PRESENT THE AGGREGATE PROCESS CONTROL DATA TO THE USER, THE AGGREGATE PROCESS CONTROL DATA OPTIONALLY INDICATING AT LEAST ONE STATISTICAL PARAMETER
1030
ADDITIONAL REQUEST RECEIVED?
YES
NO
END

FIG. 10

MANAGING BIG DATA IN PROCESS CONTROL SYSTEMS

[0001] This patent relates generally to process plants and to process control systems, and more particularly, to the storage and management of big data in process plants and in process control system.

[0002] Distributed process control systems, like those used in chemical, petroleum or other process plants, typically include one or more process controllers communicatively coupled to one or more field devices via analog, digital or combined analog/digital buses, or via a wireless communication link or network. The field devices, which may be, for example, valves, valve positioners, switches and transmitters (e.g., temperature, pressure, level and flow rate sensors), are located within the process environment and generally perform physical or process control functions such as opening or closing valves, measuring process parameters, etc. to control one or more process executing within the process plant or system. Smart field devices, such as the field devices conforming to the well-known Fieldbus protocol may also perform control calculations, alarming functions, and other control functions commonly implemented within the controller. The process controllers, which are also typically located within the plant environment, receive signals indicative of process measurements made by the field devices and/or other information pertaining to the field devices and execute a controller application that runs, for example, different control modules which make process control decisions, generate control signals based on the received information and coordinate with the control modules or blocks being performed in the field devices, such as HART®, WirelessHART®, and FOUNDATION® Fieldbus field devices. The control modules in the controller send the control signals over the communication lines or links to the field devices to thereby control the operation of at least a portion of the process plant or system.

14 06 21

[0003] Information from the field devices and the controller is usually made available over a data highway to one or more other hardware devices, such as operator workstations, personal computers or computing devices, data historians, report generators, centralized databases, or other centralized administrative computing devices that are typically placed in control rooms or other locations away from the harsher plant environment. Each of these hardware devices typically is centralized across the process plant or across a portion of the process plant. These hardware devices run applications that may, for example, enable an operator to perform functions with respect to controlling a process and/or operating the process plant, such as changing settings of the process control routine, modifying the operation of the control modules within the controllers or the field devices, viewing the current state of the process, viewing alarms generated by field devices and controllers, simulating the operation of the process for the purpose of training personnel or testing the process control software, keeping and updating a configuration database, etc. The data highway utilized by the hardware devices, controllers and field devices may include a wired communication path, a wireless communication path, or a combination of wired and wireless communication paths.

[0004] As an example, the DeltaV™ control system, sold by Emerson Process Management, includes multiple applications stored within and executed by different devices located at diverse places within a process plant. A configuration application, which resides in one or more workstations or computing devices, enables users to create or change process control modules and download these process control modules via a data highway to dedicated distributed controllers. Typically, these control modules are made up of communicatively interconnected function blocks, which are objects in an object-oriented programming

protocol that perform functions within the control scheme based on inputs thereto and that provide outputs to other function blocks within the control scheme. The configuration application may also allow a configuration designer to create or change operator interfaces which are used by a viewing application to display data to an operator and to enable the operator to change settings, such as set points, within the process control routines. Each dedicated controller and, in some cases, one or more field devices, stores and executes a respective controller application that runs the control modules assigned and downloaded thereto to implement actual process control functionality. The viewing applications, which may be executed on one or more operator workstations (or on one or more remote computing devices in communicative connection with the operator workstations and the data highway), receive data from the controller application via the data highway and display this data to process control system designers, operators, or users using the user interfaces, and may provide any of a number of different views, such as an operator's view, an engineer's view, a technician's view, etc. A data historian application is typically stored in and executed by a data historian device that collects and stores some or all of the data provided across the data highway while a configuration database application may run in a still further computer attached to the data highway to store the current process control routine configuration and data associated therewith. Alternatively, the configuration database may be located in the same workstation as the configuration application.

[0005] The architecture of currently known process control plants and process control systems is strongly influenced by limited controller and device memory, communication bandwidth and controller and device processor capability. For example, in currently known process control system architectures, the use of dynamic and static non-volatile memory in the controller is usually minimized or, at the least, managed carefully. As a result, during system configuration (e.g., *a priori*), a user typically must choose which data in the controller is to be archived or saved, the frequency at which it will be saved, and whether or not compression is used, and the controller is accordingly configured with this limited set of data rules. Consequently, data which could be useful in troubleshooting and process analysis is often not archived, and if it is collected, the useful information may have been lost due to data compression.

[0006] Additionally, to minimize controller memory usage in currently known process control systems, selected data that is to be archived or saved (as indicated by the configuration of the controller) is reported to the workstation or computing device for storage at an appropriate data historian or data silo. The current techniques used to report the data poorly utilizes communication resources and induces excessive controller loading. Additionally, due to the time delays in communication and sampling at the historian or silo, the data collection and timestamping is often out of sync with the actual process.

[0007] Similarly, in batch process control systems, to minimize controller memory usage, batch recipes and snapshots of controller configuration typically remain stored at a centralized administrative computing device or location (e.g., at a data silo or historian), and are only transferred to a controller when needed. Such a strategy introduces significant burst loads in the controller and in communications between the workstation or centralized administrative computing device and the controller.

[0008] Furthermore, the capability and performance limitations of relational databases of currently known process control systems, combined with the previous high cost of disk storage, play a large part in structuring data into independent entities or silos to meet the objectives of specific applications. For example, within the DeltaV™ system, the archiving of process models, continuous historical data, and batch and event data are

14 06 21

saved in three different application databases or silos of data. Each silo has a different interface to access the data stored therein.

[0009] Structuring data in this manner creates a barrier in the way that historized data is accessed and used. For example, the root cause of variations in product quality may be associated with data in more than one of these data silos. However, because of the different file structures of the silos, it is not possible to provide tools that allow this data to be quickly and easily accessed for analysis. Further, audit or synchronizing functions must be performed to ensure that data across different silos is consistent.

[0010] The limitations of currently known process plants and process control system discussed above and other limitations may undesirably manifest themselves in the operation and optimization of process plants or process control systems, for instance, during plant operations, trouble shooting, and/or predictive modeling. For example, such limitations force cumbersome and lengthy work flows that must be performed in order to obtain data for troubleshooting and generating updated models. Additionally, the obtained data may be inaccurate due to data compression, insufficient bandwidth, or shifted timestamps.

[0011] "Big data" generally refers to a collection of one or more data sets that are so large or complex that traditional database management tools and/or data processing applications (e.g., relational databases and desktop statistic packages) are not able to manage the data sets within a tolerable amount of time. Typically, applications that use big data are transactional and end-user directed or focused. For example, web search engines, social media applications, marketing applications and retail applications may use and manipulate big data. Big data may be supported by a distributed database which allows the parallel processing capability of modern multi-process, multi-core servers to be fully utilized.

14 06 21

[0012] Current techniques for storing, accessing, and processing big data, and especially big data associated with process plants and process control systems, are inefficient. For example, various existing process plants use relational databases configured to store process control data which, in some cases, results in too much allocated storage and long retrieval times. Further, the storage of continuous historical data does not enable users or administrators to efficiently or effectively process trends or identify parameters, or combinations of parameters, from multiple data entries. Accordingly, there is an opportunity to develop techniques to more effectively and efficiently organize, process, and manage big data associated with process plants and process control systems.

SUMMARY

[0013] A process control system or plant provides an infrastructure for supporting large-scale data mining and data analytics of process control data. A process control data network incorporates a big data schema which stores process control data and attributes thereof using lightweight non-relational database storage techniques. Using these techniques, the big data schema need not allocate storage for various process control attributes that are not present in the process control data. Further, the big data schema organizes the process control data into tables having rowkeys and column families to enable users and administrators to efficiently locate, access, and analyze the stored data.

[0014] The big data schema may create the rowkeys and the column families using various combinations of the process control data and attributes thereof. Generally, the rowkeys are unique key values that

organize the data within the big data schema and that users may use to query and retrieve specific data. For example, some rowkeys incorporate timestamps (or portions thereof) corresponding to when the process control data is recorded. Each column family includes one or more column qualifiers that the big data schema creates using process control data attributes. The big data schema stores relevant measurements or values based on the corresponding rowkeys and column qualifiers. Some rowkeys may have multiple associated measurements (and multiple associated column qualifiers), thus resulting in a three-dimensional storage schema.

[0015] The big data schema also includes techniques for periodically creating and storing "snapshot" data corresponding to the underlying stored data. For example, for every elapsed minute, the big data schema may calculate and store the minimum, maximum, mean, and standard deviation for the underlying data having a timestamp within that minute. The big data schema thus enables a user to access specified data (e.g., a specific process variable) over a specified time period (e.g., hourly, weekly, monthly). The big data schema then presents the data to the user in an interface to enable the user to efficiently and effectively assess the snapshot data and perform desired data analyses.

[0016] An aspect of the present invention provides a system for storing process control data associated with a process control plant, comprising: a data storage device configured to store process control data corresponding to at least one process that is controlled by a plurality of devices in the process control plant; and a data receiver configured to receive the process control data from the plurality of devices, the process control data including a plurality of portions, and each portion of the data associated with a respective process variable; and a processor adapted to interface with the data storage device and the data receiver and configured to, for the each portion of the data: identify a rowkey stored in the data storage device, the rowkey based on (i) at least a portion of a first attribute of the each portion of the data and (ii) a first portion of a third attribute of the each portion of the data, identify a column qualifier stored in the data storage device, the column qualifier based on (i) at least a portion of a second attribute of the each portion of the data and (ii) a second portion of the third attribute, and store, within a portion of the data storage device associated with the rowkey and the column qualifier, a measurement value associated with the respective process variable corresponding to the each portion of the data.

14 06 21

[0017] The first attribute may include one of: a timestamp corresponding to the measurement value associated with the respective process variable, an identification of the respective process variable, a type of data corresponding to the measurement value, or a status of the respective process variable; and the second attribute may include another one of: the timestamp corresponding to the measurement value associated with the respective process variable, the identification of the respective process variable, the type of data corresponding to the measurement value, or the status of the respective process variable.

[0018] The system may be as above, wherein at least one of: the rowkey may be a concatenation of the at least the portion of the first attribute and the first portion of the third attribute; or the column qualifier may be a concatenation of the at least the portion of the second attribute and the second portion of the third attribute.

[0019] The third attribute may be a timestamp corresponding to the measurement value associated with the respective process variable.

[0020] The first portion of the timestamp may correspond to an upper timestamp of the timestamp and the second portion of the timestamp may correspond to a lower timestamp of the timestamp.

[0021] The timestamp may be represented as a UNIX epoch value, the first portion of the timestamp may be the timestamp rounded down by a certain degree, and the second portion of the timestamp may be a remainder of the timestamp that was rounded down by the certain degree.

[0022] The column qualifier may be a first column qualifier of a plurality of column qualifiers, each column qualifier corresponding to at least a portion of a different attribute of the each respective portion of the data; the rowkey may be associated with each column qualifier of the plurality of column qualifiers; and the measurement value associated with the each respective portion of the data may be stored within a portion of the data storage device associated with the row key and each column qualifier of the plurality of column qualifiers.

[0023] The plurality of column qualifiers may be associated with a first column family of a plurality of column families, each column family indicating a respective category of a respective plurality of column qualifiers; the rowkey may be associated each column family of the plurality of column families; and the measurement value associated with the each respective portion of the data may be stored within a portion of the data storage device associated with the rowkey, the column qualifier, and the first column family.

[0024] The first attribute may be an identification of the respective process variable.

[0025] The second attribute may be a status of the respective process variable.

[0026] The processor may be further configured to, for the each portion of the data, identify another column qualifier based on a fourth attribute of the each portion of the data, the fourth attribute being a type of data corresponding to the measurement value associated with the respective process variable; and the measurement value associated with the each respective portion of the data may be stored within a portion of the data storage device associated with the rowkey, the column qualifier, and the another column qualifier.

[0027] The measurement value associated with the respective process variable may be a statistical measurement value of the respective process variable determined from a plurality of measurement values of the respective process variable.

[0028] An aspect of the present invention provides a method for storing process control data associated with a process control plant, the method comprising: receiving process control data from a plurality of devices configured to control at least one process in the process control plant, the process control data including a plurality of portions, and each portion of the data associated with a respective process variable; and for the each portion of the data: identifying, by a processor based on (i) at least a portion of a first attribute of the each portion of the data and (ii) a first portion of a third attribute of the each portion of the data, a rowkey stored in a data storage device; identifying, by the processor based on (i) at least a portion of a second attribute of the each portion of the data and (ii) a second portion of the third attribute, a column qualifier stored in the data storage device; and storing, within a portion of the data storage device associated with the rowkey and the column qualifier, a measurement value associated with the respective process variable corresponding to the each portion of the data.

[0029] Identifying the rowkey based on the at least the portion of the first attribute may include identifying the rowkey based on one of: a timestamp corresponding to the measurement value associated with the

14 06 21

respective process variable, an identification of the respective process variable, a type of data corresponding to the measurement value, or a status of the respective process variable; and identifying the column qualifier based on the at least the portion of the second attribute may include identifying the column qualifier based on another one of: the timestamp corresponding to the measurement value associated with the respective process variable, the identification of the respective process variable, the type of data corresponding to the measurement value, or the status of the respective process variable.

[0030] The method may be as above, wherein at least one of: the rowkey may be a concatenation of the at least the portion of the first attribute and the first portion of the third attribute; or the column qualifier may be a concatenation of the at least the portion of the second attribute and the second portion of the third attribute.

[0031] The third attribute may be a timestamp corresponding to the measurement value associated with the respective process variable.

[0032] The first portion of the timestamp may correspond to an upper timestamp of the timestamp and the second portion of the timestamp may correspond to a lower timestamp of the timestamp; or the timestamp may be represented as a UNIX epoch value, the first portion of the timestamp may be the timestamp rounded down by a certain degree, and the second portion of the timestamp may be a remainder of the timestamp that was rounded down by the certain degree.

[0033] The column qualifier may be a first column qualifier of a plurality of column qualifiers; the rowkey may be associated with each column qualifier of the plurality of column qualifiers; each column qualifier may be based on at least a portion of a different attribute of the each respective portion of the data; and storing the measurement value within the portion of the data storage device associated with the rowkey and the column qualifier may include storing the measurement value within a portion of the data storage device associated with the row key and each column qualifier of the plurality of column qualifiers.

[0034] The plurality of column qualifiers may be associated with a first column family of a plurality of column families, each column family indicating a respective category of a respective plurality of column qualifiers; the rowkey may be associated each column family of the plurality of column families; and storing the measurement value within the portion of the data storage device associated with the rowkey and the column qualifier may include storing the measurement value within a portion of the data storage device associated with the rowkey, the column qualifier, and the first column family.

[0035] Identifying the rowkey based on at least a portion of the first attribute may include identifying the rowkey based on an identification of the respective process variable.

[0036] Identifying the column qualifier based on at least a portion of the second attribute may include identifying the column qualifier based on a status of the respective process variable.

[0037] The method may further comprise, for the each portion of the data, identifying another column qualifier based on a fourth attribute of the each portion of the data, the fourth attribute being a type of data corresponding to the measurement value associated with the respective process variable; and wherein storing the measurement value within the portion of the data storage device associated with the rowkey and the column qualifier may include storing the measurement value within a portion of the data storage device associated with the rowkey, the column qualifier, and the another column qualifier.

14 06 21

[0038] Storing the measurement value associated with the respective process variable may include storing a statistical measurement value of the respective process variable determined from a plurality of measurement values of the respective process variable.

[0039] Aspects of the present invention are described herein, by way of example only, with reference to the accompanying drawings, in which:

[0040] FIG. 1 is a block diagram of an example big data network for a process plant or process control system;

[0041] FIG. 2 is a block diagram illustrating an example arrangement of provider nodes included in the process control system big data network of FIG. 1;

[0042] FIG. 3 is a block diagram illustrating an example use of appliance data receivers to store or historize data at the process control system big data appliance of FIG. 1;

[0043] FIG. 4 is a block diagram illustrating an example use of appliance request servicers to access historized data stored at the process control system big data appliance of FIG. 1;

[0044] FIG. 5A is an example table for organizing and storing big data;

[0045] FIG. 5B is an example structure of an example table for organizing and storing big data;

[0046] FIG. 5C is an example big data record configured for storage according to a big data schema;

[0047] FIG. 5D is an example structure for storing "snapshot" data in a big data schema;

[0048] FIG. 6 depicts example user interfaces associated with retrieving and accessing process control data snapshots;

[0049] FIG. 7 is a block diagram illustrating an organization technique for big data;

[0050] FIG. 8 is a flow diagram of an example method of storing big data using rowkeys;

[0051] FIG. 9 is a flow diagram of an example method of storing process control data from multiple data records; and

[0052] FIG. 10 is a flow diagram of an example method of retrieving process control data from multiple data records and performing calculations therefrom.

[0053] FIG. 1 is a block diagram of an example big data network 100 within or associated with a process plant or process control system 10. The example process control system big data network 100 includes a process control system big data apparatus or appliance 102, a process control system big data network backbone 105, and a plurality of nodes 108 that are communicatively connected to the backbone 105. The plurality of nodes 108 may collect and cache process-related data, plant-related data, and other types of data, and the network backbone 105 may deliver the data to the process control system big data apparatus or appliance 102 for long-term storage (e.g., "historization") and processing. In one implementation, at least some of the data may be delivered between nodes of the network 100, e.g., to control a process in real-time.

14 06 21

[0054] The process control system big data network 100 may collect and store any type of data related to the process control system 10. For example, the process control system big data network 100 collects and stores real-time process data such as continuous, batch, measurement and event data that is generated while a process is being controlled in the process plant 10 (and, in some cases, is indicative of an effect of a real-time execution of the process). Process definition, arrangement or set-up data such as configuration data and/or batch recipe data, as well as data corresponding to the configuration, execution and results of process diagnostics may also be collected and stored. Of course, other types of process data may also be collected and stored.

[0055] In addition, the process control system big data network 100 may collect and store data highway traffic and network management data of the backbone 105 and of various other communication networks of the process plant 10. Still further, the process control system big data network 100 may collect and store user-related data such as data related to user traffic, login attempts, queries and instructions, as well as text data (e.g., logs, operating procedures, manuals, etc.), spatial data (e.g., location-based data) and multi-media data (e.g., closed circuit TV, video clips, etc.).

[0056] In addition, the process control system big data network 100 may collect and store data that is related to the process plant 10 (e.g., to physical equipment included in the process plant 10 such as machines and devices) but that may not be generated by applications that directly configure, control, or diagnose a process. For example, the process control system big data network 100 may collect and store vibration data, steam trap data, data indicative of a value of a parameter corresponding to plant safety (e.g., corrosion data, gas detection data, etc.), and/or data indicative of an event corresponding to plant safety. In some cases, the process control system big data network 100 may collect and store data corresponding to the health of machines, plant equipment and/or devices. For example, equipment data (e.g., pump health data determined based on vibration data and other data) may be collected. In some cases, the process control system big data network 100 may collect and store data corresponding to the configuration, execution and results of equipment, machine, and/or device diagnostics.

14 06 21

[0057] Moreover, the process control system big data network 100 may collect and store data generated by or transmitted to entities external to the process plant 10, such as data related to costs of raw materials, expected arrival times of parts or equipment, weather data, and other external data. If desired, all data that is generated, received, or observed by all nodes 108 that are communicatively connected to the network backbone 105 may be collected and caused to be stored at the process control system big data appliance 102.

[0058] As illustrated in FIG. 1, the process control system big data network 100 includes a process control system big data studio 109 configured to provide a primary interface into the process control system big data network 100 for configuration and data exploration, e.g., a user interface or an interface for use by other applications. The process control system big data studio 109 may connect to the big data appliance 102 via the process control system big data network backbone 105, or may directly connect to the process control system big data appliance 102.

[0059] Process Control Big Data Network Nodes

[0060] As illustrated in FIG. 1, the plurality of nodes 108 of the process control big data network 100 may include several different groups of nodes 110, 112, 115. A first group of nodes 110, referred to herein as

"provider nodes 110" or "provider devices 110," may include one or more nodes or devices that generate, route, and/or receive process control data to enable processes to be controlled in real-time in the process plant environment 10. Examples of provider devices or nodes 110 may include devices whose primary function is directed to generating and/or operating on process control data to control a process, e.g., wired and wireless field devices, controllers, or input/output (I/O devices). Other examples of provider devices 110 may include devices whose primary function is to provide access to or routes through one or more communication networks of the process control system (of which the process control big network 100 is one), e.g., access points, routers, interfaces to wired control busses, gateways to wireless communication networks, gateways to external networks or systems, and other such routing and networking devices. Still other examples of provider devices 110 may include devices whose primary function is to temporarily store process data and other related data that is accumulated throughout the process control system 10 and to cause the temporarily stored data to be transmitted for historization at the process control system big data appliance 102.

[0061] At least one of the provider devices 110 may communicatively connect to the process control big data network backbone 105 in a direct manner. In addition, at least one of the provider devices 110 may communicatively connect to the backbone 105 in an indirect manner. For example, a wireless field device may communicatively connect to the backbone 105 via a router, and access point, and a wireless gateway. Typically, provider devices 110 do not have an integral user interface, although some of the provider devices 100 may have the capability to be in communicative connection with a user computing device or user interface, e.g., by communicating over a wired or wireless communication link, or by plugging a user interface device into a port of the provider device 110.

14 06 21

[0062] As illustrated in FIG. 1, a second group of nodes 112, referred to herein as "user interface nodes 112" or user interface devices 112," may include one or more nodes or devices that each has an integral user interface via which a user or operator may interact with the process control system or process plant 10 to perform activities related to the process plant 10 (e.g., configure, view, monitor, test, analyze, diagnose, order, plan, schedule, annotate, and/or other activities). Examples of these user interface nodes or devices 112 may include mobile or stationary computing devices, workstations, handheld devices, tablets, surface computing devices, and any other computing device having a processor, a memory, and an integral user interface. Integrated user interfaces may include a screen, a keyboard, keypad, mouse, buttons, touch screen, touch pad, biometric interface, speakers and microphones, cameras, and/or any other user interface technology. Each user interface node 112 may include one or more integrated user interfaces. The user interface nodes 112 may include a direct connection to the process control big data network backbone 105, or may include in indirect connection to the backbone 105, e.g., via an access point or a gateway. The user interface nodes 112 may communicatively connect to the process control system big data network backbone 105 in a wired manner and/or in a wireless manner.

[0063] Of course, the plurality of nodes 108 of the process control big data network 100 is not limited to only provider nodes 110 and user interface nodes 112. One or more other types of nodes 115 may optionally be included in the plurality of nodes 108. For example, a node of a system that is external to the process plant 10 (e.g., a lab system or a materials handling system) may communicatively connect to the network backbone 105 of the system 100. A node or device 115 may communicatively connect to the

backbone 105 via a direct or an indirect connection. In addition, a node or device 115 may communicatively connect to the backbone 105 via a wired or a wireless connection.

[0064] At least some of the nodes 108 of the process control system big data network 100 may include an integrated firewall. Further, any number of the nodes 108 (e.g., zero nodes, one node, or more than one node) may each include respective memory storage (denoted in FIG. 1 by the icons M_X) to store or cache tasks, measurements, events, and other data in real-time. A memory storage M_X may comprise high density memory storage technology, for example, solid state drive memory, semiconductor memory, optical memory, molecular memory, biological memory, or any other suitable high density memory technology, as well as flash memory. The memory storage M_X (and, in some cases, the flash memory) may be configured to temporarily store or cache data that is generated by, received at, or otherwise observed by its respective node 108. The flash memory M_X of at least some of the nodes 108 (e.g., a controller device) may also store snapshots of node configuration, batch recipes, and/or other data to minimize delay in using this information during normal operation or after a power outage or other event that causes the node to be off-line. If desired, all of the nodes 110, 112 and any number of the nodes 115 may include high density memory storage M_X. It is understood that different types or technologies of high density memory storage M_X may be utilized across the set of nodes 108, or across a subset of nodes included in the set of nodes 108.

[0065] Any number of the nodes 108 (e.g., zero nodes, one node, or more than one node) may each include respective multi-core hardware (e.g., a multi-core processor or another type of parallel processor), as denoted in FIG. 1 by the icons P_{MCX}. At least some of the nodes 108 may designate one of the cores of its respective processor P_{MCX} for caching real-time data at the node and for causing the cached data to be transmitted for storage at the process control system big data appliance 102. Additionally or alternatively, at least some of the nodes 108 may designate more than one of the multiple cores of its respective multi-core processor P_{MCX} for caching real-time data. The one or more designated cores for caching real-time data (and, in some cases, for causing the cached data to be stored at big data appliance 102) may be exclusively designated as such (e.g., the one or more designated cores may perform no other processing except processing related to caching and transmitting big data). In particular, at least some of the nodes 108 may designate one of its cores to perform operations to control a process in the process plant 10. Additionally or alternatively, one or more cores may be designated exclusively for performing operations to control a process, and may not be used to cache and transmit big data. It is understood that different types or technologies of multi-core processors P_{MCX} may be utilized across the set of nodes 108, or across a subset of nodes of the set of nodes 108. All of the nodes 110, 112 and any number of the nodes 115 may include some type of multi-core processor P_{MCX}.

[0066] It is noted, though, that while FIG. 1 illustrates the nodes 108 as each including both a multi-core processor P_{MCX} and a high density memory M_X, each of the nodes 108 is not required to include both a multi-core processor P_{MCX} and a high density memory M_X. For example, some of the nodes 108 may include only a multi-core processor P_{MCX} and not a high density memory M_X, some of the nodes 108 may include only a high density memory M_X and not a multi-core processor P_{MCX}, some of the nodes 108 may include both a multi-core processor P_{MCX} and a high density memory M_X, and/or some of the nodes 108 may include neither a multi-core processor P_{MCX} nor a high density memory M_X.

[0067] Examples of real-time data that may be cached or collected by provider nodes or devices 110 may include measurement data, configuration data, batch data, event data, maintenance data, and/or continuous

14 06 21

data. For instance, real-time data corresponding to configurations, batch recipes, setpoints, outputs, rates, control actions, diagnostics, alarms, events and/or changes thereto may be collected. Other examples of real-time data may include process models, statistics, status data, and network and plant management data.

[0068] Examples of real-time data that user interface nodes or devices 112 may cache or collect may include, for example, user logins, user queries, data captured by a user (e.g., by camera, audio, or video recording device), user commands, creation, modification or deletion of files, a physical or spatial location of a user interface node or device, results of a diagnostic or test performed by the user interface device 112, and other actions or activities initiated by or related to a user interacting with a user interface node 112.

[0069] Collected data may be dynamic or static data. Collected data may include, for example, database data, streaming data, and/or transactional data. Generally, any data that a node 108 generates, receives, or observes may be collected or cached with a corresponding timestamp or indication of a time of collection/caching. In some cases, all data that a node 108 generates, receives, or observes is collected or cached in its memory storage (e.g., high density memory storage M_X) with a respective indication of a time of each datum's collection/caching (e.g., a timestamp).

[0070] Each of the nodes 110, 112 (and, optionally, at least one of the other nodes 115) may be configured to automatically collect or cache real-time data and to cause the collected/cached data to be delivered to the big data appliance 102 and/or to other nodes 108 without requiring lossy data compression, data sub-sampling, or configuring the node for data collection purposes. Unlike prior art process control systems, the identity of data that is collected at the nodes or devices 108 of the process control system big data network 100 need not be configured into the devices 108 *a priori*. Further, the rate at which data is collected at and delivered from the nodes 108 also need not be configured, selected or defined. Instead, the nodes 110, 112 (and, optionally, at least one of the other nodes 115) of the process control big data system 100 may automatically collect all data that is generated by, received at, or obtained by the node at the rate at which the data is generated, received or obtained, and may cause the collected data to be delivered in high fidelity (e.g., without using lossy data compression or any other techniques that may cause loss of original information) to the process control big data appliance 102 and, optionally, to other nodes 108 of the network 100.

[0071] A detailed block diagram illustrating example provider nodes 110 connected to process control big data network backbone 105 is illustrated in FIG. 2. As previously discussed, provider nodes 110 may include devices whose main function is to automatically generate and/or receive process control data that is used to perform functions to control a process in real-time in the process plant environment 10, such as process controllers, field devices and I/O devices. In a process plant environment 10, process controllers receive signals indicative of process measurements made by field devices, process this information to implement a control routine, and generate control signals that are sent over wired or wireless communication links to other field devices to control the operation of a process in the plant 10. Typically, at least one field device performs a physical function (e.g., opening or closing a valve, increasing or decreasing a temperature, etc.) to control the operation of a process, and some types of field devices may communicate with controllers using I/O devices. Process controllers, field devices, and I/O devices may be wired or wireless, and any number and combination of wired and wireless process controllers, field devices and I/O devices may be nodes 110 of the process control big data network 100.

14 06 21

[0072] FIG. 2 illustrates a controller 11 that is communicatively connected to wired field devices 15-22 via input/output (I/O) cards 26 and 28, and that is communicatively connected to wireless field devices 40, 42, 44, 46 via a wireless gateway 35 and the network backbone 105. (In some cases, though, the controller 11 may communicatively connect to the wireless gateway 35 using a communications network other than the backbone 105, such as by using another wired or a wireless communication link.) In FIG. 2, the controller 11 is shown as being a node 110 of the process control system big data network 100, and is directly connected to the process control big data network backbone 105.

[0073] The controller 11, which may be, by way of example, the DeltaV™ controller sold by Emerson Process Management, may operate to implement a batch process or a continuous process using at least some of the field devices 15-22 and 40-46. The controller 11 may communicatively connect to the field devices 15-22 and 40-46 using any desired hardware and software associated with, for example, standard 4-20 mA devices, I/O cards 26, 28, and/or any smart communication protocol such as the FOUNDATION® Fieldbus protocol, the HART® protocol, the WirelessHART® protocol, etc. The controller 11 may additionally or alternatively communicatively connect with at least some of the field devices 15-22 and 40-46 using the big data network backbone 105. In the system illustrated in FIG. 2, the controller 11, the field devices 15-22 and the I/O cards 26, 28 are wired devices, and the field devices 40-46 are wireless field devices. Of course, the wired field devices 15-22 and wireless field devices 40-46 could conform to any other desired standard(s) or protocols, such as any wired or wireless protocols, including any standards or protocols developed in the future.

14 06 21

[0074] The controller 11 of FIG. 2 includes a processor 30 that implements or oversees one or more process control routines (stored in a memory 32), which may include control loops. The processor 30 may communicate with the field devices 15-22 and 40-46 and with other nodes (e.g., nodes 110, 112, 115) that are communicatively connected to the backbone 105. It should be noted that any control routines or modules (including quality prediction and fault detection modules or function blocks) described herein may have parts thereof implemented or executed by different controllers or other devices if so desired. Likewise, the control routines or modules described herein which are to be implemented within the process control system 10 may take any form, including software, firmware, hardware, etc. Control routines may be implemented in any desired software format, such as using object-oriented programming, ladder logic, sequential function charts, function block diagrams, or using any other software programming language or design paradigm. The control routines may be stored in any desired type of memory, such as random access memory (RAM), or read only memory (ROM). Likewise, the control routines may be hard-coded into, for example, one or more EPROMs, EEPROMs, application specific integrated circuits (ASICs), or any other hardware or firmware elements. Thus, the controller 11 may be configured to implement a control strategy or control routine in any desired manner.

[0075] The controller 11 may also implement a control strategy using what are commonly referred to as function blocks, wherein each function block is an object or other part (e.g., a subroutine) of an overall control routine and operates in conjunction with other function blocks (via communications called links) to implement process control loops within the process control system 10. Control based function blocks typically perform one of an input function, such as that associated with a transmitter, a sensor or other process parameter measurement device, a control function, such as that associated with a control routine that performs PID, fuzzy logic, etc. control, or an output function which controls the operation of some device,

such as a valve, to perform some physical function within the process control system 10. Of course, hybrid and other types of function blocks exist. Function blocks may be stored in and executed by the controller 11, which is typically the case when these function blocks are used for, or are associated with standard 4-20 mA devices and some types of smart field devices such as HART devices, or may be stored in and implemented by the field devices themselves, which can be the case with Fieldbus devices. The controller 11 may include one or more control routines 38 that may implement one or more control loops. Each control loop is typically referred to as a control module, and may be performed by executing one or more of the function blocks.

[0076] The wired field devices 15-22 may be any types of devices, such as sensors, valves, transmitters, positioners, etc., while the I/O cards 26 and 28 may be any types of I/O devices conforming to any desired communication or controller protocols. As illustrated in FIG. 2, the field devices 15-18 are standard 4-20 mA devices or HART devices that communicate over analog lines or combined analog and digital lines to the I/O card 26, while the field devices 19-22 are smart devices, such as FOUNDATION® Fieldbus field devices, that communicate over a digital bus to the I/O card 28 using a Fieldbus communications protocol. In other implementations, though, at least some of the wired field devices 15-22 and/or at least some of the I/O cards 26, 28 may communicate with the controller 11 using the big data network backbone 105. Further, at least some of the wired field devices 15-22 and/or at least some of the I/O cards 26, 28 may be nodes of the process control system big data network 100.

[0077] In the system illustrated in FIG. 2, the wireless field devices 40-46 communicate in a wireless network 70 using a wireless protocol, such as the WirelessHART protocol. Such wireless field devices 40-46 may directly communicate with one or more other nodes 108 of the process control big data network 100 that are also configured to communicate wirelessly (using the wireless protocol, for example). To communicate with one or more other nodes 108 that are not configured to communicate wirelessly, the wireless field devices 40-46 may utilize a wireless gateway 35 connected to the backbone 105 or to another process control communication network. In some cases, at least some of the wireless field devices 40-46 may be nodes of the process control system big data network 100.

[0078] The wireless gateway 35 is an example of a provider device 110 that may provide access to various wireless devices 40-58 of a wireless communication network 70. In particular, the wireless gateway 35 provides communicative coupling between the wireless devices 40-58, the wired devices 11-28, and/or other nodes 108 of the process control big data network 100 (including the controller 11 of FIG. 2). For example, the wireless gateway 35 may provide communicative coupling by using the big data network backbone 105 and/or by using one or more other communications networks of the process plant 10.

[0079] The wireless gateway 35 provides communicative coupling, in some cases, by the routing, buffering, and timing services to lower layers of the wired and wireless protocol stacks (e.g., address conversion, routing, packet segmentation, prioritization, etc.) while tunneling a shared layer or layers of the wired and wireless protocol stacks. In other cases, the wireless gateway 35 may translate commands between wired and wireless protocols that do not share any protocol layers. In addition to protocol and command conversion, the wireless gateway 35 may provide synchronized clocking used by time slots and superframes (sets of communication time slots spaced equally in time) of a scheduling scheme associated with the wireless protocol implemented in the wireless network 70. Furthermore, the wireless gateway 35 may provide network management and administrative functions for the wireless network 70, such as

14 06 21

resource management, performance adjustments, network fault mitigation, monitoring traffic, security, and the like. The wireless gateway 35 may be a node 110 of the process control system big data network 100.

[0080] Similar to the wired field devices 15-22, the wireless field devices 40-46 of the wireless network 70 may perform physical control functions within the process plant 10, e.g., opening or closing valves or taking measurements of process parameters. The wireless field devices 40-46, however, are configured to communicate using the wireless protocol of the network 70. As such, the wireless field devices 40-46, the wireless gateway 35, and other wireless nodes 52-58 of the wireless network 70 are producers and consumers of wireless communication packets.

[0081] In some scenarios, the wireless network 70 may include non-wireless devices. For example, a field device 48 of FIG. 2 may be a legacy 4-20 mA device and a field device 50 may be a traditional wired HART device. To communicate within the network 70, the field devices 48 and 50 may be connected to the wireless communication network 70 via a wireless adaptor (WA) 52a or 52b. Additionally, the wireless adaptors 52a, 52b may support other communication protocols such as Foundation® Fieldbus, PROFIBUS, DeviceNet, etc. Furthermore, the wireless network 70 may include one or more network access points 55a, 55b, which may be separate physical devices in wired communication with the wireless gateway 35 or may be provided with the wireless gateway 35 as an integral device. The wireless network 70 may also include one or more routers 58 to forward packets from one wireless device to another wireless device within the wireless communication network 70. The wireless devices 32-46 and 52-58 may communicate with each other and with the wireless gateway 35 over wireless links 60 of the wireless communication network 70.

14 06 21

[0082] Accordingly, FIG. 2 includes several examples of provider devices 110 which primarily serve to provide network routing functionality and administration to various networks of the process control system. For example, the wireless gateway 35, the access points 55a, 55b, and the router 58 include functionality to route wireless packets in the wireless communication network 70. The wireless gateway 35 performs traffic management and administrative functions for the wireless network 70, as well as routes traffic to and from wired networks that are in communicative connection with the wireless network 70. The wireless network 70 may utilize a wireless process control protocol that specifically supports process control messages and functions, such as WirelessHART.

[0083] The provider nodes 110 of the process control big data network 100, though, may also include other nodes that communicate using other wireless protocols. For example, the provider nodes 110 may include one or more wireless access points 72 that utilize other wireless protocols, such as WiFi or other IEEE 802.11 compliant wireless local area network protocols, mobile communication protocols such as WiMAX (Worldwide Interoperability for Microwave Access), LTE (Long Term Evolution) or other ITU-R (International Telecommunication Union Radiocommunication Sector) compatible protocols, short-wavelength radio communications such as near field communications (NFC) and Bluetooth, or other wireless communication protocols. Typically, such wireless access points 72 allow handheld or other portable computing devices (e.g., user interface devices 112) to communicate over a respective wireless network that is different from the wireless network 70 and that supports a different wireless protocol than the wireless network 70. In some scenarios, in addition to portable computing devices, one or more process control devices (e.g., controller 11, field devices 15-22, or wireless devices 35, 40-58) may also communicate using the wireless protocol supported by the access points 72.

[0084] Additionally or alternatively, the provider nodes 110 may include one or more gateways 75, 78 to systems that are external to the immediate process control system 10. Typically, such systems are customers or suppliers of information generated or operated on by the process control system 10. For example, a plant gateway node 75 may communicatively connect the immediate process plant 10 (having its own respective process control big data network backbone 105) with another process plant having its own respective process control big data network backbone. If desired, a single process control big data network backbone 105 may service multiple process plants or process control environments.

[0085] In another example, a plant gateway node 75 may communicatively connect the immediate process plant 10 to a legacy or prior art process plant that does not include a process control big data network 100 or backbone 105. In this example, the plant gateway node 75 may convert or translate messages between a protocol utilized by the process control big data backbone 105 of the plant 10 and a different protocol utilized by the legacy system (e.g., Ethernet, Profibus, Fieldbus, DeviceNet, etc.).

[0086] The provider nodes 110 may include one or more external system gateway nodes 78 to communicatively connect the process control big data network 100 with the network of an external public or private system, such as a laboratory system (e.g., Laboratory Information Management System or LIMS), an operator rounds database, a materials handling system, a maintenance management system, a product inventory control system, a production scheduling system, a weather data system, a shipping and handling system, a packaging system, the Internet, another provider's process control system, or other external systems.

14 06 21

[0087] FIG. 2 illustrates a single controller 11 with a finite number of field devices 15-22 and 40-46. If desired, the provider nodes 110 of the process control big data network 110 may include any number of controllers 11, and any of the controllers 11 may communicate with any number of wired or wireless field devices 15-22, 40-46 to control a process in the plan 10. Furthermore, the process plant 10 may also include any number of wireless gateways 35, routers 58, network access points 55, wireless process control communication networks 70, wireless access points 72, and/or gateways 75, 78.

[0088] As previously discussed, one or more of the provider nodes 110 may include a respective multi-core processor P_{MCX}, a respective high density memory storage M_X, or both a respective multi-core processor P_{MCX} and a respective high density memory storage M_X (denoted in FIG. 2 by the icon BD). Each provider node 100 may utilize its memory storage M_X (and optionally its flash memory) to collect and cache data. Each of the nodes 110 may cause its cached data to be transmitted to the process control system big data appliance 102. For example, a node 110 may cause at least a portion of the data in its cache to be periodically transmitted to the big data appliance 102. Alternatively or additionally, the node 110 may cause at least a portion of the data in its cache to be streamed to the big data appliance 102. For example, the process control system big data appliance 102 may be a subscriber to a streaming service that delivers the cached or collected data from the node 110, and the provider node 110 may host the streaming service.

[0089] For nodes 110 that have a direct connection with the backbone 105 (e.g., the controller 11, the plant gateway 75, the wireless gateway 35), the respective cached or collected data may be transmitted directly from the node 110 to the process control big data appliance 102 via the backbone 105. For at least some of the nodes 110, though, the collection and/or caching may be leveled or layered, so that cached or collected data at a node that is further downstream (e.g., is further away) from the process control big data

appliance 102 is intermediately cached at a node that is further upstream (e.g., is closer to the big data appliance 102).

[0090] To illustrate layered or leveled data caching, an example scenario is provided. In this example scenario, referring to FIG. 2, a field device 22 caches process control data that it generates or receives, and causes the contents of its cache to be delivered to an "upstream" device included in the communication path between the field device 22 and the process control big data appliance 102, such as the I/O device 28 or the controller 11. For example, the field device 22 may stream the contents of its cache to the I/O device 28, or the field device 22 may periodically transmit the contents of its cache to the I/O device 28. The I/O device 28 caches the information received from the field device 22 in its memory M_5 (and, in some cases, may also cache data received from other downstream field devices 19-21 in its memory M_5) along with other data that the I/O device 28 directly generates, receives and observes. The data that the I/O device 28 collects and caches (including the contents of the cache of the field device 22) may then be periodically transmitted and/or streamed to the upstream controller 11. Similarly, at the level of the controller 11, the controller 11 caches information received from downstream devices (e.g., the I/O cards 26, 28, and/or any of the field devices 15-22) in its memory M_6, and aggregates, in its memory M_6, the downstream data with data that the controller 11 itself directly generates, receives and observes. The controller 11 may then periodically deliver and/or stream the aggregated collected or cached data to the process control big data appliance 102.

[0091] In a second example scenario of layered or leveled caching, the controller 11 controls a process using wired field devices (e.g., one or more of the devices 15-22) and at least one wireless field device (e.g., wireless field device 44). In a first implementation of this second example scenario, the cached or collected data at the wireless device 44 is delivered and/or streamed directly to the controller 11 from the wireless device 44 (e.g., via the big data network 105), and is stored at the controller cache M_6 along with data from other devices or nodes that are downstream from the controller 11. The controller 11 may periodically deliver or stream the data stored in its cache M_6 to the process control big data appliance 102.

14 06 21

[0092] In another implementation of this second example scenario, the cached or collected data at the wireless device 44 may be ultimately delivered to the process control big data appliance 102 via an alternate leveled or layered path, e.g., via the device 42a, the router 52a, the access point 55a, and the wireless gateway 35. In this case, at least some of the nodes 41a, 52a, 55a or 35 of the alternate path may cache data from downstream nodes and may periodically deliver or stream its cached data to a node that is further upstream.

[0093] Accordingly, the process control system big data network 100 may cache different types of data at different nodes using different layering or leveling arrangements. For example, data corresponding to controlling a process may be cached and delivered in a layered manner using provider devices 110 whose primary function is control (e.g., field devices, I/O cards, controllers), whereas data corresponding to network traffic measurement may be cached and delivered in a layered manner using provider devices 110 whose primary function is traffic management (e.g., routers, access points, and gateways). In some cases, data may be delivered via provider nodes or devices 110 whose primary function (and, in some scenarios, sole function) is to collect and cache data from downstream devices (referred to herein as "historian nodes"). For example, a leveled system of historian nodes or computing devices may be located throughout the network 100, and each node 110 may periodically deliver or stream cached data to a historian node of a similar level, e.g., using the backbone 105. Downstream historian nodes may deliver or stream cached data to upstream

historian nodes, and ultimately the historian nodes that are immediately downstream of the process control big data appliance 102 may deliver or stream respective cached data for storage at the process control big data appliance 102.

[0094] If desired, nodes 110 that communicate with each other may perform layered caching using the process control system big data network backbone 105, and at least some of the nodes 110 may communicate cached data to other nodes 110 at a different level using another communication network and/or other protocol, such as HART, WirelessHART, Fieldbus, DeviceNet, WiFi, Ethernet, or other protocols.

[0095] Of course, while leveled or layered caching has been discussed with respect to provider nodes 110, the concepts and techniques may apply equally to user interface nodes 112 and/or to other types of nodes 115 of the process control system big data network 100. Still further, a subset of the nodes 108 may perform leveled or layered caching, while another subset of the nodes 108 may cause their cached/collected data to be directly delivered to the process control big data appliance 102 without being cached or temporarily stored at an intermediate node. If desired, historian nodes may cache data from multiple different types of nodes, e.g., from a provider node 110 and from a user interface node 112.

[0096] Process Control System Big Data Network Backbone

[0097] Returning to FIG. 1, the process control system big data network backbone 105 may include a plurality of networked computing devices or switches that are configured to route packets to/from various nodes 108 of the process control system big data network 100 and to/from the process control big data appliance 102 (which is itself a node of the process control system big data network 100). The plurality of networked computing devices of the backbone 105 may be interconnected by any number of wireless and/or wired links. The process control system big data network backbone 105 may additionally include one or more firewall devices.

14 06 21

[0098] The big data network backbone 105 may support one or more suitable routing protocols, e.g., protocols included in the Internet Protocol (IP) suite (e.g., UDP (User Datagram Protocol), TCP (Transmission Control Protocol), Ethernet, etc.), or other suitable routing protocols. At least some of the nodes 108 may utilize a streaming protocol such as the Stream Control Transmission Protocol (SCTP) to stream cached data from the nodes to the process control big data appliance 102 via the network backbone 105. Typically, each node 108 included in the process data big data network 100 may support at least an application layer (and, for some nodes, additional layers) of the routing protocol(s) supported by the backbone 105. Each node 108 may be uniquely identified within the process control system big data network 100, e.g., by a unique network address.

[0099] At least a portion of the process control system big data network 100 may be an ad-hoc network. As such, at least some of the nodes 108 may connect to the network backbone 105 (or to another node of the network 100) in an ad-hoc manner. Typically, each node that requests to join the network 100 must be authenticated; however authentication is discussed in more detail in later sections.

[00100] Process Control System Big Data Appliance

[00101] Continuing with FIG. 1, in the example big data process control network 100, the process control big data apparatus or appliance 102 is centralized within the network 100, and is configured to receive data

(e.g., via streaming and/or via some other protocol) from the nodes 108 of the network 100 and to store the received data. As such, the process control big data apparatus or appliance 102 may include a data storage area 120 for historizing or storing the data that is received from the nodes 108, a plurality of appliance data receivers 122, and a plurality of appliance request servicers 125. Each of these components 120, 122, 125 of the process control big data appliance 102 is described in more detail below.

[00102] The process control system big data storage area 120 may comprise multiple physical data drives or storage entities, such as RAID (Redundant Array of Independent Disks) storage, cloud storage, or any other suitable data storage technology that is suitable for data bank or data center storage. However, the data storage area 120 has the appearance of a single or unitary logical data storage area or entity to the nodes 108 of the network 100. As such, the data storage 120 may be viewed as a centralized big data storage area 120 for the process control big data network 100 or for the process plant 10. In some cases, a single logical centralized data storage area 120 may service multiple process plants (e.g., the process plant 10 and another process plant). For example, a centralized data storage area 120 may service several refineries of an energy company. If desired, the centralized data storage area 120 may be directly connected to the backbone 105, via for example at least one high-bandwidth communication link. Additionally, the centralized data storage area 120 may include an integral firewall.

[00103] The structure of the unitary, logical data storage area 120 may support the storage of all process control system related data. For example, each entry, data point, or observation of the data storage entity may include an indication of the identity of the data (e.g., source, device, tag, location, etc.), a content of the data (e.g., measurement, value, etc.), and a timestamp indicating a time at which the data was collected, generated, received or observed. As such, these entries, data points, or observations are referred to herein as "time-series data." The data may be stored in the data storage area 120 using a common format including a schema that supports scalable storage, streamed data, and low-latency queries, for example.

14 06 21

[00104] If desired, the schema may include storing multiple observations in each row, and using a rowkey with a custom hash to filter the data in the row. The hash may be based on the timestamp and a tag. For example, the hash may be a rounded value of the timestamp, and the tag may correspond to an event or an entity of or related to the process control system. Additionally, the data storage area 120 may also store metadata corresponding to each row or to a group of rows, either integrally with the time-series data or separately from the time-series data. For example, the metadata may be stored in a schema-less manner separately from the time-series data.

[00105] The schema used for storing data at the appliance data storage 120 may also be utilized for storing data in the cache M_X of at least one of the nodes 108. Accordingly, the schema may be maintained when data is transmitted from the local storage areas M_X of the nodes 108 across the backbone 105 to the process control system big data appliance data storage 120.

[00106] In addition to the data storage 120, the process control system big data appliance 102 may further include one or more appliance data receivers 122, each of which is configured to receive data packets from the backbone 105, process the data packets to retrieve the substantive data and timestamp carried therein, and store the substantive data and timestamp in the data storage area 120. The appliance data receivers 122 may reside on a plurality of computing devices or switches, for example. Multiple appliance data

receivers 122 (and/or multiple instances of at least one data receiver 122) may operate in parallel on multiple data packets.

[00107] If the received data packets include the schema utilized by the process control big data appliance data storage area 120, the appliance data receivers 122 may populate additional entries or observations of the data storage area 120 with the schematic information (and, may optionally store corresponding metadata, if desired). In contrast, if the received data packets do not include the schema utilized by the process control big data appliance data storage area 120, the appliance data receivers 122 may decode the packets and populate time-series data observations or data points of the process control big data appliance data storage area 120 (and, optionally corresponding metadata) accordingly.

[00108] Additionally, the process control system big data appliance 102 may include one or more appliance request servicers 125, each of which is configured to access time-series data and/or metadata stored in the process control system big data appliance storage 120, e.g., per the request of a requesting entity or application. The appliance request servicers 125 may reside on a plurality of computing devices or switches, for example. At least some of the appliance request servicers 125 and the appliance data receivers 122 may reside on the same computing device or devices (e.g., on an integral device), or are included in an integral application.

[00109] Multiple appliance request servicers 125 (and/or multiple instances of at least one appliance request servicer 125) may operate in parallel on multiple requests from multiple requesting entities or applications. As such, a single appliance request servicer 125 may service multiple requests, such as multiple requests from a single entity or application, or multiple requests from different instances of an application.

14 06 21

[00110] FIGS. 3 and 4 are example block diagrams illustrating further techniques that the appliance data receivers 122 and the appliance request servicers 125 of the process control system big data appliance 102 may implement.

[00111] FIG. 3 is an example block diagram illustrating the use of the appliance data receivers 122 to transfer data (e.g., streamed data) from the nodes 108 of the process control big data network 100 to the big data appliance 102 for storage and historization. FIG. 3 illustrates four example nodes 108 of FIG. 1, i.e., the controller 11, a user interface device 12, the wireless gateway 35, and a gateway to a third party machine or network 78. However, the techniques and concepts discussed with respect to FIG. 3 may be applied to any type and any number of the nodes 108. Additionally, although FIG. 3 illustrates only three appliance data receivers 122a, 122b and 122c, the techniques and concepts corresponding to FIG. 3 may be applied to any type and any number of appliance data receivers 122.

[00112] As illustrated in FIG. 3, each of the nodes 11, 12, 35 and 78 includes a respective scanner S_{11}, S_{12}, S_{35}, S_{78} to capture data that is generated, received or otherwise observed by the node 11, 12, 35 and 78. A respective processor P_{MCX} of the respective node 11, 12, 35, 78 may execute the functionality of each scanner S_{11}, S_{12}, S_{35}, S_{78}. The scanner S_{11}, S_{12}, S_{35}, S_{78} may cause the captured data and a corresponding timestamp to be temporarily stored or cached in a respective local memory M_{11}, M_{12}, M_{35}, M_{78}, for example, in a manner such as previously described. As such, the captured data includes time-series data or real-time data. Each of the memories M_{11}, M_{12}, M_{35} and M_{78} may store and cache the captured data using the schema utilized by the process control big data storage area 120.

[00113] Each node 11, 12, 35 and 78 may transmit at least some of the cached data to one or more appliance data receivers 122a-122c (e.g., using the network backbone 105). For example, at least one node 11, 12, 35, 78 may push at least some of the data from its respective memory M_X when the cache is filled to a particular threshold. The threshold of the cache may be adjustable, and at least one node 11, 12, 35, 78 may push at least some of data from its respective memory M_X when a resource (e.g., a bandwidth of the network 105, the processor P_{MCX}, or some other resource) is sufficiently available. An availability threshold of a particular resource may be adjustable.

[00114] Moreover, at least one node 11, 12, 35, 78 may push at least some of the data stored in the memories M_X at periodic intervals. The periodicity of a particular time interval at which data is pushed may be based on a type of the data, the type of pushing node, the location of the pushing node, and/or other criteria. The periodicity of a particular time interval may be adjustable, and at least one node 11, 12, 35, 78 may provide data in response to a request (e.g., from the process control big data appliance 102).

[00115] At least one of the nodes 11, 12, 35, 78 may stream at least some of the data in real-time as the data is generated, received or otherwise observed by each node 11, 12, 35, 78 (e.g., the node may not temporarily store or cache the data, or may store the data for only as long as it takes the node to process the data for streaming). For example, at least one of the nodes 11, 12, 35, 78 may stream at least some of the data to the one or more appliance data receivers 122 using a streaming protocol. Hence, a node 11, 12, 35, 78 may host a streaming service, and at least one of the data receivers 122 and/or the data storage area 120 may subscribe to the streaming service.

14 06 21

[00116] Accordingly, transmitted data may be received by one or more appliance data receivers 122a-122c, e.g., via the network backbone 105. A particular appliance data receiver 122 may be designated to receive data from one or more particular nodes, or a particular appliance data receiver 122 may be designated to receive data from only one or more particular types of devices (e.g., controllers, routers, or user interface devices). Further, a particular appliance data receiver 122 may be designated to receive only one or more particular types of data (e.g., network management data only or security-related data only).

[00117] The appliance data receivers 122a-122c may cause the big data appliance storage area 120 to store or historize the data. For example, the data storage area 120 may store the data received by each of the appliance data receivers 122a-122c using the process control big data schema. As illustrated in FIG. 3, the time series data 120a is stored separately from corresponding metadata 120b, although in some implementations, at least some of the metadata 120b may be integrally stored with the time series data 120a.

[00118] The data storage area 120 may integrate data that is received via the plurality of appliance data receivers 122a-122c so that data from multiple sources may be combined (e.g., into a same group of rows of the data storage area 120). Data that is received via the plurality of appliance data receivers 122a-122c may be cleaned to remove noise and inconsistent data. An appliance data receiver 122 may perform data cleaning and/or data integration on at least some of the received data before the received data is stored, and/or the process control system big data appliance 102 may clean some or all of the received data after the received data has been stored in the storage area 102. A device or node 110, 112, 115 may cause additional data related to the data contents to be transmitted, and the appliance data receiver 122 and/or the big data appliance storage area 120 may utilize this additional data to perform data cleaning. A node 110,

112, 115 may clean (at least partially) at least some data prior to the node 110, 112, 115 causing the data to be transmitted to the big data appliance storage area 120 for storage.

[00119] FIG. 4 depicts an example block diagram illustrating the use of appliance request servicers 125 to access the historized data stored at the data storage area 120 of the big data appliance 102. FIG. 4 includes a set of appliance request servicers or services 125a-125e that are each configured to access time-series data 120a and/or metadata 120b per the request of a requesting entity or application, such as a data requester 130a-130c or a data analysis engine 132a-132b. While FIG. 4 illustrates five appliance request servicers 125a-125e, three data requesters 130a-130c, and two data analysis engines 132a, 132b, the techniques and concepts discussed herein with respect to FIG. 4 may be applied to any number and any types of appliance request servicers 125, data requesters 130, and/or data analysis engines 132.

[00120] At least some of the appliance request servicers 125 may each provide a particular service or application that requires access to at least some of the data stored in the process control big data storage area 120. For example, the appliance request servicer 125a may be a data analysis support service, and the appliance request servicer 125b may be a data trend support service. Other examples of services 125 that may be provided by the process control system big data appliance 102 may include a configuration application service 125c, a diagnostic application service 125d, and an advanced control application service 125e. An advanced control application service 125e may include, for example, model predictive control, batch data analytics, continuous data analytics or other applications that require historized data for model building and other purposes. The process control system big data appliance 102 may include other request servicers 125 to support other services or applications, e.g., a communication service, an administration service, an equipment management service, a planning service, and other services.

14 06 21

[00121] A data requester 130 may be an application that requests access to data that is stored in the process control system big data appliance storage area 120. Based on a request of the data requester 130, the corresponding data may be retrieved from the process control big data storage area 120, and may be transformed and/or consolidated into data forms that are usable by the requester 130. One or more appliance request servicers 125 may perform data retrieval and/or data transformation on at least some of the requested data. The big data appliance 102 further supports casual data access, such as via a user requesting data access casually and repeatedly with variances. In particular, the big data appliance 102 may support privileged APIs that enable more granular and versatile access to the process control big data storage area 120.

[00122] At least some of the data requesters 130 and/or at least some of the request servicers 125 may be web services or web applications that are hosted by the process control system big data appliance 102 and that are accessible by nodes of the process control system big data network 100 (e.g., user interface devices 112 or provider devices 110). Accordingly, at least some of the devices or nodes 108 may include a respective web server to support a web browser, web client interface, or plug-in corresponding to a data requester 130 or to a request servicer 125. For user interface devices 112 in particular, a data requester 130 or a request servicer 125 may pull displays and stored data through a User Interface (UI) service layer 135. The UI service layer 135 includes a data visualization service 136 that facilitates the display of various process control data. In particular, the data visualization service 136 may represent various portions of process control data in pictures, charts, maps, reports, presentations, and/or the like. Accordingly, a user accessing any of the data visualization channels may be able to quickly ascertain certain data, trends,

relationships, or conclusions associated with the process control data. The data visualization service 136 supports dynamic updating whereby the data visualization service 136 may update corresponding charts or visualizations based on user input, added or removed data, and/or other factors.

[00123] A data analysis engine 132 may be an application that performs a computational analysis on at least some of the time-series data points stored in the appliance storage area 120 to generate knowledge or observations. As such, a data analysis engine 132 may generate a new set of data points or observations. The new knowledge, new observations, or new data points may provide *a posteriori* analysis of aspects of the process plant 10 (e.g., diagnostics or trouble shooting), and/or may provide *a priori* predictions (e.g., prognostics) corresponding to the process plant 10. In one case, a data analysis engine 132 may perform data mining on a selected subset of the stored data 120, and may perform pattern evaluation on the mined data to generate the new knowledge or new set of data points or observations. Of course, multiple data analysis engines 132 or instances thereof may cooperate to generate the new knowledge or new set of data points.

[00124] The new knowledge or set of data points may be stored in (e.g., added to) the appliance storage area 120, for example, and may additionally or alternatively be presented at one or more user interface devices 112. The new knowledge may also be incorporated into one or more control strategies operating in the process plant 10, if desired. A particular data analysis engine 132 may be executed when indicated by a user (e.g., via a user interface device 112), and/or the particular data analysis engine 132 may be executed automatically by the process control system big data appliance 102.

14 06 21

[00125] Generally, the data analysis engines 132 of the process control system big data appliance 102 may operate on the stored data to determine time-based relationships between various entities and providers within and external to the process plant 10, and may utilize the determined time-based relationship to control one or more processes of the plant 10 accordingly. As such, the process control system big data appliance 102 allows for one or more processes to be coordinated with other processes and/or to be adjusted over time in response to changing conditions and factors. The process control system big data appliance 102 may automatically determine and execute the coordination and/or adjustments as conditions and events occur, thus greatly increasing efficiencies and optimizing productivity over known prior art control systems.

[00126] Examples of possible scenarios in which the knowledge discovery techniques of data analysis engines 132 abound. In one example scenario, a certain combination of events leads to poor product quality when the product is eventually generated at a later time (e.g., several hours after the occurrence of the combination of events). The operator is usually ignorant of the relationship between the occurrence of the events and the product quality. Rather than detecting and determining the poor product quality several hours hence and trouble-shooting to determine the root causes of the poor product quality (as is currently done in known process control systems), the process control system big data appliance 102 (and, in particular, one or more of the data analysis engines 132 therein) may automatically detect the combination of events at or shortly after their occurrence, e.g., when the data corresponding to the events' occurrences is transmitted to the appliance 102. The data analysis engines 132 may predict the poor product quality based on the occurrence of these events, may alert an operator to the prediction, and/or may automatically adjust or change one or more parameters or processes in real-time to mitigate the effects of the combination of events. For example, a data analysis engine 132 may determine a revised set point or revised parameter values and cause the revised values to be used by provider devices 110 of the process plant 10. In this

manner, the process control system big data appliance 102 allows problems to be discovered and potentially mitigated much more quickly and efficiently as compared to currently known process control systems.

[00127] In another example scenario, at least some of the data analysis engines 132 may be utilized to detect changes in product operation. For instance, the data analysis engines 132 may detect changes in certain communication rates, and/or from changes or patterns of parameter values received from a sensor or from multiple sensors over time which may indicate that system dynamics may be changing. In yet another example scenario, the data analysis engines 132 may be utilized to diagnose and determine that a particular batch of valves or other supplier equipment are faulty based on the behavior of processes and the occurrences of alarms related to the particular batch across the plant 10 and across time.

[00128] In another example scenario, at least some of the data analysis engines 132 may predict product capabilities, such as vaccine potency. In yet another example scenario, the data analysis engines 132 may monitor and detect potential security issues associated with the process plant 10, such as increases in log-in patterns, retries, and their respective locations. In still another example scenario, the data analysis engines 132 may analyze data aggregated or stored across the process plant 10 and one or more other process plants. In this manner, the process control system big data appliance 102 allows a company that owns or operates multiple process plants to glean diagnostic and/or prognostic information on a region, an industry, or a company-wide basis.

[00129] Big Data Schema for Process Control Data

[00130] The big data appliance 102 is configured to use non-relational database mechanisms of a big data schema to store process control data. The non-relational database mechanisms enable design simplicity, horizontal scaling, and finer control over data availability. Generally, the non-relational structure of the big data schema leverages one or more tables to store process control data received from various control system components or modules. The structure of the big data schema enables efficient storage as a result of the tables only storing actual measurements or values (i.e., the tables do not have empty cells), thereby reducing the amount of total storage required by the tables. Further, the organization of the tables enables the use of multiple types of queries to efficiently locate and access stored data.

[00131] Generally, each table includes one or more rowkeys, column families, and column qualifiers. Each rowkey serves as a primary key for the corresponding table. The big data appliance 102 examines one or more fields of a received process variable to determine the rowkey to which the process variable should be associated. A column family groups one or more related columns that specify how the process variable should be associated with the rowkey. In particular, each column of a column family specifies one or more column qualifiers corresponding to fields or attributes of the received process variables. An administrator or user may specify the various column families and the column qualifiers thereof. The column qualifiers have one or more values that can result in multiple entries for the same rowkey, therefore resulting in a three-dimensional storage scheme. In some cases, the column qualifiers can correspond to fields or attributes already included in the received process control data. In other cases, the big data appliance 102 can determine or identify the column qualifiers upon receipt of the process control data.

[00132] The rowkeys and the column qualifiers can include one or more fields or attributes, or combinations thereof, of the process control data, such as one or more of a timestamp (or a portion thereof), an identification of a process variable, the measurement or value of the process control data, a type of data

14 06 21

(e.g., Boolean, integer, etc.), a status of the process variable (e.g., "good," "bad," "absent," etc.), and/or others. For example, each rowkey of a table can be a concatenation of an identification of a process variable and a portion of the timestamp corresponding to when the process variable was recorded, and the column qualifiers of the table can be a concatenation of the type of the process variable, the status of the process variable, and an additional portion of the timestamp, whereby the table stores the measurement value of the process variable in the appropriate data field. For further example, each rowkey of a table can be a concatenation of the type of the process variable and the measurement value of the process variable, and the column qualifiers can be a concatenation of the identification of the process variable and the timestamp, whereby the table stores the status of the process variable in the appropriate data field.

[00133] Generally, the big data appliance 102 collects many types of data (e.g., continuous, batch, event, operator-entered values, etc.) from a process control plant and from other sources such as lab systems and material handling systems. For example, the big data appliance 102 collects data such as process variable values, setpoints, discrete inputs and outputs, process alarms, maintenance alarms, operator actions, batch actions, end of batch data, insight models and statistics, and/or the like. The big data appliance 102 automatically buffers the collected data in local memory or storage without requiring any user input or configuration, and periodically transfers the data to a real-time database. Because the big data appliance 102 collects data at the rate at which the associated module of the process control plant is executed, the big data appliance 102 enables a complete history of the process control plant operation to be available to support various analyses.

14 06 21

[00134] The big data appliance 102 further leverages a time series database server (TSDS) of the data storage area 120 to store, index, and serve process control data and other related data collected from various control system components (e.g., control strategies, control system equipment, devices, lab systems, applications, etc.) at a large scale and to enable effective retrieval of the data. The TSDS is able to serve up data for traditional applications such as operational historians, and to collect and serve up infrastructure data related to the process control devices and equipment.

[00135] The time-series data may be thought of as a collection of data points or tuples, whereby each data point can have a timestamp and a measurement. The TSDS may collect the measurements at regular or irregular intervals, for example at the execution rate of the associated control module. For instance, the TSDS may collect a process variable and an associated status for all analog input points. In some cases, the data points can include metadata indicating the measurement, such as the fully-qualified tag generating the time series, the range on the data, and other data. By appending a timestamp to a value or to a measurement and its status, patterns and differences between and among measurement values over time can be better ascertained. For example, if a current temperature at a specific location is measured every hour, future temperatures can be more easily predicted based on one or more of the measured temperatures. Further, by maintaining the timestamp, location, and measurement information as part of a control hierarchy, the TSDS may store these relationships in the database as metadata and update the relationships as the hierarchy is updated.

[00136] As discussed herein, the big data appliance 102 implements the big data storage schema using one or more tables. FIG. 5A depicts a structure of an example table 500 configured to store process control data. The big data appliance 102 stores the example table 500 in the process control system big data storage 120. The example table 500 includes a first rowkey 505 ("ROWKEY 1") and a second rowkey 510

("ROWKEY 2"). Each of the first rowkey 505 and the second rowkey 510 has an associated first column family 506 ("CF1") and a second column family 507 ("CF2"). The first column family 506 has three (3) associated column qualifiers ("CQ1," "CQ2," and "CQ3") and the second column family 507 has two (2) associated column qualifiers ("CQ4" and "CQ5"). Because each of the rowkeys 505, 510 has multiple column qualifiers with multiple potential values, each rowkey 505, 510 may contiguously store multiple entries, as illustrated in FIG. 5A.

[00137] The big data appliance 102 examines received process control data to determine how to store the data. In some cases, the big data appliance 102 examines the process control data to identify attributes corresponding to the rowkey (or a portion thereof), or to one or more column qualifiers. For example, the process control data can include an identification of the process variable and a status of the process variable. In other cases, the big data appliance 102 appends data to the received process control data, wherein the appended data corresponds to the rowkey (or a portion thereof), or to one or more column qualifiers. For example, the big data appliance 102 may generate a timestamp corresponding to when the big data appliance 102 received the process control data. Accordingly, the big data appliance 102 may build the rowkeys and/or column qualifiers using the identified or generated attributes of the process control data. Further, the big data appliance 102 may store the process control data (or measurements or values thereof) according to the built rowkeys and column qualifiers.

[00138] Generally, the big data appliance 102 does not allocate memory segments of the table 500 prior to storing data in the table 500. Instead, the big data appliance 102 is configured to store data associated with the process variable (e.g., the measurement or value) in the table 500 according to the corresponding rowkey, column family, and column qualifiers. Further, the big data appliance 102 stores the data as the process control data is received and processed. For example, the big data appliance 102 stores data 509 in a memory segment corresponding to the first rowkey 505, the first column family 506, and CQ1 of a certain value; and the big data appliance 102 stores data 511 in a memory segment corresponding to the second rowkey 501, the second column family 507, and CQ5 of a certain value. The remaining fields of the table are null or otherwise unallocated, thus conserving memory space in the big data schema.

14 06 21

[00139] In an example implementation, the big data appliance 102 orders the rowkeys according to a timestamp of the corresponding process control data and process variables thereof. The timestamp may correspond to when a device recorded the process control data, when the device transmitted the process control data, when the big data appliance 102 received the process control data, or other times. Further, the big data appliance 102 may round the timestamp down or up by a predetermined degree. For example, timestamp may be rounded down or up to the nearest minute, hour, day, or the like. As a result, the big data appliance 102 may store multiple columns having timestamps that are included in the rounded timestamp of the corresponding rowkey. By storing multiple columns per rowkey, searching the process control system big data storage 120 is more efficient and effective. In particular, this structure enables more data to be disqualified in a single exclusion and the overall number of rows that are tracked by rowkey to be reduced. Further, by using the rounded time as a part of the hash, an administrator may partition the big data schema more effectively.

[00140] FIG. 5B illustrates an example rowkey and column qualifier for an example table, as well as example process control data related thereto. As illustrated in FIG. 5B, an example rowkey structure 510 includes an identification 511 and a first timestamp portion 512. The identification 511 is a certain amount of

bytes (e.g., 3 bytes) and uniquely identifies the received process variable. In some cases, the big data appliance 102 may generate the identification 511 and store the identification 511 in a look-up table. The first timestamp portion 512 is a certain amount of bytes (e.g., 4 bytes) and corresponds to a timestamp that is rounded by a predetermined degree. For example, the first timestamp portion 512 may be a "high order" timestamp resulting from rounding a timestamp down to the nearest hour. Because the big data appliance 102 may receive a particular process variable having the same identification at multiple times, the big data appliance 102 may store a grouping of rows for the particular identification and order the rows according to the timestamps. Accordingly, a single row can store a "bucket" of measurements (e.g., if the first timestamp portion 812 corresponds to an hour, the associated rowkey 810 can store measurements collected within that hour).

[00141] FIG. 5B further includes an example column structure 515 having a column family 516 and three column qualifiers: a type 517, a status 518, and a second timestamp portion 519. As discussed above, the column family 516 identifies or categorizes the column qualifiers. Further, the column qualifiers (i.e., the type 517, the status 518, and the second timestamp portion 519) are attributes associated with the corresponding process control data. In particular, the type 517 identifies the data type of the process variable (e.g., integer, string, number, Boolean, etc.), the status 518 indicates the status of the process variable (e.g., good, bad, absent, etc.), and the second timestamp portion 519 is a remainder of the timestamp resulting from the rounded first timestamp portion 512. The column family 516, the type 517, the status 518, and the second timestamp portion 519 may be of various sizes. For example, the column family 516, the type 517, and the status 518 can each be 1 byte, and the second timestamp portion 519 can be 4 bytes.

14 06 21

[00142] FIG. 5B also illustrates an additional example rowkey 520 and an example column 525 populated with example data. As illustrated in FIG. 5B, an identification 521 of the rowkey 520 is "001," a column family 526 of the column 825 is "Time," a type 527 of the column 525 is "INT," and a status 528 of the column 525 is "GOOD." The rowkey 520 and the column 525 also store timestamp data corresponding to a timestamp of July 1, 2013 at 10:02:03 GMT (which produces a corresponding UNIX epoch value of 1372672923). When the timestamp is rounded down to the nearest hour (i.e., rounded to 10:00:00), the timestamp can be split into an upper timestamp 522 having a UNIX epoch value of 1372672800 and a lower timestamp 529 of 123 that represents the "remainder" from rounding down. Accordingly, any data having a timestamp between or including 10:00:00 GMT and 10:59:59 GMT on July 1, 2013 can be stored in a column associated with the rowkey 520 (i.e., any data having an upper timestamp of 1372672800). FIG. 5B further illustrates an example cell value 530 storing a measurement value associated with the column 525 and corresponding process variable. As shown in FIG. 5B, the measurement value of "476" can represent a value of the process variable as collected within the process plant 10.

[00143] FIG. 5C depicts an example storage table 531 including example rowkeys 520 and example column qualifiers 535 for an example column family "Time." As illustrated in FIG. 5C, the example rowkeys 520 include a concatenation of a process variable identification and an upper timestamp (e.g., "098.1372672800"). In the example storage table 531, the upper timestamp is rounded to sixty (60) minutes. Each of the example column qualifiers 535 is a concatenation of a type of value, a status of the process variable, and a lower timestamp. For example, the column qualifier REAL.GOOD.+20 has a type REAL, a status GOOD, and a time offset of 20 (i.e., 20 seconds after the rounded upper timestamp of 1372672800). Further, the cells of the storage table 531 store measurements corresponding to a given rowkey and column

qualifier. For example, the measurement associated with rowkey 057.1372672800 and column qualifier REAL.GOOD.+40 is 45.4. The table 531 may contiguously store any data corresponding to the same rowkey and column family. For example, the table 531 contiguously stores data corresponding to a particular process variable identification (e.g., 098 or 057) over a 60-minute period.

[00144] Generally, to write a measurement, the big data appliance 102 builds the appropriate rowkey with unique data and determines the associated column family and column qualifier(s). Further, the big data appliance 102 identifies which bytes to store in the cell corresponding to the column family and column qualifier(s), and writes the associated record. For example, the table 531 writes the values of the process variables in the associated records. A user or administrator associated with the process control system 10 may query the big data appliance 102 for stored data. In particular, the query may specify a rowkey range as well as define applicable filter criteria, such as an upper timestamp range. The big data appliance 102 may execute the query and return identified results to the querying user.

[00145] The big data appliance 102 can employ "snapshot" or "zoom" features to enable users to gauge a long-term view and overall context of the data while still enabling more detailed views of the data. The snapshot or zoom features further enable users to identify snapshots of data corresponding to various timestamp ranges. Using various techniques, the big data appliance 102 can create and store aggregates of the data for specific time periods. For example, the big data appliance 102 can store the minimum, maximum, and closing measurement values for each hour of data (i.e., from measurement values corresponding to a common upper timestamp).

14 06 21

[00146] The big data appliance 102 may support Apache Hadoop for storage and large scale processing of the associated data. In some cases, the big data appliance 102 may implement the MapReduce framework associated with the Apache HBase database, which enables users or administrators to reduce data while at the same time enabling the HBase infrastructure to utilize parallel distributed programs. The MapReduce framework enables the division of a dataset and to run it in parallel over multiple nodes. Specifically, the users or administrators may divide the storage problem into simpler Map() and Reduce() functions for filtering, sorting, and summary operations, while the MapReduce framework automatically marshalls the distributed servers, runs the various tasks in parallel, manages communications and data transfers between the various parts of the system, provides for redundancy and failures, and manages the overall process.

[00147] FIG. 5D depicts an example data record created using the "snapshot" techniques as discussed herein. For example, the big data appliance 102 may use the MapReduce framework to create the data record. As illustrated in FIG. 5D, an example rowkey 534 and an example column 535 each includes various fields or attributes of the process control data. In particular, the example rowkey 534 includes an identification of the process variable and an upper timestamp in a scale of year, and the example column 535 includes a column family, a type, a status, and a lower timestamp in a scale of day. In other words, any column having a lower timestamp corresponding to a day in the year specified by the upper timestamp is associated with that corresponding rowkey. FIG. 5D also depicts an example cell value 540 that the big data appliance 102 may store according to a corresponding rowkey and column pair. The example cell value 540 includes an aggregated high value 541, an aggregated low value 542, and an aggregated close value 543. The aggregated high value 541 and the aggregated low value 542 respectively represent the highest and lowest measurement values among the data records with a timestamp corresponding to the upper timestamp

of the rowkey 534. Further, the aggregated close value 543 represents the last measurement value (by time) among the data records with a timestamp corresponding to the upper timestamp of the rowkey 534. Although the aggregated high value 541, the aggregated low value 542, and the aggregated close value 543 are depicted in FIG. 5D, it should be appreciated that other snapshot values are envisioned such as, for example, average value, mode value, median value, standard deviation, and/or other values.

[00148] FIG. 6 depicts example interfaces associated with enabling users to view various snapshots of process control data. The big data appliance 102 may employ a data explorer module (or similar module) to enable a user or administrator to select parameters for a snapshot view and then view the corresponding snapshot view. A first interface 680 corresponds to a selection screen to enabling the user to select parameters for a snapshot view. The first interface 680 includes a process variable selection option 681 that enables the user to select one or more process variables with data that the user wishes to view or analyze. The first interface 680 further includes a time period selection option 682 that enables the user to select one or more time periods for the selected process variable data. Further, the first interface 680 includes a metric selection option 683 that enables the user to select one or more various parameters corresponding to the desired "zoom" value for the selected time periods. The first interface 608 depicts selections of process variables "PV008" and "PV059", time periods for hour and month, and minimum, maximum, and average metrics.

[00149] A second interface 685 depicts charts resulting from the selections of the first interface 680. In particular, the second interface 685 depicts hourly and monthly charts 686, 687 for process variable "PV008," and hourly and monthly charts 688, 689 for process variable "PV059." The respective hourly charts 686, 688 depict hourly minimum, maximum, and average values for "PV008" and "PV059" over a period of six (6) hours. The respective monthly charts 687, 689 depict monthly minimum, maximum, and average values for "PV008" and "PV059" over a period of five (5) months. Using the first interface 680 and the second interface 685, the user or administrator may effectively and efficiently assess and analyze snapshots of process control data and parameters thereof without having to filter through all of the recorded data.

[00150] FIG. 7 depicts a flow diagram 700 related to the big data appliance 102 employing a "snapshot" technique (e.g., the MapReduce technique) to aggregate raw data. In 705, the big data appliance 102 maps input data according to the identification for each associated timestamp and then sorts the mapped data according to the identification to generate compiled measurement data corresponding to each identification. In 710, the big data appliance shuffles, merges, and reduces the data into the target output data. For example, the big data appliance can output, for each of the identifications, the high, low, and close measurement values for a specific time period. Thus, the big data appliance can effectively generate the snapshots and provide the snapshots to a requesting user or administrator.

[00151] FIGS. 8-10 illustrate flow diagrams of example methods 800, 900, 1000 for retrieving, presenting, or otherwise storing big data in a process control system or process plant according to the big data schema as discussed herein. A big data appliance (e.g., the big data appliance 102 of FIG. 1) can implement the methods 800, 900, 1000 in the process control system big data network 100 of FIG. 1, or in any other suitable network or system that supports big data in a process control system or process plant. For illustrative (and non-limiting) purposes, the methods 800, 900, 1000 are discussed below with simultaneous reference to FIGS. 1-7.

14 06 21

[00152] Referring to FIG. 8, the method 800 relates to storing process control data in a data storage device. In particular, the method 800 stores the process control data via leveraging rowkeys generated according to timestamps associated with the process control data. At a block 805 of the method 800, the big data appliance receives the process control data, for example via one or more data receivers 122. The data may correspond to a process plant and/or to a process being controlled by a process plant, and may include a plurality of portions or segments. For example, the data may include real-time data generated while controlling a process in the process plant, configuration data, batch data, network management and traffic data of various networks included in the process plant, data indicative of user or operator actions, data corresponding to the operation and status of equipment and devices included in the plant, data generated by or transmitted to entities external to the process plant, and other data.

[00153] At a block 810, the big data appliance identifies various data measurements, indications, and other attributes from each respective portion of the received data. For example, the big data appliance may identify, for each respective portion of the data, an identification of a respective process variable, a timestamp associated with the respective portion of the data, a measurement value associated with the respective process variable, a data type of the measurement, and/or a status associated with the measurement value. The timestamp may be represented as a UNIX epoch value, and may include a first timestamp portion reflecting the timestamp rounded down by a certain degree, and a second timestamp portion reflecting a remainder of the timestamp that was rounded down by the certain degree. Accordingly, the first timestamp portion may correspond to an upper timestamp of the timestamp and the second timestamp portion may correspond to a lower timestamp of the timestamp.

14 06 21

[00154] At block 815, the big data appliance identifies, for each respective portion of the received data, a rowkey based on the respective process variable and the first timestamp portion. In some cases, the big data appliance may generate the rowkey (e.g., if the particular rowkey does not exist) by concatenating an identification of the respective process variable and the first timestamp portion, and store the rowkey in a data storage device. In other cases, the big data appliance may identify a rowkey already stored in a data storage device that corresponds to the respective process variable and the first timestamp portion.

[00155] At block 820, the big data appliance stores, for each respective portion of the data, various data within a portion of the data storage device associated with the rowkey. In some cases, the big data appliance may store the second timestamp portion, the measurement value, optionally the type of the measurement value, and optionally the status of the measurement value. In this regard, the rowkey can include recorded process control data having a timestamp corresponding to the upper timestamp of the rowkey (i.e., data recorded within a period of time indicated by the rowkey), thus reducing the amount of storage necessary to store the process control data and reducing the amount of time needed to access and retrieve the data.

[00156] At block 825, the big data appliance determines if additional process control data is received. For example, the additional process control data can be additional process control data recorded by the process control system. If additional process control data is received ("YES"), processing can return to 810 or proceed to any other functionality. If additional process control data is not received ("NO"), processing can end, repeat, or proceed to any other functionality.

[00157] Referring to FIG. 9, the method 900 relates to storing process control data according to snapshot conventions that enable users to retrieve snapshot data associated with the process control data. At a block 905 of the method 900, the big data appliance receives process control data, for example via one or more data receivers 122. The data may correspond to a process plant and/or to a process being controlled by a process plant, and may include a plurality of portions. For example, the data may include real-time data generated while controlling a process in the process plant, configuration data, batch data, network management and traffic data of various networks included in the process plant, data indicative of user or operator actions, data corresponding to the operation and status of equipment and devices included in the plant, data generated by or transmitted to entities external to the process plant, and other data.

[00158] At a block 910, the big data appliance, for each record, examines the data to identify (1) a respective process variable, (2) a measurement value associated with the respective process variable, and (3) a timestamp including a first timestamp portion and second timestamp portion. The timestamp may be represented as a UNIX epoch value, and may include a first timestamp portion reflecting the timestamp rounded down by a certain degree, and a second timestamp portion reflecting a remainder of the timestamp that was rounded down by the certain degree. Accordingly, the first timestamp portion may correspond to an upper timestamp of the timestamp and the second timestamp portion may correspond to a lower timestamp of the timestamp.

[00159] At a block 915, the big data appliance determines that a set period of time based on the first timestamp portion has elapsed. For example, if the first timestamp portion specifies 10:00:00 AM and is rounded to the nearest hour, the set period of time elapses at 11:00:00 AM. In one case, the big data appliance can determine that the set period of time has elapsed by comparing a current time to first timestamp portion.

14 06 21

[00160] When the set period of time has elapsed at block 920, the big data appliance identifies at least one statistical parameter from one or more of the plurality of records having a timestamp within the set period of time. The at least one statistical parameter may be one or more of: a high value of the respective one or more measurement values, a low value of the respective one or more measurement values, a most recent value of the respective one or more measurement values, a standard deviation of the respective one or more measurement values, an average of the respective one or more measurement values, and a median of the respective one or more measurement values. It should be appreciated that other statistical parameters associated with the plurality of records are envisioned.

[00161] At block 925, the big data appliance filters the at least one statistical parameter according to the respective process variable. In particular, the big data appliance can separate the identified statistical parameter(s) according to the respective process variable such that a user can access or retrieve statistical parameter data according to the specified process variable.

[00162] At block 930, the big data appliance stores the at least one statistical parameter in a time period data record associated with the set period of time and the respective process variable. Therefore, the time period data record can store any relevant data corresponding to a specific time period and a user need not review or access individual rowkeys or records to analyze aggregate process control data.

[00163] Referring to FIG. 10, the method 1000 relates to retrieving “shapshots” of process control data that enables users to analyze the process control data across specified time periods. At a block 1005 of the

method 1000, the big data appliance receives a request from a user to retrieve a portion of process control data, wherein the request indicates a set period of time. The process control data may correspond to a process plant and/or to a process being controlled by a process plant, and may include a plurality of portions. For example, the data may include real-time data generated while controlling a process in the process plant, configuration data, batch data, network management and traffic data of various networks included in the process plant, data indicative of user or operator actions, data corresponding to the operation and status of equipment and devices included in the plant, data generated by or transmitted to entities external to the process plant, and other data.

[00164] At block 1010, the big data appliance retrieves the portion of the process control data corresponding to the set period of time wherein the portion of the process control data includes a plurality of records. The set period of time may be defined by an upper timestamp associated with the plurality of records, wherein the plurality of records each indicate a lower timestamp that falls within the set period of time defined by the upper timestamp.

[00165] At block 1015, the big data appliance, for each of the plurality of records, identifies (1) a respective process variable, (2) a measurement value associated with the respective process variable, and (3) a timestamp that falls within the set period of time. The timestamp may be represented as a UNIX epoch value, and may include a first timestamp portion reflecting the timestamp rounded down by a certain degree, and a second timestamp portion reflecting a remainder of the timestamp that was rounded down by the certain degree. Accordingly, the first timestamp portion may correspond to an upper timestamp of the timestamp and the second timestamp portion may correspond to a lower timestamp of the timestamp.

[00166] At block 1020, the big data appliance generates aggregate process control data from the plurality of records. The big data appliance may aggregate the plurality of records according to at least one statistical parameter associated with each of the plurality of records. For example, the at least one statistical parameter may be one or more of: a high value of the respective one or more measurement values, a low value of the respective one or more measurement values, a most recent value of the respective one or more measurement values, a standard deviation of the respective one or more measurement values, an average of the respective one or more measurement values, and a median of the respective one or more measurement values. The big data appliance may also calculate the at least one statistical parameter from the plurality of records, such as in cases in which the at least one statistical parameter is not explicitly indicated in the aggregate process control data. For example, the big data appliance may calculate averages, standard deviations, high values, low values and/or other metrics to generate the aggregate process control data. In some cases, a user may specify a desired statistical parameter, for example as part of a request or command, whereby the big data appliance calculates the appropriate statistical parameter from the plurality of records.

[00167] At block 1025, the big data appliance presents the aggregate process control data to the user. For example, the big data appliance may present the aggregate process control data as numeric data, a chart, a graph, or any other type of numeric data or graphical indication. Further, the aggregate process control data may indicate any statistical parameters that are identified or calculated from the plurality of records.

14 06 21

[00168] At block 1030, the big data appliance determines if an additional request is received. For example, the user may wish to narrow or expand the aggregated process control data, or may wish to perform other calculations on the process control data. If the additional request is received ("YES"), processing can return to 1010 or proceed to any other functionality. If the additional request is not received ("NO"), processing can end, repeat, or proceed to any other functionality.

[00169] When implemented in software, any of the applications, services, and engines described herein may be stored in any tangible, non-transitory computer readable memory such as on a magnetic disk, a laser disk, solid state memory device, molecular memory storage device, or other storage medium, in a RAM or ROM of a computer or processor, etc. Although the example systems disclosed herein are disclosed as including, among other components, software and/or firmware executed on hardware, it should be noted that such systems are merely illustrative and should not be considered as limiting. For example, it is contemplated that any or all of these hardware, software, and firmware components could be embodied exclusively in hardware, exclusively in software, or in any combination of hardware and software. Accordingly, while the example systems described herein are described as being implemented in software executed on a processor of one or more computer devices, persons of ordinary skill in the art will readily appreciate that the examples provided are not the only way to implement such systems.

[00170] Thus, while the present invention has been described with reference to specific examples, which are intended to be illustrative only and not to be limiting of the invention, it will be apparent to those of ordinary skill in the art that changes, additions or deletions may be made to the disclosed embodiments without departing from the spirit and scope of the invention.

[00171] In the present specification "comprise" means "includes or consists of" and "comprising" means "including or consisting of".

[00172] The features disclosed in the foregoing description, or the following claims, or the accompanying drawings, expressed in their specific forms or in terms of a means for performing the disclosed function, or a method or process for attaining the disclosed result, as appropriate, may, separately, or in any combination of such features, be utilised for realising the invention in diverse forms thereof.

14 06 21

Claims

1. A system for storing process control data associated with a process control plant, comprising:

a data storage device configured to store process control data corresponding to at least one process that is controlled by a plurality of devices in the process control plant; and

a data receiver configured to receive the process control data from the plurality of devices, the process control data including a plurality of portions, and each portion of the data associated with a respective process variable; and

a processor adapted to interface with the data storage device and the data receiver and configured to, for the each portion of the data:

identify a rowkey stored in the data storage device, the rowkey based on (i) at least a portion of a first attribute of the each portion of the data and (ii) a first portion of a third attribute of the each portion of the data,

identify a column qualifier stored in the data storage device, the column qualifier based on (i) at least a portion of a second attribute of the each portion of the data and (ii) a second portion of the third attribute, and

store, within a portion of the data storage device associated with the rowkey and the column qualifier, a measurement value associated with the respective process variable corresponding to the each portion of the data.

14 06 21

2. The system of claim 1, wherein:

the first attribute includes one of: a timestamp corresponding to the measurement value associated with the respective process variable, an identification of the respective process variable, a type of data corresponding to the measurement value, or a status of the respective process variable; and

the second attribute includes another one of: the timestamp corresponding to the measurement value associated with the respective process variable, the identification of the respective process variable, the type of data corresponding to the measurement value, or the status of the respective process variable.

3. The system of any preceding claim, wherein at least one of:

the rowkey is a concatenation of the at least the portion of the first attribute and the first portion of the third attribute; or

the column qualifier is a concatenation of the at least the portion of the second attribute and the second portion of the third attribute.

4. The system of any preceding claim, wherein the third attribute is a timestamp corresponding to the measurement value associated with the respective process variable.

5. The system of claim 4, wherein the first portion of the timestamp corresponds to an upper timestamp of the timestamp and the second portion of the timestamp corresponds to a lower timestamp of the timestamp.

6. The system of claim 4, wherein the timestamp is represented as a UNIX epoch value, the first portion of the timestamp is the timestamp rounded down by a certain degree, and the second portion of the timestamp is a remainder of the timestamp that was rounded down by the certain degree.

7. The system of any one of the preceding claims, wherein:

the column qualifier is a first column qualifier of a plurality of column qualifiers, each column qualifier corresponding to at least a portion of a different attribute of the each respective portion of the data;

the rowkey is associated with each column qualifier of the plurality of column qualifiers; and

the measurement value associated with the each respective portion of the data is stored within a portion of the data storage device associated with the row key and each column qualifier of the plurality of column qualifiers.

14 06 21

8. The system of the preceding claim, wherein:

the plurality of column qualifiers is associated with a first column family of a plurality of column families, each column family indicating a respective category of a respective plurality of column qualifiers;

the rowkey is associated each column family of the plurality of column families; and

the measurement value associated with the each respective portion of the data is stored within a portion of the data storage device associated with the rowkey, the column qualifier, and the first column family.

9. The system of any one of the preceding claims, wherein the first attribute is an identification of the respective process variable.

10. The system of any one of the preceding claims, wherein the second attribute is a status of the respective process variable.

11. The system of the preceding claim, wherein:

the processor is further configured to, for the each portion of the data, identify another column qualifier based on a fourth attribute of the each portion of the data, the fourth attribute being a type of data corresponding to the measurement value associated with the respective process variable; and

the measurement value associated with the each respective portion of the data is stored within a portion of the data storage device associated with the rowkey, the column qualifier, and the another column qualifier.

12. The system of any one of the preceding claims, wherein the measurement value associated with the respective process variable is a statistical measurement value of the respective process variable determined from a plurality of measurement values of the respective process variable.

13. A method for storing process control data associated with a process control plant, the method comprising:

receiving process control data from a plurality of devices configured to control at least one process in the process control plant, the process control data including a plurality of portions, and each portion of the data associated with a respective process variable; and

for the each portion of the data:

identifying, by a processor based on (i) at least a portion of a first attribute of the each portion of the data and (ii) a first portion of a third attribute of the each portion of the data, a rowkey stored in a data storage device;

identifying, by the processor based on (i) at least a portion of a second attribute of the each portion of the data and (ii) a second portion of the third attribute, a column qualifier stored in the data storage device; and

storing, within a portion of the data storage device associated with the rowkey and the column qualifier, a measurement value associated with the respective process variable corresponding to the each portion of the data.

14 06 21

14. The method of claim 13, wherein:

identifying the rowkey based on the at least the portion of the first attribute includes identifying the rowkey based on one of: a timestamp corresponding to the measurement value associated with the respective process variable, an identification of the respective process variable, a type of data corresponding to the measurement value, or a status of the respective process variable; and

identifying the column qualifier based on the at least the portion of the second attribute includes identifying the column qualifier based on another one of: the timestamp corresponding to the measurement value associated with the respective process variable, the identification of the respective process variable, the type of data corresponding to the measurement value, or the status of the respective process variable.

15. The method of claim 13 or 14, wherein at least one of:

the rowkey is a concatenation of the at least the portion of the first attribute and the first portion of the third attribute; or

the column qualifier is a concatenation of the at least the portion of the second attribute and the second portion of the third attribute.

16. The method of any one of claims 13-15, wherein the third attribute is a timestamp corresponding to the measurement value associated with the respective process variable.

17. The method of claim 16, wherein:

the first portion of the timestamp corresponds to an upper timestamp of the timestamp and the second portion of the timestamp corresponds to a lower timestamp of the timestamp; or

the timestamp is represented as a UNIX epoch value, the first portion of the timestamp is the timestamp rounded down by a certain degree, and the second portion of the timestamp is a remainder of the timestamp that was rounded down by the certain degree.

14 06 21

18. The method of any one of claims 13-17, wherein:

the column qualifier is a first column qualifier of a plurality of column qualifiers;

the rowkey is associated with each column qualifier of the plurality of column qualifiers;

each column qualifier is based on at least a portion of a different attribute of the each respective portion of the data; and

storing the measurement value within the portion of the data storage device associated with the rowkey and the column qualifier includes storing the measurement value within a portion of the data storage device associated with the row key and each column qualifier of the plurality of column qualifiers.

19. The method of the preceding claim, wherein:

the plurality of column qualifiers is associated with a first column family of a plurality of column families, each column family indicating a respective category of a respective plurality of column qualifiers;

the rowkey is associated each column family of the plurality of column families; and

storing the measurement value within the portion of the data storage device associated with the rowkey and the column qualifier includes storing the measurement value within a portion of the data storage device associated with the rowkey, the column qualifier, and the first column family.

20. The method of any one of claims 13-19, wherein identifying the rowkey based on at least a portion of the first attribute includes identifying the rowkey based on an identification of the respective process variable.

21. The method of any one of claims 13-20, wherein identifying the column qualifier based on at least a portion of the second attribute includes identifying the column qualifier based on a status of the respective process variable.

22. The method of claim 21,

further comprising, for the each portion of the data, identifying another column qualifier based on a fourth attribute of the each portion of the data, the fourth attribute being a type of data corresponding to the measurement value associated with the respective process variable; and

wherein storing the measurement value within the portion of the data storage device associated with the rowkey and the column qualifier includes storing the measurement value within a portion of the data storage device associated with the rowkey, the column qualifier, and the another column qualifier.

23. The method of any one of claims 13-22, wherein storing the measurement value associated with the respective process variable includes storing a statistical measurement value of the respective process variable determined from a plurality of measurement values of the respective process variable.

14 06 21