



(51) International Patent Classification:
G06F 11/34 (2006.01)

(21) International Application Number:
PCT/CN2017/096358

(22) International Filing Date:
08 August 2017 (08.08.2017)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **ALIBABA GROUP HOLDING LIMITED**
[GB/CN]; Fourth Floor, One Capital Place, P.O. Box 847,
George Town, Grand Cayman (KY).

(72) Inventors: **KINGSUM, Chow**; Hangzhou, Zhejiang (US).
ZHU, Wanyi; Hangzhou, Zhejiang (CN). **LI, Chengdong**;
Hangzhou, Zhejiang (CN). **WU, Puyuan**; Hangzhou, Zhe-
jiang (CN).

(74) Agent: **BEIJING TSINGYUANHUI INTELLECTUAL
PROPERTY LAW FIRM**; Suzhou Street No.362 Haidian
District, Beijing 100080 (CN).

(81) Designated States (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,
KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

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

(54) Title: PROCESSING PERFORMANCE DATA FOR MACHINE LEARNING

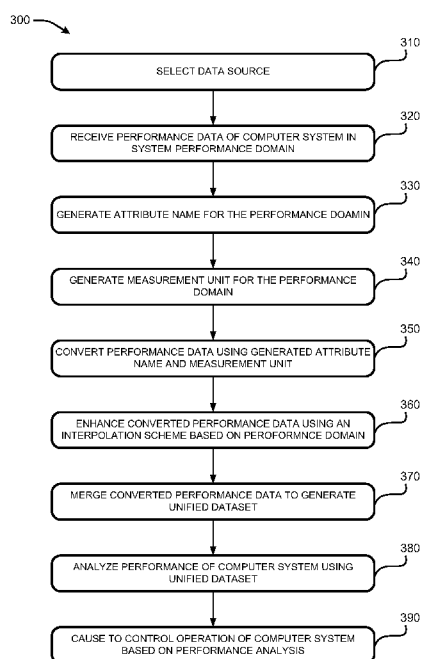


FIG. 3

(57) Abstract: Methods and computing devices for processing system performance data for machine learning are provided. System performance data in system performance domains are received and converted under naming conventions each including an attribute name and a measurement unit. The converted system performance data are enhanced using interpolation and merged into a unified dataset with a unified time scale.



PROCESSING PERFORMANCE DATA FOR MACHINE LEARNING

BACKGROUND

[0001] Cloud service providers and data centers usually process large number of computer system performance data to improve data center operation efficiency such as software application migrations and capacity planning. Approaches to implement trained machine learning to solve large scale performance optimization include manual examination of relations among variables.

BRIEF DESCRIPTION OF THE DRAWINGS

[0002] The detailed description is set forth with reference to the accompanying figures. In the figures, the left-most digit(s) of a reference number identifies the figure in which the reference number first appears. The use of the same reference numbers in different figures indicates similar or identical items or features.

[0003] FIG. 1 illustrates an example system for processing performance data of a computing system.

[0004] FIG. 2 illustrates an example operation environment of a system of FIG. 1 for processing performance data of a computing system.

[0005] FIG. 3 illustrates an example operation process of the system of FIG. 1.

[0006] FIG. 4 illustrates another example operation process of the system of FIG. 1.

[0007] FIG. 5 illustrates another example operation process of the system of FIG. 1.

DETAILED DESCRIPTION

[0008] The disclosure provides a solution to process performance data of a computing system. In the following detailed description, references are made to the accompanying drawings that form a part hereof, and in which are shown by way of illustration specific configurations or examples, in which like numerals represent like elements throughout the several figures.

1. Example Devices

[0009] Referring to FIG. 1, an example system 10 for processing performance data of a computing system is illustrated. As shown in FIG. 1, system 10 may include one or more memory 100 which stores computer executable instructions, which when executed by one or more processing units, configure the processing units and the related computing device(s) to implement a data receiving unit 110, a naming convention unit 120, a data enhancing unit 130, a unification unit 140 and a machine learning unit 150.

[0010] Data receiving unit 110 may further include a data source selection unit 112; naming convention unit 120 may further include an attribute name generation unit 122 (including a measurement unit generation unit 124) and a converting unit 126; data enhancing unit 130 may further include a mapping unit 132, an interpolation scheme selection unit 134 an interpolation operation unit 136, and a confidence interval determination unit 138; unification unit 140 may further include a unified time scale generation unit 142 and a

unified dataset generation unit 144; and machine learning unit 150 may further include an analysis unit 152 and a control unit 154.

[0011] System 10 may also include one or more processing units (PU) 160, interfacing units 170, communication units 180 and/or other components 190.

[0012] It should be appreciated that although system 10 and the units thereof, e.g., data receiving unit 110, naming convention unit 120, data enhancing unit 130, unification unit 140 and machine learning unit 150 are illustrated as a single system, it does not necessarily mean that all components of system 10 are physically or functionally located within a single computing system. All the units as illustrated in system 10 of FIG. 1 may be located on separate computing systems communicating and functioning together in a distributed computing environment.

[0013] FIG. 2 illustrates an example operation environment 200 of system 10. Note that FIG. 2 illustrates operation environment 200 in a distributed computing scheme, which is not necessary for the implementation. Some or all of the communication shown in FIG. 2 may be implemented through data bus (i.e., local communication) within a single computing system.

[0014] As shown in FIG. 2, units of system 10 may communicate with one another through a network 210. One or more units of system 10 may further communicate with data source(s) 230 (230-1 to 230-N) and/or target computing system(s) 240 (240-1 to 240-M) through a network 220. Each data source 230 may acquire/collect performance data of one or more target computing systems 240 in one or more system performance domains. A system performance domain may be any domain that is relevant to the

performance of a target computing system 240 and/or relevant to an analysis thereof. For example, a system performance domain may include a CPU usage, throughput rate, system bandwidth, system response time, etc.

[0015] System 10 and/or components thereof, e.g., data receiving unit 110, may be configured to communicate with data source 230 to receive the system performance data acquired and/or processed by data source 230. It is understood that data source 230 may engage different mechanisms to acquire data and/or process acquired data of target computing system(s) 240 because the data from different data sources may include different formats such as attribute name and measurement units. The sampling frequency of data sources 230 may also vary, which is reflected in the data entry points of the dataset acquired.

[0016] In an example, as illustrated in FIG. 2, the acquired data may be initially processed locally in different data sources 230 in a distributed computing scheme within system 10. The locally and initially processed system performance data may be received by system 10, through data communication via network 220, for further processing. Such a tiered processing of the acquired system performance data in the distributed computing environment may save communication bandwidth, improve capacity allocation, and enable architecture flexibility and adaptability, which are all technical advantages.

[0017] It should be appreciated that although FIG. 2 illustrates network 210 and network 220 with different reference numerals, it does not necessarily mean that network 210 and network 220 are two separate network systems. Each of networks 210 and 220 may include one or more network systems for

data communication, which may overlap and/or be the same physical network systems.

[0018] In operation, data receiving unit 110 may be generally configured to receive system performance data from data sources 230. Data source selection unit 112 may be configured to select multiple data sources 230, among all the pool of data sources 230, to receive performance data. Any criteria may be used in the selection of data source 230 and all are included in the disclosure. For example, selection criteria may be partially based on the quality of the performance data acquired by a data source 230. For another example, selection criteria may be based on a data acquisition approach used by a data source 230 in acquiring performance data of target computing system(s) 240. For another example, selection criteria may be based on whether the performance data from the multiple data sources 230 are sufficiently consistent in some parameters, e.g., sampling frequency in data acquisition, for further processing. In an example implementation, the selection may be made based partially on at least one of a frequency, a time interval, a data metric, and a measurement unit associated with the performance data collected by data sources 230.

[0019] Naming convention unit 120 may be generally configured to generate an attribute name and a measurement unit for a system performance domain and convert the received performance data in the performance domain using the standardized attribute name and measurement unit. In an example, attribute name generation unit 122 is configured to generate the attribute name and measurement unit, and converting unit 126 is configured to perform the

conversion of received performance data using the generated attribute name and measurement unit(s).

[0020] For example, for each system performance domain, e.g., a network bandwidth, attribute name generation unit 122 may generate a unified naming convention for a system performance domain, which may include attribute name and measurement unit. Therefore, system performance data in the system performance domain of, e.g., network bandwidth, that are received from different data sources 230 may be converted under the unified naming convention including the attribute name and the measurement unit. For example, a name convention may be:

<attribute_name>_<unit_of_measurement>. As an illustrative example, with respect to system performance domain of network bandwidth, an example name convention may be *network_bandwidth_kB_s*, which indicates that the units of measurements for network bandwidth are kilobytes per second.

[0021] In an example, depending on different system analysis scenarios, e.g., the selected sources 230 and their data acquiring approaches, the units of measurement for an attribute name may vary. Measurement unit generation unit 124 of attribute name generation unit 122 may be configured to generate a specific measurement unit for a scenario. As an illustrative example of network bandwidth, the measurement unit may also be packets per second and the name convention may be *network_bandwidth_pack_s*.

[0022] Measurement unit generation unit 124 may also be configured to generate a measurement unit used for data unification. For example, still with the illustrative example of network bandwidth, in a case that data receiving unit 110 receives network bandwidth data in both kilobytes per second format and

packet per second format, a new measurement unit of kilobytes per packet may be generated and consequently a new name convention of:

<new_name>_kB_pack.

[0023] Data enhancing unit 130 may be generally configured to enhance the converted performance data with an interpolation operation using an interpolation scheme determined based on a type of the performance domain.

[0024] Within data enhancing unit 130, mapping unit 132 may be configured to map the converted system performance data (e.g., converted by converting unit 126) with a unified time scale. The performance data originally acquired by data source 230 may include different time scales due to the different acquisition procedures/approaches, e.g., different data sampling frequency/interval, sampling points mismatch, etc. Conversion of the performance data using the attribute name and measurement units may not completely resolve the issue of inconsistent time scales among performance data in the same performance domain.

[0025] In the mapping operation, the converted data may be mapped with a unified time scale. Data entries that do not fit into the unified time scale may be removed from further operations (but may not need to be deleted and may be maintained for other analytical purposes).

[0026] Interpolation scheme selection unit 134 may be configured to select an interpolation scheme based on different performance domains. For performance data in different performance domains, different interpolation schemes may be selected to determine an estimated value to impute for a missing value. The interpolation scheme may include a zero interpolation, an arithmetic mean substitution, a cubic spline interpolation or a Kalman filter

interpolation. For example, a cubic spline interpolation scheme may be selected as a default interpolation scheme. A cubic spline interpolation scheme generally tends to be more stable than polynomial interpolation and tends to include less possibility of wild oscillations between the tabulated points.

[0027] Besides or instead of the default scheme, in the case that the system performance domain includes CPU frequency data, an arithmetic mean substitution interpolation scheme may be selected for an interpolation operation. In the case that the system performance domain includes instantaneous performance data, a Kalman filter interpolation scheme may be selected for the interpolation operation. In the case that the system performance domain includes event log data, a Zero interpolation scheme may be selected for the interpolation operation. In an example, interpolation scheme selection unit 134 may select multiple interpolation schemes as candidates for a performance domain. The selected multiple candidates may be used in an interpolation operation with an order of priority.

[0028] Interpolation operation unit 136 may be configured to perform an interpolation operation to obtain an estimated value for a missing value using an interpolation scheme(s) selected by the interpolation scheme selection unit 134. In a case that multiple interpolation schemes are selected, interpolation operation unit 136 may first use an interpolation scheme that has a higher priority in the order.

[0029] Confidence interval determination unit 138 may be configured to determine whether a value estimated by the operation of interpolation operation unit 136 is within a confidence interval. The confidence interval may be determined dynamically by confidence interval determination unit 138 for a

performance domain and/or a performance data of the performance domain, or may be received, e.g., through user inputs.

[0030] Unification unit 140 may be configured to merge converted performance data in one or more system performance domains with a unified time scale to generate a unified dataset. Specifically, for example, unified time scale generation unit 142 may be configured to generate a unified time scale for one or more performance domains. The unified time scale(s) may be used for the generation of a unified dataset and/or may be used for other operations. For example, a unified time scale generated by unified time scale generation unit 142 may be used in the operation of mapping unit 132 in mapping the converted performance data.

[0031] In an example, unified time scale generation unit 142 may generate a single unified time scale for all the performance data enhanced and merged into the unified dataset. That is, the same unified data scale may be used in the mapping operation of mapping unit 132 and may be used in the operation of unification. In another example, a time scale(s) is generated separately for the data enhancing operation(s) and another time scale is generated separately for the dataset unification operation.

[0032] Unified dataset generation unit 144 may be configured to merge multiple converted performance data for one or more system performance domains to generate a unified dataset using a unified time. In the merging, data items (values) that do not map into the unified time scale may not enter the unified dataset, but may be kept separately for potential other analytical uses.

[0033] In an example, unified dataset generation unit 144 may be configured to merge the converted performance data after the enhancing

operation by data enhancing unit 130. However, the scope of the disclosure is not limited to this example and unification unit 140 may be configured to merge converted performance data without the enhancing operation(s).

[0034] Machine learning unit 150 may be configured to learn a target computing system using the unified dataset(s), wherein analysis unit 152 may be configured to analyze/cause to analyze the unified dataset to determine the performance status of the target computing system, and control unit 154 may be configured to control/cause to control the target computing system 240 based on a result of the analysis. For example, based on a result of the analysis, a configuration of the target computing system 240, e.g., operation parameters and/or computing capacity allocation, may be adjusted.

2. Example Processes

[0035] Referring to FIG. 3, an example operation process 300 of system 10 of FIG. 1 is illustrated. In example operation 310, data source selection unit 112 may select multiple data sources 230, among all the pool of data sources 230, to receive performance data. Any criteria may be used in the selection of data source 230 and all are included in the disclosure. In an example implementation, the selection may be made based on at least one of a frequency, a time interval, a data metric, and a measurement unit associated with the performance data collected by data sources 230.

[0036] In example operation 320, data receiving unit 110 may receive performance data of a target computing system(s) 240 from the selected multiple data sources 230. Any approach may be adopted in the data receiving and all are included in the disclosure. Although FIG. 1 illustrates that data

source 230 are separate from data receiving unit 110, the disclosure is not limited to such a specific example. Data receiving unit 110 may physically reside together with and/or function together with one or more data source(s) 230. As such, the receiving performance data from multiple data sources 230 includes the scenario(s) that data receiving unit 110 causes/controls data source 230 to acquire system performance data from target computing system(s) 240.

[0037] In example operation 330, attribute name generation unit 122 may generate an attribute name for a system performance domain.

[0038] In example operation 340, measurement unit generation unit 124 may generate a measurement unit for the system performance domain associated with the generated attribute name. In an example, attribute name and measurement unit are generated together, e.g., an attribute name includes measurement unit. For an illustrative example performance domain of network bandwidth, an example attribute name may be *network_bandwidth_pack_s*, wherein “network bandwidth” is attribute name and “packet per second” is the measurement unit. Accordingly, operations 330 and 340 may be one operation and/or may be operated together.

[0039] In another example, a measurement unit may be determined separately from the attribute name for different scenarios. For example, an attribute name may be relatively stable for a system performance domain, but the measurement unit may be updated based on different scenarios of, e.g., different performance analysis requirements, different data source 230, etc.

[0040] In an example, in generating the measurement unit for a system performance domain, measurement unit generation unit 124 may select a

measurement unit that tends to minimize data value missing in the following data conversion for multiple data sources. For example, measurement unit generation unit 124 may evaluate the sampling interval/frequency adopted by each of the selected data sources 230 in collecting performance data of a target computing system(s) 240 and select a measurement unit that tends to minimize the detrimental effect(s) of different sampling frequencies/intervals of different data sources 230 in the subsequent conversion operation(s). Measurement unit generation unit 124 may also evaluate the mismatching in time scales among system performance data obtained from different data sources 230, and may select a measurement unit that tends to minimize the detrimental effect of such mismatched time scales in the subsequent conversion operations.

[0041] In an example, measurement unit generation unit 124 may receive, e.g., from unified time scale generation unit 142, a unified time scale configured to be used for one or more of the data enhancing operation or the data unification operation. Measurement unit generation unit 124 may select a measurement unit based on the unified time scale to, e.g., minimize missing value .

[0042] In example operation 350, converting unit 126 may convert the received system performance data from multiple data sources 230 using the determined attribute name and measurement unit.

[0043] In example operation 360, data enhancing unit 130 may enhance the converted performance data with an interpolation operation using an interpolation scheme based on a type of the performance domain. Any existing and/or future developed interpolation scheme(s) may be used and all are included in the disclosure. The interpolation scheme may include a zero

interpolation, an arithmetic mean substitution, a cubic spline interpolation or a Kalman filter interpolation. For example, a cubic spline interpolation scheme may be selected as a default imputation scheme. A cubic spline interpolation scheme generally tends to be more stable than polynomial interpolation and tends to include less possibility of wild oscillations between the tabulated points.

[0044] Besides or instead of the default scheme, in a case that the system performance domain includes CPU frequency data, an arithmetic mean substitution interpolation scheme may be selected for an interpolation operation. In a case that the system performance domain includes instantaneous performance data, a Kalman filter interpolation scheme may be selected for the interpolation operation. In a case that the system performance domain includes event log data, a Zero interpolation scheme may be selected for the interpolation operation. In an example, interpolation scheme selection unit 134 may select multiple interpolation schemes as candidates for a performance domain.

[0045] FIG. 4 illustrates an example operation flow 400 including example details of operation 360 of FIG. 3.

[0046] Referring now to FIG. 4, in example operation 410, mapping unit 132 of data enhancing unit 130 may map the converted system performance data (e.g., converted by converting unit 126) with a unified time scale. The unified time scale may be received from unified time scale generation unit 142 and/or may be received from other sources, e.g., through a user input. In the mapping operation, data entries that do not fit into the unified time scale may be removed from the further enhancing operations (but may not need to be deleted).

[0047] In example operation 420, interpolation scheme selection unit 134 may select an interpolation scheme based on a type of the performance domain. For performance data in different performance domains, different interpolation schemes may be selected for estimating a missing value.

[0048] In example operation 430, confidence interval determination unit 138 may determine a confidence interval for an estimated value to impute into the converted performance data to replace a missing value for a system performance domain. The confidence interval may be determined dynamically by confidence interval determination unit 138 for a performance domain and/or a performance data of the performance domain, or may be received, e.g., through user inputs.

[0049] In example operation 440, interpolation operation unit 136 may perform an interpolation operation to obtain an estimated value for a missing value using an interpolation scheme(s) selected by the interpolation scheme selection unit 134. In an example, for converted system performance data of different system performance domains, different interpolation schemes may be used in the interpolation operation, corresponding to the interpolation scheme selection operation of operation 410.

[0050] In example operation 450, confidence interval determination unit 138 may determine whether an estimated value determined by the operation of interpolation operation unit 136 is within the confidence interval. If an estimated value is within the confidence interval, the operation flow proceeds to example operation 460, where the estimated value will be imputed into the system performance data to substitute for a missing value. If the estimated value is not within the confidence interval, the operation flow may revert back to example

operation 420, where another interpolation scheme may be selected for the interpolation operation.

[0051] Referring back to FIG. 3, in example operation 370, unification unit 140 may merge converted performance data in one or more system performance domains with a unified time scale to generate a unified dataset. FIG. 5 illustrates an example operation flow 500 of example details of operation 370.

[0052] Referring now to FIG. 5, in example operation 510, unified time scale generation unit 142 generates a unified time scale for one or more performance domains. In an example, unified time scale generation unit 142 may generate a single unified time scale for all the performance data enhanced and to be merged into a unified dataset. In an example, unified time scale generation unit 142 may select a time scale that minimizes data value missing in the mapping of data (converted data and/or enhanced data) with the unified time scale.

[0053] In example operation 520, unified dataset generation unit 144 may merge multiple covered performance data for one or more system performance domains using the unified time scale to generate a unified dataset. In the merging, data items (values) that do not map into the unified time scale may not enter the unified dataset, but may be kept separately for potential other uses.

[0054] In an example, unified dataset generation unit 144 may merge the converted performance data after the enhancing operation by data enhancing unit 130. However, the scope of the disclosure is not limited to this example

and unification unit 140 may merge converted performance data without the enhancing operation.

[0055] Referring back to FIG. 3, in example operation 380, analysis unit 152 may analyze/cause to analyze the unified dataset to determine the performance status of the target computing system. Any existing and/or future developed performance analysis may be used and all are included in the disclosure.

[0056] In example operation 390, control unit 154 may control/cause to control the target computing system 240 based on a result of the analysis. For example, based on a result of the analysis, a configuration of an associated target computing system 240, e.g., operation parameters and/or computing resource allocation, may be adjusted.

[0057] The processes described above in association with FIGS. 3-5 can be implemented in hardware, software, or a combination thereof. In the context of software, the operations represent computer-executable instructions stored on one or more computer-readable storage media that, when executed by one or more processors, perform the recited operations. Generally, computer-executable instructions include routines, programs, objects, components, data structures, and the like that perform particular functions or implement particular abstract data types. In other embodiments, hardware components perform one or more of the operations. Such hardware components may include or be incorporated into processors, application-specific integrated circuits (ASICs), programmable circuits such as field programmable gate arrays (FPGAs), or in other ways. The order in which the operations are described is not intended to

be construed as a limitation, and any number of the described operations can be combined in any order and/or in parallel to implement the processes.

[0058] The memory may include computer readable media such as a volatile memory, a Random Access Memory (RAM), and/or non-volatile memory, e.g., Read-Only Memory (ROM) or flash RAM, and so on. The memory is an example of a computer readable medium.

[0059] Computer readable media include non-volatile, volatile, mobile and non-mobile media, and can implement information storage through any method or technology. The information may be computer readable instructions, data structures, program modules or other data. Examples of storage media of a computer include, but not limited to, Phase-change RAMs (PRAMs), Static RAMs (SRAMs), Dynamic RAMs (DRAMs), other types of RAMs, ROMs, Electrically Erasable Programmable Read-Only Memories (EEPROMs), flash memories or other memory technologies, Compact Disk Read-Only Memories (CD-ROMs), Digital Versatile Discs (DVDs) or other optical memories, cassettes, cassette and disk memories or other magnetic memory devices or any other non-transmission media, and can be used for storing information accessible to the computation device. According to the definitions herein, the computer readable media exclude transitory media, such as modulated data signals and carriers.

[0060] It should be further noted that, the terms "include", "comprise", or any variants thereof are intended to cover a non-exclusive inclusion, such that a process, a method, a product, or a device that includes a series of elements not only includes such elements but also includes other elements not specified expressly, or may further include inherent elements of the process, method,

product, or device. In the absence of more restrictions, an element limited by "include a/an..." does not exclude other same elements existing in the process, method, product, or device that includes the element.

[0061] Described above are merely the examples of the present application, which are not used to limit the present application. For those skilled in the art, the present application may have various alterations and changes. Any modification, equivalent replacement, improvement, and the like made within the spirit and principle of the present application shall be included in the scope of the claims of the present application.

[0062] The disclosure may be appreciated with the following clauses:

[0063] Clause 1: a computer implemented method, comprising: receiving performance data of a computing system in a system performance domain from multiple sources; generating an attribute name and a measurement unit for the system performance domain; converting the performance data received from the sources using the attribute name and the measurement unit; and enhancing the converted performance data with an interpolation operation using an interpolation scheme determined based on a type of the performance domain.

[0064] Clause 2: the method of clause 1, wherein the generating the measurement unit for the system performance domain includes evaluating a sampling interval in collecting performance data from each of the multiple sources.

[0065] Clause 3: the method of claim 1, wherein the generating the measurement unit for the system performance domain includes selecting a measurement unit that minimizes missing data values in the converting.

[0066] Clause 4: the method of claim 1, wherein the generating the measurement unit for the system performance domain include evaluating mismatching in time scales among system performance data obtained from two or more different sources.

[0067] Clause 5: the method of claim 4, wherein the generating the measurement unit for the system performance domain includes selecting a measurement unit that minimizes the mismatching in time scales among the system performance data from the two or more different sources.

[0068] Clause 6: the method of claim 1, further comprising selecting the multiple sources from a pool of system performance data sources, the selecting being made based at least partially on at least one of a frequency, a time interval, a data metric, and a measurement unit associated with performance data collected in each system performance data source in the pool.

[0069] Clause 7: the method of claim 1, wherein the receiving the performance data includes receiving the performance data on multiple system performance domains, and the enhancing the converted performance data with the interpolation operation includes using a first interpolation scheme for converted performance data in a first system performance domain and a second different interpolation scheme for converted performance data in a second different system performance domain.

[0070] Clause 8: the method of claim 1, wherein the interpolation scheme includes at least one of a zero interpolation, an arithmetic mean substitution, a cubic spline interpolation or a Kalman filter interpolation.

[0071] Clause 9: the method of claim 1, on a condition that the system performance domain includes CPU frequency data, an arithmetic mean substitution interpolation scheme is used for the interpolation operation.

[0072] Clause 10: the method of claim 1, on a condition that the system performance domain includes instantaneous performance data, a Kalman filter interpolation scheme is used for the interpolation operation.

[0073] Clause 11: the method of claim 1, on a condition that the system performance domain includes event log data, a Zero interpolation scheme is used for the interpolation operation.

[0074] Clause 12: the method of claim 1, further comprising generating a unified timescale for one or more system performance domains.

[0075] Clause 13: the method of claim 12, further comprising merging converted performance data in the one or more system performance domains with the unified timescale to generate a unified dataset.

[0076] Clause 14: the method of claim 13, further comprising analyzing a performance of the computing system using the unified dataset.

[0077] Clause 15: the method of claim 1, wherein the enhancing the converted performance data with the interpolation operation includes using a interpolation scheme to impute a missing value of the converted performance data with an estimated value within a confidence interval.

[0078] Clause 16: the method of claim 1, wherein the enhancing the converted performance data with the interpolation operation includes mapping the covered performance data with a unified time scale.

[0079] Clause 17: a method, comprising: receiving multiple datasets; providing a unified time scale; mapping the multiple datasets with the unified time scale; enhancing a mapped dataset in the multiple dataset with an interpolation scheme based on a type of the dataset to generate an enhanced dataset; and merging the mapped multiple datasets including the enhanced dataset into a unified dataset.

[0080] Clause 18: the method of claim 17, wherein the interpolation scheme includes at least one of a zero interpolation, an arithmetic mean substitution, a cubic spline interpolation or a Kalman filter interpolation.

[0081] Clause 19: the method of claim 1, wherein the providing the unified time scale includes selecting a time scale that minimizes data value missing in the mapping.

[0082] Clause 20: a computing device, comprising: one or more processors; memory; and a plurality of programing instructions stored on the memory and executable by the one or more processors to implement: a data receiving unit operable to receive performance data of a computing system in a system performance domain from multiple sources, an attribute name generation unit operable to generate an attribute name and a measurement unit for the system performance domain, a converting unit operable to convert the performance data obtained from the multiple sources using the attribute name and the measurement unit; and a data enhancing unit operable to enhance the converted performance data with an interpolation operation using an interpolation scheme determined based on a type of the performance domain.

CLAIMS

What is claimed is:

1. A computer implemented method, comprising:
receiving performance data of a computing system in a system performance domain from multiple sources;
generating an attribute name and a measurement unit for the system performance domain;
converting the performance data received from the sources using the attribute name and the measurement unit; and
enhancing the converted performance data with an interpolation operation using an interpolation scheme determined based on a type of the performance domain.
2. The method of claim 1, wherein the generating the measurement unit for the system performance domain includes evaluating a sampling interval in collecting performance data from each of the multiple sources.
3. The method of claim 1, wherein the generating the measurement unit for the system performance domain includes selecting a measurement unit that minimizes missing data values in the converting.
4. The method of claim 1, wherein the generating the measurement unit for the system performance domain include evaluating mismatching in time scales among system performance data obtained from two or more different sources of the multiple sources.

5. The method of claim 4, wherein the generating the measurement unit for the system performance domain includes selecting a measurement unit that minimizes the mismatching in time scales among the system performance data from the two or more different sources of the multiple sources.

6. The method of claim 1, further comprising selecting the multiple sources from a pool of system performance data sources, the selecting being made based at least partially on at least one of a frequency, a time interval, a data metric, and a measurement unit associated with performance data collected in each system performance data source in the pool.

7. The method of claim 1, wherein the receiving the performance data includes receiving the performance data on multiple system performance domains, and the enhancing the converted performance data with the interpolation operation includes using a first interpolation scheme for converted performance data in a first system performance domain and a second different interpolation scheme for converted performance data in a second different system performance domain.

8. The method of claim 1, wherein the interpolation scheme includes at least one of a zero interpolation, an arithmetic mean substitution, a cubic spline interpolation or a Kalman filter interpolation.

9. The method of claim 1, on a condition that the system performance domain includes CPU frequency data, an arithmetic mean substitution interpolation scheme is used for the interpolation operation.
10. The method of claim 1, on a condition that the system performance domain includes instantaneous performance data, a Kalman filter interpolation scheme is used for the interpolation operation.
11. The method of claim 1, on a condition that the system performance domain includes event log data, a Zero interpolation scheme is used for the interpolation operation.
12. The method of claim 1, further comprising generating a unified timescale for one or more system performance domains.
13. The method of claim 12, further comprising merging converted performance data in the one or more system performance domains with the unified timescale to generate a unified dataset.
14. The method of claim 13, further comprising analyzing a performance of the computing system using the unified dataset.
15. The method of claim 1, wherein the enhancing the converted performance data with the interpolation operation includes using a

interpolation scheme to impute a missing value of the converted performance data with an estimated value within a confidence interval.

16. The method of claim 1, wherein the enhancing the converted performance data with the interpolation operation includes mapping the covered performance data with a unified time scale.

17. A method, comprising:

- receiving multiple datasets;
- providing a unified time scale;
- mapping the multiple datasets with the unified time scale;
- enhancing a mapped dataset in the multiple dataset with an interpolation scheme based on a type of the dataset to generate an enhanced dataset; and
- merging the mapped multiple datasets including the enhanced dataset into a unified dataset.

18. The method of claim 17, wherein the interpolation scheme includes at least one of a zero interpolation, an arithmetic mean substitution, a cubic spline interpolation or a Kalman filter interpolation.

19. The method of claim 1, wherein the providing the unified time scale includes selecting a time scale that minimizes data value missing in the mapping.

20. A computing device, comprising:
- one or more processors;
 - memory; and
 - a plurality of programming instructions stored on the memory and executable by the one or more processors to implement:
 - a data receiving unit operable to receive performance data of a computing system in a system performance domain from multiple sources,
 - an attribute name generation unit operable to generate an attribute name and a measurement unit for the system performance domain,
 - a converting unit operable to convert the performance data obtained from the multiple sources using the attribute name and the measurement unit;
 - and
 - a data enhancing unit operable to enhance the converted performance data with an interpolation operation using an interpolation scheme determined based on a type of the performance domain.

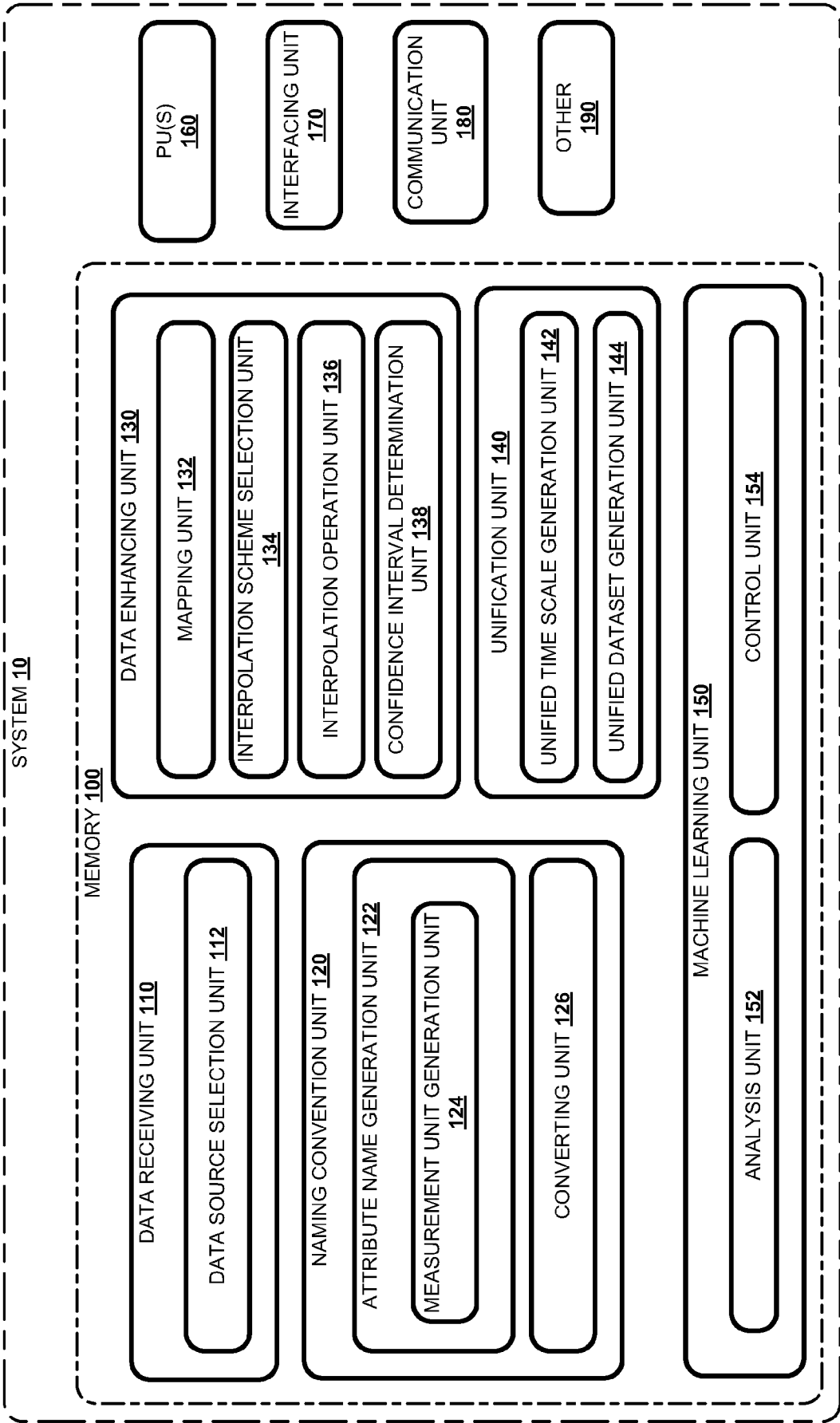
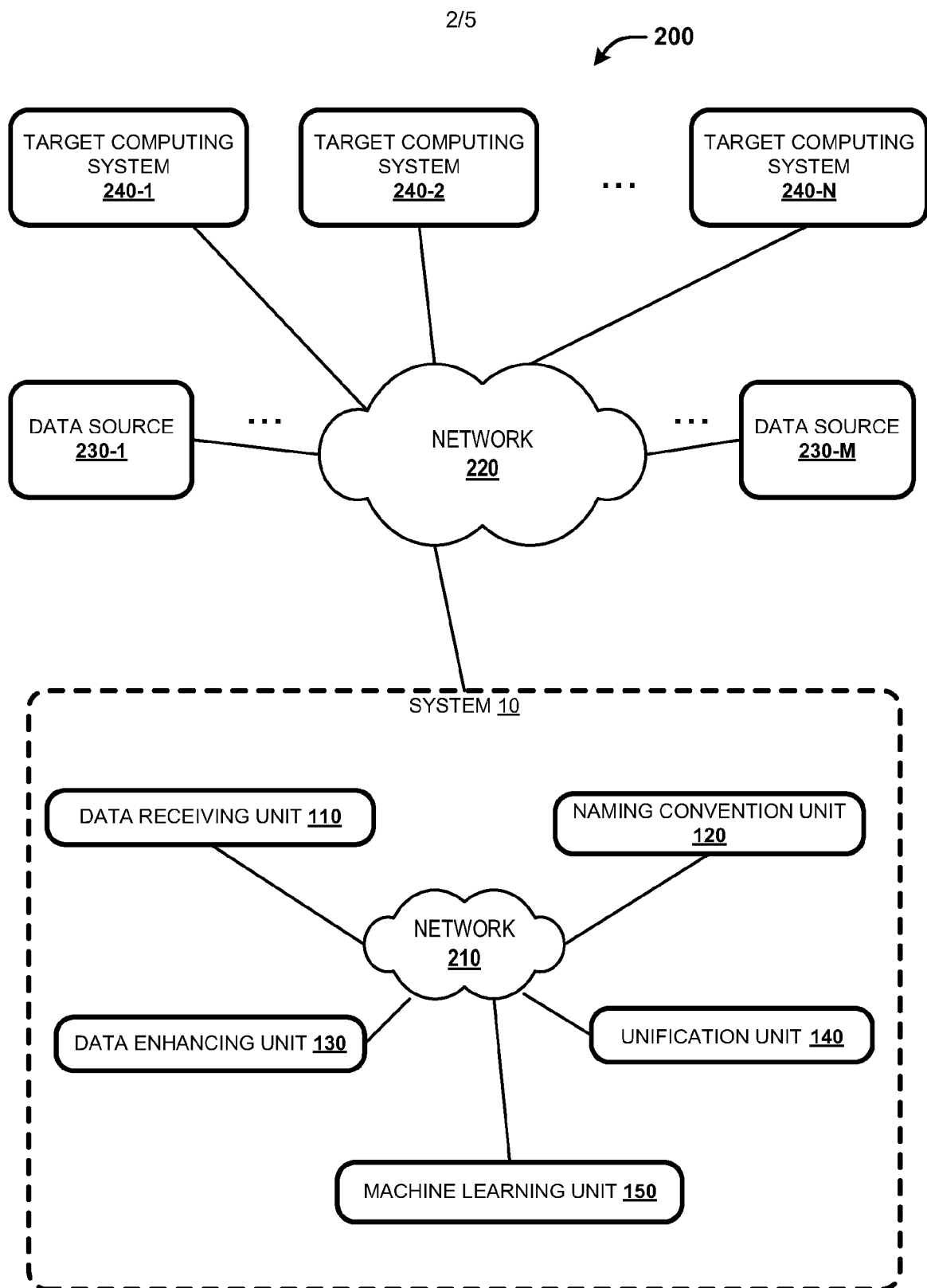
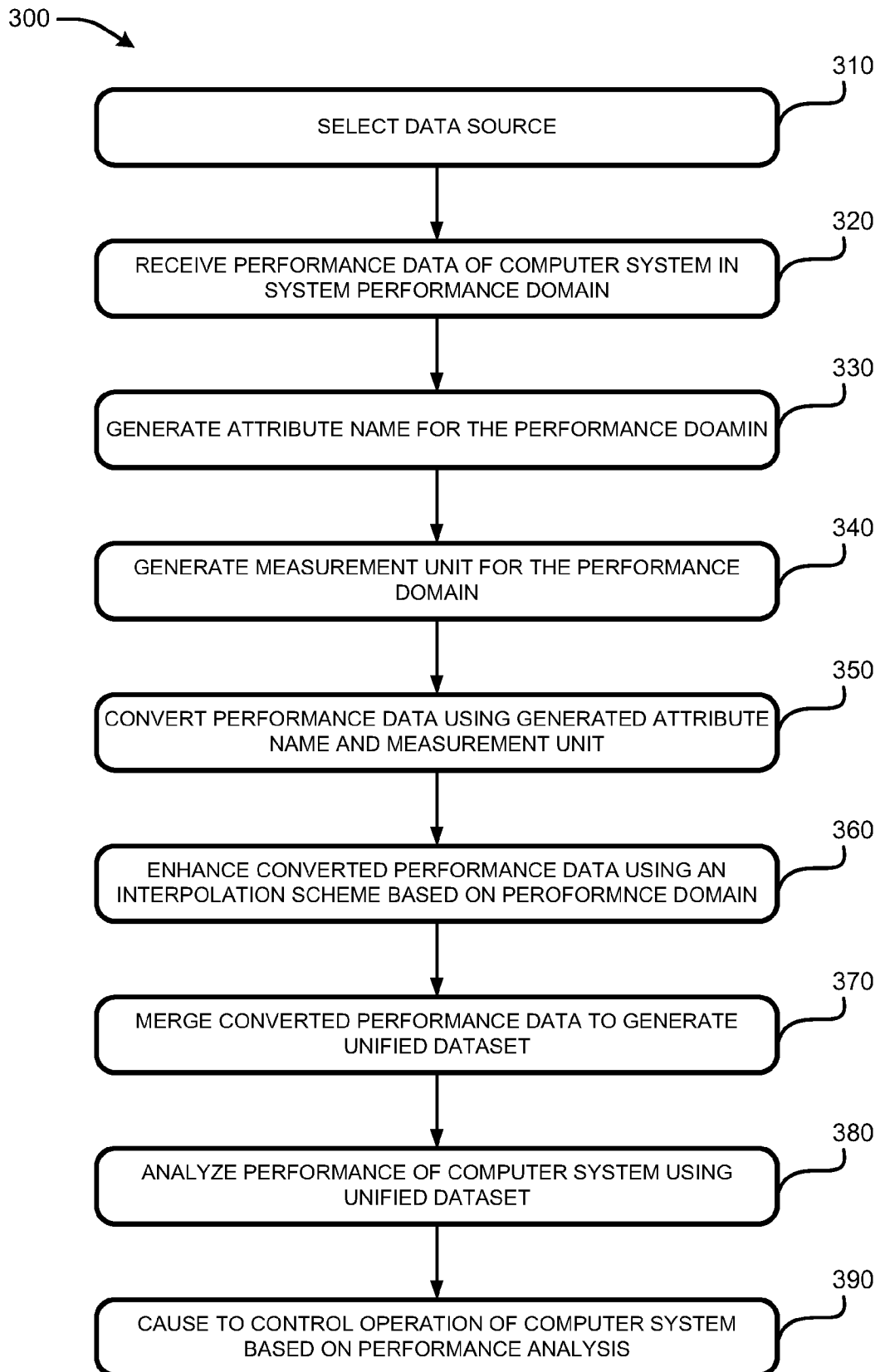


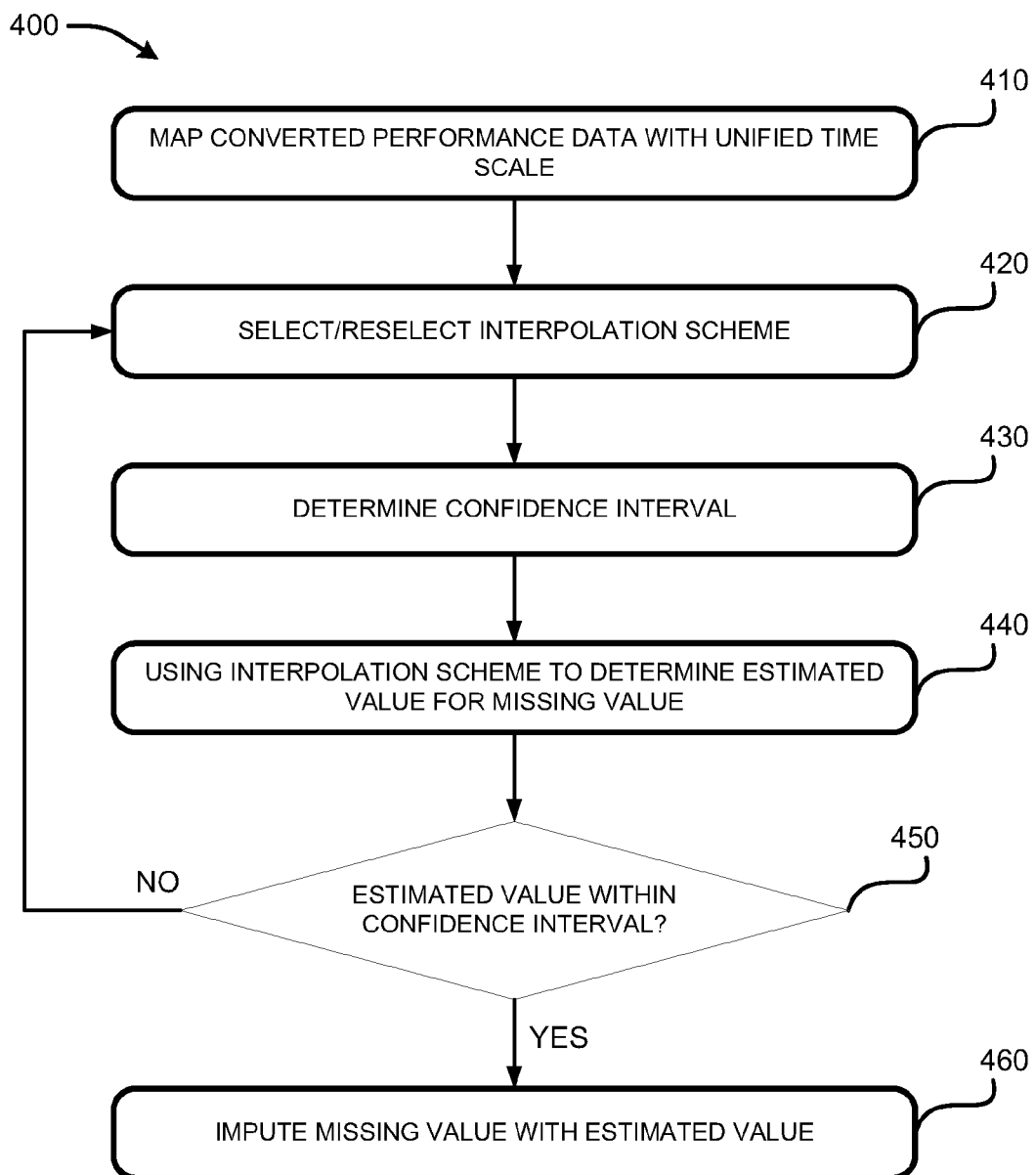
FIG. 1

**FIG. 2**

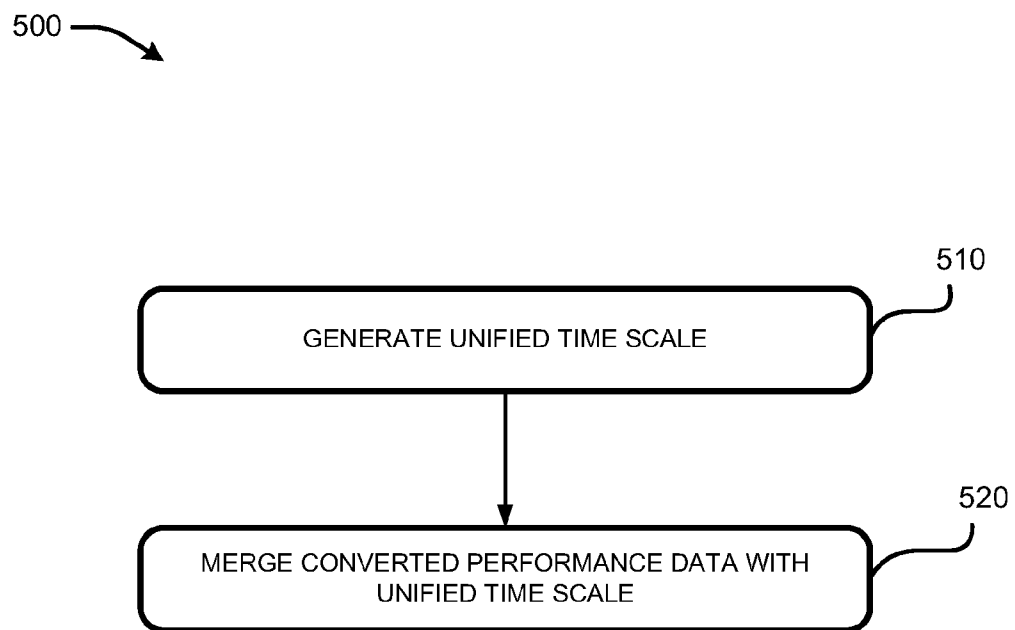
3/5

**FIG. 3**

4/5

**FIG. 4**

5/5

**FIG. 5**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2017/096358

A. CLASSIFICATION OF SUBJECT MATTER

G06F 11/34(2006.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

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

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

CNPAT,WPI,EPODOC,CNKI:performance, data, bandwidth, measur+, unit, interpolat+, type,enhanc+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP 2960797 A1 (INTEL CORP.) 30 December 2015 (2015-12-30) description, paragraphs[0023]-[0024], [0050]-[0051], figure 4	1-20
Y	CN 101790092 A (CHANGZHOU CAMPUS OF HOHAI UNIVERSITY) 28 July 2010 (2010-07-28) description, paragraphs[0072]-[0076]	1-20
A	CN 101533366 A (INSPUR ELECTRONIC INFORMATION INDUSTRY CO., LTD.) 16 September 2009 (2009-09-16) the whole document	1-20
A	US 2006074823 A1 (HEUMANN, JOHN M.) 06 April 2006 (2006-04-06) the whole document	1-20

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

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

03 May 2018

Date of mailing of the international search report

10 May 2018

Name and mailing address of the ISA/CN

STATE INTELLECTUAL PROPERTY OFFICE OF THE
P.R.CHINA
6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing
100088
China

Authorized officer

HU,Lili

Facsimile No. (86-10)62019451

Telephone No. 86-(10)-53961386

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2017/096358

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
EP	2960797	A1	30 December 2015	US	2015378861	A1	31 December 2015
				CN	105224434	A	06 January 2016
CN	101790092	A	28 July 2010	None			
CN	101533366	A	16 September 2009	None			
US	2006074823	A1	06 April 2006	TW	200609798	A	16 March 2006
				CN	1749987	A	22 March 2006
				SG	121074	A1	26 April 2006