(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 January 2017 (05.01.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/003454 A1

(51) International Patent Classification:
G06Q 10/06 (2012.01)

(21) International Application Number:
PCT/US2015/038555

(22) International Filing Date:
30 June 2015 (30.06.2015)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **HEWLETT PACKARD ENTERPRISE DEVELOPMENT LP** [US/US]; 11445 Compaq Center Drive West, Houston, TX 77070 (US).

(72) Inventors: **CHUA, Freddy**; 1501 Page Mill Road, Palo Alto, California 94304 (US). **HUBERMAN, Bernardo**; 1501 Page Mill Road, Palo Alto, California 94304 (US).

(74) Agents: **PAGAR, Preetam B.** et al.; Hewlett Packard Enterprise, 3404 E. Harmony Road, Mail Stop 79, Fort Collins, CO 80528 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i))*

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

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

(54) Title: PARTITIONING A WORKFLOW

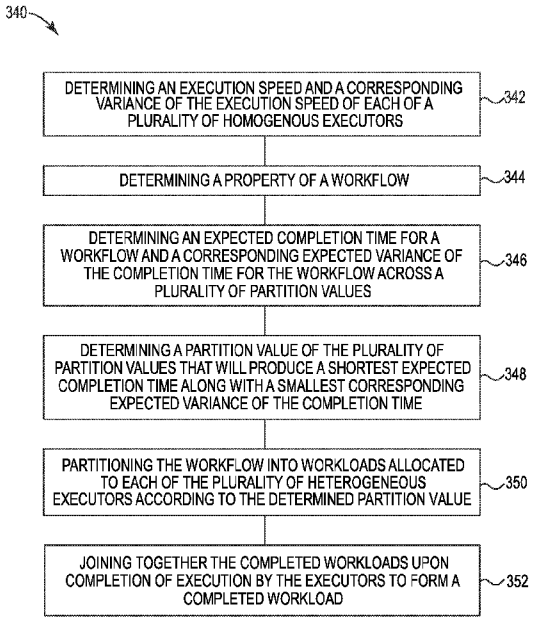


FIG. 3

(57) Abstract: Example implementations relate to partitioning a workflow. For example, a system can include a workflow engine to determine a partition of a workflow into a plurality of workloads, wherein the determined partition reduces an expected workflow completion time and reduced an expected variance corresponding to the expected workflow completion time. The system can include a partition engine to partition the workflow into workloads according to the determined partition.

WO 2017/003454 A1

PARTITIONING A WORKFLOW

Background

[0001] Completion of a workflow can include a defined series of tasks or operations upon inputs which may be executed by an executor. In some instances, rapid completion of a workflow may be targeted. To accelerate the completion of large workflows, more efficient completion algorithms and/or faster execution speeds of an executor may be utilized. However, these options may be costly and may be limited by the capabilities of available executors.

[0002] Workflows may also be partitioned into smaller workloads. Partitioning of workloads may accelerate the completion of the full workflow. For example, parallel execution of workloads may cause a more rapid completion of a workflow than a serial execution of the workloads. This is the assumption underlying the parallelization of large processing jobs, such as the use of map-reduce for indexing documents, parallel algorithms for machine learning, decentralization of load balancing in networks, and many other complex processes. Parallel executions may assume that that the executors are homogenous. An assumption of homogenous executors may produce partitioning of the workflow into equal sized workloads among the available executors.

[0003] Executors may have different capabilities. For example, in a large data processing center that undergoes multiple iterations of hardware upgrades, the processing capabilities of executor machines may deviate and become non-identical. In such an example, the parallel execution of an equally sized workload across executors with different capabilities may limit the completion time of the total workflow to the longest completion time associated with the slowest executor.

Brief Description of the Drawings

[0004] Figure 1 illustrates an example system, according to the disclosure;

[0005] Figure 2 illustrates an example of a computing device, according to the disclosure;

[0006] Figure 3 illustrates a flow chart of an example of a method, according to the disclosure; and

[0007] Figure 4 illustrates a flow chart of an example of a method, according to the disclosure.

Detailed Description

[0008] A workflow can be any job, tasks, and/or set of operations. An example of a workflow can be a large file to be transmitted over the Internet, a large print job capable of being executed on more than one printer, a manufacturing process capable of being broken into different streams, traffic capable of being routed over more than one route, and/or a number of other job, tasks, and/or set of operations capable of execution on a plurality of executors.

[0009] An executor can be any mechanism capable of executing the workflow or a workload portion thereof. For example, an executor can be a processor of a computing device, a virtual machine, a printer, a manufacturing stream, a traffic route, and/or a number of other mechanisms by which to execute a portion of a workflow.

[0010] An executor can have a property associated therewith. For example, the executor can have an associated execution speed. An execution speed can be a rate of performing the workflow and/or a workload portion thereof. For example, an execution speed can be a processing speed of a processor of a computing device, an operations rate of a virtual machine, a page printing rate of a printer, a widget producing rate of a manufacturing stream, a traffic capacity and/or traffic speed associated with a traffic route, and/or any other metric related to the speed at which an executor is capable of executing a workflow.

[0011] Executors can be heterogeneous. Executors can have different properties, such as execution speed, from one another. Additionally, each executor can possess a distinct amount of uncertainty or variance associated with its execution speed. This variance can be a result of fluctuations in the properties

and/or characteristics of the executor and/or demands placed on the executor and/or its resources.

[0012] Partitioning a workflow into workloads allocated among heterogeneous executors can be accomplished through equal partitioning of the workflow. However, this can lead to a workflow only reaching completion upon completion by the slowest executor of its workload portion of the workflow.

[0013] In contrast, examples herein relate to determining an expected completion time for a workflow and an expected corresponding variance of the completion time for the workflow associated with each of a number of partitions of the workflow among a number of executors each executing a respective workload portion of the workflow in parallel. Then, partitioning the workflow into workloads allocated among the number of executors based on the determined expected completion time for the workflow and the determined expected corresponding variance of the completion time for the workflow associated with each of the number of partitions of the workflow. That is, in contrast to assuming homogeneity among executors and equally dividing a workflow into equally sized workloads for executors, examples herein relate to partitioning the workflow into workloads based on expected completion times and corresponding variance thereof associated with the partition.

[0014] The parallelization procedure examples herein include a decision on how to partition a workflow so that its full completion process takes the shortest amount of time to complete with a relatively minimal amount of uncertainty and/or variance in the completion time. Uncertainty and/or variance is a relevant variable because, for example, unavoidable fluctuations in executing a workload that each executor (e.g., processing unit, channel, virtual machine, etc.) undergoes when having to time share with other processes and/or workflows. This can introduce a stochastic component into the execution of any workflow, which can at times increase the amount of time it takes for a given workflow and/or workload portion of a workflow to complete. Therefore, incorporating these fluctuations into a partitioning procedure can result in shorter times to completion than an original non-partitioned workflow.

[0015] The parallelization procedure can include partitioning a workflow into a plurality of workloads in such a way that their outputs are joined and their full completion time takes less time and uncertainty and/or variance than when running it

in one executor. The parallelization procedure can be implemented to effect a partitioning for joint parallel execution of workload portions of a workflow with an implication that the overall execution time of the workflow is determined by the longest running executor.

[0016] Figure 1 illustrates an example of a workflow partition system 100, according to examples of the present disclosure. As illustrated in Figure 1, the workflow partition system can include a database 102, a partition manager 104, a number of engines 106 and 108. The partition manager 104 can include the number of engines 106 and 108 in communication with the database 102 via a communication link. The system can include additional or fewer engines than illustrated to perform the various tasks described herein. .

[0017] The number of engines 106 and 108 can include a combination of hardware and instructions (e.g. programming) which are executed by the hardware to perform tasks including those described herein (e.g., determine a partition of a workflow into a plurality of workloads, wherein the determined partition reduces an expected workflow completion time and reduced an expected variance corresponding to the expected workflow completion time, etc.). The instructions can include instructions (e.g., software (e.g., instructions that can be executed by hardware such as a processor), hardware with instructions, etc.) stored in a memory resource (e.g., computer readable medium (CRM), machine readable medium (MRM), etc.) as well as hard-wired program (e.g., logic).

[0018] The workflow engine 106 can include hardware and/or a combination of hardware and instructions (e.g., programming) which are executed by the hardware to perform tasks to determine a partition of a workflow into a plurality of workloads, wherein the determined partition reduces an expected workflow completion time and reduced an expected variance corresponding to the expected workflow completion time. That is, the workflow engine 106 can include hardware and/or a combination of hardware and instructions (e.g., programming) to perform tasks to measure statistical properties of a workflow allocation system and determine a partition of the workflow based on the measure of statistical properties such that an expected workflow completion time and an expected variance corresponding to the expected workflow completion time are reduced.

[0019] Determining the partition can include determining an expected workflow completion time associated with each of a plurality of partitions of the

workflow among a plurality of parallel executors of the workflow. That is, determining the expected workflow completion time can be based on a rate of performing a workload and a property of the workload. The expected completion time for the workflow can be determined based on the rate of performing the workload and a property of the workload. The rate of performing the workflow can include an execution speed associated with one or more of the executors. The execution speed can be a known stored value and/or can be measured across repeated trials of a partitioning process conducted over a period of time utilizing various partitions. The property of the workload can include a metric of how intensive the workload will be on an executor/executors. A metric of how intensive the workload will be can include a size of the workflow/workload, a time associated with executing the workflow/workload, a complexity of the workflow/workload, an amount of resource demand associated with executing the workflow/workload, an amount of operations involved in executing the workflow/workload. The expected completion time can additionally be determined based on the expected corresponding variance of the completion time for the workflow.

[0020] Determining the partition can further include determining an expected variance corresponding to each expected workflow completion time. That is, determining the expected workflow completion time and an expected corresponding variance of the completion time can include numerically deriving values of the expected completion time for a workflow and an expected corresponding variance of the completion time for each of a number of potential partitions of the workflow. The expected workflow completion time and an expected corresponding variance of the workflow completion time can be determined based on the fluctuations in an execution speed associated with each of the plurality of parallel executors. Parallel executors can include executors that separately perform the joint execution of workload portions of a workflow in a parallel. The parallel executors can be heterogeneous in that they each are associated with distinct execution speeds and fluctuations of those execution speeds brought on by characteristics of the executor, characteristics properties of the executor resources, and/or competition/depletion of executor resources, among other demands. The expected variance of the workflow completion time can be a numerical categorization of the fluctuation in the workflow and/or workload completion time brought on by the distinct fluctuations associated

with utilizing particular executors to execute particular workload portions of the workflow.

[0021] As described above, the expected variance corresponding to each expected workflow completion time can be determined based on fluctuations in the execution speed associated with the each number of executors executing a respective workload portion of the workflow in parallel. These fluctuations can be stored as fluctuation values associated with each executor and/or be measured through repeated trials of a partitioning process conducted over a period of time utilizing various partitions and/or measured by inducing competition for an executor resource during test trials.

[0022] Further, the numerical values for the expected workflow completion time and an expected corresponding variance of the workflow completion time can be derived according to a model of the executors of the workflow and how they can execute a workflow based on their properties. For example, a workflow *D*, can be partitioned into two workloads D_i and D_j (while this example only characterizes the partitioning of the workflow into two workloads to be allocated across two executors the examples included herein are not so limited are equally applicable to additional partitions and additional executors), each of which can be executed on different executors *i* and *j*, with different execution speeds and fluctuating performance parameters. Once the slowest workload has been completed, the two outputs from execution of the workloads can be joined together and the workflow can be considered complete.

[0023] It can be assumed that the completion time, t_i for the full workflow *D* executing on executor *i* is a continuous variable which is Normally distributed with mean μ_i and standard deviation σ_i,

$$p(t_i | D, \mu_i, \sigma_i) \sim N(\mu_i, \sigma_i^2) \qquad \textbf{(Eq. 1.1)}$$

[0024] If the workload on executor *i* , D_i, is smaller than *D* by a factor of *f*, i.e., $|D_i| = f|D|$, the resulting distribution of completion times t_i for executor *i* can be given by,

$$p(t_i | D_i, \mu_i, \sigma_i) \sim N(f\mu_i, [f\sigma_i]^2) \qquad \textbf{(Eq. 1.2)}$$

and similarly for executor *j* that executes workload D_j so that $|D_j| = (1\text{-}f)|D|$,

$$p(t_j | D_j, \mu_j, \sigma_j) \sim N([1-f]\mu_j, [(1-f)\sigma_j]^2) \quad \textbf{(Eq. 1.3)}$$

[0025] The workflow can be complete when both executors *i* and *j* finish executing their assigned workloads D_1 and D_j, respectively. Thus, the cumulative density function for the completion time *t* can be the probability that both workloads t_i and t_j complete within a time ε.

$$p(\mathrm{t} \le \varepsilon \| f, D, \mu_i, \sigma_i, \mu_j, \sigma_j) = p(\mathrm{t}_i \le \varepsilon \| D_i, \mu_i, \sigma_i) \cdot p(\mathrm{t}_j \le \varepsilon \| D_j, \mu_j, \sigma_j) \quad \textbf{(Eq. 1.4)}$$

[0026] The decision as to how to partition the workflow can include choosing the value of *f* such that the workflow will execute with a lowest expected completion time μ(*f*) and variance $\sigma^{2}(f)$ relative to the other expected completion times and variances. This can include understanding the behavior of μ and variance σ^2 as a function of *f*. μ and variance σ^2 can be derived from their probability density function as,

$$\mu(f) = E(t|\Theta) = \int_0^\infty t \cdot p(t|\Theta) dt \quad \Theta$$
$$\sigma^2(f) = Var(t|\Theta) = \int_0^\infty t^2 \cdot p(t|\Theta) dt - [E(t|\Theta)]^2 \quad \textbf{(Eq. 1.5)}$$

with

$$\Theta \equiv \{f, D, \mu_i, \sigma_i, \mu_j, \sigma_j\} \quad \textbf{(Eq. 1.6)}$$

with the probability density function given by the first order derivative of the cumulative density function shown in Equation (Eq.) 1.4,

$$p(t|\Theta) = \frac{d}{dt} P(t' \le t|\Theta) \quad \textbf{(Eq. 1.7)}$$

[0027] However, since there may not be a closed form solution for the probability density function, the expected completion time μ(f) can be expressed in terms of the cumulative density function shown in Equation Eq. 1.4,

$$\mu(f) = \int_0^\infty 1 - P(t \le \varepsilon | \Theta) d\varepsilon \quad \textbf{(Eq. 1.8)}$$

[0028] Similarly, the variance $\sigma^2(f)$ can be given by,

$$\sigma^2(f) = \{2\int_0^\infty \varepsilon[1 - P(t - \varepsilon | \Theta)] d\varepsilon\} - [\mu(f)]^2 \quad \textbf{(Eq. 1.9)}$$

[0029] Utilizing the above stated models, the expected workflow completion time associated with each of a plurality of partitions of the workflow among the plurality of parallel executors of the workflow and an expected variance corresponding to each workflow completion time, can be determined. That is, the

expected workflow completion time and an expected variance corresponding to each workflow completion time can each be calculated as a function of a plurality of partition values *f* of the workflow among the plurality of parallel executors. In this manner, a numerical value for the expected workflow completion time and an expected variance corresponding to each workflow completion time can be derived corresponding to each of a plurality of possible partitionings.

[0030] The partition engine 108 can include hardware and/or a combination of hardware and instructions which are executed by the hardware to partition the workflow into workloads according to the determined priority. The workloads can be workloads allocated among the plurality of executors. The partition can be based on the determined partition which can be based on the determined expected completion time associated with each of the plurality of partitions of the workflow and the determined expected variance corresponding to each expected work flow completion time. For example, utilizing the above stated models the expected completion time $\mu(f)$ and expected corresponding variance of the completion time for the workflow $\sigma^2(f)$ can be numerically derived for different values of *f* representing various potential partitions of the workflow which can be graphed as μ with respect to each value of *f* and σ^2 with respect to each value of *f*. An efficient frontier can be identified representing range of partition values of *f* which provide a targeted combination of μ and σ^2 values. For example, the efficient frontier can be a target value can be a minima of μ and σ^2 on these graphs, but since the value of *f* that provides a minimum point of μ and σ^2 can be different, then the efficient frontier can include the range of values of *f* which include at least both of these minima. This can provide a range of choices of workload size for partitioning the workflow to achieve the target combination of μ and σ^2 values for the workflow.

[0031] The efficient frontier can additionally be identified in a graph where μ and σ^2 are graphed parametrically as a function of each other resulting in a parabolic curve. The statistical distribution assumptions of the completion times for the two parallel workloads characterized in the above stated model can permit selection of an appropriate value of *f* which minimizes μ and σ^2 for the full workflow execution.

[0032] Therefore, determining the partition can include determining a partition of the workflow into workloads by selecting a particular partition value (e.g., *f* which minimizes both μ and σ^2 relative to other values of *f*) from a range of partition values, of the number of partitions, corresponding to a minima for the determined completion

time for the workflow and a minima the determined variance of the completion time (e.g., the efficient frontier and/or the range of values including the minima of μ and σ^2).

[0033] Further examples of the application of the above stated workflow partition system 100 can include a parallel partitioning of a least squares error function which is quadratic and therefore convex. Although a map-reduce application can be applicable to such a partitioning, it does not produce the targeted solution. The input *D* to the convex function, in this example, can be partitioned into two workloads D_i and D_j. A classical operation can then be applied to each of the workloads D_i and D_j to obtain a globally optimized solution θ_i and θ_j based on each of their inputs. The solution θ for the original workflow *D* can be obtained as a linear combination of the solution from each of the workloads.

$$\Theta = f \quad \Theta \quad (1 \quad f) \quad \Theta \quad \textbf{(Eq. 1.10)}$$

[0034] The parallel global algorithm can be processed on two executors (e.g., virtual machines, each one with a CPU core). The mean completion times μ and variance σ^2 can be obtained at each partition value of *f* by repeating a plurality of trials of the process over a period of time using different partitions (values of *f*). The mean completion times μ and variance σ^2 can be graphed as a function of *f* and a performance curve similar to the performance curve derived in the above examples can be obtained. A curve can be fitted to the observed behavior of mean completion times μ and variance σ^2 across various partition values *f* to identify an efficient frontier and/or a partition value *f* to achieve a target mean completion time μ and variance σ^2.

[0035] Another example of the application of the above stated workflow partition system 100 can include performing a file transmission of a fixed size file in parallel from a source node to a destination node over two network paths. Since the TCP network protocol may not allow fine grain control of how file packets travel through the Internet, an intermediate overlay can be created to redirect a fraction of the file packets through a different path. In this example, the source node can be hosted in New York City, while the destination node can be hosted in Singapore. A traceroute can be utilized to demonstrate that network packets can be transmitted through the west coast of the United States before reaching the destination in

Singapore, which implies that the network packet from New York City route through the Pacific Ocean to Singapore.

[0036] In order to determine if utilizing an alternate route of sending file packets from New York through the Atlantic Ocean using Europe as the overlay before finally reaching Singapore is a more efficient network channel, another host can be created in London to act as an overlay to receive file packets from New York City and forward them to Singapore. A file can be partitioned into workloads based on different values of *f* and sent across the two different networks (e.g., Atlantic Ocean, Pacific Ocean). Disk I/O delays can be excluded to ensure that network transmissions times' contribution to completion times are analyzed alone.

[0037] To measure the completion time of the two parallel file transfers, the node at the destination can measure the time of the last packet (from either channel) and then subtract the time of the request for the first packet (e.g., from both channels).

[0038] In order to measure the mean completion times μ and variance σ^2 thereof, the file transfer can be repeated a number of times over a time period. For each trial, a randomized value of *f* can be used. The results obtained for this file transmission example can be consistent with the performance curve derived in the above examples.

[0039] Figure 2 illustrates an example computing device 220 according to examples of the disclosure. The computing device 220 can utilize instructions (software and/or hardware with instructions) hardware, and/or logic to perform a number of tasks including those described herein. The computing device 220 can be a combination of hardware and program instructions to share information. The hardware, for example, can include a processing resource 222 and/or a memory resource 224 (e.g., CRM, MRM, database, etc.).

[0040] A processing resource 222, as used herein, can include a processor capable of executing instructions stored by a memory resource 224. Processing resource 222 can be implemented in a single device or distributed across multiple devices. The program instructions (e.g., computer readable instructions (CRI)) can include instructions stored on the memory resource 224 and executable by the processing resource 222 to implement a desired task (e.g., derive an expected completion time of a workflow and a corresponding expected variance of the

completion time of the workflow across a plurality of workflow partitions among a plurality of heterogeneous executors, etc.).

[0041] The memory resource 224 can be in communication with the processing resource 222 via a communication link (e.g., a path) 226. The communication link 226 can be local or remote to a machine (e.g., a computing device) associated with the processing resource 222. Examples of a local communication link 226 can include an electronic bus internal to a machine (e.g., a computing device) where the memory resource 224 is one of volatile, non-volatile, fixed, and/or removable storage medium in communication with the processing resource 222 via the electronic bus.

[0042] A module and/or modules 228, 230, and 232 can include CRI that when executed by the processing resource 222 can perform a number of tasks including those described herein. A number of modules 228, 230, and 232 can be sub-modules of other modules. For example, the derive module 228 and the determine module 230 can be sub-modules and/or contained within the same computing device. In another example, the number of modules 228, 230, and 232 can comprise individual modules at separate and distinct locations (e.g., CRM, etc.).

[0043] Each of the number of modules 228, 230, and 232 can include instructions that when executed by the processing resource 222 can operate as a corresponding engine described herein. For example, the derive module 228 and determine module 230 can include instructions that when executed by the processing resource 222 can operate as the workflow engine 106. The partition module 232 can include instructions that when executed by the processing resource 222 can operate as the partition engine 108.

[0044] The derive module 228 can include instructions that when executed by the processing resource 222 can derive an expected completion time of a workflow and a corresponding expected variance of the completion time of the workflow across a plurality of workflow partitions among a plurality of heterogeneous executors. The expected completion time of a workflow and the corresponding expected variance of the completion time of the workflow can be derived based on a size of the workflow, a portion of the workflow to be allocated to a particular executor, and an execution speed of the particular executor. That is, the expected completion time and expected variance can be derived by stored values of a size of a workflow, a workload portion of the workflow to be allocated to a particular

executor, and an execution speed associated with the particular executor and/or values of a size of a workflow, a workload portion of the workflow to be allocated to a particular executor, and/or an execution speed associated with the particular executor obtained from trials with the executors handling various workloads. The expected completion time of the workflow can be determined by a derived amount of time involved in totally executing a workflow obtained from times involved in executing partitioned workloads of a given size on an executor with a given execution speed, and a given variance of that execution speed. The derived expected completion time and expected variance thereof can be expressed as a function of the partition (*f*) at various values. The executors can be heterogeneous in that they each are associated with a distinct execution speed and/or a distinct variance thereof.

[0045] The determine module 230 can include instructions that when executed by the processing resource 222 can determine a number of workflow partitions of the plurality of workflow partitions that achieve a target expected completion time of the workflow and a corresponding target expected variance of the completion time of the workflow. That is, the determine module 230 can include instructions that when executed by the processing resource 222 can determine the number of workflow partitions of the plurality of workflow partitions that achieve the target values by application of the models described above in more detail with regard to Figure 1. That is, the determine module 230 can include instructions that when executed by the processing resource 222 can compare the derived expected completion time of a workflow and the corresponding expected variance of the completion time of the workflow across a plurality of workflow partitions with a target value for each. For example, this can include deriving the expected completion time of a workflow and the corresponding expected variance of the completion time of the workflow as a function of the partition value *f* and identifying a target range of values of partition values *f* including the minima of the expected completion time of a workflow as a function of *f* and/or the minima of the corresponding expected variance of the completion time of the workflow as a function of *f*.

[0046] The partition module 232 can include instructions that when executed by the processing resource 222 can partition the workflow into workloads to be executed in parallel by each of the number of executors according to a selected partition of the determined number of partitions. Partitioning the workflow can

include splitting the workflow into workloads to be executed in parallel by each of the number of executors. The partition can be a partition value *f* selected from the target range of values of partition values *f* including the minima of the expected completion time of a workflow as a function of *f* and/or the minima of the corresponding expected variance of the completion time of the workflow as a function of *f*. The selected partition value *f* can be the partition that produces a shortest expected completion time with a smallest corresponding expected variance relative to the other derived expected completion times and corresponding expected variances of the completion time of other workflow partitions of the plurality of workflow partitions.

[0047] The partition module 232 can further include instructions that when executed by the processing resource 222 can piece completely executed workloads together to output a completed workflow.

[0048] Figure 3 illustrates a flowchart of an example method 340 according to examples of the disclosure. The example method 340 can be implemented using the system 100 illustrate in Figure 1 and/or the computing device 220 illustrated in Figure 2.

[0049] At 342, the method 340 can include determining an execution speed of a plurality of heterogeneous executors and/or a corresponding variance thereof. Determining the execution speed and/or corresponding variance thereof can be accomplished by obtaining corresponding measurements throughout trials of the executor and/or from stored values associated with the executor.

[0050] At 344, the method 340 can include determining a property of the workflow. Determining a property of the workflow can include determining a size, technical complexity, and/or resource demand of a particular workflow. Determining a property of the workflow can also include determining how a workflow may be partitioned into workloads and/or compatible executors to execute the workload.

[0051] At 346, the method 340 can include determining an expected completion time of a workflow and an expected variance of the completion time of the workflow across a plurality of partition values. That is, the method 340 can include numerically deriving an expected completion time of a workflow and an expected variance of the completion time of the workflow as a function of various partition values *f* corresponding to potential partitions of the workflow into workloads to be allocated to the plurality of heterogeneous executors.

[0052] At 348, the method 340 can include determining a partition value of the plurality of partition values that will produce a shortest expected completion time along with a smallest expected variance of completion time as compared to the others associated with the plurality of partition values. The determined partition value can be a range of values including the minima of the expected completion time and the expected variance of completion time as a function of various partition values *f*. Alternatively, the partition value can be a single value of *f* corresponding to the shortest expected completion time along with a smallest expected variance of completion time.

[0053] At 350, the method 340 can include partitioning the workflow into workloads allocated to the plurality of heterogeneous executors according to the determined partition value from 348 of the method 340.

[0054] At 352, the method 340 can include joining together the completed workloads upon completion by the executors to form a completed workflow.

[0055] Figure 4 illustrates a flowchart of an example method 480 according to examples of the disclosure. The example method 480 can be implemented using the system 100 illustrated in figure 1 and/or the computing device 220 illustrated in figure 2.

[0056] At 482, the method 480 can include determining an expected completion time and a corresponding expected variance of the completion time of an execution of a portion of a workflow for each of a plurality of heterogeneous executors based on a corresponding execution speed of the executor and a property of the portion of the workflow. The execution speed of the executor can be determined based on estimates of the execution speed collected during deployment of the executor in executing workloads.

[0057] At 484, the method 480 can include determining a partition of the workflow into workloads to be allocated among the plurality of executors that achieves a workflow completion time under a threshold completion time. The determined partition of the workflow can be a partition of the workflow into workloads allocated to executors that produces a workflow completion time under a threshold completion time wherein the threshold completion time is a predetermined threshold completion times based on a Quality-of-Service level agreement. For example, the determined expected completion time and a corresponding determined expected variance of the completion time of an execution of a portion of a workflow for each of

a plurality of heterogeneous executors can be utilized as the basis for formulating a pricing model for different levels of Quality-of-Service. The determined expected completion times and corresponding determined expected variances of the completion time of an execution of a portion of a workflow for each of a plurality of heterogeneous executors can be segmented into different pricing structures which a customer may select from. A selection of a particular pricing structure can correspond to a particular Quality-of-Service level agreement. A threshold completion time and/or a threshold variance of the completion time can be indicated by the selected pricing structure. Accordingly, the determined partition of the workflow can be a partition that has an expected completion time and/or a corresponding expected variance of the completion time that is at least under the threshold completion time and/or a threshold variance of the completion time specified in the Quality-of-Service level agreement associated with the pricing structure.

[0058] At 486, the method 480 can include allocating the workloads to each of the plurality of executors according to the determined partition. Allocating the workloads can include splitting the workflow into workloads determined by the determined partition value and disseminating the workloads to the executors for execution.

[0059] In the disclosure, reference is made to the accompanying drawings that form a part hereof, and in which is shown by way of illustration how a number of examples of the disclosure can be practiced. These examples are described in sufficient detail to enable those of ordinary skill in the art to practice the examples of this disclosure, and it is to be understood that other examples can be used and that process, electrical, and/or structural changes can be made without departing from the scope of the disclosure.

[0060] The figures herein follow a numbering convention in which the first digit corresponds to the drawing figure number and the remaining digits identify an element or component in the drawing. Elements shown in the various figures herein can be added, exchanged, and/or eliminated so as to provide a number of additional examples of the disclosure. In addition, the proportion and the relative scale of the elements provided in the figures are intended to illustrate the examples of the disclosure, and should not be taken in a limiting sense.

[0061] As used herein, “logic” is an alternative or additional processing resource to perform a particular action and/or function, etc., described herein, which includes hardware, e.g., various forms of transistor logic, application specific integrated circuits (ASICs), etc., as opposed to computer executable instructions, e.g., software firmware, etc., stored in memory and executable by a processor. Further, as used herein, “a” or “a number of” something can refer to one or more such things. For example, “a number of widgets” can refer to one or more widgets. Also, as used herein, “a plurality of” something can refer to more than one of such things.

[0062] The above specification, examples and data provide a description of the method and applications, and use of the system and method of the disclosure. Since many examples can be made without departing from the spirit and scope of the system and method of the disclosure, this specification merely sets forth some of the many possible configurations and implementations.

What is claimed is:

1. A system, comprising:

a workflow engine to determine a partition of a workflow into a plurality of workloads, wherein the determined partition reduces an expected workflow completion time and reduced an expected variance corresponding to the expected workflow completion time; and

a partitioning engine to partition the workflow into workloads according to the determined partition.

2. The system of claim 1, wherein each of the plurality of parallel executors is associated with an execution speed.

3. The system of claim 2, wherein the execution speeds of each of the plurality of parallel executors are heterogeneous across the plurality of parallel executors.

4. The system of claim 1, wherein to determine the partition includes to determine an expected workflow completion time associated with each of a plurality of partitions of the workflow among the plurality of parallel executors of the workflow.

5. The system of claim 4, wherein the expected workflow completion time for the workflow is determined based on the rate of performing the workload and a metric of how intensive execution of the workload will be on a particular executor of the plurality of executors.

6. The system of claim 4, wherein to determine the partition includes to determine an expected variance corresponding to each expected workflow completion time.

7. The system of claim 6, wherein to determine the partition includes to select a particular partition from a range of partitions that includes a partition producing a minima for the determined expected completion time and a partition producing a minima for the determined expected variance corresponding to each expected workflow completion time.

8. The system of claim 6, wherein the expected variance corresponding to each expected workflow completion time is determined based on a fluctuation in an execution speed associated with each of the plurality of parallel executors executing a respective portion of the workflow.

9. A non-transitory computer readable medium storing instructions executable by a processing resource to:

derive an expected completion time of a workflow and a corresponding expected variance of the completion time of the workflow across a plurality of workflow partitions among a plurality of heterogeneous executors;

determine a number of workflow partitions of the plurality of workflow partitions that achieve a target expected completion time of the workflow and a corresponding target expected variance of the completion time of the workflow; and

partition the workflow into workloads to be executed in parallel by each of the number of executors according to a selected partition of the determined number of partitions.

10. The medium of claim 9, wherein the selected partition value is a partition value that produces a shortest expected completion time with a smallest corresponding expected variance relative to the derived expected completion times and corresponding expected variances of the completion time of other workflow partitions of the plurality of workflow partitions.

11. The medium of claim 9, further comprising instructions executable by a processing resource to piece the executed workloads together to output a completed workflow.

12. The medium of claim 9, wherein an expected completion time and a corresponding expected variance of the completion time of the workflow are derived based on a size of the workflow, a portion of the workflow to be allocated to a particular executor, and an execution speed of the particular executor.

13. A method, comprising:

determining an expected completion time and a corresponding expected variance of the completion time of an execution of a portion of a workflow for each of a plurality of heterogeneous executors based on a corresponding execution speed of the executor and a property of the portion of the workflow;

determining a partition of the workflow into workloads to be allocated among the plurality of executors that achieves a workflow completion time under a threshold completion time; and

allocating the workloads to each of the plurality of executors according to the determined partition.

14. The method of claim 13, further comprising determining the execution speed of the executor based on estimates collected during deployment of the executor.

15. The method of claim 13, further comprising determining the threshold completion time based on a selected Quality-of-Service level.

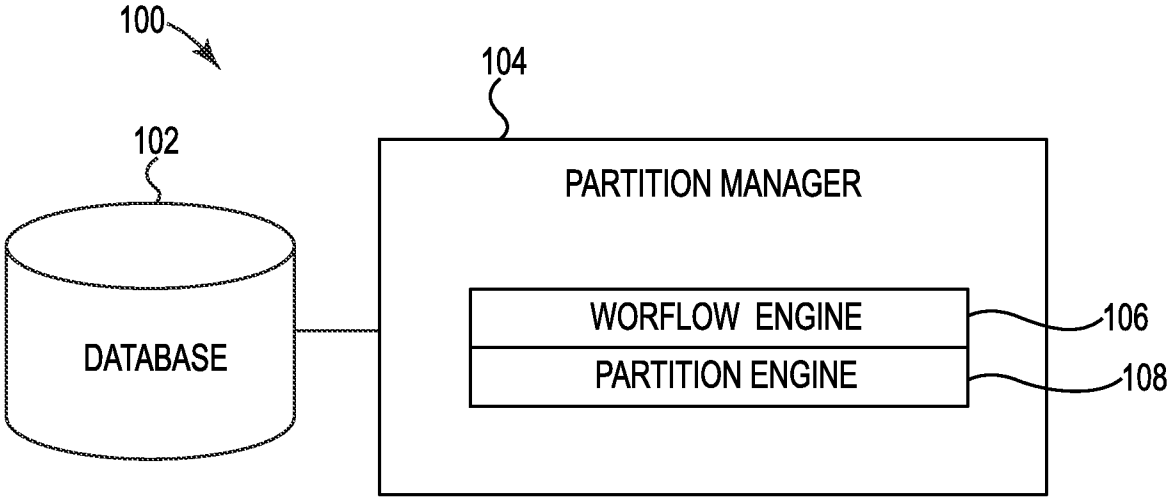

100
104
102
PARTITION MANAGER
DATABASE
WORFLOW ENGINE
106
PARTITION ENGINE
108

FIG. 1

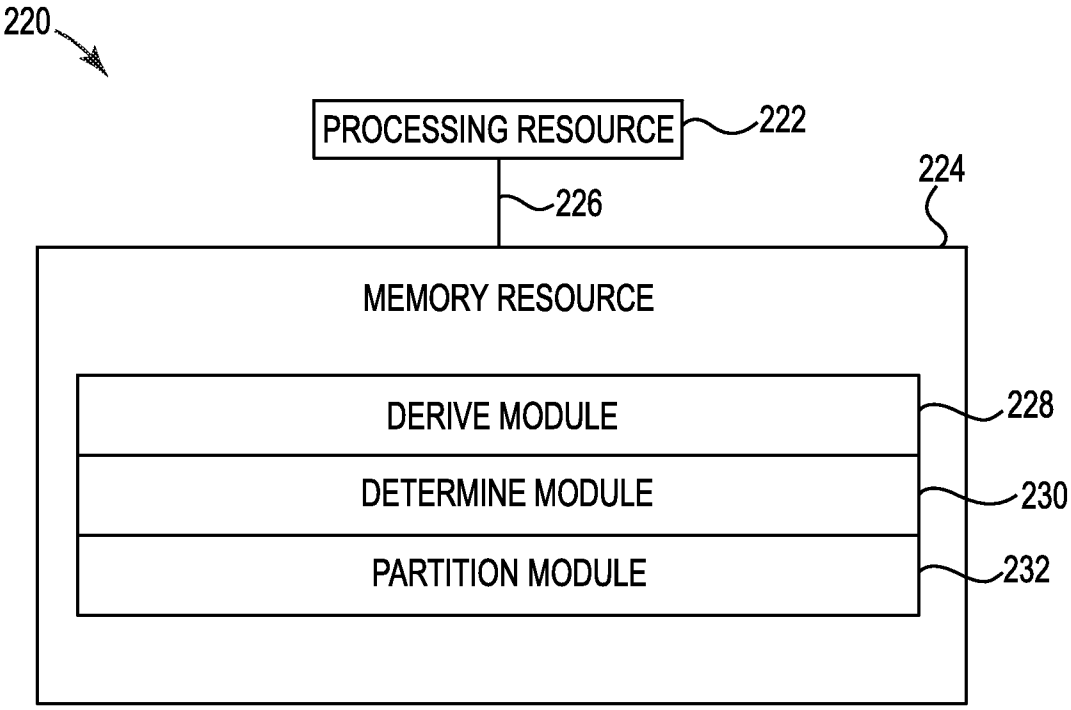

220
PROCESSING RESOURCE
222
224
226
MEMORY RESOURCE
DERIVE MODULE
228
DETERMINE MODULE
230
PARTITION MODULE
232

FIG. 2

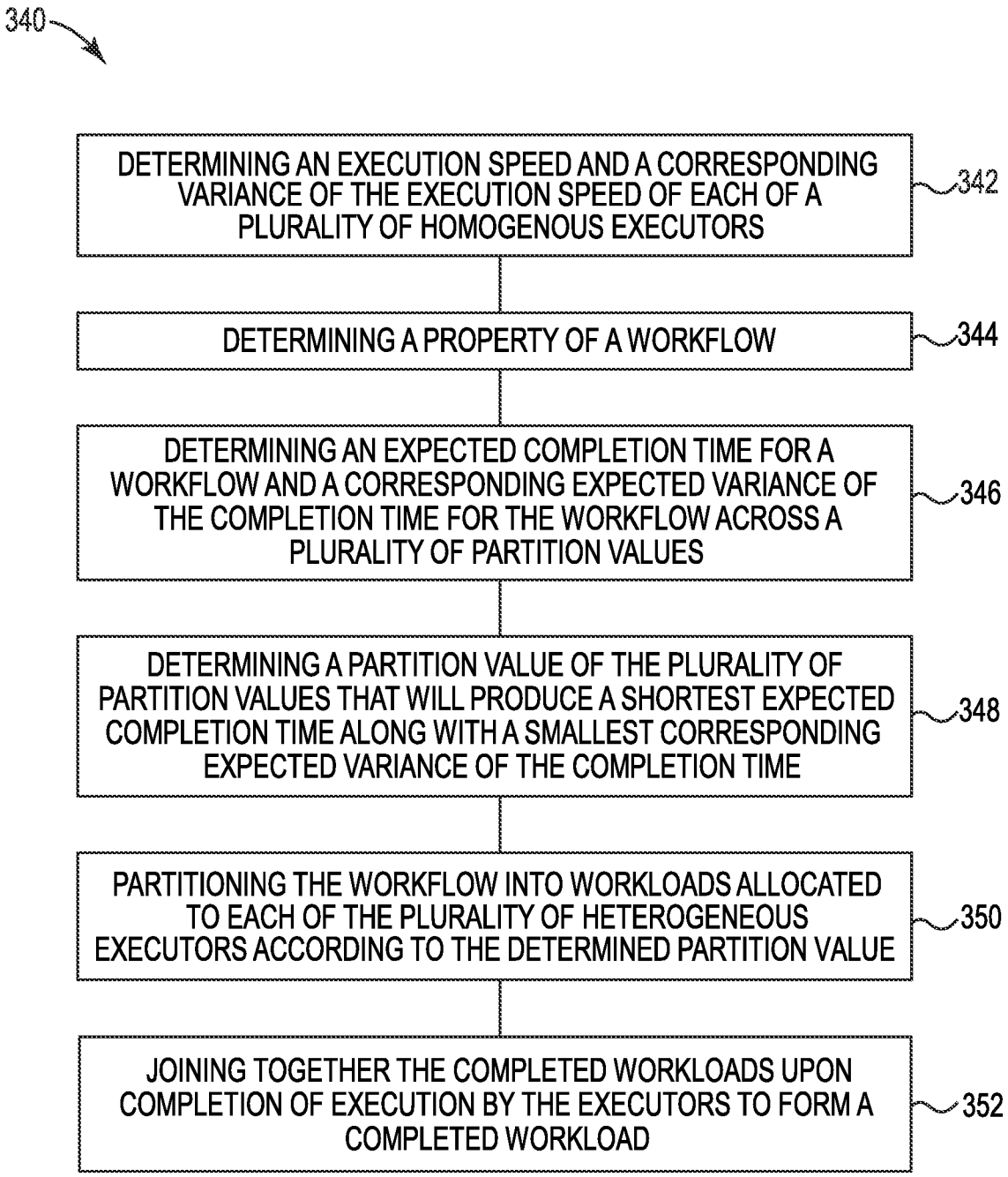

340
DETERMINING AN EXECUTION SPEED AND A CORRESPONDING VARIANCE OF THE EXECUTION SPEED OF EACH OF A PLURALITY OF HOMOGENOUS EXECUTORS
342
DETERMINING A PROPERTY OF A WORKFLOW
344
DETERMINING AN EXPECTED COMPLETION TIME FOR A WORKFLOW AND A CORRESPONDING EXPECTED VARIANCE OF THE COMPLETION TIME FOR THE WORKFLOW ACROSS A PLURALITY OF PARTITION VALUES
346
DETERMINING A PARTITION VALUE OF THE PLURALITY OF PARTITION VALUES THAT WILL PRODUCE A SHORTEST EXPECTED COMPLETION TIME ALONG WITH A SMALLEST CORRESPONDING EXPECTED VARIANCE OF THE COMPLETION TIME
348
PARTITIONING THE WORKFLOW INTO WORKLOADS ALLOCATED TO EACH OF THE PLURALITY OF HETEROGENEOUS EXECUTORS ACCORDING TO THE DETERMINED PARTITION VALUE
350
JOINING TOGETHER THE COMPLETED WORKLOADS UPON COMPLETION OF EXECUTION BY THE EXECUTORS TO FORM A COMPLETED WORKLOAD
352

FIG. 3

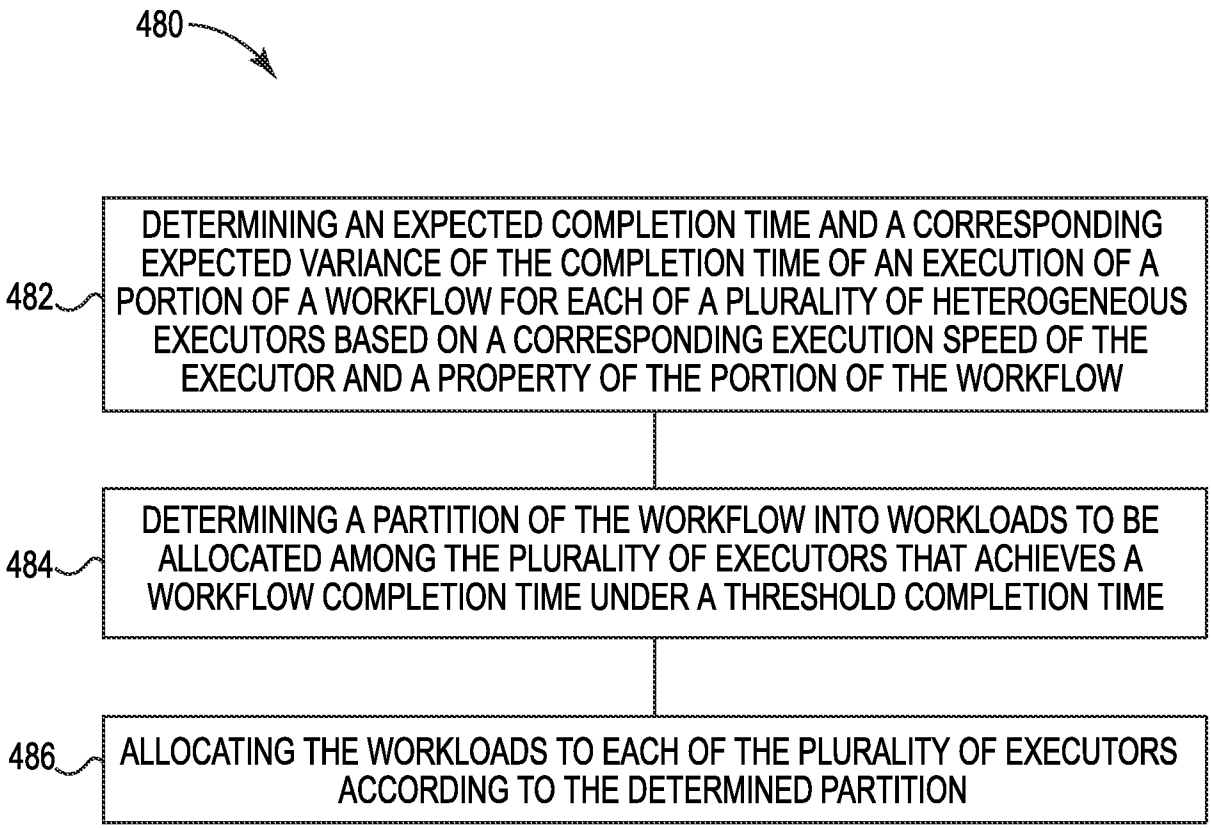

480
482
DETERMINING AN EXPECTED COMPLETION TIME AND A CORRESPONDING EXPECTED VARIANCE OF THE COMPLETION TIME OF AN EXECUTION OF A PORTION OF A WORKFLOW FOR EACH OF A PLURALITY OF HETEROGENEOUS EXECUTORS BASED ON A CORRESPONDING EXECUTION SPEED OF THE EXECUTOR AND A PROPERTY OF THE PORTION OF THE WORKFLOW
484
DETERMINING A PARTITION OF THE WORKFLOW INTO WORKLOADS TO BE ALLOCATED AMONG THE PLURALITY OF EXECUTORS THAT ACHIEVES A WORKFLOW COMPLETION TIME UNDER A THRESHOLD COMPLETION TIME
486
ALLOCATING THE WORKLOADS TO EACH OF THE PLURALITY OF EXECUTORS ACCORDING TO THE DETERMINED PARTITION

FIG. 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/038555

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 10/06(2012.01)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)
G06Q 10/06; G06Q 30/02; G06Q 30/00; G06F 9/46; G06Q 10/00; G06F 3/12

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & Keywords: workflow partition, expected completion time, allocate, workload

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2012-0158530 A1 (MOHAMMAD REZA SHAFIEE et al.) 21 June 2012 See abstract, paragraphs [0033]-[0036],[0038],[0039],[0051],[0056],[0063], [0073],[0075],[0076],[0091] and claims 1,3-7,12,13.	1-6,8,9,11-15
Y		7,10
Y	US 9020829 B2 (WEN-SYAN LI et al.) 28 April 2015 See abstract and claim 1.	7,10
A	US 2013-0218618 A1 (JOSHUA L. PURCELL et al.) 22 August 2013 See abstract and claim 1.	1-15
A	US 2009-0021774 A1 (SUDHENDU RAI) 22 January 2009 See abstract and claims 1-5.	1-15
A	US 2006-0224432 A1 (YANG LI) 05 October 2006 See abstract and claims 1-10.	1-15

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 March 2016 (28.03.2016)	**28 March 2016 (28.03.2016)**

Name and mailing address of the ISA/KR
International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea
Facsimile No. +82-42-481-8578

Authorized officer
KANG, Min Jeong
Telephone No. +82-42-481-8131

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2015/038555

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2012-0158530 A1	21/06/2012	US 8768748 B2	01/07/2014
US 9020829 B2	28/04/2015	None	
US 2013-0218618 A1	22/08/2013	US 2011-0238458 A1 US 8504400 B2	29/09/2011 06/08/2013
US 2009-0021774 A1	22/01/2009	US 2012-0010742 A1 US 8134743 B2 US 8400679 B2	12/01/2012 13/03/2012 19/03/2013
US 2006-0224432 A1	05/10/2006	None	