

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
16 August 2001 (16.08.2001)

PCT

(10) International Publication Number
WO 01/59561 A1

(51) International Patent Classification⁷: G06F 9/00, 12/00

(21) International Application Number: PCT/US01/03801

(22) International Filing Date: 6 February 2001 (06.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/502,692 11 February 2000 (11.02.2000) US

(71) Applicant: GOTO.COM, INC. [US/US]; 140 W. Union Street, Pasadena, CA 91103 (US).

(72) Inventors: CARRASCO, John, Joseph; 2841 St. George Street, Los Angeles, CA 90027 (US). DOLIOV, Stephan; 45 Crestline Drive #4, San Francisco, CA 94131 (US). EHRENFRIED, Frank, B.; 1150 Glenwood Drive, Millbrae, CA 94030 (US).

(74) Agent: GNOFFO, Vincent, J.; Brinks Hofer Gilson & Lione, P.O. Box 10087, Chicago, IL 60610 (US).

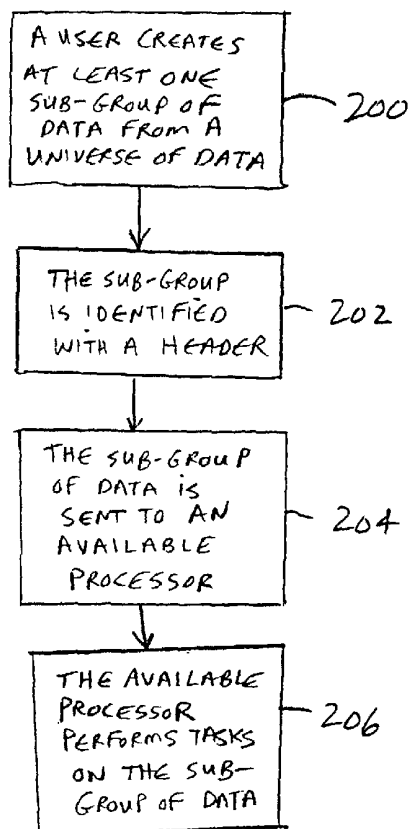
(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search report

[Continued on next page]

(54) Title: SYSTEM AND METHOD FOR CONTROLLING THE FLOW OF DATA ON A NETWORK



(57) Abstract: A method for running tasks on a network, comprising: creating (200) at least one sub-group of data from a universe of data; identifying (202) the sub-group of data with a header, the header containing executable code; sending (204) the sub-group of data to an available processor; and performing (206) tasks with the available processor to obtain result data using the sub-group of data and instructions contained in the executable code in the header.



WO 01/59561 A1



-
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

-1-

SYSTEM AND METHOD FOR CONTROLLING THE FLOW OF DATA ON A NETWORK

REFERENCE TO APPENDIX

5 Appendix A is included of a computer routine listing. The total number of pages is forty.

BACKGROUND

Many commercial enterprises require that large volumes of data be processed in as short a time frame as possible. In recent years, businesses needing to process such large volumes of data have purchased very
10 expensive, specialized multi-processor hardware, often referred to as mainframe computers, supercomputers or massively parallel computers. The cost of such hardware is often in the millions of dollars, with additional costs incurred by support contracts and the need to hire specialized personnel to maintain these systems. Not only is such supercomputing power expensive,
15 but it does not afford the user much control over how any given task gets distributed among the multiple processors. How any computing task gets distributed becomes a function of the operating system of such a supercomputer.

In the field of data processing, often very similar operations are
20 performed on different groups of data. For example, one may want to count the unique instances of a class, e.g., a group of data, for several different classes, know what the arithmetic mean of a given class is, or know what the intersection of two classes may be. In a supercomputing environment, one has to rely on the operating system to make sound decisions on how to
25 distribute the various parts of a task among many central processing units (CPUs). Today's operating systems, however, are not capable of this kind of decision making in a data processing context.

Thus, there is a need for a system and method that overcomes these deficiencies.

-2-

BRIEF SUMMARY OF THE PRESENTLY PREFERRED EMBODIMENTS

According to the preferred embodiments, described is a system and method for allowing multiple processors, for example, a computer's central processing unit(s) (CPU), on a network to perform varying number and type of data processing tasks. Data processing tasks are provided to any of the available CPU's on a network equipped for the system and method. The system and method choose the first available CPU for performance of the data processing task. The system and method provide the task performing CPU with the minimum amount of data needed to complete the task and the necessary software instructions to complete the task.

More specifically, the preferred system and method allow for improved efficiency, in terms of both cost and time, of data processing. The user of the software considers a given data processing task, provides a set of definitions for the type of data required for each task, and specifies the task for a given group of data. The system and method then divide up the input file into the sub-task data files and ships the given data and task specifications to any available computer on the network. The CPU performs the task and returns the completed result to the computer that requested the task.

Thus, large amounts of data to be processed quickly, by ordinary, commodity personal computers, running conventional operating systems such as Windows NT or Unix. A small cluster of, for example, twelve dual processor computers or twenty-four single processor computers, running the software of the preferred embodiments, can equal, if not exceed the performance of a supercomputer with an equivalent number of CPUs.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows an exemplary system for carrying out the present method in accordance with the preferred embodiments.

FIG. 2 shows an overview of the present method, in accordance with the preferred embodiments.

-3-

FIG. 3 shows a process flow for how the method constructs a work queue, in accordance with the preferred embodiments.

FIG. 4 shows a process of sending task data and task instructions to a computer on the network, in accordance with the preferred embodiments.

5 FIG. 5 shows a process for reassembling file results once the worker routines have completed their tasks, in accordance with the preferred embodiments.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

10 A preferred embodiment of the present system and method consists of two major components – a plurality of processing units on a network to perform data processing, and software to control the computers to each work on one discreet task within a larger task, also known as a master task. A task is a routine, such as a grabber routine, a worker routine and a reassembling routine, all described in more detail below. Preferred embodiments for both
15 the network hardware and software are described herein. The present system can be used with search engines. One suitable search engine is described in co-pending application Serial No. 09/322,677, filed May 28, 1999 entitled “System and Method for Influencing a Position on a Search Result List Generated by a Computer Network Search Engine,” which is incorporated by
20 reference herein.

 Referring to the drawings, and particularly FIG. 1, as an overview, a network, generally 100, of processors 102 are shown which perform the routines used in the present system. Those skilled in the art will appreciate that, in addition to a network, the processors 102 can be connected on a LAN,
25 over the Internet, or the like. The processors 102 can be CPUs of commodity computers, super computers, laptops or the like. At least one of the processors 102 acts as a main processor 104. The main processor 104 can be any of the processors 102.

 Referring to FIG. 2, the above system can be used to perform the
30 present method. The present method takes a large universe of data and breaks the data down to smaller groups of data that are task specific (block

-4-

200). In other words, the smaller groups of data are tailored to include only that data which is needed for a particular task. For example, a user can prepare a task queue and a task data queue. The task queue defines which tasks are to be executed and the order in which the tasks are to be executed. The task data queue defines the data elements required for the completion of data processing tasks and how the data for the data processing tasks are laid out in a file so that the individual sub-tasks may readily be performed. Thereafter, the sub-group is identified with a header (block 202). The header is described in more detail below. The sub-group of data is then sent to an available processor (block 204). The available processor can then perform the required task on the sub-group of data (block 206).

Referring to FIG. 3, a grabber routine, referred to as monkeyGrab, located at the CPU which performs the grabbing task, receives instructions on which data elements to grab from a task data file created previously (block 300). The task data file is defined by a combination of a file header and data used by the routine. The grabber routine creates the file header that describes output data so that the worker routine can act on elements in the data file (block 302). The grabber routine insures that the only data elements in the data file are those which are required by a subsequent worker task (block 304). The grabber routine is described in more detail below.

Referring to FIG. 4, shown is the process for sending the tasks to the remote host. Software, for example, scans a memory database of available CPUs and the state of the available CPUs (block 400). Preferably, the first CPU in an available state is sent the task instruction file (block 402) and the task data file (block 404). The task instructions include a worker routine such as monkeyCount, described in more detail below. The task data file could be the results of a routine like monkeyDeJournal, described below. Preferably, the controller routine creates a data file and memory entries to keep track of which available host is performing with task (block 406). Upon completing the task, the task routine makes the result data available to the controller routine (block 408). The controller routine waits for the task routine to indicate completion. Upon successful completion, the controller routine updates the

-5-

memory data files of available hosts (block 410). Upon unsuccessful completion, however, the controller routine updates the data file and memory entries and reassigns the task to another available CPU.

As shown in FIG. 5, the data file which results from completing the task is sent back to the requesting computer, i.e., the main processor, to be reassembled with resulting data files from other completed tasks (block 500). The main processor constructs a reassembling command to specify the sub-task data files that will be reassembled into one file. (block 502). For each sub-task, reassembling code, e.g., monkeyJoin described below, reads and evaluates the file header to identify data columns defining keys and data (block 504). After having read the result file header for each sub-task, the reassembling routine reads a data line from each of the files and outputs the data values for the key in the data line if the keys match (block 506). If the keys do not match, the reassembling routine reads the file header for how to represent and handle the missing data values. Preferably, all task result files come from the same input data file, so there is no need to handle duplicate keys. The task result data files will contain only one of any given key. Likewise, it is preferred that the task routines are written to output the keys in sorted order, so not further sorting is required. Thereafter, the data is reassembled (block 508).

I. Hardware and Network.

Turning to a more specific example of the hardware, the preferred embodiment of the hardware is a set of commodity personal computers. So that a user works within a known environment, it is preferred that each computer contains a current model CPU, the same amount of memory, and at least one high-speed network interface card. Of course, other combinations could be used. For explanatory purposes, a preferred embodiment consists of twelve, dual CPU computers, with each CPU running at 500MHz clock speed, 1Gigabyte of Random Access Memory, and two 100BaseT ethernet network adapter cards. Each computer in the cluster runs the RedHat Linux operating system. Those skilled in the art will appreciate that other operating systems could be used, such as Microsoft Windows and Window NT, Sun

Solaris, and Macintosh. The twelve computers are connected to each other via a high-speed network switch. Each computer is also equipped with, for example, a SCSI II interface and a nine Gigabyte hard drive. Of course, many more computers could be added to the group without impairing performance.

5 II. Software

A preferred embodiment of the software consists of several separate computer programs, implemented in a language that is interpreted at run time, for example, the Perl programming language. Those skilled in the art will appreciate that other programming languages that are interpreted at run time could be used such as Java, PHP3, Python and Tcl/Tk.

10 As an overview, a first computer routine, referred to as the data preprocessing routine, reads the input data file, reads the task data queue control file, performs any number of arbitrary instructions which are specified in the task data queue file, and creates a file header and data file. A second
15 computer routine extracts only the data fields required for a particular data processing task from the data file created by the first routine. The grabber routine creates a file that includes a file header segment and a data segment that a worker routine utilizes. The file header segment of the output data file includes executable or evaluable code. The third routine, the worker routine,
20 performs computations on the data file created by the grabber routine. Preferably, the worker routine parses the file header created by the grabber routine and then performs a requested operation on the data portion of the file that follows the file header. Parsing includes reading the file header and evaluating the programmed code to generate a hash table that defines the
25 structure of the data to follow. A hash table is a data structure that uses keys that map to data values.

30 Since many of the tasks performed by the software are tasks related to some kind of aggregation, e.g., counting the number of instances of a given class, the preferred embodiment of this invention includes a fourth software routine, which allows various task files to be merged. The task files are merged, so that, for a given class, all the aggregate measures are in one data

-7-

row, which makes loading databases an easy task. Each of the four software routines is treated in more detail below.

A. Data Preprocessing routine.

The data preprocessing routine, referred to in the Appendix as monkeyDeJournal, formats data selected by the user so that the data may easily be rendered into discrete parts required for processing. A preferred embodiment of the data preprocessing routine reads an input data file and a control file, returns groupings of data specified in the control file and outputs a new file which will be read by the other routines of this invention.

Reading the input data file is handled via an abstraction mechanism--the data preprocessing routine is pointed to an existing data file parsing routine. The only two assumptions made about the data file parsing routine is that it is capable of 1) identifying groupings of data and 2) returning such groupings of data in a hash table data structure. For example, if the input data file contains groups of data for user clicks on a web site (group: userClick) and groups of data for searches performed by a search engine (group: search), 'userClick' and 'search' would be keys in the hash table of all known data groupings for the input data file. These two grouping keys, in turn, each need to point to another hash table. The names of the keys in this hash table would be the names of the data elements, and the values would be the actual data value for that data element within a particular instance of a group. The group, 'userClick', could for example, contain data elements named 'timeStamp', 'IPAddress', 'rank', and 'AccountID'; the group named 'search' could for example, contain data elements named 'query', 'timeStamp', and 'resultsFound'. As the data file is read groups and their individual elements are returned based on the data in the control file which the data preprocessing routine must read.

The reading of the control file, and acting upon its data, is handled by the code of the data preprocessing routine. The control file is specified as a hash table evaluable at run time. The keys in this hash table are arbitrary names, with each name representing a task queue. The values for these keys are hash tables also. This nested hash table has three required

-8-

keys and any number of optional keys. The required keys are named 'event', 'columns', and 'delimiter'. The optional keys will be described later. The value for the key named 'event' specifies a group of data from the input file that contains the data elements of interest. The input data file parser identifies the group name. For example, the value for the key named 'event' could be 'userClick'. The key named columns points to another hash table. The keys of this nested hash table are the arbitrarily assigned column names that any later data processing tasks may need. The values of the column name keys are one innermost hash table, which allows the user to specify how to construct the columns of interest for a data processing task. This is done via two keys, one named 'source_args', the other named 'source_routine'. The 'source_args' key specifies the data elements from the data group which are to be used in constructing the desired output field; the 'source_routine' is specified as a valid Perl subroutine. Lastly, the key named 'delimiter' specifies how the data columns of interest are to be separated in the data segment of the output file. In the preferred embodiment, this control file would, at a minimum, be presented as follows:

```
{
  'AdvertiserQueue' => {
    'event' => 'userClick',
    'columns' => {
      'AdListingX' => {
        'source_args' => ['AccountID','rank'],
        'source_routine' => 'sub {
          my($ald)=shift @_ ;
          my($rank)=shift @_ ;
          my($x)="$ald###$rank";
          return $x;
        },
      },
    },
    'IP_address' => {
```

-9-

```

        'source_args' => ['clientIP'],
        'source_routine' => 'sub {
                                return shift@$_;
                            },
5          },
          },
          'delimiter' => '\t',
          },
        }
10

```

In this example, 'AdvertiserQueue' is the arbitrary name assigned to the task data queue. The data group of interest comes from the group 'userClick'. Two output columns are desired: one referred to as 'AdListingX', the other as 'IP_Address'. The output column 'AdListingX' would consist of the data elements 'AccountID' and 'rank', which are part of the 'userClick' data group. The final presentation of the 'AdListingX' data column would be in the form of ABC###3, supposing the advertiser ID was 'ABC' and the rank was '3'. The output column 'IP_Address' would merely contain the data value that the data element 'clientIP' has from an instance of the data group 'userClick'. Those skilled in the art will recognize that any valid Perl syntax can be used in the 'source_routine' value to create derivations and modifications to input data fields of interest. The key named 'delimiter' having the value '\t' indicates that the output fields should be separated by a tab character.

In the present instance of this invention, the control file may include three additional, optional keys. These optional keys have the following names: 'deltaT', 'restriction_args', and 'restriction'. The 'deltaT' provides information to the data preprocessing routine about the earliest minute within an hour for which the output of the data preprocessing file should contain data. Legal values for this key are numbers between 0 and 59 inclusive; single digit numbers are preceded with a zero (i.e. if 1 is the value, spell out '01'). The 'restriction_args' key works just like the 'source_args' key previously mentioned. This value for this key provides input arguments to a

-10-

user defined function. The elements of the list must be names of data elements within the data group of interest. The 'restriction' key value is a valid Perl subroutine. For example,

```
{  
  ...  
  'deltaT' => '09',  
  'restriction_args' => ['bid'],  
  'restriction' => 'sub {  
    my($x)=shift @$_;  
    return ($x > 0);  
  }',  
  ...  
}
```

specifies that the first data groups of interested to be included in the output file should occur no sooner than nine minutes after the first hour of data seen by the file parsing routine. The only data group instances to be returned are those whose 'bid' element has a value greater than zero. The 'bid' element is passed to the user defined function specified via the 'restriction' key.

Once the data preprocessing routine has read the control file and evaluated the contents, it creates an output file that begins with a file header segment. The file header segment is written in the form of a Perl evaluable hash table. This file header segment has four required keys and three optional keys that are discussed in the next section (B. Grabber Routine).

After having output the fileheader, the data preprocessing routine enters a token named "EndOfHeader" on the next output line. At this point, any instances of data groupings which meet any restriction criteria are assembled according the rules specified in the control file and then written out to the data portion of the output file, with the columns being delimited by the delimiting character specified in the control file.

B. Grabber routine.

-11-

According to a preferred embodiment of the grabber routine, referred to as monkeyGrab, the preprocessor routine utilizes the data for which the user wishes computations to be performed on and places these data in a file. Thereafter, other routines that can read the file header with executable code, and execute the code, can read the data to be acted upon. Thus, a requirement of the grabber routine is to be able to read the file header created by the preprocessing routine so it can extract the appropriate data elements. Since the grabber routine also writes the minimum amount of data needed for a given task to a task file, the grabber routine can write out a file header which is bound by similar rules as the preprocessing routine's file headers.

The grabber routine grabs data columns based on column names that are provided through input arguments delivered to the grabber routine. For example, a data file may contain a column containing a price value, such as a search result bid column, and a column containing a class descriptor, a key field that may consist of one or many fields. In such a case, the grabber routine could be invoked as follows: "grab -g price -g myKeyField". The data file from which the grabber routine works has in its file header entries for the columns named "price" and "myKeyField". The file header from the preprocessed routine should contain the appropriate header entries, i.e., the key names for the hash table which describes the data columns. The grabber routine reads the header information from the data file to obtain information about the location within the data file of the column locations, the character(s) which delimit the data columns, and any special handling rules, such as how to treat or value an empty column location. Once the grabber has ascertained the column locations and what processing rules are required for a given column, the grabber routine loads those columns of interest and places them in an output file.

The output file, written by the grabber routine, has header information pertinent to the extracted columns. For explanatory purposes, the header is in the form of a Perl hash table with four keys in it. The four keys correspond to four of the seven keys included with the output of the preprocessor routine. These four keys are for the data columns, which of the data columns, or group

-12-

of columns, make a given data row unique, the data column labels, and the output data field delimiter. The key for the data columns points to a second hash table. The second hash table has as its keys column names, and as its values a hash table with two keys. The first key of the innermost hash table describes the data column's location in the data portion of the file, where the innermost hash table is the most embedded hash table in a series of hash tables. The second key of the innermost hash table describes how to represent null data values.

The key for the column which uniquely describes any given data row must have the name of the column that describes the data row. This name is a key in the hash table of columns. The key for the data column labels has the value of the Perl list. Lastly, the key describing the column delimiter has a value corresponding to the column delimiter. If this column delimiter includes any characters which are escape sequences in Perl, then these escape characters are preceded by a backslash --"\ " -- character. The preferred embodiment places a token at the end of the file header, so that the software knows when it is done reading header information and when it can begin reading data.

C. Worker routine.

A preferred embodiment of the worker routine(s), referred to as monkeyCount, much like the preferred embodiment of the preprocessor and the grabber routines, reads the data file headers and also output such file headers. Like the grabber routine, the worker routine reads the file header to determine which columns it is reading and in what order those columns appear in the data segment of the input file, and which of the input data columns constitute the class, or key, definition. Upon reading this information, the worker routine performs the operation desired on the input data. If the worker routine is a counting routine, it will output, for each class, a count of the number of rows which belong to this class, along with a descriptor of a class. For example, if the input data consists of seven rows and two columns, a key and an item to be counted for that key, is as follows:

Key	Search Term
-----	-------------

-13-

5 Adv01 dog
 Adv03 cat
 Adv05 house
 Adv03 mouse
 Adv01 travel
 Adv05 music
 Adv01 sound

the output data file contains the following data rows:

10 Key # searches
 Adv01 3
 Adv03 2
 Adv05 2

15 Likewise, if the worker routine were an adding program, it would sum up the data values in the column of interest for each key of interest.

20 The worker routines minimize any given task in terms of file input/output and complexity, for example, if a count is requested, the worker routine only receives a key and the data element to be counted columns. This allows many worker assignments to occur on any arbitrary number of machines. One worker routine might count the total searches for each advertiser; another might count the number of unique IP addresses that clicked on any given advertisers listing within a specific time period. When the worker routine finishes its assignment, it writes out a file with the same header format as the input file, however, the values in the hash table describing the columns will be the key descriptor from the input file and name of the worked upon data. For example, if the input data file had a key of "Advertiser ID" and a column of "Search Term" and the worker routine was set up to count the number of searches which returned a given advertiser, the output file would have a key of "Advertiser ID" and a column of "Count of Searches".

D. Data Reconstruction routine.

-14-

According to a preferred embodiment of the data reconstruction routine, referred to as monkeyJoin, all fields are reconstructed into one file, organized by key. The reconstruction occurs after the data to be worked on has been preprocessed, i.e., broken up into smaller work units, and the small work units are sent to machines on the network for processing. The data reconstruction facilitates convenient database loading. To accomplish data reconstruction, the data reconstruction routine is given input as to which data files need to be merged into one data file. Each of the data files that will become part of the database load file is supplied as an argument to the data reconstruction routine, in list format. For example, the data reconstruction routine, e.g., monkeyJoin, is called as follows:

```
reconstruct file1 file2 file3 ... fileN
```

For each of the files supplied as an input argument, the reconstruction routine reads the file header information and stores the header information and data in a hash table. Once all of the headers and data have been read, each of the key values are cycled through. For every input file that had a matching key, the corresponding output columns are written. If one of the input files did not have a key entry or a value, the handling of missing or undefined values is invoked and the reconstruction routine supplies an appropriate value, per the notation in the input file header hash table. This file is written out, as the other files, with header information in the format of a Perl hash table. The hash table contains the same four keys as the hash table headers supplied by the grabber and worker routines. The values for the keys of this hash table include the same basic four keys required by this application: the columns hash table, the key hash table, the column delimiter specification and the hash table of column labels.

E. Dispatching routine.

According to a preferred embodiment of the workload distributing routine, referred to as monkeyDispatcher, the CPU and computer memory intensive work occurs in the worker routines which perform the operations of interest, for example, counting unique instances within a class. This CPU and

-15-

memory intensive work ideally is distributed to a number of computers.

Preferably, the present system and method dispatches work to available computers on a network based on the distributing software's known usage load of the computers on the network. The dispatch routine allows one
5 worker or grabber routine to run for each CPU of a computer attached to the network. For example, if there are twenty four counting operations to be performed, and there are twelve computers each equipped with two CPUs, two worker operations can be farmed off to each of the twelve computers for simultaneous processing to ensure the most rapid possible completion of the
10 counting tasks. Likewise, the dispatcher routine needs information regarding which tasks or task components can be done simultaneously and which ones first require the completion of some other task component.

Thus, the dispatching routine needs data about the machines which are capable of receiving the work orders, and how many work orders they
15 may receive at once. For example, a four CPU machine could receive four orders at a time, a one CPU machine only one. The routine also stores data about 1) which machine is currently performing how many tasks at any given point in time and 2) which task(s) any of these machines is performing at any given point in time. Lastly, the dispatch routine can initiate the launching of
20 code for processing the data on a remote machine.

The preferred embodiment begins with a data file, written as a Perl hash table, which specifies the names of the available machines on the network and the total number of CPUs on the given machine. Also specified is the last known "busy/idle" state of each of the CPUs on a given machine,
25 the last known start time. For example, in integer format, an integer indicates that a task was started on a CPU and a blank value indicates that no job is currently running on a given machine's CPU(s). Each machine on the network has one key in the hash table. Each key in this hash table points to a second hash table, the second hash table having key entries for the number
30 of CPUs known for the given machine and the number of CPUs currently occupied doing work for that machine. On the first construction of the hash table, the value for CPUs currently occupied doing work is zero.

-16-

Also in the data file the task queues are specified. The outermost key in this set of tasks hash table points to one or more hash tables which specify the components of a task, and whether or not these sub-tasks can be performed simultaneously. The keys of this outermost task hash table are simply integers, beginning with the number one, and incrementing by one for each task-set. Each of these numbered tasks points to yet another hash table, which contains keys to represent aspects of the task (such as data preprocessing, data grabbing/counting, data joining etc.). A preferred embodiment wraps these individual task keys inside a hash table whose single present key is named, for example, 'parms'. The 'parms' key points to a hash table with four key entries: 'key', 'name', 'tasks' and 'masterTaskFile'. These keys have the corresponding values, a descriptor of the column which constitutes the class level key, e.g., Advertiser ID, a tokenized, that is, a predefined representation of the data preprocessing task (for example, dejournal.lineads to represent the task list pertaining to the reduction of advertiser listings in an internet search engine), the list of paired "grabbing" and "counting" tasks which can be performed simultaneously, and the name of the output file of the data preprocessing routine.

According to a preferred embodiment, the dispatch routine reads in the control file to identify available machines on the network and the machine's availability. As a task is dispatched to a machine, the dispatching software updates its memory copy of the machine availability hash table. Thus, if a machine has two CPUs and the dispatching routine sent a task to the machine with two CPUs, the dispatching software would increment the number of busy CPUs from 0 to 1, to indicate that one job has been sent to the machine on a network. When the machine performing the worker routine task finishes the task, the dispatcher decrements the busy CPU value by one for the machine that performed the task.

With this mechanism in place, prior to assigning tasks to machines on the network, the dispatching software sorts the available machines by current tasks assigned to a machine. If machine X on the network has 0 busy CPUs and machine Y has 1 busy CPU, and both machines X and Y have a total of

-17-

two CPUs, then the dispatching software will preferably first assign a task to machine X. This occurs because machine X has no busy CPUs as far as the dispatching software can determine. Machine X could be running some CPU intensive software without the dispatching routines knowledge. It is preferred that the computers having the CPUs only have the necessary operating system software running, to prevent the problem of work being sent to a computer whose processor is tied up with a non germane task, such as a word processing task. In the preferred embodiment, the computers only include an operating system, a program interpreter, such as a Perl interpreter, and a secure copying program, if necessary. If all machines on the network are equally busy, the dispatching software sorts the stack of available machines by machine name and assigns tasks in that way. If a machine is fully occupied, the dispatching software removes this machine from the available machine stack until the busy machine reports that it has finished at least one of its assigned tasks.

If all machines are busy, the dispatching software waits for a first time period, for example, a few minutes, to retry task dispatching. If the dispatcher software has tasks queued but cannot find an available machine after a second time period, for example, fifteen minutes, the dispatcher software creates a warning message. This condition might indicate a larger system failure that would require resetting the software system and the tasks. The preferred embodiment of this invention supplies enough hardware on a network so that all pieces of hardware on this network are not likely to be completely busy for any given fifteen minutes.

Once the dispatching software identifies the machine assigned a given task, the dispatching software begins assembling a set of commands to be executed on a remote computer. The command set the software assembles is specific to a task, guided by the information provided in the dispatching software's control file. The construction of these commands is specified as follows. The machine creates a name that will uniquely identify a task and the machine on which the task is to be run. This name then gets used as the directory entry mark which the software uses as an indication that a task is

-18-

either running or completed. After constructing the unique name, the dispatching software uses the syntax of the freely available secure shell utility (also known as ssh) to create the command which will launch a program on a remote computer. Those skilled in the art will recognize that other existing utilities, such as remote shell execution (also known as rsh) could as readily be used.

In its present form, the preferred embodiments have the computers on the network access shared disk space, so that the remote computer references the shared disk space for program code and data. Again, those skilled in the art will recognize that using existing network and remote execution tools, both program code and data could be copied to a remote computer's private disk. Thereafter, a remote execution utility could point the worker computer to the new location of the program code and data. Lastly, the dispatching software adds the syntax of the commands to remove the file and mark it created upon completion of the remotely executed task. Once this syntax is constructed, the dispatcher routine creates a copy of itself (known as forking) and overwrites this copy with a call to the constructed commands (known as a fork-exec combination). Pseudo-code is used to illustrate this process:

```
$task = "NumUniqueUsers";
$machineToUse = "machine07";
$programToLaunch = "monkeyGrab";
$dataFileToUse = "AdvertiserReport"
$programArguments = "-g AdvertiserID -g $task";
$programLocation = "/shared/disk/space/code";
$dataLocation = "/shared/disk/space/data";
$dirEntryFileMark = "$task . $machineToUse. system(\"date\")";
$remoteExecutionTool = "ssh";
$remoteExToolArgs = "-secret /secrets/myKeyFile";
$commandSet = "touch $dirEntryMark; x=`$remoteExecutionTool
$remoteExToolArgs $machineToUse '$programToLaunch
$programArguments $dataFileToUse`; if [ $x eq 0]; then rm $dirEntryMark; fi";
```

-19-

fork()
exec("\$commandSet");

If the dispatcher routine's control file indicates that a particular process or process pair can be executed simultaneously, it loops over the steps just described to launch as many processes as are required by the control file and that can be handled by the existing network. If the control file indicates that a process must finish before one or more other processes must begin, the dispatcher routine waits for such a serial task to finish before launching more serial or parallel tasks within a task queue.

F. Special Cases and general extensibility

A preferred embodiment includes a special case of a worker routine, referred to as monkeyLoad. Like all the other worker routines that can be created in the framework of the present system and method, monkeyLoad has the capability of parsing file headers which are in the form of Perl evaluable code. This monkeyLoad routine takes the file header information and creates a set of SQL (structured query language) statements to insert the data which follows the file header into a database. Through a set of standardized and freely available database interfaces for the Perl language, this routine can read the data lines in the output file and insert these as rows into a database. Those skilled in the art will recognize that the worker routine could also read and evaluate the file header, for example, to produce a control file for another routine, which might have more efficient interactions with a database routine, such as Oracle's SQL loader routine (sqlldr). The special requirement of this routine is that the database columns match the column labels provided in the file header. This detail is attended to at the beginning of the process, in the initial control file which a user creates and where the user can specify arbitrary data column labels. Any given worker routine functions by reading in the file header information, evaluating it, and upon output, creating another file header which has a minimum number of keys, e.g., four which any of the worker routines needs to function.

The one special instance of a worker routine demonstrates that the present system and method can be generalized. Since the worker routines

-20-

can parse the file header information, worker routines can accomplish many useful things. One can readily recognize that instead of being instructed to count unique instances, a worker routine could be written to add, subtract, divide, multiply, compute standard deviation and so forth. The unique
5 functionality any such routine requires is the ability to evaluate the header information in a data file as executable code that translates into a hash table with a minimum number of keys.

Although the invention has been described and illustrated with reference to specific illustrative embodiments thereof, it is not intended that
10 the invention be limited to those illustrative embodiments. Those skilled in the art will recognize that variations and modifications can be made without departing from the true scope and spirit of the invention as defined by the claims that follow. It is therefore intended to include within the invention all such variations and modifications as fall within the scope of the appended
15 claims and equivalents thereof.

-21-

WHAT IS CLAIMED IS:

1. A method for running tasks on a network, comprising:
creating at least one sub-group of data from a universe of data;
identifying the sub-group of data with a header, the header
5 containing executable code;
sending the sub-group of data to an available processor; and
performing tasks with the available processor to obtain result
data using the sub-group of data and instructions contained in the executable
code in the header.

10 2. The method according to claim 1, further including returning the
result data files for each data sub-group to a main processor for possible
further processing.

3. The method according to claim 2, wherein returned result data is
reconstructed to form a single result data file which consists of many
15 individual result data files.

4. A method for preprocessing data, comprising:
reading an input data file;
placing the input data into a data structure;
selecting data elements from the data structure;
20 outputting a file header that describes the selected data
elements into a file; and
outputting the selected data elements into the file.

5. The method according to claim 4, wherein the data structure is a
hash table.

25 6. The method according to claim 4, wherein the data elements
selected are chosen by evaluating a block of program code such as a hash
data at run time.

-22-

7. The method according to claim 4, wherein the file header is a hash table that can be evaluated at run time.

8. A method for extracting data, comprising:
reading a file header containing executable code;
5 executing the code to determine data to extract from a universe of data; and
obtaining the data to be extracted from a data structure.

9. The method according to claim 8, further including loading the extracted data to an output file.

10. The method according to claim 8, wherein the data structure is a hash table.

11. A method for processing data, comprising:
reading a file header that contains executable code;
executing the executable code;
15 determining a desired operation from the executed code; and
performing the desired operation on data.

12. The method according to claim 11, further including outputting result data after performing the desired operation.

13. The method according to claim 11, wherein the desired operation is written in a language that is executed at run time.

14. The method according to claim 13, wherein the desired operation is any one selected from the group of addition, subtraction, multiplication, division, counting total number of instances in a class, enumerating the unique instances of a class, computing descriptive statistics such as the standard deviation, standard error of the mean, median,
25 arithmetic mean, variance, covariance, correlation coefficient, and odds ratio.

-23-

15. A method for dispatching a task and data to a central processing unit located on a network, comprising:

executing code placed in a file header;

determining if at least one, if any, central processing unit is

5 available on the network; and

dispatching the task to the central processing unit based on availability.

16. The method according to claim 15, wherein the availability of the central processing unit is determined by reading a control file.

10 17. The method according to claim 16, wherein the control file is formatted for evaluation as a hash table at run time.

18. The method according to claim 16, further including updating the control file by re-writing a new control file which indicates the status of the central processing units.

15 19. The method according to claim 16, further including:
assembling at least one command to be executed by the central processing unit; and
sending the at least one command to the central processing unit for execution.

20 20. The method according to claim 19, wherein the at least one command includes creating a name to identify the task.

21. The method according to claim 15, further including marking a state of the task.

25 22. The method according to claim 21, wherein the state of the task is either running or completed.

-24-

23. The method according to claim 15, further including retrying, after a time period has elapsed, to determine if the availability of the at least one central processing unit if all central processing units were busy.

24. The method according to claim 23, wherein the time period can be specified by a user.

25. The method according to claim 24, wherein the time period is fifteen minutes.

26. A method for loading data, comprising:
reading a file header containing executable code;
executing the code to obtain file header information; and
creating a structured query language statement based on the file header information.

27. A method for running tasks on a network, comprising:
creating at least one sub-group of data from a universe of data;
identifying the sub-group of data with a header, the header containing executable code;
sending the sub-group of data to an available processor;
performing tasks with the available processor to obtain result data using the sub-group of data and instructions contained in the executable code in the header; and
returning the result data to a main processor, wherein returned result data is reconstructed to form a result.

28. A method for running tasks on a network, comprising:
reading an input data file;
placing the input data into a data structure;
selecting data elements from the data structure;
outputting a file header that describes the selected data elements into a file;
outputting the selected data elements into the file;

-25-

reading the file header containing executable code; and
executing the code to determine data to extract from a universe
of data.

29. The method according to claim 28, further including loading the
extracted data to an output file.

30. The method according to claim 28, wherein the data structure is
a hash table.

31. A method for running tasks on a network, comprising:
reading a file header that contains executable code;
executing the executable code;
determining a desired operation from the executed code;
performing the desired operation on data;
outputting result data after performing the operation.

32. The method according to claim 31, wherein the desired
operation is written in a language that is executed at run time.

33. The method according to claim 32, wherein the desired
operation is any one selected from the group of addition, subtraction,
multiplication, division, counting total number of instances in a class,
enumerating the unique instances of a class, computing descriptive statistics
such as the standard deviation, standard error of the mean, median,
arithmetic mean, variance, covariance, correlation coefficient, and odds ratio.

34. A method for dispatching a task and data to a central processing
unit located on a network, comprising:
executing code placed in a file header;
determining if at least one, if any, central processing unit is
available on the network;

-26-

retrying, after a time period has elapsed, to determine if the availability of the at least one central processing unit if all central processing units were busy;

dispatching the task to the central processing unit based on availability; and

assembling at least one command to be executed by the central processing unit, wherein the at least one command includes creating a name to identify the task.

35. The method according to claim 34, wherein the availability of the central processing unit is determined by reading a control file.

36. The method according to claim 35, further including the step of updating the control file.

37. The method according to claim 34, further including marking a state of the task.

38. The method according to claim 37, wherein the state of the task is either running or completed.

39. The method according to claim 34, wherein the time period can be determined by a user.

40. The method according to claim 39, wherein the time period is fifteen minutes.

41. A system for running tasks on a network, comprising:
a data preprocessing routine to create at least one sub-group of data from a universe of data;

a data header containing executable code that identifies the sub-group of data; and

a dispatch routine to send the sub-group of data to an available processor, wherein the available processor performs to obtain result data

-27-

using the sub-group of data and instructions contained in the executable code in the header.

42. The system according to claim 41, further including a main processor to collect the result data files for each data sub-group.

5 43. The system according to claim 42, further including a data reconstructing routine to reconstruct the result data to form a result.

 44. A system for preprocessing data, comprising:
 a processor to read an input data file; and
 a preprocessing routine to place the input data into a data
10 structure, select data elements from the data structure, output a file header that describes the selected data elements into a file, and output the selected data elements into the file.

 45. The system according to claim 44, wherein the data structure is a hash table.

15 46. The system according to claim 45, wherein the data elements selected are chosen by evaluating a block of program code such as a hash table.

 47. The system according to claim 41, wherein the file header is a hash table that can be evaluated at run time.

20 48. A system for extracting data, comprising:
 a processor to read a file header containing executable code, wherein the processor executes the code to determine data to extract from a universe of data; and
 a data grabber routine to obtain the data to be extracted from a
25 data structure.

 49. The system according to claim 48, further including a dispatch routine to load the extracted data to an output file.

-28-

50. The system according to claim 48, wherein the data structure is a hash table.

51. A system for processing data, comprising:
a processor to read a file header that contains executable code;

5

and

an operating system supporting an interpreted language to execute the executable code, wherein the processor determines a desired operation from the executed code and performs a desired operation on data.

10

52. The system according to claim 51, wherein the processor outputs result data after performing the desired operation.

53. The system according to claim 51, wherein the desired operation is written in a language that is executed at run time.

15

54. The system according to claim 53, wherein the desired operation is any one selected from the group of addition, subtraction, multiplication, division, counting total number of instances in a class, enumerating the unique instances of a class, computing descriptive statistics such as the standard deviation, standard error of the mean, median, arithmetic mean, variance, covariance, correlation coefficient, and odds ratio.

20

55. A system for dispatching a task and data to a central processing unit located on a network, comprising:

an operating system supporting an interpreted language to execute code placed in a file header; and

a dispatching routine to determine at least one, if any, central processing unit available on the network, and to dispatch the task to the central processing unit based on availability.

25

56. The system according to claim 55, wherein the availability of the central processing unit is determined by reading a control file.

-29-

57. The system according to claim 56, wherein the control file is formatted for evaluation as a hash table at run time.

58. The system according to claim 56, wherein the dispatching routine updates the control file by re-writing a new control file which indicates the status of the central processing units.

59. The system according to claim 55, wherein the dispatching routine assembles at least one command to be executed by the central processing unit, and sends the at least one command to the central processing unit for execution.

60. The system according to claim 59, wherein the command includes creating a name to identify the task.

61. The system according to claim 55, wherein the dispatching routine marks a state of the task.

62. The system according to claim 61, wherein the state of the task is either running or completed.

63. The system according to claim 55, wherein the dispatching routine retries, after a time period has elapsed, to determine the availability of the at least one central processing if all central processing units were busy.

64. The system according to claim 63, wherein the time period can be determined by a user.

65. The system according to claim 64, wherein the time period is fifteen minutes.

66. A system for loading data, comprising:
a processor to read a file header containing executable code and executing the code to obtain file header information; and
a preprocessing routine to create a structured query language based on the file header information.

-30-

67. A system for running tasks on a network, comprising:
a preprocessing routine to create at least one sub-group of data
from a universe of data;
a data header containing executable code that identifies the
sub-group of data; and
a dispatch routine to send the sub-group of data to an available
processor, wherein the available processor performs to obtain result data
using the sub-group of data and instructions contained in the executable code
in the header; and
a main processor to collect returned result data, wherein
returned result data is reconstructed to form a result.

68. A system for running tasks on a network, comprising:
a processor to read an input data file;
a preprocessing routine to place the input data into a data
structure, select data elements from the data structure, output a file header
that describes the selected data elements into a file, and output the selected
data elements into the file, and the processor reads the file header containing
executable code, wherein the processor executes the code to determine data
to extract from a universe of data; and
a data grabber routine to obtain the data to be extracted from a
table.

69. The system according to claim 68, further including a dispatch
routine to load the extracted data to an output file.

70. The system according to claim 68, wherein the data structure is
a hash table.

71. A system for running tasks on a network, comprising:
a processor to read a file header that contains executable code;
and

-31-

an operating system supporting an interpreted language to execute the executable code, wherein the processor determines a desired operation from the executed code and performs the desired operation on data, wherein the processor outputs result data after performing the operation.

5 72. The system according to claim 71, wherein the desired operation is written in a language that is executed at run time.

73. The system according to claim 72, wherein the desired operation is any one selected from the group of addition, subtraction, multiplication, division, counting total number of instances in a class, enumerating the unique instances of a class, computing descriptive statistics such as the standard deviation, standard error of the mean, median, arithmetic mean, variance, covariance, correlation coefficient, and odds ratio.

10

74. A system for dispatching a task and data to a central processing unit located on a network, comprising:

15 an operating system supporting an interpreted language to execute code placed in a file header; and

 a dispatching routine to determine at least one, if any, central processing unit available on the network, and to dispatch the task to the central processing unit based on availability, wherein the dispatching routine retries, after a time period has elapsed, to determine the availability of the at

20 least one central processing if all central processing units were busy.

75. The system according to claim 74, wherein the availability of the central processing unit is determined by reading a control file.

76. The system according to claim 75, wherein the dispatching routine updates the control file.

25

77. The system according to claim 74, wherein the dispatching routine marks a state of the task.

-32-

78. The system according to claim 77, wherein the state of the task is either running or completed.

79. The system according to claim 74, wherein the time period can be determined by a user.

5 80. The system according to claim 79, wherein the time period is fifteen minutes.

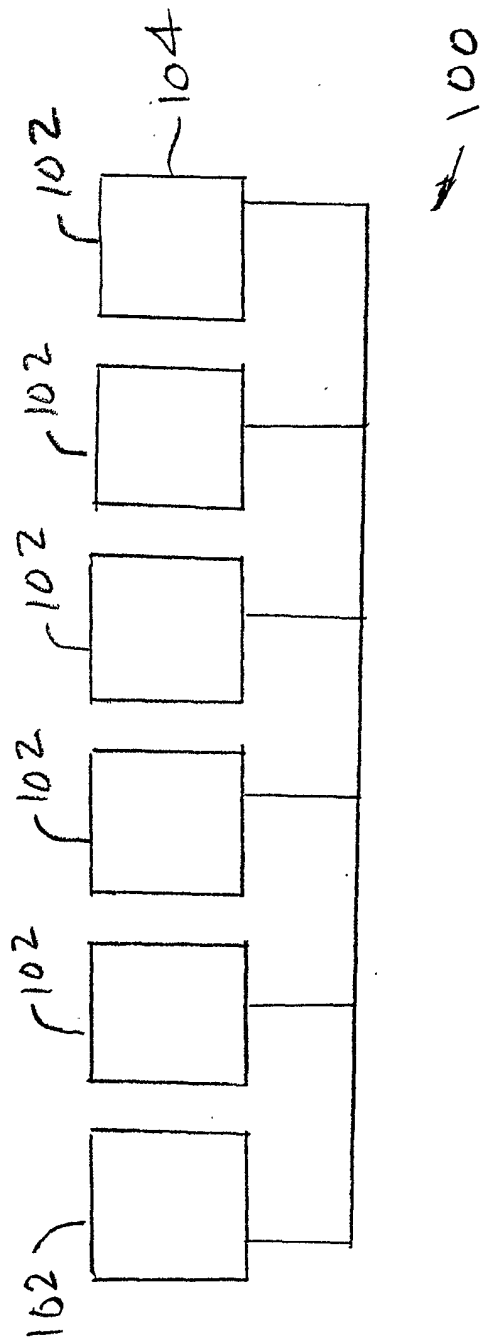


FIG. 1

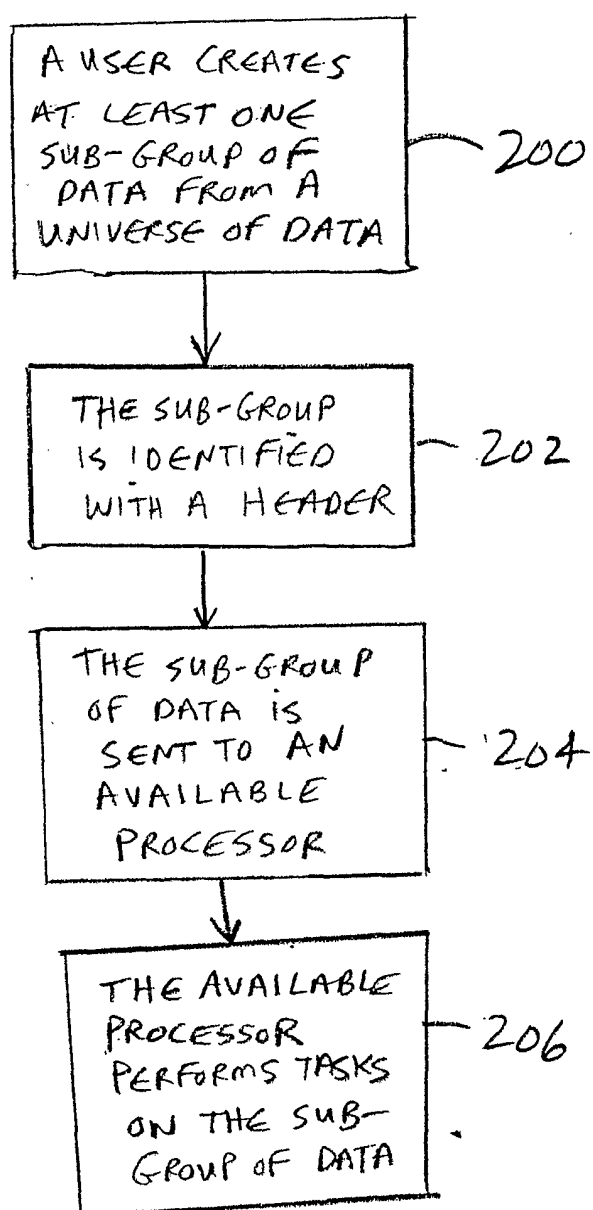


FIG. 2

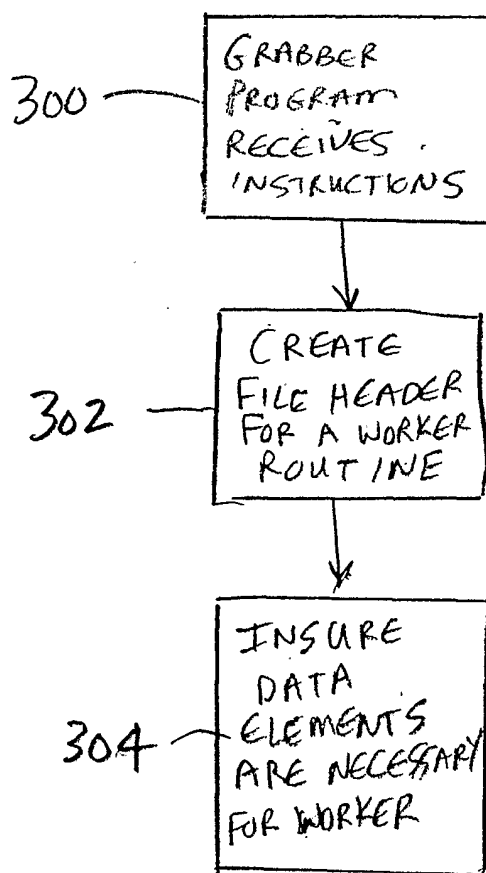


FIG. 3

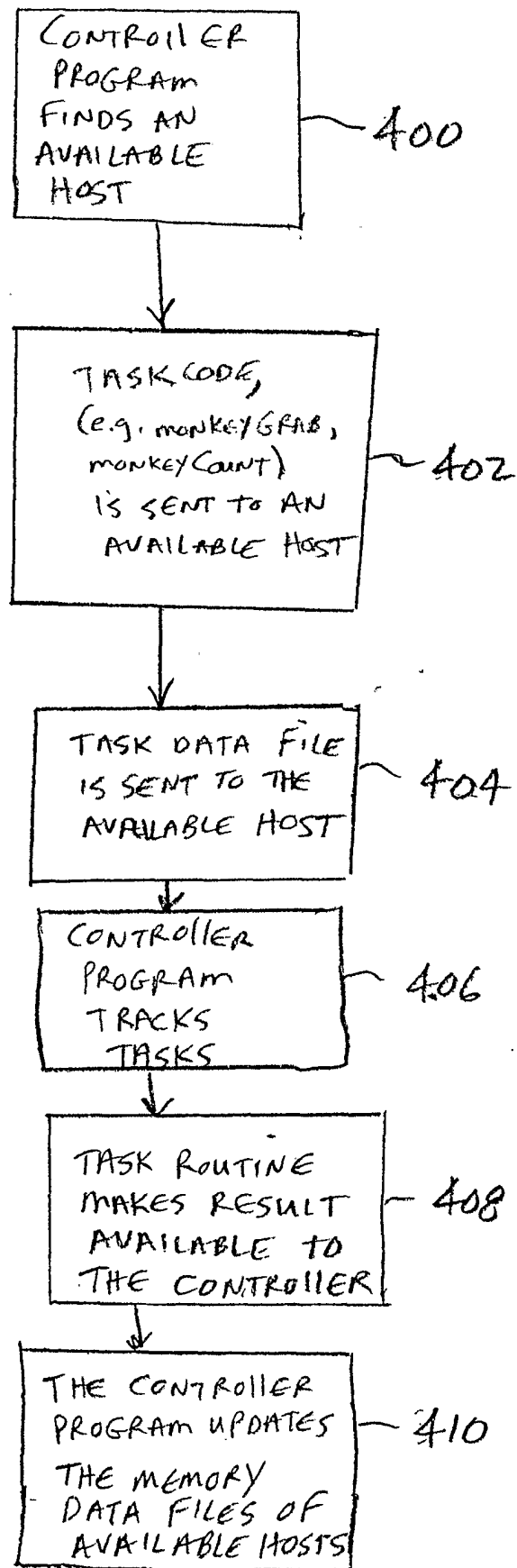


FIG. 4

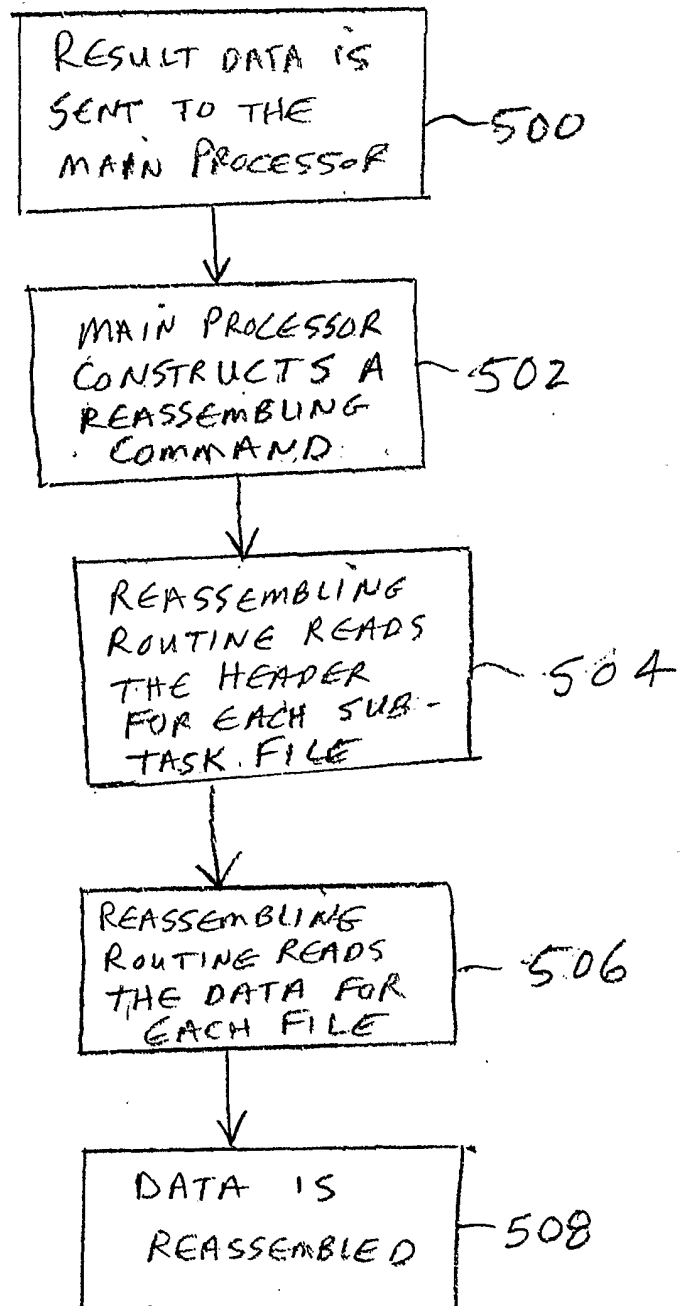


FIG. 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/03801

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G06F 9/00, 12/00
US CL : 709/100, 104; 711/149

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)

U.S. : 709/100, 104; 711/149

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)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 3,913,070 A (MALCOLM et al.) 14 October 1975 (14.10.1975), col.10-col.25.	1-80
Y	US 5,995,996 A (VENABLE) 30 November 1999 (30.11.1999), see Summary of the Invention.	1-80
Y	US 4,972,314 A (GETZINGER et al.) 20 November 1990 (20.11.1990), see Summary of the Invention.	1-80

☐ 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	"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
"E" earlier application or patent published on or after the international filing date	"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
"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)	"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
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

11 May 2001 (11.05.2001)

Date of mailing of the international search report

107 JUN 2001

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. 703 305-3230

Authorized officer

ALVIN OBERLEY *James R. Matthews*
Telephone No. 703 305-3665