US009674033B2

(12) **United States Patent**
AlShaikh et al.

(10) **Patent No.: US 9,674,033 B2**
(45) **Date of Patent: *Jun. 6, 2017**

(54) **HIGH PERFORMANCE AND GRID COMPUTING WITH LIVELINESS AND DEADLINES FAULT TOLERANT DATA DISTRIBUTOR QUALITY OF SERVICE**

(71) Applicants: **Saudi Arabian Oil Company**, Dhahran (SA); **KING FAHD UNIVERSITY OF PETROLEUM AND MINERALS**, Dhahran (SA)

(72) Inventors: **Raed Abdullah AlShaikh**, Dhahran (SA); **Sadiq M. Sait**, Dhahran (SA)

(73) Assignees: **Saudi Arabian Oil Company**, Dhahran (SA); **King Fahd University of Petroleum and Minerals**, Dhahran (SA)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **15/202,044**

(22) Filed: **Jul. 5, 2016**

(65) **Prior Publication Data**

US 2016/0315800 A1 Oct. 27, 2016

Related U.S. Application Data

(63) Continuation-in-part of application No. 14/835,279, filed on Aug. 25, 2015, now Pat. No. 9,429,677,
(Continued)

(51) **Int. Cl.**
G06F 3/00 (2006.01)
H04L 12/24 (2006.01)
(Continued)

(52) **U.S. Cl.**
CPC ***H04L 41/0668*** (2013.01); ***G01V 1/28*** (2013.01); ***G01V 1/282*** (2013.01); ***G01V 11/00*** (2013.01);
(Continued)

(58) **Field of Classification Search**
CPC G06F 9/5072; G01V 1/28; G01V 11/00; G01V 2200/14; G01V 1/282; H04L 67/42;
(Continued)

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,526,418 B2 4/2009 Pita et al.
7,562,143 B2 * 7/2009 Fellenstein G06F 9/5011 370/254
(Continued)

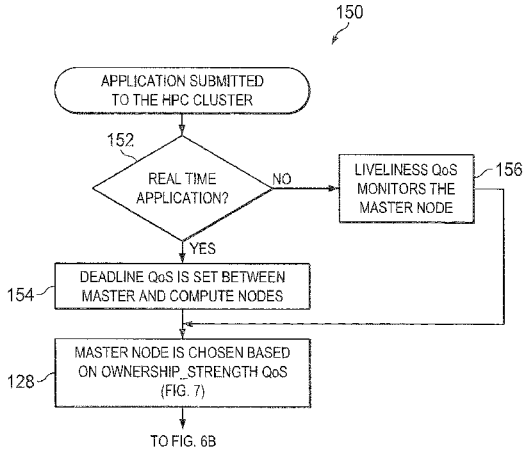
OTHER PUBLICATIONS

A. Agbaria et al., "Starfish: Fault-Tolerant Dynamic MPI Programs on Clusters of Workstations", In the 8th IEEE International Symposium on High Performance Distributed Computing, 1999; pp. 167-176.
(Continued)

Primary Examiner — Jing-Yih Shyu
(74) *Attorney, Agent, or Firm* — Bracewell LLP; Constance G. Rhebergen; Albert B. Kimball, Jr.

(57) **ABSTRACT**

High performance computing (HPC) and grid computing processing for seismic and reservoir simulation are performed without impacting or losing processing time in case of failures. A Data Distribution Service (DDS) standard is implemented in High Performance Computing (HPC) and grid computing platforms, to avoid the shortcomings of current Message Passing Interface (MPI) communication between computing modules, and provide quality of service (QoS) for such applications. QoS properties of the processing can be controlled. Multiple data publishers or master nodes of a cluster have access to the same data source. Each of these publishers has an ownership strength quality of service, and the publisher with the highest ownership strength number is the designated publisher of the data to subscriber processor nodes of the cluster. If the designated data publisher prematurely terminates or crashes for some
(Continued)



reason, then the publisher node with the next highest ownership strength measure is designated as data publisher and continues publishing data to subscribers. The QoS properties include ability to switch to a different designated publisher when the input data is being processed in real time, and a specified deadline time has passed during which data has not been published.

21 Claims, 8 Drawing Sheets

Related U.S. Application Data

which is a continuation-in-part of application No. 14/490,894, filed on Sep. 19, 2014, now Pat. No. 9,128,211, which is a continuation-in-part of application No. 14/490,862, filed on Sep. 19, 2014, now Pat. No. 9,134,455, which is a continuation of application No. 13/649,286, filed on Oct. 11, 2012, now Pat. No. 8,874,804.

(60) Provisional application No. 61/545,766, filed on Oct. 11, 2011.

(51) **Int. Cl.**

H04L 29/08	(2006.01)
G01V 1/28	(2006.01)
G01V 11/00	(2006.01)
G06F 9/50	(2006.01)

(52) **U.S. Cl.**

CPC ***G06F 9/5072*** (2013.01); ***H04L 41/14*** (2013.01); ***H04L 67/10*** (2013.01); ***H04L 67/12*** (2013.01); *G01V 2200/14* (2013.01)

(58) **Field of Classification Search**

CPC H04L 67/12; H04L 41/0668; H04L 67/10; H04L 41/14

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,596,480	B2	9/2009	Fung et al.	
7,620,534	B2	11/2009	Pita et al.	
7,660,711	B2	2/2010	Pita et al.	
7,809,537	B2	10/2010	Hemanthkumar et al.	
8,577,660	B2	11/2013	Wendt et al.	
2004/0162677	A1*	8/2004	Bednar	G01V 1/28 702/14
2006/0248182	A1*	11/2006	Glassco	H04L 67/322 709/223
2007/0118842	A1*	5/2007	Weida	G06F 9/465 719/330
2009/0187391	A1*	7/2009	Wendt	G01V 1/28 703/7
2010/0114494	A1	5/2010	Higginbotham et al.	
2010/0146183	A1*	6/2010	Fleming	G06F 13/387 710/305
2010/0254217	A1*	10/2010	Chu	G01V 1/30 367/38
2011/0138396	A1*	6/2011	Chen	G06F 9/5066 718/105
2011/0153300	A1*	6/2011	Holl	E21B 47/022 703/10
2011/0295923	A1*	12/2011	de Campos Ruiz	G06F 9/542 709/201
2012/0136996	A1*	5/2012	Seo	H04L 41/5058 709/224

OTHER PUBLICATIONS

Anonymous "OpenDDS Developer's Guide" OpenDDS Version 3.0, Aug. 19, 2011, 192 pages.

Corsaro "Quality of Service in Publish/Subscribe Middleware" Global Data Management 19 (2006): 20, Apr. 26, 2006, pp. 1-22.

International Search Report and Written Opinion for related PCT Application No. PCT/US2012/059630 dated Sep. 12, 2013.

J. Neelamegam et al., "PromisQoS: An Architecture for Delivering QoS to High-Performance Applications on Myrinet Clusters", Proceedings of the 28th Annual International Conference on Local Computer Networks, Oct. 2003; 8 pages.

Rizou et al., "Providing Qos Guarantees in Large-Scale Operator Networks", International Conference on High Performance Computing and Communications, 2010, pp. 1-9, IEEE.

Schneider et al. "Is DDS for You? A Whitepaper by Real-Time Innovations" Real-Time Innovations, Inc., Apr. 29, 2009, 12 pages.

* cited by examiner

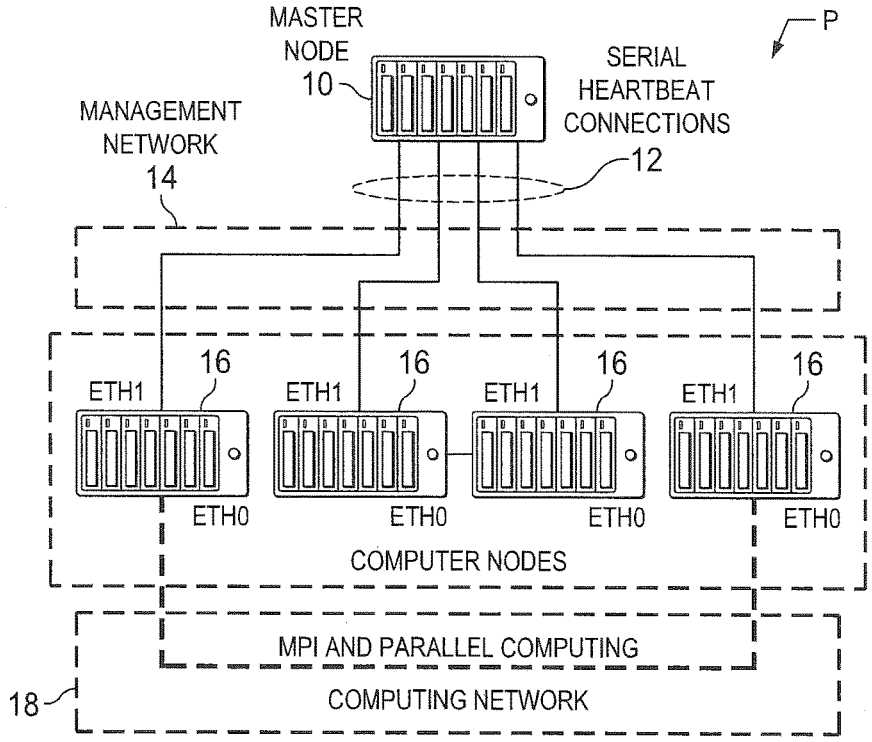


FIG. 1
(PRIOR ART)

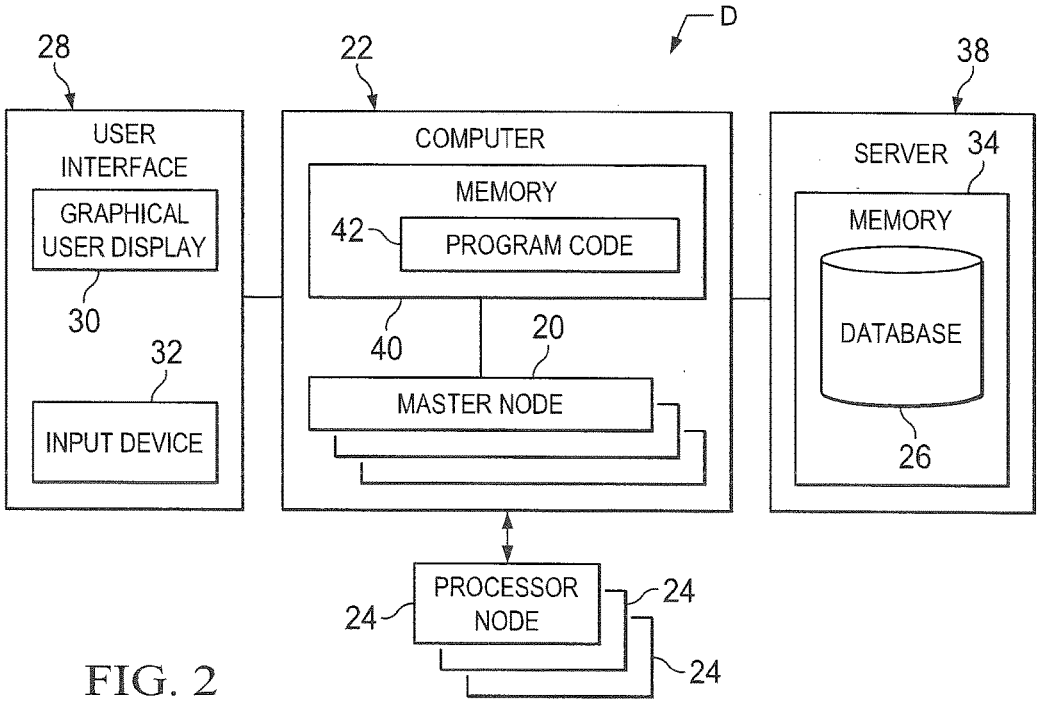


FIG. 2

FIG. 3

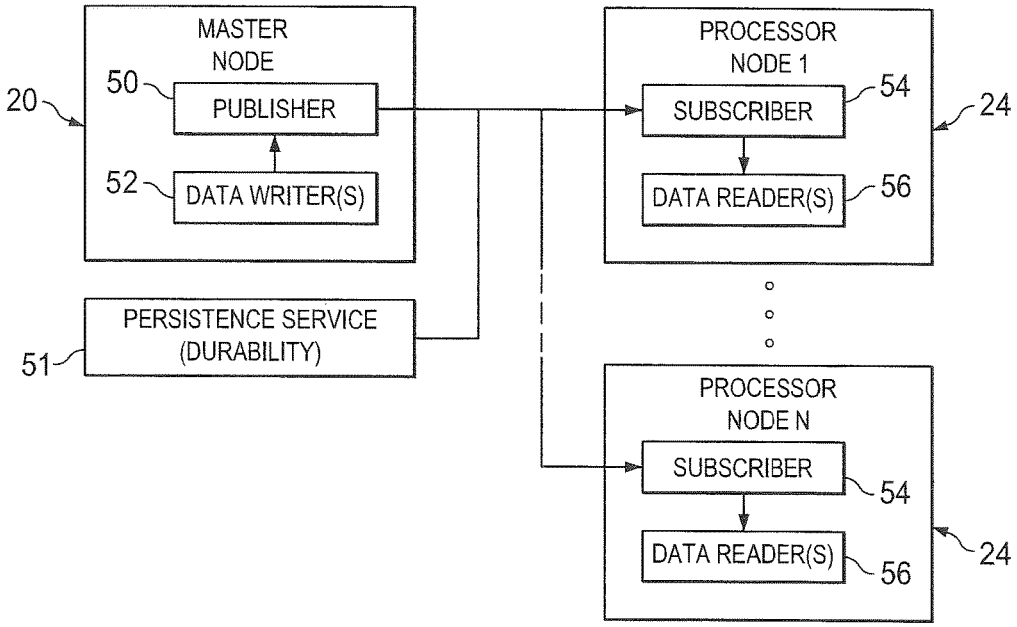
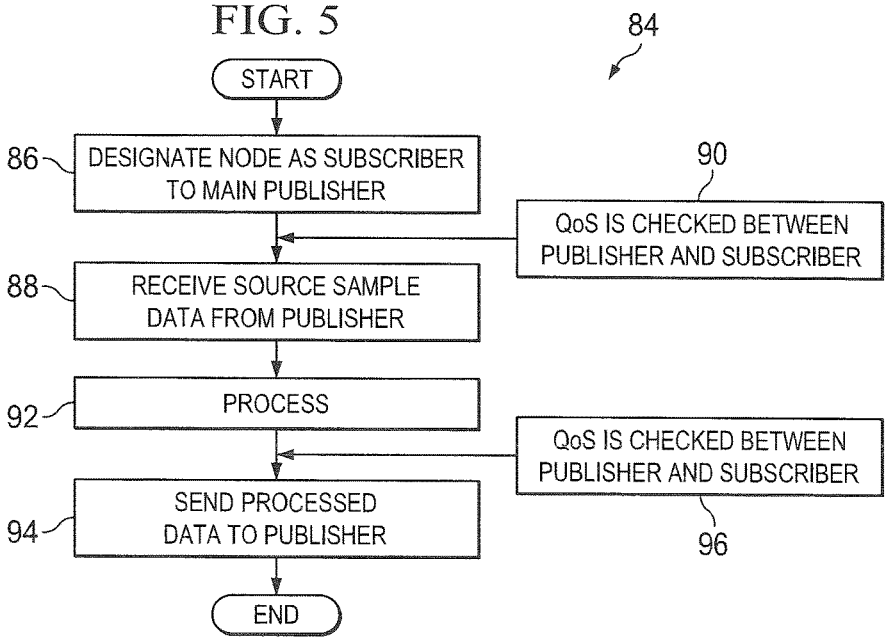

MASTER NODE
20
50
PUBLISHER
52
DATA WRITER(S)
PERSISTENCE SERVICE (DURABILITY)
51
PROCESSOR NODE 1
SUBSCRIBER
54
DATA READER(S)
56
24
PROCESSOR NODE N
SUBSCRIBER
54
DATA READER(S)
56
24

FIG. 5


84
START
86
DESIGNATE NODE AS SUBSCRIBER TO MAIN PUBLISHER
90
QoS IS CHECKED BETWEEN PUBLISHER AND SUBSCRIBER
88
RECEIVE SOURCE SAMPLE DATA FROM PUBLISHER
92
PROCESS
QoS IS CHECKED BETWEEN PUBLISHER AND SUBSCRIBER
94
SEND PROCESSED DATA TO PUBLISHER
96
END

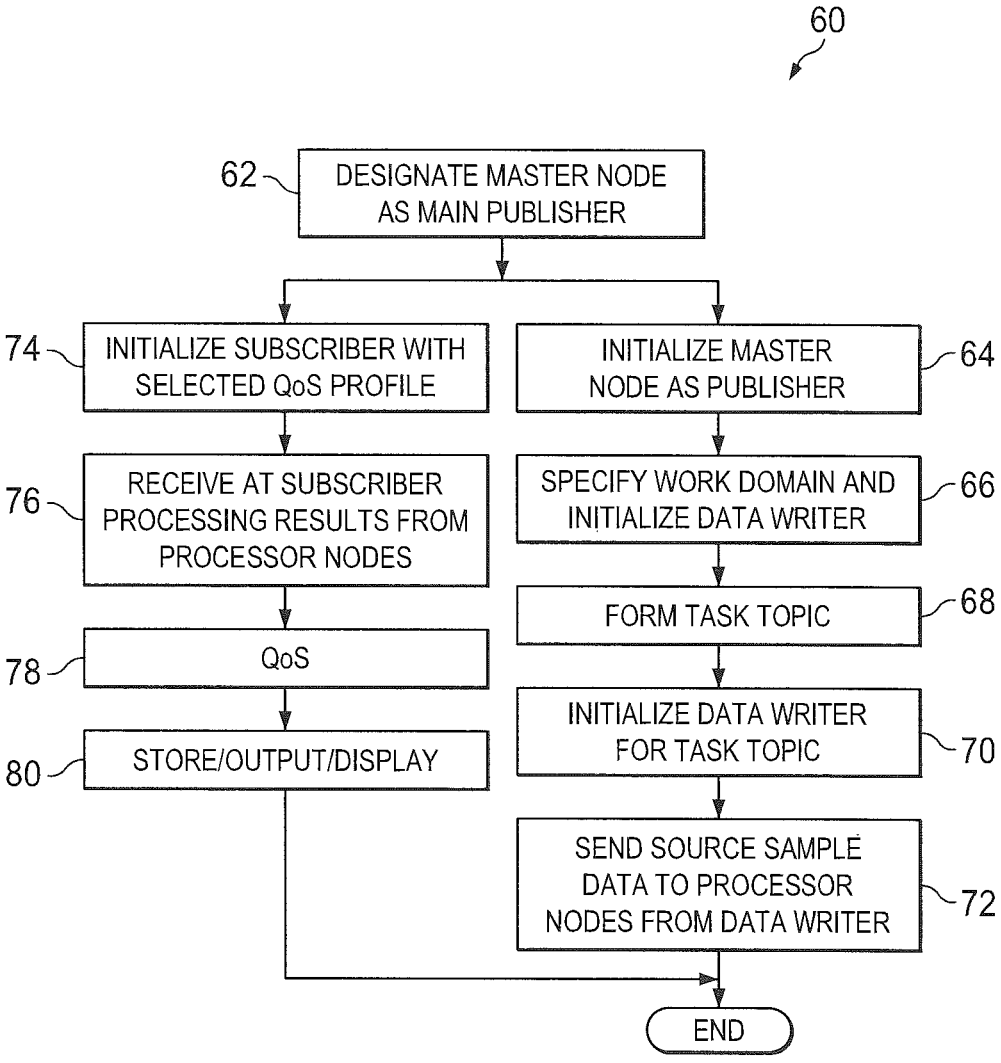

60
62
DESIGNATE MASTER NODE AS MAIN PUBLISHER
74
INITIALIZE SUBSCRIBER WITH SELECTED QoS PROFILE
64
INITIALIZE MASTER NODE AS PUBLISHER
76
RECEIVE AT SUBSCRIBER PROCESSING RESULTS FROM PROCESSOR NODES
66
SPECIFY WORK DOMAIN AND INITIALIZE DATA WRITER
68
FORM TASK TOPIC
78
QoS
70
INITIALIZE DATA WRITER FOR TASK TOPIC
80
STORE/OUTPUT/DISPLAY
72
SEND SOURCE SAMPLE DATA TO PROCESSOR NODES FROM DATA WRITER
END

FIG. 4

FIG. 6

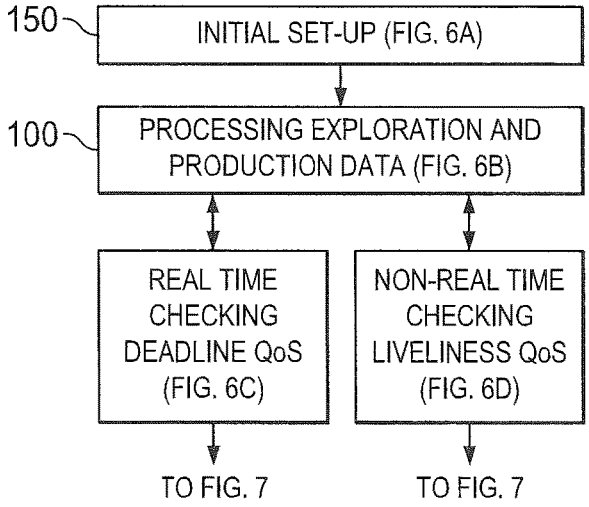

150
INITIAL SET-UP (FIG. 6A)
100
PROCESSING EXPLORATION AND PRODUCTION DATA (FIG. 6B)
REAL TIME CHECKING DEADLINE QoS (FIG. 6C)
NON-REAL TIME CHECKING LIVELINESS QoS (FIG. 6D)
TO FIG. 7
TO FIG. 7

FIG. 6A

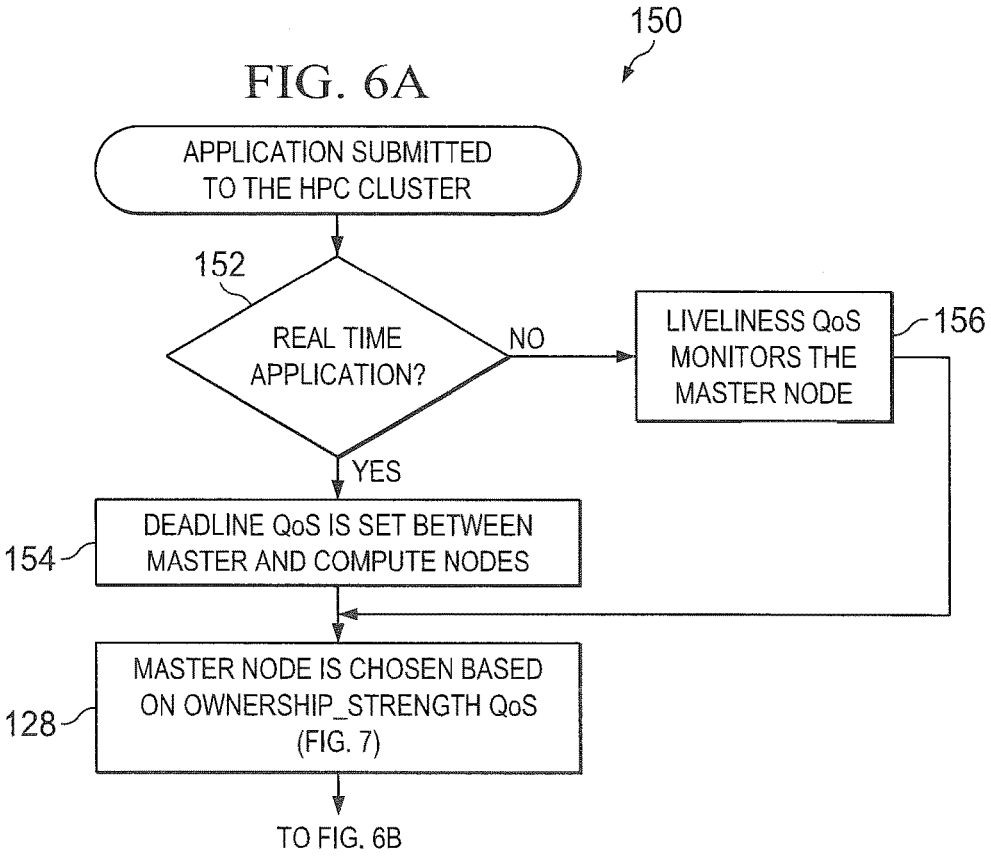
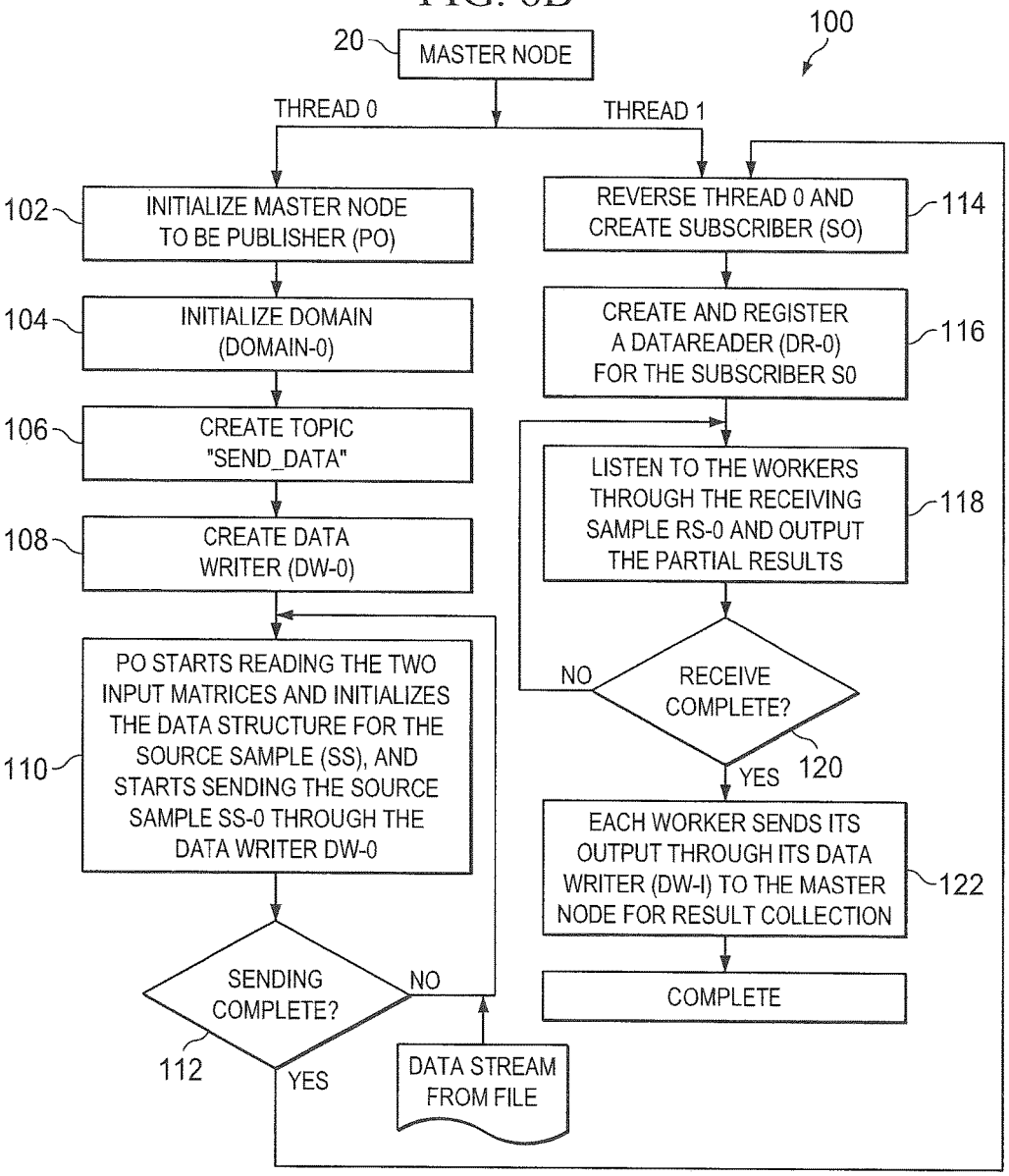

150
APPLICATION SUBMITTED TO THE HPC CLUSTER
152
REAL TIME APPLICATION?
NO
LIVELINESS QoS MONITORS THE MASTER NODE
156
YES
154
DEADLINE QoS IS SET BETWEEN MASTER AND COMPUTE NODES
128
MASTER NODE IS CHOSEN BASED ON OWNERSHIP_STRENGTH QoS (FIG. 7)
TO FIG. 6B

FIG. 6B


100
20
MASTER NODE
THREAD 0
THREAD 1
102
INITIALIZE MASTER NODE TO BE PUBLISHER (PO)
104
INITIALIZE DOMAIN (DOMAIN-0)
106
CREATE TOPIC "SEND_DATA"
108
CREATE DATA WRITER (DW-0)
110
PO STARTS READING THE TWO INPUT MATRICES AND INITIALIZES THE DATA STRUCTURE FOR THE SOURCE SAMPLE (SS), AND STARTS SENDING THE SOURCE SAMPLE SS-0 THROUGH THE DATA WRITER DW-0
SENDING COMPLETE?
NO
YES
112
DATA STREAM FROM FILE
114
REVERSE THREAD 0 AND CREATE SUBSCRIBER (SO)
116
CREATE AND REGISTER A DATAREADER (DR-0) FOR THE SUBSCRIBER S0
118
LISTEN TO THE WORKERS THROUGH THE RECEIVING SAMPLE RS-0 AND OUTPUT THE PARTIAL RESULTS
RECEIVE COMPLETE?
NO
YES
120
122
EACH WORKER SENDS ITS OUTPUT THROUGH ITS DATA WRITER (DW-I) TO THE MASTER NODE FOR RESULT COLLECTION
COMPLETE

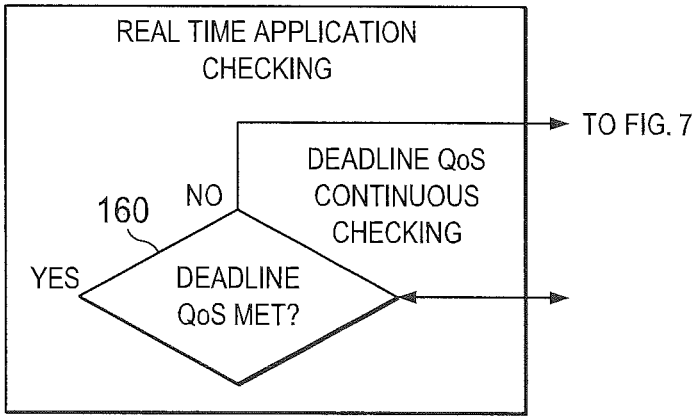

REAL TIME APPLICATION CHECKING
TO FIG. 7
DEADLINE QoS CONTINUOUS CHECKING
NO
160
YES
DEADLINE QoS MET?

FIG. 6C

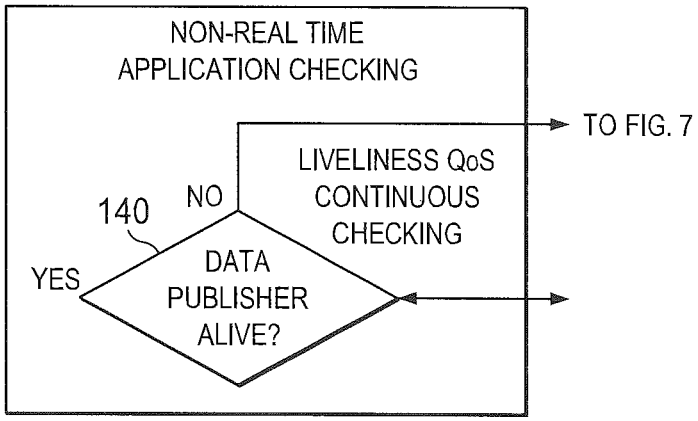

NON-REAL TIME APPLICATION CHECKING
TO FIG. 7
LIVELINESS QoS CONTINUOUS CHECKING
NO
140
YES
DATA PUBLISHER ALIVE?

FIG. 6D

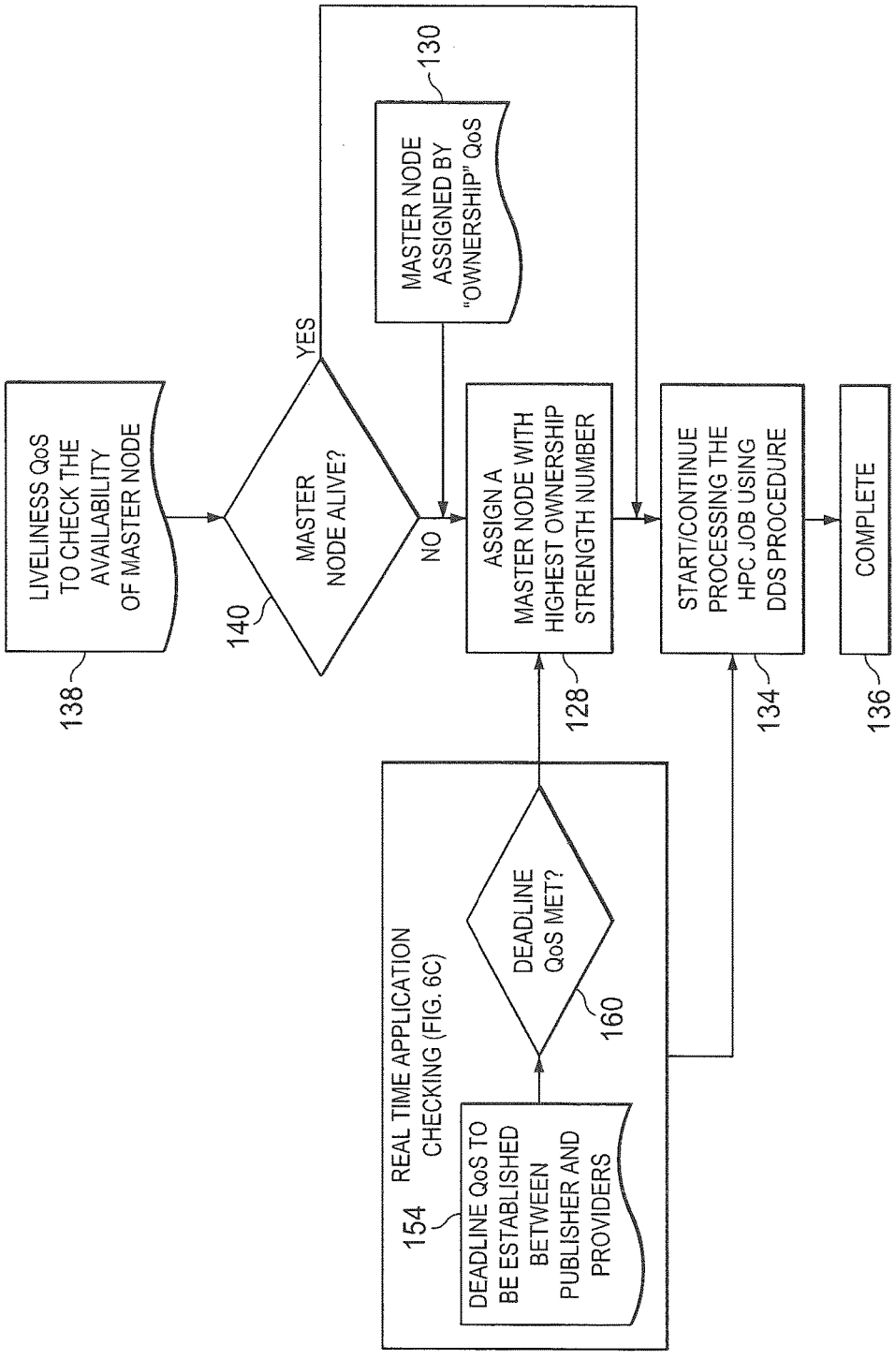

LIVELINESS QoS TO CHECK THE AVAILABILITY OF MASTER NODE
138
MASTER NODE ALIVE?
140
YES
NO
MASTER NODE ASSIGNED BY "OWNERSHIP" QoS
130
ASSIGN A MASTER NODE WITH HIGHEST OWNERSHIP STRENGTH NUMBER
128
START/CONTINUE PROCESSING THE HPC JOB USING DDS PROCEDURE
134
COMPLETE
136
REAL TIME APPLICATION CHECKING (FIG. 6C)
154
DEADLINE QoS TO BE ESTABLISHED BETWEEN PUBLISHER AND PROVIDERS
DEADLINE QoS MET?
160

FIG. 7

FIG. 8

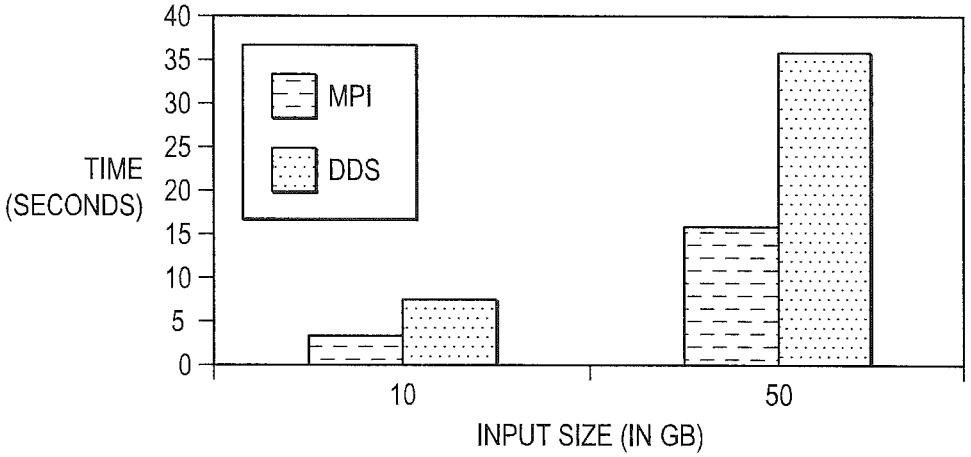
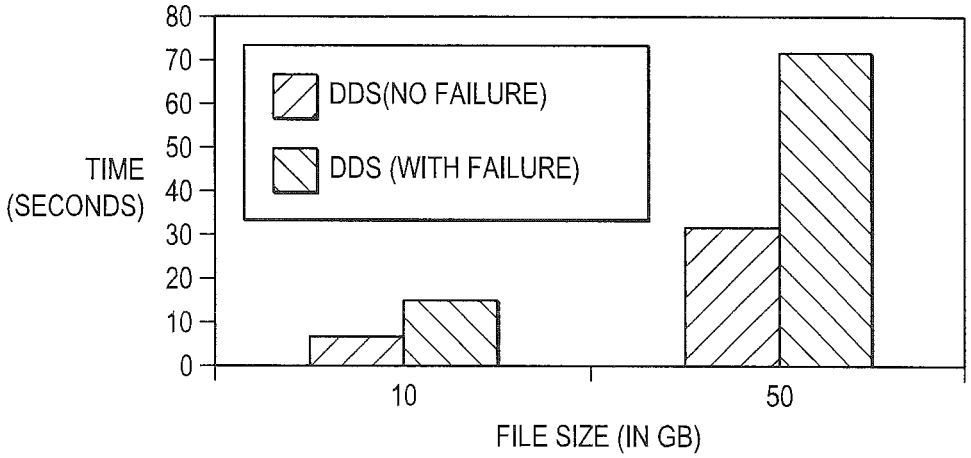

40
35
30
25
20
15
10
5
0
TIME
(SECONDS)
MPI
DDS
10
50
INPUT SIZE (IN GB)

FIG. 9


80
70
60
50
40
30
20
10
0
TIME
(SECONDS)
DDS(NO FAILURE)
DDS (WITH FAILURE)
10
50
FILE SIZE (IN GB)

HIGH PERFORMANCE AND GRID COMPUTING WITH LIVELINESS AND DEADLINES FAULT TOLERANT DATA DISTRIBUTOR QUALITY OF SERVICE

CROSS REFERENCED TO RELATED APPLICATIONS

This application is a continuation-in-part of, and claims priority to: commonly-owned U.S. patent application Ser. No. 14/835,279, titled "High Performance And Grid Computing With Fault Tolerant Data Distributors Quality Of Service Control," filed Aug. 25, 2015 and now U.S. Pat. No. 9,429,677; and to: each of earlier filed, commonly-owned: U.S. patent application Ser. No. 14/490,894, filed Sep. 19, 2014, titled "High Performance and Grid Computing With Reliability Quality of Service Control", now U.S. Pat No. 9,128,211; U.S. patent application Ser. No. 14/490,862, filed Sep. 19, 2014, titled "High Performance and Grid Computing With History Quality of Service Control", now U.S. Pat. No. 9,134,455; and U.S. patent application Ser. No. 13/649,286, filed Oct. 11, 2012, claiming priority to U.S. Provisional Patent Application No. 61/545,766 filed Oct. 11, 2011, and titled, "High Performance and Grid Computing With Quality of Service Control", now issued as U.S. Pat. No. 8,874,804.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to high performance and grid computing of data for exploration and production of hydrocarbons, such as computerized simulation of hydrocarbon reservoirs in the earth, geological modeling, and processing of seismic survey data, and in particular to quality of service (QoS) control of such computing.

2. Description of the Related Art

In the oil and gas industries, massive amounts of data are required to be processed for computerized simulation, modeling and analysis for exploration and production purposes. For example, the development of underground hydrocarbon reservoirs typically includes development and analysis of computer simulation models of the reservoir. These underground hydrocarbon reservoirs are typically complex rock formations which contain both a petroleum fluid mixture and water. The reservoir fluid content usually exists in two or more fluid phases. The petroleum mixture in reservoir fluids is produced by wells drilled into and completed in these rock formations.

A geologically realistic model of the reservoir, and the presence of its fluids, also helps in forecasting the optimal future oil and gas recovery from hydrocarbon reservoirs. Oil and gas companies have come to depend on geological models as an important tool to enhance the ability to exploit a petroleum reserve. Geological models of reservoirs and oil/gas fields have become increasingly large and complex.

In simulation and geological models, the reservoir is organized into a number of individual cells. Seismic data with increasing accuracy has permitted the cells to be on the order of 25 meters areal (x and y axis) intervals. For what are known as giant reservoirs, the number of cells is the least hundreds of millions, and reservoirs of what is known as giga-cell size (a billion cells or more) are encountered.

Similar considerations of data volume are also presented in seismic data processing. Seismic data obtained from surveys over large areas of the earth's surface such as above giant reservoirs, has been acquired and made available in increased volumes. In processing vast amounts of data of all three of the types described above, processing time was an important consideration.

Three types of computer systems have been available for processing the vast amounts of data of the types encountered in petroleum exploration and production. These are supercomputers, high performance computing (HPC) and grid computing. Typically, supercomputers are specially designed for particular calculation intensive tasks. An HPC system takes the form of a group of powerful workstations or servers, joined together as a network to function as one supercomputer. Grid computing involves a more loosely coupled, heterogeneous and often dispersed network of workstations or servers than HPC.

So far as is known, existing distributed memory HPC and grid computing systems did not provide proper quality of service (QoS) based communication because of two limitations. First, standard communication libraries such as Message Passing Interface (MPI) and Parallel Virtual Machine (PVM) did not provide a capability for applications to specify service quality for computation and communication. Second, modern high-speed interconnects such as InfiniBand, Myrinet, Quadrics and Gigabit Ethernet were optimized for performance rather than for predictability of communication latency and bandwidth.

There has been, so far as is known, little attention given to QoS control in high performance and grid computing. HPC users have witnessed a dramatic increase in performance over the last ten years with regard to the HPC systems. What used to take one month of HPC computation time in the ten years ago, is now taking only a few hours to run in current systems.

In view of this, the simplest remedy to users for a data accuracy failure rate during a routine computation run has been resubmitting the processing data run after disregarding or offlining (or what is known as fencing) the problematic node/core. However, the hours of the crashed job were thus discarded and wasted. Moreover, in the case of a more extensive data processing run which would require several days or even weeks to perform, resubmitting the entire data set for processing was required. This was duplicative and time consuming.

U.S. Pat. No. 7,526,418, which is owned by the assignee of the present application, relates to a simulator for giant hydrocarbon reservoirs composed of a massive number of cells. The simulator mainly used high performance computers (HPC). Communication between the cluster computers was performed according to conventional, standard methods, such as MPI mentioned above and Open MP.

U. S. Published Patent Application No. 2011/0138396 related to a data distribution mechanism in HPC clusters. The focus of the system described was methodology to enable data to be distributed rapidly to various computation nodes in an HPC cluster. Thus, the focus and teachings of this system were improving processing speed by more rapidly distributing data to the cluster nodes.

SUMMARY OF THE INVENTION

Briefly, the present invention provides a new and improved computer implemented method of computerized processing in a data processing system of data for exploration and production of hydrocarbons. The data processing system includes a plurality of master nodes, each with an established quality of service standard profile. The quality of service profile includes assigned ownership strength, with the master node of the plurality of master nodes having a

highest assigned ownership strength being designated master publisher node for the exploration and production data being processed. The quality of service profile further includes a specified deadline time interval for periodic production of data by the designated master publisher node when the data being processed is being produced in real time. The data processing system also includes a plurality of processor nodes established as subscribers to receive exploration and production data from the designated master publisher, and a data memory. According to the computer implemented method, the established quality of service standard profile is transmitted from the designated master publisher to the subscriber processor nodes. The designated master publisher establishes a domain for exploration and production processing by the designated master publisher and designated ones of the pluralities of processor nodes as subscriber processor nodes. The exploration and production data is produced from the designated master publisher to the designated subscriber processor nodes of the domain, and a determination made if the data being processed is indicated to be real time data. The transmitted exploration and production data are processed in the designated subscriber processor nodes of the domain, and the processed exploration and production data of the designated subscriber processor nodes is monitored at the designated master publisher of the domain. The processed exploration and production data received at the designated master publisher is assembled in the data memory of the data processing system. Performance of the designated master publisher node is monitored during the processing of the exploration and production data to determine if the specified time interval has passed without periodic production of data from the designated master publisher node. If the specified time interval has not passed, monitoring of performance of the designated master publisher node is continued to determine that the designated master publisher node is operating and processing of the exploration and production data is continued. If the designated master publisher node is not operating, the ownership strength of the others of the plurality of master nodes according to their respective ownership strength quality of service profile is determined. A master node having the highest ownership strength quality of service profile is then established as the designated master publisher node the processing of the exploration and production data is resumed.

The present invention also provides a new and improved data processing system for computerized processing of data for exploration and production of hydrocarbons. The data processing system includes a plurality of master nodes, each with an established quality of service standard profile including assigned ownership strength. The master node of the plurality of master nodes having highest assigned ownership strength is designated master publisher node for the exploration and production data being processed. The quality of service profile also includes a specified deadline time interval for periodic production of data by the designated master publisher node when the data being processed is being produced in real time. The data processing system also includes a plurality of processor nodes and a data memory. In the data processing system, the designated master publisher node transmits the established quality of service standard profile from the designated master publisher to the subscriber processor nodes. The designated master publisher node also establishes a domain for exploration and production processing by the designated master publisher and designated ones of the plurality of processor nodes as subscriber processor nodes, and produces the exploration and production data to the designated subscriber processor nodes of the domain. The master node also determines if the data being processed is indicated to be real time data. The processor nodes established as subscribers receive exploration and production data from the designated master publisher, and receive in the designated subscriber processor nodes of the domain the exploration and production data and the established quality of service standard profile from the designated master publisher node. The designated subscriber processor nodes of the domain process the transmitted exploration and production data. The master node further monitors the processed exploration and production data of the designated subscriber processor nodes of the domain, and assembles in the data memory the processed exploration and production data from the designated subscriber processor nodes which comply with the transmitted established quality of service standard profile. The master node also monitors performance of the designated master publisher node during the processing of the exploration and production data to determine if the specified time interval has passed without periodic production of data from the designated master publisher node. If the specified time interval has not passed, monitoring performance of the designated master publisher node to determine that the designated master publisher node is operating, processing of the exploration and production data are continued. If the designated master publisher node is not operating, ownership strength of the others of the plurality of master nodes is determined according to their respective ownership strength quality of service profile. The master node having the highest ownership strength quality of service profile is established as designated master publisher node and the processing of the exploration and production data is resumed.

The present invention further provides a new and improved data storage device having stored in a computer readable medium computer operable instructions for causing a data processing system to perform computerized processing of data for exploration and production of hydrocarbons. The data processing system includes a plurality of master nodes, each with an established quality of service standard profile including an assigned ownership strength, and with the master node of the plurality of master nodes having a highest assigned ownership strength being designated master publisher node for the exploration and production data being processed. The quality of service profile further includes a specified deadline time interval for periodic production of data by the designated master publisher node when the data being processed is being produced in real time. The data processing system includes a plurality of processor nodes established as subscribers to receive the exploration and production data from the designated master publisher node, and a data memory. The instructions stored in the data storage device causing the data processing system to transmit the established quality of service standard profile from the designated master publisher to the subscriber processor nodes, and to establishing with the designated master publisher a domain for exploration and production processing by the designated master publisher and designated ones of the plurality of processor nodes as subscriber processor nodes. The instructions also cause production of the exploration and production data from the designated master publisher to the designated subscriber processor nodes of the domain. The instructions also cause a determination to be made if the data being processed is indicated to be real time data. The instructions cause processing of the transmitted exploration and production data in the designated subscriber processor nodes of the domain, and moni-

toring at the designated master publisher the processed exploration and production data of the designated subscriber processor nodes of the domain. The instructions also cause assembling in the data memory of the data processing system the processed exploration and production data received at the designated master publisher. The instructions further cause monitoring of performance of the designated master publisher node during the processing of the exploration and production data to determine if the specified time interval has passed without periodic production of data from the designated master publisher node. If the specified time interval has not passed, the instructions cause continuing the step of monitoring performance of the designated master publisher node to determine that the designated master publisher node is operating, and continuing the processing of the exploration and production data. If the designated master publisher node is not operating, the instructions cause ownership strength of the others of the plurality of master nodes to be determined according to their respective ownership strength quality of service profile, and establishing the master node having the highest ownership strength quality of service profile as designated master publisher node, and resuming the processing of the exploration and production data.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a schematic block diagram of a prior art data processing system for high performance computing.

FIG. **2** is a schematic block diagram of a data processing system for high performance and grid computing with quality of service control according to the present invention.

FIG. **3** is a functional block diagram of the data processing system of FIG. **2** configured for high performance processing with quality of service control according to the present invention.

FIG. **4** is a functional block diagram of a set of data processing steps performed in the data processing system of FIGS. **2** and **3** for high performance processing with quality of service control according to the present invention.

FIG. **5** is a functional block diagram of a set of data processing steps performed in the data processing system of FIGS. **2** and **3** for high performance processing with quality of service control according to the present invention.

FIG. **6** is a functional block diagram indicating the functional interrelationship between the interactive operations for quality of service control according to the present invention.

FIGS. **6**A, **6**B, **6**C and **6**D are functional block diagrams indicating in more detail the interactive operation of processors of the data processing system of FIGS. **2** and **3** during performance of the data processing steps of FIGS. **4** and **5**.

FIG. **7** is a functional block diagram indicating the operation of processors of the data processing system of FIGS. **2** and **3** according to the ownership, deadlines and liveliness strength quality of service controls according to the present invention.

FIG. **8** is a plot of runtime for the data processing system shown in FIGS. **2** and **3** for high performance processing with quality of service control with different sizes of input data in comparison with prior art MPI communication protocols.

FIG. **9** is a plot of network time delay in engaging a new processor node for different file sizes in the data processing system of FIGS. **2** and **3**.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention relates to high performance and grid computing of data for exploration and production of hydrocarbons, such as computerized simulation of hydrocarbon reservoirs in the earth, geological modeling, processing of seismic survey data, and other types of data gathered and processed to aid in the exploration and production of hydrocarbons. For the purposes of the present invention data of the foregoing types are referred to herein as exploration and production data. The present invention is particularly adapted for processing exploration and production data where vast amounts of such data are present, such as in or around what are known as giant reservoirs.

In the drawings, FIG. **1** represents an example prior art high performance computing network P. The high performance computing network P is configured for parallel computing using the message passing interface (MPI) with a master node **10** transferring data through what are known as serial heartbeat connections over data links **12** of a management network **14** to a number of processor nodes **16**. The processor nodes **16** are configured to communicate with each other as indicated at **18** according to the message passing interface (MPI) standard communication library during parallel computing and processing of data. As has been set forth, so far as is known, standard communication libraries such as Message Passing Interface (MPI) and Parallel Virtual Machine (PVM) did not provide a capability for applications to specify service quality for computation and communication.

With the present invention, as is shown schematically in FIG. **2** in a data processing system D a plurality of master nodes **20** of a CPU **22** and a group of processor or worker nodes **24** operating as a network arranged for high performance or grid computing, depending on the configuration of the network, of exploration and production data. As will be set forth, the data processing system D processes exploration and production data with a controllable specified quality of service (QoS) for the processing applications. Data processing system D operates according to the processing techniques which are shown schematically in FIGS. **4**, **5**, **6**A through **6**D, and **7**. Thus, high performance computing (HPC) and grid computing processing of exploration and production data are performed without impacting or losing processing time in case of failures. A data distribution service (DDS) standard is implemented in the high performance computing (HPC) and grid computing platforms of the data processing system D, to avoid shortcomings of message passing interface (MPI) communication between computing modules, and provide quality of service (QoS) for such applications.

Considering now the data processing system according to the present invention, as illustrated in FIG. **2**, the data processing system D is provided as a processing platform for high performance computing (HPC) and grid computing of exploration and processing data. The data processing system D includes one or more central processing units or CPU's **22**. The CPU or CPU's **22** have associated therewith a reservoir memory or database **26** for general input parameters, of a type and nature according to the exploration and production data being processed, whether reservoir simulation, geological modeling, seismic data or the like.

A user interface **28** operably connected with the CPU **22** includes a graphical display **30** for displaying graphical images, a printer or other suitable image forming mechanism and a user input device **32** to provide a user access to

manipulate, access and provide output forms of processing results, database records and other information.

The reservoir memory or database **26** is typically in a memory **34** of an external data storage server or computer **38**. The reservoir database **26** contains data including the structure, location and organization of the cells in the reservoir model, data general input parameters, as well as the exploration and production data to be processed, as will be described below.

The CPU or computer **22** of data processing system D includes the master nodes **20** and an internal memory **40** coupled to the master nodes **20** to store operating instructions, control information and to serve as storage or transfer buffers as required. The data processing system D includes program code **42** stored in memory **40**. The program code **42**, according to the present invention, is in the form of computer operable instructions causing the master nodes **20** and processor nodes **24** to transfer the exploration and production data and control instructions back and forth according to data distribution service (DDS) intercommunication techniques, as will be set forth.

It should be noted that program code **42** may be in the form of microcode, programs, routines, or symbolic computer operable languages that provide a specific set of ordered operations that control the functioning of the data processing system D and direct its operation. The instructions of program code **42** may be stored in memory **40** or on computer diskette, magnetic tape, conventional hard disk drive, electronic read-only memory, optical storage device, or other appropriate data storage device having a computer usable medium stored thereon. Program code **42** may also be contained on a data storage device as a computer readable medium.

The processor nodes **24** are general purpose, programmable data processing units programmed to perform the processing of exploration and production data according to the present invention. The processor nodes **24** operate under control of the master node **20** and the processing results obtained are then assembled in memory **34** where the data are provided for formation with user interface **28** of output displays to form data records for analysis and interpretation.

Although the present invention is independent of the specific computer hardware used, an example embodiment of the present invention is preferably based on master nodes **20** and processor nodes **24** of an HP Linux cluster computer. It should be understood, however, that other computer hardware may also be used.

According to the present invention, the data processing system D whose components are shown in FIG. **2** is configured as shown in FIG. **3** according to the data distribution service (DDS) techniques. As indicated in FIG. **3**, a designated one of the master nodes **20** also shown in FIG. **2** is established as a designated master publisher node as indicated at **50** in that such master node **20** is the node responsible for dissemination and distribution of the exploration and production data to be processed by the processor nodes **24**. The designated master publisher node **20** includes a persistence service capability as shown at **51** to preserve data samples so that they can be furnished to processor nodes which replace failed processor nodes, as will be described. The designated master publisher node **20** is also as indicated at **52** in FIG. **3** established as a data writer for the exploration and production data distributed for processing, so that values of the data which is distributed can be established.

The processor nodes **24** in operating according to data distribution service (DDS) are each individually established as subscribers, as indicated at **54** in FIG. **3**. Thus the processor nodes as subscribers are configured to operate as the processors responsible assigned to receive the exploration and production data received as a result of the subscriber relationship to the publisher **50** of the master node **20**. The processor nodes **24** are also configured as data readers as indicated at **56** in FIG. **3**. As data readers **56**, the processor nodes **24** receive an allocated portion of the exploration and production data from the data writer **52** of the master node **20**.

The DDS techniques of the present invention are further explained in the data distribution service (DDS) standard. The DDS standard provides a scalable, platform-independent, and location-independent middleware infrastructure to connect information producers to consumers (i.e. nodes to nodes). DDS also supports many quality-of-service (QoS) policies, such as asynchronous, loosely-coupled, time-sensitive and reliable data distribution at multiple layers (e.g., middleware, operating system, and network).

The DDS methodology implements a publish/subscribe (PS) model for sending and receiving data, events, and commands among the participant or designated master node **20** and processor nodes **24** in the processing of exploration and production data according to the present invention. As will be set forth, the master node serves as a primary publisher and the processor nodes **24** function as the primary subscribers. The processor nodes **24** are also configured to transfer the processed exploration and production data results to the designated master node **20**, and the designated master node **20** is configured for this purpose as a subscriber function. Thus each node of the data processing system D can serve as a publisher, subscriber, or both simultaneously.

Nodes of the data processing system D that are producing information (publishers) create "topics" which are parameters of interest for the data being processed, depending on the type of exploration and production data being processed. The nodes operating in the DDS mode take care of delivering the data samples to those subscribers that indicate an interest in that topic. In a preferred computer network according to the present invention, example implementations of DDS provide low latency messaging (as fast as 65 microseconds between nodes) and high throughput (up to 950 Mbps).

As has been set forth, the present invention processes exploration and production data for giant reservoirs where vast amounts of data need to be processed. An example type of processing performed is two-way data streaming between master and processor nodes. Examples of data streaming usage in exploration and production data for giant reservoirs are U.S. Pat. Nos. 7,596,480; 7,620,534; 7,660,711; 7,526,418; and 7,809,537. It should be understood that the present invention may also be used in connection with communication between nodes for HPC or grid computing of exploration and production data for other types of processing in addition to data streaming. The data communication according to the present invention between nodes for HPC and grid computing may also be used for other types of data, such as bioinformatics processing and computational fluid dynamics. The present invention is adapted for use in processing of large amounts of data which consume considerable time and thus has an increased likelihood of system failure during the course of processing.

FIG. **4** is a functional block diagram of a set **60** of data processing steps performed by the master node **20** in the data processing system D according to the present invention. As shown at step **62** of FIG. **4**, the master node **20** is designated as the master publisher. Master node **20** then spawns two

threads using OpenMP to parallelize two functionalities. In the first thread, the master node **20** initializes during step **64** to be a publisher (P**0**) with selected QoS profile, which is predefined in an XML file. For example, in the data streaming embodiment described below, three main QoS policies are adopted. These are: durability, reliability, and history.

The "durability" of the QoS profile saves the published data topics so that the data topics can be delivered to subscribing nodes that join the system at a later time, even if the publishing node has already terminated. The persistence service can use a file system or a relational data base to save the status of the system.

The second QoS profile or policy, which is reliability, indicates the level of reliability requested by a DataReader or offered by a DataWriter. Publisher nodes may offer levels of reliability, parameterized by the number of past issues they can store for the purpose of retrying transmissions. Subscriber nodes may then request different levels of reliable delivery, ranging from fast-but-unreliable "best effort" to highly reliable in-order delivery. Thus providing per data stream reliability control. In case the reliability type is set to "RELIABLE", the write operation on the DataWriter may be blocked if the modification would cause data to be lost or else cause one of the resource limited to be exceeded.

The third policy, history, controls the behavior of the communication when the value of a topic changes before it is finally communicated to some of its existing DataReader entities. If the type is set to "KEEP_LAST", then the service will only attempt to keep the latest values of the topic and discard the older ones. In this case, a specified value of depth of data retention regulates the maximum number of values the service will maintain and deliver. The default (and most common setting) for depth is one, indicating that only the most recent value should be delivered.

If the history type is set to "KEEP_ALL", then the service will attempt to maintain and deliver all the values of the sent data to existing subscribers. The resources that the service can use to keep this history are limited by the settings of the RESOURCE_LIMITS QoS. If the limit is reached, then the behavior of the service will depend on the RELIABILITY QoS. If the reliability kind is "BEST_EFFORTS", then the old values will be discarded. If the reliability setting is "RELIABLE", then the service will block the DataWriter until it can deliver the necessary old values to all subscribers.

In the data streaming embodiment described herein, "DURABILITY" was used and specified, because it was desirable for compute nodes to continue joining the system whenever there is a failure in another node, and thus avoid the consequences of a node failure during an extended processing run. In the third policy of history, "KEEP_LAST" was selected since the streamed seismic or simulation data are not expected to change. Specifically, the seismic shots and simulation cell values are fixed and not dynamic. In reliability QoS policy, RELIABLE, was specified as all the values need to reach the subscribers and have complete answers of the data streaming. The requirement of data accuracy prohibited missing or losing elements while transmitting.

In high performance and grid computing, multiple data publishers (master nodes) can exist in a cluster in which all of the master nodes function as publishers and have access to the same data source. With the present invention, a methodology is provided to enable fault tolerant continuity of the high performance and grid computing. According to the present invention, the quality of service (QoS) profile includes an OWNERSHIP_STRENGTH quality of service. Each of the data master publisher nodes has an assigned or delegated "OWNERSHIP_STRENGTH" quality of service, where the publisher with the highest ownership strength number is the designated publisher of the data. By default, OWNERSHIP_STRENGTH is assigned based on the time the master node is joining the domain (that is, the cluster). In other words, the first node joining the domain would be assigned the highest OWNERSHIP_STRENGTH integer. This default behavior can be manually changed by a user who may alter the OWNERSHIP_STRENGTH for a specific node (set it higher or lower) if it is preferred to re-arrange the priorities of the master nodes.

There are types of exploration and production data which require that the high performance cluster of data processing system D perform applications with real-time processing. In these situations, a periodic data stream is a main component as an input to the cluster. Examples of such data and applications include GeoSteering applications where the stream of data is processed real time in the cluster to provide processed output to steer drilling bits in real time drilling. Another example is real time reservoir simulation cluster where the processing of data happens in real time and simulation data input observed at periodic time intervals.

In real time processing applications, the data subscribers (compute nodes) of the cluster require and are programmed to expect that the stream of data be received in specific "periodical" time from the master node (Data publishers). This quality of service according to the present invention referred to as a "Deadline" QoS. With the Deadline quality of service or QoS, the present invention provides a capacity for cases where a specified deadline time is passed and data is not received. In such an event, the currently designated Data master publisher node is considered offline, and another data writer (master node) is chosen based on its Ownership Strength quality of service ranking under the Ownership Strength QoS functionality.

However, with the present invention, if the application does not need to be real time processing, the Deadline quality of service is neglected. In such cases, if the data master publisher node crashes, the compute nodes are not required to wait for data to be processed. The present invention still provides fault tolerant QoS, wherein a LIVELINESS quality of service determines the next data publisher to take the function of distributor master node based on its OWNERSHIP_STRENGTH quality of service, as discussed above.

If a master node currently serving as designated master publisher node and functioning as data publisher prematurely terminates or crashes for any reason, then according to the present invention, a substitute or replacement master publisher node with the next highest ownership strength is designated as master publisher node. The replacement master publisher node is then in charge of processing to continue publishing data to subscribers. Thus, the present invention provides fault tolerance between multiple data publishers. Detection of the active master node being prematurely terminated is accomplished as shown in FIG. **7**, as will be described below.

The present invention also provides a "LIVELINESS" or operating status inquiry quality of service profile. According to established selected time intervals, an inquiry is made of the operating status of the master node **20** presently serving as designated master publisher node. The inquiry is made in order to confirm the availability or presently active operating status of the currently serving designated master publisher node. The inquiry happens in the form of "network heartbeats" from the Data Distribution Service (DDS) to data

writer (publisher) that has the strongest (OWNERSHIP) integer. If this publisher does not acknowledge these network heart-beat packets, then the LIVELINESS QoS assigns the next available master node that has the highest OWNERSHIP integer to be the publisher.

As indicated at step **66**, the thread also specifies the domain where all the publishers and subscribers would work on, which is domain-**0** in the present embodiment. Next, as indicated at step **68**, a topic with an identifying name is specified and a DataWriter (DW-**0**) is initialized during step **70** under P**0** using that topic. After that, the master node **20** as publisher begins reading the data matrices from input to be processed, and initializes the data structure for the source sample (SS) by defining the matrices dimension and the number of processor nodes **24** which are designated as processor or worker nodes for the processing of the exploration and production data. The master node then during step **72** starts sending the Source sample SS-**0** to the designated worker processor nodes **24** through the DataWriter DW-**0**.

The second thread of master node **20** reverses the function of thread **0** by creating an instance of a subscriber S**0** as indicated at step **74** with the selected QoS profile in Domain-**0**, in preparation to receive the partial results from the worker nodes **24** (designated worker nodes **24** act as subscribers at the beginning and then as publishers at the end of their processing). The master node **20** then listens to the worker nodes **24** to receive processing results as indicated at **76** through the receiving sample RS-**0** and verifies compliance with the selected QoS profile as indicated at step **78**. The master node **20** checks the QoS profile while it is receiving data from the worker nodes. It only receives data from those working nodes **24** which are matching in their QoS. It will not receive and will not negotiate with other worker nodes which have incompatible QoS settings. The master node **20** then during step **80** stores the verified processing results in data memory, and the stored results are available for output and display.

FIG. **5** is a functional block diagram of a set **84** of data processing steps performed by the processor nodes **24** in the data processing system D according to the present invention. On the subscribers' side (i.e., the workers), each node **24** during step **86** initiates itself as a subscriber to the main publisher P**0**, assigns an ID to itself (Wi), and during step **88** starts receiving the data for computational processing. The distribution of which data goes to which node **24** is done dynamically in a way that is determined by first identifying the data range taken by each node according to the format of the data. As indicated at step **90**, the QoS checking is done during the nodes' receiving process. The data is then processed during step **92** according to the processing required. Each worker node **24** during step **94** then sends its output with verified QoS through its DataWriter (DW-i) to the master node **20** for result collection. As indicated at step **96**, QoS checking is done during the sending of data in step **94**. Thus, when there is communication between the publisher and subscriber, QoS checking is done to establish this connection.

FIG. **6** is a functional block diagram indicating the functional interrelationship between the interactive operations shown in more detail in FIGS. **6**A through **6**D, respectively, for quality of service control according to the present invention. As shown in FIG. **6**, FIG. **6**B is a functional block diagram **100** illustrating the interaction of the data processing nodes according to the present invention during the processing of exploration and production data. FIG. **6**A is a functional block diagram **150** of an initial set-up methodology performed prior to the actions shown in diagram **100**. FIG. **6**C and FIG. **6**D are functional block diagrams of alternative control of the interactions of the data processing depending on the Quality of Service profile which is indicated to be in effect during the initial set-up.

Thus, as shown in FIG. **6**A, prior to step **102**, the initial set-up methodology **150** is performed. The data processing application is submitted to the HPC cluster of the data processing system D as shown at **150** in FIG. **6**A. As indicated at step **152**, a determination is made whether the data processing application is identified as one requiring real time processing. If real time processing is indicated during step **152**, as indicated at step **154**, the above described Deadline Quality of Service is established between the master node(s) **20** as published and the processor nodes **24**. For the Deadline Quality of Service, a specified time is identified within which the currently designated master publisher node **20** must publish data to the worker nodes **24**.

If it is determined during step **152** that real time processing is not indicated, then the Liveliness Quality of Service is established during step **156** between the master publisher node(s) **20** and the processor nodes **24**. As a result of either of steps **154** or **156**, processing proceeds to the step **128** (FIGS. **6**A and **7**) for selection of one of the nodes **20** as designated master node and designated master publisher as a result of having the highest ranking Ownership Strength QoS profile.

FIG. **6**B is a diagram **100** illustrating the interaction of the master node **20** selected during step **128** (FIG. **6**A and **7**) and worker processor nodes **24** in an example implementation of a data streaming processing of exploration and production data. The implementation starts at step **102** by designating the master node **20** of the cluster as the main publisher.

After being so designated, the master node **20**, in turn, spawns two threads using OpenMP to parallelize its two main functions: initializing the node to be a publisher (P**0**) with the selected QoS profile; and specifying during step **104** the domain which the publishers and subscribers are to work on, which is domain-**0** in the example implementation. Specifying the domain is necessary in order to allow multiple groups of publishers and subscribers to work independently, segmenting the cluster into several smaller subclusters, if needed. Different algorithms may require different topics (i.e. datasets) to be sent independently by the same publisher, and each of these topics may have several DataWriters for redundancy.

Next, during step **106**, a topic with the name "SEND_DATA" is created and a DataWriter (DW-**0**) is initialized during step **108** under P**0** using the created topic. The reason for this hierarchy is that different topics (i.e. datasets) may be required to be sent independently by the same publisher, and each of these topics may have several DataWriters for redundancy. After that, the publisher during step **110** starts reading the seismic shots or simulation cells data from input, and initializes the data structure for the source sample (SS) by defining the matrices data size and the number of workers. The publisher then during step **110** starts sending the source sample SS-**0** through the DataWriter DW-**0**. The procedure continues until step **112** indicates all sending has been completed. Processing then is continued at step **114**.

As indicated at step **114**, the second thread from master node **20** reverses the function of thread **0** by creating an instance of a subscriber S**0** with selected QoS profile in Domain-**0**, in preparation to receive the partial results from the worker or processor nodes **24**. Specifically, a DataReader (DR-**0**) is configured during step **116** at master node **20** for the subscriber S**0** (the workers) that uses topic "RECV_RE-

SULT". Master node **20** then as indicated at step **118** listens to the workers through the receiving sample RS-**0** and outputs the partial results. This processing continues as indicated at step **120** until receiving of sample RS-**0** is completed.

On the subscribers' side (i.e., the workers), during step **122** each worker node **24** as discussed above and as shown in FIG. **5** initiates itself as a subscriber to the main publisher P**0** , assigns an ID to itself (Wi), and starts receiving the seismic or simulation data for processing and subsequent result collection. The distribution of which data goes to which node is done dynamically in a way that is determined by first identifying the data size taken by each node according to the format of the data.

In case of a node failure of a processor node **24** on the workers' side, the system administrator of master node **20** may initiate a new processor node **24** with the same ID of the failed worker. The new worker would read the written checkpointed status as defined in the policy, re-read the sample from the persistence service **51**, and resume the operation of the system.

It is important to mention that as a requirement for the durability QoS, all sent topics require DataWriters to match the configuration of the persistent QoS policy configuration with the DataReaders. As a consequence, a DataWriter that has an incompatible QoS with respect to what the topic specified will not send its data to the persistent service, and thus its status will not be saved. Similarly, a DataReader that has an incompatible QoS with respect to the specified in the topic will not get data from it.

FIGS. **6**C and **6**D indicate the methodology according to the present invention of providing either Deadline Quality of Service for real time applications checking of real time data. (FIG. **6**C) or Liveliness Quality of Service (FIG. **6**D) for non-real time application checking of data. As mentioned, the selection is based on the determination made during step **152** (FIG. **6**A). During the data stream processing **100** shown in FIG. **6**B, either Deadline or Liveliness Quality of Service is continuously monitored according to the results of step **152**.

If Deadline Quality of Service is so indicated, as shown at step **160** in FIG. **6**C, a determination is continuously made whether the Deadline QoS is met. If the deadline QoS is indicated not to be met during step **160**, a specified time has passed during which real time data has not been published from master publisher node **20**. If so, as indicated in FIG. **6**C, step **128** (FIG. **7**) is involved and a replacement master node **20** is substituted based on the highest Ownership QoS among the available master nodes **20**, as will be described. If during the periodic Deadline QoS processing shown in FIG. **6**C real time data is being provided continuously by the master publisher node **20** according to the Deadline QoS and thus, within the specified period, the processing during step **100** of FIG. **6**C continues uninterrupted.

If during step **152**, it is determined that the Liveliness QoS is to be in effect, a determination is continuously made during step **140** (FIGS. **6**D and **7**) whether the master publisher node is performing properly. If so, processing during step **100** continues uninterrupted as shown in FIG. **6**D. If not, a replacement master node **20** is substituted during step **128** based on the highest Ownership QoS among the available other master nodes **20**, as will be described in connection with FIG. **7**.

Thus FIG. **6** as well as FIGS. **6**A through **6**D illustrate an example parallel processing often encountered with exploration and production data according to the present invention. The results were as expected: QoS is controlled with fault-tolerance enabled when using the DDS techniques as adopted middleware in such computer processing of exploration and production data. Specifically, compute nodes on the cluster are turned off and back on again, and the jobs continued to run with no need for a restart.

FIG. **7** illustrates the role of the Ownership QoS, the Liveliness QoS, and the Deadline QoS in choosing one of the available master nodes **20** in the cluster. The master node **20** with the highest "OWNERSHIP_STRENGTH" number is chosen as the designated master publisher node **20** as indicated at step **128** based on an "ownership strength QoS" assigned as indicated at step **130**. The assignment during step **130** is according to the criteria described.

Under control of the designated master publisher node **20**, processing of the exploration and production data continues to process exploration and production data, as indicated as step **134** in the manner shown in FIG. **5** until the processing job is completed, as indicated at step **136**, so that the processed data can be sent to the publisher as shown at step **94** (FIG. **5**).

Periodically, based on the "Liveliness" quality of service profile, established in the manner set forth above, as indicated at step **138**, an inquiry is made as indicated at step **140** of the operating status of the designated master publisher node **50**. If it is determined during step **140** that the designated master publisher node **50** is not operating or functioning, a master node **20** having the next highest ownership strength number is introduced at the designated master publisher node according to step **128** and processing of the exploration and production data resumed in step **134**. If during step **140**, the present designated master publisher node **50** is indicated operable and functioning, processing according to step **134** continues.

The "LIVELINESS" QoS service according to the present invention thus continuously checks the health of the master node **50** in the absence of the Deadline QoS service being in effect. This is done through sending status inquiry or heartbeat signals. If the node is unavailable, the "OWNERSHIP" QoS **130** is instructed to assign the next available master node with the highest "OWNERSHIP_STRENGTH" to continue the job. It is important to note that all available master nodes have access to the same data source. This is to allow them all to continue the stopped process.

As also shown in FIG. **7**, if the Deadline QoS service is invoked for real time processing, and such data is not being timely received, a replacement master node is substituted. As indicated at step **160** in FIGS. **6**C and **7**, if the deadline time specified for the Deadline is not met, step **128** is also invoked to assign the available master node **20** with the highest ownership strength number. Thereafter, step **134** is implemented in the manner described above.

The present invention provides the ability to control QoS properties on HPC and Grids that affect predictability, overhead, resource utilization, and aligns the scarce resources to the most critical requirements to these jobs.

To evaluate the performance of the present invention over conventional HPC and compare it with processing using MPI, data streaming was performed on the same data set using both paradigms (i.e., DDS with the present invention and MPI) and evaluated them on the data processing system clusters of FIGS. **1** and **2**. The data streaming processing algorithm is computationally intensive with iterations, and it was chosen since it is a fundamental operation in many numerical linear algebra applications used in processing of exploration and production data. An efficient implementa-

tion on parallel computers is an issue of prime importance when providing such systems for processing of exploration and production data.

FIG. **8** shows benchmarks to test the scalability and runtime of MPI and DDS by streamlining Reservoir Simulation data. It can be seen that the MPI version outperformed in terms of speed by taking around 6.76 seconds to stream 10 GB size of file, compared with 7.9 seconds using DDS. This is expected since the QoS is adding more computations and checks to the communication level. As has been mentioned, however, MPI and PVM are focused on processing speed, with no effort to monitor accuracy or possible processing deficiencies.

FIG. **9** shows the delay in engaging a new node according to the present invention, replacing a crashed node, while using the persistent service and durability, reliability and history QoS. This test is not applicable to a prior art MPI implementation. Since MPI implementations do not provide a capability of specifying service quality for computation or communication. As indicated in FIG. **9**, the delay in engaging a new node is proportional to the size of the matrices, since the persistent service needs to resend all the previously published instances to this new node. During the benchmark testing, test results depicted in FIG. **8** indicate that it took 15.2 seconds in a 10 GB data stream between nodes, while it took 72.2 seconds in a 50 GB test. This is to be compared with the time required when it was necessary to resubmit the entire data set for processing for data in large quantities.

The present invention thus adds several benefits of having quality of service in high performance computing that are not available in the traditional method (i.e. by using MPI as a middleware). Among the benefits are that periodic publishers can indicate the speed at which they can publish by offering guaranteed update deadlines. By setting a deadline, a compliant publisher promises to send a new update at a minimum rate. Subscribers may then request data at that or any slower rate.

Another benefit is that the continuing participation or activity of entities can be monitored. The selected Liveliness and Durability QoS offered with the present invention determine whether an entity or a node is "active" (i.e., alive). The application can also be informed via a listener when an entity is no longer responsive. A further benefit is that the present invention also permits the data processing system D to automatically arbitrate between multiple publishers of the same topic with a parameter called "strength." Subscribers receive from the strongest active publisher. This provides automatic failover; if a strong publisher fails, all subscribers immediately receive updates from the backup (weaker) publisher.

It is also to be noted that QoS parameters exist with the DDS employment to control the resources of the entire system, suggest latency budgets, set delivery order, attach user data, prioritize messages, set resource utilization limits and partition the system into namespaces. The present invention thus provides the ability to control QoS properties on HPC and grids that affect predictability, overhead, resource utilization, and align the computational resources to the most critical requirements.

The present invention is a feasible option for those applications in which QoS is considered a priority, or for those HPC batch jobs that would run for several days on commodity hardware, where the probability of failure is not negligible. Accordingly, the present invention provides the ability to control QoS properties on HPC and grids that affect predictability, overhead, resource utilization, and aligns the scarce resources to the most critical requirements.

The invention has been sufficiently described so that a person with average knowledge in the matter may reproduce and obtain the results mentioned in the invention herein Nonetheless, any skilled person in the field of technique, subject of the invention herein, may carry out modifications not described in the request herein, to apply these modifications to a determined structure, or in the manufacturing process of the same, requires the claimed matter in the following claims; such structures shall be covered within the scope of the invention.

It should be noted and understood that there can be improvements and modifications made of the present invention described in detail above without departing from the spirit or scope of the invention as set forth in the accompanying claims.

What is claimed is:

1. A computer implemented method of computerized processing in a data processing system of data for exploration and production of hydrocarbons, the data processing system including a plurality of master nodes, each with an established quality of service standard profile including an assigned ownership strength, with the master node of the plurality of master nodes having a highest assigned ownership strength being designated master publisher node for the exploration and production data being processed, the quality of service standard profile further including a specified deadline time interval for periodic production of data by the designated master publisher node when the data being processed is being produced in real time, the data processing system further including a plurality of processor nodes established as subscribers to receive exploration and production data from the designated master publisher node, and a data memory, the method comprising the computer processing steps of:

(a) transmitting the established quality of service standard profile from the designated master publisher node to the processor nodes;

(b) establishing with the designated master publisher node a domain for exploration and production data processing by the designated master publisher node and designated ones of the plurality of processor nodes as subscriber processor nodes;

(c) producing the exploration and production data from the designated master publisher node to the designated subscriber processor nodes of the domain;

(d) determining if the data being processed is indicated to be real time data;

(e) processing the received exploration and production data in the designated subscriber processor nodes of the domain;

(f) monitoring at the designated master publisher node the processed exploration and production data of the designated subscriber processor nodes of the domain;

(g) assembling in the data memory of the data processing system the processed exploration and production data received at the designated master publisher node;

(h) monitoring performance of the designated master publisher node during the processing of the exploration and production data to determine if the specified deadline time interval has passed without periodic production of data from the designated master publisher node; and

(i) if the specified deadline time interval has not passed, continuing the step of monitoring performance of the designated master publisher node to determine that the

designated master publisher node is operating, continuing the processing of the exploration and production data; and,

(j) if the designated master publisher node is not operating, determining ownership strength of the others of the plurality of master nodes according to their respective ownership strength quality of service standard profile; and

(k) establishing as designated master publisher node the master node having the highest ownership strength quality of service standard profile and resuming the processing of the exploration and production data.

2. The computer implemented method of claim **1**, wherein the established quality of service standard profile further includes a liveliness quality of service control to periodically determine the operating status of the designated master publisher node, and wherein the step of determining if the data being processed is indicated to be real time data, and further including the steps of:

(a) monitoring performance of the designated master publisher node during the processing of the exploration and production data to determine operating status of the designated master publisher node;

(b) if the step of monitoring performance of the designated master publisher node indicates the designated master publisher node is operating, continuing the processing of the exploration and production data; and, if not

(c) determining ownership strength of the others of the plurality of master nodes according to their respective ownership strength quality of service standard profile; and

(d) establishing as designated master publisher node the master node having the highest ownership strength quality of service standard profile and resuming the processing of the exploration and production data.

3. The computer implemented method of claim **1**, wherein the data processing system further includes a data display, and further including the computer processing step of:

forming an output display of the assembled processed exploration and production data.

4. The computer implemented method of claim **1**, wherein the exploration and production data comprises a reservoir simulation model.

5. The computer implemented method of claim **1**, wherein the exploration and production data comprises a geological model of a reservoir.

6. The computer implemented method of claim **1**, wherein the exploration and production data comprises a seismic survey of the earth in a reservoir.

7. The computer implemented method of claim **1**, further including the steps of:

(a) determining in the designated master publisher node whether the designated subscriber processor nodes of the domain comply with the transmitted established quality of service standard profile from the designated master publisher node; and

(b) if so, receiving at the designated master publisher node the processed exploration and production data from the designated subscriber processor nodes which comply with the transmitted established quality of service standard profile; and

(c) if not, inhibiting at the designated master publisher node transfer to the designated master publisher of the processed exploration and production data from the designated subscriber processor nodes which do not comply with the transmitted established quality of service standard profile;

(d) joining additional processor nodes as designated subscriber processor nodes of the domain during processing of exploration and production data; and

(e) delivering the produced exploration and production data to the joined additional processor nodes.

8. A data processing system for computerized processing of data for exploration and production of hydrocarbons, the data processing system comprising a plurality of master nodes, each with an established quality of service standard profile including an assigned ownership strength, with the master node of the plurality of master nodes having a highest assigned ownership strength being designated master publisher node for the exploration and production data being processed, the quality of service standard profile further including a specified deadline time interval for periodic production of data by the designated master publisher node when the data being processed is being produced in real time, a plurality of processor nodes and a data memory, the data processing system further comprising:

(a) the designated master publisher node performing the steps of:

(1) transmitting the established quality of service standard profile from the designated master publisher node to the processor nodes;

(2) establishing with the designated master publisher node a domain for exploration and production data processing by the designated master publisher node and designated ones of the plurality of processor nodes as subscriber processor nodes;

(3) producing the exploration and production data from the designated master publisher node to the designated subscriber processor nodes of the domain;

(4) determining if the data being processed is indicated to be real time data;

(b) the plurality of processor nodes established as subscribers to receive exploration and production data from the designated master publisher node, the plurality of processor nodes performing the steps of:

(1) receiving in the designated subscriber processor nodes of the domain the exploration and production data and the established quality of service standard profile from the designated master publisher node;

(2) processing the received exploration and production data in the designated subscriber processor nodes of the domain; and

(c) the designated master publisher node further performing the steps of:

(1) monitoring the processed exploration and production data of the designated subscriber processor nodes of the domain;

(2) assembling in the data memory the processed exploration and production data from the designated subscriber processor nodes which comply with the transmitted established quality of service standard profile;

(3) monitoring performance of the designated master publisher node during the processing of the exploration and production data to determine if the specified deadline time interval has passed without periodic production of data from the designated master publisher node;

(4) if the specified deadline time interval has not passed, continuing the step of monitoring performance of the designated master publisher node to

determine that the designated master publisher node is operating, continuing the processing of the exploration and production data;

(5) if the designated master publisher node is not operating, determining ownership strength of the others of the plurality of master nodes according to their respective ownership strength quality of service standard profile; and

(6) establishing as designated master publisher node the master node having the highest ownership strength quality of service standard profile and resuming the processing of the exploration and production data.

9. The data processing system of claim **8**, wherein the established quality of service standard profile further includes a liveliness quality of service control to periodically determine the operating status of the designated master publisher node, and wherein the step of determining if the data being processed is indicated to be real time data, and wherein the processor nodes further performs the steps of:

(a) monitoring performance of the designated master publisher node during the processing of the exploration and production data to determine operating status of the designated master publisher node;

(b) if the step of monitoring performance of the designated master publisher node indicates the designated master publisher node is operating, continuing the processing of the exploration and production data; and, if not

(c) determining ownership strength of the others of the plurality of master nodes according to their respective ownership strength quality of service standard profile; and

(d) establishing as designated master publisher node the master node having the highest ownership strength quality of service standard profile and resuming the processing of the exploration and production data.

10. The data processing system of claim **8** further including a data display, and wherein the master node further performs the step of:

forming an output display of the assembled processed exploration and production data.

11. The data processing system of claim **8**, wherein the exploration and production data comprises a reservoir simulation model.

12. The data processing system of claim **8**, wherein the exploration and production data comprises a geological model of a reservoir.

13. The data processing system of claim **8**, wherein the exploration and production data comprises a seismic survey of the earth in a reservoir.

14. The data processing system of claim **8**, wherein the processor nodes further perform the steps of:

(1) determining whether the designated subscriber processor nodes of the domain comply with the transmitted established quality of service standard profile from the designated master publisher node;

(2) if so, receiving the processed exploration and production data from the designated subscriber processor nodes which comply with the transmitted established quality of service standard profile;

(3) if not, inhibiting at the designated master publisher node transfer of the processed exploration and production data from the designated subscriber processor nodes which do not comply with the transmitted established quality of service standard profile;

(4) joining additional processor nodes as designated subscriber processor nodes of the domain during processing of exploration and production data;

(5) delivering the produced exploration and production data to the joined additional processor nodes.

15. A data storage device having stored in a non-transitory computer readable storage medium computer operable instructions for causing a data processing system to perform computerized processing of data for exploration and production of hydrocarbons, the data processing system including a plurality of master nodes, each with an established quality of service standard profile including an assigned ownership strength, and with the master node of the plurality of master nodes having a highest assigned ownership strength being designated master publisher node for the exploration and production data being processed, the quality of service standard profile further including a specified deadline time interval for periodic production of data by the designated master publisher node when the data being processed is being produced in real time, the data processing system including a plurality of processor nodes established as subscribers to receive the exploration and production data from the designated master publisher node, and a data memory, the instructions stored in the data storage device causing the data processing system to perform the following steps:

(a) transmitting the established quality of service standard profile from the designated master publisher node to the processor nodes;

(b) establishing with the designated master publisher node a domain for exploration and production data processing by the designated master publisher node and designated ones of the plurality of processor nodes as subscriber processor nodes;

(c) producing the exploration and production data from the designated master publisher node to the designated subscriber processor nodes of the domain;

(d) determining if the data being processed is indicated to be real time data;

(e) processing the received exploration and production data in the designated subscriber processor nodes of the domain;

(f) monitoring at the designated master publisher node the processed exploration and production data of the designated subscriber processor nodes of the domain;

(g) assembling in the data memory of the data processing system the processed exploration and production data received at the designated master publisher node;

(h) monitoring performance of the designated master publisher node during the processing of the exploration and production data to determine if the specified deadline time interval has passed without periodic production of data from the designated master publisher node; and

(i) if the specified time deadline interval has not passed, continuing the step of monitoring performance of the designated master publisher node to determine that the designated master publisher node is operating, continuing the processing of the exploration and production data; and,

(j) if the designated master publisher node is not operating, determining ownership strength of the others of the plurality of master nodes according to their respective ownership strength quality of service standard profile; and

(k) establishing as designated master publisher node the master node having the highest ownership strength

quality of service standard profile and resuming the processing of the exploration and production data.

16. The data storage device of claim **15**, wherein the established quality of service standard profile further includes a liveliness quality of service control to periodically determine the operating status of the designated master publisher node, and wherein the data being processed is indicated to be real time data, and further include instructions to perform the steps of:

(a) monitoring performance of the designated master publisher node during the processing of the exploration and production data to determine operating status of the designated master publisher node;

(b) if the step of monitoring performance of the designated master publisher node indicates the designated master publisher node is operating, continuing the processing of the exploration and production data; and, if not

(c) determining ownership strength of the others of the plurality of master nodes according to their respective ownership strength quality of service standard profile; and

(d) establishing as designated master publisher node the master node having the highest ownership strength quality of service standard profile and resuming the processing of the exploration and production data.

17. The data storage device of claim **15**, wherein the data processing system further includes a data display, and further including instructions for performing the step of:

forming an output display of the assembled processed exploration and production data.

18. The data storage device of claim **15**, wherein the exploration and production data comprises a reservoir simulation model.

19. The data storage device of claim **15**, wherein the exploration and production data comprises a geological model of a reservoir.

20. The data storage device of claim **15**, wherein the exploration and production data comprises a seismic survey of the earth in a reservoir.

21. The data storage device of claim **15**, further including instructions for performing the steps of:

(a) determining in the designated master publisher node whether the designated subscriber processor nodes of the domain comply with the transmitted established quality of service standard profile from the designated master publisher node; and

(b) if so, receiving at the designated master publisher node the processed exploration and production data from the designated subscriber processor nodes which comply with the transmitted established quality of service standard profile; and

(c) if not, inhibiting at the designated master publisher node transfer to the designated master publisher node of the processed exploration and production data from the designated subscriber processor nodes which do not comply with the transmitted established quality of service standard profile;

(d) joining additional processor nodes as designated subscriber processor nodes of the domain during processing of exploration and production data; and

(e) delivering the produced exploration and production data to the joined additional processor nodes.

* * * * *