



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК

G06F 12/00 (2019.08); G06F 9/00 (2019.08); G06F 17/00 (2019.08); G06F 3/06 (2019.08)

(21)(22) Заявка: 2018144176, 13.12.2018

(24) Дата начала отсчета срока действия патента:
13.12.2018

Дата регистрации:
14.02.2020

Приоритет(ы):

(22) Дата подачи заявки: 13.12.2018

(45) Опубликовано: 14.02.2020 Бюл. № 5

Адрес для переписки:

119021, Москва, ул. Льва Толстого, 16,
патентный отдел

(72) Автор(ы):

Станкевичус Алексей Алексеевич (RU),
Трифонов Сергей Владимирович (RU)

(73) Патентообладатель(и):

ОБЩЕСТВО С ОГРАНИЧЕННОЙ
ОТВЕТСТВЕННОСТЬЮ "ЯНДЕКС" (RU)

(56) Список документов, цитированных в отчете

о поиске: US 5845316 A1, 01.12.1998. US

9304694 B2, 05.04.2016. EP 1508850 A3,

18.03.2015. RU 2547705 C2, 10.04.2015.

(54) Способ и система для планирования выполнения операций ввода/вывода

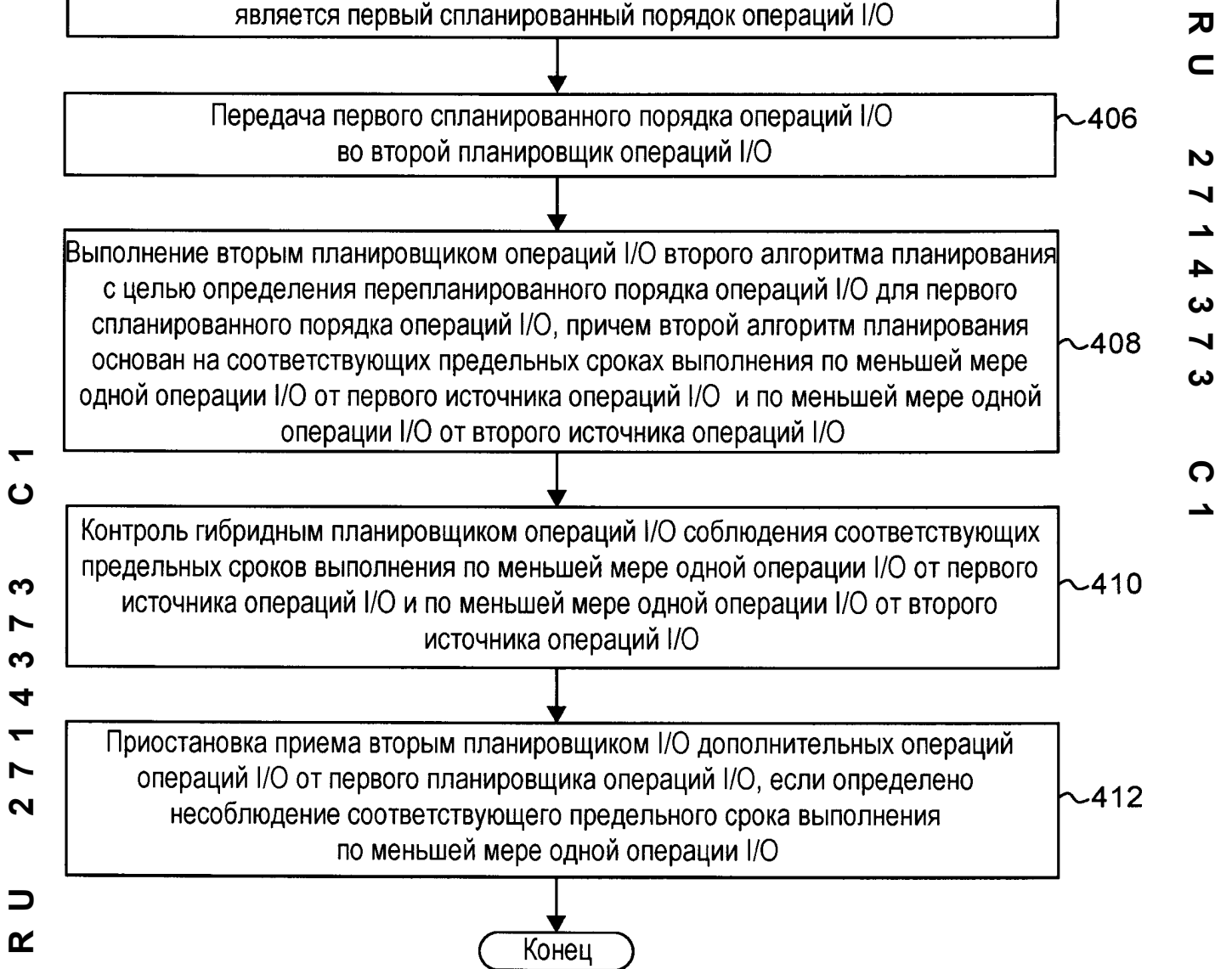
(57) Реферат:

Группа изобретений относится к средствам планирования операций ввода/вывода (I/O) в накопителях. Технический результат – обеспечение возможности сбалансированного управления накопителем за счет планирования операций I/O. Для этого описан гибридный планировщик и способ выполнения этим планировщиком операций I/O. Гибридный планировщик содержит два планировщика. Способ включает в себя прием операций I/O от двух источников операций I/O и выполнение первым планировщиком первого алгоритма для определения спланированного порядка операций I/O на основе производительности накопителя.

Способ также включает в себя передачу этого порядка операций I/O во второй планировщик, который выполняет второй алгоритм для определения перепланированного порядка операций I/O на основе соответствующих предельных сроков выполнения. Способ также включает в себя контроль соблюдения соответствующих предельных сроков выполнения операций I/O и приостановку приема дополнительных операций I/O вторым планировщиком от первого планировщика, если определено несоблюдение соответствующего предельного срока выполнения операции I/O. 2 н. и 28 з.п. ф-лы, 4 ил.

RU 2 714 373 C1

RU 2 714 373 C1



Фиг. 4



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(12) **ABSTRACT OF INVENTION**

(52) CPC

G06F 12/00 (2019.08); G06F 9/00 (2019.08); G06F 17/00 (2019.08); G06F 3/06 (2019.08)(21)(22) Application: **2018144176, 13.12.2018**

(24) Effective date for property rights:
13.12.2018

Registration date:
14.02.2020

Priority:

(22) Date of filing: **13.12.2018**(45) Date of publication: **14.02.2020** Bull. № 5

Mail address:

**119021, Moskva, ul. Lva Tolstogo, 16, patentnyj
otdel**

(72) Inventor(s):

**Stankevichus Aleksey Alekseevich (RU),
Trifonov Sergey Vladimirovich (RU)**

(73) Proprietor(s):

**OBSHCHESTVO S OGRANICHENNOJ
OTVETSTVENNOSTYU "YANDEKS" (RU)**

(54) **METHOD AND SYSTEM FOR SCHEDULING EXECUTION OF INPUT/OUTPUT OPERATIONS**

(57) Abstract:

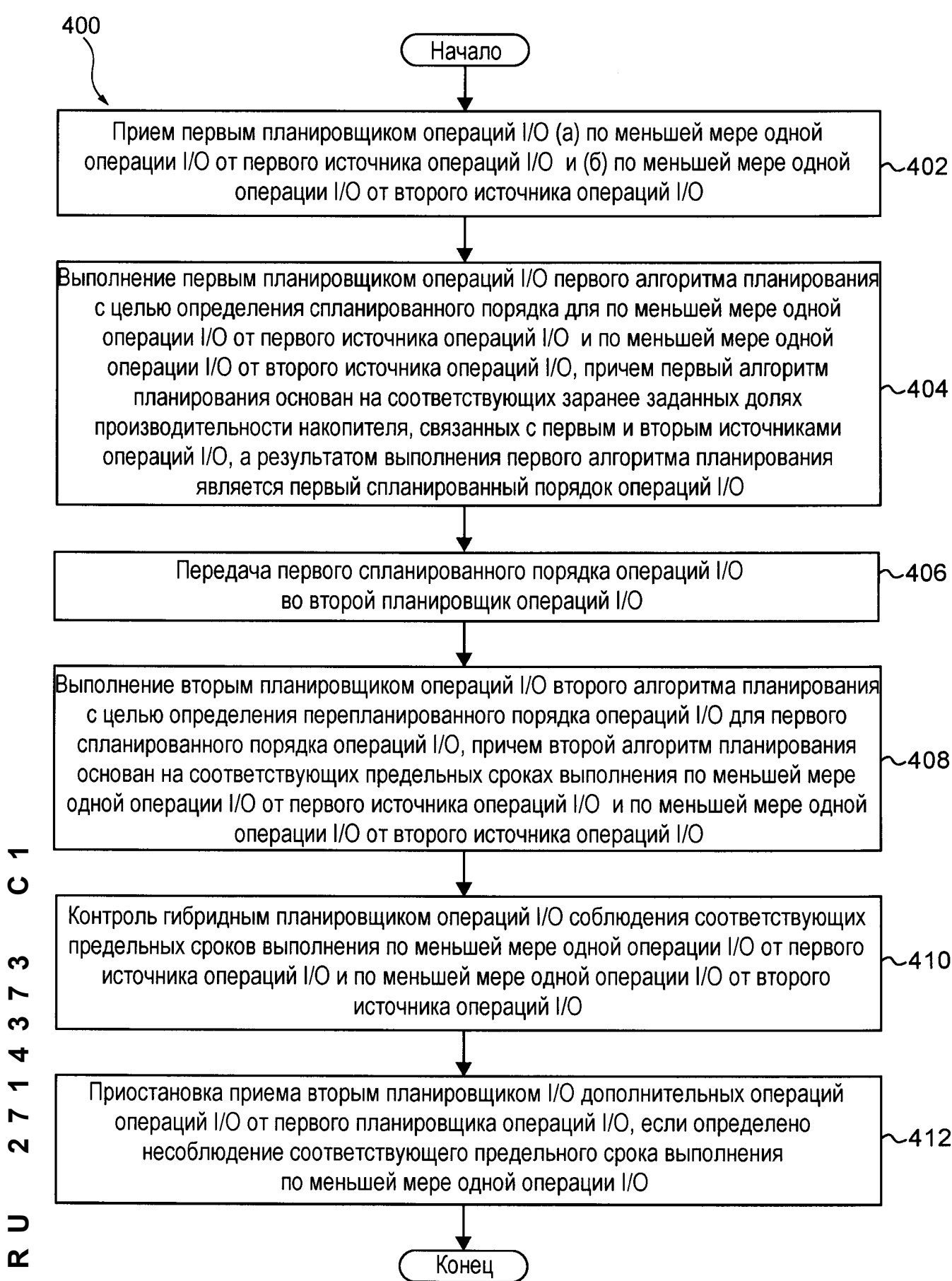
FIELD: computer equipment.

SUBSTANCE: group of inventions relates to devices for scheduling input/output (I/O) operations in storage units. A hybrid scheduler and a method of executing I/O by said scheduler are described. Hybrid scheduler comprises two schedulers. Method includes receiving I/O operations from two I/O operation sources and executing the first algorithm by the first scheduler to determine the planned I/O sequence based on the storage capacity. Method also includes transmitting that order of I/O operations to a second scheduler, which

executes a second algorithm to determine the rescheduled order of I/O operations based on the corresponding execution time limits. Method also includes monitoring compliance with the respective I/O execution time limits and stopping the receipt of additional I/O operations by the second scheduler from the first scheduler if non-compliance with the corresponding I/O execution deadline is detected.

EFFECT: providing possibility of balanced control of accumulator due to planning of I/O operations.

30 cl, 4 dwg



Фиг. 4

Область техники

Данная технология относится к распределенной обработке данных и, в частности, к способам и системам для планирования выполнения операций ввода/вывода.

Уровень техники

- 5 Требования к запоминающим устройствам для хранения цифровых данных постоянно повышаются, поскольку большие объемы цифровых данных создаются ежедневно. Например, хранения могут требовать пользовательские данные, данные организаций и/или данные приложений различных видов. Это повышает требования к емкости хранилищ данных. Такие потребности пользователей и/или организаций способны
- 10 удовлетворить «облачные» системы удаленного хранения данных.

- В общем случае «облачное» хранилище данных представляет собой вид компьютерного запоминающего устройства, в котором цифровые данные хранятся в логических пулах. Физическая память, в которой фактически хранятся цифровые данные, распределяется между несколькими серверами, возможно, расположенными в различных
- 15 местах (т.е. в различных центрах обработки данных), и обычно управляется компанией, предоставляющей услуги «облачного» хранения данных. Пользователи и/или организации обычно покупают или арендуют некоторый объем памяти у поставщиков услуг «облачного» хранения данных, чтобы хранить свои цифровые данные. Со своей стороны, поставщики услуг «облачного» хранения данных отвечают за доступность
- 20 цифровых данных и защиту физической памяти с целью предотвращения утраты данных.

Раскрытие изобретения

Разработчики данной технологии обнаружили определенные технические недостатки, связанные с существующими решениями для обеспечения доступности и предотвращения утраты данных.

- 25 В соответствии с первым аспектом данной технологии реализован способ планирования операций ввода/вывода (I/O, Input/Output) для их выполнения накопителем. Накопитель представляет собой часть системы распределенной компьютерной обработки информации, содержащей множество источников операций I/O, от которых поступают операции I/O. Множество источников операций I/O содержит первый источник
- 30 операций I/O и второй источник операций I/O. Первый источник операций I/O и второй источник операций I/O связаны с соответствующими заранее заданными долями производительности накопителя для выполнения соответствующих операций I/O, поступивших от первого источника операций I/O и второго источника операций I/O. Система распределенной компьютерной обработки информации содержит гибридный
- 35 планировщик операций I/O. Гибридный планировщик операций I/O содержит первый планировщик операций I/O и второй планировщик операций I/O. Способ выполняется гибридным планировщиком операций I/O. Способ включает в себя прием первым планировщиком операций I/O (а) по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и (б) по меньшей мере одной операции I/O,
- 40 сформированной вторым источником операций I/O. Способ включает в себя выполнение первым планировщиком операций I/O первого алгоритма планирования с целью определения спланированного порядка для по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O. Первый алгоритм планирования
- 45 основан на соответствующих заранее заданных долях производительности накопителя, связанных с первым источником операций I/O и вторым источником операций I/O. Результатом выполнения первого алгоритма планирования является первый спланированный порядок операций I/O.

Способ включает в себя передачу первого спланированного порядка операций I/O во второй планировщик операций I/O. Способ также включает в себя выполнение вторым планировщиком операций I/O второго алгоритма планирования с целью определения перепланированного порядка операций I/O для первого спланированного
 5 порядка операций I/O. Второй алгоритм планирования основан на соответствующих предельных сроках выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O. Способ включает в себя контроль гибридным планировщиком операций I/O соблюдения соответствующих предельных
 10 сроков выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O. Способ включает в себя приостановку приема вторым планировщиком операций I/O дополнительных операций I/O от первого планировщика операций I/O, если выявлено несоблюдение соответствующего предельного срока
 15 выполнения по меньшей мере одной операции I/O.

В некоторых вариантах осуществления способа контроль соблюдения соответствующих предельных сроков выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O, включает в себя
 20 контроль вторым планировщиком операций I/O возможности выполнения до истечения соответствующих предельных сроков по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O.

В некоторых вариантах осуществления способа определение несоблюдения соответствующего предельного срока выполнения по меньшей мере одной операции I/O включает в себя определение вторым планировщиком операций I/O невозможности выполнения накопителем до истечения соответствующего предельного срока по меньшей
 25 мере одной операции I/O, передаваемой в накопитель согласно перепланированному порядку операций I/O.

В некоторых вариантах осуществления способа определение несоблюдения соответствующего предельного срока выполнения по меньшей мере одной операции I/O включает в себя определение вторым планировщиком операций I/O невозможности выполнения накопителем до истечения соответствующего предельного срока по меньшей
 30 мере одной операции I/O, если от первого планировщика операций I/O принимается
 35 дополнительная операция I/O.

В некоторых вариантах осуществления способа контроль соблюдения соответствующих предельных сроков выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O, включает в себя
 40 контроль гибридным планировщиком операций I/O фактического выполнения до истечения соответствующих предельных сроков по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O.

В некоторых вариантах осуществления способа определение несоблюдения соответствующего предельного срока выполнения по меньшей мере одной операции I/O включает в себя определение гибридным планировщиком операций I/O фактического невыполнения накопителем до истечения соответствующего предельного срока по
 45 меньшей мере одной операции I/O.

В некоторых вариантах осуществления способа приостановка приема включает в себя прекращение приема дополнительных операций I/O вторым планировщиком операций I/O.

5 В некоторых вариантах осуществления способа приостановка приема включает в себя передачу вторым планировщиком операций I/O в первый планировщик операций I/O сигнала обратной связи, при получении которого первый планировщик операций I/O прекращает отправку дополнительных операций I/O во второй планировщик операций I/O.

10 В некоторых вариантах осуществления способ дополнительно включает в себя прием первым планировщиком операций I/O по меньшей мере одной дополнительной операции I/O, сформированной первым источником операций I/O, и прием первым планировщиком операций I/O по меньшей мере одной дополнительной операции I/O, сформированной вторым источником операций I/O.

15 В некоторых вариантах осуществления способ дополнительно включает в себя возобновление вторым планировщиком операций I/O приема дополнительных операций I/O от первого планировщика операций I/O.

В некоторых вариантах осуществления способа возобновление приема выполняется, если гибридный планировщик операций I/O определяет, что дополнительные операции I/O, принятые вторым планировщиком операций I/O, выполняются с соблюдением
20 соответствующих предельных сроков.

В некоторых вариантах осуществления способа первый планировщик операций I/O представляет собой планировщик на основе циклической диспетчеризации по дефициту (DRR, Deficit Round Robin).

25 В некоторых вариантах осуществления способа второй планировщик операций I/O представляет собой планировщик на основе диспетчеризации по ближайшему предельному сроку завершения (EDF, Earliest Deadline First).

В некоторых вариантах осуществления способа гибридный планировщик операций I/O работает в обычном режиме, когда второй планировщик операций I/O принимает дополнительные операции I/O от первого планировщика операций I/O, или в режиме
30 высокой нагрузки, когда второй планировщик операций I/O приостанавливает прием дополнительных операций I/O от первого планировщика операций I/O.

В некоторых вариантах осуществления способа в обычном режиме работы гибридный планировщик операций I/O обеспечивает «справедливое» распределение
35 производительности накопителя для первого и второго источников операций I/O и выполнение требования первого и второго источников операций I/O относительно режима реального времени. В режиме высокой нагрузки гибридный планировщик операций I/O обеспечивает «справедливое» распределение производительности накопителя для первого и второго источников операций I/O.

40 Согласно второму аспекту данной технологии реализован гибридный планировщик операций I/O для планирования операций I/O с целью их выполнения накопителем, представляющим собой часть системы распределенной компьютерной обработки информации, содержащей множество источников операций I/O, от которых поступают операции I/O. Множество источников операций I/O содержит первый источник операций I/O и второй источник операций I/O. Первый источник операций I/O и второй источник
45 операций I/O связаны с соответствующими заранее заданными долями производительности накопителя для выполнения соответствующих операций I/O, поступающих от первого источника I/O и второго источника I/O. Система распределенной компьютерной обработки информации включает в себя гибридный

планировщик операций I/O, содержащий первый планировщик операций I/O и второй планировщик операций I/O. Гибридный планировщик операций I/O выполнен с возможностью приема первым планировщиком I/O (а) по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и (б) по меньшей мере одной операции I/O, сформированной вторым источником операций I/O. Гибридный планировщик операций I/O выполнен с возможностью выполнения первым планировщиком операций I/O первого алгоритма планирования с целью определения спланированного порядка для по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O. Первый алгоритм планирования основан на соответствующих заранее заданных долях производительности накопителя, связанных с первым источником операций I/O и вторым источником операций I/O. Результатом выполнения первого алгоритма планирования является первый спланированный порядок операций I/O. Гибридный планировщик операций I/O выполнен с возможностью передачи первого спланированного порядка операций I/O во второй планировщик операций I/O. Гибридный планировщик операций I/O выполнен с возможностью выполнения вторым планировщиком операций I/O второго алгоритма планирования с целью определения перепланированного порядка операций I/O для первого спланированного порядка операций I/O. Второй алгоритм планирования основан на соответствующих предельных сроках выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O. Гибридный планировщик операций I/O выполнен с возможностью контроля соблюдения соответствующих предельных сроков выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O. Гибридный планировщик операций I/O выполнен с возможностью приостановки приема вторым планировщиком операций I/O дополнительных операций I/O от первого планировщика операций I/O, если выявлено несоблюдение соответствующего предельного срока выполнения по меньшей мере одной операции I/O.

В некоторых вариантах осуществления способность гибридного планировщика операций I/O контролировать соблюдение соответствующих предельных сроков выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O, включает в себя способность контролировать возможность выполнения до истечения соответствующих предельных сроков по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O.

В некоторых вариантах осуществления способность гибридного планировщика операций I/O определять превышение соответствующего предельного срока выполнения по меньшей мере одной операции I/O включает в себя способность определять невозможность выполнения накопителем до истечения соответствующего предельного срока по меньшей мере одной операции I/O, передаваемой в накопитель согласно перепланированному порядку операций I/O.

В некоторых вариантах осуществления способность гибридного планировщика операций I/O определять превышение соответствующего предельного срока выполнения по меньшей мере одной операции I/O, включает в себя способность определять невозможность выполнения накопителем до истечения соответствующего предельного

срока по меньшей мере одной операции I/O, если от первого планировщика операций I/O принимается дополнительная операция I/O.

В некоторых вариантах осуществления способность гибридного планировщика операций I/O контролировать соблюдение соответствующих предельных сроков выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O, включает в себя способность контролировать фактическое выполнение до истечения соответствующих предельных сроков по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O.

В некоторых вариантах осуществления способность гибридного планировщика операций I/O определять превышение соответствующего предельного срока выполнения по меньшей мере одной операции I/O включает в себя способность определять фактическое невыполнение накопителем до истечения соответствующего предельного срока по меньшей мере одной операции I/O.

В некоторых вариантах осуществления способность гибридного планировщика операций I/O приостанавливать прием включает в себя способность прекращать прием дополнительных операций I/O вторым планировщиком операций I/O.

В некоторых вариантах осуществления способность гибридного планировщика операций I/O приостанавливать прием включает в себя способность второго планировщика операций I/O передавать сигнал обратной связи в первый планировщик операций I/O, при получении которого первый планировщик операций I/O прекращает отправку дополнительных операций I/O во второй планировщик операций I/O.

В некоторых вариантах осуществления гибридного планировщика операций I/O он выполнен с возможностью приема первым планировщиком операций I/O по меньшей мере одной дополнительной операции I/O, сформированной первым источником операций I/O, и приема первым планировщиком операций I/O по меньшей мере одной дополнительной операции I/O, сформированной вторым источником операций I/O.

В некоторых вариантах осуществления гибридного планировщика операций I/O он дополнительно выполнен с возможностью возобновления приема вторым планировщиком операций I/O дополнительных операций I/O от первого планировщика операций I/O.

В некоторых вариантах осуществления гибридного планировщика операций I/O он выполнен с возможностью возобновления приема, если гибридный планировщик операций I/O определяет, что дополнительные операции I/O, принятые вторым планировщиком операций I/O, выполняются с соблюдением соответствующих предельных сроков.

В некоторых вариантах осуществления гибридного планировщика операций I/O первый планировщик операций I/O представляет собой планировщик типа DRR.

В некоторых вариантах осуществления гибридного планировщика операций I/O второй планировщик операций I/O представляет собой планировщик типа EDF.

В некоторых вариантах осуществления гибридного планировщика операций I/O он работает в обычном режиме, если второй планировщик операций I/O принимает дополнительные операции I/O от первого планировщика операций I/O, и в режиме высокой нагрузки, если второй планировщик операций I/O приостанавливает прием дополнительных операций I/O от первого планировщика операций I/O.

В некоторых вариантах осуществления гибридного планировщика операций I/O в обычном режиме работы он обеспечивает «справедливое» распределение

производительности накопителя для первого и второго источников операций I/O и выполнение требований первого и второго источников операций I/O относительно режима реального времени. В режиме высокой нагрузки гибридный планировщик операций I/O обеспечивает «справедливое» распределение производительности накопителя для первого и второго источников операций I/O.

В контексте настоящего описания термин «сервер» означает компьютерную программу, выполняемую соответствующими аппаратными средствами и способную принимать запросы (например, от клиентских устройств) через сеть и выполнять эти запросы или инициировать их выполнение. Аппаратные средства могут представлять собой один физический компьютер или одну компьютерную систему и это не существенно для данной технологии. В настоящем контексте выражение «сервер» не означает, что каждая задача (например, принятая команда или запрос) или некоторая определенная задача принимается, выполняется или запускается одним и тем же сервером (т.е. одними и теми же программными и/или аппаратными средствами); это выражение означает, что любое количество программных средств или аппаратных средств может принимать, отправлять, выполнять или запускать выполнение любой задачи или запроса либо результатов любых задач или запросов; все эти программные и аппаратные средства могут представлять собой один сервер или несколько серверов, причем оба эти случая подразумеваются в выражении «по меньшей мере один сервер».

В контексте настоящего описания термин «клиентское устройство» означает любое компьютерное аппаратное средство, способное выполнять программы, подходящие для решения поставленной задачи. Таким образом, некоторые (не имеющие ограничительного характера) примеры клиентских устройств включают в себя персональные компьютеры (настольные, ноутбуки, нетбуки и т.п.), смартфоны и планшеты, а также сетевое оборудование, такое как маршрутизаторы, коммутаторы и шлюзы. Следует отметить, что в данном контексте устройство, функционирующее как клиентское устройство, также может функционировать как сервер для других клиентских устройств. Использование выражения «клиентское устройство» не исключает использования нескольких клиентских устройств для приема, отправки, выполнения или инициирования выполнения любой задачи или запроса либо результатов любых задач или запросов, либо шагов любого описанного здесь способа.

В контексте настоящего описания выражение «информация» включает в себя информацию любого рода или вида, допускающую хранение в базе данных. Таким образом, информация включает в себя аудиовизуальные произведения (изображения, фильмы, звукозаписи, презентации и т.д.), данные (данные о местоположении, числовые данные и т.д.), текст (мнения, комментарии, вопросы, сообщения и т.д.), документы, электронные таблицы, списки слов и т.д., но не ограничивается ими.

В контексте настоящего описания значение выражения «компонент» включает в себя программное обеспечение (подходящее для определенных аппаратных средств), необходимое и достаточное для выполнения соответствующей определенной функции (или функций).

В контексте настоящего описания выражение «пригодный для использования в компьютере носитель информации» означает носители любого типа и вида, включая ОЗУ, ПЗУ, диски (CD-ROM, DVD, гибкие диски, жесткие диски и т.д.), USB-накопители, твердотельные накопители, накопители на магнитных лентах и т.д.

В контексте настоящего описания числительные «первый», «второй», «третий» и т.д. используются лишь для указания различия между существительными, к которым они относятся, но не для описания каких-либо определенных взаимосвязей между этими

существительными. Например, должно быть понятно, что использование терминов «первый сервер» и «третий сервер» не подразумевает какого-либо определенного порядка, вида, хронологии, иерархии или классификации, в данном случае, серверов, и что их использование (само по себе) не подразумевает обязательного наличия «второго сервера» в любой ситуации. Кроме того, как встречается в настоящем описании в другом контексте, ссылка на «первый» элемент и «второй» элемент не исключает того, что эти два элемента могут быть одним и тем же реальным элементом. Таким образом, например, в некоторых случаях «первый» сервер и «второй» сервер могут представлять собой одно и то же программное и/или аппаратное средство, а в других случаях - различные программные и/или аппаратные средства.

Каждый вариант осуществления данной технологии имеет отношение к по меньшей мере одной из вышеупомянутых целей и/или одному из вышеупомянутых аспектов, но не обязательно ко всем ним. Должно быть понятно, что некоторые аспекты данной технологии, связанные с попыткой достижения вышеупомянутой цели, могут не соответствовать этой цели и/или могут соответствовать другим целям, явным образом здесь не упомянутым.

Дополнительные и/или альтернативные признаки, аспекты и преимущества вариантов осуществления данной технологии содержатся в дальнейшем описании, в приложенных чертежах и в формуле изобретения.

Краткое описание чертежей

Дальнейшее описание приведено для лучшего понимания данной технологии, а также других аспектов и их признаков, и должно использоваться совместно с приложенными чертежами.

На фиг. 1 представлена схема системы, пригодной для реализации вариантов осуществления данной технологии, не имеющих ограничительного характера.

На фиг. 2 представлено соответствующее некоторым вариантам осуществления данной технологии устройство для хранения данных подсистемы распределенного хранения данных с фиг. 1.

На фиг. 3 изображен гибридный планировщик для планирования операций I/O, предназначенных для выполнения накопителем с фиг. 1.

На фиг. 4 приведена блок-схема способа, выполняемого гибридным планировщиком с фиг. 3 и реализованного согласно вариантам осуществления данной технологии, не имеющим ограничительного характера.

Осуществление изобретения

Дальнейшее подробное описание представляет собой лишь описание примеров, иллюстрирующих данную технологию. Это описание не предназначено для определения объема или границ данной технологии.

Кроме того, если в некоторых случаях примеры модификаций не описаны, это не означает, что модификации невозможны и/или что описание содержит единственный вариант осуществления определенного аспекта данной технологии. Кроме того, следует понимать, что настоящее подробное описание в некоторых случаях касается упрощенных вариантов осуществления данной технологии, и что такие варианты представлены для того, чтобы способствовать лучшему ее пониманию. Различные варианты осуществления данной технологии могут быть значительно сложнее.

На фиг. 1 представлена система 100 распределенной компьютерной обработки информации или сокращенно система 100 распределенной обработки информации. Система 100 распределенной обработки информации позволяет реализовать не имеющие ограничительного характера варианты осуществления данной технологии. Очевидно,

что система 100 распределенной обработки информации приведена лишь в качестве иллюстративного варианта осуществления данной технологии. Таким образом, дальнейшее описание системы представляет собой описание примеров, иллюстрирующих данную технологию. Это описание не предназначено для определения объема или

5 границ данной технологии.

В некоторых случаях также приводятся полезные примеры модификаций системы 100 распределенной обработки информации. Они способствуют пониманию, но также не определяют объем или границы данной технологии. Эти модификации не составляют исчерпывающего списка. Специалисту в данной области должно быть понятно, что

10 возможны и другие модификации. Кроме того, если в некоторых случаях примеры модификаций не описаны, это не означает, что модификации невозможны и/или что описание содержит единственный вариант осуществления определенного элемента данной технологии. Специалисту в данной области должно быть понятно, что это не так. Кроме того, следует понимать, что система 100 распределенной обработки

15 информации в некоторых случаях может представлять собой упрощенную реализацию данной технологии, и что такие варианты представлены, чтобы способствовать лучшему ее пониманию. Специалисту в данной области должно быть понятно, что различные варианты осуществления данной технологии могут быть значительно сложнее.

Система 100 распределенной обработки информации содержит источник 102 запросов, сеть 103 связи, подсистему 104 предварительной обработки запросов, подсистему 105

20 обработки транзакций, подсистему 106 маршрутизации транзакций, подсистему 108 распределенного хранения данных, подсистему 110 базы данных и операционную подсистему 111.

Далее описана реализация указанных выше компонентов системы 100 распределенной

25 обработки информации в соответствии с различными не имеющими ограничительного характера вариантами осуществления данной технологии.

Источник запросов

Источник 102 запросов может представлять собой электронное устройство, связанное с конечным пользователем (например, клиентское устройство), или любую другую

30 подсистему системы 100 распределенной обработки информации, способную обеспечивать пользовательские запросы к системе 100 распределенной обработки информации. Несмотря на то, что на фиг. 1 показан лишь один источник 102 запросов, система 100 распределенной обработки информации может содержать несколько источников 102 запросов. Как показано в данном документе, источник 102 запросов

35 входит в состав системы 100 распределенной обработки информации. Тем не менее, в некоторых вариантах осуществления данной технологии источник 102 запросов может быть внешним по отношению к системе 100 распределенной обработки информации и может подключаться с использованием линии связи (не обозначена).

На практике типичный вариант осуществления системы 100 распределенной

40 обработки информации может включать в себя большое количество источников 102 запросов (сотни, тысячи, миллионы и т.д.).

В некоторых вариантах осуществления данной технологии, когда система 100 распределенной обработки информации используется в среде бизнес-потребитель (B2C, Business-to-Customer), источник 102 запросов может представлять собой клиентское

45 устройство, такое как смартфон, связанное с пользователем системы 100 распределенной обработки информации. Например, система 100 распределенной обработки информации может обеспечивать услуги «облачного» хранения данных для клиентского устройства определенного пользователя.

В некоторых вариантах осуществления данной технологии, когда система 100 распределенной обработки информации используется в среде бизнес-бизнес (B2B, Business-to-Business), источник 102 запросов может представлять собой подсистему, такую как удаленный сервер, обеспечивающую пользовательские запросы к системе 100 распределенной обработки информации. Например, в некоторых вариантах осуществления данной технологии система 100 распределенной обработки информации может обеспечивать услуги отказоустойчивой обработки и/или хранения данных для оператора такой подсистемы.

В целом, независимо от реализации системы 100 распределенной обработки данных в виде системы B2C или B2B (или любого другого варианта системы), источник 102 запросов может представлять собой клиентское устройство или другую подсистему, которые могут быть внутренними или внешними по отношению к системе 100 распределенной обработки информации.

Как указано выше, источник 102 запросов способен выдавать множество запросов 180, каждый из которых далее называется запросом 180. Характер запроса 180 зависит от вида источника 102 запросов. В частности, один из примеров запроса 180 представляет собой запрос, сформированный на языке структурированных запросов (SQL, Structured Query Language). Поэтому предполагается, что в некоторых вариантах осуществления данной технологии запрос 180 может быть сформирован на языке декларативного программирования, т.е. запрос 180 может представлять собой запрос декларативного вида.

В общем случае декларативное программирование относится к построению структуры и элементов компьютерных программ, которые отражают логику вычислений без описания потока управления. Распространенные языки декларативного программирования включают в себя SQL, XQuery и другие языки запросов к базам данных. В целом, запрос декларативного типа определяет действие в выражениях «что требуется выполнить», а не «как это требуется выполнить».

Это означает, что запрос декларативного типа может быть связан с условием выполнения определенного действия. В качестве условия, например, может быть указано, в каком элементе должно выполняться это действие или где следует получать значения для выполнения этого действия.

В качестве не имеющих ограничительного характера примеров можно привести следующие запросы декларативного вида: «Вставить значение 5 в ячейку, связанную с ключом, который равен значению ячейки, связанной с ключом А» и «Для всех ключей, связанных с ячейкой, имеющей значение 5, заменить это значение на значение 10». Тем не менее, должно быть понятно, что представленные выше примеры декларативных языков и примеры запросов декларативного вида приведены исключительно для лучшего понимания, и что другие декларативные языки и другие запросы декларативного вида могут использоваться источником 102 запросов в пределах объема данной технологии.

В некоторых вариантах осуществления данной технологии источник 102 запросов также способен принимать множество ответов 181, каждый из которых далее называется ответом 181. В общем случае в ответ на запрос 180, обработанный или не обработанный системой 100 распределенной обработки информации, система 100 распределенной обработки информации может формировать ответ 181, предназначенный для источника 102 запросов, связанного с соответствующим запросом 180. Характер ответа 181, среди прочего, зависит от вида источника 102 запросов, вида соответствующего запроса 180 и от того, обработала или не обработала система 100 распределенной обработки

информации соответствующий запрос 180. В некоторых вариантах осуществления данной технологии система 100 распределенной обработки информации может формировать ответ 181, только если не удалось обработать запрос, или только в случае успешной обработки запроса, или в обоих случаях.

5 Например, во время обработки запроса 180 система 100 распределенной обработки информации может быть способна запрашивать дополнительные данные от источника 102 запросов для продолжения или завершения обработки запроса 180. В этом случае система 100 распределенной обработки информации может быть способна формировать
10 ответ 181 в виде сообщения с запросом данных, в котором указываются дополнительные данные, запрашиваемые системой 100 распределенной обработки информации для продолжения или завершения обработки запроса 180.

В другом примере, если система 100 распределенной обработки информации успешно обработала соответствующий запрос 180, система 100 распределенной обработки информации может быть способна формировать ответ 181 в виде сообщения об
15 успешном завершении, оповещающего об успешной обработке соответствующего запроса 180.

В другом примере, если системе 100 распределенной обработки информации не удалось успешно обработать соответствующий запрос 180, система 100 распределенной
20 обработки информации может быть способна формировать ответ 181 в виде сообщения о неудачном завершении, оповещающий о том, что соответствующий запрос 180 обработать не удалось. В этом случае источник 102 запроса может быть способен выполнять дополнительные действия, такие как повторное направление запроса 180, выполнение диагностического контроля для определения причины неудачного
25 завершения обработки запроса 180 системой 100 распределенной обработки информации, направление нового запроса, предназначенного для системы 100 распределенной обработки информации, и т.д.

Сеть связи

Источник 102 запросов связан с сетью 103 связи для направления запроса 180 в систему 100 распределенной обработки информации и для приема ответа 181 от системы
30 100 распределенной обработки информации. В некоторых не имеющих ограничительного характера вариантах осуществления данной технологии в качестве сети 103 связи может использоваться сеть Интернет. В других не имеющих ограничительного характера вариантах осуществления данной технологии сеть 103 связи может быть реализована иначе, например, в виде любой глобальной сети связи, локальной сети связи, частной
35 сети связи и т.п. Реализация линии связи (отдельно не обозначена) между источником 102 запросов и сетью 103 связи зависит, среди прочего, от реализации источника 102 запросов.

Лишь в качестве примера, не имеющего ограничительного характера, в тех вариантах осуществления данной технологии, где источник 102 запросов реализован как
40 беспроводное устройство связи (такое как смартфон), линия связи может быть реализована как беспроводная линия связи (такая как канал сети связи 3G, канал сети связи 4G, Wireless Fidelity или сокращенно WiFi[®], Bluetooth[®] и т.п.). В тех примерах, где источник 102 запросов реализован как удаленный сервер, линия связи может быть
45 беспроводной (такой как Wireless Fidelity или сокращенно WiFi[®], Bluetooth[®] и т.д.) или проводной (такой как соединение по сети Ethernet).

Следует отметить, что сеть 103 связи, среди прочего, способна передавать пакет данных запроса, содержащий запрос 180, из источника 102 запросов в подсистему 104 предварительной обработки запросов системы 100 распределенной обработки

информации. Например, этот пакет данных запроса может содержать выполняемые компьютером команды, составленные на языке программирования декларативного типа, на котором представлен запрос 180. Сеть 103 связи, среди прочего, также способна передавать пакет данных ответа, содержащий ответ 181, из системы 100 распределенной
 5 обработки информации в источник 102 запросов. Например, этот пакет данных ответа может содержать выполняемые компьютером команды, представляющие ответ 181.

Тем не менее, предполагается, что в некоторых вариантах осуществления данной технологии, где источник 102 запросов представляет собой, например, подсистему системы 100 распределенной обработки информации, сеть 103 связи может быть
 10 реализована способом, отличным от описанного выше, или в некоторых случаях может вовсе отсутствовать в пределах объема данной технологии.

Операционная подсистема (улей)

Как указано выше, система 100 распределенной обработки информации содержит операционную подсистему 111 или, сокращенно, улей. В общем случае улей 111
 15 представляет собой программное приложение (например, конечный автомат), способное управлять по меньшей мере некоторыми подсистемами системы 100 распределенной обработки информации, такими как подсистема 104 предварительной обработки запросов и подсистема 105 обработки транзакций. Можно сказать, что улей 111 может быть реализован как конечный автомат (SM, State Machine), способный формировать,
 20 удалять и/или выравнивать нагрузку других автоматов SM, образующих по меньшей мере некоторые подсистемы системы 100 распределенной обработки информации.

Должно быть понятно, что автомат SM представляет собой вычислительную модель, реализованную в виде компьютерной системы и определяемую списком состояний. Автомат SM может изменять свое текущее состояние в ответ на некоторые внешние
 25 входные данные и может в конкретный момент времени находиться лишь в одном состоянии. Переход автомата SM из одного состояния в другое состояние называется изменением состояния.

Следует отметить, что в контексте данной технологии автоматы SM, образующие по меньшей мере некоторые подсистемы системы 100 распределенной обработки
 30 информации, являются детерминированными по своей природе, т.е. каждое изменение состояния каждого автомата SM однозначно определено (а) текущим состоянием соответствующего автомата SM и (б) внешними входными данными соответствующего автомата SM. Иными словами, для текущего состояния автомата SM и для внешних входных данных существует единственное следующее состояние этого автомата SM.
 35 Этот детерминированный характер изменения состояния не зависит от того, у какого именно автомата SM системы 100 распределенной обработки информации происходит изменение состояния.

Поэтому, как описано ниже, в некоторых вариантах осуществления данной технологии система 100 распределенной обработки информации должна принимать
 40 внешние входные данные определенного вида, соответствующие свойству детерминированности автоматов SM, от по меньшей мере некоторых подсистем системы 100 распределенной обработки информации.

Подсистема распределенного хранения данных

Как упомянуто ранее, система 100 распределенной обработки информации также
 45 содержит подсистему 108 распределенного хранения данных. В общем случае подсистема 108 распределенного хранения данных способна, среди прочего, хранить системные данные, указывающие состояния, изменения состояний, внешние входные и/или выходные данные по меньшей мере некоторых автоматов SM системы 100

распределенной обработки информации. Например, системные данные, связанные с автоматом SM системы 100 распределенной обработки информации, могут храниться в форме журнала, содержащего хронологический список состояний, изменений состояний, внешних входных и/или выходных данных данного автомата SM.

5 Подсистема 108 распределенного хранения данных также способна хранить клиентские данные, т.е. данные, связанные с внешними входными данными, обрабатываемыми системой 100 распределенной обработки информации. Например, в некоторых вариантах осуществления данной технологии клиентские данные могут храниться как часть системных данных в распределенной подсистеме 108 хранения
10 данных в пределах объема данной технологии.

Для хранения системных и/или клиентских данных подсистема 108 распределенного хранения данных содержит множество устройств 112 хранения данных, каждое из которых далее называется устройством 112 хранения данных. Согласно различным вариантам осуществления данной технологии, некоторые или все устройства из
15 множества устройств 112 хранения данных могут располагаться в одном или в различных местах. Например, некоторые или все устройства из множества устройств 112 хранения данных могут располагаться в одной серверной стойке, и/или в одном центре обработки данных, и/или могут быть распределены среди множества серверных стоек в одном или нескольких центрах обработки данных.

20 В некоторых вариантах осуществления данной технологии системные и/или клиентские данные, хранящиеся в одном устройстве 112 хранения данных, могут дублироваться и храниться в нескольких других устройствах 112 хранения данных. В некоторых вариантах осуществления такое дублирование и хранение системных и/или клиентских данных может обеспечивать отказоустойчивое хранение системных и/или
25 клиентских данных в системе 100 распределенной обработки информации. Отказоустойчивое хранение системных и/или клиентских данных способно предотвратить утрату данных, когда устройство 112 хранения данных подсистемы 108 распределенного хранения данных становится временно или постоянно недоступным для хранения и извлечения данных. Такое отказоустойчивое хранение системных и/или
30 клиентских данных также может предотвращать утрату данных, когда автомат SM системы 100 распределенной обработки информации становится временно или постоянно недоступным.

Предполагается, что устройство 112 хранения данных может быть реализовано в виде компьютерного сервера. Компьютерный сервер содержит по меньшей мере одно
35 физическое запоминающее устройство (т.е. накопитель 126) и одно или несколько программных приложений, способных выполнять машиночитаемые команды. Накопитель 126 может представлять собой твердотельный накопитель (SSD, Solid State Drive), накопитель на жестких дисках (HDD, Hard Disk Drive) и т.п. Поэтому можно сказать, что по меньшей мере одно физическое запоминающее устройство может быть
40 реализовано как устройство с подвижным или неподвижным диском.

Например, как показано на фиг. 1, данное устройство 112 хранения данных способно содержать, помимо прочего, следующие программные приложения: приложение 114 виртуального накопителя (Vdrive), приложение 116 физического накопителя (Pdrive), по меньшей мере одно приложение 118 моделирования накопителя, по меньшей мере
45 одно приложение 120 планирования работы, приложение 122 обеспечения работы в реальном времени и по меньшей мере один прокси-сервер 124 автомата SM. Функции указанных выше программных приложений и накопителя 126, обеспечивающие хранение по меньшей мере некоторых системных и/или клиентских данных, более подробно

описаны ниже со ссылками на фиг. 2.

Подсистема предварительной обработки запросов

Как указано выше, подсистема 105 обработки транзакций может содержать несколько детерминированных автоматов SM, которые должны получать входные данные
 5 определенного вида и которые соответствуют свойству детерминированности автоматов SM. Также следует еще раз отметить, что источник 102 запросов выдает запрос 180 в виде запроса декларативного типа.

Соответственно, подсистема 104 предварительной обработки запросов способна принимать запрос 180 декларативного типа от источника 102 запросов и предварительно
 10 обрабатывать или преобразовывать запрос 180 во множество детерминированных транзакций 182, которые соответствуют свойству детерминированности нескольких автоматов SM, образующих подсистему 105 обработки транзакций.

В целом, подсистема 104 предварительной обработки запросов предназначена для предварительной обработки или преобразования запроса 180 во множество
 15 детерминированных транзакций 182, которые могут быть обработаны детерминированными автоматами SM из подсистемы 105 обработки транзакций.

Следует отметить, что подсистема 104 предварительной обработки запросов также способна формировать ответ 181 для передачи в источник 102 запросов. Разумеется, что подсистема 104 предварительной обработки запросов связана с подсистемой 105
 20 обработки транзакций не только для передачи множества детерминированных транзакций 182, но и для приема информации об обработке множества детерминированных транзакций 182. В некоторых не имеющих ограничительного характера вариантах осуществления данной технологии множество детерминированных транзакций 182 может содержать одну или несколько транзакций вида «считывание»
 25 или «запись».

В некоторых вариантах осуществления данной технологии подсистема 104 предварительной обработки запросов реализована в виде по меньшей мере одного автомата SM в пределах объема данной технологии.

В некоторых вариантах осуществления данной технологии предполагается, что
 30 представленная на фиг. 1 система 100 распределенной обработки информации способна поддерживать транзакции ACID (Atomicity, Consistency, Isolation and Durability). В целом, ACID - это аббревиатура для набора свойств транзакции (атомарность, согласованность, изолированность, долговечность), которые обеспечивают надежность базы данных при выполнении транзакций. Поэтому в некоторых вариантах осуществления данной
 35 технологии предполагается, что транзакции, предназначенные для подсистемы 105 обработки транзакций, могут быть атомарными, согласованными, изолированными и долговечными в пределах объема данной технологии.

Подсистема обработки транзакций

В общем случае подсистема 105 обработки транзакций способна принимать и
 40 обрабатывать множество детерминированных транзакций 182 и, таким образом, обрабатывать запрос 180 от источника 102 запросов. Подсистема 105 обработки транзакций содержит (а) подсистему 106 маршрутизации транзакций и (б) подсистему 110 базы данных, которые описаны далее.

Подсистема 110 базы данных содержит множество мест назначения транзакции (TDL, Transaction Destination Location) и разделена на множество сегментов 109, каждый из
 45 которых далее называется сегментом 109. В не имеющем ограничительного характера примере подсистема 110 базы данных может содержать базу данных, включающую в себя одну или несколько таблиц базы данных. Таблица базы данных может состоять

из по меньшей мере двух столбцов, таких как первый столбец, содержащий ключи, и второй столбец, содержащий записи, в которых хранятся данные, связанные с соответствующими ключами. В этом не имеющем ограничительного характера примере место TDL может соответствовать строке таблицы базы данных, т.е. место TDL может соответствовать ключу и соответствующей записи в таблице базы данных.

В этом не имеющем ограничительного характера примере каждый сегмент 109 подсистемы 110 базы данных содержит часть таблицы базы данных. Следовательно, множество мест TDL, сопоставленных с соответствующими строками таблицы базы данных, разделяется между множеством сегментов 109 так, чтобы каждый сегмент 109 содержал соответствующее подмножество (например, диапазон) множества мест TDL.

В некоторых вариантах осуществления данной технологии предполагается, что каждый сегмент из множества сегментов 109 может быть реализован с использованием соответствующего детерминированного автомата SM. Это означает, что после приема транзакции, предназначенной для места TDL сегмента 109, реализованного с использованием автомата SM, этот автомат SM может обработать транзакцию и перейти из текущего состояния в новое состояние в соответствии с транзакцией, как описано выше.

Подсистема 106 маршрутизации транзакций способна маршрутизировать транзакции из множества детерминированных транзакций 182 в соответствующие места TDL и, таким образом, направлять их в соответствующие сегменты 109 подсистемы 110 базы данных. С этой целью подсистема 106 маршрутизации транзакций может включать в себя множество портов, которые обычно способны (а) принимать множество детерминированных транзакций 182 из подсистемы 104 предварительной обработки запросов, (б) разделять множество детерминированных транзакций 182 на подмножества детерминированных транзакций, предназначенных для соответствующих сегментов 109, и (в) формировать для каждого сегмента централизованный порядок выполнения детерминированных транзакций соответствующими сегментами 109.

Следует отметить, что каждый порт из множества портов, образующих подсистему 106 маршрутизации транзакций, может быть реализован как соответствующий автомат SM. В некоторых вариантах осуществления изобретения предполагается, что множество портов может включать в себя порты двух различных видов для маршрутизации транзакций из множества детерминированных транзакций 182 в соответствующие сегменты 109. В других вариантах осуществления изобретения по меньшей мере некоторые функции множества портов могут выполняться автоматами SM, соответствующими множеству сегментов 109.

Кроме того, как показано на фиг. 1, по меньшей мере некоторые из автоматов SM подсистемы 105 обработки транзакций могут быть связаны с подсистемой 108 распределенного хранения данных с использованием соответствующей линии 160 связи. В целом, линия 160 связи предназначена для передачи системных данных, таких как данные о состояниях, изменениях состояний, внешних входных и/или выходных данных соответствующих автоматов SM и т.д., в подсистему 108 распределенного хранения данных для хранения. Далее со ссылкой на фиг. 2 более подробно описана реализация линий 160 связи и порядок конфигурирования подсистемы 108 распределенного хранения данных для хранения системных данных.

На фиг. 2 представлено устройство 112 хранения данных, входящее в состав подсистемы 108 распределенного хранения данных. Как описано выше, устройство 112 хранения данных содержит по меньшей мере один прокси-сервер 124 автомата SM. Прокси-сервер автомата SM предназначен для управления связью между автоматом

SM и подсистемой 108 распределенного хранения данных. В некоторых вариантах осуществления данной технологии предполагается, что по меньшей мере один прокси-сервер 124 автомата SM устройства 112 хранения данных может представлять собой прикладной программный интерфейс (API, Application Programing Interface), управляющий связью между автоматом SM и устройством 112 хранения данных. В других вариантах осуществления данной технологии по меньшей мере один прокси-сервер 124 автомата SM сам может быть реализован как автомат SM. В других вариантах осуществления данной технологии по меньшей мере один прокси-сервер 124 автомата SM может быть реализован как программный модуль (не как автомат SM) для выполнения описанных выше функций.

В некоторых вариантах осуществления данной технологии прокси-сервер 124 автомата SM может быть способен (а) принимать системные данные обновления журнала этого автомата SM с использованием соответствующей линии 160 связи, (б) обрабатывать системные данные и (в) передавать обработанные системные данные соответствующему приложению 114 Vdrive для дальнейшей обработки.

По меньшей мере один прокси-сервер 124 автомата SM может быть способен обрабатывать системные данные, например, для обеспечения целостности и отказоустойчивости системных данных. В некоторых вариантах осуществления данной технологии предполагается, что по меньшей мере один прокси-сервер 124 автомата SM может быть способен выполнять стойкое к потере кодирование (erasure coding) системных данных. В целом, стойкое к потере кодирование представляет собой способ кодирования, предусматривающий получение данных с избыточностью и разделение их на несколько фрагментов. Такие избыточность и фрагментация способствуют восстановлению данных в случае потери одного или нескольких фрагментов вследствие отказов в системе.

Предполагается, что обработанные таким образом по меньшей мере одним прокси-сервером 124 автомата SM системные данные принимаются по меньшей мере одним соответствующим приложением 114 Vdrive устройства 112 хранения данных. Приложение 114 Vdrive предназначено для обработки системных данных, полученных от по меньшей мере одного прокси-сервера 124 автомата SM, и для ответного формирования соответствующих операций I/O, которые должны выполняться накопителем 126 для сохранения системных данных в накопителе 126 устройства 112 хранения данных. После формирования по меньшей мере одним приложением 114 Vdrive операций I/O, соответствующих принятым системным данным, по меньшей мере одно приложение 114 Vdrive передает эти операции I/O приложению 116 Pdrive.

Иными словами, предполагается, что устройство 112 хранения данных может содержать несколько прокси-серверов 124 автомата SM для обработки и передачи системных данных нескольким соответствующим приложениям 114 Vdrive, которые обрабатывают системные данные, формируют соответствующие операции I/O и передают соответствующие операции I/O в единое приложение 116 Pdrive устройства 112 хранения данных.

В целом, приложение 116 Pdrive предназначено для управления работой накопителя 126. Например, приложение 116 Pdrive может быть способным выполнять кодирование операций I/O, которые должны выполняться в накопителе 126, и различные другие функции, способствующие надежному хранению данных в накопителе 126.

Приложение 116 Pdrive связано с приложением 120 планирования работы для передачи операции I/O. Приложение 120 планирования работы способно планировать передачу операций I/O в накопитель 126. Предполагается, что приложение 120 планирования

работы или сокращенно планировщик может использовать различные схемы планирования для определения порядка передачи операций I/O в накопитель 126 для их дальнейшего выполнения.

Предполагается, что в некоторых вариантах осуществления данной технологии планировщик 120 может быть реализован как часть приложения 116 Pdrive. Иными словами, приложением 116 Pdrive могут выполняться различные схемы планирования в пределах объема данной технологии.

Предполагается, что планировщик 120 может обеспечивать гибридную схему планирования. Например, планировщик 120 может обеспечивать схему «справедливого» планирования, которая при определенных условиях также представляет собой схему планирования «для реального времени».

Следует отметить, что для данного устройства 112 хранения данных может потребоваться хранение операций I/O, соответствующих системным данным, связанным с несколькими автоматами SM. Кроме того, каждый из нескольких автоматов SM связан с заранее заданной долей производительности накопителя, которую накопитель 126 может выделить для выполнения операций I/O, связанных с соответствующим автоматом SM. В целом, схемы «справедливого» планирования способны упорядочивать операции I/O, передаваемые в накопитель 126, так, чтобы производительность накопителя 126, предназначенная для выполнения упорядоченных операций I/O, использовалась согласно заранее заданным долям, выделенным нескольким автоматам SM.

Следует еще раз отметить, что система 100 распределенной обработки информации может использоваться для предоставления услуг «облачного» хранения данных. Во многих таких вариантах осуществления может быть желательным обрабатывать и сохранять системные данные в реальном времени, т.е. в течение очень короткого интервала времени. Чтобы обеспечить выполнение требования относительно режима реального времени к системе 100 распределенной обработки информации, операции I/O могут быть связаны с соответствующими предельными сроками, которые указывают момент времени, по прошествии которого соответствующие операции I/O уже не попадают в пределы допустимого времени, соответствующие требованиям относительно режима реального времени к системе 100 распределенной обработки информации. В целом, схемы планирования реального времени способны упорядочивать операции I/O, передаваемые в накопитель 126, так, чтобы операции I/O выполнялись накопителем 126 с соблюдением соответствующих предельных сроков.

В целом, планировщик 120 способен обеспечивать гибридную схему планирования, позволяющую упорядочивать операции I/O, передаваемые для выполнения в накопитель 126, так, чтобы учитывались и заранее заданные доли производительности накопителя для каждого соответствующего SM, и соответствующие предельные сроки операций I/O.

Как было упомянуто ранее, накопитель 126 представляет собой носитель информации для выполнения операций I/O и, соответственно, для хранения системных данных, переданных в устройство 112 хранения данных. Например, накопитель 126 может быть реализован в виде накопителя HDD или накопителя SSD. Накопитель 126 включает в себя внутреннее логическое устройство 250 накопителя для выбора операции I/O среди всех переданных в накопитель операций I/O для выполнения в текущий момент времени.

Следует отметить, что операции I/O могут направляться для выполнения в накопитель 126 по одной, но это может приводить к увеличению задержек между накопителем 126 и другими компонентами устройства 112 хранения данных. Поэтому операции I/O могут передаваться в накопитель 126 пакетами или группами. После приема пакета или группы

операций I/O накопителем 126 внутреннее логическое устройство 250 накопителя способно выбрать для выполнения наиболее эффективную операцию I/O среди имеющихся в пакете операций I/O.

Например, наиболее эффективная операция I/O может выбираться на основе различных критериев, таких как место выполнения предыдущей операции I/O в накопителе 126 и место выполнения операций I/O, имеющихся в накопителе 126. Иными словами, внутреннее логическое устройство 250 накопителя способно выбирать для текущего выполнения наиболее эффективную операцию (с точки зрения накопителя 126) среди всех операций I/O, имеющихся в накопителе 126 в текущий момент времени.

Поэтому в некоторых случаях, несмотря на то, что планировщик 120 упорядочил операции I/O для передачи определенным образом с учетом требований относительно режима реального времени к системе 100 распределенной обработки информации, внутреннее логическое устройство 250 накопителя 126 может выдавать в накопитель 126 команды для формирования порядка выполнения операций I/O, отличного от порядка передачи, выбранного планировщиком 120. Поэтому порядок выполнения иногда может не соответствовать требованиям относительно режима реального времени системы 100 распределенной обработки информации (в частности, когда от планировщика 120 принимаются дополнительные операции I/O, которые могут быть более «эффективными» с точки зрения накопителя 126 и могут выбираться среди еще не выполненных операций I/O).

Чтобы обеспечить работу устройства 112 хранения данных в реальном времени и предотвратить описанную выше проблему, также известную как «торможение работы» (operation stagnation), устройство 112 хранения данных может содержать приложение 122 обеспечения работы в реальном времени. В целом, приложение 122 обеспечения работы в реальном времени позволяет управлять тем, какие операции I/O из числа уже упорядоченных планировщиком 120 передаются в текущий момент времени для выполнения в накопитель 126.

Предполагается, что в некоторых вариантах осуществления данной технологии приложение 122 обеспечения работы в реальном времени может быть реализовано как часть приложения 116 Pdrive. Иными словами, вышеупомянутые функции приложения 122 обеспечения работы в реальном времени могут выполняться приложением 116 Pdrive в пределах объема данной технологии.

Устройство 112 хранения данных также может содержать по меньшей мере одно соответствующее приложение 118 моделирования накопителя для каждого накопителя 126 в устройстве 112 хранения данных. В целом, приложение 118 моделирования накопителя способно эмулировать идеальную работу накопителя 126 для контроля накопителя 126 в диагностических целях. Тем не менее, в других вариантах осуществления изобретения планировщик 120 также может использовать приложение 118 моделирования накопителя, чтобы упорядочивать операции I/O для их передачи в накопитель 126.

Предполагается, что в некоторых вариантах осуществления данной технологии по меньшей мере одно соответствующее приложение 118 моделирования накопителя может быть реализовано как часть приложения 116 Pdrive. Иными словами, вышеупомянутые функции по меньшей мере одного соответствующего приложения 118 моделирования накопителя могут выполняться приложением 116 Pdrive в пределах объема данной технологии.

Гибридный планировщик операций I/O

Как указано выше, приложение 116 Pdrive может быть связано с планировщиком

120. В других вариантах осуществления предполагается, что планировщик 120 может быть реализован как часть приложения 116 Pdrive, т.е. различные схемы планирования могут выполняться приложением 116 Pdrive в пределах объема данной технологии.

Кроме того, как было упомянуто ранее, планировщик 120 способен обеспечивать гибридную схему планирования. Иными словами, планировщик 120 может обеспечивать схему планирования, позволяющую упорядочивать операции I/O, передаваемые для выполнения в накопитель 126 (или передаваемые в приложение 122 обеспечения работы в реальном времени, которое затем передает операции I/O в накопитель 126), так, чтобы учитывались и заранее заданные доли производительности накопителя для каждого соответствующего SM, и соответствующие предельные сроки операций I/O.

На фиг. 3 представлен гибридный планировщик 300 для операций I/O. В одном не имеющем ограничительного характера примере планировщик 120, использующий гибридную схему планирования, может быть реализован как гибридный планировщик 300. Разумеется, это означает, что гибридный планировщик 300 может быть связан приложением 116 Pdrive и/или может быть реализован как часть приложения 116 Pdrive.

Гибридный планировщик 300 содержит первый планировщик 310 и второй планировщик 320. В общем случае первый планировщик 310 способен обеспечивать схему «справедливого» планирования, а второй планировщик 320 способен обеспечивать схему планирования «для реального времени». С этой целью первый планировщик 310 может быть способен реализовывать алгоритм «справедливого» планирования, а второй планировщик 320 может быть способен реализовывать алгоритм планирования «для реального времени».

В общем случае тот или иной алгоритм планирования может использоваться планировщиком, чтобы упорядочивать принятые им операции I/O на основе некоторых заранее заданных критериев и, следовательно, обеспечивать схему планирования для операций I/O на основе некоторых заранее заданных критериев. Например, алгоритм «справедливого» планирования может использоваться первым планировщиком 310, чтобы упорядочивать операции I/O (принятые первым планировщиком 310) с учетом, среди прочего, производительности накопителя 126. Например, алгоритм планирования «для реального времени» может использоваться вторым планировщиком 320, чтобы упорядочивать операции I/O (принятые вторым планировщиком 320) с учетом, среди прочего, предельных сроков выполнения соответствующих операций I/O.

Далее описана реализация первого планировщика 310 для обеспечения схемы «справедливого» планирования для операций I/O, принятых первым планировщиком 310, и реализация второго планировщика 320 для обеспечения схемы планирования «для реального времени» для операций I/O, принятых вторым планировщиком 320.

Следует отметить, что операции I/O могут передаваться в гибридный планировщик 300, например, из нескольких приложений 114 Vdrive. Поэтому можно сказать, что гибридный планировщик 300 способен принимать операции I/O от нескольких источников операций I/O, которые могут быть реализованы в виде соответствующих приложений 114 Vdrive. Например, первое приложение 114 Vdrive может быть способно формировать операции I/O на основе системных данных, сформированных первым автоматом SM. В другом примере второе приложение 114 Vdrive может быть способно формировать операции I/O на основе системных данных, сформированных вторым автоматом SM. Иными словами, гибридный планировщик 300 может быть способен принимать операции I/O, сформированные несколькими источниками I/O (например, первым приложением 114 Vdrive и вторым приложением 114 Vdrive) на основе системных данных, сформированных соответствующими автоматами SM.

В представленном на фиг. 3 не имеющем ограничительного характера примере первый планировщик 310 может быть способен принимать первый набор 302 операций I/O и второй набор 304 операций I/O. Первый планировщик 310 способен принимать операции I/O из первого набора 302 операций I/O по первой линии 312 связи, которая
 5 соответствует операциям I/O, сформированным первым источником операций I/O. В этом примере первая линия 312 связи может связывать первое приложение 114 Vdrive с приложением 116 Pdrive для передачи первого набора 302 операций I/O в приложение 116 Pdrive, реализующее гибридный планировщик 300.

Первый планировщик 310 также может быть способен принимать операции I/O из
 10 второго набора 304 операций I/O по второй линии 314 связи, которая соответствует операциям I/O, сформированным вторым источником операций I/O. В этом примере вторая линия 314 связи может связывать второе приложение 114 Vdrive с приложением 116 Pdrive для передачи второго набора 304 операций I/O в приложение 116 Pdrive, реализующее гибридный планировщик 300.

Следует отметить, что первый источник операций I/O связан с первой заранее заданной долей производительности накопителя, которую накопитель 126 может выделить для выполнения операций I/O из первого набора 302 операций I/O. Аналогично, второй источник операций I/O связан со второй заранее заданной долей
 15 производительности накопителя, которую накопитель 126 может выделить для выполнения операций I/O из второго набора 304 операций I/O.

Предполагается, что первая и вторая заранее заданные доли могут быть определены оператором системы 100 распределенной обработки информации. Также следует отметить, что в пределах объема данной технологии в первый планировщик 310 могут направляться операции I/O из более чем двух источников операций I/O подобно тому,
 20 как первый набор 302 операций I/O и второй набор 304 операций I/O направляются в первый планировщик 310.

Первый планировщик 310 способен реализовать схему «справедливого» планирования, которая позволяет упорядочивать операции I/O из первого набора 302 операций I/O и из второго набора 304 операций I/O таким образом, чтобы
 30 производительность накопителя 126 для выполнения упорядоченных операций I/O использовалась согласно первой и второй заранее заданным долям. Иными словами, схема «справедливого» планирования упорядочивает операции I/O из первого набора 302 операций I/O и из второго набора 304 операций I/O таким образом, чтобы производительность накопителя 126 использовалась «справедливо», т.е. согласно
 35 соответствующим первой и второй заранее заданным долям производительности накопителя, по меньшей мере в некоторых вариантах осуществления данной технологии.

Можно сказать, что первый планировщик 310, как упомянуто ранее, использует алгоритм «справедливого» планирования для формирования «справедливо» спланированного порядка операций I/O на основе первой и второй заранее заданных
 40 долей производительности накопителя 126.

На реализуемый первым планировщиком 310 алгоритм «справедливого» планирования не накладывается каких-либо особых ограничений. Тем не менее, в одном не имеющем ограничительного характера примере первый планировщик 310 может быть реализован с использованием алгоритма планирования DRR, т.е. первый
 45 планировщик 310 может быть реализован как планировщик типа DRR.

Независимо от реализуемого первым планировщиком 310 конкретного вида алгоритма «справедливого» планирования, первый планировщик 310 способен определять порядок, в котором операции I/O из первого набора 302 операций I/O и из

второго набора 304 операций I/O должны передаваться, чтобы производительность накопителя 126 использовалась в соответствии с (а) первой заранее заданной долей производительности, выделенной первому источнику I/O, и (б) второй заранее заданной долей производительности, выделенной второму источнику I/O.

5 Первый планировщик 310 способен передавать операции I/O в «справедливо» спланированном порядке, как описано выше, во второй планировщик 320 по линии 316 связи. Поэтому можно сказать, что второй планировщик 320 способен принимать операции I/O от первого планировщика 310. Также можно сказать, что операции I/O принимаются вторым планировщиком 320 по линии 316 связи в «справедливо»
10 спланированном порядке, определенном первым планировщиком 310.

Следует отметить, что операции I/O могут связываться с соответствующими предельными сроками, которые указывают момент времени, после которого соответствующие операции I/O уже не попадают в пределы допустимого времени, соответствующие требованиям относительно режима реального времени к системе 100
15 распределенной обработки информации.

Для выполнения требований относительно режима реального времени к системе 100 распределенной обработки информации, второй планировщик 320 способен обеспечивать схему планирования «для реального времени», которая позволяет выполнять операции I/O до истечения соответствующих предельных сроков. Иными
20 словами, второй планировщик 320 способен упорядочивать (или в данном случае переупорядочивать) операции I/O для их передачи в накопитель 126 так, чтобы операции I/O могли выполняться накопителем 126 до истечения соответствующих предельных сроков.

Можно сказать, что, как упомянуто выше, второй планировщик 320 использует
25 алгоритм планирования «для реального времени» для формирования спланированного «для реального времени» порядка операций I/O на основе соответствующих предельных сроков выполнения операций I/O.

На реализуемый вторым планировщиком 320 алгоритм планирования «для реального времени» не накладывается каких-либо особых ограничений. Тем не менее, в одном не
30 имеющем ограничительного характера примере второй планировщик 320 может быть реализован с использованием алгоритма планирования EDF, т.е. второй планировщик 320 может быть реализован как планировщик типа EDF.

Независимо от реализуемого вторым планировщиком 320 конкретного вида алгоритма планирования «для реального времени», второй планировщик 320 способен
35 определять спланированный «для реального времени» порядок принятых операций I/O, передаваемых далее в накопитель 126 для их выполнения до истечения соответствующих предельных сроков.

Второй планировщик 320 способен передавать операции I/O в спланированном «для реального времени» порядке, как описано выше, в накопитель 126 (или в приложение
40 122 обеспечения работы в реальном времени, которое затем передает операции I/O в накопитель 126) по линии 318 связи.

Следует отметить, что гибридный планировщик 300 передает единый поток операций I/O. Иными словами, предполагается, что гибридный планировщик 300 не способен «демультиплексировать» принятые операции I/O в несколько потоков операций I/O.
45 Поэтому можно сказать, что гибридный планировщик 300 представляет собой гибридный планировщик без демультиплексирования, обеспечивающий единый поток операций I/O (упорядоченных вторым планировщиком 320).

В некоторых вариантах осуществления данной технологии предполагается, что

гибридный планировщик 300 способен селективно работать в обычном режиме или в режиме высокой нагрузки. Далее описаны реализация гибридного планировщика 300 для работы в обычном режиме, реализация гибридного планировщика 300 для работы в режиме высокой нагрузки и условия селективного переключения гибридного планировщика 300 из обычного режима в режим высокой нагрузки.

Обычный режим работы

В обычном режиме работы гибридный планировщик 300 способен принимать операции I/O по нескольким линиям связи, таким как первая линия 312 связи и вторая линия 314 связи, от нескольких приложений 114 Vdrive (например, от источников операций I/O).

Каждая принятая гибридным планировщиком 300 операция I/O может быть связана с соответствующим предельным сроком выполнения и заранее заданной долей производительности накопителя, выделенной соответствующему источнику I/O.

Принятые первым накопителем 300 операции I/O, такие как операции I/O из первого набора 302 операций I/O и из второго набора 304 операций I/O, передаются в первый планировщик 310. Первый планировщик 310 использует алгоритм «справедливого» планирования, чтобы упорядочивать принятые операции I/O на основе заранее заданных долей производительности накопителя, выделенных соответствующим источникам I/O, как описано выше. В результате в обычном режиме работы первый планировщик 310 передает принятые операции I/O в «справедливо» спланированном порядке во второй планировщик 320 по линии 316 связи.

В обычном режиме работы второй планировщик 320 принимает операции I/O в «справедливо» спланированном порядке. Второй планировщик 320 использует алгоритм планирования «для реального времени», чтобы упорядочивать принятые операции I/O на основе соответствующих предельных сроков, как описано выше. В результате второй планировщик 320 передает операции I/O, спланированные «для реального времени», по линии 318 связи.

Можно сказать, что работающий в обычном режиме гибридный планировщик 300 выполняет следующие действия: (а) прием операций I/O из нескольких источников операций I/O, (б) их «справедливое» упорядочивание с использованием первого планировщика 310, (в) передача их в «справедливом» порядке во второй планировщик 320, (г) упорядочивание их для режима «реального времени» с использованием второго планировщика 320, (д) передача их по линии 318 связи в накопитель 126 для выполнения (или в приложение 122 обеспечения работы в реальном времени, которое затем передает их в накопитель 126 для выполнения).

Предполагается, что в обычном режиме работы гибридный планировщик 300 способен обеспечить такое выполнение операций I/O, чтобы (а) производительность накопителя 126 использовалась в соответствии с заранее заданными долями производительности, выделенными соответствующим источникам I/O, и (б) эти операции выполнялись до истечения соответствующих предельных сроков. Иными словами, работающий в обычном режиме гибридный планировщик 300 способен обеспечить, чтобы связанные с несколькими источниками I/O операции I/O выполнялись так, чтобы производительность накопителя 126 «справедливо» использовалась для выполнения операций I/O, связанных с несколькими источниками I/O, и одновременно выполнялись требования относительно режима реального времени к системе 100 распределенной обработки информации.

Условия для селективного переключения режимов работы

В определенный момент времени гибридный планировщик 300 может принимать

для планирования большое количество операций I/O для их дальнейшего выполнения накопителем 126. Таким образом, в некоторых случаях, даже если накопитель 126 выполняет операции I/O, накопитель 126 может не соответствовать высоким требованиям к обработке из-за большого количества операций I/O, которые необходимо

5 выполнить.

Следует еще раз отметить, что в обычном режиме работы первый планировщик 310 принимает операции I/O и отправляет их в «справедливо» спланированном порядке во второй планировщик 320. Таким образом, чем больше операций I/O принимается гибридным планировщиком 300, тем больше операций I/O передается во второй

10 планировщик 320. Поэтому второй планировщик 320 должен быть способен применять алгоритм планирования «для реального времени» в отношении большого количества операций I/O с целью определения спланированного «для реального времени» порядка операций I/O. Иными словами, алгоритм планирования «для реального времени» требуется, чтобы до известной степени учитывать большее количество операций I/O и

15 их соответствующие предельные сроки выполнения с целью определения порядка операций I/O «для реального времени».

Предполагается, что во время пиков высокой нагрузки (когда в гибридный планировщик 300 направляется довольно большое количество операций I/O) гибридный планировщик 300 способен селективно переключать режимы работы. Это означает,

20 что во время пиков высокой нагрузки гибридный планировщик 300 способен селективно переключаться из обычного режима в режим высокой нагрузки.

Предполагается, что селективное переключение между режимами работы может выполняться гибридным планировщиком 300 за счет контроля ситуации, в которой по меньшей мере одна операция I/O выполняется без соблюдения соответствующего

25 предельного срока. Например, второй планировщик 320 может быть способен контролировать выполнение операций I/O, сформированных первым и вторым источниками I/O, согласно соответствующим срокам.

Этот осуществляемый гибридным планировщиком 300 контроль может включать в себя априорный анализ выполнения операций I/O и/или апостериорный анализ

30 выполнения операций I/O. Далее описаны априорный и апостериорный анализ, которые могут выполняться в процессе контроля, осуществляемого гибридным планировщиком 300 с целью селективного переключения режимов работы.

В процессе контроля с использованием априорного анализа второй планировщик 320 может контролировать возможность выполнения принятых операций I/O до

35 истечения их соответствующих предельных сроков. С одной стороны, если второй планировщик 320 может определить спланированный «для реального времени» порядок принятых операций I/O, в котором возможно выполнение операций I/O до истечения их соответствующих предельных сроков, второй планировщик 320 может селективно определить, что гибридный планировщик 300 должен работать в обычном режиме. С

40 другой стороны, если второй планировщик 320 не может определить порядок принятых операций I/O для «реального времени», в котором возможно выполнение операций I/O до истечения их соответствующих предельных сроков, второй планировщик 320 может селективно определить, что гибридный планировщик 300 должен работать в режиме высокой нагрузки.

Следовательно, можно сказать, что второй планировщик 320 способен селективно определять, что гибридный планировщик 300 должен работать в режиме высокой

45 нагрузки, если определена невозможность выполнения накопителем 126 до истечения соответствующего предельного срока по меньшей мере одной принятой операции I/O,

передаваемой в накопитель 126 согласно перепланированному порядку операций I/O.

Когда контроль осуществляется с использованием априорного анализа, второй планировщик 320 дополнительно или в качестве альтернативы может селективно определить, что гибридный планировщик 300 должен работать в режиме высокой нагрузки, если определена невозможность выполнения накопителем 126 до истечения соответствующего предельного срока по меньшей мере одной операции I/O, если вторым планировщиком 320 принимается дополнительная операция I/O от первого планировщика 310. Например, второй планировщик 320 может определить, что после приема дополнительной операции I/O и добавления ее в перепланированный порядок операций I/O на основе соответствующего предельного срока выполнения накопитель 126 не будет способен выполнить до истечения соответствующего предельного срока по меньшей мере одну операцию I/O согласно перепланированному порядку.

В процессе контроля с использованием апостериорного анализа гибридный планировщик 300 может контролировать фактическое выполнение до истечения соответствующих предельных сроков переданных в накопитель 126 операций I/O согласно перепланированному порядку. Это означает, что гибридный планировщик 300 может селективно определить, что гибридный планировщик 300 должен работать в режиме высокой нагрузки, если определено фактическое невыполнение накопителем 126 до истечения соответствующего предельного срока по меньшей мере одной операции I/O.

Предполагается, что в некоторых вариантах осуществления данной технологии, когда контроль выполняется с использованием апостериорного анализа, гибридный планировщик 300 может быть способен принимать подтверждения выполнения операций I/O, переданных гибридным планировщиком 300 в накопитель 126 для выполнения. Иными словами, если выполнение данных операций I/O происходит после истечения соответствующего срока, гибридный планировщик 300 может быть способен селективно переключаться в режим высокой нагрузки.

Гибридный планировщик 300 может осуществлять этот контроль с использованием апостериорного анализа в различных ситуациях. В одном не имеющем ограничительного характера примере такая ситуация может возникать вследствие того, что описанное выше внутреннее логическое устройство 250 накопителя способно выбирать наиболее эффективный порядок выполнения операций I/O среди имеющихся операций I/O. В результате этого в некоторых случаях, даже несмотря на передачу гибридным планировщиком 300 операций I/O в порядке «для реального времени», в котором операции I/O могут быть выполнены до истечения их соответствующих предельных сроков, внутреннее логическое устройство 250 накопителя может выдавать в накопитель 126 команды для выполнения операций I/O в другом порядке, согласно которому некоторые операции I/O фактически не будут выполнены до истечения их соответствующих предельных сроков.

В целом, селективное переключение между режимами работы может выполняться гибридным планировщиком 300 с использованием контроля ситуации, в которой по меньшей мере одна операция I/O выполняется без соблюдения соответствующего предельного срока. Этот контроль может включать в себя выполнение априорного анализа с целью определения возможности выполнения принятых вторым планировщиком 320 операций I/O до истечения их соответствующих предельных сроков. Дополнительно или в качестве альтернативы этот контроль может включать в себя выполнение апостериорного анализа с целью контроля фактического выполнения накопителем 126 операций I/O (переданных гибридным планировщиком 300 в порядке

«для реального времени») до истечения их соответствующих предельных сроков.

Режим работы с высокой нагрузкой

Когда гибридный планировщик 300 определяет, что гибридный планировщик 300 должен селективно переключиться в режим высокой нагрузки, второй планировщик 320 способен приостанавливать прием дополнительных операций I/O от первого планировщика 310.

В некоторых вариантах осуществления данной технологии второй планировщик 320 может приостанавливать прием дополнительных операций I/O от первого планировщика 310 путем прекращения приема дополнительных операций I/O, которые первый планировщик 310 передает по линии 316 связи.

В других вариантах осуществления данной технологии второй планировщик 320 может приостанавливать прием дополнительных операций I/O от первого планировщика 310 путем передачи сигнала 342 обратной связи по линии 340 связи в первый планировщик 310 (см. фиг. 3). При получении сигнала 342 обратной связи первый планировщик 310 прекращает отправку дополнительных операций I/O во второй планировщик 320.

Например, сигнал 342 обратной связи может указывать на информацию, определенную вторым планировщиком 320 в ходе контроля. В другом примере сигнал 342 обратной связи может указывать на отправленные в первый планировщик 310 машиночитаемые команды, согласно которым первый планировщик 310 должен прекратить отправку дополнительных операций I/O во второй планировщик 320.

Поэтому предполагается, что в режиме высокой нагрузки второй планировщик 320 приостанавливает прием дополнительных операций I/O от первого планировщика 310. В результате этого второй планировщик 320 может обработать операции I/O, уже принятые и требующие упорядочивания для передачи, перед приемом дополнительных операций I/O от первого планировщика 310.

Тем не менее, следует отметить, что приостановка вторым планировщиком 320 приема дополнительных операций I/O от первого планировщика 310 не означает, что первый планировщик 310 должен прекратить прием операций I/O. Например, в режиме высокой нагрузки первый планировщик 310 может принять по меньшей мере одну дополнительную операцию I/O из первого набора 302 операций I/O. В другом примере в режиме высокой нагрузки первый планировщик 310 может принять по меньшей мере одну дополнительную операцию I/O из второго набора 304 операций I/O.

Предполагается, что эта по меньшей мере одна дополнительная операция I/O, принятая первым планировщиком 310 в режиме высокой нагрузки, не возвращается в соответствующий источник операций I/O. Эта по меньшей мере одна дополнительная операция I/O, принятая первым планировщиком 310 в режиме высокой нагрузки, сохраняется в первом планировщике 310, пока гибридный планировщик 300 селективно не переключится в обычный режим работы.

Предполагается, что в некоторых вариантах осуществления изобретения гибридный планировщик 300 может быть способен работать в режиме высокой нагрузки, пока гибридный планировщик 300 не определит возможность селективного переключения из режима высокой нагрузки в обычный режим.

В некоторых вариантах осуществления предполагается, что гибридный планировщик 300 способен селективно переключаться из режима высокой нагрузки в обычный режим, если гибридный планировщик 300 определяет, что операции I/O в данный момент времени выполняются с соблюдением соответствующих предельных сроков.

Например, в некоторый момент времени количество операций I/O, оставшихся во

втором планировщике 320 и требующих передачи для выполнения, может быть меньше чем в момент времени, когда гибридный планировщик 300 селективно переключился в режим высокой нагрузки. Поэтому второму планировщику 320 требуется учитывать меньшее количество операций I/O и, следовательно, меньшее количество предельных

сроков выполнения при определении перепланированного порядка операций I/O. Следовательно, в такой момент времени второй планировщик 320 может определить спланированный «для реального времени» порядок принятых операций I/O, согласно которому операции I/O могут быть выполнены до истечения их соответствующих предельных сроков. В результате гибридный планировщик 300 может селективно

переключиться из режима высокой нагрузки в обычный режим.

В другом примере в некоторый момент времени гибридный планировщик 300 может определить, что переданные для выполнения операции I/O фактически выполняются до истечения соответствующих предельных сроков (например, путем приема подтверждений выполнения из накопителя 126). В результате гибридный планировщик

300 может селективно переключиться из режима высокой нагрузки в обычный режим. В еще одном примере в некоторый момент времени гибридный планировщик 300 может определить, что все принятые вторым планировщиком 320 операции I/O переданы для выполнения. В результате гибридный планировщик 300 может селективно переключиться из режима высокой нагрузки в обычный режим.

Когда гибридный планировщик 300 селективно переключается из режима высокой нагрузки в обычный режим, второй планировщик 320 может селективно возобновлять прием дополнительных операций I/O от первого планировщика 310.

Например, когда гибридный планировщик 300 принимает решение о селективном переключении из режима высокой нагрузки в обычный режим, второй планировщик 320 может передавать другой сигнал обратной связи (не показан) в первый планировщик 310 по линии 340 связи для возобновления передачи операций I/O в «справедливо» спланированном порядке во второй планировщик 320 по линии 316 связи. В результате гибридный планировщик 300 селективно переключается в обычный режим работы.

В некоторых вариантах осуществления данной технологии гибридный планировщик 300 может быть способен выполнять представленный на фиг. 4 способ 400 для планирования операций I/O с целью их выполнения накопителем 126. Далее описан такой способ 400.

Шаг 402: прием первым планировщиком операций I/O (а) по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и (б) по меньшей мере одной операции I/O, сформированной вторым источником операций I/O.

Способ 400 начинается на шаге 402, где первый планировщик 310 принимает по меньшей мере одну операцию I/O, сформированную первым источником операций I/O. Например, первый планировщик 310 может быть способен принимать по меньшей мере одну операцию I/O из первого набора 302 операций I/O по первой линии 312 связи от первого приложения 114 Vdrive.

На шаге 402 первый планировщик 310 также принимает по меньшей мере одну операцию I/O, сформированную вторым источником операций I/O. Например, первый планировщик 310 может быть способен принимать по меньшей мере одну операцию I/O из второго набора 304 операций I/O по второй линии 314 связи от второго приложения 114 Vdrive.

Шаг 404: выполнение первым планировщиком операций I/O первого алгоритма планирования с целью определения спланированного порядка для по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей

мере одной операции I/O, сформированной вторым источником операций I/O, причем первый алгоритм планирования основан на соответствующих заранее заданных долях производительности накопителя, связанных с первым источником операций I/O и вторым источником операций I/O, а результатом выполнения первого алгоритма планирования является первый спланированный порядок операций I/O.

Способ 400 продолжается на шаге 404, где первый планировщик 310 выполняет первый алгоритм «справедливого» планирования.

Как описано выше, первый планировщик 310 способен реализовать схему «справедливого» планирования, позволяющую упорядочивать операции I/O из первого набора 302 операций I/O и из второго набора 304 операций I/O так, чтобы производительность накопителя 126 для выполнения упорядоченных операций I/O использовалась согласно первой и второй заранее заданным долям.

Иными словами, схема «справедливого» планирования упорядочивает операции I/O из первого набора 302 операций I/O и из второго набора 304 операций I/O так, чтобы производительность накопителя 126 использовалась «справедливо», т.е. согласно соответствующим первой и второй заранее заданными долям производительности накопителя, по меньшей мере в некоторых вариантах осуществления данной технологии.

Можно сказать, что первый планировщик 310, как упомянуто ранее, использует алгоритм «справедливого» планирования для формирования «справедливо» спланированного порядка операций I/O на основе первой и второй заранее заданных долей производительности накопителя 126.

Поэтому результатом выполнения первого алгоритма планирования (например, алгоритма «справедливого» планирования) является первый спланированный порядок операций I/O (например, «справедливо» спланированный порядок операций I/O).

На реализуемый первым планировщиком 310 алгоритм «справедливого» планирования не накладывается каких-либо особых ограничений. Тем не менее, в одном не имеющем ограничительного характера примере первый планировщик 310 может быть реализован с использованием алгоритма планирования DRR, т.е. первый планировщик 310 может быть реализован как планировщик типа DRR.

Независимо от реализуемого первым планировщиком 310 конкретного вида алгоритма «справедливого» планирования (т.е. независимо от конкретного первого алгоритма планирования), первый планировщик 310 способен определять порядок, в котором операции I/O из первого набора 302 операций I/O и из второго набора 304 операций I/O должны передаваться для выполнения накопителем 126, чтобы производительность накопителя 126 использовалась «справедливо».

Шаг 406: передача первого спланированного порядка операций I/O во второй планировщик операций I/O.

Способ продолжается на шаге 406, где первый планировщик 310 способен передать первый спланированный порядок («справедливо» спланированный порядок) операций I/O во второй планировщик 320. Например, первый спланированный порядок операций I/O может передаваться первым планировщиком 310 во второй планировщик 320 по линии 316 связи. Как упомянуто выше, первый спланированный порядок операций I/O может включать в себя «справедливо» спланированные операции I/O из первого набора 302 операций I/O и из второго набора 304 операций I/O.

Предполагается, что в некоторых вариантах осуществления данной технологии первый планировщик 310 может передавать первый спланированный порядок («справедливо» спланированный порядок) операций I/O непосредственно во второй планировщик 320.

Например, вместо передачи первого спланированного порядка операций I/O в промежуточный стек операций I/O, расположенный между первым планировщиком 310 и вторым планировщиком 320, и последующей передачи операций I/O из промежуточного стека операций I/O во второй планировщик 320, первый планировщик 310 может передавать первый спланированный порядок операций I/O непосредственно во второй планировщик 320 без использования промежуточного стека операций I/O.

Следовательно, предполагается, что линия 316 связи может представлять собой прямую линию связи между первым планировщиком 310 и вторым планировщиком 320, чтобы операции I/O, передаваемые от первого планировщика 310 во второй планировщик 320 не принимались или не перехватывались каким-либо промежуточным стеком операций I/O, расположенным между первым планировщиком 310 и вторым планировщиком 320. Предполагается, что в гибридном планировщике 300 может не требоваться промежуточный стек операций I/O между первым планировщиком 310 и вторым планировщиком 320.

Шаг 408: выполнение вторым планировщиком операций I/O второго алгоритма планирования с целью определения перепланированного порядка операций I/O для первого спланированного порядка операций I/O, причем второй алгоритм планирования основан на соответствующих предельных сроках выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O.

Способ 400 продолжается на шаге 408, где второй планировщик 320 выполняет второй алгоритм планирования «для реального времени».

Как описано выше, чтобы выполнить требования относительно режима реального времени к системе 100 распределенной обработки информации, второй планировщик 320 способен обеспечивать схему планирования «для реального времени», позволяющую выполнять операции I/O до истечения соответствующих предельных сроков. Иными словами, второй планировщик 320 способен упорядочивать (или в данном случае переупорядочивать) операции I/O для их передачи в накопитель 126 с целью выполнения так, чтобы операции I/O могли выполняться накопителем 126 до истечения соответствующих предельных сроков.

Можно сказать, что, как упомянуто выше, второй планировщик 320 использует алгоритм планирования «для реального времени» с целью формирования спланированного «для реального времени» порядка операций I/O на основе соответствующих предельных сроков выполнения операций I/O.

Поэтому результатом выполнения второго алгоритма планирования (например, алгоритма планирования «для реального времени») на основе первого спланированного порядка операций I/O (например, «справедливо» спланированного порядка операций I/O) является перепланированный порядок операций I/O (например, спланированный «для реального времени» порядок операций I/O).

На реализуемый вторым планировщиком 320 алгоритм планирования «для реального времени» не накладывается каких-либо особых ограничений. Тем не менее, в одном не имеющем ограничительного характера примере второй планировщик 320 может быть реализован с использованием алгоритма планирования EDF, т.е. второй планировщик 320 может быть реализован как планировщик типа EDF.

Независимо от реализуемого вторым планировщиком 320 конкретного вида алгоритма планирования «для реального времени» (т.е. независимо от конкретного второго алгоритма планирования), второй планировщик 320 способен определять спланированный «для реального времени» порядок принятых операций I/O,

позволяющий накопителю 126 выполнять операции I/O до истечения их соответствующих предельных сроков.

Шаг 410: контроль гибридным планировщиком операций I/O соблюдения соответствующих предельных сроков выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O.

Способ 400 продолжается на шаге 410, где гибридный планировщик 300 осуществляет контроль соблюдения соответствующих предельных сроков выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O.

В некоторых вариантах осуществления это означает, что гибридный планировщик 300 способен контролировать возможность выполнения до истечения соответствующих предельных сроков по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O (например, этот контроль, помимо прочего, может включать в себя априорный анализ, как описано выше).

Это также означает, что гибридный планировщик 300 дополнительно или в качестве альтернативы способен контролировать фактическое выполнение до истечения соответствующих предельных сроков по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O (например, этот контроль, помимо прочего, может включать в себя апостериорный анализ, как описано выше).

Шаг 412: приостановка приема вторым планировщиком операций I/O дополнительных операций I/O от первого планировщика операций I/O, если определено несоблюдение соответствующего предельного срока выполнения по меньшей мере одной операции I/O.

Способ продолжается на шаге 412, где приостанавливается прием вторым планировщиком 320 дополнительных операций I/O от первого планировщика 310, если определено несоблюдение соответствующего предельного срока выполнения по меньшей мере одной операции I/O.

В некоторых вариантах осуществления гибридный планировщик 300, выполняющий на шаге 410 контроль с использованием априорного анализа, может определить невозможность выполнения накопителем 126 до истечения соответствующего предельного срока по меньшей мере одной операции I/O, передаваемой в накопитель 126 согласно перепланированному порядку операций I/O.

Например, гибридный планировщик 300 может определить невозможность выполнения накопителем 126 до истечения соответствующего предельного срока по меньшей мере одной операции I/O, передаваемой в накопитель 126 согласно перепланированному порядку операций I/O, когда второй планировщик 320 неспособен определить спланированный «для реального времени» порядок, согласно которому операции I/O могут быть выполнены до истечения их соответствующих предельных сроков.

В других вариантах осуществления гибридный планировщик 300, выполняющий на шаге 410 контроль с использованием априорного анализа, может определить невозможность выполнения накопителем 126 до истечения соответствующего предельного срока по меньшей мере одной операции I/O, если от первого планировщика 310 принимается дополнительная операция I/O.

Например, второй планировщик 320 может определить, что после приема дополнительной операции I/O и добавления ее в перепланированный порядок операций I/O на основе соответствующего предельного срока выполнения накопитель 126 будет неспособен выполнить до истечения соответствующего предельного срока по меньшей мере одну операцию I/O согласно перепланированному порядку.

Гибридный планировщик 300, выполняющий на шаге 410 контроль с использованием априорного анализа, дополнительно или в качестве альтернативы может определять выполнение по меньшей мере одной операции I/O без соблюдения соответствующего предельного срока путем определения фактического невыполнения накопителем 126 до истечения соответствующего предельного срока по меньшей мере одной операции I/O.

Например, гибридный планировщик 300 может быть способен принимать подтверждения выполнения операций I/O, ранее переданных гибридным планировщиком 300 в накопитель 126 для их выполнения. Если выполнение этих операций I/O происходит после истечения соответствующего срока, гибридный планировщик 300 может определить несоблюдение соответствующего предельного срока выполнения по меньшей мере одной операции I/O.

Следует отметить, что приостановка приема вторым планировщиком 320 дополнительных операций I/O от первого планировщика 310 выполняется в результате селективного переключения гибридного планировщика 300 из обычного режима в режим высокой нагрузки.

В некоторых вариантах осуществления изобретения приостановка приема вторым планировщиком 320 дополнительных операций I/O от первого планировщика 310 включает в себя прекращение приема вторым планировщиком 320 дополнительных операций I/O.

В других вариантах осуществления изобретения приостановка приема вторым планировщиком 320 дополнительных операций I/O от первого планировщика 310 включает в себя передачу вторым планировщиком 320 сигнала 342 обратной связи в первый планировщик 310. При получении сигнала 342 обратной связи первый планировщик 310 прекращает отправку дополнительных операций I/O во второй планировщик 320, как описано выше.

Предполагается, что гибридный планировщик 300, работающий в обычном режиме или в режиме высокой нагрузки, способен дополнительно принимать (а) по меньшей мере одну дополнительную операцию I/O, сформированную первым источником операций I/O, и (б) по меньшей мере одну дополнительную операцию I/O, сформированную вторым источником операций I/O.

Предполагается, что эта по меньшей мере одна дополнительная операция I/O, принятая первым планировщиком 310 в режиме высокой нагрузки, не возвращается в соответствующий источник операций I/O. Эта по меньшей мере одна дополнительная операция I/O, принятая первым планировщиком 310 в режиме высокой нагрузки, сохраняется в первом планировщике 310, пока гибридный планировщик 300 селективно не переключится в обычный режим работы.

Предполагается, что второй планировщик 320 способен возобновлять прием дополнительных операций I/O от первого планировщика 310.

В некоторых вариантах осуществления данной технологии гибридный планировщик 300 способен определять, что принятые вторым планировщиком 320 операции I/O выполняются с соблюдением соответствующих предельных сроков, как описано выше. Предполагается, что в других вариантах осуществления второй планировщик 320

способен возобновить прием дополнительных операций I/O от первого планировщика 310, если определено, что принятые вторым планировщиком 320 операции I/O выполняются с соблюдением соответствующих предельных сроков.

5 Следует отметить, что возобновление приема вторым планировщиком 320 дополнительных операций I/O от первого планировщика 310 выполняется в результате селективного переключения гибридного планировщика 300 из режима высокой нагрузки в обычный режим.

Предполагается, что в некоторых вариантах осуществления данной технологии гибридный планировщик 300 способен работать в обычном режиме, когда второй
10 планировщик 320 принимает дополнительные операции I/O от первого планировщика 310. Гибридный планировщик 300 также способен работать в режиме высокой нагрузки, когда второй планировщик 320 приостанавливает прием дополнительных операций I/O от первого планировщика 310.

Предполагается, что гибридный планировщик 300 в обычном режиме работы
15 способен обеспечивать (а) «справедливое» распределение производительности накопителя 126 для первого и второго источников операций I/O; и (б) выполнение требований первого и второго источников операций I/O относительно режима реального времени.

Также предполагается, что гибридный планировщик 300 в режиме высокой нагрузки
20 способен обеспечивать «справедливое» распределение производительности накопителя 126 для первого и второго источников операций I/O.

Для специалиста в данной области техники очевидны возможные изменения и усовершенствования описанных выше вариантов осуществления данной технологии. Предшествующее описание приведено в качестве примера, а не для ограничения объема
25 изобретения. Объем охраны данной технологии определяется исключительно объемом приложенной формулы изобретения.

(57) Формула изобретения

1. Способ планирования операций I/O (ввода/вывода) для их выполнения накопителем,
30 реализованным как часть системы распределенной компьютерной обработки информации, содержащей:

- множество источников операций I/O, от которых поступают операции I/O, включая первый источник операций I/O и второй источник операций I/O, каждый из которых связан с соответствующей заранее заданной долей производительности накопителя
35 для выполнения соответствующих операций I/O первого источника I/O и второго источника I/O; и

- гибридный планировщик операций I/O, содержащий первый планировщик операций I/O и второй планировщик операций I/O;

способ, выполняемый гибридным планировщиком операций I/O и включающий в
40 себя:

- прием первым планировщиком операций I/O (а) по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и (б) по меньшей мере одной операции I/O, сформированной вторым источником операций I/O;

- выполнение первым планировщиком операций I/O первого алгоритма планирования
45 с целью определения спланированного порядка для по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O, причем первый алгоритм планирования основан на соответствующих заранее заданных долях

производительности накопителя, связанных с первым источником операций I/O и вторым источником операций I/O, причем результатом выполнения первого алгоритма планирования является первый спланированный порядок операций I/O;

- передачу первого спланированного порядка операций I/O во второй планировщик операций I/O;

- выполнение вторым планировщиком операций I/O второго алгоритма планирования с целью определения перепланированного порядка операций I/O для первого спланированного порядка операций I/O, причем второй алгоритм планирования основан на соответствующих предельных сроках выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O;

- контроль гибридным планировщиком операций I/O соблюдения соответствующих предельных сроков выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O,

сформированной вторым источником операций I/O;

- приостановку приема вторым планировщиком операций I/O дополнительных операций I/O от первого планировщика операций I/O, если определено несоблюдение соответствующего предельного срока выполнения по меньшей мере одной операции I/O.

2. Способ по п. 1, отличающийся тем, что контроль соблюдения соответствующих предельных сроков выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O, включает в себя контроль вторым планировщиком операций I/O возможности выполнения до истечения соответствующих предельных сроков по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O.

3. Способ по п. 2, отличающийся тем, что определение несоблюдения соответствующего предельного срока выполнения по меньшей мере одной операции I/O включает в себя определение вторым планировщиком операций I/O невозможности выполнения накопителем до истечения соответствующего предельного срока по меньшей мере одной операции I/O, передаваемой в накопитель согласно перепланированному порядку операций I/O.

4. Способ по п. 2, отличающийся тем, что определение несоблюдения соответствующего предельного срока выполнения по меньшей мере одной операции I/O включает в себя определение вторым планировщиком операций I/O невозможности выполнения накопителем до истечения соответствующего предельного срока по меньшей мере одной операции I/O, если от первого планировщика операций I/O принимается дополнительная операция I/O.

5. Способ по п. 1, отличающийся тем, что контроль соблюдения соответствующих предельных сроков выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O, включает в себя контроль гибридным планировщиком операций I/O фактического выполнения до истечения соответствующих предельных сроков по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O.

6. Способ по п. 5, отличающийся тем, что определение несоблюдения

соответствующего предельного срока выполнения по меньшей мере одной операции I/O включает в себя определение гибридным планировщиком операций I/O фактического невыполнения накопителем до истечения соответствующего предельного срока по меньшей мере одной операции I/O.

5 7. Способ по п. 1, отличающийся тем, что приостановка приема включает в себя прекращение приема дополнительных операций I/O вторым планировщиком операций I/O.

8. Способ по п. 1, отличающийся тем, что приостановка приема включает в себя передачу вторым планировщиком операций I/O в первый планировщик операций I/O сигнала обратной связи, при получении которого первый планировщик операций I/O
10 прекращает отправку дополнительных операций I/O во второй планировщик операций I/O.

9. Способ по п. 1, отличающийся тем, что дополнительно включает в себя:

- прием первым планировщиком операций I/O по меньшей мере одной дополнительной
15 операции I/O, сформированной первым источником операций I/O; и
- прием первым планировщиком операций I/O по меньшей мере одной дополнительной операции I/O, сформированной вторым источником операций I/O.

10. Способ по п. 9, отличающийся тем, что дополнительно включает в себя возобновление вторым планировщиком операций I/O приема дополнительных операций
20 I/O от первого планировщика операций I/O.

11. Способ по п. 10, отличающийся тем, что возобновление приема выполняется, если гибридный планировщик операций I/O определяет, что дополнительные операции I/O, принятые вторым планировщиком операций I/O, выполняются с соблюдением соответствующих предельных сроков.

25 12. Способ по п. 1, отличающийся тем, что первый планировщик операций I/O представляет собой планировщик типа DRR.

13. Способ по п. 1, отличающийся тем, что второй планировщик операций I/O представляет собой планировщик типа EDF.

14. Способ по п. 1, отличающийся тем, что гибридный планировщик операций I/O
30 работает:

- в обычном режиме, когда второй планировщик операций I/O принимает дополнительные операции I/O от первого планировщика операций I/O; и
- в режиме высокой нагрузки, когда второй планировщик операций I/O приостанавливает прием дополнительных операций I/O от первого планировщика
35 операций I/O.

15. Способ по п. 14, отличающийся тем, что:

- в обычном режиме гибридный планировщик операций I/O обеспечивает «справедливое» распределение производительности накопителя для первого и второго источников операций I/O и соответствует требованиям первого и второго источников
40 операций I/O относительно режима реального времени; и

- в режиме высокой нагрузки гибридный планировщик операций I/O обеспечивает «справедливое» распределение производительности накопителя для первого и второго источников операций I/O.

16. Гибридный планировщик операций I/O (ввода/вывода) для планирования операций
45 I/O с целью их выполнения накопителем, реализованным как часть системы распределенной компьютерной обработки информации, содержащей множество источников операций I/O, от которых поступают операции I/O, включая первый источник операций I/O и второй источник операций I/O, связанные с соответствующими заранее

заданными долями производительности накопителя для выполнения соответствующих операций I/O первого источника I/O и второго источника I/O, содержащий первый планировщик операций I/O и второй планировщик операций I/O и выполненный с возможностью выполнения следующих действий:

- 5 - прием первым планировщиком операций I/O (а) по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и (б) по меньшей мере одной операции I/O, сформированной вторым источником операций I/O;
 - выполнение первым планировщиком операций I/O первого алгоритма планирования с целью определения спланированного порядка для по меньшей мере одной операции
10 I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O, причем первый алгоритм планирования основан на соответствующих заранее заданных долях производительности накопителя, связанных с первым источником операций I/O и вторым источником операций I/O, а результатом выполнения первого алгоритма
15 планирования является первый спланированный порядок операций I/O;
 - передача первого спланированного порядка операций I/O во второй планировщик операций I/O;
 - выполнение вторым планировщиком операций I/O второго алгоритма планирования с целью определения перепланированного порядка операций I/O для первого
20 спланированного порядка операций I/O, причем второй алгоритм планирования основан на соответствующих предельных сроках выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O;
 - контроль соблюдения соответствующих предельных сроков выполнения по меньшей
25 мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O;
 - приостановка приема вторым планировщиком операций I/O дополнительных
 операций I/O от первого планировщика операций I/O, если определено несоблюдение
30 соответствующего предельного срока выполнения по меньшей мере одной операции I/O.

17. Гибридный планировщик по п. 16, отличающийся тем, что возможность контроля гибридным планировщиком операций I/O соблюдения соответствующих предельных
сроков выполнения по меньшей мере одной операции I/O, сформированной первым
35 источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O, включает в себя возможность контроля возможности выполнения до истечения соответствующих предельных сроков по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O.

40 18. Гибридный планировщик по п. 17, отличающийся тем, что возможность определения гибридным планировщиком операций I/O несоблюдения соответствующего предельного срока выполнения по меньшей мере одной операции I/O включает в себя возможность определения невозможности выполнения накопителем до истечения соответствующего предельного срока по меньшей мере одной операции I/O,
45 передаваемой в накопитель согласно перепланированному порядку операций I/O.

19. Гибридный планировщик по п. 17, отличающийся тем, что возможность определения гибридным планировщиком операций I/O несоблюдения соответствующего предельного срока выполнения по меньшей мере одной операции I/O включает в себя

возможность определения невозможности выполнения накопителем до истечения соответствующего предельного срока по меньшей мере одной операции I/O, если от первого планировщика операций I/O принимается дополнительная операция I/O.

20. Гибридный планировщик по п. 16, отличающийся тем, что возможность контроля гибридным планировщиком операций I/O соблюдения соответствующих предельных сроков выполнения по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O, включает в себя возможность контроля фактического выполнения до истечения соответствующих предельных сроков по меньшей мере одной операции I/O, сформированной первым источником операций I/O, и по меньшей мере одной операции I/O, сформированной вторым источником операций I/O.

21. Гибридный планировщик по п. 20, отличающийся тем, что возможность определения гибридным планировщиком операций I/O несоблюдения соответствующего предельного срока выполнения по меньшей мере одной операции I/O включает в себя возможность определения фактического невыполнения накопителем до истечения соответствующего предельного срока по меньшей мере одной операции I/O.

22. Гибридный планировщик по п. 16, отличающийся тем, что возможность приостановки приема гибридным планировщиком операций I/O включает в себя возможность прекращения приема дополнительных операций I/O вторым планировщиком операций I/O.

23. Гибридный планировщик по п. 16, отличающийся тем, что возможность приостановки приема гибридным планировщиком операций I/O включает в себя возможность передачи вторым планировщиком операций I/O в первый планировщик операций I/O сигнала обратной связи, при получении которого первый планировщик операций I/O прекращает отправку дополнительных операций I/O во второй планировщик операций I/O.

24. Гибридный планировщик по п. 16, отличающийся тем, что он выполнен с возможностью выполнения следующих действий:

- прием первым планировщиком операций I/O по меньшей мере одной дополнительной операции I/O, сформированной первым источником операций I/O; и
- прием первым планировщиком операций I/O по меньшей мере одной дополнительной операции I/O, сформированной вторым источником операций I/O.

25. Гибридный планировщик по п. 24, отличающийся тем, что он дополнительно выполнен с возможностью возобновления приема вторым планировщиком операций I/O дополнительных операций I/O от первого планировщика операций I/O.

26. Гибридный планировщик по п. 25, отличающийся тем, что он выполнен с возможностью возобновления приема, если гибридный планировщик операций I/O определяет, что дополнительные операции I/O, принятые вторым планировщиком операций I/O, выполняются с соблюдением соответствующих предельных сроков.

27. Гибридный планировщик по п. 16, отличающийся тем, что первый планировщик операций I/O представляет собой планировщик типа DRR.

28. Гибридный планировщик по п. 16, отличающийся тем, что второй планировщик операций I/O представляет собой планировщик типа EDF.

29. Гибридный планировщик по п. 16, отличающийся тем, что он выполнен с возможностью работы:

- в обычном режиме, когда второй планировщик операций I/O принимает дополнительные операции I/O от первого планировщика операций I/O; и
- в режиме высокой нагрузки, когда второй планировщик операций I/O

приостанавливает прием дополнительных операций I/O от первого планировщика операций I/O.

30. Гибридный планировщик по п. 16, отличающийся тем, что:

- в обычном режиме гибридный планировщик операций I/O обеспечивает

5 «справедливое» распределение производительности накопителя для первого и второго источников операций I/O и соответствует требованиям первого и второго источников операций I/O относительно режима реального времени; и

- в режиме высокой нагрузки гибридный планировщик операций I/O обеспечивает «справедливое» распределение производительности накопителя для первого и второго

10 источников операций I/O.

15

20

25

30

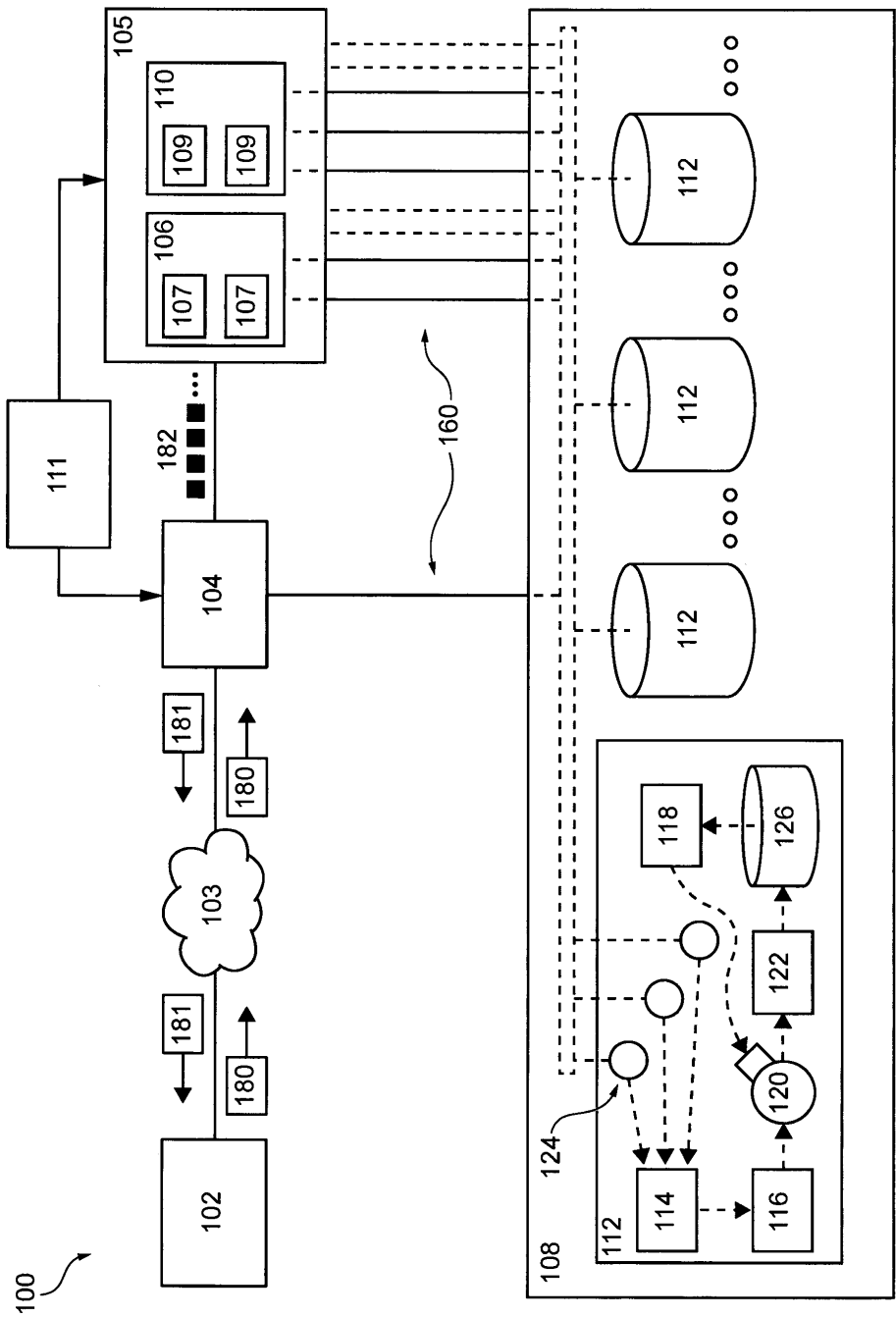
35

40

45

1

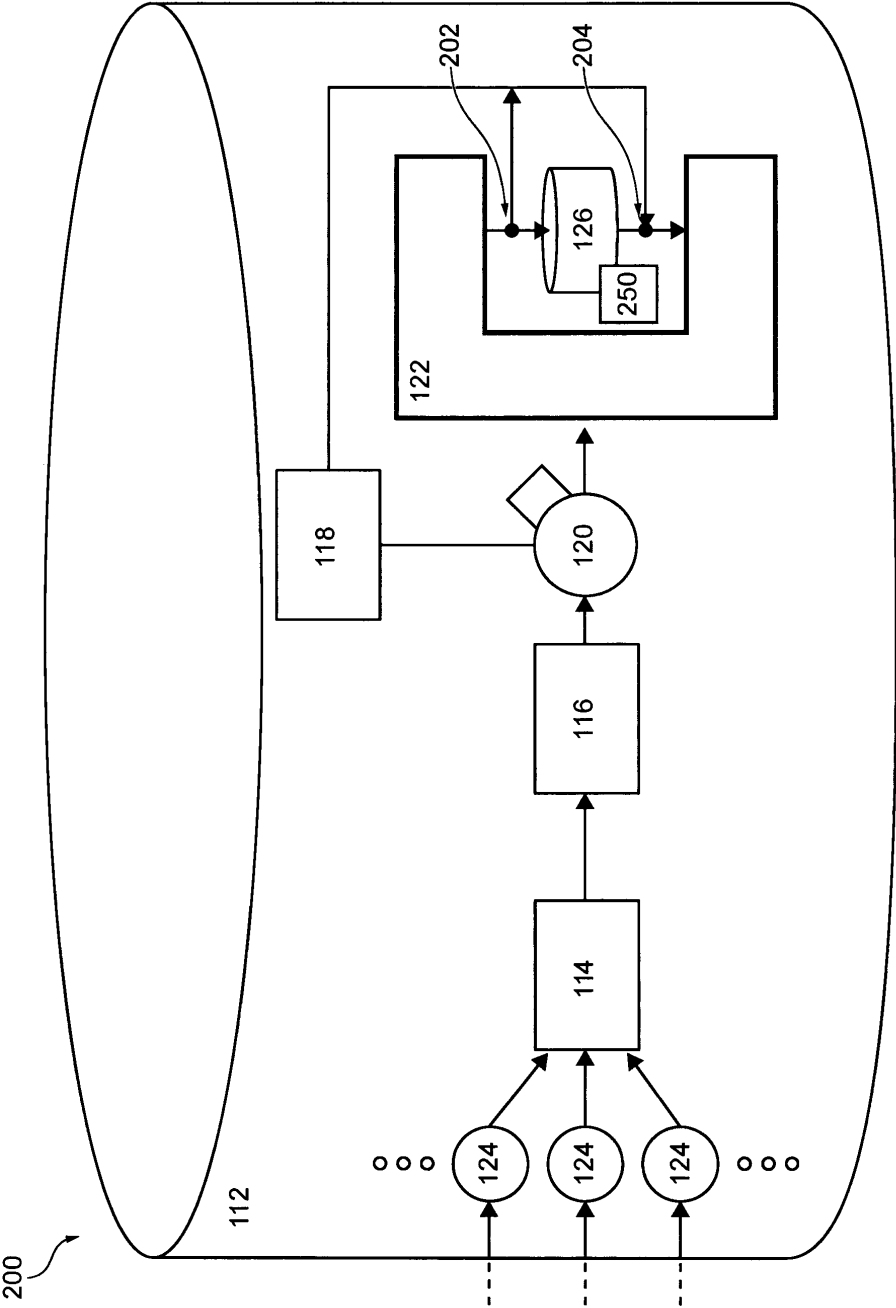
1/4



ФИГ. 1

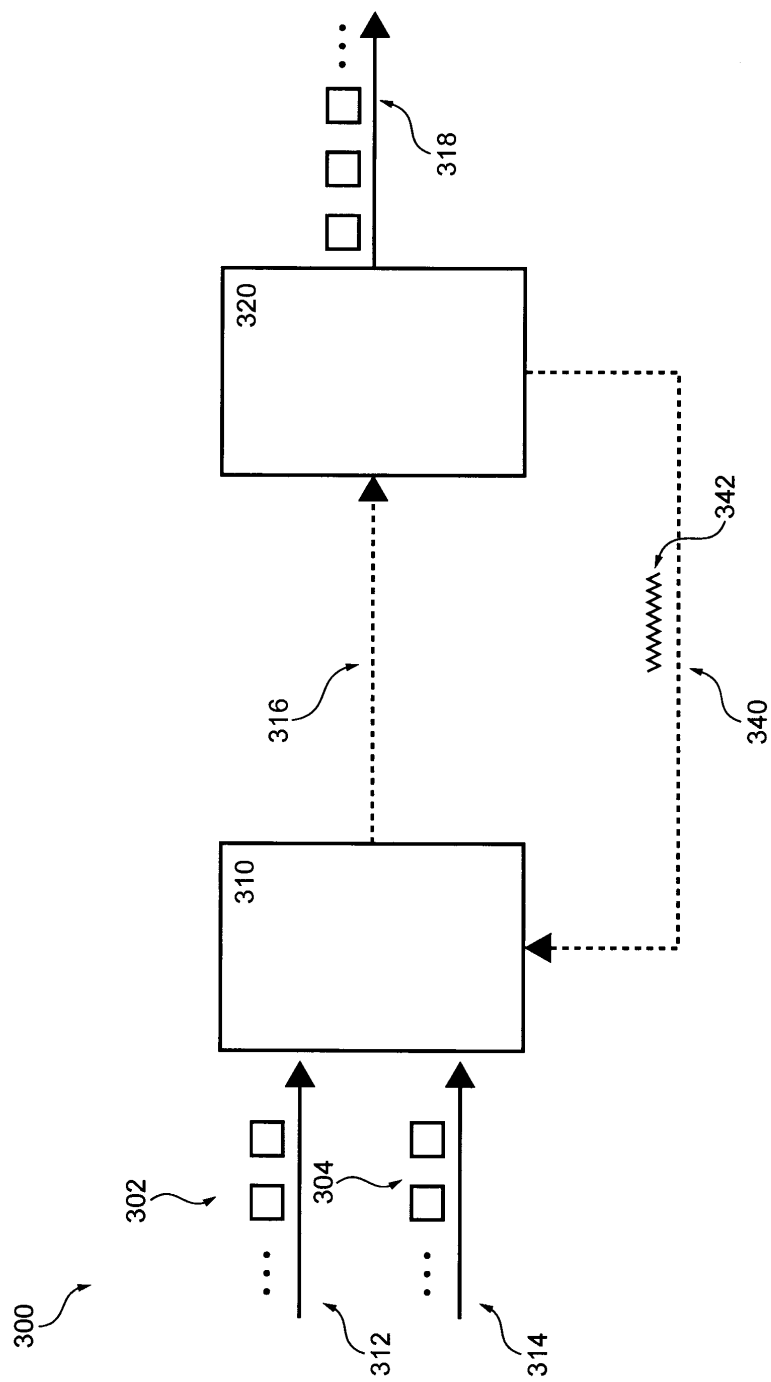
2

2/4

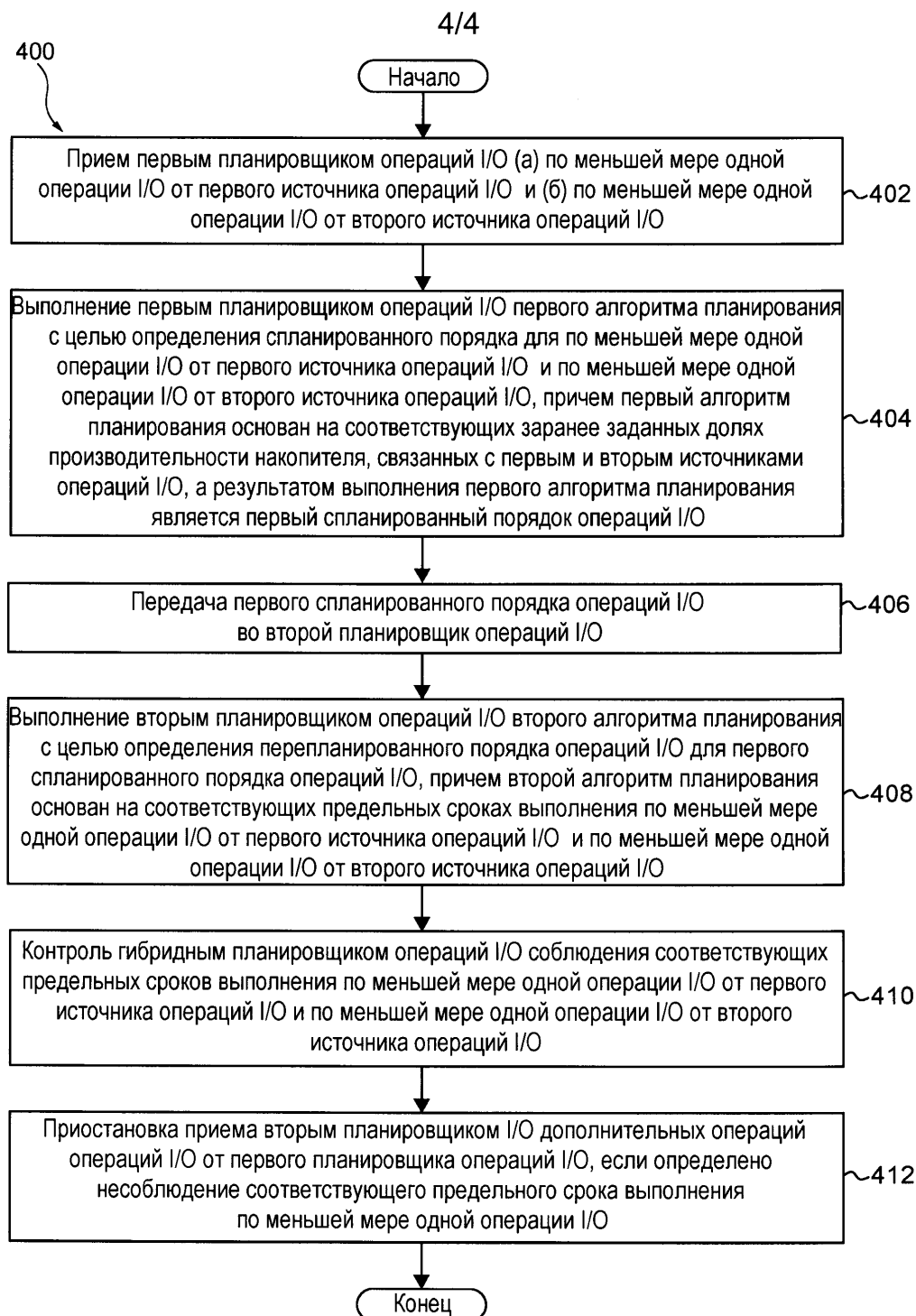


ФИГ. 2

3/4



ФИГ. 3



Фиг. 4