

a circular disc rotating at a constant angular velocity in restricted space, downfall of water column, the problem of a body falling into water, the problem of a sinking body, the problem of sinking model ship, simulation of a boat on the waves, rotation of propelling screw VP1304 in free water, etc. The physical parameters obtained according to the numerical simulation results were compared with the available experimental data and theoretical studies.

Key words: LOGOS program package, shipbuilding, verification basis, free surface, rotation.

УДК 004.416.6

КОЛЛЕКТИВНЫЙ ДОСТУП К ФАЙЛОВЫМ ДАННЫМ НА ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМАХ С РАЗЛИЧНОЙ АРХИТЕКТУРОЙ В БИБЛИОТЕКЕ ЕФР

К. К. Олесницкая, И. А. Антипин, М. А. Петрова

Российский федеральный ядерный центр –
Всероссийский НИИ экспериментальной физики, Саров

В докладе представлен обзор функциональных возможностей библиотеки ЕФР версии 3.3 для организации коллективного доступа к файловым данным на вычислительных системах с различной архитектурой.

Ключевые слова: ЕФР, библиотека ЕФР, коллективный доступ к файловым данным.

Введение

Для решения задач эффективного¹ доступа к файловым данным на вычислительных системах (ВС) с различной архитектурой в математическом отделении РФЯЦ-ВНИИЭФ разрабатывается кроссплатформенная библиотека ЕФР [1].

Для прикладных пакетов программ библиотека ЕФР предоставляет единый функциональный интерфейс работы с файловыми данными и единый формат представления расчетных данных. Что позволяет сохранять все необходимые данные как для продолжения счета в рамках одного пакета программ, так и для продолжения счета в рамках нескольких пакетов программ при сопряженных расчетах, а также для передачи в другой пакет программ при поэтапных расчетах.

Использование библиотеки ЕФР в подавляющем большинстве пакетов программ математического отделения позволяет сосредоточить решение проблем файлового доступа на программном уровне в рамках библиотеки и освободить разработчиков пакетов программ от необходимости менять свои методы чтения/записи, при переходе между вычислительными системами с различной архитектурой.

¹ Эффективный доступ – это использование стратегии доступа к файловым данным с учетом аппаратных и файловых ресурсов конкретной ВС

Для решения данной задачи в библиотеке ЕФР реализован коллективный доступ, под которым понимается доступ к файловым данным группы процессов с использованием приемлемых по скорости коммуникационных связей² и учетом наличия доступа к файловой системе и ее особенностей.

Механизмы коллективного доступа, реализованные в библиотеке ЕФР, позволяют на программном уровне регулировать нагрузку на файловую систему и скрыть от разработчиков пакетов программ особенности той или иной ВС, например, отсутствие доступа к файловой системе с ряда вычислительных устройств. Что, в свою очередь, позволяет создавать универсальные методы чтения/записи файловых данных в рамках конкретного пакета программ для работы на вычислительных системах с различной архитектурой.

Проблематика

При использовании в математическом отделении ВС на базе различных архитектур отчетливо проявились две проблемы. Во-первых, существенно возросла нагрузка на файловую систему, из-за увеличения количества клиентов файловой системы, которое обусловлено ростом вычислительных мощностей. Во-вторых, оптимальное использование ВС, которые имеют аппаратные особенности, такие как отсутствие прямого доступа к файловой системе с ряда вычислительных устройств, требует значительной доработки методов чтения/записи файловых данных в существующих прикладных пакетах программ.

Поэтому возникла необходимость в создании связующего звена между файловой системой и программой пользователя, которое снизит нагрузку на файловую систему и обеспечит независимость программы пользователя от размещения данных на физическом устройстве хранения.

Функцию такого связующего звена выполняет библиотека ЕФР, которая обеспечивает масштабируемый доступ к файловым данным и является неотъемлемым компонентом технологии проведения расчетов на ВС математического отделения.

Задачи

Таким образом, совместно со специалистами, отвечающими за аппаратные и файловые ресурсы ВС и разработчиками прикладных пакетов программ, были определены задачи, которые необходимо решить для достижения поставленной цели.

Во-первых, необходимо снизить нагрузку на параллельную файловую систему (ПФС), уменьшив: количество результирующих файлов; количество файлов, записываемых одновременно, и количество процессов, работающих с одним файлом в режиме записи.

Во-вторых, необходимо обеспечить доступ к файловым данным на ВС с аппаратными особенностями и обеспечить переносимость методов чтения/записи между ВС с различной архитектурой.

Решением поставленных задач является реализация коллективного доступа к файловым данным в библиотеке ЕФР и поддержка его в прикладных пакетах программ.

Причем реализовать коллективный доступ необходимо таким образом, чтобы это не привело к существенным изменениям в библиотеке ЕФР, а его поддержка в прикладных пакетах программ требовала минимум трудозатрат.

Коллективный доступ к файловым данным

В подавляющем большинстве пакетов программ методы чтения/записи файловых данных реализуют классическую схему работы с ЕФР (рис. 1).

² Взаимодействие процессов осуществляется через MPI [2] (стандарт интерфейса передачи информации между процессами).

То есть в параллельном режиме расчетные данные сохраняются в виде распределенного по файлам разреза, где каждый процесс независимо от других (в асинхронном режиме) формирует свой локальный файл-фрагмент, который содержит локальный фрагмент задачи с процесса. Затем управляющий процесс формирует головной файл, который содержит глобальные данные о задаче и информацию о распределении данных по файлам-фрагментам.

Созданный таким образом ЕФР является распределенным, а доступ к данным становится автоматически масштабируемым, то есть данные можно прочитать на произвольном числе процессов. Для чтения такого распределенного ЕФР каждый процесс независимо от других открывает только головной файл и работает с ним как с нераспределенным. А функции доступа к данным на основе анализа информации о распределении данных автоматически открывают необходимые файлы-фрагменты и считывают из них необходимые порции данных.

Для решения поставленных задач необходимо изменить схему работы с ЕФР, чтобы запись локальных фрагментов задачи с процессов в файлы-фрагменты и чтение локальных фрагментов задачи процессами из головного файла выполнялись в коллективном режиме (рис. 2).



Рис. 1. Классическая схема работы с ЕФР



Рис. 2. Коллективная схема работы с ЕФР

Таким образом, в рамках библиотеки ЕФР необходимо было реализовать средства управления режимом коллективной работы, реализовать средства синхронизации методов доступа к файловым данным, которые могут быть не на всех процессах и поддержать коллективный режим работы функций доступа к файловым данным.

Управление режимом коллективной работы

Под коллективным режимом работы понимается режим, когда функция выполняется группой процессов с коммуникационной связью между ними. Главное отличие коллективного режима от асинхронного состоит в том, что все процессы, входящие в коммуникационную группу, должны осуществить вызов одинаковых коллективных функций в одном и том же порядке. Если порядок вызова функций будет нарушен, это будет означать невыполнение требований коллективного режима работы функций доступа, и может привести к «зависанию» или аварийному завершению работы программы.

Для установки коллективного режима работы библиотека предоставляет пользователю специализированные функции, в которые передается информация о группе процессов, участвующих в коллективном доступе и путь к файловому ресурсу, к которому будет осуществляться доступ.

Информация о характеристиках ВС определяется динамически на этапе выполнения программы с помощью специализированной системной библиотеки (разработка математического отделения). На текущий момент используется следующая информация: количество файловых серверов ПФС, режим доступа к ПФС с процесса, информация о приемлемых по скорости коммуникационных связях между процессами и имя устройства, с которого запущен процесс.

Формирование коммуникационных групп

Используя приведенную ранее информацию, выполняется распределение процессов по коммуникационным группам. Рассмотрим алгоритм распределения процессов по группам для записи на примере (рис. 3). Пусть задана коммуникационная группа процессов, которые будут осуществлять коллективный доступ в режиме записи. Процессы, выделенные одним цветом, принадлежат одному вычислительному узлу, которое состоит из нескольких разнородных по производительности и архитектуре устройств. Над процессами указан режим доступа к ПФС (RW – полный доступ, R- – только чтение и -- – доступ к ПФС отсутствует). Для процессов, которые не имеют доступ к ПФС в режиме записи, пунктирной линией указаны приемлемые по скорости коммуникационные связи с другими процессами.

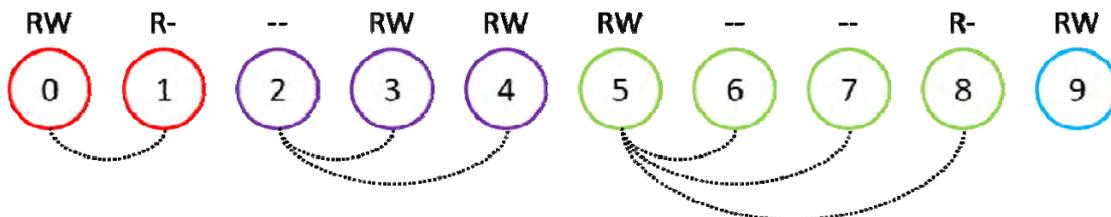


Рис. 3. Коммуникационная группа процессов на запись

Итак, формирование коммуникационных связей для коллективной записи выполняется следующим образом (рис. 4):

1. Определяется количество узлов с доступом к ПФС в режиме записи (в примере 5 процессов с рангами: 0, 3, 4, 5 и 9).
2. Определяется количество файлов-фрагментов:

а) задается пользователем (в примере задано 3 файла-фрагмента);

б) определяется автоматически (один файл-фрагмент с узла, для данного примера в автоматическом режиме было бы создано 4 файла-фрагмента).

3. Определяется количество файловых серверов ПФС (в примере ПФС управляет работой двух файловых серверов).

4. Выполняется равномерное распределение процессов без доступа к ПФС в режиме записи по обменным группам (ОГ), в соответствии с информацией о приемлемой скорости коммуникационных связей, а процесс с доступом на запись назначается управляющим и ему присваивается ранг 0 (для примера будет создано пять ОГ).

5. Обменные группы равномерно распределяются по файлам-фрагментам. Для каждого процесса определяется файл-фрагмент и коммуникатор группы процессов для работы с ним – группа на файл (ГФ) (для примера будет создано три ГФ, так как пользователем задано 3 файла-фрагмента).

6. Группы по файлам равномерно распределяются по файловым серверам. Для каждого процесса логически определяется файловый сервер и коммуникатор группы процессов для работы с ним – группа на файловый сервер (ГФС) (в примере будет создано две ГФС, так как ПФС управляет работой двух файловых серверов).

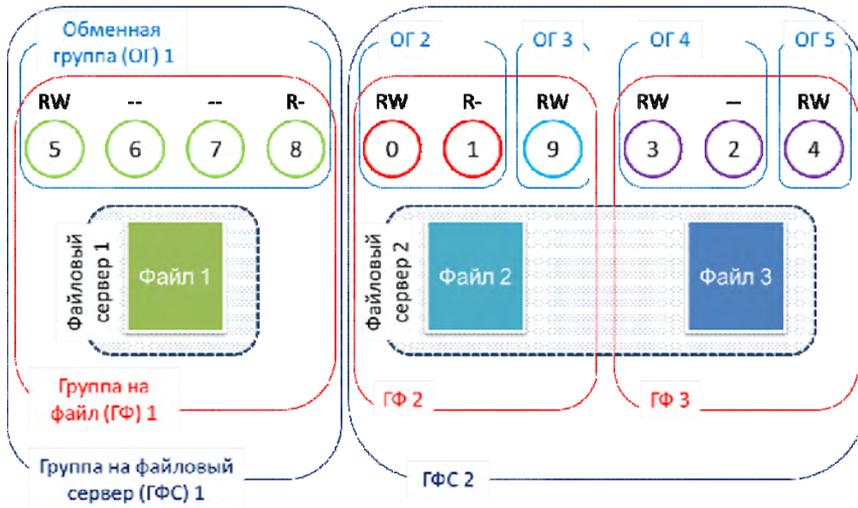


Рис. 4. Распределение процессов для коллективной записи

Таким образом, в результате распределения получены две очереди упорядоченных процессов для одновременного доступа к файловым данным в режиме записи. Первая состоит из процессов с рангами: 5, 6, 7 и 8, а вторая состоит из процессов с рангами: 0, 1, 9, 3, 2 и 4.

При формировании коммуникационной группы на чтение формируются только обменные группы (рис. 5).

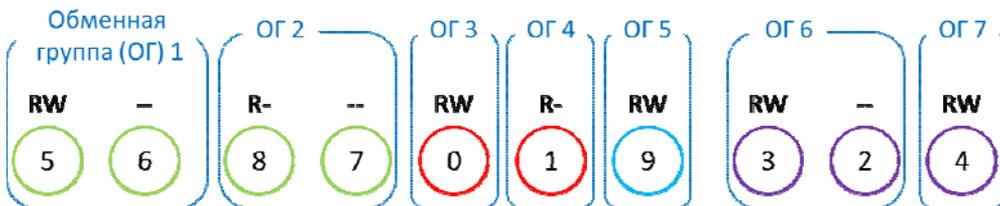


Рис. 5. Распределение процессов для коллективного чтения

Таким образом, в результате распределения получено семь очередей упорядоченных процессов для одновременного доступа к файловым данным в режиме чтения.

Средства синхронизации методов доступа к файловым данным

Для синхронизации методов доступа к файловым данным, которые могут быть не на всех процессах, пользователю предлагается вызывать на некоторых процессах функции доступа «вхолостую». При этом должно выполняться обязательное требование неизменного порядка вызова одинаковых функций доступа в коллективном режиме работы. Библиотека предоставляет средства, которые позволяют отметить как фрагмент кода, который будет работать «вхолостую», так и отдельную функцию.

Коллективный режим работы функций доступа к файловым данным

Библиотека ЕФР на текущий момент содержит более 500 функций, а поддержать коллективный доступ необходимо таким образом, чтобы это не привело к существенным изменениям программного кода. Поэтому первым этапом был определен ряд функций доступа, которые являются наиболее эффективными при работе с файлом (например, объектный доступ является более эффективным в сравнении с поэлементным). Сделано это для того, чтобы прикладные пакеты программ, поддерживав коллективный доступ, перешли на наиболее эффективные методы доступа к файловым данным.

Затем была создана оболочка для скалярных функций, которые должны работать в коллективном режиме. Для этого был разработан шаблонный класс, экземпляром которого является функциональный объект (функтор), который действует как функция, но может также хранить состояние. Цель данного класса в упаковке скалярной функции и входных параметров в функтор, который передается в схему коллективного доступа, которая может:

- Осуществлять вызов функции.
- Осуществлять прием и передачу входных (IN) параметров функции посредством коммуникационных связей.
- Осуществлять вызов функции с принятыми параметрами.
- Осуществлять прием и передачу выходных (OUT) параметров функции и возвращаемого значения посредством коммуникационных связей.

Необходимость пересылки параметров функции определяется с помощью флагов IN и OUT. Флаги устанавливаются автоматически на этапе создания функтора по атрибуту const соответствующего параметра, то есть если у параметра имеется данный атрибут, то параметру выставляется флаг IN. С целью уменьшения лишних пересылок предусмотрено ручное задание данных флагов (например, можно снять флаги, чтоб параметр не участвовал в пересылках, или установить оба флага, чтоб параметр участвовал в двухсторонней пересылке для IN-OUT параметров).

Общая схема коллективного доступа

По общей схеме коллективного доступа одновременно в параллельном режиме работает по одной обменной группе из всех групп процессов для работы с файловым сервером. Алгоритм работы процессов в рамках обменной группы представлен на рис. 6.

Кратко описать алгоритм можно следующим образом: управляющий процесс обменной группы при необходимости выполняет свою функцию, а затем при необходимости выполняет функции с других процессов, а другие процессы обменной группы при необходимости выполнения функции, передают управление вызовом своей функции управляющему процессу.

Внедрение

Для внедрения коллективного доступа в прикладные пакеты программ библиотека ЕФР предоставляет пользователям средства журналирования своих функций, которые позволяют получить полную информацию о том, как работала пользовательская программа доступа к файловым данным. В лог выводится информация о ранге процесса и имени устройства, с которого он запущен, названия функций и их входные параметры в хронологическом порядке их вызовов.

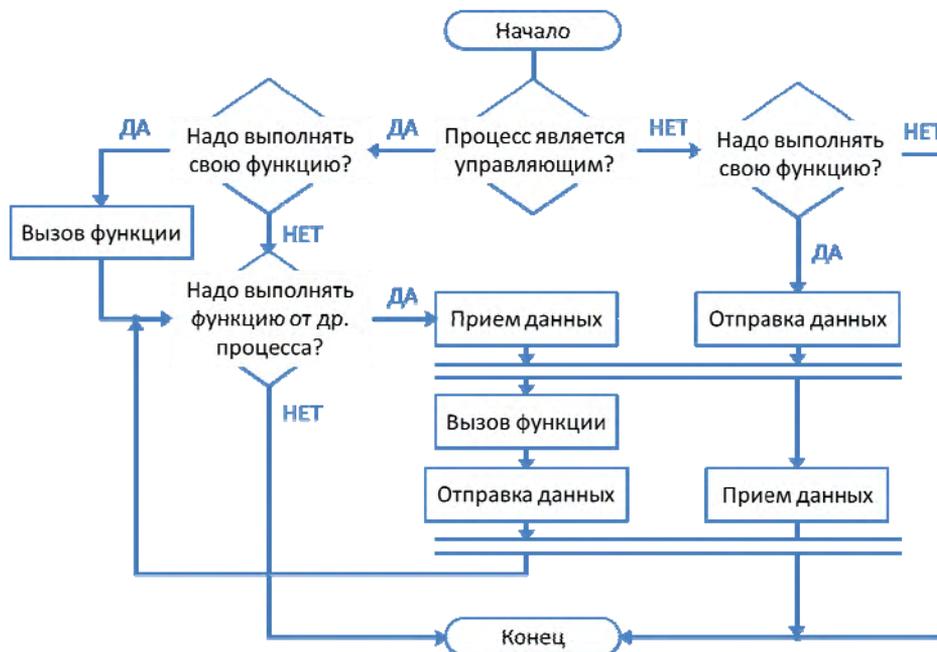


Рис. 6. Блок-схема алгоритма работы процессов в обменной группе

Также пользователям предоставлены средства эмуляции запуска программы на неоднородном вычислительном поле на ВС без аппаратных особенностей. То есть на программном уровне процессам задаются необходимые настройки коммуникационных связей и доступа к ПФС.

Коллективный доступ поддержан в ряде прикладных пакетов программ математического отделения ВНИИЭФ (например, ЛОГОС-ТМП[3], ЭГИДА[4] и др.). В ходе проведения расчетов получены приемлемые результаты. Скорость программ доступа к файловым данным напрямую зависит от количества процессов в обменных группах. Соответственно, чем больше процессов, которые не имеют доступ к ПФС, тем дольше выполняется программа доступа к файловым данным. Например, при использовании процессов в отношении 1:2 (на один процесс с полным доступом к ПФС приходится один процесс без доступа к ПФС) время выполнения программы доступа замедляется в два раза по отношению ко времени выполнения только на процессах с полным доступом к ПФС. В отношении 1:3 замедляется в три раза и т. д. Но хочется отметить, что проигрыш по времени доступа к файловым данным в пакете программ компенсируется уменьшением времени счета благодаря тому, что появилась возможность использовать более производительные устройства, имеющие особенности архитектуры.

Заключение

В докладе описываются механизмы коллективного доступа, реализованные в библиотеке ЕФР, которые позволяют на программном уровне регулировать нагрузку на файловую систему и скрыть от разработчиков пакетов программ особенности той или иной ВС, например, отсутствие

доступа к файловой системе с ряда вычислительных устройств. Что, в свою очередь, позволяет создавать универсальные методы чтения/записи файловых данных в рамках конкретного пакета программ для работы на вычислительных системах с различной архитектурой.

Реализация коллективного доступа не потребовала существенных изменений в функциональном ядре библиотеки ЕФР за счет реализации автономного модуля и оболочки для функций, которые должны работать в параллельном режиме, а его поддержка в прикладных пакетах программ не потребовала много трудозатрат.

На текущий момент методы коллективного доступа к файловым данным успешно используются в ряде параллельных решателей, входящих в пакет программ «ЛОГОС» [5].

Литература

1. Олесницкая К. К., Антипин И. А., Петрова М. А. Библиотека ЕФР как средство эффективного доступа к файловым данным на гибридных вычислительных системах и суперкомпьютерах // XV Международная конференция «Супервычисления и математическое моделирование»: сб. науч. тр. / под ред. Р. М. Шагалиева. – Саров: ФГУП «РФЯЦ-ВНИИЭФ», 2015. – С. 246 – 354.

2. MPI: A Message-Passing Interface Standard Version 2.2.[Electronic resource]. Mode of access: <http://www.mpi-forum.org/docs/mpi-2.2/mpi22-report.pdf>.

3. Козелков А. С., Дерюгин Ю. Н., Зеленский Д. К., Глазунов В. А., Голубев А. А., Денисова О. В., Лашкин С. В., Жучков Р. Н., Тарасова Н. В., Сизова М. А. Многофункциональный пакет программ ЛОГОС для расчета задач гидродинамики и тепломассопереноса на суперЭВМ. Базовые технологии и алгоритмы // XII Международный семинар «Супервычисления и математическое моделирование»: сб. докл. Саров, 11-15 октября 2010. – Саров: РФЯЦ-ВНИИЭФ, 2011. С. 215 – 230.

4. Янилкин Ю. В., Беляев С. П., Городничев А. В., Куделькин В. Г., Линяев С. А., Смоляков А. А., Тарасов В. И., Царева Т. В., Чистякова И. Н., Шабуров В. М. Код ЭГИДА для моделирования двумерных и трехмерных задач механики сплошной среды // Научно-техническая конференция «Научное программное обеспечение в образовании и научных исследованиях»: сб. науч. тр. – Санкт-Петербург: Издат-во Политехнического ун-та, 2008. – С. 310 – 313.

5. ЛОГОС. Пакет программ инженерного анализа и суперкомпьютерного моделирования [Электронный ресурс] – Саров: ФГУП «РФЯЦ-ВНИИЭФ», 2018 – Режим доступа: <http://logos.vniief.ru>.

SHARABLE ACCESS TO THE FILE DATA ON COMPUTING SYSTEMS WITH DIFFERENT ARCHITECTURES IN THE EFR LIBRARY

K. K. Olesnitskaya, I. A. Antipin, M. A. Petrova

Russian Federal Nuclear Center –
All-Russian Research Institute of Experimental Physics, Sarov

The paper provides the review of functional capabilities of the EFR library Version 3.3 to establish sharable access to the file data at the computational systems with different architectures.

Key words: EFR, EFR library, shared access to file data.