

БИБЛИОТЕКА EFR. НОВЫЕ ВОЗМОЖНОСТИ

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

ФГУП «РФЯЦ-ВНИИЭФ», г. Саров Нижегородской обл.

Введение

В рамках общего интерфейса, создаваемого в математическом отделении РФЯЦ-ВНИИЭФ, разрабатывается библиотека EFR [1]. Библиотека представляет собой инструмент разработчика программных математических комплексов, позволяющий в кратчайшие сроки создавать программы формирования результатов расчетов в виде единого файлового разреза (ЕФР), и обеспечивает:

- обмен данными между различными пакетами прикладных программ на основе стандартизированного представления сеточных данных;
- единый интерфейс взаимодействия с сервисными программами для постобработки и визуализации сеточных данных;
- проблемно-ориентированный масштабируемый ввод-вывод информации на многопроцессорных ЭВМ.

В математическом отделении РФЯЦ-ВНИИЭФ библиотека EFR является составной частью программ общего сервиса (препроцессор, постпроцессор), таких как ЛОГОС.ПРЕПОСТ, и неотъемлемым компонентом технологии проведения расчетов на многопроцессорных ЭВМ.

Постоянное развитие программ общего сервиса и совершенствование технологии проведения расчетов на многопроцессорных ЭВМ тесно связаны с развитием библиотеки EFR. На 5-й конференции «Молодежь в науке» был сделан доклад, посвященный библиотеке EFR, поэтому данный доклад кратко напомнит о назначении и внутренней организации библиотеки EFR и подробно расскажет о текущем состоянии и о наиболее значимых новых возможностях, появившихся за последнее время. К наиболее существенным изменениям можно отнести:

- переход на новое шаблонное ядро;
- оптимизацию алгоритмов доступа к данным;
- введение новой структурной единицы;
- введение нового формата описания нерегулярной сеточной топологии;
- реализацию новых технологий работы с распределенными данными.

1. Библиотека EFR

1.1. Назначение библиотеки EFR

Библиотека EFR представляет собой инструмент разработчика программных математических комплексов, позволяющий сохранять данные в виде дво-

ичного файла-разреза и считывать сохраненные данные посредством методов прямого доступа к файлу (рис. 1). Полученный ЕФР (единый файловый разрез) имеет гибкую модель представления данных, позволяющую сохранять все необходимые данные как для продолжения счета в рамках одного пакета программ, так и для передачи в другой пакет программ для полномасштабных сквозных расчетов. Также можно сохранять выборочный набор данных для визуализации, табличного анализа или в результате предобработки (РНД). Во всех случаях данные ЕФР пригодны для постобработки, табличного анализа и визуализации программами общего сервиса.

Обращение к файловому разрезу осуществляется через внешние функции и процедуры прямого доступа, объединенные в библиотеку. Внутренний формат файлового разреза имеет бинарное представление и максимально скрыт от пользователей.

Библиотека EFR является кроссплатформенной, а внешний интерфейс библиотеки ориентирован на языки Фортран 90 и C/C++, являющиеся основными языками программирования для программ общего сервиса.



Рис. 1. Назначение библиотеки EFR

1.2. Структура ЕФР

С точки зрения пользователя ЕФР представляет собой иерархическую многоуровневую структуру объектов (рис. 2).



Рис. 2. Структура данных ЕФР

Объект – это метаданные и контейнер, включающий в себя иерархическую структуру объектов следующих уровней. Количество объектов каждого уровня не ограничено и динамически увеличивается по мере необходимости. Исключением являются объекты четвертого уровня, ограничивающиеся группами идентификаторов.

Метаданные представляют собой совокупность служебной информации, предназначенной для работы функций прямого доступа к двоичному файловому разрезу.

Объектом первого уровня является структура данных, содержащая информацию о параметрах задачи и файловом разрезе в целом: номер версии ЕФР, размерность задачи, число математических областей, система единиц, учетные характеристики, время счета и т. д.

Объектами второго уровня являются структуры данных, каждая из которых содержит информацию о параметрах какой-либо математической области. Основным параметром математической области является тип топологии области (регулярная сеточная топология, регулярно-адаптивная сеточная топология, нерегулярная сеточная топология и бессеточная топология). К остальным параметрам относятся: количество временных шагов, количество веществ; количество сеточных узлов, ячеек, ребер или граней; размерность сеточной матрицы (для регулярной сетки), число вершин, образующих вспомогательные линии и триангуляционные поверхности, не связанные с сеткой, число самих вспомогательных линий и поверхностей и т. д.

Помимо характерных размеров к параметрам области относятся интегральные физические величины. С помощью данных величин можно легко сравнить (в первом приближении) результаты расчетов одной задачи в разных пакетах прикладных программ.

Объектами третьего уровня являются временные шаги. Временной шаг относится к области, максимальное количество временных шагов для области не ограничено и динамически увеличивается по мере необходимости. Каждый временной шаг представляет собой контейнер объектов данных (многомерные массивы, сеточные массивы, списки) и метаданные.

Объектами четвертого уровня являются массивы и списковые структуры данных, содержащие описание топологии сетки, значения сеточных функций, нестандартные параметры области, таблицы граничных условий и многое другое. Другими словами, объекты четвертого уровня представляют собой множество расчетных данных, относящихся к какой-либо математической области. В этом множестве данных выделено три основных класса:

- Сеточные массивы.
- Произвольные многомерные массивы, не связанные с сеткой.
- Списковые структуры (списки данных).

Данные классы различаются как по структуре представления самого объекта, так и по набору его параметров. Ключевым параметром для каждого объ-

екта третьего уровня является его целочисленный идентификатор, представляющий собой уникальный номер этого объекта в пределах математической области.

1.3. Типы математических областей

Библиотека EFR ориентирована на схему раздельного счета по математическим областям. Таким образом, область представляет собой двумерную или трехмерную пространственную сетку. На текущий момент ЕФР поддерживает следующие типы математических областей:

- область с регулярной сеточной топологией;
- область с регулярно-адаптивной сеточной топологией;
- область с нерегулярной сеточной топологией;
- область с результатами моделирования бессеточными методами.

Под регулярной топологией сетки подразумевается матричное расположение ее узлов. Ячейки двумерной сетки являются произвольными четырехугольниками, совокупность которых образует топологический параллелограмм. Ячейки трехмерной сетки являются произвольными шестигранниками, совокупность которых образует топологический параллелепипед. Описание топологии регулярной сетки сводится к заданию размерностей двумерной или трехмерной матрицы ее узлов (в зависимости от размерности задачи) и определению декартовых координат каждого из узлов. Определением координат узлов является задание двух двумерных (X, Y) или трех трехмерных вещественных массивов (X, Y, Z) в зависимости от размерности задачи. Размерности данных массивов определяются размерностями матрицы узлов.

Адаптивная сетка рассматривается как дополнение к регулярной сетке, выполняющей роль опорного шаблона, т. е. каждая ячейка адаптивной сетки является «дочерней» по отношению к какой-либо ячейке регулярной сетки.

Область с нерегулярной сеточной топологией определяется количеством узлов, ячеек, ребер или граней. Сеточные объекты, определяющие координаты узлов сетки, дополняются объектами, содержащими описание сеточной топологии (более подробно о форматах описания нерегулярных данных будет рассказано ниже).

Область с результатами моделирования бессеточными методами представляет собой множество узлов и предназначена для хранения рассчитанных данных в частицах, на кластерах и т. д.

2. Новые возможности

С момента выпуска первой версии библиотека EFR претерпела значительные изменения. Самым существенным является переход на новое шаблонное ядро, которое улучшило скоростные показатели всего функционала в целом и сделало максимально простым процесс разработки новых возможностей.

На текущий момент в библиотеке появился ряд принципиально новых вещей:

- Новый объект данных – временной шаг.
- Новый тип математической области – область с результатами моделирования бессеточными методами.
- Новый формат описания нерегулярной сеточной топологии.
- Новая технология работы с распределенными данными.
- Новый функционал для работы с объектами данных.

Далее рассмотрим наиболее важные новые возможности.

2.1. Новый объект данных – временной шаг

В процессе счета часто возникает необходимость сохранения «контрольных точек», т. е. промежуточных значений рассчитываемых величин. Для этого служит временной шаг. Данные временных шагов позволяют в реальном времени наблюдать динамику протекающих физических процессов во время счета, а также предоставляют возможность проведения повторных расчетов с нужной «контрольной точки».

Ранее для сохранения временных шагов создавался новый самодостаточный EFR-разрез, содержащий в себе полный набор данных, насчитанных на текущий момент времени. В этот набор входят и те данные, которые не изменились с предыдущего шага. Дублирование объемных одинаковых данных привело к увеличению времени счета задачи и неэффективному использованию дискового пространства.

Удачным решением стало введение временного шага как объекта данных в рамках библиотеки EFR. Таким образом, теперь математическая область представляет собой набор временных шагов. Временной шаг представляет собой метаданные и контейнер объектов данных (многомерные массивы, сеточные массивы, списки). Нулевой временной шаг является основным и содержит определяющий набор сеточных данных, необходимых в математической

счетной области. Последующие временные шаги служат для сохранения сеточных данных, изменяющихся с течением времени.

Использование временных шагов как нового объекта данных позволило сократить общее время счета задачи, позволило создавать разрезы меньших объемов и ускорило процесс визуализации сеточных данных.

2.2. Форматы описания нерегулярной сеточной топологии

Изначально для описания нерегулярной сеточной топологии в библиотеке EFR был разработан универсальный формат. В процессе использования оказалось, что при конечно-элементном моделировании формат очень избыточен, что нецелесообразно увеличивает объем данных и время формирования разрезов.

Следующим немаловажным нововведением является поддержка компактного ячеечно-узлового формата. Таким образом, в текущей версии библиотеки существует два формата описания нерегулярной сеточной топологии: ребро-граневый и ячеечно-узловой.

Ребро-граневый формат задания нерегулярной сеточной топологии является универсальным и подходит для описания ячеек в виде произвольных многогранников с неплоскими гранями. При таком подходе описание сводится к описанию узлов, граней и ячеек, где грань – это множество узлов, а ячейка – множество граней (рис. 3,а).

Ячеечно-узловой формат задания нерегулярной сеточной топологии является более компактным. В ее описании используются узлы и ячейки, причем описание ячейки происходит в соответствии с ее типом и строго определенным порядком узлов. Строго определенный порядок узлов однозначно определяет грани, поэтому их описание в данном формате не требуется. Набор типов ячеек по мере необходимости может расширяться (рис. 3,б).

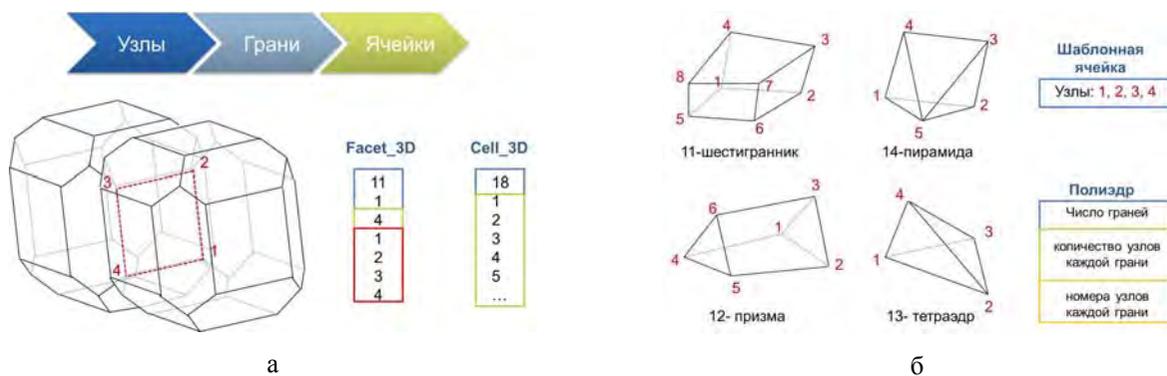


Рис. 3. Форматы описания нерегулярной сеточной топологии

2.3. Технология работы с распределенными данными

Одной из основных задач библиотеки EFR является обеспечение проблемно-ориентированного масштабируемого доступа к информации на многопроцессорных ЭВМ. В многопроцессорном режиме расчетные данные удобно сохранять в виде распределенного по файлам разреза, где каждый процессор формирует свой локальный файл-фрагмент.

Файл-фрагмент представляет собой обычный нераспределенный файл ЕФР, содержащий следующие данные:

- параметры задачи;
- локальные параметры математической области для конкретного процессора;
- локальные фрагменты сеточных данных, присутствующих в математической области.

Получаемый таким образом набор файлов-фрагментов дополняется головным файлом (рис. 4), содержащим:

- параметры задачи;
- глобальные параметры математической области;
- описание фрагментов математической области (это количество фрагментов и карта фрагментов с описанием их декомпозиции);
- параметры сеточных массивов, распределенных по фрагментам.



Рис. 4. Распределенный ЕФР

Созданный таким образом ЕФР-разрез является распределенным, а доступ к данным становится автоматически масштабируемым (рис. 5), т. е. данные можно будет прочитать на произвольном числе процессоров. Для чтения такого распределенного разреза каждому процессору будет достаточно открыть только головной файл-разрез и работать с ним как с нераспределенным. А функции доступа к сеточным данным на основе анализа карты фрагментов будут автоматически открывать необходимые файлы-фрагменты и считывать из них необходимые порции данных.

Создание распределенного разреза (рис. 6) можно выполнять одним из двух способов:

- «Один к одному», т. е. каждый процессор сохраняет собственный рассчитываемый фрагмент сеточных данных в отдельный файл-фрагмент.
- «Многие ко многим», т. е. один или более процессоров сохраняет рассчитанный фрагмент(ы) сеточных данных в один или более файлов-фрагментов. Причем M снизу ограничено единицей, а сверху –

только пространством файловых имен. На данный момент этот механизм реализован для областей с регулярной сеточной топологией. Но имеется возможность поддержать этот механизм и для областей с нерегулярной сеточной структурой. Основной задачей данного механизма является решение проблем, связанных с огромным количеством мелких файлов, например, такой, как переполнение файлового пространства имен.



Рис. 5. Масштабируемый доступ к данным ЕФР

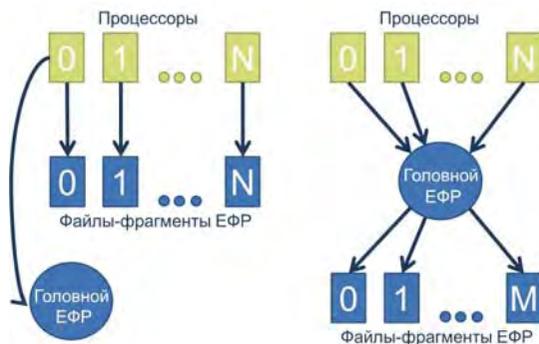


Рис. 6. Создание распределенного ЕФР

2.3.1. Функционал для работы в параллельном режиме

В качестве новой возможности библиотека EFR предоставила функционал для работы с распределенными файлами-разрезами в параллельном режиме:

- Сжатие и распаковка данных.
- Проверка стандартных объектов.
- Поиск экстремумов объектов.
- Выборка, поиск-замена данных объектов.

Это значительно ускорило процесс постобработки распределенных разрезов.

2.3.2. Новые технологии

Помимо нового функционала, в библиотеке EFR появились две новые технологии:

- технология работы с распределенной регулярной-адаптивной областью;
- технология работы с распределенной нерегулярной областью.

Основной задачей новых технологий является задача обеспечить масштабируемый доступ к сеточным данным в этих областях.

Распределение области с регулярно-адаптивной сеточной топологией базируется на распределении регулярной области, которое производится заданием соответствия глобальных сеточных фрагментов необходимым файлам-фрагментам. Таким образом, распределение адаптивных данных осуществляется на основе регулярных. А вся адаптивная информация хранится в элементах сеточных списков. Элементом такого списка является контейнер данных, структура которого зависит от типа адаптивности.

Данная технология представления адаптивных данных имеет ряд значительных преимуществ:

- легко обобщается на случай нерегулярно-адаптивных сеточных топологий;
- поддерживает возможность описания сложной(многоуровневой) адаптивной сетки;
- автоматически поддерживает работу с распределенными адаптивными данными.

Распределение нерегулярной сеточной топологии производится по ячейкам, узлам и граням. Новая технология основана на полной информации о связях локальной и глобальной нумераций сеточных элементов.

Первая связь представляет собой соответствие локальной нумерации сеточных элементов глобальной (рис. 7). Данная информация хранится в каждом файле-фрагменте.

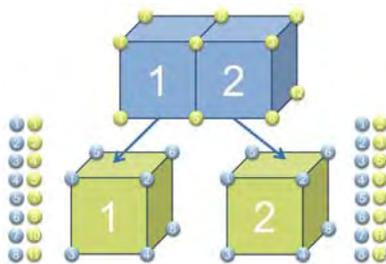


Рис. 7. Распределенная нерегулярная область

Вторая связь представляет собой соответствие глобальных номеров сеточных элементов локальным в каждом фрагменте (ведется учет сеточных «перехлестов» между фрагментами). Данная информация хранится в головном файле и является системной картой фрагментов (рис. 8).

На текущий момент пользователи имеют два инструмента для создания системной карты фрагментов, обеспечивающей масштабируемость сеточных данных:

- первый инструмент строит системную карту фрагментов на основе анализа первой связи, заданной пользователем;
- второй инструмент сначала выстраивает первую связь, анализируя сеточную топологию, затем строит системную карту фрагментов на основе анализа первой.

При отсутствии системной карты фрагментов доступ к распределенным данным будет осуществ-

ляться, как и раньше, без учета дублирующийся информации. За счет сквозной нумерации элементов (ячеек, узлов, ребер/граней) по всем файлам-фрагментам.

Новая технология работы с распределенными нерегулярными сеточными данными позволила обеспечить:

- масштабируемый доступ к сеточным данным;
- быстрый доступ к сеточным данным по глобальному номеру сеточного элемента;
- возможность чтения данных без «перехлестов» по фрагментам;
- возможность чтения распределенных топологических списков с автоматическим преобразованием локальной нумерации сеточных элементов к глобальной;
- поддержку несквозной, глобальной нумерации сеточных элементов.

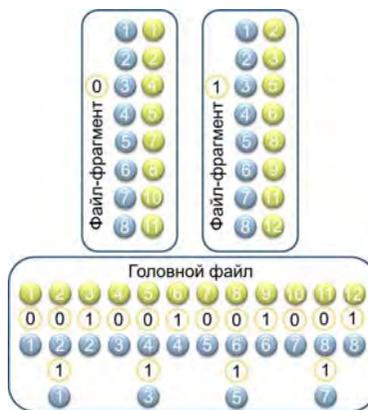


Рис. 8. Системная карта фрагментов

Заключение

В заключение хочется отметить, что библиотека EFR обеспечивает эффективное решение следующих задач:

- Оперативное сохранение расчетных данных в едином формате для всех комплексов прикладных программ математического отделения РФЯЦ-ВНИИЭФ.
- Связь между различными комплексами прикладных программ на основе стандартизированного представления расчетных данных.
- Постобработка и визуализация расчетных данных программами общего сервиса на базе единого интерфейса взаимодействия с прикладными программами.

Литература

1. Волгин А. В., Красов А. В., Кузнецов М. Ю., Тарасов В. И. Библиотека EFR для универсального представления расчетных данных // Труды РФЯЦ-ВНИИЭФ. Саров: РФЯЦ-ВНИИЭФ, 2007. Вып. 11. С. 130–135.