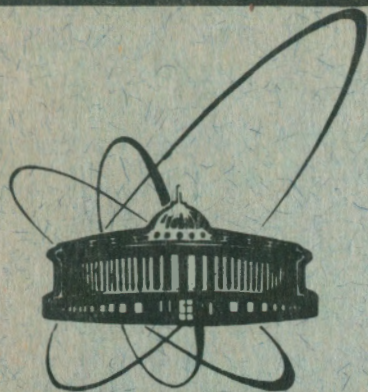


91-80



сообщения
объединенного
института
ядерных
исследований
дубна

P10-91-80

В.В.Глаголев, В.Г.Иванов, Т.Эрдэнэдэлгэр

ВОПРОСЫ ПРОГРАММНОГО СОПРОВОЖДЕНИЯ
МАТЕМАТИЧЕСКОЙ ОБРАБОТКИ
ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

1991

Современные высокоавтоматизированные системы математической обработки экспериментальных данных являются большими и сложными программными комплексами, состоящими из многих компонент, которые вызываются в определенной последовательности. При этом результаты работы одних программ являются исходными данными для одной или нескольких программ соответствующей последовательности. Объемы перерабатываемой и хранимой информации в ходе массовой обработки экспериментальных данных исчисляются сотнями мегабайт, а для ряда новых экспериментов десятками и сотнями гигабайт^{1/}. В связи с этим в системах используются различные уровни памяти (магнитные диски, магнитные ленты, картриджи и автоматические загрузчики), а сами файлы с данными довольно часто перемещаются с одного уровня на другой. Вследствие этого требуется строгое соблюдение технологической дисциплины в ходе массовой обработки данных и аккуратная регистрация статуса данных системы после окончания работы любой из программ комплекса. Традиционно результаты пропуска задач на ЭВМ ведутся в специальных журналах, что при относительно небольших объемах перерабатываемой информации и невысоких скоростях счета не вызывает особых затруднений, хотя и приводит к потерям времени из-за допускаемых пользователями ошибок.

Создание в ЦВК ОИЯИ многомашинного комплекса ЕС ЭВМ с общей дисковой памятью и включение в его состав ЭВМ ЕС-1066 позволили существенно ускорить процесс математической обработки данных. Так, например, если раньше на ЭВМ ЕС-1061 на обработку 1000 событий, полученных в ¹⁶Ор-эксперименте, требовалось 6-7 часов процессорного времени, то на ЭВМ ЕС-1066 - около полутора часов.

Наличие общей дисковой памяти ЕС ЭВМ позволяет пользователям комплекса работать на любой из его машин (ЕС-1037, ЕС-1061 и ЕС-1066), используя файлы общей памяти, независимо от того, на какой из ЭВМ были получены результаты предыдущих сеансов, а также пересылать задачи с одной ЭВМ на другую.

Для полного использования возможностей многомашинного комплекса ЕС ЭВМ в ходе массовой обработки экспериментальных данных традиционный метод ручной регистрации стал "узким местом" системы. Для его ликвидации была разработана и реализована методика программного сопровождения математической обработки фильмовой информации в рамках операционной системы СВМ^{2/}, которая позволяет автоматически регистрировать результаты пропуска задач на специальном файле и тем самым "исключить" ручное ведение "журнала обработки".

При разработке методики предполагалось, что в ходе ее применения основным рабочим инструментом пользователя будет ПЭВМ типа IBM, на диске которой будет храниться специальная база данных. В последней в компактной форме будут заноситься протоколы прохождения всех задач системы с тем, чтобы пользователь мог в любой момент получить информацию о текущем состоянии дел процесса обработки и на основе ее принять соответствующее решение. В настоящее время эта методика и соответствующее программное обеспечение используется для обработки данных ¹⁶Ор-эксперимента. В дальнейшем на основании полученного опыта будет производиться ее адаптация к другим экспериментам.

Основные этапы математической обработки фильмовой информации

Основные этапы процесса математической обработки фильмовой информации на примере анализа данных в нуклон-ядерных взаимодействиях^{/3/} и соответствующая схема движения потока экспериментальных данных показаны на рис.1.

Поступающие на вход системы магнитные ленты с результатами измерений (ЛРИ) камерных фотографий обрабатываются цепочкой программ, осуществляющих сортировку, проверку и перекодировку результатов измерений событий. Основным результатом их работы является файл результатов измерений (ФРИ) в стандартном формате системы "Гидра"^{/4/}.

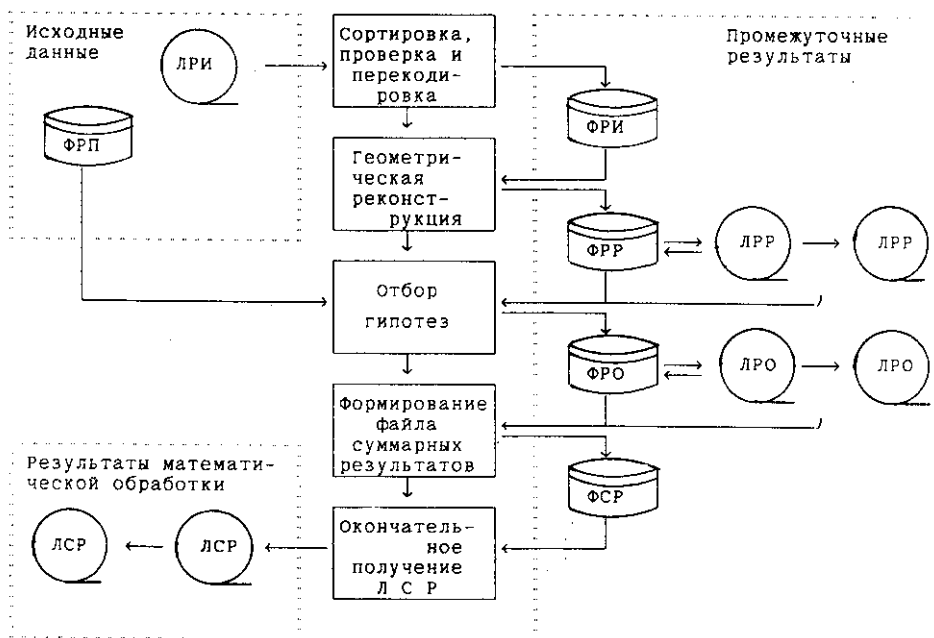


Рис. 1

Затем работает геометрическая программа, которая восстанавливает пространственную картину измеренных событий и вычисляет параметры образующих их треков заряженных частиц для задаваемых наборов гипотез об их массах. Эти данные накапливаются на файле результатов реконструкции (ФРР) и после его заполнения переписываются на магнитные ленты (ЛРР) с помощью специальной процедуры.

На следующем этапе анализа из результатов реконструкции выбираются и переписываются на файл результатов отбора (ФРО) только те данные, которые соответствуют приписанным при просмотре зарядам и массам. В дальнейшем содержимое ФРО переписывается на магнитные ленты (ЛРО). На основании ФРО (ЛРО) формируется файл суммарных результатов (ФСР). А на завершающем этапе формируются ленты суммарных результатов (ЛСР) данного эксперимента.

Для обеспечения надежности хранения данных файлы на лентах ЛРР, ЛРО и ЛСР нередко дублируются.

Из выше рассмотренного следует, что математическая обработка фильмовой информации - это достаточно сложный и динамичный процесс, и его аккуратное протоколирование необходимо для прослеживания "истории" обработки каждой из ЛРИ (группы событий которые были переписаны с МЛ на диск; этапы, через которые они прошли, где хранятся результаты счета по каждой из программ системы). Эта информация нужна не только для того, чтобы знать состояние дел с обработкой, но и формировать новые задания ЭВМ.

Программная регистрация протоколов прохождения задач на ЭВМ в ходе массовой обработки фильмовой информации

Задача программной регистрации протоколов заключается в переносе журнала обработки на машинные носители и его программное заполнение с тем, чтобы при необходимости можно было бы на экране дисплея познакомиться с текущим состоянием дел, например, на каком этапе находятся результаты анализа данных с той или иной ленты результатов измерений, где хранятся результаты счета по любой из программ системы,

номера лент, названия файлов и т.д. и т.п., и на этой основе подготовить очередное задание ЭВМ.

Для того, чтобы иметь полную и объективную информацию о ходе и результатах обработки, целесообразно фиксировать данные о каждой задаче, результатах ее окончания и ресурсах, которые потребовались для ее работы. Это позволяет сохранять всю историю обработки, встречавшиеся ошибки, узнать частоту их появления, обнаруживать "узкие места" в системе и принимать меры для их устранения.

Журнал обработки - это последовательный текстовый файл, находящийся на устройстве прямого доступа и состоящий из записей постоянной длины, в которых фиксируются в хронологическом порядке данные о запуске и результатах прохождений задач, используемых для их работы файлах и некоторая дополнительная информация, определяемая спецификой отдельных программ.

Такая организация журнала выбрана из соображений:

- простоты;
- удобства просмотра и при необходимости редактирования его на терминале с помощью штатных системных средств;
- возможности передачи протоколов через локальную сеть и накопления их на ПЭВМ для анализа и работы на рабочем месте пользователя.

Журнал обработки состоит из записей четырех типов.

Первая позиция каждой записи журнала определяет ее тип:

- * - запись о работе программы;
- + - запись о дополнительной информации;
- = - запись о файле;
- с - комментарий.

В запись о работе программы заносятся следующие данные:

- имя программы;
- идентификатор ЭВМ;
- идентификатор виртуальной машины;
- дата счета (DDMMYY);
- день недели;
- время начала работы программы (ННММ);

- астрономическое время выполнения задачи (в минутах);
- время виртуального процессора, использованное виртуальной машиной на выполнение задачи (в секундах);
- общее время реального процессора, затраченное на выполнение данной задачи (в секундах);
- причина окончания задачи (код завершения).

Запись дополнительной информации позволяет вносить в журнал некоторые дополнительные данные, специфические для отдельных программ. Например, для программы NELJA это может быть число, начальные и конечные номера массивов событий, которые нужно обрабатывать, для программы геометрической реконструкции - это заказанное процессорное время, значения счетчиков обработанных событий в начале и в конце работы программы и т. д.

Для каждого файла, который используется в ходе работы программы, в журнал включается отдельная запись (запись о файле), которая содержит:

- логический номер файла;
- тип устройства (МЛ или МД);
- статус файла (чтение - R, запись - W, чтение и запись - RW).

Кроме того, в запись о файле включаются следующие данные.

Если файл дисковый, то:

- имя минидиска;
- виртуальный адрес минидиска;
- название файла;
- тип файла;
- мода файла;
- размер блока;
- число блоков;
- дата (DDMMYY) и время (HHMMSS) создания или обновления файла.

Если файл ленточный, то:

- номер ленты;
- виртуальный адрес магнитофона;
- метка магнитной ленты;
- номер файла на ленте;

- размер блока;
- плотность записи на ленте.

Комментарий - это любой текст, который пользователь может включить в журнал обработки программным путем или вручную, используя любой штатный текстовый редактор базового программного обеспечения ЭВМ серии ЕС, например XEDIT^{/5/}.

Записи разных типов (кроме комментариев), связанные с одной задачей, в журнале располагаются в следующей последовательности: запись о работе программы, запись о дополнительной информации, если она есть, а после этого идут записи об использовавшихся файлах.

Реализация программной регистрации протоколов

Подсистема автоматизированного ведения журнала обработки предназначена для работы на ЕС ЭВМ в рамках операционной системы СВМ.

Для вызова и запуска в среде СВМ программ системы математической обработки фильмовой информации используются отдельные EXEC-процедуры, написанные на языке REXX^{/6/}.

Одними из основных функций процедур являются:

- обращение к средствам ведения журнала обработки для регистрации данных о результатах прохождения задач и использовавшихся файлах;
- запись в журнал обработки характерных для данной программы данных, если такие имеются.

Для ведения журнала обработки используются две небольшие EXEC-подпрограммы (RNINDCTV и FLINDCTV), написанные также на языке REXX.

Подпрограмма RNINDCTV предназначена для записи данных о прохождении задачи в журнал обработки. К ней имеется два обращения из процедуры. Одна непосредственно до запуска, а вторая - после выполнения соответствующей программы. При первом обращении в программном СТЭКе запоминаются данные о начале работы программы. При втором на их основе и данных об окончании задачи формируется запись о результатах работы программы, которая заносится в файл журнала обработки.

Подпрограмма FLINDCTV предназначена для регистрации данных о заданном файле в журнале обработки. Поэтому для каждого файла, который используется для работы программы и данные о котором следует включать в журнал обработки, необходимо организовать обращение к этой подпрограмме.

Заключение

Автоматизированное ведение журнала прохождения задач и обрабатываемых файлов устраняет трудоемкую работу ручного журнального учета и обеспечивает пользователей достоверной информацией о состоянии дел с обработкой экспериментальных данных. В журнале обработки отражается полная картина обработки, и это дает возможность проследить всю историю процесса, обнаруживать узкие места в системе и принимать меры для их устранения. Он также дает полные данные о ресурсах вычислительной системы, которые потребовались для обработки данных эксперимента в целом, или той или иной его части.

Иметь объективные данные о ресурсах, которые потребовались для обработки данных тех или иных экспериментов, для работы той или иной компоненты системы очень важно не только для оценки проделанной работы и работы системы, но и при планировании и проведении новых экспериментов, а также организации программного управления процессом обработки.

Рассмотренные средства автоматизированного ведения журнала обработки не зависят от прикладных программ системы и не накладывают на них никаких ограничений. Для их использования не требуется внесения изменений и в сами прикладные программы. В EXEC-процедуры, которые организуют запуск и выполнение программ, подлежащих протоколированию, необходимо включить CALL-операторы обращения к средствам ведения журнала обработки. Поэтому они вполне могут быть использованы и в других системах массовой обработки данных, вызов которых производится с помощью EXEC-процедур.

Литература

1. Newman H. Comput. Phys. Commun. 45(1987) 27-45.
2. Булко И.М. и др. Система виртуальных машин для ЕС ЭВМ. М., Финансы и статистика, 1985.
3. Балгансурен Я. и др. ОИЯИ, P10-89-40, Дубна, 1989.
4. HYDRA Topical Manual, Book MQ, CERN, Geneva, 1981.
5. VM/SP: System Product Editor User's Guide, SC24-5220.
6. VM/SP: System Product Interpreter Reference, SC24-5239.

Рукопись поступила в издательский отдел
8 февраля 1991 года.