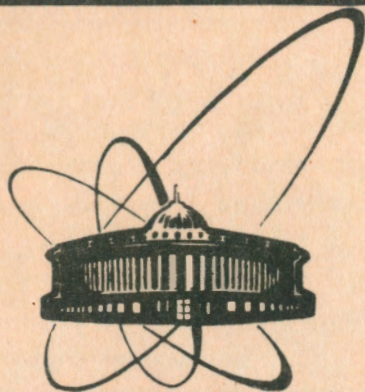


92-5



**СООБЩЕНИЯ  
ОБЪЕДИНЕННОГО  
ИНСТИТУТА  
ЯДЕРНЫХ  
ИССЛЕДОВАНИЙ  
ДУБНА**

P10-92-5

**В.Г.Иванов, Т.Эрдэнэдэлгэр**

**ПРИМЕНЕНИЕ БАЗЫ ДАННЫХ  
ДЛЯ УПРАВЛЕНИЯ ПРОЦЕССОМ  
МАССОВОЙ ОБРАБОТКИ  
ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ**

**1992**

Применение базы данных для управления процессом  
массовой обработки экспериментальных данных

Рассматривается система для организации и использования базы данных о прохождении задач и движении данных в ходе массовой обработки экспериментальных данных. Для функционирования системы требуется персональный компьютер, совместимый с IBM PC/XT/AT.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1992

---

Перевод авторов

---

Ivanov V.G., Erdenedelger T.

P10-92-5

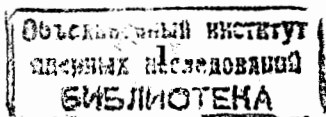
The Usage of the Data Base for the Handling  
by Experimental Data Mass Processing

A system for organizing and using the data base concerning the origin of the problem and data running during experimental data mass processing is considered. The personal computer compatible with IBM PC/XT/AT is required for operating the system.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

Обработка на ЭВМ экспериментальных результатов, получаемых в ходе проведения современных физических экспериментов, является сложным, трудоемким и длительным процессом, связанным с использованием десятков тысяч машинных носителей информации ( магнитные ленты, картриджи ) и других устройств внешней памяти<sup>/1/</sup>. В ходе массовой обработки в вычислительной среде происходит постоянное движение данных, что требует строжайшего соблюдения технологической дисциплины и автоматической регистрации процесса движения данных от одного носителя информации к другому. Для решения последней задачи нами была разработана методика программного сопровождения процесса математической обработки пленочной информации на ЭВМ серии ЕС. Ее суть заключается в том, что на дисках ЭВМ, на которой ведется обсчет данных, в специальном файле ( "журнал обработки" ) автоматически в хронологической последовательности фиксируются результаты прохождения каждой задачи системы. В них входят данные о файлах с исходными данными и результатами счета, потребовавшиеся ресурсы ( время, размеры дисковой памяти ) и код завершения задачи ( нормальное или аварийное окончание ). Организация, структура и содержание журнала обработки были рассмотрены ранее<sup>/2/</sup>.

Во многих случаях, просмотрев содержимое журнала обработки, можно получить из него интересующую пользователя информацию, т.к. этот журнал является обычным текстовым файлом, данные в котором представлены в достаточно наглядной форме. Однако, несмотря на простоту представления данных, особенно при больших размерах журнала, для оперативного получения нужной информации целесообразно сформировать журнал обработки в форме базы данных ( БД ), снабженной простыми и эффективными средствами доступа к содержащейся в ней информации. Для этого на персональном компьютере типа IBM PC



была создана специализированная база данных и соответствующие программные средства, позволяющие пользователям решать указанные выше задачи на своем рабочем месте, связанном через локальные сети передачи данных с ЭВМ ЦВК. Для управления БД нами была выбрана **DBASEIII PLUS<sup>3/</sup>**, которая является одной из самых распространенных реляционных СУБД. Организация процесса массовой обработки экспериментальных данных с использованием такой БД производится следующим образом.

На дисках большой машины ( в данном случае это ЕС ЭВМ ), на которой ведется обсчет данных, фиксируются результаты прохождения каждой из задач. Эти данные затем передаются на ПЭВМ, где, собственно, создается и обновляется соответствующая база данных. С помощью специально разработанного дружественного интерфейса пользователь имеет возможность получать интересующую его информацию о текущем состоянии дел и на ее основе принимать решения.

Данное сообщение посвящено рассмотрению вопросов организации и применения такого типа БД для управления процессами массовой обработки данных.

### Логическая структура БД

Для создания базы данных прежде всего нужно разработать ее логическую структуру. Эта задача сводится к распределению информации на несколько файлов, однородных по структуре своих записей и организации связей между ними.

Логическая структура БД журнала обработки показана на рис.1.

Файл данных о результатах запусков программ содержит информацию о всех запусках программ. Каждая его запись относится к одному запуску программы и содержит ее имя, идентификаторы ЭВМ и виртуальной машины, день недели, дату и время начала работы программы, астрономическое и процессорное время выполнения задачи, а также код ее завершения.

Каждому использовавшемуся при работе программы файлу соответствует отдельная запись в метафайле 1. Она содержит тип файла, его статус, тип устройства и указатель на соответствующую запись в метафайле 1 или 2 в зависимости от типа устройства.

Записи об использованных дисковых файлах содержат имя минидиска, где он находится, идентификатор, тип и объем файла, дату и время его создания или последнего обновления, а также число событий.

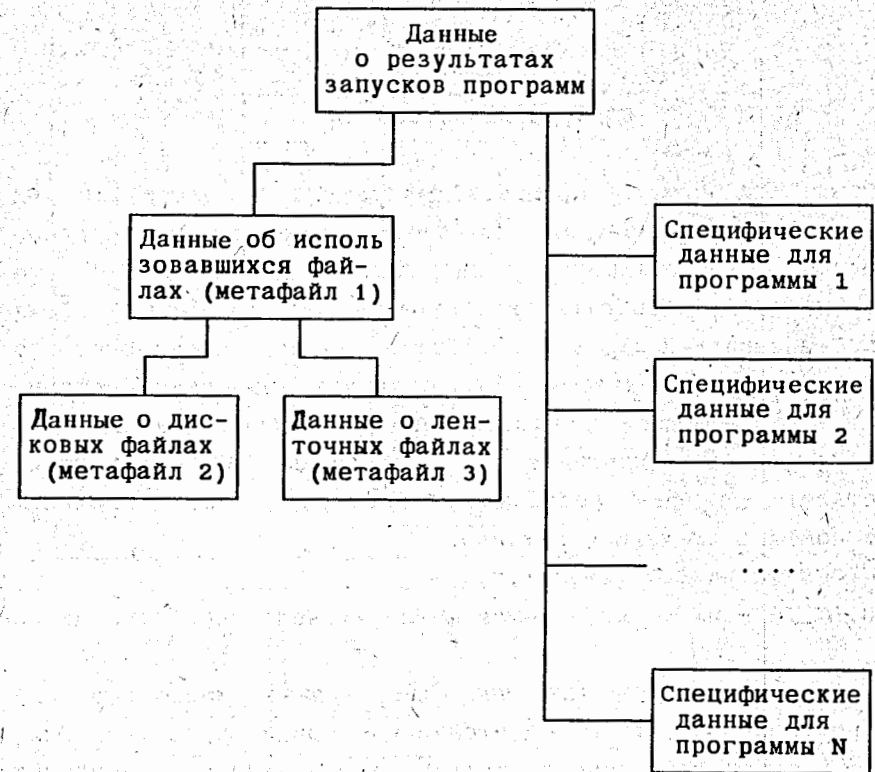


Рис.1

Записи об использованных ленточных файлах содержат номер и метку ленты, номер файла на ней, размер блока, плотность записи на ленте и число событий.

Кроме того, программам задаются разнообразные дополнительные данные (значения параметров, режим работы и т.д.). Поскольку по своей природе они для разных программ различны, для каждой программы, которой требуется дополнительная специфическая информация, организован специальный файл-БД.

Для связки файлов используются общие поля, содержащие номера запуска программы.

### Создание и использование БД журнала обработки

Общая организация процесса создания и использования БД журнала обработки показана на рис.2. Как уже отмечалось, журнал обработки создается и хранится на диске ЕС ЭВМ, где ведется массовая обработка экспериментальных данных. Одной из важных особенностей журнала обработки является постоянное его обновление во время каждого нового запуска программ системы. Поэтому для создания и постоянного обновления БД на ПЭВМ нужно передавать содержимое этого файла на ПЭВМ. Это можно осуществлять разными способами, в частности, через локальную сеть с помощью следующей команды:

```
LOGSEND LN [LT [LM]] [,N] ,
```

где LN, LT, LM - имя, тип и режим файла журнала обработки на ЕС ЭВМ;

N - определяет, как передать журнал полностью или частично. Если  $N \leq 0$ , то передается весь журнал обработки. Если  $N > 0$ , то передаются данные только о последних N запусках программ. По умолчанию передаются данные только о самом последнем запуске. Следует отметить, что когда БД создана, то для ее обновления нет необходимости каждый раз передавать

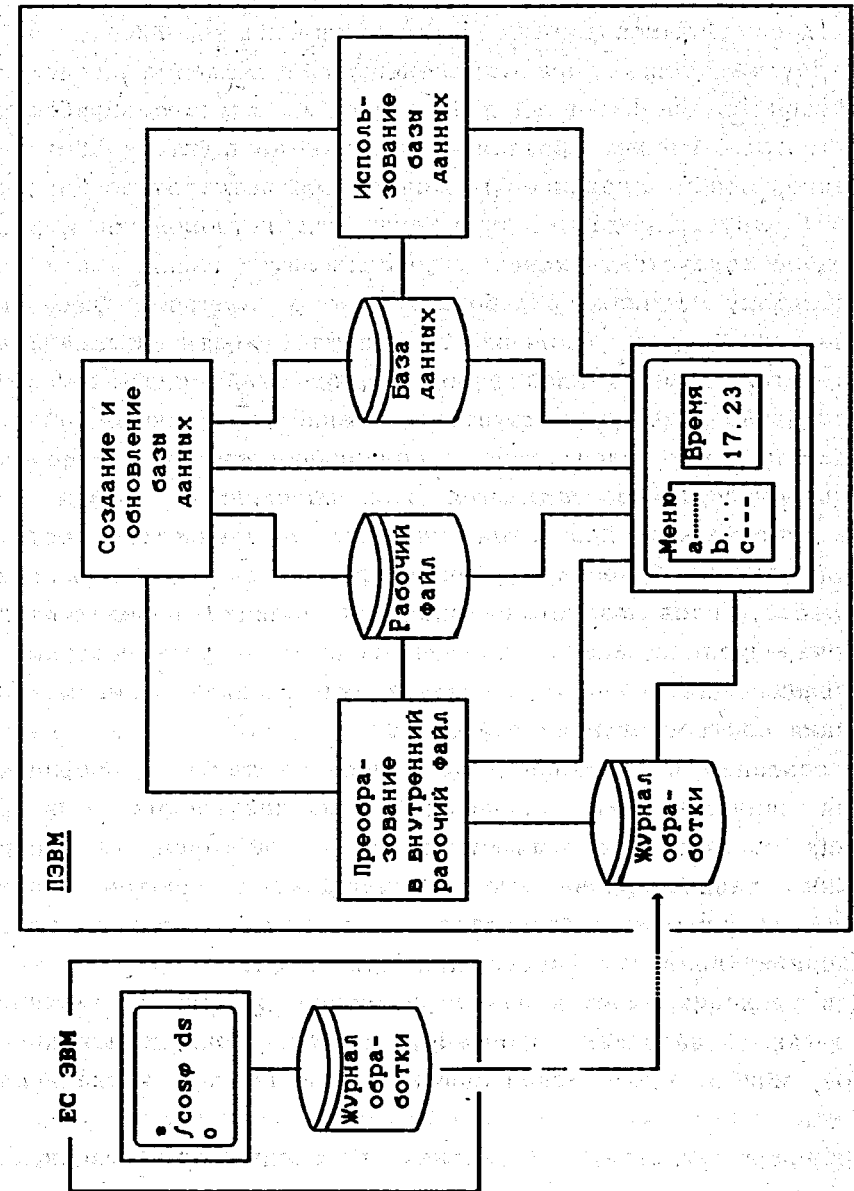


Рис. 2

весь журнал обработки, а нужно передавать только ту его часть, где содержатся данные, еще не включенные в БД.

С другой стороны, при обновлении БД делается проверка того, чтобы каждый раз в БД добавлялись из журнала обработки только новые данные. Поэтому здесь не ставится жесткое требование, чтобы значение N точно соответствовало числу последних запусков, данные о которых еще не включены в БД. Оно лишь не должно быть меньше этого числа.

Поскольку журнал обработки - это внешний файл в отношении СУБД, для упрощения дальнейшей работы он сначала преобразуется во временный рабочий файл СУБД. Одной строке журнала обработки соответствует одна запись временного файла, состоящая из двух полей: типа и содержания строки. Далее на основе рабочего файла создаются или обновляются файлы БД журнала обработки. При этом сначала определяется место, откуда начинать обработку рабочего файла, затем, начиная с этого места, последовательно читаются записи, определяется тип очередной записи и вызывается соответствующая подпрограмма для анализа содержания данной записи и обновления соответствующих файлов БД.

К созданию и обновлению БД относятся также и операции удаления данных из БД. Логической единицей данных в этих операциях является полная информация об одном запуске программы, распределенная по разным файлам. Поэтому при удалении запуска вычеркивается вся относящаяся к нему информация из всех соответствующих файлов БД.

При удалении имеются разные возможности: удалить данные об отдельных запусках программ, группы последовательных запусков, или их любую комбинацию, а также полная чистка всех файлов БД.

Операция удаления - сложная и очень ответственная. Поэтому данные удаляются не сразу, пока пользователь работает в подменю "удаления данных", а в ведущем файле метятся те запуски, которые нужно удалить. Только после выхода из

этого подменю и после повторного подтверждения пользователем они окончательно удаляются из БД. Если пользователь не подтвердит необходимость удаления, то эти данные остаются, а маркировки снимаются.

Таким образом, создаваемая и постоянно обновляемая БД может использоваться для оперативного получения любой информации, связанной с процессом обработки данных. В связи с этим для выполнения следующих наиболее часто осуществляемых и сложных операций разработаны соответствующие программы на командном языке СУБД.

#### 1. Просмотр информации о запусках программ

- Просмотр краткой информации, где выдаются имя программы, номер, дата и время запуска.

- Просмотр информации о последнем запуске выбранной программы. Здесь выдается имя программы, номер, дата и время последнего ее запуска в рассматриваемом интервале времени, названия реальной и виртуальной машин, на которых производился счет, код завершения, информация о входных и выходных файлах, а также специфические дополнительные данные, связанные с данным запуском программы, если последние имеются.

- Просмотр информации о запусках программ в заданном интервале времени. Здесь последовательно выдаются данные о всех запусках выбранной одной или всех программ в пределах рассматриваемого периода времени. Выдаваемые данные такие же, как в случае просмотра информации о последнем запуске программы.

- Просмотр информации о самом последнем запуске программы. Здесь выдается информация о самой последней задаче, зарегистрированной в журнале обработки. На основе этой информации и других данных можно принимать решение о продолжении обработки.

#### 2. Получение суммарной статистики

Здесь в пределах рассматриваемого периода времени для

каждой программы системы в отдельности выдаются число запусков, число обработанных событий, а также процессорное и астрономическое время, ресурсы дисковой памяти, которые потребовались для их работы. Кроме того выдаются их суммарные итоги.

3. Получение данных об использовавшихся магнитных лентах - 0 всех магнитных лентах, содержащих результаты измерений, которые поступали на вход системы математической обработки данных. Отсюда можно узнать, какие ленты, когда обработаны, какие массивы с них считаны, полностью ли обработана та или иная магнитная лента исходных данных.

- 0 всех магнитных лентах, на которые записывались промежуточные и окончательные результаты обработки. Из них можно узнать, какие файлы, когда, на какие ленты копировались.

#### 4. История обработки данных

Это дает возможность проследить всю историю обработки конкретной магнитной ленты с данными или заданного дискового файла. Т.е. можно узнать, какие данные прошли все этапы анализа, а какие находятся на промежуточных.

5. На основании БД также можно получать разнообразные данные. Например, с помощью специальной программы можно получить распределения числа запусков задач в задаваемом интервале времени по дням недели, по часам суток и по разным ЭВМ.

### Пользовательский интерфейс

Для обеспечения дружественного интерфейса с пользователем совместно используются многоуровневое меню, экранные бланки, разноцветные окна и словесные сообщения. Главное меню для работы с БД журнала обработки показано на рис.3. После выбора одного из предложенных вариантов на дисплей выдается соответствующее подменю, выбор из которого, в свою очередь, выдает следующее подменю и т.д.

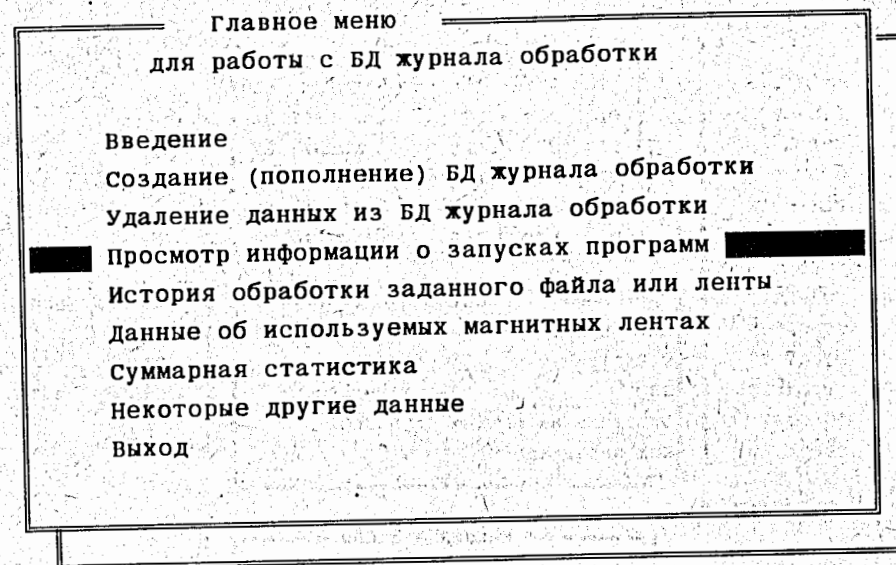


Рис.3

Во многих случаях система выдает на экран бланки в естественном для человека виде, с поясняющими текстами и разноцветными полями для их заполнения пользователем. Это дает возможность пользователю одновременно задавать значения разных параметров, ограничений и т. д. путем заполнения готовой формы. На этих бланках всегда высвечиваются значения по умолчанию. Это, с одной стороны, показывает образец ее заполнения, а с другой - позволяет заполнять только те места на бланке, где нужно менять заданные по умолчанию значения.

Следует отметить, что использование разноцветного многооконного метода представления информации для выдачи на экран меню, бланков и других выдаваемых данных в ответ на запросы дает дополнительную ясность и большие удобства. В качестве примера на рис.4 показан вид экрана после запроса выдачи данных о последнем запуске программы COLL DST в июле 1991 года.

ПРОГРАММЫ	
NEIJA	
ZINPUT	
CHECK	
GEOMOP	
CLPHOP	
PASPORT	
<b>COLLDST</b>	
DSTWRTOP	
SCPSPTOP	
DTFQXCOP	
TDFQXCOP	
FQXSORT	
ВЫХОД	

Входные файлы	
DISK PEVC PAS65EXP CATALOG B1	
DISK SSF CLP1 CLPHRSLT B1	
DISK SSF CLP2 CLPHRSLT B1	
DISK SSF CLP3 CLPHRSLT B1	
DISK SSF CLP4 CLPHRSLT B1	
DISK SSF CLP5 CLPHRSLT B1	

COLLDST	
Дата и время запуска: 07/03/91 13.36	
Номер запуска : 83	
ЭВМ : J1NR66A	
Виртуальная машина : H803002	
Код завершения : 0	

Доп. данные	
EVENT 2316	
EXACT 0.7	

Выходные файлы	
DISK PRN COLL DST LISTING B1	
DISK DSF COLL DST DTFQX B1	

Интервал	
Дата	
Начальная 07/01/91	
Конечная 07/31/91	
Продолжить - (Y/N)? Y	

Рис. 4

### Заключение

В рамках рассматриваемого подхода пользователь имеет возможность оперативно получить практически любую интересующую его информацию о самом процессе массовой обработки данных, текущем его состоянии, уже использованных ресурсах и т.д. и на ее основе принимать решение, в частности планировать дальнейшие действия, подготовить очередное задание ЭВМ и т.п.

Весьма полезным представляется возможность проследить историю обработки массивов данных, так как она показывает весь путь получения тех или иных результатов.

Система была реализована на персональном компьютере типа

IBM PC. Испытания системы проводились в ходе массовой обработки данных <sup>16</sup>Ор-эксперимента<sup>4/</sup>, и ее можно без особых усилий адаптировать на другие задачи.

### Литература

1. Newman H. Comput. Phys. Commun. 45 (1987) 27-45.
2. Глаголев В.В. и др. ОИЯИ, P10-91-80, Дубна, 1991.
3. Каррабис Д.Д. Программирование в dBASEIII Plus. Пер. с англ. -М.: Финансы и статистика, 1991.
4. Балгансурэн Я. и др. ОИЯИ, P10-89-40, Дубна, 1989.

Рукопись поступила в издательский отдел  
6 января 1992 года.