Московский физико-технический институт
(государственный университет)

Факультет общей и прикладной физики

Кафедра физики высоких энергий

Particle Physics Data System/GNU/Linux/x86

Выпускная квалификационная работа на степень бакалавра

Руководитель:
О.В. Зенин

Выполнил:
Студент IV курса Д.С. Корниенко

Москва 2007

Содержание

Обозначения и сокращения

ФВЭ – физика высоких энергий.
БД – база данных.
СУБД – система управления базой данных.
ОС – операционная система.
PPDS – Particle Physics Data System: система БД по физике высоких энергий, поддерживаемая в ГНЦ ИФВЭ. Содержит архивные базы данных DOCuments (DOC), ReactionData (RD), EXperiments (EX), DiscoverY (DY), компиляции оцененных данных ParticleProperties (PP,свойства частиц) и CrossSections (CS, интегральные сечения реакций) и дополнительную базу VocabularY (VY, контролируемый словарь языка описания данных, используемого в базах данных PPDS).
RD (ReactionData) – БД в составе PPDS; содержит табличные экспериментальные данные, извлеченные из экспериментальных публикаций: дифференциальные и полные сечения реакций, структурные функции, поляризационные наблюдаемые и другие измеряемые величины физики частиц.
CS (CrossSections) – БД в составе PPDS; содержит формально оцененные данные по интегральным сечениям и средним множественностям адронов в различных столкновениях.
DG (DataGuide) – БД в составе PPDS; содержит описательную информацию о результатах эксперимента, извлеченную из экспериментальных публикаций (но не сами экспериментальные данные)
VY (VocabularY) – БД в составе PPDS; содержит определения терминов, имена и синонимы частиц, ускорителей, детекторов и многие другие лексические единицы языка описания данных физики частиц – PPDL, используемого во всех базах данных системы PPDS.
PP (ParticleProperties) – БД в составе PPDS; содержит оцененные данные по свойствам частиц из обзора физики частиц RPP.
BDMS – Berkeley Database Management System, Берклеевская СУБД.
BDMS/4.0 – версия BDMS для платформ Alpha/OpenVMS, VAX/OpenVMS, x86/GNU/Linux (α-версия), основа PPDS.
PPDL – Particle Physics Data Language, язык описания данных ФВЭ, используемый в PPDS.
Запись – минимальный набор данных в BDMS, которым производится обмен между СУБД и дисковым(и) файлом(ами) БД.
Документ – внешнее представление *записи*.
ЭД – элемент данных: наименьшая логическая единица данных в BDMS.
FDT – File Definition Table – набор данных, определяющий структуру *записи*, характерную для данной БД. Необходим для инициализации БД и последующей загрузки в нее данных.
КОМПАС – программа КОМПиляции, Анализа, Систематизации данных и моделей физики высоких энергий, развиваемая в ГНЦ ИФВЭ.

Аннотация

Данная работа посвящена модификации системы баз данных по физике частиц (Particle Physics Data System, далее PPDS [1]), разработанной на базе СУБД BDMS и эксплуатируемой в настоящее время в ИФВЭ. Конечная цель модификации – перенос PPDS с устаревшей и дорогостоящей платформы DEC Alpha/OpenVMS на массовую платформу x86/GNU/Linux. Успех представленной работы – адаптация СУБД BDMS для POSIX-систем – позволяет:

1. Продлить эксплуатацию PPDS на практически неограниченный срок.
2. Создать развитые пользовательские интерфейсы к PPDS (WWW, консоль, и т.д.), пользуясь современным инструментарием ОС GNU/Linux.
3. Использовать BDMS как встраиваемую БД в физических приложениях.

Полученные наработки могут быть использованы при миграции унаследованных программ с эксплуатируемой в ИФВЭ платформы DEC Alpha/OpenVMS на платформу x86/GNU/Linux.

Введение

Физика высоких энергий – это обширный комплекс проблем и задач, одной из которых является систематическое сравнение совокупности мировых экспериментальных данных с теоретическими утверждениями и феноменологическими моделями для выяснения текущего уровня согласия теоретического и эмпирического знания.

Решение данной задачи невозможно без создания “банка эмпирического знания”, одной из реализаций которого является разработанная и эксплуатируемая в ИФВЭ система баз данных по физике частиц – Particle Physics Data System (PPDS).

Система баз данных PPDS

Система PPDS содержит архивные базы данных DOCuments, ReactionData, EXperiments, DiscoverY, компиляции оцененных данных ParticleProperties (свойства частиц) и CrossSections (интегральные сечения реакций) и дополнительную базу VocabularY (контролируемый словарь языка описания данных, используемого в базах данных PPDS).

PPDS в настоящее время имеет $> 10^4$ уникальных по IP пользователей, причем их число увеличивается на $\sim 2 \times 10^3$ ежегодно. Число содержательных запросов к PPDS через WWW-интерфейс `http://wwwppds.ihep.su:8001/` $\gtrsim 1500$ в месяц.

Базы данных PPDS накапливаются и сопровождаются в ИФВЭ в сотрудничестве с группами систематизации данных из LBL(США), Durham University(Великобритания), ИТЭФ, ОИЯИ, и в сотрудничестве с библиотекой SLAC.

PPDS включает следующие БД:

DG (DataGuide) – содержит информацию, извлеченную из экспериментальных публикаций (но не сами экспериментальные данные).

Поиск данных возможен по номеру эксперимента, авторам публикации, ускорителю, детектору, реакции, частице пучка, импульсу пучка, частице мишени, частице конечного состояния реакции, наблюдаемой величине, более ранним публикациям эксперимента и другим атрибутам эксперимента и публикации.

Интервал поиска: 1895 $\rightarrow$ современность.

RD (ReactionData) – содержит табличные экспериментальные данные, извлеченные из экспериментальных публикаций (дифференциальные и полные сечения реакций, структурные функции, поляризационные наблюдаемые и другие измеряемые величины физики частиц).

Поиск данных возможен по номеру эксперимента, первому автору публикации, детектору, реакции, частице пучка, импульсу пучка, частице мишени, частице конечного состояния реакции, наблюдаемой величине и другим атрибутам эксперимента и публикации.

Интервал поиска: 1952 $\rightarrow$ современность.

CS (CrossSections) – содержит формально оцененные данные по интегральным сечениям и средним множественностям адронов в различных столкновениях. Включает данные из компиляций CERN—HERA и из старых компиляций UCLRL и LBL. Все данные проверены повторно в сравнении с оригинальными публикациями. Предварительные данные помечены. Модельно зависимые данные помечены. Записи дополнены данными о систематических погрешностях из оригинальных публикаций.

Поиск данных возможен по первому автору публикации, реакции, частице пучка, импульсу пучка, частице мишени, частице конечного состояния реакции, наблюдаемой величине и другим атрибутам эксперимента и публикации. Пополняется регулярно данными из архивной базы RD после повторной сверки с оригинальными публикациями.

Интервал поиска: 1950 $\rightarrow$ современность.

PP (ParticleProperties) – содержит оцененные данные по свойствам частиц из Review of Particle Properties Summary Tables группы PDG.

Поиск данных возможен по имени частицы, квантовым числам, модам распада, частице конечного состояния в распаде и по другим атрибутам частицы.

На данный момент не актуальна, последняя синхронизация с RPP в 1993 г.

VY (VocabularY) – содержит определения терминов, имена и синонимы частиц, ускорителей, детекторов и многие другие лексические единицы формального языка описания данных физики частиц – PPDL, используемого во всех базах данных PPDS.

EX (EXperiments) – содержит данные, описывающие одобренные предложения экспериментов большинства исследовательских лабораторий по физике частиц.

Поиск данных возможен по номеру эксперимента, авторам предложения, ускорителю, детектору, реакции, частице пучка, импульсу пучка, частице мишени, частице конечного состояния реакции, по библиографии публикаций эксперимента и другим атрибутам эксперимента.

Интервал поиска: 1961 $\rightarrow$ современность.

DY (DYscovery) – содержит описательную информацию о ключевых научных публикациях в ФВЭ с 1895 г. по настоящее время. В записях DY приводятся аннотации (выдержки) и формулировки “описательных формул” открытий. Используется также для подготовки публикации аннотированной Хронологии физики частиц. Первое издание Хронологии опубликовано как совместное издание ИФВЭ и LBNL (США) к столетию физики частиц в 1996 году [?].

Наборы данных по наблюдаемым ФВЭ накапливаются и сопровождаются в системе PPDS ИФВЭ в рамках программы КОМПАС [1]. В качестве источников данных используются открытые публикации в реферируемых научных журналах. Сопровождение данных ведется по единой методике, включающей повторяемый с публикацией новых экспериментальных данных цикл:

1. Контроль полноты данных по тематикам.

2. Перенос табличных данных из публикаций в БД.

3. Выделение предварительных данных.

4. Выделение модельно-зависимых данных.

5. Проверка соответствия содержимого БД оригинальным публикациям.

По некоторым тематикам (полные адронные сечения в столкновениях *адрон адрон*, γ *адрон*, e^+e^-) проводится периодическая проверка соответствия накопленных в PPDS данных современным теоретическим моделям.

СУБД BDMS

Поисковая часть системы PPDS реализована в ИФВЭ на ЭВМ DEC Alpha, работающих под управлением ОС VMS. Базы данных используют иерархическую СУБД BDMS/4 – Berkeley Database Management System, первоначально разработанную в рамках Particle Data Group (LBNL, США) [2] и существенно развитую в ИФВЭ [3]. В BDMS/4 значительно развиты диалоговые возможности BDMS, система сравнима по возможностям с наиболее распространенными в настоящее время СУБД.

Развитие оригинальной BDMS/2.2 [2] до BDMS/4 проводилось без перестройки архитектуры системы. В набор инструкций BDMS/4 включены все инструкции предыдущих версий. BDMS/4 поддерживает фиксированные в версии BDMS/2.2 схемы и форматы обмена данными между разными “экземплярами” PPDS.

Из других эксплуатировавшихся вплоть до недавнего времени экземпляров системы нам известна версия BDMS, адаптированная группой HEPDATA (Даремский ун-т, Великобритания) для ЭВМ Sun [7].

СУБД BDMS/4 – программа на языке ФОРТРАН 77 (существенно используются расширения, допускаемые компилятором DEC FORTRAN, и подпрограммы, специфические для библиотеки времени исполнения DEC FORTRAN).

СУБД BDMS/4 можно разделить на две части:

1. *ЯДРО*: набор процедур для поиска в базе данных, обмена записями с дисковой памятью, преобразований между внутренним и внешним представлениями записей, создания и поддержания инвертированных списков;

2. прикладные *ПРОЦЕССОРЫ* управления вводом/выводом и преобразованием данных.

Процессоры реализованы на языке ФОРТРАН, и совместно с ядром системы образуют *ЗАГРУЗОЧНЫЙ МОДУЛЬ* (исполняемый образ – executable image, по терминологии ОС VMS). Каждому прикладному процессору соответствует один загрузочный модуль. Примеры загрузочных модулей (имена соответствуют эксплуатируемой на `MX.IHEP.SU` системе):

RD.EXE – включает прикладные процессоры управления описанием измерений наблюдаемых ФВЭ;

CS.EXE – включает прикладные процессоры управления описанием измерений полных сечений;

VOC.EXE – включает прикладной процессор, управляющий единым словарем обозначений PPDL, использующимся во всех БД PPDS;

BDMS.EXE – загрузочный модуль, не имеющий прикладных процессоров [1]

Конструирование БД, ориентированной на определенную структуру данных, включает в себя написание прикладных процессоров и создание специализированного загрузочного модуля.

Единицей хранения в БД BDMS является *запись*, имеющая иерархическую структуру *элементов данных* (ЭД), по инвертированному списку которых осуществляется поиск в БД. Каждая БД в системе PPDS имеет специфическую структуру ЭД (см. ниже). Запись – как правило, логически обособляемая из всего информационного массива часть. Например, для БД CS, содержащей данные по интегральным сечениям реакций, запись содержит: набор сечений, измеренных в данном эксперименте для данной реакции по единой методике, вообще говоря, при нескольких значениях импульсов частиц пучка и мишени; данные о статистических и систематических ошибках измерений; идентификаторы ускорителя и детектора; библиографические данные публикации; комментарии специалиста, подготовившего данные к занесению в БД.

Как было отмечено выше, наименьшей логической единицей базы данных BDMS является *элемент данных* (ЭД). Ядром BDMS/4 структура самого ЭД не рассматривается. При необходимости такую задачу решают *прикладные процессоры.*

Основные функции ЭД в базе данных:

1. разделение данных в записи на минимально значимые элементы, классификация данных;

2. объединение данных из различных записей в общей для них базе данных по общему признаку;

3. введение иерархического описания данных в документе (записи), структурирование данных.

[1] Именно он перенесен в первую очередь на платформу x86/GNU/Linux.

Конструируя специфическую БД, администратор:

- определяет понятие “запись” в данной БД;
- вводит описание записи как упорядоченного набора логически выделяемых единиц информации – элементов данных;
- фиксирует структуру записи, введя значения *атрибутов* [4] всех ее элементов данных.

Один и тот же ЭД может быть неоднократно представлен своими *генерациями* в данной записи. Например, если содержимым ЭД является имя одного из авторов публикации, а запись – библиографическое описание публикации, то ЭД "автор" в записи будет представлен столько раз, сколько соавторов содержит публикация.

Процедуры создания, заполнения, редактирования БД BDMS, а также поиска и выдачи данных из БД подробно рассмотрены в руководстве [3].

БД физически размещена в файлах прямого доступа на дисковом носителе. Набор файлов БД вместе с файлами загрузочных модулей системы, вспомогательными файлами управления и конфигурации, файлами описаний и самоописаний образует *систему файлов* системы BDMS/4, отображенную в *дистрибутив* PPDS.

Необходимость миграции PPDS на POSIX-платформу

В настоящее время основным препятствием для продолжительной эксплуатации PPDS является ее зависимость от аппаратно-программной платформы DEC Alpha/OpenVMS. Учитывая бо́льшую эффективность PPDS по сравнению с широко известной связкой БД SPIRES/HEP [8] + HEPDATA [9], было принято решение о продолжении эксплуатации PPDS в ИФВЭ, с последующей разработкой современного WWW-интерфейса. Дальнейшая разработка PPDS и интерфейсов к ней на базе DEC Alpha/OpenVMS нецелесообразна в силу прекращения в ближайшие несколько лет поддержки данной платформы правопреемником фирмы DEC – фирмой Hewlett Packard Co. (США). Разумным решением является перенос PPDS на массовую POSIX-совместимую платформу GNU/Linux/x86.

Миграция на GNU/Linux значительно упрощает дальнейшее развитие БД, создание современного пользовательского WWW-интерфейса к БД, включающего, в перспективе, интерактивные средства занесения данных силами пользователей, не являющихся специалистами по BDMS.

В свете вышеизложенного, тема данной работы весьма **актуальна**.

Перенос BDMS на платформу GNU/Linux/x86

Первоочередными задачами при переносе PPDS на POSIX-платформу являются:

1. Адаптация исходного кода *ядра* BDMS к свободно распространяемому компилятору языка Фортран 77 GNU g77 [16] и среде выполнения ОС.
   Это порождает несколько подзадач:
   (a) Перевод исходного текста программы с DEC Fortran на стандартный Фортран 77[10].
   (b) Приведение используемых системных вызовов и вызовов библиотеки времени выполнения в соответствие с POSIX; эмуляция вызовов, не имеющих аналогов в стандарте POSIX.
   (c) Дополнительная проверка совместимости машинно-зависимой части программы с архитектурой x86: порядок байт в машинном слове, разрядность слова.
2. Отладка перенесенного ядра на целевой платформе и запуск минимального загрузочного модуля BDMS.
3. Прохождение минимальным загрузочным модулем BDMS набора тестов, подтверждающих его правильное функционирование:
   (a) загрузка File Definition Table – таблицы, определяющей структуру записи тестовой БД
   (b) создание пустой БД
   (c) загрузка нескольких записей в созданную БД
   (d) проверка поиска по тестовой БД
   (e) проверка выдачи записей (dump) из тестовой БД
   (f) проверка возможностей удаления, добавления, редактирования записей в тестовой БД.

В процессе выполнения данной работы все вышеперечисленные этапы были успешно пройдены. Реализованный на данный момент на DEC Alpha/OpenVMS набор функций минимального загрузочного модуля BDMS воспроизведен на POSIX-совместимой платформе GNU/Linux/x86.

Вышеперечисленные этапы описаны более подробно в последующих подразделах.

Адаптация исходного кода BDMS к GNU g77

Поскольку значительное число изменений однотипны, так как порождены несоответствием формата/лексики и не требуют дополнительной проверки – возможна автоматизация. Для этого была написана программа, "разрезающая" монолитный исходный файл минимального загрузочного модуля BDMS на подпрограммы/функции и исправляющая несоответствия форматов DEC Fortran и Фортран 77 (см. Таб. 1)

Таблица 1: Список исправлений

1. Замена `INCLUDE`
2. Добавление запятых в операторы `FORMAT`
3. Исправление неверного `COMMON /DBMACH/`
   и переименование `COMMON /ICOMPCOM/`
4. Исправление старых функций преобразования типов:
   `DFLOTJ` $\rightarrow$ `DBLE`; `IIDINT` $\rightarrow$ `INT`;
   Приведение `COMMON /DEF_DAT/` к единому размеру.
5. Исправление вызовов функций/подпрограмм с текстовыми параметрами, текстовые параметры перенесены в отдельные переменные.
6. Исправление аргументов-меток для вызова функций/подпрограмм.
7. Ручные правки.

п/п	Файл	1	2	3	4	5	6	7
1.	block1		1			2	5	8
2.	block2			1				2
4.	interst		3			6	8	8
5.	bdms_spy					14	7	8
6.	execb		1				1	
7.	filon					4	1	9
8.	load						1	
9.	compar			2			1	
10.	reclst	8				23	1	3
11.	ident	8		2			1	7
16.	ccompar			4				3
17.	checkcon		1			9		
18.	precont							3
19.	hollerit	8		3			1	
21.	exec	8		2	1	96	34	
22.	message		1					
23.	outedit	8	3			8	7	15

Таблица 1: продолжение

п/п	Файл	1	2	3	4	5	6	7
24.	login					14	4	3
25.	default_parameters				1	7		
26.	check_par					17		
29.	nrecnset	8			1	6	2	
31.	eval	8				33		1
32.	query	8	1	2		60	3	4
34.	rpnset	8						
35.	rpnrec	8						
36.	rpnde	8		2				
37.	recid	8						
39.	sortbdms	8		2				
40.	shells			2				
41.	or	8						
42.	and	8						
43.	andnot	8						
44.	not	8						
45.	bufin	8						
47.	ifound	8						
48.	purge	8						
51.	sysfnal							6
52.	cont_inp		1				1	
53.	bdms_releas		3				2	
54.	bdms_input		2				4	
55.	bdms_rename						2	
56.	call_rn_sys		3				2	6
57.	file_look		1				4	
58.	from_to						3	
59.	type_file					2	1	
63.	bdms_user_host		1			2	1	2
64.	lfcom_decod						2	4
66.	lf_frtodc		1					
68.	call_chan_sys		1					
69.	call_rl_sys		4				1	
70.	call_in_sys		5				1	9
71.	call_er_sys		1				1	2
72.	call_lf_sys		1				3	
73.	call_ld_sys						1	
74.	sys7_8_monitor		1				5	
75.	www		2					2
76.	sys_dirname						1	

Таблица 1: продолжение

п/п	Файл	1	2	3	4	5	6	7
77.	list_dir					2	1	
79.	file_mask		1					
80.	file_name						1	
81.	list_of_de	8				7		3
82.	unddenm					9	1	1
83.	selist	8	1			5	3	
84.	lvalue			2				
85.	outinv	8						
86.	kmod	8						
109.	part_of_text		1					1
110.	ifind_two							8
111.	comm_monitor			1		20		4
112.	start_macro			1				
114.	if_stop		2				1	
115.	noesccop							4
116.	icompcom			1		26	1	15
117.	find_syn			1				
118.	find_ncom			1				
119.	print_syn			1		5		
120.	set_synonym			1		23	3	1
121.	set_c_object			1				
122.	if_level			1				
123.	del_command			1				
124.	restart_command		1	1				
125.	move_cm			1				
126.	set_macro		1	1				
127.	set_level			1				
128.	find_comm			1				
129.	com_prin		1	1		15	1	3
132.	list	8		2		29	2	12
133.	mainflag					4	7	3
134.	dopen		1				1	
135.	dsopen							8
139.	dclose		1					
140.	lgetar		1					
142.	ddfree		1					
144.	drd		1					
150.	tread					4	1	4
155.	outerr			2				
156.	wrtstop		1				1	3

Таблица 1: продолжение

п/п	Файл	1	2	3	4	5	6	7
157.	inln						3	
160.	cominp							1
161.	monitor					2	15	6
167.	enfloa				5			
169.	decodr				4			
171.	fdtcom						2	6
172.	fdtprint	8	2			41	4	23
174.	directory					5	1	
175.	dirline					2	1	
176.	bittext					2		
179.	dirset							3
180.	delfil							1
183.	help_me		2			2	3	6
184.	helptext		3				4	15
185.	lengfn							3
186.	outname							4
187.	show_proc		1			2	1	8
188.	menu	8	4			6		7
189.	enmenu	8					1	4
190.	title						2	
195.	rea					2	1	1
198.	srtmrg	8	4					
199.	setind	8						
200.	outbuf	8						
206.	merin		1					
207.	merging						2	
208.	icompar	8						
209.	wrrt		1					
213.	tree	8						
214.	page	8						
217.	clean	8				12		
218.	postcard						1	3
219.	user_name							4
220.	def_mail		2			2	5	10
222.	mail		2			4	14	22
223.	delete_mail		1			4	6	6
224.	read_mail						2	
225.	write_mail		1			12	6	24
227.	helpbase						2	
231.	bdms_merg		4			2	4	5

Таблица 1: продолжение

п/п	Файл	1	2	3	4	5	6	7
233.	sort_output	8				6	2	3
234.	sorting_set	8					2	9
235.	to_output	8	1			31	4	3
236.	setsort		3	3				3
237.	wrt33		1		1		1	
243.	iop							4
244.	dalgo		1			15	4	
245.	ralgo					15	4	
246.	ialgo						1	
248.	deguidline	8				10	3	6
249.	autocoder	8				8	11	6
250.	encoment	8					1	3
251.	enmacro	8	1					
252.	readde	8	4			2	9	12
253.	coder	8	1			6	4	
254.	guidance	8				4	1	8
255.	enc_init	8	4			12	6	11
256.	setgdat	8				4		
257.	outguid	8				7	1	
258.	getdename	8						
259.	intercall		1				1	1
260.	guidedit	8	2			2	7	11
261.	encolist						4	
262.	entrans	8	1				2	
263.	calcul							9
264.	edtran					6	5	8
265.	editbyencod	8	1			2	6	3
266.	eedelist	8	1					
267.	eechange		1				2	5
268.	eememo		1			4	1	4
269.	eeacting	8					3	10
270.	eecheck	8					1	
271.	deprot	8						3
272.	setencodd						2	1
273.	enconst	8				2		26
274.	part_text		1					
275.	toiproc	8						
276.	textcheck						1	
277.	editwords					2		1
278.	wordlist					23		1

Таблица 1: продолжение

п/п	Файл	1	2	3	4	5	6	7
279.	findwrd						2	
280.	initspell		1			8	4	3
282.	slowo						2	6
284.	workwcheck						1	7
285.	inde_if	8						
287.	if_sequence	8	1				1	
288.	invindcom	8		3		31	4	
291.	invertind	8	3		1		5	
292.	index_newset				1			
293.	invwrt			3			1	4
294.	inde_edit	8	1		1	4	1	
295.	i_compar							3
298.	saveset		3			5		
300.	edit	8		2		84	4	9
301.	insert	8						
302.	get	8		2				
306.	lset	8						
307.	delete	8						
308.	chain	8						
309.	repl	8						
310.	ifirst	8						
311.	in	8						
312.	define	8		4		17	6	
313.	fdt	8		2		42		
315.	loader	8	1			22	1	
316.	upddir	8						
317.	balance	8				19		
318.	dump	8				11	1	
319.	loads	8						
320.	outkey	8						
321.	init	8		2				1
322.	bopen	8						
323.	closer	8						
325.	fetch	8				24		
326.	store	8						
327.	bremove	8						
328.	create	8						
329.	packr	8						
330.	unpckr	8						
331.	empty	8						

Таблица 1: продолжение

п/п	Файл	1	2	3	4	5	6	7
332.	index7	8		2				
334.	addkey	8						
335.	branch	8						
336.	delkey	8						
337.	matchr	8						
338.	matchk	8						
339.	fopn		3				1	4
340.	lptext					2	1	1
355.	wro							16
356.	rdo							14
357.	wroa							14
358.	uread		1					
361.	datime					12		

Из несоответствий стандарту Фортран 77, часто встречающихся в исходном коде BDMS, следует также упомянуть

1. несоответствие синтаксиса при передаче меток в подпрограмму/ф-цию;
2. различие синтаксиса объявления параметра (`PARAMETER`);
3. способ разделения спецификаторов в операторе `FORMAT`;
4. способ инициализации переменных, задаваемых массивом (`DIMENSION`) и строкой (`CHARACTER*(*)`);
5. объявление переменных в восьмеричной форме, не поддерживаемое стандартом Фортран 77.

Также исправлены различия в синтаксисе имен файлов в ОС OpenVMS и POSIX-системах.

Несколько примеров замен при преобразовании исходного кода приведены в Таб. 2.

Ручные изменения коснулись в основном процедур открытия и чтения из файлов, а также взаимодействия со средой выполнения программы:

- полностью переписана подпрограмма `SHOW_SYS`, получающая параметры среды;
- в подпрограммах `WRO`, `WROA`, `RDO` изменен размер и тип переменных, в которых хранятся пути к файлам БД.
- в подпрограмму `RDO` вставлена проверка на правильность имени файла;
- написаны подпрограммы и функции `VZERO, VBLANK, IULAST, IUCOMP, VFILL, UCOPY`

Таблица 2: Примеры замен при преобразовании исходного кода BDMS

#	Dec Fortran	Фортран 77
1	function1(arg1,arg2,&lb1,&lb2)	function1(arg1,arg2,*lb1,*lb2)
2	parameter MAXLEN=4096	parameter (MAXLEN=4096)
3	format('STOP "'A4'"TIME:')	format('STOP "',A4,'"TIME :')
4	byte b66(8) data B66/'SERVICE$'/	byte b66(8) data B66/'S','E','R','V','I','C','E','$ '/
	character*36 CHARR data CHARR/'COMPASHELP '/	character*36 CHARR data CHARR/'COMPASHELP '/
5	data LARGE/"17 777 777 777/	data LARGE/2147483647/
6	include 'BDMS$nuc:EQUI.'	include 'bdms_nuc.inc'

- Заменены системные вызовы VMS:
  - LIB$ERASE_PAGE → scls(integer function) (внешняя)
  - LIB$SPAWN → ispawn (integer function) (внешняя)
  - RMS$_RENAME → rename (integer function) (внутренняя для g77)
  - RMS$_ERASE → unlink (integer function) (внутренняя для g77)
  - RMS$_GETSIZE → cGETSIZE(subroutine) (внешняя)
  - lib$get_symbol не используется
- В подпрограмме DSOPEN изменен размер записей (параметр RECL) с 512 на 2048, в связи с тем, что в DEC Fortran при открытии файла с параметрами ACCESS='DIRECT', FORM='UNFORMATTED' размер записи выражается в словах (4 байта), в то время как в g77 RECL задается в байтах.
- Изменена процедура получения даты: теперь дата вычисляется функциями DATE_Y2KBUG, VXTIDATE_Y2KBUG.
- Изменена процедура получения времени в формате hh:mm:ss, вместо time для VMS, используется ctime(time).

Отладка минимального загрузочного модуля BDMS

Отладка минимального загрузочного модуля BDMS проводилась в основном с использованием так называемой "тестовой печати", также активно использовался отладчик gdb.

Основным результатом отладки явились некоторые исправления логики, а также типа открытия файлов.

Тестирование загрузочного модуля BDMS

При тестировании загрузочного модуля BDMS было проверена правильность выполнения им следующих процедур:

1. загрузка File Definition Table – таблицы, определяющей структуру записи тестовой БД
2. создание пустой БД
3. загрузка нескольких записей в созданную БД
4. проверка поиска по тестовой БД
5. проверка выдачи записей (dump) из тестовой БД во внешних машино- и человеко-читаемых форматах
6. проверка возможностей удаления, добавления, редактирования записей в тестовой БД.

Все перечисленные этапы пройдены успешно.

План дальнейших работ

Дальнейшее развитие PPDS на GNU/Linux требует перенести прикладные процессоры обработки ЭД для БД Vocabulary, DataGuide, ReactionData, CrossSections, ParticleProperties. По оценке снизу, требуемый объем работ ~ 1 человеко-год. В результате выполнения этого этапа работы возможен полный отказ от эксплуатации PPDS на VMS-кластере ИФВЭ `mx.ihep.su`, и полный переход PPDS на используемую в ИФВЭ платформу GNU/Linux/x86.

После этого планируется устранить наибольшую помеху широкому использованию PPDS – разработать современный WWW-интерфейс, в том числе, и решающий задачи по заполнению БД и их “содержательному” администрированию (см. RPP Encoder Interface [17]). Оценка объема работ – ~ 1 человеко-год.

Результаты работы

В результате проделанной работы успешно осуществлен перенос СУБД BDMS на POSIX-совместимую платформу `/GNU/Linux/x86`. Функциональность BDMS на ОС GNU/Linux соответствует требованиям, предъявляемым к иерархической СУБД. Этот основной результат позволяет сохранить эксплуатируемую в ИФВЭ систему БД PPDS на неограниченный срок, а также значительно ее усовершенствовать, в том числе, в сторону "дружественности" к пользователям.

Полученные наработки могут быть использованы при миграции унаследованного программного обеспечения (DEC Fotran) с ОС OpenVMS на POSIX-совместимые системы.

BDMS также может быть использована в качестве встраиваемой СУБД для физических приложений.

Благодарности

Считаю своим долгом выразить благодарность:

научному руководителю Зенину О.В., предложившему тему работы и оказавшему помощь в ее выполнении;

сотруднику ИФВЭ Куянову Ю.В., оказавшему неоценимую помощь на первом этапе работы;

соавтору BDMS, сотруднику ИФВЭ Луговскому С.Б., чьи консультации многократно сокращали на несколько дней время отладки;

сотруднику ИФВЭ Ежеле В.В. за плодотворные обсуждения данной работы;

Кафедре физики высоких энергий МФТИ, за создание благоприятных условий для выполнения работы.

Список литературы

[1] V. V. Ezhela, “Particle Physics Data Systematization At IHEP”, Comput. Phys. Commun. **33** (1984) 225.

[2] D.R. Richards, “BDMS User's Manual (Version 2.2)”, LBL-4683, [http://hermes.ihep.su:8001/bdms2users_manual.txt].
D.R. Richards, “BDMS Programmer's Manual (Version 2.2)”, LBL-4684, [http://hermes.ihep.su:8001/bdms2users_manual.txt].

[3] “Руководство по BDMS/4”, ЛСФ ОАФ ИФВЭ, 1997, [http://hermes.ihep.su:8001/bdms4manual.ps].

[4] См. [3], раздел 1.3.1.

[5] “Руководство по БД ReactionData”, ЛСФ ОАФ ИФВЭ, 1992, [http://hermes.ihep.su:8001/rd.ps].

[6] “БД PPDS.CS”, ЛСФ ОАФ ИФВЭ, 1993, [http://hermes.ihep.su:8001/cs.ps].

[7] V.V. Ezhela et al., AIP Press, Woodbury, New York, 1996 PARTICLE PHYSICS. *One Hundread Years of Discoveries. An Annotated Chronological Bibliography.*

[8] Группа HEPDATA (Даремский ун-т, Великобритания) поддерживает БД REACTION DATA, содержащую экспериментальные данные по полным и дифференциальным сечениям реакций и другим наблюдаемым ФВЭ [http://durpdg.dur.ac.uk/hepdata/reac.html]. БД REACTION DATA (Дарем) значительно уступает RD (ИФВЭ) по возможностям интерактивного использования. К достоинствам БД REACTION DATA (Дарем) следует отнести развитый WWW-интерфейс, облегчающий ее использование неквалифицированными и эпизодическими пользователями.

[9] http://www.slac.stanford.edu/spires/

[10] http://durpdg.dur.ac.uk/HEPDATA/

[11] ANSI X3J3/90.4, http://www.fortran.com/fortran/F77_std/f77_std.html

[12] Hewlett Packard Corp., “OpenVMS Programming Concepts Manual”, *June 2002*

[13] Compaq Corp., “OpenVMS System Services Reference Manual”, *November 1996*

[14] Compaq Corp., “OpenVMS Record Management Services Reference Manual”, *January 1999*

[15] Digital Equipment Corp., “OpenVMS Obsolete Features Manual”, *May 1993*

[16] SunSoft, Sun Microsystems Inc., “FORTRAN 77 4.0 Reference Manual”, *November 1995*, Part No.: 802-2998-10, Rev. A

[17] Free Software Foundation, “On-line g77 documentation”, `http://gcc.gnu.org/onlinedocs/gcc-3.4.6/g77/`

[18] К.С.Луговский, В.С.Луговский, С.Б.Луговский, “pdgLive. Интернет версия PDG-обзора физики частиц. (Руководство для пользователей)”, Препринт ИФВЭ 2007-10