

УДК 519.6

ЕДИНАЯ СИСТЕМА УПРАВЛЕНИЯ ЗАДАНИЯМИ НА ЭВМ НЕОДНОРОДНОГО ВЫЧИСЛИТЕЛЬНОГО КОМПЛЕКСА

А. Б. Киселев, Ю. Г. Бартнев, А. М. Варгин, С. Н. Киселев, С. И. Колпаков
(РФЯЦ-ВНИИЭФ)

Описываются архитектура и возможности единой системы управления счетом параллельных приложений на ЭВМ ФГУП РФЯЦ-ВНИИЭФ. Система создана для организации централизованного ввода, контроля и управления счетом заданий на вычислительных мощностях неоднородного вычислительного комплекса.

Введение

Еще в советское время Вычислительный центр (ВЦ) РФЯЦ-ВНИИЭФ эксплуатировал неоднородный вычислительный комплекс (НВК), состоящий из моделей БЭСМ, СВС, ЕС ЭВМ и "Эльбрус". Для его эффективного использования была создана система общего ввода заданий ЛМ [1]. Новые ЭВМ и связанные с ними изменения в технологии работы пользователей потребовали реализации современного, универсального программного обеспечения, в полной мере удовлетворяющего требованиям производственного счета. Оно было создано и эксплуатируется в ВЦ ВНИИЭФ с 2005 года.

Единая система управления заданиями (ЕСУЗ) [2] не только организует общий ввод заданий, но и выполняет задачи централизации управления растущим парком ЭВМ, формирует единый вычислительный комплекс: ежедневно на вычислительных мощностях ВНИИЭФ считается большое количество заданий, составленных для выполнения на ЭВМ с разными операционными системами, коммуникационным программным обеспечением и локальными системами пакетной обработки заданий. Задание для таких систем представляет собой текстовый файл, содержащий набор таких реквизитов, как время счета программы, исполняемый файл, количество узлов и процессоров, переменные окружения и т. д.

ЕСУЗ позволяет охватить потоки неоднородной по характеру информации, отображать ее в унифицированном и понятном для разных групп пользователей виде, формализовать син-

таксис языка описания (создания) заданий, сделать "прозрачным" управление счетом.

Система реализована с применением современных технологий: графические программы загружаются с Web-сервера, работа с ними производится в Web-браузере, использованы языки программирования Java, XML и стандартная система управления базами данных (СУБД).

В зарубежных компьютерных вычислительных центрах эксплуатируются подобные системы, например Livermore Computing Resource Management System (LCRM) [3] и Globus [4], но сравнивать с ними ЕСУЗ затруднительно, поскольку цели этих систем значительно отличаются. В Globus, в отличие от ЕСУЗ, много внимания уделено защите информации, поскольку система функционирует в среде Интернет. Целью LCRM является приоритетное предоставление вычислительных ресурсов пользователям, разбитым на группы, подразделения или организации. Работа с системой в основном осуществляется в режиме командной строки. LCRM поддерживает только автоматическое распределение заданий по вычислительным ресурсам, ЕСУЗ же допускает и ручное управление. В отличие от указанных систем ЕСУЗ функционирует в замкнутой сети ВНИИЭФ, где не требуется мощных средств разграничения доступа.

Возможности ЕСУЗ

В ЕСУЗ заложены следующие возможности:
– централизованное управление заданиями;

- централизованный контроль выполнения заданий и вычислительных ресурсов;
- автоматизация выбора подходящей для счета задания ЭВМ;
- автоматизация планирования счета заданий;
- единое операторское управление потоками заданий на НВК;
- единообразное формирование заданий для всех ЭВМ;
- автоматический рестарт заданий, завершившихся с ненулевым кодом возврата¹, поддержка одношаговых и многошаговых вариантов заданий;
- идентификация пользователей, контроль разрешений, защита от несанкционированного доступа;
- сбор и обработка статистической информации.

¹Код возврата — числовое значение, формируемое программой или операционной системой после завершения программы и указывающее характер обнаруженных ошибок или их отсутствие.

Архитектура системы

Ввод заданий на НВК осуществляется с помощью подсистемы командной строки и графического интерфейса, использующих доступ к базе данных (БД) ЕСУЗ, которая необходима для подключения, идентификации пользователей, хранения и обработки информации. Архитектуру ЕСУЗ можно представить в виде схемы, изображенной на рис. 1. Потоки информации движутся в направлениях сверху вниз и снизу вверх.

Верхний уровень — это пользователи ЕСУЗ, которые разделяются на следующие группы: диспетчеры, операторы, администраторы, математики и руководители.

Графический интерфейс — это набор программ, с которыми пользователь может работать с помощью Web-браузеров Internet Explorer и Netscape.

Для возможности формирования заданий с индивидуальными атрибутами в ЕСУЗ создана подсистема командной строки. Все запускаемые в режиме командной строки задания обрабаты-

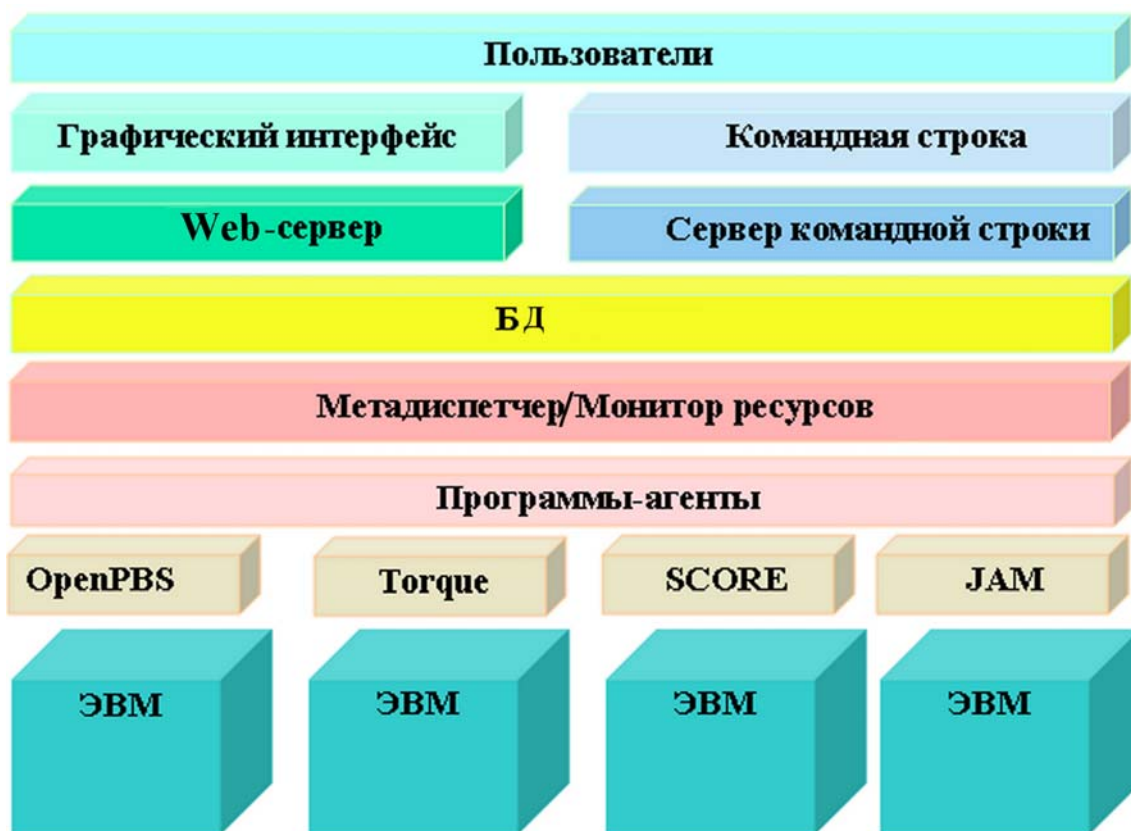


Рис. 1. Архитектура ЕСУЗ

вает сервер командной строки. Сервер проверяет правильность указанных реквизитов задания и фиксирует их в БД.

Метадиспетчер является главной программой ЕСУЗ. Функции метадиспетчера следующие: запуск, управление, планирование, обработка одношаговых и многошаговых заданий. Кроме того, его работа тесно связана с монитором ресурсов — программой, обеспечивающей обратную связь метадиспетчера с пакетными системами обработки заданий. Прием/передача информации осуществляется с помощью программных агентов.

Для контроля момента запуска/останова заданий, а также доставки кодов возврата программ в каждой поддерживаемой системе пакетной обработки заданий используются *пролог* и *эпилог*. В прологе регистрируется время старта, а в эпилоге осуществляется доставка в ЕСУЗ кода возврата программы. Таким образом, для обеспечения функциональных возможностей ЕСУЗ реализовано несколько подсистем, выполняющих следующие функции:

- ввод заданий, контроль и управление счетом с помощью графических программ;
- поддержку командной строки (сервер и конвертеры);
- обеспечение целостности, хранение информации (БД);

– обработку заданий, контроль ресурсов НВК (метадиспетчер и монитор ресурсов).

Технология создания и запуска заданий

ЕСУЗ поддерживает два варианта создания/запуска заданий — в режиме командной строки и посредством графического интерфейса.

Графический интерфейс. Графический интерфейс обеспечивает обращение к программам, разработанным для всех пяти групп пользователей. Интерфейс унифицирует процесс создания заданий для выполнения на НВК. Вместе с тем работа пользователя с отдельной ЭВМ и ее файловыми ресурсами совершенно прозрачна, так как ЕСУЗ при успешной авторизации автоматически предоставляет права доступа ко всему НВК.

В графическом окне создания и запуска заданий (рис. 2) пользователь вводит необходимые для пакетной системы значения: число процессоров, время выполнения, другие реквизиты. Он также указывает атрибуты, нужные диспетчерской службе: заказчика данной задачи, номер темы и название методики. Кроме того, пользователь может ввести другие специфические параметры: значение "пульса" считающейся прог-

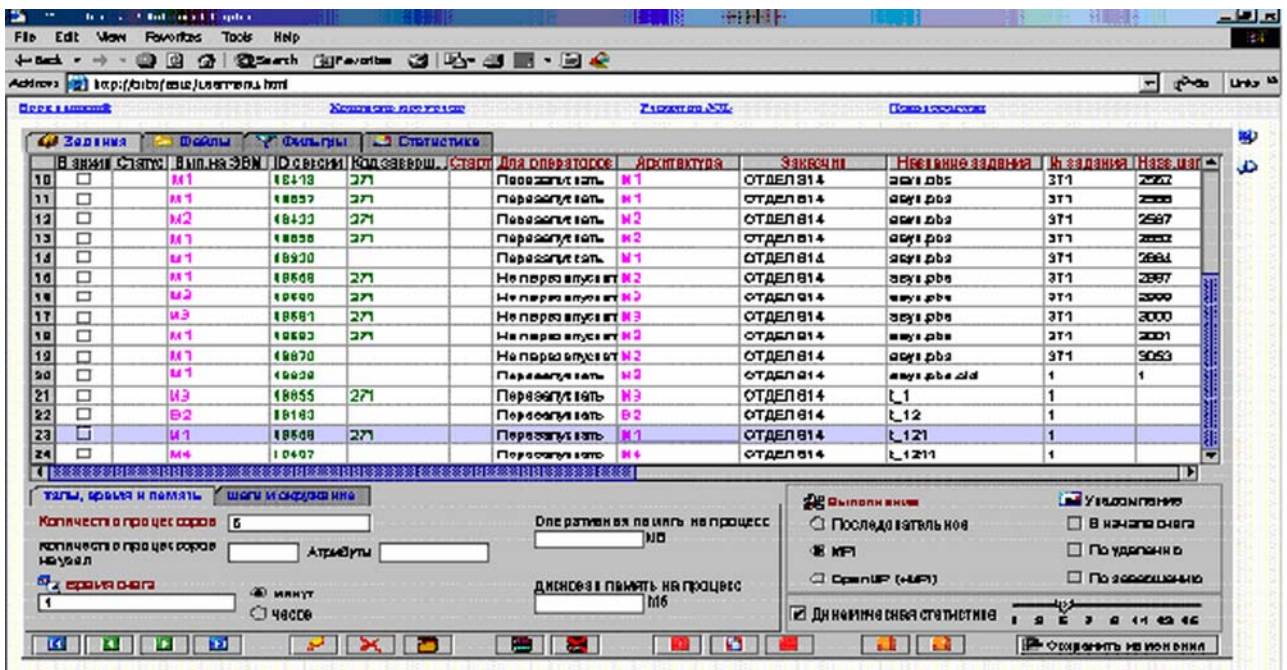


Рис. 2. Графическое окно создания и запуска заданий

раммы², минимальное количество процессоров, время вычислительного шага.

Графическое окно предоставляет возможность просмотра, редактирования и удаления файлов, находящихся на любой ЭВМ НВК, с помощью щелчка мышью на закладке "Файлы" (рис. 3).

Описываемый графический интерфейс не позволяет использовать все уникальные возможности поддерживаемых пакетных систем обработки — для унификации формирования заданий взяты только их основные особенности.

Графическое окно позволяет простым нажатием кнопки копирования без количественных ограничений создавать однотипные задания, одновременно запускать их и останавливать; при этом не имеет значения, на какой ЭВМ они находятся. Эта возможность очень удобна для подготовки и ведения массового счета из серий коротких заданий. Все изменения пользователь должен сохранять в БД ЕСУЗ.

²Имеется в виду интервал времени, через который выдаются сообщения о функционировании программы.

Графический интерфейс предоставляет возможность контролировать задания на любой ЭВМ. Он информирует пользователя о загрузке и доступности НВК — его работе, выключении или профилактике.

На закладке "Статистика" по выбранному заданию в виде таблиц и графиков отображаются статистические данные, связанные с использованием МРІ-функций в параллельном пользовательском приложении.

Для интерактивного взаимодействия с выполняющейся программой предназначена закладка "Информация". С помощью этой закладки пользователь может передавать программе, в которой используются функции из библиотеки ЕСУЗ, приказы и получать от нее ответы, например, направить заданию приказ записать в БД ЕСУЗ графическое изображение, получившееся в результате счета.

Режим командной строки. Для обеспечения запуска заданий в режиме командной строки

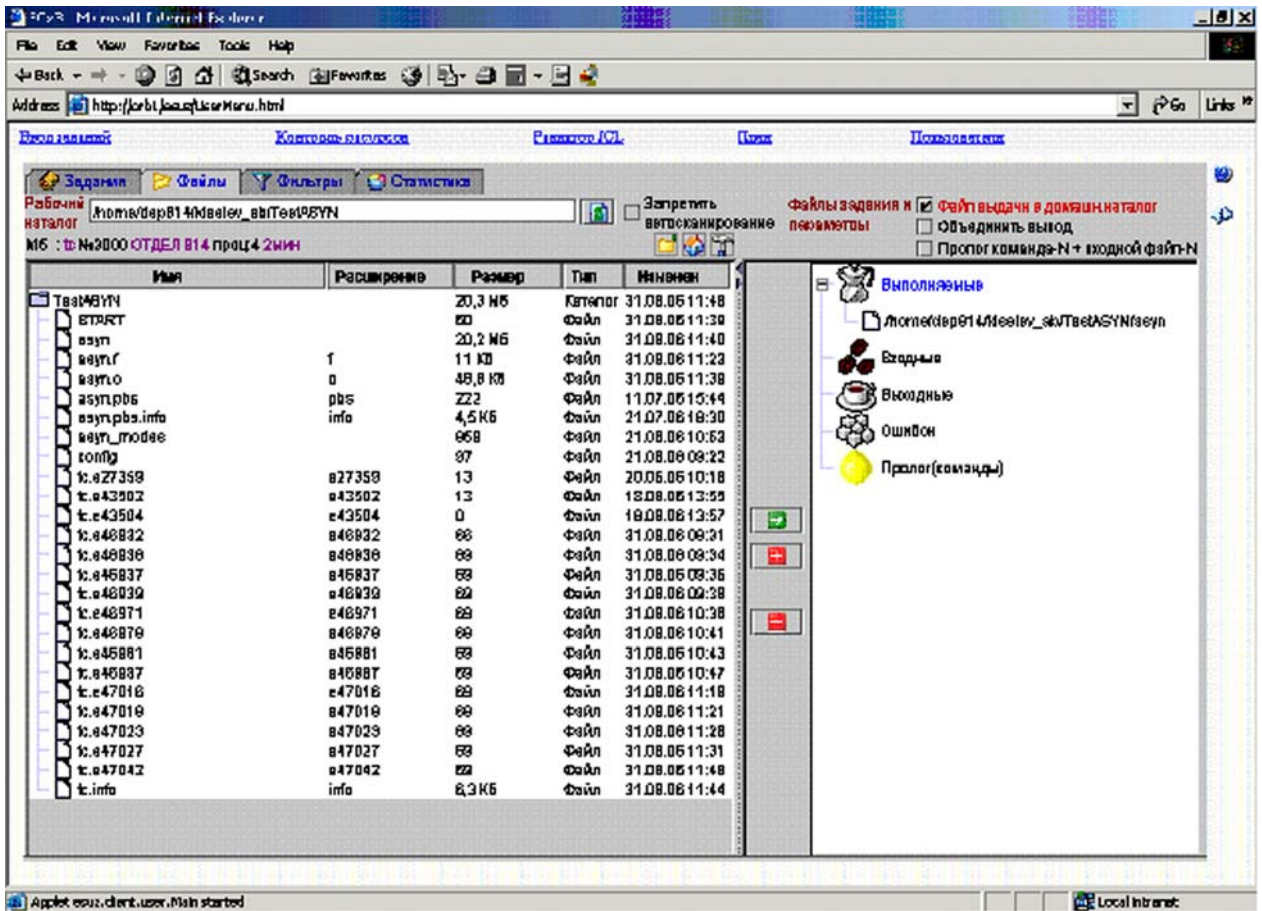


Рис. 3. Файловый менеджер. Окно создания и запуска заданий

в ЕСУЗ создана соответствующая подсистема, которая содержит конвертеры пользовательских сценариев систем пакетной обработки заданий Open PBS, Torque, SCORE и JAM.

Для каждой системы пакетной обработки реализован соответствующий конвертер информации, который обрабатывает и передает в БД ЕСУЗ реквизиты заданий. Вызов конвертера прозрачен для пользователя, так как при запуске утилиты qsub или maui submit фактически происходит обращение к конвертеру. Основное отличие состоит в том, что вместо идентификатора задачи в пакетной системе обработки заданий пользователь получает сообщение о допущенной ошибке или об успешной обработке реквизитов.

Так, для Open PBS, Torque и SCORE используется одна программа преобразования данных, которая разбирает управляющие тэги этих систем, например строку заказа ресурсов

```
#PBS -l nodes=20:ppn=6
```

Для системы пакетной обработки заданий JAM предназначен другой конвертер, так как для заказа тех же ресурсов она использует иные синтаксические конструкции:

```
Nodes == 20
```

```
TaskPerNode == 6
```

После разбора информации и проверки значений заказываемых ресурсов конвертеры посылают обработанную информацию в структурированном виде серверу командной строки. Структурированный вид задания обеспечивается с помощью унифицированного языка описания заданий ЕСУЗ JCL-XML. JCL-XML — это универсальный язык, разработанный для создания и запуска заданий не только на больших ЭВМ, но и с рабочих компьютеров пользователей.

Программы серверной части ЕСУЗ

Программы серверной части ЕСУЗ состоят из сервера командной строки, метадиспетчера, программ-агентов выполнения команд на ЭВМ.

Сервер командной строки. Сервер командной строки предназначен для обеспечения работы пользователей в режиме командной строки. Его функции сводятся к проверке, обработке и регистрации в БД ЕСУЗ заданий, получаемых от конвертеров. Благодаря использованию JCL-XML процедура проверки выполняется на уровне стандартного обработчика документов XML, который сравнивает полученную информацию на соответствие генерируемому в

процессе работы шаблону. Значения, указанные в реквизитах задания, должны соответствовать эталонной информации, присутствующей в шаблоне.

После регистрации заданий в БД ЕСУЗ их обрабатывает метадиспетчер.

Метадиспетчер. Метадиспетчер реализует централизованный подход к диспетчеризации заданий на НВК. Задача метадиспетчера — автоматическая обработка множества заданий, зарегистрированных в БД ЕСУЗ, их распределение по доступным ресурсам, запуск и мониторинг их выполнения.

Программное обеспечение состоит из собственно метадиспетчера и монитора ресурсов НВК.

Метадиспетчер работает в соответствии со следующим алгоритмом:

- поиск наиболее подходящего для выполнения задания кластера в группе ЭВМ одной архитектуры (для ЕСУЗ архитектура ЭВМ — это совокупность таких параметров, как операционная система, тип процессора и коммуникационная среда);
- проверка выполнения условия запуска шага в многошаговом задании;
- генерация файла с реквизитами задания для системы пакетной обработки выбранной ЭВМ с учетом режима счета задания (последовательный, MPI- или OpenMP-распараллеливание);
- передача файла с реквизитами в систему пакетной обработки выбранной для задания ЭВМ.

Метадиспетчер работает с каждой обслуживаемой системой пакетного запуска в установленном администратором режиме. Поддерживаются два режима работы:

- 1) *пассивный* — устанавливается, если внутренние очереди пакетной системы обработки заданий закрыты, т. е. все поступающие в такую систему задания не будут выполнены до тех пор, пока оператор не запустит их вручную;
- 2) *активный* — используется тогда, когда внутренние очереди в пакетной системе обработки заданий открыты. Метадиспетчер специально задерживает задания, предназначенные для выполнения, в своей внутренней очереди, передавая их пакетной системе по мере освобождения вычислительных ресурсов.

Таким образом, в метадиспетчере реализован алгоритм оперативного планирования счета заданий, учитывающий количество свободного вычислительного ресурса на ЭВМ. Для развития возможностей планирования по времени выполнения разработана система пакетной обработки заданий ЖАМ, которой будет посвящена отдельная статья.

Основная функция монитора ресурсов НВК — обеспечение метадиспетчера информацией для принятия решения при работе с заданиями. Монитор контролирует:

- состояние ресурсов ЭВМ (вычислительный элемент полностью загружен, свободен, доступен, исключен из рабочего списка, отключен);
- изменение состояний заданий. Поддерживается четыре основных статуса задания: 1) ожидает запуска; 2) задание "заморожено"; 3) выполняется; 4) задание завершилось, но находится в режиме освобождения вычислительных ресурсов;
- состояние очередей систем пакетного запуска заданий (закрыты/открыты);
- значение параметра системы качества обслуживания (QoS).

При изменении состояния отдельного задания или вычислительного ресурса монитор всегда делает запись в журнале событий, который располагается в БД ЕСУЗ. Информация автоматически появляется в графическом окне оператора НВК.

БД ЕСУЗ. Работа составляющих ЕСУЗ компонентов почти целиком построена на взаимодействии с БД, которая организована с помощью СУБД, обеспечивающей:

- целостность данных;
- единое информационное пространство между пользователями и подсистемами ЕСУЗ. БД выступает в роли оперативного буфера для передачи пользовательских приказов параллельной программе при интерактивной работе, а также получения двоичной/текстовой информации в течение выполнения программы на ЭВМ;
- реализацию механизма циклического перезапуска заданий при их ненормальном завершении;
- мониторинг работы метадиспетчера и его перезапуск при отказе;

- ведение журнала событий, произошедших с заданием, с вычислительными элементами ЭВМ в результате команд пользователя, диспетчера или оператора;
- накопление статистических данных, связанных с работой НВК.

В СУБД реализована первичная обработка информации и такие операции, как:

- подготовка для отображения статистических данных, собранных с использованием инструментального средства сбора и анализа статистической информации счета программ пользователей (STK). STK — это интегрированный с ЕСУЗ отдельный программный комплекс;
- ведение статистики полезной загрузки НВК, подсчет коэффициентов целостности, готовности и обслуживания ЭВМ;
- удаление информации из БД по заданиям недельной давности, созданным в режиме командной строки;
- удаление из БД цепочек многошаговых заданий недельной давности;
- контроль зависания заданий и информирование об этом оператора НВК.

Таким образом, применение СУБД позволяет не только легко решать целый ряд технических проблем, связанных с многопользовательским режимом работы, но и обеспечивать централизацию хранения, доступа к информации, проводить ресурсоемкие операции с информацией непосредственно в СУБД, выполнять отправку писем пользователю по началу и концу выполнения, удалению и "зависанию" задания.

Программные агенты. С помощью программных агентов метадиспетчер, монитор ресурсов и часть графических программ могут выполнять команды на НВК, а также получать информацию о названиях и размерах файлов, находящихся на разных ЭВМ.

Управление и планирование счета на вычислительном комплексе

Планирование счета на НВК осуществляет диспетчерская служба, которая в соответствии с некоторыми критериями назначает заданиям плановые номера. Задача подсистем ЕСУЗ состоит в том, чтобы связать плановый номер с

приоритетом задания в системе пакетной обработки на ЭВМ.

Управление счетом и ресурсами на вычислительном комплексе осуществляется с помощью графического интерфейса оператора, который позволяет выполнять все необходимые операции в одном графическом окне. Все операторы НВК могут работать параллельно, загрузив данную программу, каждый на свой компьютер.

Библиотека функций для взаимодействия с БД ЕСУЗ

Для организации интерактивного взаимодействия пользователя со своим выполняющимся на ЭВМ приложением создана библиотека функций ЕСУЗ. На рис. 4 приведена схема взаимодействия пользователя и программы с помощью этой библиотеки.

БД ЕСУЗ выступает в этой схеме в качестве буфера данных. С помощью библиотеки выполняющаяся на ЭВМ программа считывает приказ пользователя из БД и выполняет заказанное пользователем действие. В библиотеку входят как функции передачи заданию управляющих команд, так и процедуры записи/чтения дополнительной информации, например, предназначенной для оповещения оператора НВК и пользователя о событиях, произошедших в программе.



Рис. 4. Схема интерактивной работы пользователя со своей считающейся программой

Заключение

Основными свойствами, которые выделяют ЕСУЗ среди систем, обеспечивающих выполнение параллельных программ на комплексе ЭВМ,

являются унификация и централизация. Кроме того, ЕСУЗ имеет следующие индивидуальные особенности:

- поддерживает автоматический циклический перезапуск заданий;
- предоставляет пользователю возможность с помощью библиотеки доступа к информации, хранящейся в БД, организовать интерактивное взаимодействие со своим приложением;
- обеспечивает ручное и автоматическое управление счетом;
- позволяет в одном графическом окне формировать задания для любой ЭВМ с любой поддерживаемой архитектурой;
- имеет собственную систему пакетной обработки заданий (JAM).

ЕСУЗ — расширяемая система, способная удовлетворить специфические требования отдельных групп пользователей ВЦ РФЯЦ-ВНИИЭФ.

Список литературы

1. Кибкало А. А. Система ввода заданий для комплекса ЭВМ // Вопросы атомной науки и техники. Сер. Математическое моделирование физических процессов. 1989. Вып. 1. С. 52—55.
2. Варгин А. М., Кульнев Д. В., Семенов Г. П., Олесницкий А. Б., Холушкин В. С., Романова М. Д., Фролова Н. В., Бармин А. М., Оразмагомедов Р. А., Киселев А. Б. Принципы построения единой системы управления заданиями неоднородного вычислительного комплекса ИТМФ // Сб. тез. докл. на Межд. семинаре "Супервычисления и математическое моделирование". Саров, 6—11 октября 2003 г. Саров: РФЯЦ-ВНИИЭФ, 2003.
3. Livermore Computing Resource Management System (LCRM). LCRM Reference Manual. <http://www.llnl.gov/LCdocs/dpcs>.
4. Globus Metacomputing Infrastructure Toolkit. <http://www.globus.org>.

Статья поступила в редакцию 28.09.07.