

Формализация информации и Big Data

«02.03.03 - Математическое обеспечение и администрирование информационных систем
направленность разработка и администрирование информационных систем»

<http://vikchas.ru>

Лекция 11 «СУБД Adabas Big Data»

Часовских Виктор Петрович

д-р техн. наук, профессор кафедры ШИиКМ

ФГБОУ ВО «Уральский государственный экономический
университет»

Екатеринбург 2023

Функциональные возможности системы

СУБД ADABAS и ее окружение

ADABAS — постреляционная СУБД, поддерживающая различные модели данных, включающая вложенные отношения и иерархические поля. С помощью ADABAS можно строить иерархические, сетевые и реляционные SQL базы данных, а также сложные текстовые информационно-поисковые и интегрированные системы. Разумеется, ADABAS дает пользователям возможность работать с данными в рамках чисто реляционной модели. ADABAS работает на разных платформах — Windows, Unix, Open VMS, AS/400, BS2000, MVS, VM, OS/390 и др. Обладает возможностью выборочного страхового сохранения данных в процессе обычного сеанса, что важно при работе с большими объемами данных и большим количеством активных пользователей. Также параллельно могут выполняться административные и другие задачи обслуживания баз данных. В то же самое время использует относительно небольшое количество вычислительных ресурсов, и затраты на обслуживание баз данных невелики. ADABAS прекрасно соответствует требованиям в области мультимедиа-данных, в области управления документами и др. (поддерживает динамические переменные). В отличие от систем, имеющих одну модель для всех типов данных, поддерживает различные модели и структуры данных, которые могут быть спе

циально спроектированы для работы с разными видами приложений. На рис. 1 показаны основные возможности СУБД ADABAS в виде схемы.

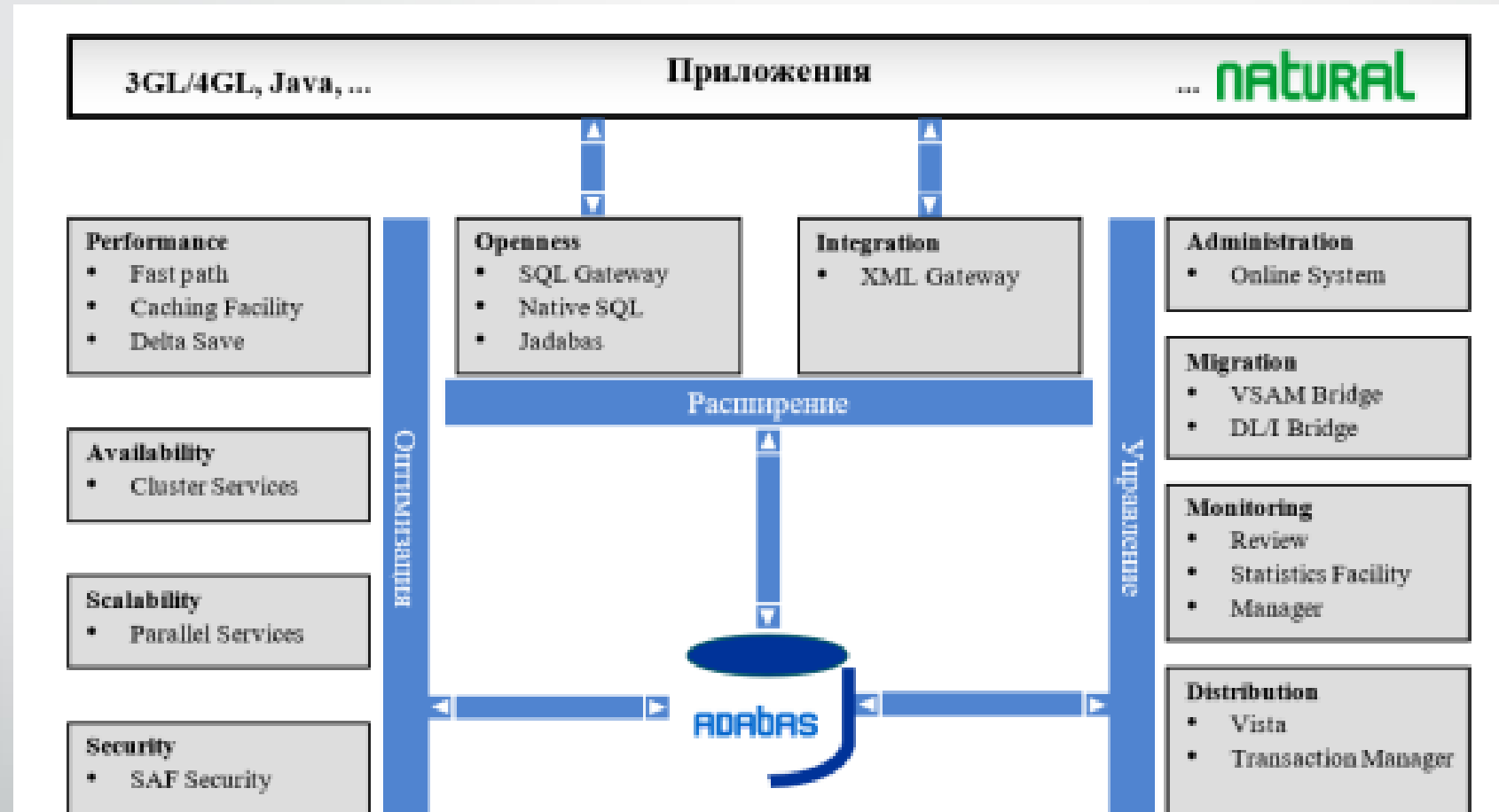


Рис. 1. СУБД ADABAS и ее окружение

Основные характеристики СУБД ADABAS:

- высокая надежность и исполнение;
- полная доступность 24×7;
- минимальная потребность в администрировании;
- доступность дополнительных расширений;
- доступность разных платформ;
- реляционные и вложенные реляционные структуры;
- различные методы доступа;
- меню-управляемые утилиты;
- администрирование, когда база данных активная.

Результаты тестов, проводившихся в лабораториях IBM:

- 5 000 транзакций в секунду;
- 160 000 вызовов в секунду;
- 40 000 параллельных пользователей

Главные конструктивные особенности, обеспечивающие высокую эффективность ADABAS:

- вложенные отношения позволяют уменьшать схему базы данных и увеличивать количество данных, передаваемых за одну операцию ввода-вывода;
- автоматическое, независимое от платформы сжатие данных требует меньше объема памяти для хранения и позволяет оптимизировать процедуры ввода-вывода;
- обусловленное спецификой приложений использование различных типов памяти ЭВМ сокращает время обработки данных и уменьшает число операций ввода-вывода;
- блокировки на уровне строки (записи) в многопользовательском режиме максимально снижают проблемы доступа (клинча) к базам данных и улучшают условия для параллельной обработки данных;
- способы хранения и доступа к данным отделены от особенностей конкретных физических носителей, что делает СУБД гибкой и эффективной

Полнота функций

ADABAS относится к классу функционально развитых СУБД, которые обеспечивают пользователей большим набором функциональных возможностей. Она рассчитана на широкий класс практических применений, требующих коллективного доступа к базам данных при решении интерактивных прикладных задач приложений информационных систем.

Предназначена для накопления, интегрированного хранения, поиска, обработки и представления данных в форме, удобной для подготовки и принятия решений. Функции и технические характеристики этой системы позволяют применять ее для решения широкого класса информационных задач без разработки дополнительных прикладных программ. Различные виды поиска, широкие возможности логической обработки выбранной информации позволяют получать разнообразные типы выходных документов.

В процессе диалогового взаимодействия с системой ADABAS производится форматированный ввод информации

в БД, модификация хранимых данных, поиск и обработка данных, а также подготовка выходных отчетов.

Средства описания сценариев диалога позволяют реализовать не только отдельные слабо связанные элементарные процессы поиска, ввода, корректировки и обработки данных, но и совокупности взаимосвязанных информационных или технологических процессов, которые могут быть каталогизированы в словаре данных системы и в дальнейшем выполнены в автоматическом режиме.

Программы ADABAS обеспечивают:

- создание и ведение больших БД сетевой или иерархической структуры;
- диалоговое программирование прикладных задач;
- обработку данных и форматирование отчетов по запросам пользователей;

- централизованное управление доступом к БД;
- сохранность информации в БД;
- поддержание целостности и непротиворечивости БД;
- ведение словаря данных;
- настройку на программно-аппаратную среду и управление функционированием системы;
- диалоговый и пакетный режимы работы.

По совокупности реализуемых функций ADABAS может быть отнесена к классу интегрированных систем, обеспечивающих ввод, обработку, корректировку и редактирование исходных данных с возможностью управления поиском и анализом информации в БД. Обладает средствами гибкой перестройки и развития структуры БД в соответствии с изменяющимися потребностями прикладных задач.

ADABAS включает в себя гибкие и простые в освоении языковые и технологические средства создания и эксплуатации БД информационных систем коллективного пользования, позволяет создавать эффективную унифицированную технологию их разработки, развертывания и ведения.

Естественность модели данных

Используемая в СУБД модель данных должна обеспечивать возможность отображения объектов реального мира в естественных терминах и понятиях, не навязывающих пользователю непривычные для него представления.

Модель данных ADABAS обладает одновременно простотой и значительной гибкостью.

База данных состоит из совокупности связанных между собой файлов.

Логическая структура записей файлов обеспечивает естественное отображение объектов предметной области с иерархической структурой.

Возможность динамического связывания различных файлов позволяет создавать иерархические или сетевые структуры, предоставляя механизм для отображения иерархических структур объектов предметной области или установления между ними более сложных взаимосвязей. При этом не требуются другие средства, кроме средств связывания для физического представления таких объектов в БД.

Указатели, поддерживающие иерархические или сетевые структуры, не хранятся в БД вместе с записями, а формируются динамически на основе связей, которые, например, могут быть созданы или переопределены после загрузки БД.

В модели данных ADABAS отсутствуют понятия «корневой» и «подчиненный» файлы. В зависимости от точки зрения пользователя один и тот же файл в разных запросах может быть как корневым, так и подчиненным.

Основной особенностью модели данных системы ADABAS является наличие механизмов для целенаправленного разбиения БД на классы ассоциаций, доступ к которым обеспечивается на основе значений атрибутов объектов предметной области. В результате создается возможность предварительной статической структуризации и выделения подмножеств записей БД на этапе ее создания.

Логическое разделение БД и выделение ассоциативных подмножеств, содержащих информацию о разбиении на непересекающиеся множества, создает конструктивную основу для ассоциативного доступа к объектам предметной области и обеспечивает доступ к этим объектам только на основе информации о тех свойствах, которыми они характеризуются в рамках соответствующей предметной области.

Более того, в рамках такого представления возникает естественная возможность динамического выделения подмножества объектов (динамическая структуризация), удовлетворяющих запросу, без непосредственного обращения к записям БД. Пользователю достаточно задать требуемые предикаты, определяющие подмножества объектов с заданными свойствами, и связать их логическими операциями, которые необходимо выполнить над выделенными подмножествами объектов

Такая технология работы с данными позволяет обеспечить:

- равнодоступность всех объектов и динамический характер связей между объектами БД;
- доступ к записям БД на основе значений нескольких атрибутов, определяемых в запросе и связанных логическими операциями;
- создание мощных высокоуровневых языков обработки данных;
- быструю реакцию системы на запросы пользователей за счет статического выделения подмножества объектов, ориентированных на последующую обработку данных;
- построение интегрированных документально-фактографических систем.

В модели данных ADABAS имеются два типа команд — команды манипулирования записями и команды обработки ассоциаций, обеспечивающие доступ к записям по значениям атрибутов и связям с другими записями, обладающими заданными свойствами.

Среда хранения БД ADABAS физически разделена на накопитель и ассоциатор.

Накопитель содержит записи БД, а ассоциатор — ассоциативные списки внутрисистемных номеров записей.

Такое построение во многих случаях обеспечивает возможность выдачи результата на основе информации, содержащейся только в ассоциаторе, без непосредственного обращения к записям БД.

ADABAS предоставляет пользователям также ряд внешних моделей, поддерживаемых соответствующими языковыми средствами. Эти модели в большей степени ориентированы на конкретные категории пользователей и обеспечивают им более удобное взаимодействие с системой. В частности, администратор БД имеет возможность с помощью средств описания интерфейсных модулей определить не только подсхему, доступную прикладной программе, но и допустимые в рамках этой подсхемы операции

Надежность функционирования

Надежность работы СУБД может рассматриваться в двух аспектах:

- надежность организации вычислительного процесса;
- обеспечение целостности БД.

Надежность организации вычислительного процесса обеспечивается в ADABAS в первую очередь путем отделения программ, управляющих мультидоступом к БД, от прикладных программ и компонентов системы, реализующих обработку запросов пользователей. Такое построение системы полностью исключает возможность разрушения программ управления БД в случае ошибок в работе прикладных программ путем использования предоставляемых операционной системой программных и аппаратных средств защиты памяти.

ADABAS предоставляет широкие возможности для сохранения целостности данных в случае сбоев программного обеспечения, отказов аппаратуры и ошибок в прикладных программах.

Для этой цели в системе предусмотрено:

- восстановление БД со страховой копии или на заданную контрольную точку;
- автоматическое восстановление БД на синхронную контрольную точку или транзакцию;

- удержание записей БД на время их модификации; автоматический контроль взаимной согласованности ассоциатора и накопителя;
- автономный контроль корректности ассоциатора; синтаксический и семантический контроль значений атрибутов.

Таким образом, целостность БД поддерживается на уровнях:

- всей БД;
- файлов;
- контрольных точек;
- транзакций;
- отдельных команд и атрибутов.

Механизмы поддержания целостности БД данных на уровне транзакции, команды и атрибута работают в автоматическом режиме. Этим обеспечивается быстрое восстановление БД в целостное состояние и повышается эффективность функционирования системы.

Самостоятельную группу средств, поддерживающих надежное функционирование системы, составляют компоненты обеспечения сохранности информации в БД. Сохранность данных гарантируется в ADABAS на уровне доступа в систему, доступа к файлам и атрибутам записей

Производительность системы

Высокая производительность ADABAS при выполнении операций поиска данных является следствием принятой организации БД, в основе которой лежит принцип отделения механизма реализации путей доступа от самих данных.

Информация о путях доступа к данным содержится в ассоциативных списках, обработка которых с учетом логических условий поиска обеспечивает идентификацию релевантных записей только на основе значений поисковых атрибутов без необходимости доступа к файлам накопителя БД. При этом значительно сокращается объем данных, которые требуется подвергнуть обработке, и, как следствие, уменьшается время обработки для тех приложений, где количество операций доступа и обработки данных превосходит количество операций по ведению БД.

Допускается также рандомизированный доступ к записям по значению атрибута, уникально идентифицирующего записи БД. В этом случае искомая запись выделяется за время, приблизительно равное одному доступу к магнитному диску.

Для массовой обработки данных (ввод значительных объемов данных и формирование отчетов путем последовательной обработки большого количества записей БД) предназначены специальные компоненты системы ADABAS, реализующие массовую загрузку БД и пакетный режим обработки данных, обеспечивающий одновременное формирование не менее 32 отчетов за один просмотр БД. При этом производительность системы может возрасти (в пересчете на одну запись) на порядок по сравнению с произвольным доступом к каждой записи.

При диалоговом взаимодействии с системой важнейшей характеристикой для пользователя является время компиляции запроса. Так, транслятор диалогового языка программирования задач проверяет синтаксис запроса по мере ввода строк, что сокращает время реакции системы как в период трансляции, так и в период выполнения запроса.

С целью повышения пропускной способности системы организуется несколько параллельно работающих процессов взаимодействия с БД, причем один из них обрабатывает команды модификации данных, а остальные — команды, не изменяющие состояние БД. Такая организация вычислительного процесса наряду с повышением производительности работы системы обеспечивает возможность эффективного восстановления БД при аппаратных или программных сбоях.

Кроме того, в ADABAS организуется параллельная работа нескольких процессов на основе применения средств операционной системы (запуск нескольких задач в независимых разделах памяти) и мониторов управления вычислительным процессом (инициализация задачи для каждого видеотерминала). Благодаря параллельной работе на нескольких уровнях достигается эффективное использование центрального процессора.

Эффективность использования вычислительных ресурсов

Записи БД ADABAS хранятся в сжатой форме. За счет сжатия значительно сокращается объем внешней памяти, требуемой для хранения файлов БД, и уменьшается количество операций ввода-вывода. За счет сжатия записей экономится около 50 % внешней памяти. Каждой записи БД присваивается уникальный внутрисистемный номер, который служит ее логическим адресом на все время существования записи в БД.

Использование логической адресации с помощью внутрисистемных номеров позволяет сравнительно просто учитывать и распределять память, которая освобождается в процессе работы, и исключает реорганизацию БД.

Обмен информацией между внешней и оперативной памятью осуществляется через системный буфер. Использование системного буфера и динамического механизма управления им обеспечивает сохранение в оперативной памяти тех блоков БД, к которым наиболее часто обращались в процессе функционирования системы. Благодаря этому существенно сокращается количество необходимых обращений к БД при поиске записей и их модификации. Эффективность использования системного буфера значительно возрастает за счет применения механизма сжатия данных.

ADABAS использует записи переменной длины, которые размещаются в блоках накопителя и ассоциатора, имеющих фиксированный размер. В каждом блоке может быть зарезервирована свободная память.

Резервирование свободной памяти позволяет избежать частых перемещений информации при увеличении размера записей. При этом отпадает необходимость корректировки индексов или таблиц, благодаря чему не уменьшается эффективность доступа к БД по мере ее заполнения. Память, которая освобождается внутри блока, может быть немедленно использована для других записей.

Благодарю за внимание!

