**Тема 2. Бази данни**

* 1. ***Определения и основни характеристики на база данни***

1/ Базата данни е место за съхраняване на електронни данни в компютърната система чрез специални структури (най-често таблици) и с помощта на файлове (един или няколко файла – в зависимост от системата за управление на базата данни).

2/ Базата данни е специфичен начин за оптимално съхраняване на много голямо количество информация, която трябва да се ползва многократно за решаване на много задачи на много потребители.

3/ Базата данни е цялостна компютърна система за регистриране и поддържане на логически свързани помежду си структури от данни.

Базата данни прилича на файловата система на компютъра по съхраняването на електронни данни в компютъра чрез файлове, т.е. на физическо ниво тези два елемента са подобни. Съществената разлика между тях се състои в начините на достъп и на обработване на съхранените електронни данни.

1/ При файловата система достъпът до данните и тяхната обработка се осъществява чрез команди за операционната система на компютъра. Примерно във всяко приложение, работещо под управление на операционната система Windows, е налице меню File, съдържащо команди за опериране с файлове.

2/ При базите данни достъпът до данните и тяхната обработка се осъществява чрез команди от специален език за програмиране, насочен към приложения в базите данни – **SQL** (език за структурирани заявки). За разлика от процедурно-ориентираните езици за програмиране заявките на SQL задават „какво” трябва да се изпълни, а не „как” да се изпълни.

* 1. ***Основните характеристики на базите данни са:***

1/ Базата данни е ценен информационен ресурс на една институция, затова изискванията към нея са много големи. Тя се планира, проектира, създава, съхранява, управлява, използва и поддържа с особено голямо внимание.

2/ Базата данни трябва да съхранява общодостъпни данни, т.е. тя е предназначена за обслужване на изискванията на много потребители за решаване на много и разнообразни техни задачи. По принцип достъпът до данните може да се управлява с пароли и с права за достъп.

3/ Данните в базата данни трябва да се структурират логически по такъв начин, че да бъдат адекватни на бизнес модела на управление на съответната институция, т.е. внедряването на една информационна система с база данни не трябва да променя успешен бизнес модел.

4/ Базите данни следва да се изграждат с минимален излишък на информация, т.е. минимално да се дублира определена част от данните, която обикновено служи за еднозначно определяне на търсен обект. Излишна информация има във всеки модел на управление и по принцип се дублират кодове, които улесняват търсенето на обектите.

5/ Базата данни обикновено се съхранява върху дискови устройства (твърди дискове), но при необходимост може да се архивира и върху друг тип устройства – например магнитни ленти.

6/ Базите данни представляват модели на процеси от реалния свят, като реалните процеси се представят в базата данни чрез: обекти, връзки между обектите („едно към едно”, „едно към много”, „много към много”) и свойства на обектите и на техните връзки.

Обикновено всеки обект, връзка и свойство се означава и управлява с име. Чрез количествено измерване, качествено определяне или по изкуствен път едно свойство може да получи стойност от множество с допустими стойности за това свойство.

* 1. ***Операции над базите данни***

Потребителите на една база данни разполагат с набор от технологични средства (заявки), които им дават възможност да реализират следните операции над данните:

1/ Създаване на нова база данни – ползва се изключително рядко.

2/ Създаване на нова структура (таблица) към налична база данни – много рядко.

3/ Допълване на таблица от базата данни с нов ред – запис за нов обект– сравнително често.

4/ Изтегляне на данни от таблици на базата данни по зададен от потребителя критерий (генериране на справка) – най-често използваната операция в базата данни.

5/ Променяне на налични данни в таблици на базата данни – рядко.

6/ Променяне на структурата на таблица от базата данни – много рядко.

7/ Изтриване на данни от таблици на базата данни – рядко.

8/ Изтриване на таблица от базата данни – изключително рядко.

9/ Изтриване на база данни – по принцип никога, освен в учебни и тестови примери.

* 1. ***Системи за управление на бази данни - определение и свойства***

Една система за управление на база данни (СУБД) в тесен смисъл е софтуерен продукт.

Ролята на СУБД може да се визуализира чрез следната схема:

База данни

СУБД

бодигард,мениджър,

изпълнител

СУБД изпълнява три основни дейности:

1/ Управлява достъпа на потребителите до данните в базата данни, като проверява правата на потребителите, ако достъпът се осъществява чрез пароли и права на достъп.

2/ Проверява коректността на заявките на оторизираните потребители и изпълнява правилно зададените заявки.

3/ Поддържа базата данни след изпълнение на заявки за актуализиране на данните (допълване, променяне, изтриване).

* 1. ***Основните характеристики на СУБД са:***

1/ По заявки на потребител за обработване на данни СУБД осъществява две основни операции над данните в БД: **запис** и **четене**.

2/ Чрез операцията четене СУБД реализира и трета основна операция над данните в БД: **търсене**. Предпоставка за успешно търсене в БД е наличие на два вида информация във всеки ред на всяка таблица от БД:

а) **идентификаторна** – уникална информация, която служи за еднозначно определяне на обектите в БД и се ползва при търсене. Тази информация не може да се обработва.

б) **съдържателна** – същинската информация, заради която се създава БД и служи за описване на обектите. Тази информация се обработва.

3/ СУБД предоставя на потребителя само **логическата структура** на таблиците от БД и скрива от него физически детайли по тяхното съхраняване. Това означава, че достъпът до данните в таблиците на БД е възможен само чрез имена на таблици и на колони, а не могат да се ползват номера на таблици, на колони или на редове.

4/ СУБД трябва да поддържа свойството **цялостност** (пълнота) на данните в БД – това означава, че СУБД трябва да обезпечи информационно всякаква заявка на потребител за обработване на данни, особено заявка за генериране на справки. Липсата на данни за някаква заявка се дължи на лошо проектиране на базата данни.

5/ СУБД разпределя достъпа на потребителите до различните таблици на БД. При заявки за **едновременен достъп** до една и съща таблица на БД от различни потребители (при работа в мрежов режим) СУБД предоставя на всички потребители копие на желаната таблица и те работят с моментното състояние на БД. След приключване на обработката на заявките СУБД проверява дали е реализирана операция по актуализиране на някоя таблица и прави от съответното копие оригинал, който тя съхранява и предоставя за следващи обработки.

6/ СУБД притежава и известни **недостатъци**:

а) разработването на една СУБД е много скъпо и изисква много усилия и време, поради което и продаваемият софтуерен продукт е скъп;

б) СУБД изисква мощен хардуер – бърз процесор, скоростен твърд диск, бърз обмен на данни между оперативната и външната памет – необходими са разходи за закупуване на скъпи компютърни системи.

***2.6. Релационни бази данни – модели, понятия, структура, свойства***

Съществуват няколко модела за организиране на базите данни: мрежов, йерархичен, релационен, обектно-ориентиран, дедуктивен, Web-базиран и др. Най-разпространен е релационният модел на БД, който се изгражда на строга математическа основа. В този модел е заложена релационната алгебра, която предоставя операции над специални обекти, наречени релации. В релационния модел на БД са дефинирани релации и са дефинирани операции за манипулиране с релации.

Нека е налице база данни, състояща се от няколко таблици, които са съхранени в хардуера на БД (твърд диск). Тези структури от данни ще наричаме физически таблици на БД.

Една **релация** представлява визуализирана таблица върху екрана на компютъра, която се получава чрез изтегляне на данни от физическите таблици на БД по зададен от потребителя критерий. Възможни са три случая:

1/ Релацията е пълно копие на една физическа таблица.

2/ Релацията е празно копие, състоящо се само от заглавен ред с имена на колони без стойности под тях.

3/ Релацията е частично копие на физическа таблица – визуализирани са някои от колоните на съответната физическа таблица(и) и само няколко от редовете й(им) (в частност можа да са изведени и всички редове).

За опериране с таблиците, колоните и редовете им в релационния модел са дефинирани съответни операции, които се реализират с помощта на заявки от езика SQL.

***Основните понятия на релационния модел на БД са:***

1/ Релация – визуализирана върху екрана на компютъра таблица, която може да бъде, пълно, празно или частично копие на една или повече физически съхранени върху твърдия диск таблици на БД.

2/ Атрибут – колона от релацията, която съответства на колона от физическа таблица и се състои от име и стойности под името.

3/ Наредено множество – ред със стойности от релацията, като наредбата на стойностите зависи от наредбата на атрибутите. По-често се използват понятията ред или запис.

4/ Първичен ключ - стойност на поле от атрибут на релацията със следното свойство: във всеки момент от съществуване на релацията няма повече от един ред с една и съща стойност в този атрибут. Първичният ключ се ползва за търсене, защото в една физическа таблица стойността на първичния ключ може да се появи 0 или 1 път(и).

5/ Област – множество от допустими стойности за един атрибут от релацията. Областта може да бъде крайно множество (съвкупността от курсовете в таблицата СТУДЕНТ) или безкрайно множество (съвкупността от имената на студентите в същата таблица). Областта не се съхранява в компютъра. Тя играе дефинираща роля за БД (данни за областите се храняват заедно с дефиницията на БД) и служи за проверка на коректността на въвеждани и изчислявани стойности за полетата на съответните атрибути.

***Структура и свойства на релациите***

Една релация е визуализирана таблица върху екрана на компютъра. Тя се състои от две части:

1/ Заглавие – фиксирано множество от имена на атрибути.

2/ Тяло – множество от променящи се във времето редове със стойности за полетата на избраните атрибути.

Ще въведем две дефиниции:

1/ Броят на атрибутите в една релация се нарича степен на релацията. С течение на времето степента на релацията не се променя.

2/ Броят на редовете в тялото на релацията се нарича мощност на релацията. С течение на времето мощността на релацията се променя - при заявки за допълване или изтриване на данни от физическите таблици на базата данни.

Ако за една релация са налице следните 4 **свойства**:

1/ няма дублирани редове (по принцип е невъзможно в една физическа таблица да има два еднакви записа поради уникалността на избраната за първичен ключ колона);

2/ няма наредба в атрибутите (наредбата зависи от потребителя);

3/ няма наредба в редовете;

4/ няма съставни стойности в тялото на релацията, т.е. всички стойности са прости, неделими (елементарни, атомарни),

то релацията се нарича нормализирана. По принцип всички физически таблици се подготвят да отговарят на четвъртото свойство и всички релации са нормализирани.

Под **релационна база данни** се разбира множество от променящи се във времето нормализирани релации с фиксирани степени. Променят се мощностите на релациите. Кореспондиращите на атрибутите области никога не трябва да се променят, защото промяната на една област води до нова релационна база данни – например ако числовите оценки от колоната **Uspeh** в примерната таблица се заменят с буквени оценки (каквито са изискванията на кредитната система), то се променя и базата данни, и операциите над нея – примерно не може да се изчислява среден успех.