

DATABÁZOVÉ SYSTÉMY, DATOVÉ MODELOVÁNÍ

Databázový systém = systém řízení báze dat + databáze

Databáze

Databáze je strukturovaná množina dat ve formě databázového souboru. Důležité je slovo strukturovaná, protože databáze může být prázdná, a přesto můžeme, říci, že existuje, pokud byl definován popis tvaru a struktury jejích dat. K těmto datům potom připisují jednotliví uživatelé různé úrovně prostřednictvím aplikací, které jim umožní zadávat data a dotazy pomocí uživatelského rozhraní jako jsou formuláře nebo dialogová okna pro zadávání dat nebo tiskové sestavy pro zobrazení a tisk výsledků dotazů.

Datové typy:

Textový typ – textový řetězec

Číselné typy – celé a reálná čísla s pevnou nebo plovoucí desetinnou tečkou

Logický typ – slouží k uložení logické hodnoty ano/ne

Memo – pro uložení textu proměnné délky

Datumový typ – datum nebo datum a čas

Systém řízení báze dat

SŘBD je program nebo soubor programů, který data organizuje – umožňuje data definovat, ukládat, měnit, mazat, zobrazovat. SŘBD dodávají výrobci programového vybavení podobně jako operační systémy, textové editory nebo programy pro kreslení.

Databázový systém

Databázový systém spojuje vlastní údaje ukládané v databázi s nástroji umožňujícími jejich správu. Databázový systém zahrnuje: technické prostředky, data, programové vybavení, uživatele. Databázové systémy se objevily počátkem 60. let 20. století.

Charakteristika databázové systému:

- Struktury aplikačních programů a vlastních datových souborů musí být oddělené.
- K datům je možno přistupovat pouze prostřednictvím databázové systému.
- Dotazy lze tvořit i v průběhu práce se systémem, nemusí být formulovány předem.
- Je umožněn současný přístup k datům více uživatelům.

Rozdělení databází podle typu:

- Hierarchická databáze
- Síťová databáze
- Objektová databáze
- Relační databáze

1. Vysvětlíte pojmy redundance, konzistence, integrita, transakce.

Redundance

Redundance znamená přetékat či přebývat, v IT problematice znamená informační nadbytek, vícenásobný výskyt dat v systému. Je to situace, kdy je použit jeden prvek více, než je nezbytně nutné. V databázi redundance vzniká duplikováním už jednou nasbíraných dat a informací. Takto vznikají nepřírozené duplicity neboli nepotřebná a často zbytečně rozsáhlá data, zápisy čísel, programů či kódů. Např. z rodné číslo obsahuje datum narození.

Konzistence

Konzistence (integrita) dat odráží skutečný reálný stav. Data odpovídají vlastnostem příslušného popisovaného objektu reálného světa. Např. únor má 28 dní, datum 30. února je vždy porušením integrity, ale datum 29. února je porušením integrity jen někdy. Referenční integrita představuje konzistenci vazeb mezi záznamy tabulek. Při vymazání záznamu z jedné tabulky musí být vymazány odpovídající záznamy z propojených tabulek.

Transakce

Transakce je logický nedělitelný celek, který se buď provede celý, nebo se neprovede vůbec. Od zahájení transakce až do jejího konce se uchovávají původní data a zapisují se veškeré změny v nich. Teprve po dokončení transakce je tento fakt zapsán do systému a původní data mohou být odstraněna. Pokud by v průběhu transakce došlo k výpadku počítače, nezapíše se její ukončení a databázový systém po obnově své funkce obnoví původní data, která má stále uložena.

2. K čemu slouží primární klíč?

Primární klíč

Primární klíč je pole, které má pro každý řádek (záznam) jedinečnou hodnotu, tj. pro žádné dva řádky v tabulce nemůže nastat situace, že by hodnota primárního klíče byla totožná. Většinou je to pomocné pole s číslem řádku (s názvem např. ID), které má tu vlastnost, že při vytvoření nového záznamu se jeho hodnota doplní automaticky. Primární klíč může být v tabulce jeden, má také minimální délku. Primární klíč nesmí obsahovat hodnotu NULL.

Př. Rodné číslo, ID

3. Co je to cizí klíč?

Cizí klíč

Je to atribut, který je v jiné tabulce primárním klíčem. Na rozdíl od primárního klíče, který je pro jednu tabulku vždy jen jeden, cizích klíčů může být více, pokud je tabulka ve vztahu s více tabulkami.

4. Vysvětlete pojem normalizace. Definujte první, druhou a třetí normální formu.

Normalizace, normální formy

Normalizace je proces dekompozice dat na jednotlivé tabulky, jež by měla vést k dobře navrženému datovému modelu. Principy a pravidla jsou nazývány jako tzv. normální formy. Ty se běžně uplatňují na několika úrovních. Vyšší norma vždy staví na normalizaci provedené na nižší úrovni.

1NF

Tabulka je v první normální formě, pokud všechny její atributy jsou atomické tj. dále nedělitelné a obsahují skalární hodnoty.

2NF

Tabulka je v druhé normální formě, pokud splňuje podmínky 1NF a každý její atribut, který není primárním klíčem je na primárním klíči úplně závislý. To znamená, že se nesmí v řádku tabulky objevit položka, která by byla závislá jen na části primárního klíče. Jde o to, definovat v tabulce pouze atributy popisující jednu entitu.

3NF

Relace je v třetí normální formě, pokud je ve druhé normální formě a navíc všechny její neklíčové atributy jsou vzájemně nezávislé, tedy žádný atribut nezávisí na jiném atributu, kromě primárního klíče.

5. K čemu slouží E-R diagram?

Entity-Relationship diagram

E-R model je jedním z nejčastěji využívaných návrhových prostředků.

Slouží k modelování „oblasti zájmu“ (např. podnik) jako kolekcí entit a vztahů (relationships) mezi nimi. Je to metoda datového modelování, která vytváří jeden z typů datových modelů systému. E-R diagramy nezobrazují samostatné entity nebo jednoduché vztahy, zobrazují množiny entit a množiny vztahů.

V E-R diagramech se používají symboly představující základní typy informací: boxy (obdélníky) reprezentují entity, atributy jsou reprezentované elipsami, které jsou spojené s entitou, nebo je název atributu přímo vepsaný do značky pro entitu. Množiny vztahů jsou reprezentovány kosočtverci.

Entita

je specifickým předmětem zájmu z oblasti reálného světa, která danou oblast popisuje. Jako příklad uvedeme účtový systém, kde jsou nevyhnutné dvě věci: účty a transakce. Každá taková jednotka základních informací se nazývá 'entita', která má určitý název a obsahuje atributy (vlastnosti entity).

Atribut

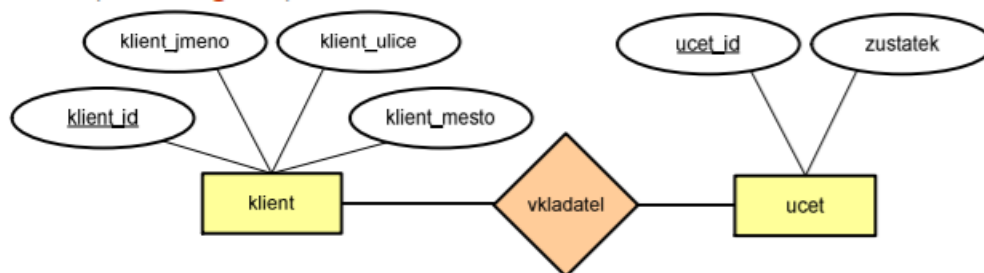
Každá entita obsahuje jeden nebo víc atributů. Atributy si můžeme představit jako malé kusy informací v rámci entity. Spolu popisují entitu do takové míry, jak je to nevyhnutné.

Příkladem může být entita zákazník:

- Entita: zákazník
- Atributy: zákaznické číslo, křestní jméno

Relace, vztah

Vztah popisuje, jakým způsobem jsou dvě nebo více entit vztaženy mezi sebou. V relační databázi mají všechny subjekty mezi sebou určitou vazbu, která je vyjádřena jako vztah. Relace je vazba mezi subjekty.



6. K modelování jakého typu vztahu používáme vazebnou tabulku?

Vazebnou tabulku používáme k modelování vztahu M:N.

Tento vztah umožňuje každému záznamu z jedné tabulky přiřadit libovolný počet záznamů z druhé tabulky, přičemž záznam v druhé tabulce přiřazením k záznamu k první tabulce svou možnost přiřazení „nespotřebuje“, takže jej lze přiřadit k libovolnému počtu záznamů první tabulky.

Příkladem z reálného života by mohl být vztah výrobku a jeho vlastnosti – výrobek může mít více vlastností a jednu vlastnost může mít více výrobků. Dalším příkladem je vztah herce a filmu – jeden herec může hrát ve více filmech a v jednom filmu může hrát více herců.

7. Co to je a k čemu slouží databázový index?

Co to je databázový index

- nazývaný i databázová konstrukce

Slouží ke zrychlení vyhledávacích a dotazovacích procesů v databázi. Databázový index je obvykle definován výběrem tabulky a jednoho konkrétního sloupce (nebo více sloupců), nad kterými si analytik nebo designér databáze přeje dotazování urychlit. Dále je definován technickým určením datového typu a typu indexu. Chování a způsoby uložení indexů se mohou významně i výrazně lišit podle použité databázové technologie.

K čemu indexy slouží

Slouží ke zrychlení přístupu k datům, měly by se používat u všech sloupců, podle kterých se vyhledává, třídí nebo podle kterých se spojují tabulky.

Když chceme data z tabulky později vybrat podle nějakého kritéria, poslouží k tomu, aby při výběru několika záznamů nebylo potřeba procházet všechny ostatní, protože právě v indexech jsou data organizována tak, aby bylo možné rychle vybrat pouze relevantní záznamy.

8. Co to je Business Intelligence? Jaké nástroje používá?

BI je sada procesů, aplikací a technologií, jejichž cílem je účinně a účelně podporovat rozhodovací procesy ve firmě. Podporují analytické a plánovací činnosti podniků a organizací a jsou postaveny na specifických tzv. OLAP (On-line Analytical Processing) technologiích a jejich modifikacích, tedy na principech multidimenzionálních pohledů na podniková data.

Základním rysem práce s daty v aplikacích ERP je využití relačních databázových systémů. **Relační databáze** jsou vhodné pro systémy na operativní a taktické úrovni. Z hlediska analytických a plánovacích činností manažerů však mají **jistá omezení**:

- Neumožňují rychle a pružně měnit analytická kritéria
- V obrovských objemech dat se obtížně provádějí agregace dat na nejrůznějších úrovních
- Analytické úlohy nadměrně zatěžují a zpomalují klasické zpracování ERP
- Firmy jsou zahlceny redundantními a nekonzistentními daty, které v procesu rozhodování není možné využít

9. Co to je datový sklad (data warehouse)? V čem se odlišuje od klasické relační databáze?

Datový sklad (anglicky Data Warehouse, případně DWH) je zvláštní typ relační databáze, která umožňuje řešit úlohy zaměřené převážně na analytické dotazování nad rozsáhlými soubory dat. K definici rozdílu mezi „běžnou“ relační databází a datovým skladem se obvykle používá následujících charakteristik popsaných Williamem Inmonem:

Orientace na subjekt

U běžné relační databáze je obvyklá snaha o co nejmenší redundanci uložení dat, které je dosahováno jejich normalizací do třetí normální formy a vnitřním provázáním jednotlivých logických funkčních celků. V datovém skladu je naproti tomu řešení vždy vedeno snahou o jasnou vnitřní separaci jednotlivých funkčních celků – výsledkem je struktura, která je čitelnější pro uživatele (manažera, business analytika) za cenu zvýšených nároků na paměťový prostor.

Integrovanost

Běžná provozní aplikace nad relační databází řeší určitý specifický okruh úloh nad „svými“ specifickými daty. V datovém skladu je třeba naproti tomu shromáždit informace z mnoha různých zdrojů a seskupit je nikoliv podle původu, ale podle logického významu. Data jsou do datového skladu obvykle nahrávána ve větších dávkách (například v denních nebo týdenních intervalech) a pak již nejsou nijak modifikována.

Historizace

Data jsou v datovém skladu obvykle udržována v historické podobě, nikoliv pouze v aktuálním stavu. To je dáno nutností provádění analýz zaměřených na vývoj v čase. V běžné relační databázi je z pohledu uživatelů obvykle zajímavý pouze aktuální stav datových objektů.

10. Definujte pojmy OLAP a OLTP. Uveďte základní rozdíly mezi OLAP a OLTP.

OLAP (Online Analytical Processing) je technologie uložení dat v databázi, která umožňuje uspořádat velké objemy dat tak, aby byla data přístupná a srozumitelná uživatelům zabývajícím se analýzou obchodních trendů a výsledků (Business Intelligence). Způsob uložení dat se svým zaměřením liší od běžněji užívaného OLTP (Online Transaction Processing), kde je důraz kladen především na snadné a bezpečné ukládání změn v datech v konkurenčním (víceuživatelském) prostředí.

Základní rozdíly mezi OLAP a OLTP vyplývají z rozdílného použití – u OLAP se jedná o jednorázově nahrávaná data, nad kterými jsou prováděny složité dotazy, u OLTP jsou data průběžně a často modifikována a přidávána, a to obvykle mnoha uživateli zároveň:

- OLAP nepoužívá na rozdíl od OLTP normalizované uložení dat v 3NF formě – data jsou v uložena tak, aby umožňovala rychlou realizaci složitých dotazů, časté je zdvojené (redundantní) uložení, které by v případě OLTP komplikovalo provádění změn v datech,
- OLAP používá podstatně více indexů než OLTP – opět to souvisí se zaměřením, kde indexy umožňují rychlé provedení složitých dotazů,
- OLAP na rozdíl od OLTP často používá předpočítané agregované a odvozené hodnoty.

Obě tyto technologie se však v systémech podnikové inteligence využívají.

Zdroje:

<http://vychodil.inf.upol.cz/kmi/data1/data1pr01.pdf>

<http://www.ststeplice.cz/download/studenti/studijni%20materialy/ikt-vyukoveprezentace/15%20-%20databaze.pdf>

<https://phoenix.inf.upol.cz/esf/ucebni/databa.pdf>

<http://books.fs.vsb.cz/dbacc20/dbacc01.htm>

<http://bech.mzf.cz/accessok.pdf>

http://www.ped.muni.cz/wtech/03_studium/cvt4/databaze.pdf

http://www.fce.vutbr.cz/aiu/vojkuvka.m/bu006/BU04_M01.pdf

<http://www.sprava-site.eu/redundance/>

<https://sites.google.com/site/xvinformatika/databaze/databazove-systemy>