

Datové modely. Konceptuální datový model, Databázové modely, Fyzický pohled na data. Relační model, Relace, Relační algebra, Normální formy.

Model dat

- integrovaná kolekce konceptů pro popis dat, relací mezi daty a omezení používaných dat,
- snaží se reprezentovat datové požadavky organizace, kterou má modelovat, účelem je reprezentovat data srozumitelně,
- poskytuje základní koncepty a notaci umožňující návrhářům a uživatelům komunikovat o organizaci dat organizace,
- má tři složky:
 - strukturální část - množina pravidel pro konstrukci databáze,
 - manipulační část - typy na datech přípustných operací, transakcí (aktualizace dat, struktura databáze),
 - množina pravidel integrity - zajišťuje přesnost dat,
- možné tři úrovně abstrakce: konceptuální, databázová, fyzická.

Databázový (logický) model

- aparát, způsob vytváření struktur (pojmy k modelování) a způsob manipulace s nimi,
- výsledkem databázového modelování je databázové schéma (popis struktury dat),
- střední úroveň abstrakce (entity strukturami logického modelu).

Síťový databázový model

- vazby mezi záznamy souborů,
- grafické schéma - síť typů záznamů - orientovaný graf,
- záznam může mít několik předků i potomků (svazové uspořádání).

Hierarchický databázový model

- grafické schéma tvoří strom (jeden předek, více potomků),
- záznam nazýván segmentem,
- odkazy lze vytvářet pouze od předků směrem k potomkům (nelze zapojovat jeden typ záznamu do více hierarchií).

Relační databázový model

- data modelována tabulkami/relacemi (řádky=entity, sloupce=atributy entit),
- jednoduché dotazování deklarativním SQL,
- normalizace schémat (bez redundance a aktualizčních anomálií),
- jednoduchá implementace, vysoký výkon.

Objektově databázový model

- data modelována třídami (zapouzdřené datové entity objektu),
- přímé asociace (vztahy) mezi objekty (nativní modelování grafů objektů),
- nevýhody - složitá implementace, nevykonné.

Objektově-relační model

- relační databáze obohacená o objektové prvky,
- relace (tabulky), objektové datové typy (atributem třída).

Konceptuální model

- nejvyšší úroveň abstrakce vytvoření popisu dat v databázi,

- datové entity modelovány jako objekty reálného světa (bez ohledu na implementaci, architekturu),
- modelování dat přítomných v systému, pohled uživatele na danou část reálného světa,
- principy konceptuálních modelů:
 - orientace na objekty - myslí se v objektech,
 - funkcionální podstata vztahů,
 - ISA hierarchie (nadtypy, podtypy),
 - hierarchický mechanismus konstrukce jiných objektů z objektů (agregace, skládání),
- nejznámější E-R konceptuální model = množina pojmů na konceptuální úrovni popisující aplikaci za účelem specifikace struktury databáze:
 - entity = objekty reálného světa (schopné samostatné existence),
 - vztahy, relace = vazby mezi entitami,
 - atributy = blíže popisují entity nebo vztahy,
 - integritní omezení = soulad schématu s realitou.

Fyzický pohled na data

- fyzický model = nízká úroveň abstrakce, implementace logického modelu ve specifických technických podmínkách,
- optimalizace výkonu SŘBD a uložení dat pro manipulaci rychlou, bezpečnou, škálovatelnou,
- management datových souborů, indexování, transakční zpracování.

Relační model (neformálně)

- model uchování entit/vztahů v tabulkách (max. 1 tabulka na entitní/vztahový typ),
- datová entita/vztah E je reprezentována řádkem tabulky,
- atribut A entity je reprezentován sloupcem tabulky,
- buňka tabulky (x,y) uchovává hodnotu atributu A(x) v kontextu entity/vztahu E(y),
- schéma tabulky = popis struktury tabulky (vše, kromě dat) $S(A1:T1, A2:T2, \dots)$, kde S je název tabulky, A_i atribut a T_i datový typ atributu,
- schéma relační databáze = množina schémat tabulek + integritní omezení + další věci,
- základní integritní omezení:
 - neexistují dva stejné řádky (unikátní identifikace entit)
 - rozsah hodnot atributu dán typem
 - všechny hodnoty tabulky jsou defionvané (možno NULL)
- každý řádek tabulky identifikovatelný - nadklíč (skupina atributů), klíč (minimální počet atributů),
- speciální "mezitabulkové" integritní omezení - cizí klíč (klíčem v jiné tab.), zachycuje relaci mezi tabulkami,
- přínosy relačního datového modelu:
 - oddělení dat od implementace - při manipulaci s daty nejsou zajímavé

přístupové mechanismy

- prostředky pro dotazování - relační algebra, relační kalkul
- vhodný návrh schémat pomocí normalizací

Relace

- množina prvků daná schématem relace $R(A1:D1, A2:D2, \dots, An:Dn)$, A_i - atributy, D_i atributové domény,
- jsou to konkrétní data v tabulce.
Prvek relace = konkrétní řádek v tabulce
Schéma relace = schéma tabulky
Domény relace = typy atributů

Relační algebra

- matematický formalismus použitelný na relacích pro dotazování,
- množina operací na relacích se schématy, s relací jako výsledkem:

- přejmenování atributů (unární operace)
- množinové operace (binární) - sjednocení, průnik, rozdíl, kartézský součin
- projekce (unární) = "odstranění" části atributů
- selekce (unární) = "výběr" prvků z relace dle podmínky
- přirozené spojení (binární) = "spojení" prvků přes stejné hodnoty
- vnitřní spojení (binární) = spojení přes předpoklad kladený na atributy
- dělení relací (binární)
- vnější spojení = doplnění NULL k prvkům nespojitelným vnitřním spojením,
- logické pořadí vyhodnocení operací v dotazu, též prioritizace operací,
- ekvivalentní výrazy = tentýž dotaz různými výrazy,
- relační úplnost = dotazovací jazyk vyjadřující všechny konstrukce relační algebry,
- vlastnosti:
 - deklarativní dotazovací jazyk
 - výsledkem vždy konečná relace - bezpečné operace
 - vlastnosti operací - asociativita, konektivita.

Normální formy

- problémy návrhu relačního schématu:
 - redundance = stejná informace uložena vícekrát
 - aktualizací anomálie - změna všech kopií, vložení a mazání i příslušných
přídavných entit,
- řešení - normalizace.

1. normální forma

- každý atribut schématu relace je elementárního typu a je nestrukturovaný (tabulka je dvourozměrné pole, ne strom nebo graf). Laicky: tabulka, v níž každý průsečík sloupce a záznamu (řádku) obsahuje jen jednu hodnotu.

2. normální forma

- neexistují částečné závislosti neklíčových atributů na (libovolném) klíči. Laicky: tabulka v 1.NF, ve které jsou všechny hodnoty sloupců netvořících součást primárního klíče určeny všemi hodnotami sloupců primárního klíče (plná funkční závislost na PK).

3. normální forma

- žádný neklíčový atribut není tranzitivně závislý na žádném klíči. Laicky: tabulka v 1. a 2. NF, ve které všechny hodnoty ve sloupcích nepatřících primárnímu klíči jsou určeny pouze sloupci primárního klíče a žádnými (!) jinými sloupci.

Boyce-Coddova normální forma

- každý atribut je netranzitivně závislý na klíči.

Další normální formy: 4.NF - multizávislosti, 5.NF