

Návrh schématu DB

Robert Haken [MVP ASP.NET/IIS, MCT]
Software Architect, HAVIT, s.r.o.

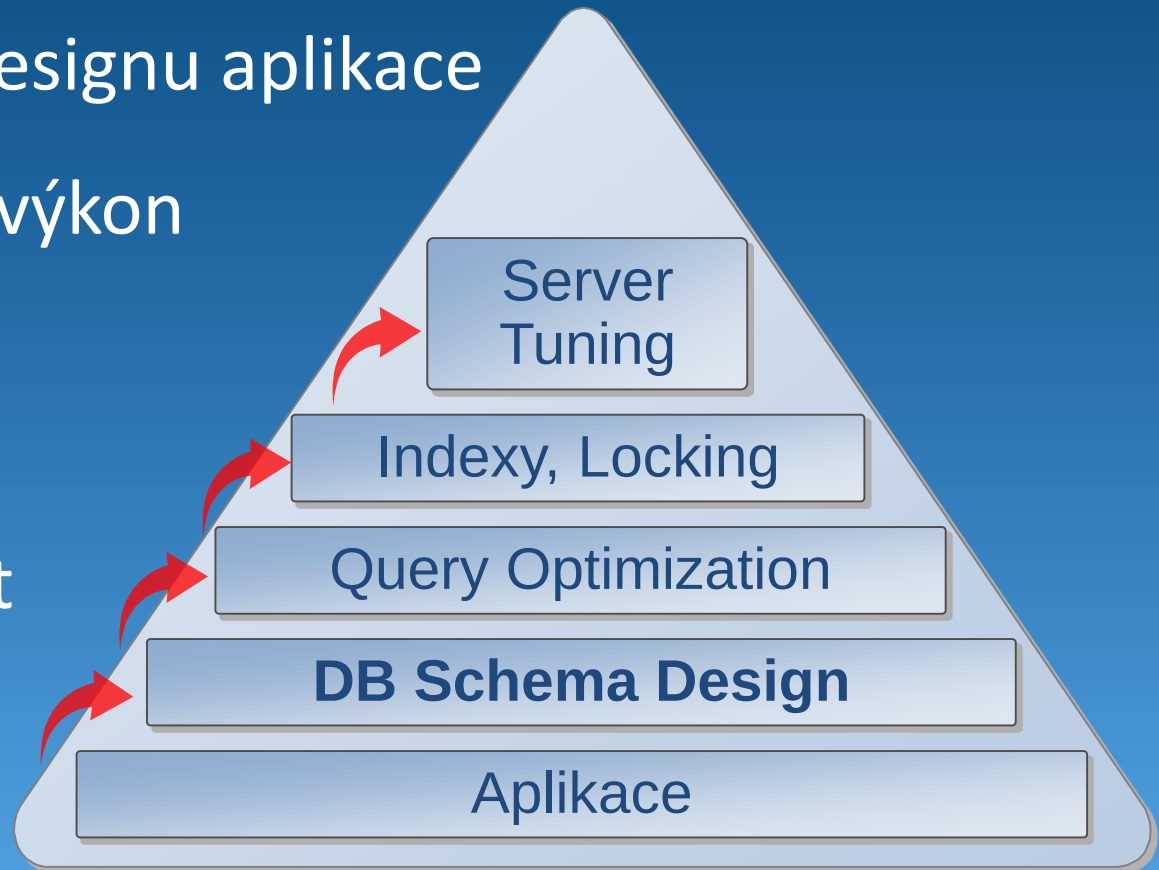
TECHNICAL EDUCATION & DEVELOPER DATABASE CONFERENCE 2011

Tech·Ed
DevCon 
Praha 2011

Proč je schéma DB důležité?

- ▶ klíčový prvek designu aplikace
- ▶ zásadní vliv na výkon
 - ▶ rychlost
 - ▶ objem dat
- ▶ konzistence dat





DAQUAS

Microsoft

DEMO

Ukázka cílové podoby DB schématu

Hlavní zásady tvorby schématu

- Konvence a standardy
- Přirozená míra normalizace
- Respektujte scénáře použití
- Vlastní návrhové vzory
- Standardní návrhové vzory
- Maximálně self-documented + přehlednost

Konvence – Naming Conventions

- ▶ název tabulky jednotným číslem
- ▶ výstižná podstatná a přídavná jména
- ▶ PascalCase
- ▶ bez národních, speciálních znaků, mezer, atp.
- ▶ raději nepřekrývat s keywords (Order, Text, ...)
- ▶ zkratky lowercase (např. SazbaDph, HtmlText)
- ▶ jednotná terminologie

Naming conventions II.

- ▶ suffix ID pro všechny PK i FK sloupce (ZakaznikID)
- ▶ PK vždy v podobě *NazevTabulkyID*
- ▶ FK kopírujeme PK, pokud není potřeba rozlišit
 - ▶ InvoiceItem.InvoiceID
 - ▶ Invoice.DeliveryAddressID, Employee.BossID
- ▶ Varianta konvencí – trigram prefixy
 - ▶ tabulka Invoice, sloupce invID, invNumber,...



DAQUAS

Microsoft

DEMO

Naming Conventions, Trigramy

Standardy – Výchozí schéma, PK

- ▶ každá tabulka jednoduchý umělý PK typu `int`
 - ▶ s výjimkou čistě vazebních tabulek M:N (PK = 2xFK)
- ▶ PK – umělý jednoduchý typu `int`, `clustered`
 - ▶ PRO: umělý = nemění se
 - ▶ PRO: velmi rychlé SQL operace (JOIN atp.)
 - ▶ PROTI: nejednoznačný v distrib. systémech => GUID
- ▶ PK – auto-increment (pokud není tabulka R/O)
 - ▶ PRO: monotónní růst = vhodný pro clustered index
 - ▶ PRO: zajišťuje unikátnost klíčů pro nové záznamy

Přirozená míra normalizace

- ▶ přiměřená atomicita atributů
 - ▶ Employee.Name
 - ▶ Employee.FirstName, Employee.Surname
 - ▶ vždy záleží na nejzazším dohledném scénáři + 100%
- ▶ NE vícehodnotové sloupce (...nebo ano?)

PracovníkID	Jmeno	TelefonniCisla
156	Josef Novák	608 123 456, 602 123 456
157	Marta Nová	601 123 456

Přirozená míra normalizace II.

● NIKDY vícenásobné sloupce

PracovníkID	Jmeno	Telefon1	Telefon2
156	Josef Novák	608 123 456	602 123 456
157	Marta Nová	601 123 456	NULL

● správná závislost na klíči (PK) je klíčová

PracovníkID	Jmeno	Oddeleni	ManagerOddeleniID
156	Josef Novák	IT	158 (<i>Karel Malý</i>)
157	Marta Nová	IT	158 (<i>Karel Malý</i>)
180	Helena Malá	HR	180 (<i>Jan Bílý</i>)

● každá duplicita dat je velmi podezřelá !!!

Denormalizace

- ▶ denormalizace pouze, pokud požadovaného výkonu v klíčových scénářích nejsme schopni dosáhnout normalizovaným schématem !!!
- ▶ vždy něco ztrácíme, abychom jinde získali
- ▶ obtížnější údržba (nejlépe automatizace)
- ▶ potenciální nekonzistence

Standardy - další doporučení

- ▶ procenta – ukládat jako koeficient 100% => 1.00
- ▶ stabilizace významu a používání pojmů
 - ▶ např. marže / přírážka / rabat / sleva / koeficient / ...
 - ▶ User / Login / LoginAccount / Uživatel / ...
- ▶ používejte Database Diagrams
- ▶ Dokumentujte! – xp attribute MS_Description
 - ▶ DB>doc generátor HTML/XML/WikiPlex



DAQUAS

Microsoft

DEMO

Standardní schéma, Dokumentace,
Denormalizace

Schema Killers

▶ hodnoty v názvech sloupců

PracovníkID	Jmeno	Prodej2010	Prodej2011
156	Josef Novák	123 456,38	348 260,30
157	Marta Nová	2 650 456,42	123 128,40

▶ chybné datové typy

int EmployeeID	varchar(20) DatumNastupu	bigint BossID	char(1) Aktivni
156	20.3.2011	157	A
157	17.6.1989	NULL	N

Schema Killers II.

- ▶ nedefinovaný PK tabulky
- ▶ chybějící reference mezi tabulkami (FK)
- ▶ vazba na jiný sloupec, než PK

DirectoryEntryID	EntryCode	ParentEntryCode
123	{123-XYZ-....-ABC-123}	{124-XYZ-....-ABC-124}
234	{124-XYZ-....-ABC-124}	<i>NULL</i>

Bad Practices

- ▶ magic numbers = chybějící enum-tabulky
 - ▶ PROTI: chybí validace hodnot, konzistence
- ▶ podřizování schématu lidské čitelnosti dat
 - ▶ přirozené PK/FK (Username, číslo faktury, ...)
 - ▶ PROTI: nejsou stabilní
 - ▶ PROTI: nejsou výkonné pro typické operace (JOIN atp.)
 - ▶ PROTI: nejsou obv. monotónně rostoucí (CLUSTERED)
 - ▶ chybějící číselníky, hodnoty přímo místo FK

Bad Practices II.

➤ hromadné číselníky (popř. složený PK)

CiselnikItemID	CiselnikID	Value
1	1	aktivní
2	1	neaktivní
3	2	hnědá
4	2	červená
5	2	modrá

- PROTI: obtížná rozšiřitelnost o další vlastnosti
- PROTI: komplikované FK odkazy, validace konzistence
- komplikovanější indexování (složené/filtrované)
- PRO: unifikované použití/editace z aplikace



DAQUAS

Microsoft

DEMO

Schema Killers, Bad Practices

Patterns – Dědičnost objektů

- ▶ Fowler: Patterns of EAA
 - ▶ Class Table Inheritance
 - ▶ tabulka pro každou třídu i abstraktní předky
 - ▶ pouze atributy definované třídou (nutný JOIN)
 - ▶ Single Table Inheritance
 - ▶ jedna tabulka pro všechny možné potomky
 - ▶ Concrete Table Inheritance
 - ▶ tabulka pro každou neabstraktní třídu
 - ▶ sloupce všech předků + třídy

Patterns – Dynamické datové schéma

- Atributová databáze
 - uložení všech objektů v cca 5 tabulkách
 - PRO: snadná dynamická změna schématu za provozu
 - PRO: relativně slušný výkon
 - PRO: snadná implementace dědičnosti / interface
 - PROTI: komplikované validace FK
 - PROTI: komplikovanější dotazování, výkonové ztráty
- => atributová rozšíření statického schématu



DAQUAS

Microsoft

DEMO

Design Patterns

Nebud'me jen on-line, potkejme se na TechEdu!

Tech·Ed
DevCon 
Praha 2011



 **GOPAS[®]**

 **DAQUAS**

Microsoft[®]

