

Normální tabulky, entity a relace

Tentokráté nás čeká dokončení problematiky klíčů a vyšších normálních forem a na závěr si řekneme něco málo o entitách a relacích.

Jednoduchý klíč je sloupec tabulky, který obsahuje vzájemně různé, neprázdné hodnoty. V tabulce č. 1 bohužel ani sloupec TRPASLIK, sloupec ROKNAR, ani sloupec VLASTNOST nemohou být samy o sobě klíčem. ROKNAR není vždy znám, a navíc se opakuje hodnota 1830. SMUDLA se vyskytuje třikrát v prvním sloupci a ve třetím sloupci se několikrát opakuje vlastnost HODNY. V tomto slova smyslu je tabulka č. 1 zdánlivě

TRPASLIK	ROKNAR	VLASTNOST
SMUDLA	1830	SPINAVY
SMUDLA	1830	HODNY
SMUDLA	1830	ZAVISLY
PROFA	1807	UCENY
BRUMLA		ZADUMCIVY
KEJCHAL	1857	HODNY

Tab. č. 1: Začínáme přemýšlet.

nedokonalá. Tuto nedokonalost odstraníme definicí složeného klíče:

Složený klíč je skupina sloupců tabulky, které obsahují vzájemně různé kombinace neprázdných hodnot.

Obr. č. 1. Jednoduchý klíč.

V tabulce č. 1 je složený klíč tvořený prvním a třetím sloupcem, tj. sloupci TRPASLIK a VLASTNOST. Klíč obsahuje dvojice hodnot, které se neopakují. SMUDLA sám o sobě je uveden třikrát a slovo HODNY dvakrát, ale kombinace SMUDLA-HODNY se objevuje právě jednou. Podobně tomu je s ostatními dvojicemi SMUDLA-SPINAVY, SMUDLA-

-ZAVISLY, PROFA-UCENY, BRUMLA-ZADUMCIVY, KEJCHAL-HODNY. Na obrázku č. 1 vidíme graficky znázorněnou situaci. Modrá barva je použita k vyznačení všech různých hodnot klíče v prvním sloupci tabulky a v prvním sloupci záhlaví je uvedeno slovo KEY pro znázornění, že jde o sloupec, který tvoří jednoduchý klíč.

Obr. č. 2. Složený klíč.

Ve druhém a třetím sloupci na obrázku č. 1 jsou nepovinně uloženy neklíčové hodnoty, které jsou na klíči závislé. Známe-li hodnotu klíče, například křestní jméno trpaslíka, pak známe i jeho další vlastnosti. To je velká výhoda jednoduchého a složeného klíče. Na obrázku č. 2 vidíme příklad tabulky, jejíž první dva sloupce tvoří složený klíč.

Oba klíčové sloupce jsou vyznačeny modrou barvou. Ve třetím sloupci zbyly nepovinné neklíčové hodnoty. Někdy může dojít k situaci, kdy všechny sloupce tabulky jsou

Obr. č. 3. Jenom klíč?

zahrnuty do složeného klíče, jak je to uvedeno na obrázku č. 3.

Obr. č. 4. Částečná závislost.

Takovéto tabulky se běžně v databázových systémech vyskytují a mají charakter spojek mezi jinými tabulkami. Má-li tabulka jednoduchý nebo složený klíč, pak rozlišujeme její sloupce podle toho, zda jsou nebo nejsou součástí tohoto klíče. Je-li sloupec součástí klíče, hovoříme o klíčovém sloupci. Není-li sloupec součástí klíče, hovoříme o neklíčovém sloupci. Při převodu tabulky do vyšších normálních forem se chtít nechtět budeme muset zabývat jednotlivými klíčovými a neklíčovými sloupci tabulky a jejich nebezpečnými závislostmi. Například v tabulce č. 1 je vztah mezi sloupci TRPASLIK a ROKNAR velmi nebezpečný.

Například v tabulce č. 1 je vztah mezi sloupci TRPASLIK a ROKNAR velmi nebezpečný.

Nebezpečné závislosti

Samotná vzájemná závislost sloupců v tabulce není záporným jevem. Uvažujme tabulku s osobními údaji lidí. Pak klíčem by mohlo být rodné číslo, které můžeme akceptovat jako neopakovatelnou charakteristiku jednotlivých osob žijících v konkrétním státě. Jestliže známe rodné číslo, pak jsem schopen najít jednoznačně řádek tabulky a z něj zjistit křestní jméno. Tedy křestní jméno je funkcí rodného čísla. Obdobně je to s trvalým bydlištěm, průměrnou

Obr. č. 5. Neklíčová závislost.

nou roční mzdou a datem narození. Všechny sloupce jsou závislé na klíči. Existují však jiné zálučné a nebezpečné závislosti. Pokud je klíč složen z několika sloupců, může se klidně stát, že některé neklíčové sloupce jsou funkcí nikoli celého klíče, ale pouze jeho části, například jen jednoho z klíčových sloupců. Krásnou ukázkou je tabulka č. 1. Zde je klíč tvořen dvěma sloupci. Potíž je v tom, že ROKNAR, jako druhý a neklíčový sloupec obsahující rok narození trpaslíka, není funkcí celého dvojdi-

ného klíče, ale pouze jeho části, prvního sloupce nazvaného TRPASLIK. V tom je úskalí tabulky č. 1. Na obrázku č. 4 je schematicky znázorněna tato částečná závislost.

Jsou použity dva odstíny modré barvy. První dva sloupce tvoří složený klíč a třetí sloupec obsahuje hodnoty, které nejsou závislé na celém klíči, ale pouze na jeho části. Tato část klíče, která předurčuje svou hodnotou hodnotu neklíčové položky ve třetím sloupci, je odlišena tmavou barvou. Tento jev je velmi negativní, ale zároveň pozitivně alarmující. Nenápadně jsme upozorněni na to, abychom závislost na části klíče zcela osamostatnili do jiné tabulky. Z tabulky na obrázku č. 4 vytvoříme tabulky dvě, z nichž první musí obsahovat první a třetí sloupec. Druhá tabulka musí obsahovat pouze první a druhý sloupec. Hrozí nám ještě další nebezpečí. Pokud mimo klíč leží více než jeden sloupec, mohlo by se stát, že skrytá závislost bude mezi dvěma neklíčovými sloupci. To by ovšem znamenalo, v souladu s obrázkem č. 5, že máme provést dokořenější rozbor problému.

Zde je první sloupec tvořen jednoduchým klíčem, zatímco druhý a třetí sloupec obsahuje neklíčové položky. Tmavě zelenou barvou je naznačen neklíčový sloupec, jehož hodnota jednoznačně určuje hodnotu ve třetím sloupci. Tedy třetí sloupec je funkcí druhého a ani jeden z nich není součástí klíče. Hovoříme potom o neklíčové závislosti. Tabulka z obrázku č. 5 je snadno dekomponovatelná na dvě tabulky, z nichž jedna obsahuje první a druhý sloupec, zatímco druhá obsahuje první a třetí. První tabulka má už jenom jeden klíčový a jeden neklíčový sloupec a těžko můžeme hovořit o neklíčové závislosti. Druhá tabulka má stejnou vlastnost. Nyní jsme už zvládli nebezpečné závislosti v tabulkách a je na místě definovat vyšší normální formy tabulek.

Druhá a třetí normální forma

Platí následující definice:

Tabulka je v **druhé normální formě 2NF**, jestliže je v první normální formě, zároveň existuje klíč a současně všechna neklíčová pole jsou funkcí celého klíče, a nikoli jen jeho části.

Definice tabulky ve třetí normální formě jde ještě o krok dál:

Tabulka je v **třetí normální formě 3NF**, jestliže ve druhé normální formě a zároveň neexistují závislosti neklíčových sloupců tabulky.

Tolik poněkud chladná a suchá řeč definic. Nyní se zabýváme jednoduchými případy. Tabulka z obrázku č. 1 obsahuje jednoduchý klíč. Jestliže tabulka obsahuje jednoduchý klíč, pak je zřejmé, že neklíčová pole jsou funkcí celého klíče, a tudíž je nejen v 1NF, ale rovnou v 2NF. To, zda je ve 3NF formě, není z obrázku č. 1 poznatelné. Tabulka na obrázku č. 2 má už složený klíč a jedno neklíčové pole. Proto pokud se prokáže, že neklíčové pole je funkcí celého klíče, je tabulka nejen v 2NF, ale i ve 3NF, protože v případě jednoho neklíčového pole nemůžeme hovořit

o vzájemné závislosti neklíčových polí. Na obrázku č. 3 je tabulka, která je automaticky nejen v 1NF, ale hned i v 2NF a 3NF, protože nemá vůbec žádné neklíčové sloupce. Tabulka na obrázku č. 4 je pouze v 1NF, protože první sloupec coby část klíče předurčuje hod-

TRPASLIK	VLASTNOST
SMUDLA	SPINAVY
SMUDLA	HODNY
SMUDLA	ZAVISLY
PROFA	UCENY
BRUMLA	ZADUMCIVY
KEJCHAL	HODNY

Tab. č. 2. Triumf.

noty neklíčového třetího sloupce. Na obrázku č. 5 je tabulka ve 2NF, protože existuje klíč a je jednoduchý, a tím pádem veškeré neklíčové položky jsou funkcí celého klíče, a navíc mezi druhým a třetím sloupcem je prokázána

TRPASLIK	ROKNAR
SMUDLA	1830
PROFA	1807
BRUMLA	
KEJCHAL	1857

Tab. č. 3. Co zbylo?

skrytá neklíčová závislost. Pro zajímavost uvedu možné definice tabulek ve čtvrté a páté normální formě:

Tabulka je ve **čtvrté normální formě 4NF**, je-li ve třetí normální formě a popisuje pouze jednu příčinnou souvislost. Někdy hovoříme o tom, že popisuje jeden fakt.

Tabulka je v **páté normální formě 5NF**, pokud je ve čtvrté normální formě a není možné do ní přidat nový sloupec nebo skupinu sloupců tak, aby se vlivem skrytých závislostí rozpadla na několik dílčích tabulek.

U tabulek nestačí jen určovat, v jaké jsou normální formě. Je třeba usilovat o transformaci tabulky do 3NF až 5NF. Toho docílíme velmi snadno. Pokud se tabulka dostala z 0NF do 1NF namnožením řádků tak, jak to udělala Sněhurka v tabulce č. 1, potom se z 1NF do vyšších forem tabulek dostaneme tak, že si stanovíme unikátní klíč a nepříjemné závislosti osamostatníme do jiných tabulek. Takovou automatickou dekompozici dost čas-

TRPASLIK	DEN	SACHTA	RUDA	PLAN	SKUTECNOST
BRUMLA	1.1.98	5	SMOLINEC	3000	4000
PROFA	1.1.98	3	UHLI	3000	2000
KEJCHAL	1.1.98	8	UHLI	3500	5000
BRUMLA	2.1.98	5	SMOLINEC	4000	1000
PROFA	2.1.98	7	ZLATO	2	0
KEJCHAL	2.1.98	5	SMOLINEC	3000	6000

Tab. č. 4. A do práce.

to a zcela přirozeně vzniknou číselníky. Pokračujme v analýze Sněhurčiny úlohy.

Polepšení

Rozložením tabulky č. 1 vznikly dvě tabulky: č. 2 a 3.

Tabulka č. 2 s názvem TRV obsahuje pouze informace o vlastnostech jednotlivých trpaslíků. Je tvořena dvěma sloupci TRPASLIK a VLASTNOST, a je dokonce v 5NF. Co ovšem potom zbyde z původní tabulky č. 1? Nezbyvá než v tabulce č. 3 shromáždit zbylé rozumné souvislosti. Tak vznikla tabulka s názvem TRNR, ve které jsou dva sloupce TRPASLIK a ROKNAR. Vzhledem k tomu, že rok narození není vždy snadno zjistitelný, stačí uvést do klíče pouze první sloupec s názvem TRPASLIK. Klíč je unikátní a tabulka č. 3 má méně řádků než tabulka č. 2, protože vlastně popisuje trpaslíky jako takové a jejich význačné unikátní vlastnosti. Neklíčový sloupec je také jeden, popisuje jenom jednu souvislost, a proto vzhledem k nemožnosti různých skrytých závislostí je také v 5NF a těžko lze ještě do ní vnést další jasno. S tabulkami č. 2 a 3 můžeme být spokojeni, protože nám umožní snadnou aktualizaci a vyhledávání. Pod vlivem metod, které Sněhurka aplikovala na povahové vlastnosti trpaslíků, se rozhodla trpaslíky více pracovně zapojit. Proto začala uvažovat o evidenci jejich pracovních povinností. Trpaslíci se převážně zabývali dolováním. Rozhodla se, že jim bude pravidelně přidělovat denní úkoly v těžbě. V tabulce č. 4 je uvedena první představa.

První představa může být jakákoliv, proto je tabulka č. 4 doslova odfláknutá. Je zde několik sloupců. Sloupec TRPASLIK zachycuje jméno trpaslíka, kterému vznikla pracovní povinnost, sloupec DEN znázorňuje den, kdy má jít do práce, sloupec SACHTA obsahuje číslo šachty, ve které má ten den dolovat, sloupec RUDA označuje, jakou rudu má v dané šachtě těžít, sloupec PLAN říká, kolik kilogramů rudy má za den nakutat, sloupec SKUTECNOST potom slouží k objektivní evidenci, kolik této rudy nakutal. Pokud bychom u tabulky č. 4 hledali nějaké známky normality, zjistíme, že všechna pole obsahují právě jednu hodnotu, a tím pádem tabulka není v 0NF. Po chvíli zjistíme, že sloupec TRPASLIK se nehodí jako klíč a že je třeba sloupec TRPASLIK spojit se sloupcem DEN, a tak vytvořit složený klíč. Při pozorném pohledu zjistíme, že tabulka č. 4 je v 1NF, a dokonce

SACHTA	RUDA
3	UHLI
5	SMOLINEC
7	ZLATO
8	UHLI

Tab. č. 5. Číselník šachet.

v 2NF, protože zbylá neklíčová pole SACHTA, RUDA, PLAN a SKUTECNOST jsou funkcí celého klíče. Nemůže se totiž stát, že by plán souvisel jenom se dnem nebo jenom se jménem trpaslíka. Při pozorném prozkoumání celé tabulky zjistíte skutečně, že všechna neklíčová pole jsou funkcí celého klíče, a nikoli jen jeho části. Při studiu tabulky č. 4 zjistíte některé zajímavé manažerské vlastnosti Sněhurky, které vám možná připomenou některé epizody z vašeho vlastního života. Položíme-li si otázku, zda tabulka č. 4 by nemohla být ve 3NF, musíme zkoumat, zda neexistují nějaké skryté souvislosti mezi neklíčovými poli. Souvislost mezi plánem a skutečností určitě neexistuje. Souvislost mezi rudou a plánem těžby také nenajdeme, protože Sněhurka má dost subjektivní představu o zařezávání jednotlivých skřítků prací. Jediná souvislost, kterou můžeme velmi snadno vytipovat, je mezi číslem šachty a rudou, která se v ní těží. To patrně znamená, že jednotlivé šachty jsou specializovány na těžbu jednotlivých nerostů. Pro nás to znamená, že tabulka č. 4 není ještě vrcholem dokonalosti. Do 3NF a výše se může tabulka dostat jedině rozpadem na další dvě tabulky, při kterém se závislost mezi číslem šachty a těžbou rudou zcela osamostatní. Tak vznikne tabulka č. 5, která není nic jiného než úplně obyčejný číselník šachet. Zde pole SACHTA tvoří klíč a pole RUDA neklíčovou položku. Tabulka č. 5 je poté již v 5NF, protože na ní lze těžko ještě něco zlepšit. Ze staré tabulky č. 4 potom zbyde torzo tak, jak je uvedeno v tabulce č. 6.

Navíc je zde pro zpřesnění práce zavedeno nové pole PRIKAZ, které není nic jiného než unikátní číslo příkazu pro jednotlivé trpaslíky. Další pole CITRP vlastně obsahuje evidenční číslo trpaslíka. Všimněte si, že Sněhurka si své trpaslíky konečně očíslovala. DEN je datum fárání, SACHTA je číslo šachty, PLAN a SKUTECNOST jsou opět plánované a vytěžené kilogramy rudy. Tabulka č. 6 už má smysl, protože zavádí číslování příkazů, trpaslíků a šachet, jako to, co se v budoucnu vyplácí při práci s daty. Sněhurka se doopravdy zakousla do problematiky dekompozice a pochopila velký význam číselníků a jednoznačného číslování pojmů. K tabulce by měl existovat číselník trpaslíků, který zde není konkrétně zobrazen.

Snadná aktualizace

Největší výhodou dekompozice tabulek do vyšších normálních forem počínaje třetí normální formou je to, že takové tabulky jdou

snadno aktualizovat. Aktualizace vlastně znamená přidávání nových hodnot do tabulky nebo rušení hodnot v tabulce či oprava jednotlivých údajů v tabulce. Pokud si uvědomíme, že tabulka ve třetí normální formě obsahuje unikátní klíč, pak v jazyce SQL máme možnost se účinně bránit přidávání nových položek s neúplným nebo duplicitním klíčem. Přidávání dat do tabulky ve 3NF je s použitím SQL velmi snadnou záležitostí. Chceme-li naopak některou položku zrušit nebo opravit, potom stačí si uvědomit hodnotu klíče, která popisuje příslušný řádek, kde se položka

PRIKAZ	CITRP	DEN	SACHTA	PLAN	SKUTECNOST
1	2	1. 1. 98	5	3000	4000
2	3	1. 1. 98	3	3000	2000
3	7	1. 1. 98	8	3500	5000
4	2	2. 1. 98	5	4000	1000
5	3	2. 1. 98	7	2	0
6	7	2. 1. 98	5	3000	6000

Tab. č. 6. Dolují naplno.

vykazuje, a napsat příslušný příkaz SQL. Nedivme se Sněhurce, že zavedla nové pole PRIKAZ, protože toto pole je vlastně jednoduchým klíčem do celé tabulky. Pokud byste se chtěli pocvičit v určování normálních forem tabulek, stačí si ještě jednou přečíst druhý díl tohoto seriálu a u všech tabulek určit, jaká je jejich normální forma.

Entity a relace

V žádném případě vás nechci hned v úvodu zaskočit odbornými výrazy. Úplně bude stačit, když si dvě nová slova přeložíme do češtiny. Entita neznamená nic jiného než jsoucno, tedy všechno to, co existuje.

Naproti tomu relace není nic jiného než vzájemný vztah mezi dvěma existujícími jsoucny. V předchozích dílech tohoto seriálu jsme se zatím zabývali pouze tabulkami, kde každá z nich popisovala právě jedno samostatné jsoucno. Nyní nás čeká vyhledávání vztahů mezi tabulkami.

Pokud jsou tabulky ve vysoké normální formě, nebude vůbec žádným problémem mezi nimi ustanovit smysluplné relace. Abychom se blíže seznámili s pojmem entity, našeho českého jsoucna, zamysleme se nejprve nad naším prvním tématem receptů na jídla. Takový RECEPT, DAVKA, SUROVINA, KROK nebo FRAZE jsou typická jsoucna, která prostě existují.

Pokud raději chcete spolu se Sněhurkou sledovat trpaslíky, ať už v pracovní době nebo mimo ni, potom důležité entity budou TRPASLIK, PLAN, SACHTA, RUDA a VLASTNOST. Stálo by za to se zamyslet nad tím, jak budovat vzájemné vztahy neboli relace mezi jednotlivými entitami. Tyto relace by potom měly mít co nejčistší a nejjednodušší tvar. Vztah mezi tabulkami můžeme totiž chápat jako vztah mezi jednotlivými položkami re-

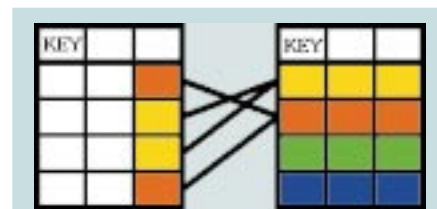
spektive řádky těchto tabulek. Takovýto vztah může být tří typů.

První typ je označován jako 1 : 1. V tomto případě si můžeme představit dvě tabulky, kde každému řádku jedné tabulky odpovídá právě jeden řádek tabulky následující. Jako příklad bych uvedl vztah dvou tabulek, kde v jedné je plánovaná výroba a v druhé skutečná výroba.

Ke každému dni a ke každému stroji existuje v první tabulce plánovaný výkon, zatímco v druhé tabulce ke každému dni a ke každému stroji je znám skutečný výkon. Takového

tabulky přirozeně vznikají v různých fázích práce s databázovým systémem. První tabulka při plánování, zatímco druhá tabulka při evidenci dosažených výsledků. Není proto špatné je vnímat odděleně, ovšem z hlediska databázových systémů patrně je vztah mezi nimi velmi málo produktivní a užitečný, protože si jistě dovedete představit jednu tabulku, kde jsou plán a skutečnost uvedeny současně, a přesto bude také v 5NF. Z tohoto důvodu se vzájemné vztahy, tedy relace typu 1 : 1 vyskytují v databázových systémech zřídka a považují se za nadbytečné.

Druhým extrémem je relace typu M : N. Ta popisuje úplný chaos ve vzájemných vztazích mezi dvěma tabulkami. Představme si například v jedné tabulce číselník řek v 5NF a v druhé tabulce číselník států též v 5NF. Jaký je vzájemný vztah entit REKA a STAT? Jde o vztah typu M : N. To znamená, že každá řeka může protékat několika, tedy M státy, zatímco v každém státě klidně může být několik, například N řek. Takový vztah je sice zajíma-



Obr. č. 6. Relace N : 1 mezi dvěma tabulkami.

vý svou univerzálností, ale patrně příliš komplikovaný na to, aby nám usnadnil vzájemné prohledávání tabulek.

V tomto slova smyslu jsou ideální relace třetího typu, tedy relace typu N : 1. Ty potom předurčují, že tabulky spolu souvisí organizovanějším způsobem.

Jaromír Kukal