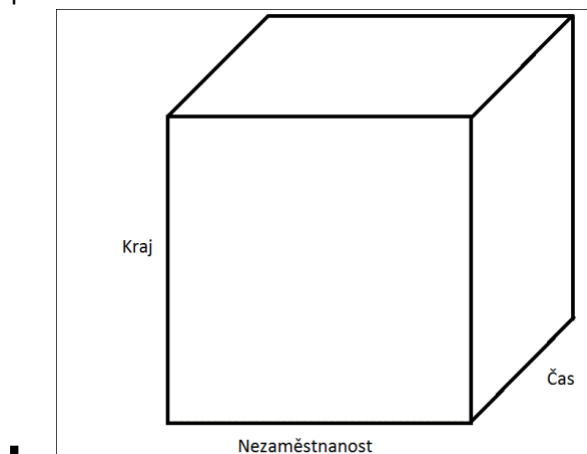


Normalizace databáze

OLAP, OLTP, Normální formy,
normalizace, optimalizace, 0. NF, 1. NF,
2. NF, 3. NF

OLAP (online analytical processing)

- Je technologie uložení dat v databázi
- Umožňuje uspořádat velké objemy dat tak, aby byla data přístupná a srozumitelná uživatelům zabývajícím se analýzou obchodních trendů a výsledků
 - Chceme hlavně získat souhrnná data
 - Důležitá je rychlost
- **OLAP kostka**
 - Způsob organizace dat, který rozšiřuje dvojrozměrně tabulkové uspořádání tak, že každá datová dimenze je uložena v jedné ose kostky
 - Tím překonává některá omezení relačních databází



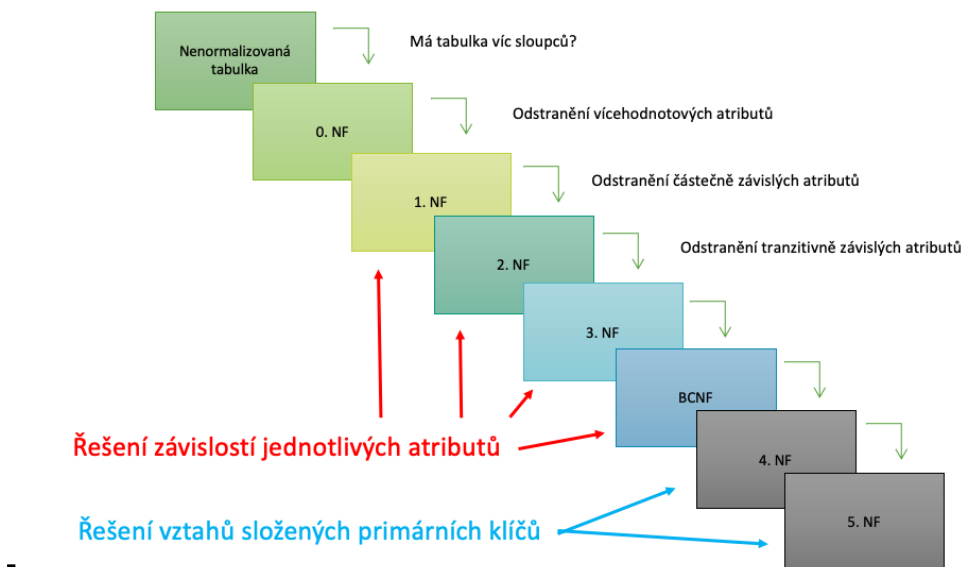
OLTP (online transaction processing)

- Technologie uložení dat v databázi
- Umožňuje co nejjednodušší a nejbezpečnější modifikaci dat v mnohauživatelském prostředí
 - Chceme ukládat a organizovat nová data
 - Důležitá je integrita

Normalizace a optimalizace

- Doporučení pro ideální návrh databáze
- Čím je tabulka ve vyšší normální formě (stupeň doporučení), tím kvalitněji je tabulka navržena = je menší šance porušení integrity dat (stav, kdy přechtená data jsou totožná s daty uloženými)
- Ale zároveň vzrůstá zátěž na databázový stroj = pomalejší výsledky
- Kompromis: Dosáhnout co nejvyšší formy a pak úroveň snižovat na zvládnutelnou úroveň
- **Normalizace:**
 - Proces, při kterém se relace rozkládají za účelem.
 - Jednodušší práce s daty
 - Jejich lepší manipulace
 - Zabránění redundance dat (dat, která se zbytečně opakují)
 - Lepší konzistence dat (souvisí s redundancí)
 - Při jednoduchých klíčích nemá smysl uvažovat normu vyšší, než druhou
 - Obvykle nemá smysl usilovat o splnění norem nad třetí

- Běžně se dosáhne co nejvyšší normy, pak se během optimalizace výkonu DB pravidla poruší
- **Stupně normalizace:**



0. NF

- Tabulka je v nulté normální formě právě tehdy, existuje-li alespoň jedno pole, které obsahuje více než jednu hodnotu

1. NF

- Tabulka je v první normální formě, jestliže lze do každého pole dosadit pouze jednoduchý datový typ (jsou dále nedělitelné)
- Všechna data jsou atomická (v jedné položce více dat: "+420 777 888 999, +420 156 321 489")

Jméno	Příjmení	ID (PK)	Telefon
Jan	Polesný	1	454789789
Martin	Dvořák	2	489159159, 607777777
Jana	Pokorná	3	478478896

Jméno	Příjmení	ID (PK)	ID (PK)	Telefon
Jan	Polesný	1	1	454789789
Martin	Dvořák	2	2	489159159
Jana	Pokorná	3	2	607777777
			3	478478896

○

2. NF

- Tabulka je ve druhé normální formě, jestliže je v první a navíc platí, že existuje klíč a všechny neklíčové atributy jsou plně závislé na celém klíči
- Data závisí na celém klíči, nejen na jeho části
- Je zajímavá u tabulek se složeným klíčem, jinak je splněna

Značka (PK)	Model (PK)	Telefon	Počet
Škoda	Superb	458458458	5
Škoda	Yeti	458458458	10
Fiat	Punto	478489458	2
VW	Passat	468469467	1

ZnačkaID (PK)	Model (PK)	Počet
1	Superb	5
1	Yeti	10
2	Punto	2
3	Passat	1

ZnačkaID (PK)	Značka	Telefon
1	Škoda	458458458
2	Fiat	478489458
3	VW	468469467

3. NF

- Tabulka je ve třetí normální formě, jestliže každý neklíčový atribut není transitivně závislý na žádném klíči schématu
- Tabulka je ve třetí normální formě, je-li ve druhé normální formě a zároveň neexistuje ani jediná závislost neklíčových sloupců tabulky
- Neklíčová data jsou závislá jen na klíči a ne mezi sebou

RČ (PK)	Jméno	Příjmení	Pozice	Plat
8011124545	Petr	Svoboda	CEO	100000
7703144748	Jiří	Vomáčka	Technik	17000
7812124898	Alena	Černá	Technik	17000
8207070509	Jana	Zimová	Sekretářka	15000

RČ (PK)	Jméno	Příjmení	PoziceID
8011124545	Petr	Svoboda	CEO
7703144748	Jiří	Vomáčka	Technik
7812124898	Alena	Černá	Technik
8207070509	Jana	Zimová	Sekretářka

Název (PK)	Plat
CEO	100000
Technik	17000
Sekretářka	15000

Boyceho-Coddova normální forma (BCNF)

- Dodatečná norma řešící vzácné situace
- Tabulka je v Boyce-Coddově normální formě, jestliže je ve 3. NF a pro každou závislost $X \rightarrow Y$ platí, že X obsahuje klíč schématu

- Ve vztahu $X \rightarrow Y$ se nesmí stát, že X není klíčovým atributem, pokud Y je klíčovým atributem
- Lze ji aplikovat pouze za podmínek:
 - Relace má více kandidátů na klíč
 - Nejméně 2 z kandidátů na klíč jsou složené
 - Kandidáti na klíč se v některých attributech musí překrývat
- Relace je v BCNF, pokud není mezi kandidáty na klíč žádná funkční závislost

Předmět	Den	Učitel	Učebna
MAT	Po	Adam	A
MAT	Út	Adam	C
ANJ	Po	Beáta	B

Učitel, Den \rightarrow Předmět, Učebna

Předmět, Den \rightarrow Učebna, Učitel

Učitel \rightarrow Předmět

Učitel	Předmět
Adam	MAT
Beáta	ANJ

Učitel	Den	Učebna
Adam	Po	A
Adam	Út	C
Beáta	Po	B

○

4. NF

- Tabulka je ve čtvrté normální formě, jestliže žádný vztah klíčových atributů není nazávislý na ostatních vztazích
- Tabulka má klíč složený z nejméně tří atributů
- V praxi se téměř nevyskytuje
- Obvykle není vyučována

Restaurace (PK)	Pizza (PK)	Oblast
Fary	Hawai	Františkov
Fary	Bismarck	Jeřáb
Fary	Salami	Harcov
A1	Salami	Jeřáb
A1	Bismarck	Františkov

Restaurace (PK)	Pizza (PK)
Fary	Hawai
Fary	Bismarck
Fary	Salami
A1	Salami
A1	Bismarck

Restaurace (PK)	Oblast (PK)
Fary	Františkov
Fary	Jeřáb
Fary	Harcov
A1	Jeřáb
A1	Františkov

○

5. NF

- Tabulka je v páté normální formě, pokud je ve čtvrté a není možné ji dále rozdělovat bez ztráty informací

Prodejce (PK)	Produkt (PK)	Místo (PK)
Alfa	Nůžky	Liberec
Alfa	Lepidlo	Liberec
Alfa	Papír	Praha
Beta	Papír	Liberec

Prodejce (PK)	Produkt (PK)	Místo (PK)
Alfa	Nůžky	Liberec
Alfa	Lepidlo	Liberec
Alfa	Papír	Liberec
Alfa	Nůžky	Praha
Alfa	Lepidlo	Praha
Alfa	Papír	Praha
Beta	Papír	Liberec

↓

Prodejce (PK)	Produkt (PK)
Alfa	Nůžky
Alfa	Lepidlo
Alfa	Papír
Beta	Papír

↓

Prodejce (PK)	Místo (PK)
Alfa	Liberec
Alfa	Praha
Beta	Liberec

○