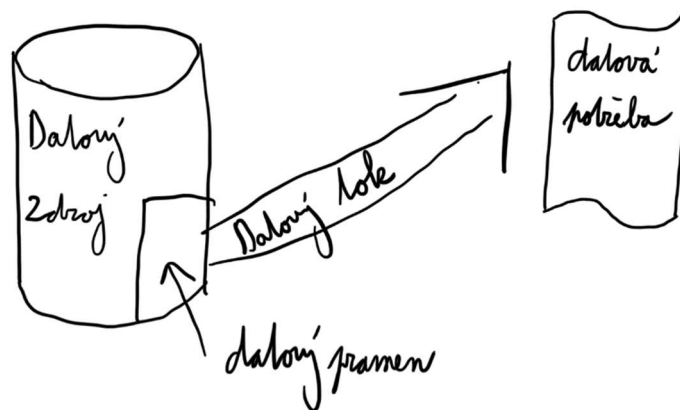


# 24. DB pojmy

## Zásady pro práci s daty

- Minimalizovat vstupy
- Maximalizovat výstupy
- Vkládat pouze primární údaje
  - Stav tachometru před jízdou a po jízdě
- Každý údaj pouze jednou
- Kompromis mezi četností a přesností
  - Měření teploty venku každé 3 sekundy je zbytečné

## Závislosti zdroje dat



## Klíče a číselníky

### Entita

- vše co jde jednoznačně určit

### Primární klíč

- jednoznačná identifikace entity

### Částečný klíč

- část primárního klíče, která v kombinaci s jinou částí tvoří primární klíč

### Cizí klíč

- sekundární
- Odkaz na primární klíč

### Číselník

- Primární klíče s popisem v tabulce
  - PSČ s názvem města
  - Zdravotní pojišťovny

## Kontrola a ochrana dat

### Formální konstanty

- Vizuální wtf
- Opakovaným zápisem
- Typem dat
- Rozsahem údaje
- Povinnost výskytu

### Logické

- Podle masky
- Podle seznamu
- Kontrola intervalem
  - Číselný dosah a datum
- Matematickými pravidly
  - Dělení 11 beze zbytku

### Ochrana dat před ztrátou

- Zálohování
- Distribuované databáze X centralizovaná báze dat
- Cloudy

### Ochrana dat před zneužitím

- Zničení
- Autorizovaný přístup
- Hesla
- Biometrie

Nároky na ochranu dat rostou s množstvím uživatelů a interaktivitou jejich práce

### Nároky na kvalitu dat

- Chyba dat může vést k chybě výstupu
- Dominový efekt
  - Jedna chyba vede k další chybě
- Řetězový efekt
  - Jedna chyba vytvoří více chyb
- Synergetický efekt
  - 2 špatné věci vyvolají jednu extrémní věc
  - Hoří plynárna + zamrzá voda kvůli mrazům

### Hlediska k posuzování kvality dat

- Dostupnost
  - V čase
  - V místě
  - V metadatech
    - Požadovaný formát
- Přesnost

- Platnost dat v kontextu
- Úplnost
  - Stupeň úplnosti v kontextu
- Konzistence

### **Obvyklé chyby v datech**

#### **Uživatel**

- Pravopisné chyby
- Prohozené číslice
- Špatně zadávané kódy
- Chyby číselníků po konsolidaci databází
- Neoprávněné testování
- Rozdílná interpretace dat
- Chyby z existenčních zdrojů
- Nedodržení bezpečnostní politiky

#### **Normální formy**

##### **Normalizace**

- Sada pravidel popisující postup transformace struktury entit a reakcí E-R modelu na strukturu
- Cílem normalizace je:
  - Odstranění redundantních dat
  - Omezení složitosti
  - Zabránění anomálií
    - Po smazání řidiče se smažou všechny jeho jízdy

##### **0 NF**

- Tabulky bez normalizace
- Tabulka ve wordu
  - Jednoduchá
  - nepoužitelná

##### **1 NF**

- každý atribut obsahuje pouze atomické hodnoty
- NEEXISTENCE PRIMÁRNÍCH KLÍČŮ
- PŘ. Jméno a Přímení

##### **2 NF**

- Každý atribut je závislý na primárním klíči
- PŘ. ID zaměstnance

##### **3 NF**

- Všechny neklíčové atributy musí být vzájemně nezávislé
- PŘ. Odstranění redundantních PSČ

## **BCNF**

- Mezi 3 a 4 formou
- Atributy, které jsou součástí primárního klíče musí být vzájemně nezávislé

## **4 NF**

- Relace popisuje pouze příčinou souvislost mezi klíčem a atributem
- PŘ. Kontakt tabulka – práce tabulka

## **5 NF**

- Relace již nelze beztrátově rozložit
- PŘ. Spojení číselníků při strojovém návrhu