

## Analýza a vizualizace dat

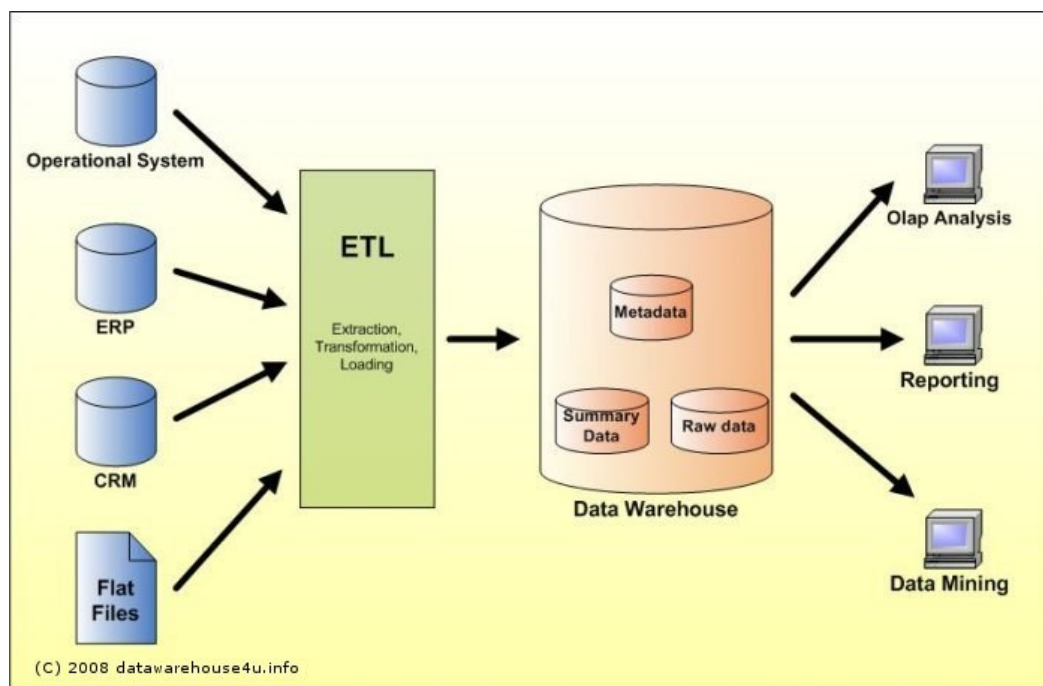
### Business intelligence

Jednou z dalších oblastí, která spadá do sféry systémové integrace, je návrh a implementace řešení, spadajících do oblasti nazývané Business Intelligence (BI). Jedná se o **integraci informací a znalostí, uložených v databázích, jejich následná analýza a prezentace řídicím pracovníkům pro podporu rozhodovacích procesů**. Pojem byl zaveden v roce 1989 (Howard Dresner) jakožto souhrn konceptů, metodik a nástrojů, které zlepšují rozhodovací proces.

Do této kategorie patří následující softwarové prostředky:

- Datový sklad (Data Warehouse)
- OLAP (On-line Analytical Processing)
- Dolování dat (Data Mining)

Důvod vzniku BI je různorodost a značná velikost provozních databází (OLTP; On-line Transaction Processing). Získání spontánních informací pro podporu manažerského rozhodování je sice z OLTP databází možné, avšak je hodně pracné a neefektivní. Databáze také mohou obsahovat informace, které jsou nějakým způsobem zahaleny a je třeba je „vydolovat“.



Obr. 9.1 Schéma prostředků business intelligence [Zdroj: datawarehouse4u]

Typy problémů, které lze řešit pomocí dolování dat

1. Klasifikace: Proč ztrácíme zákazníky? Bude produkt úspěšný?
2. Regrese – závislost mezi dvěma proměnnými. Regrese může být lineární nebo logická. Do této kategorie úloh patří odhad tržby, příčina platební neschopnosti, odhad ceny bytů v závislosti na lokalitách apod.

3. Shlukování (segmentace) – matematicky jde o rozdělení do množin dle společných znaků, např. segmentace zákazníků
4. Přiřazování (asociace) – např. analýza nákupního koše – odhalení produktů, které se prodávají společně
5. Prognózování – trendy, cykly
6. Analýza časových řad – předpověď trendu, forecasting
7. Risk management – pravděpodobnost výskytu událostí, např. podvodného jednání v bance
8. Analýza sekvencí – nalezení vzorců v řadách diskrétních hodnot (stavů) – např. sekvence procházení firemního webu návštěvníkem
9. Analýza odchylek – nalezení objektů, které jsou velmi odlišné od ostatních; využití např. při detekci podvodů (banky) nebo detekce výrobních vad
10. Analýza textu (nestrukturovaných dat).

Při analýze dat posuzujeme:

- **Spolehlivost** – dává metoda při opakovaném nasazení stejné výsledky?
- **Platnost** – zachycuje měření objekt přesně?
- **Použitelnost** – můžeme na základě informací provádět rozhodnutí?

## Big Data

Objem dat pořizovaných a zpracovávaných v informačních systémech neustále roste. Zároveň se objevují úlohy, kdy je nutné zpracovat data, která nemají striktně strukturovaný charakter. Vznikl nový fenomén, označovaný termínem „Big Data“. Existuje celá řada definic, podstatu ale nejlépe vystihuje definice dle Gartner: „Big Data jsou soubory dat, jejichž velikost je mimo schopnosti zachycovat, spravovat a zpracovávat data běžně používanými softwarovými nástroji v rozumném čase“.

Základní charakteristika Big Data:

- **Objem (Volume)** – množství dat vznikajících v rámci provozu firem roste exponenciálně každý rok,
- **Typ (Variety)** – různorodost typů dat vzrůstá, například nestrukturované textové soubory, semi-strukturovaná data (XML), data o geografické poloze, data z logů,
- **Rychlost (Velocity)** – rychlost s jakou data vznikají a potřeba jejich **analýzy v reálném čase** vzrůstá díky pokračující digitalizaci většiny transakcí, mobilním zařízení a vzrůstajícímu počtu internetových uživatelů.

Řešení pro Big Data je dvojí:

- a) HW a navyšování zdrojů
- b) Distribuované řešení dat a jejich managementu

## Prezentace a vizualizace dat

Prezentační vrstva BI může být realizována například prostřednictvím **Excelu** (kontingenční tabulka) nebo **podnikového portálu** (např. SharePoint od Microsoftu, Alfresco, Google Cloud). V souvislosti s prezentací firemních informací se můžeme setkat také s termíny **dashboard** nebo **scorecard**.

**Dashboard** – zobrazování sumárních podnikových dat pomocí interaktivního GUI. Dashboard si můžeme představit jako „palubní desku“. Zatímco u automobilu nám dashboard dává přehled informací o řízení auta, firemní dashboard informuje manažera o stavu řízení a ekonomických výsledcích firmy.

**Scorecard** – ukazuje srovnání skutečnosti oproti plánu.

### Balanced Scorecard

Jde o systém sledování a managementu podniku ve čtyřech oblastech:

- finanční hodnoty
- perspektiva zákazníka
- podnikové procesy
- inovace, učení se, flexibilita a růst

Životní cyklus reportů se skládá z:

návrh – správa - doručení

## Grafy

Kategorie grafů:

- **Distribuce** – sloupcové, bodové grafy
- **Kompozice** - *jak se rozkládá celek do dílů* - koláčové, stromové, sloupcové
- **Trendy** – spojnicové, plošné grafy
- **Vztahy a srovnání** – bublinové, bullet grafy

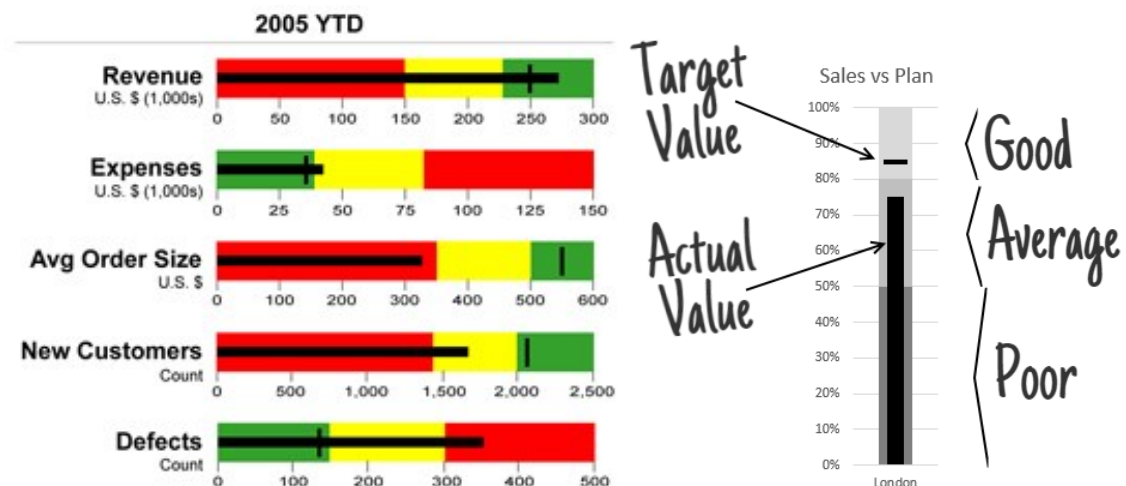
Schémata grafů:

- Sekvenční - seřazení hodnot od nízkých po vysoké
- Divergentní - existuje klíčový bod (např. průměr)
- Kategorická - rozdílné skupiny: kontrastní barva

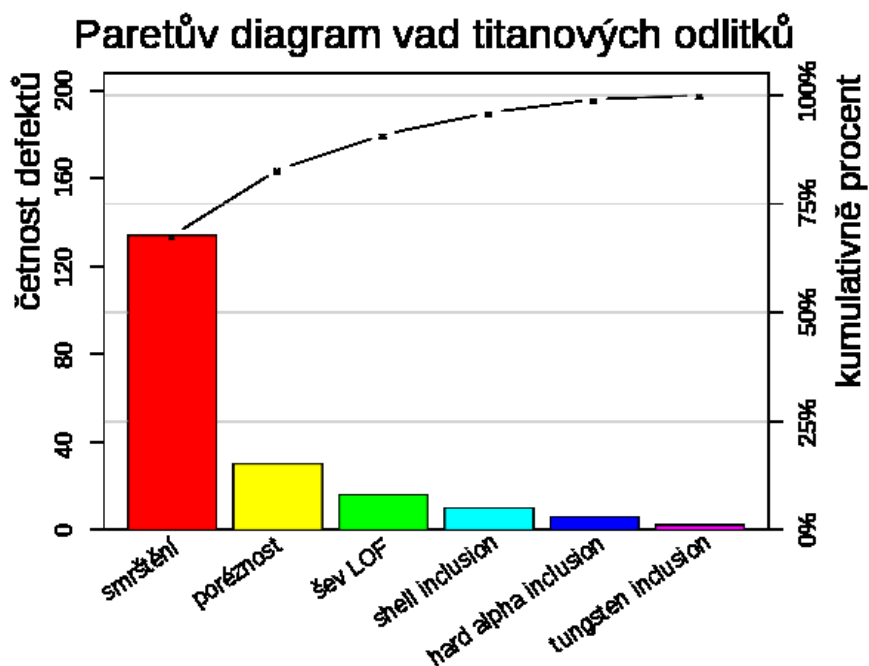
Typy grafů:

- Koláčový – pro znázornění jak se celek rozpadá na části; při velkém počtu částí nepřehledný
- Spojnicový – hodí se pro změny hodnot v čase a trendy; omezení – velký počet řad brání rozlišení

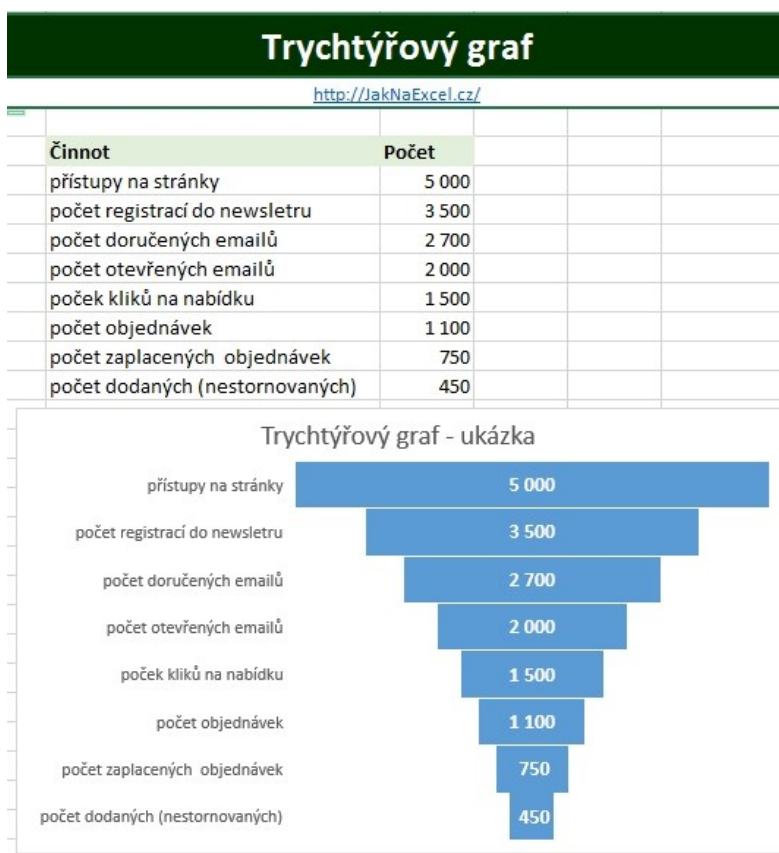
- Sloupcový – vhodný pro srovnání hodnot u řady položek
- Bublinový graf – hledání mezních hodnot, zobrazení extrémů
- Bullet graf – sloupcový graf obsahující aktuální hodnotu a referenční (kam míříme)
- Paretův diagram – kombinace sloupcového a čárového grafu; sloupce znázorňující četnost pro jednotlivé kategorie jsou seřazeny podle velikosti (nejvyšší vlevo) – čára představuje kumulativní četnost; používá se pro znázornění důležitosti jednotlivých kategorií
- Trychtýřový graf (Funnel chart)
- Paprskový graf
- Prstencový graf
- Dendrogram – shluková analýza – horizontální osa vyjadřuje vzdálenost mezi jednotlivými shluky
- Cladogram – použití v biologii pro klasifikaci organismů
- Polární graf – zaznamenávání dat, která vykazují určitou cykličnost, či se u nich zjišťuje směr
- Vodopádový graf (Waterfall/Bridge chart)
- Svícový graf (Candlestick chart) – vhodný pro znázornění trendu, zobrazuje nejvyšší a nejnižší hodnotu (např. cenu)
- Histogram – rozložení hodnot, znázornění distribuce dat; výška sloupců vyjadřuje četnost veličiny v daném intervalu
- Myšlenková mapa – grafické uspořádání klíčových slov
- Průběhový diagram (Run chart)
- Regulační diagram (Control chart) - zobrazuje změny určité sledované veličiny v čase; typicky se používá pro sledování a analýzu procesů.
- Korelační diagram (Scatter diagram) – znázorňuje statistickou závislost a vztah dvou naměřených hodnot (např. pro opakované srovnávání naměřených hodnot, spokojenost zákazníků, údaje o kvalitě výroby...)



Obr. 9.2 Bullet chart



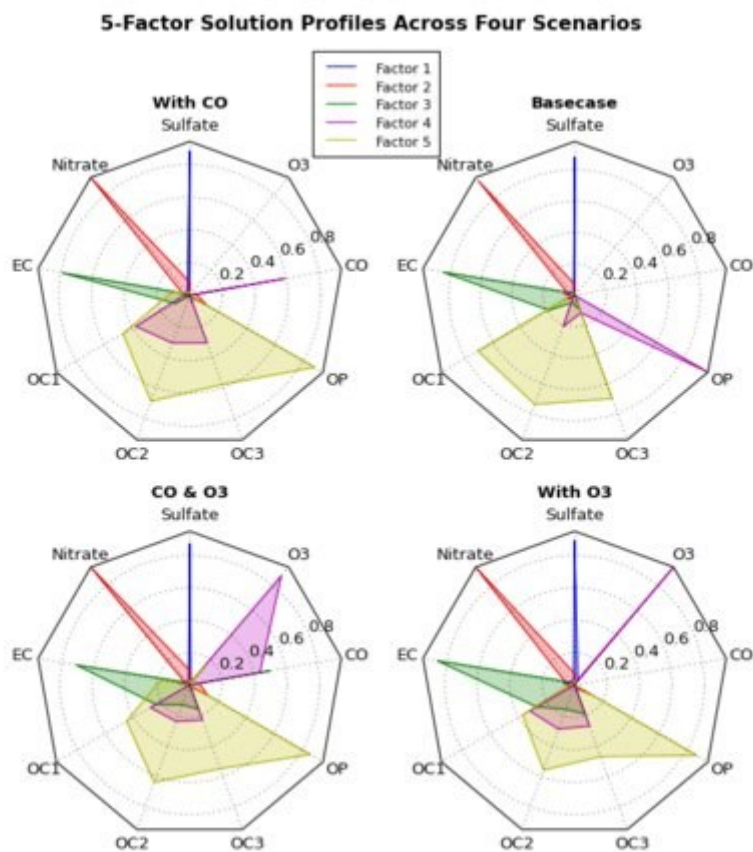
Obr. 9.3 Paretův diagram [Zdroj: Wikipedie]



Obr. 9.4 Trychtýřový graf



Obr. 9.5 Candlestick chart (svícový graf)



Obr. 9.6 Polární graf

[Zdroj: <http://office.lasakovi.com/excel/grafy/polarni-grafy-excel/>]