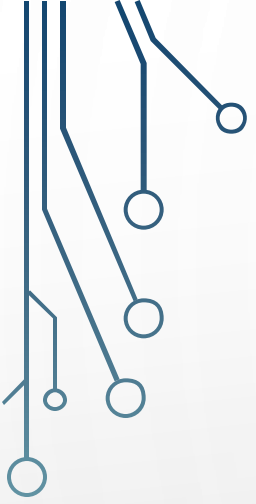


The background of the slide is a dark blue/black field filled with a dense, slightly blurred pattern of white and light blue binary digits (0s and 1s). Overlaid on this are thin, light blue lines that resemble circuit traces or data paths, with small circles at the junctions. A semi-transparent, light beige rectangular box with rounded corners is centered horizontally and vertically, serving as a container for the title and author information.

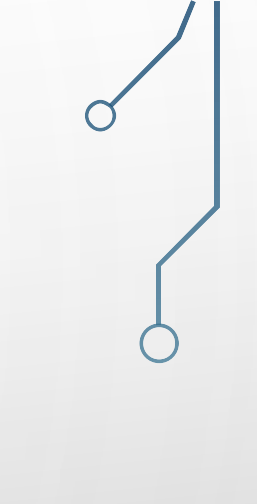
BIG DATA

DOC. ING. MARCELA HALLOVÁ, PHD.



402,74 miliónov terabytov

**Odhadované množstvo dát,
ktoré sa vyprodukuje
každý deň**

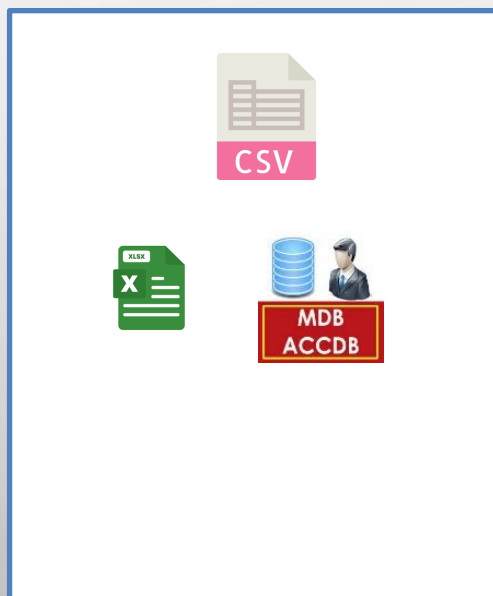




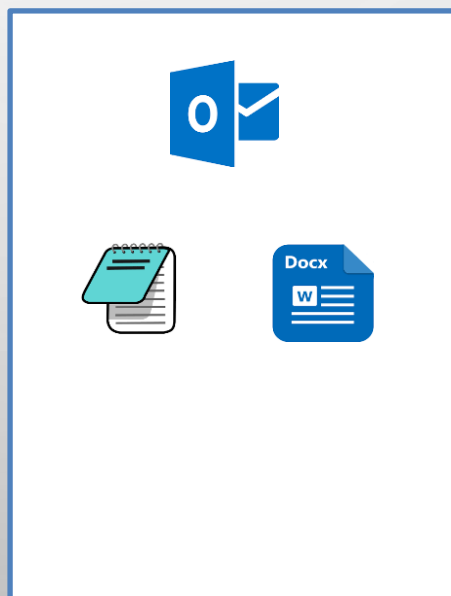
EVOLÚCIA BIG DATA

1. Fáza – štruktúrovaný obsah Obdobie 1970 - 2000	2. Fáza – neštruktúrovaný obsah založený na údajoch z webov Obdobie 2000 - 2010	3. Fáza – Údaje založené na mobilných a senzorických dátach Obdobie 2010 - súčasnosť
✓ Relačné databázové systémy, dátové sklady ✓ Analytické spracovanie online ✓ Dashboardy a výsledkové karty ✓ Data mining a štatistické analýzy	✓ Získavanie a extrakcia údajov ✓ Získavanie názorov – opinion mining ✓ Webové analýzy a web inteligencia ✓ Analýza sociálnych médií ✓ Analýza sociálnych sietí ✓ Priestorová časová analýza	✓ Lokalizačné analýzy ✓ Personálne analýzy ✓ Analýzy relevancie obsahu ✓ Mobilné vizualizácie ✓ Interakcie človek-počítač

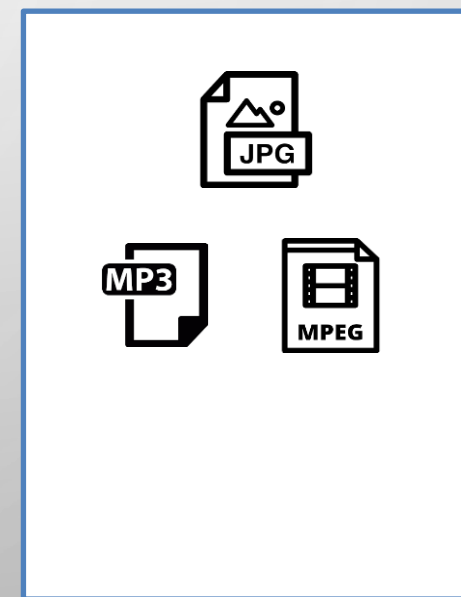
ŠTRUKTÚROVANÉ



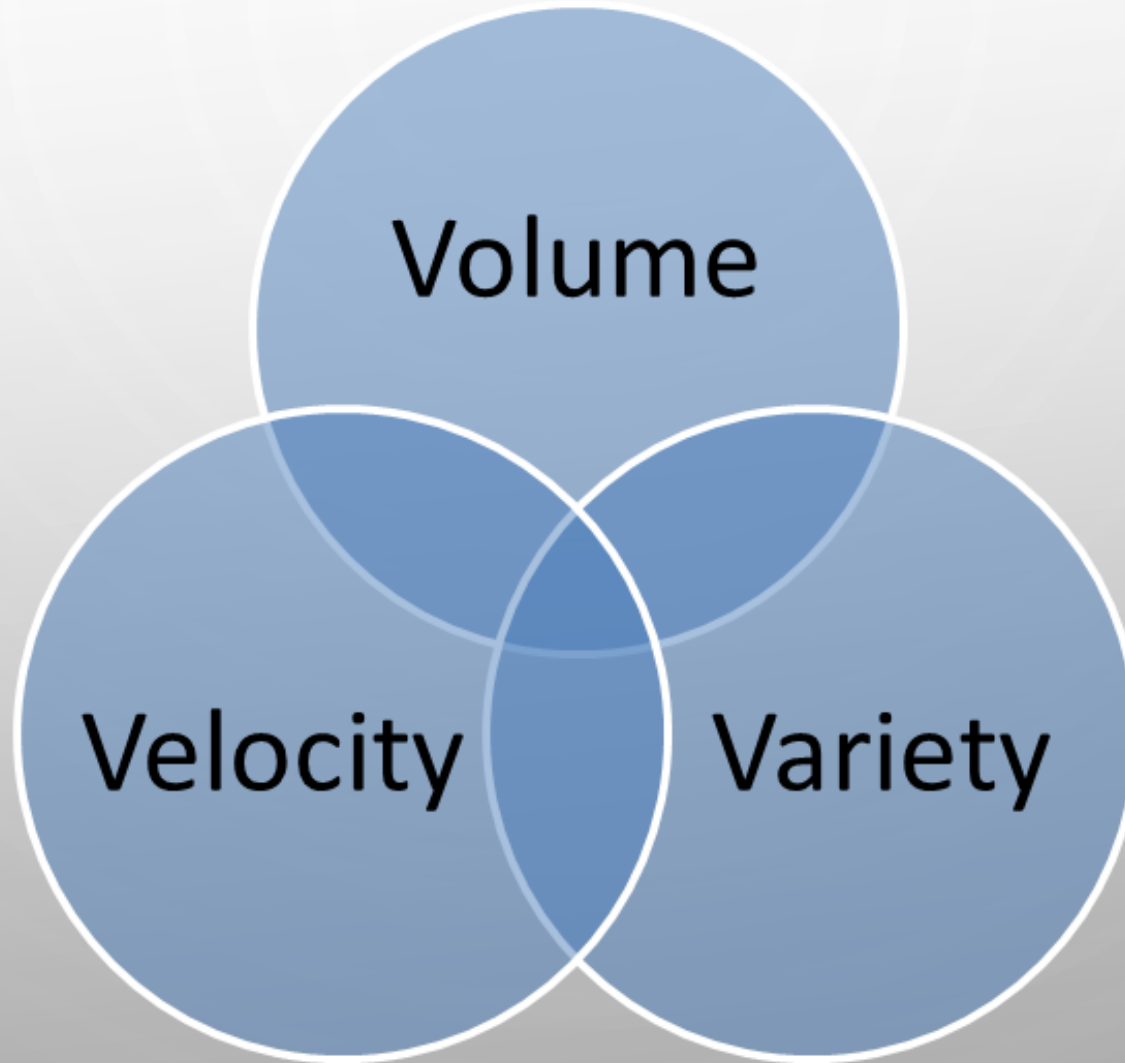
POLO - ŠTRUKTÚROVANÉ



NEŠTRUKTÚROVANÉ



CHARAKTERISTICKÉ VLASTNOSTI BIG DATA



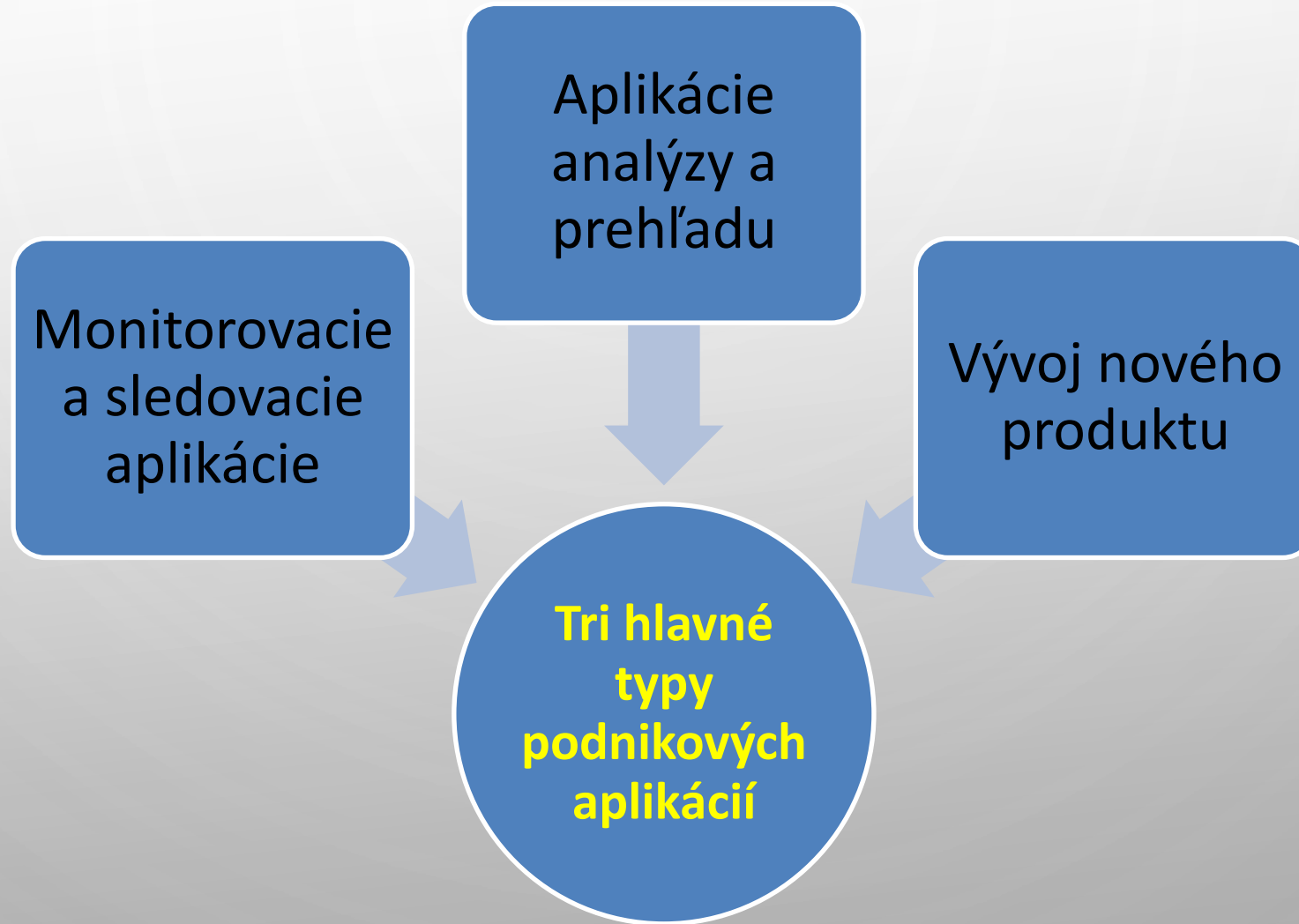
CHARAKTERISTICKÉ VLASTNOSTI BIG DATA

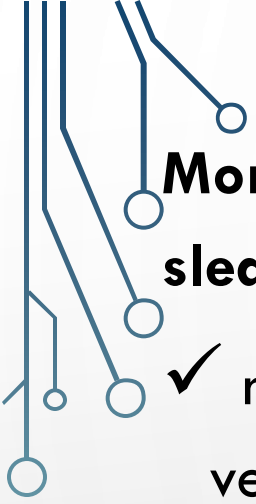
- ✓ **Volume** – **Množstvo** - veľké množstvo dát. Veľkosť údajov hrá rozhodujúcu úlohu pri určovaní ich hodnoty.
- ✓ **Velocity** – **Rýchlosť tvorby** - dáta prichádzajú vysokou rýchlosťou zo strojov, sietí, sociálnych médií, mobilných telefónov a iných zdrojov rýchlosťou veľkých dát.
- ✓ **Variety** – **Rôznorodosť** – sem patrí rozdelenie dát na štruktúrované, polo-štruktúrované a neštruktúrované.

VÝHODY SPRACOVANIA BIG DATA

- ✓ Podniky môžu pri prijímaní rozhodnutí využívať externé informácie.
- ✓ Prístup k sociálnym údajom z vyhľadávačov a stránok ako Facebook, Twitter umožňujú organizáciám doladiť ich obchodné stratégie.
- ✓ Vylepšený zákaznícky servis.
- ✓ Včasná identifikácia rizika pre produkt/službu, ak nejaké existuje.
- ✓ Lepšia prevádzková efektivita.

APLIKÁCIE SÚVISIACE S BIG DATA





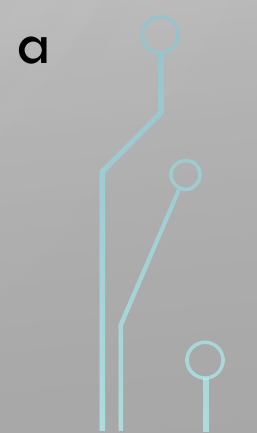
Monitorovacie a sledovacie aplikácie:

- ✓ monitorovanie verejného zdravia,
- ✓ sledovanie aktív,
- ✓ monitorovanie dodávateľského reťazca,
- ✓ preventívna údržba strojov a zariadení.

Aplikácie analýzy a prehľadu:

- ✓ prediktívna polícia – kontrola kriminality,
- ✓ víťazstvo v politických voľbách,
- ✓ osobné zdravie.

Aplikácie analýzy a prehľadu:

- ✓ flexibilné poistenie auta,
 - ✓ maloobchodná propagácia,
 - ✓ odporúčanie služieb a tovarov.
- 

NAJPOUŽÍVANEJŠIE NÁSTROJE PRE SPRACOVANIE BD

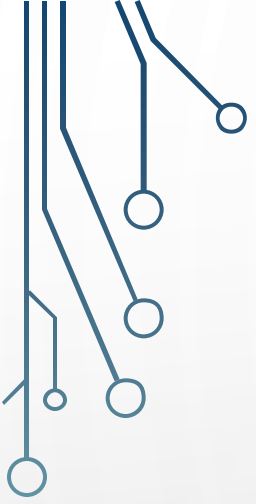
- ✓ **Apache Hadoop** - umožňuje distribuované spracovanie rozsiahlych súborov údajov naprieč klastrami počítačov.
- ✓ **HPCC Systems** - spracovanie údajov na jedinej platforme, jedinej architektúre a jedinom programovacím jazyku.
- ✓ **Apache STORM** - jeden z najlepších nástrojov pre veľké dáta, ktorý ponúka distribuovaný systém spracovania v reálnom čase odolný voči chybám.
- ✓ **Qubole** - nástroj s otvoreným zdrojovým kódom pre veľké dáta, ktorý sa sám spravuje, sám sa optimalizuje.
- ✓ **Statwing** - je jednoducho použiteľný štatistický nástroj.

VÝZVY PRI SPRACOVANÍ BIG DATA

- ✓ Nedostatok správneho pochopenia Big Data.
- ✓ Problémy s rastom údajov.
- ✓ Zmätok pri výbere nástroja na spracovanie Big Data.
- ✓ Nedostatok dátových profesionálov.
- ✓ Zabezpečenie údajov.
- ✓ Integrácia údajov z rôznych zdrojov.

UKLADANIE BIG DATA

- ✓ **Distribúcia** - distribúcia blokov množín údajov medzi viacerými uzlov. Každý uzol má svoj výpočtový výkon, čo dáva schopnosť systému paralelne spracovávať dátové bloky.
- ✓ **Replikácia** - distribuované súborové systémy budú tiež replikovať dátové bloky na rôznych klastroch kopírovaním rovnakých častí informácií do viacerých klastrov v rôznych stojanoch.



Kilobajt 1 000 bajtov

Megabajt 1 000 000 bajtov

Gigabajt 1 000 000 000 bajtov

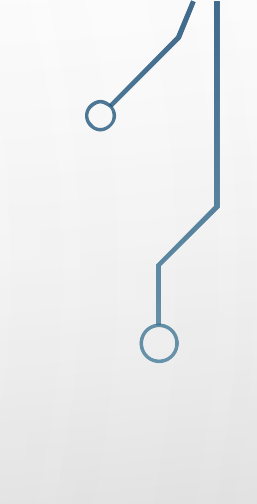
Terabajt 1 000 000 000 000 bajtov

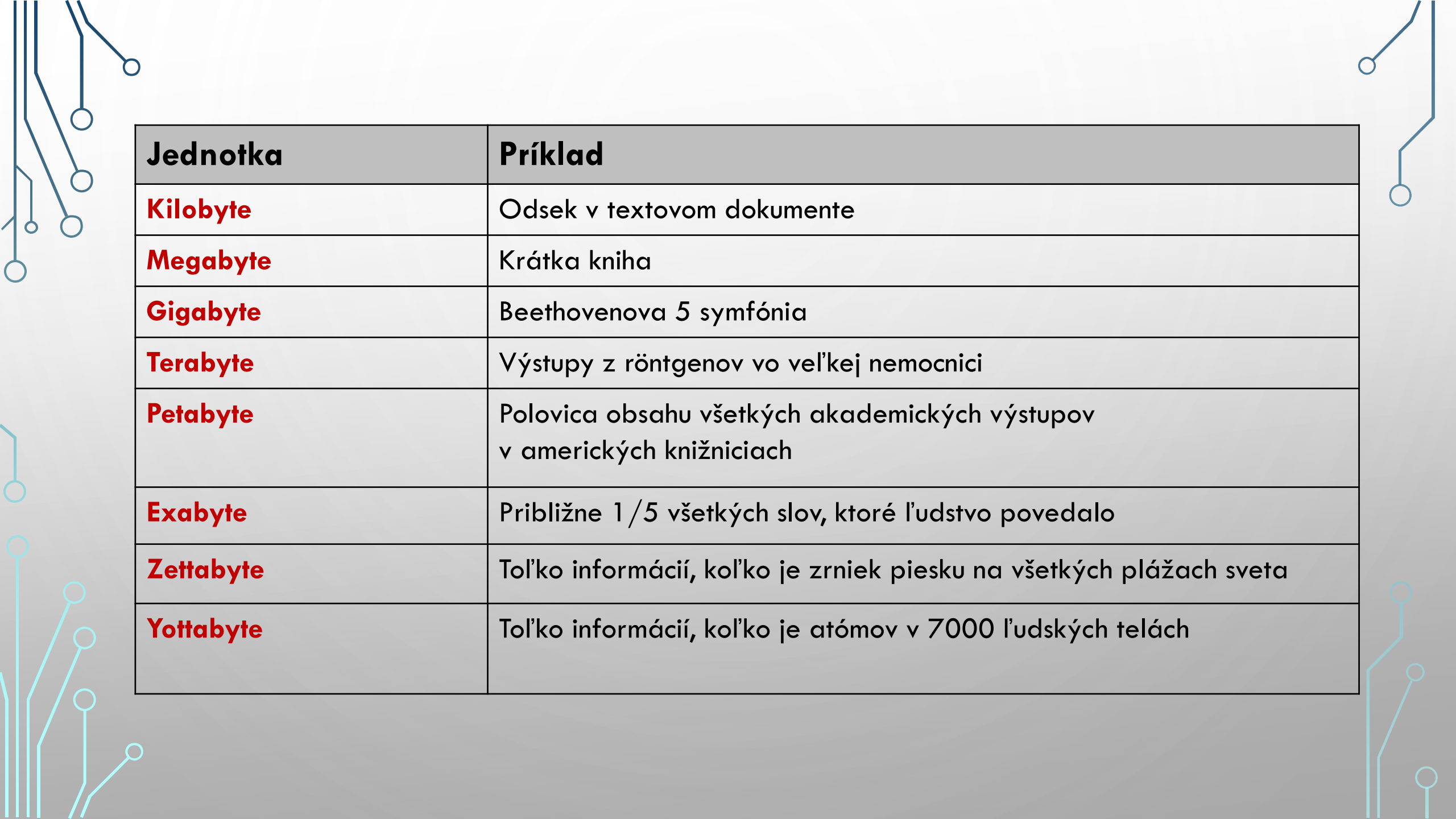
Petabajt 1 000 000 000 000 000 bajtov

Exabajt 1 000 000 000 000 000 000 bajtov

Zetabajt 1 000 000 000 000 000 000 000 bajtov

Yottabajt 1 000 000 000 000 000 000 000 000 bajtov





Jednotka	Príklad
Kilobyte	Odsek v textovom dokumente
Megabyte	Krátka kniha
Gigabyte	Beethovenova 5 symfónia
Terabyte	Výstupy z röntgenov vo veľkej nemocnici
Petabyte	Polovica obsahu všetkých akademických výstupov v amerických knižniciach
Exabyte	Približne 1 / 5 všetkých slov, ktoré ľudstvo povedalo
Zettabyte	Tol'ko informácií, koľko je zrníek piesku na všetkých plážach sveta
Yottabyte	Tol'ko informácií, koľko je atómov v 7000 ľudských telách

ŠKÁLOVATEĽNÁ VÝPOČTOVÁ TECHNIKA CEZ INTERNET

- ✓ Vek internetovej výpočtovej techniky.
- ✓ Vysokovýkonná výpočtová technika (High Performance Computing).
- ✓ Tri nové počítačové paradigmy – RFID, GPS, IoT.
- ✓ Rozdiely výpočtovej paradigmy.



NOVÉ DIMENZIE ÚDAJOV

- ✓ Údaje v reálnom čase.
- ✓ Zdieľané údaje.
- ✓ Prepojené údaje.
- ✓ Vysoko dôverné údaje.



ÚLOŽISKÁ ÚDAJOV

- ✓ **Skladové úložisko** - je jedným z najbežnejších spôsobov ukladania veľkého množstva údajov, má však svoje nevýhody. Ak napríklad potrebujeme okamžitý prístup k svojim údajom a chceme sa vyhnúť oneskoreniam alebo problémom s prístupom k nim cez internet, môžu existovať lepšie možnosti. Skladovanie môže byť tiež drahé, ak hľadáme dlhodobé zmluvy alebo potrebujeme ďalších pracovníkov na správu skladových priestorov.
- ✓ **Cloudové úložisko** - je čoraz populárnejšou možnosťou, pretože je jednoduchšie ako kedykoľvek predtým použiť túto metódu vďaka pokrokom v technológii, ako sú napríklad Amazon Web Services (AWS).

KDE SÚ ULOŽENÉ VŠETKY DÁTA Z INTERNETU?



**China Telecom Data
Centre - Honhote**



**The Citadel – Tahoe Reno
- Nevada**

REÁLNE PRÍKLADY BIG DATA

Big Data v marketingu a reklame

- ✓ Netflix
- ✓ Amazon

Big Data v zdravotníctve

- ✓ Elektronické zdravotné záznamy
- ✓ Zariadenia na každodenné nosenie

Big Data v cestovaní a logistike

- ✓ Logistika
- ✓ Mestská mobilita

Big Data v poľnohospodárstve

- ✓ Precízne poľnohospodárstvo

BIG DATA A ZBIERANIE INFORMÁCIÍ O SÚKROMÍ

- ✓ Zdieľať menej informácií s aplikáciami a službami.
- ✓ Používať silné a jedinečné heslá s dvojfaktorovým overením.
- ✓ Sprísniť nastavenia ochrany osobných údajov na svojich účtoch sociálnych médií.
- ✓ Odstrániť nepoužívané mobilné aplikácie a rozšírenia prehliadača.
- ✓ Neignorovať aktualizácie softvéru alebo operačného systému.

Global Data Generated Annually

