EKONOMICKÁ UNIVERZITA

Fakulta hospodárskej informatiky

Katedra aplikovanej informatiky

distribuované spracovanie údajov

a Hadoop

**Meno študentov:**

**Stupeň/ročník:** druhý/2.ročník

**Školský rok:** 2022/2023

**Meno prednášajúceho:** Ing. Jaroslav Kultán PhD.

**Meno cvičiaceho:** SCHMIDT, Peter, Ing. Mgr., PhD.

Obsah

**Nenašli sa žiadne položky obsahu.**

# Úvod

# Jadro

Kapitoly odlíšiť farbou + to iste zdroje

Témy - distribuované spracovanie dát

Hadoop

Cvičenia Big Data

Existuje niekoľko spôsobov spracovania veľkých údajov v závislosti od veľkosti, štruktúry a formátu údajov, ako aj od konkrétneho prípadu použitia a dostupných zdrojov. Niektoré bežné prístupy zahŕňajú:

· **Distribuované výpočty**: použitie klastra počítačov na paralelné spracovanie údajov, ako napríklad s Apache Hadoop alebo Apache Spark.

· **Cloud computing**: používanie cloudových služieb, ako sú Amazon Web Services (AWS) alebo Microsoft Azure, na spracovanie údajov.

· **Streamovanie:** spracovanie údajov v reálnom čase pri ich generovaní alebo prijímaní, ako napríklad pomocou Apache Kafka alebo Apache Flink.

· **NoSQL databázy**: používanie nerelačných databáz, ako je MongoDB alebo Cassandra, na ukladanie a dopytovanie veľkého množstva neštruktúrovaných údajov.

**· Dátové sklady**: pomocou špecializovaných riešení dátových skladov, ako je Amazon Redshift alebo Google BigQuery, na ukladanie a dopytovanie veľkého množstva štruktúrovaných údajov.

· **Strojové učenie**: pomocou algoritmov a knižníc strojového učenia, ako sú TensorFlow alebo scikit-learn, získate poznatky z veľkého množstva údajov.

Je dôležité poznamenať, že v závislosti od konkrétneho prípadu použitia a dostupných zdrojov možno na spracovanie veľkých údajov použiť kombináciu týchto metód.

Možme si aj štatnice rovno spraviť :D

6. Charakterizjte distribuované technológie na spracovanie veľkého množstva dát (big data). Opíšte základné vlastnosti **Hadoop (dfs, mapreduce, yarn)a doplnkov ako Pig, Hive, Spark atď.**

7. Opíšte ecosystém na spracovanie Big Data, charakterizujte clusterovú technológiu Hadoop, spôsob spracovania dát, vysvetlite **5V**.(štátnice)

# Záver

# Bibliografia

**V aktuálnom dokumente neexistujú zdroje.**

Každý svoja s farbou