



**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

**COVID-19**

UKLADANIE A PRÍPRAVA DÁT (UPA)

Bc. Matej Otčenáš (xotcen01), Bc. Ján Jakub Kubík (xkubik32), Bc. Ján Kačur (xkacur04)

3. novembra 2021

# Obsah

<b>1</b>	<b>Zvolené dotazy</b>	<b>2</b>
<b>2</b>	<b>Dátové sady a ich uloženie do DB</b>	<b>4</b>
2.1	Spôsob uloženia dátových sád do databáze . . . . .	4
2.2	Stručná charakteristika zvolených dátových sád . . . . .	4
<b>3</b>	<b>Implementácia</b>	<b>6</b>
<b>4</b>	<b>Lokálne spustenie projektu</b>	<b>7</b>
	<b>Literatúra</b>	<b>9</b>

# 1 Zvolené dotazy

- **Dotazy zo skupiny A**

- **Dotaz 1:** Vytvořte čárový (spojnicový) graf zobrazující vývoj covidové situace po měsících pomocí následujících hodnot: počet nově nakažených za měsíc, počet nově vyléčených za měsíc, počet nově hospitalizovaných osob za měsíc, počet provedených testů za měsíc. Pokud nebude výsledný graf dobře čitelný, zvažte logaritmické měřítko, nebo rozdělte hodnoty do více grafů.
- **Dotaz 2:** Vytvořte krabicové grafy zobrazující rozložení věku nakažených osob v jednotlivých krajích.

- **Dotaz zo skupiny B:** Sestavte 4 žebříčky krajů "best in covid" za poslední 4 čtvrtletí (1 čtvrtletí = 1 žebříček). Jako kritérium volte počet nově nakažených přepočtený na jednoho obyvatele kraje. Pro jedno čtvrtletí zobrazte výsledky také graficky. Graf bude pro každý kraj zobrazovat celkový počet nově nakažených, celkový počet obyvatel a počet nakažených na jednoho obyvatele. Graf můžete zhotovit kombinací dvou grafů do jednoho (jeden sloupcový graf zobrazí první dvě hodnoty a druhý, čárový graf, hodnotu třetí).

- **Vlasté dotazy**

- **Dotaz 1 (Matej):** Na základe vybraného datasetu o počte hospitalizovaných s ohľadom na vykázané očkovania budú zostrojené spojnicové a stĺpcové grafy s dennou i mesačnou incidenciou, ktoré budú zobrazovať závislosti medzi počtami hospitalizovaných bez očkovania, s očkovaním po prvej a druhej dávke. Podobne budú vytvorené grafy pre počty umrtí pre ľudí, ktorí neboli očkovaní alebo boli očkovaní jednou či dvomi dávkami. Následne bude prevedená analýza dennej incidence počtu nakazených, ktorí neboli očkovaní, alebo boli očkovaní jednou či dvomi dávkami. Z týchto štatistík bude vytvorená predikcia, za akých okolností je väčšia šanca nákazy. Na záver bude vyvodený konečný dôsledok zo všetkých zozbieraných štatistík, ktoré budú napovedať o celkovej efektívnosti očkovania proti ochoreniu COVID-19.
- **Dotaz 2 (Ján):** Budem skúmať vývin positivity testov a jej závislosť na počte testov na obyvateľa. Plánujem tam dať nejaké grafy s tým, ako sa mení pozitivita, potom možno nejakú koreláciu positivity a počtu testov na obyvateľa. Budem to pravdepodobne skúmať najprv celoštátne, a potom podľa okresov. Možno by som tam vybral nejakých TOP 5 okresov, v ktorých je tá korelácia (alebo samotná pozitivita) nízka/vysoká, možno potom vyvodiť záver, kde sa testuje dostatočne/nedostatočne. Presnejšie to ešte premyslím keď to budem riešiť, každopádne myslím že toto sú všetky potrebné dáta.

- **Dotaz zo skupiny C:** Hledání skupin podobných měst z hlediska vývoje covidu a věkového složení obyvatel. Atributy: počet nakažených za poslední 4 čtvrtletí, počet očkovaných za poslední 4 čtvrtletí, počet obyvatel ve věkové skupině 0..14 let, počet obyvatel ve věkové skupině 15 - 59, počet obyvatel nad 59 let. Pro potřeby projektu vyberte libovolně 50 měst, pro které najdete potřebné hodnoty (můžete např. využít nějaký žebříček 50 nejlidnatějších měst v ČR).

## 2 Dátové sady a ich uloženie do DB

### 2.1 Spôsob uloženia dátových sád do databáze

Rozhodli sme sa použiť InfluxDB. InfluxDB je open-source databáza na ukladanie údajov v časových radoch. Hlavný dôvod bol, že väčšina dostupných dátových sád je do veľkej miery závislá a vyvíjajúca sa v čase.

Všetky dátové sady sú uložené v databáze **covid-19**. Každá dátová sada používa vlastný measurement. Measurement by sa dalo prirovnať k tabuľke v SQL databázach.

### 2.2 Stručná charakteristika zvolených dátových sád

Zvolené dátové sady sme tématicky rozdelili podľa dotazov do 5 celkov.

- **1. celok - dotaz 1 zo skupiny A**
  - COVID-19: Přehled hospitalizací
  - COVID-19: Celkový (kumulativní) počet osob s prokázanou nákazou dle krajských hygienických stanic včetně laboratoří, počet vyléčených, počet úmrtí a provedených testů (v2)
- **2. celok - dotaz 2 zo skupiny A + dotaz zo skupiny B**
  - Obyvatelstvo podle pětiletých věkových skupin a pohlaví v krajích a okresech
  - COVID-19: Přehled epidemiologické situace dle hlášení krajských hygienických stanic podle okresu
- **3. celok - vlastný dotaz 1**
  - COVID-19: Přehled úmrtí s ohledem na vykázaná očkování
  - COVID-19: Přehled osob s prokázanou nákazou s ohledem na vykázaná očkování
- **4. celok - vlastný dotaz 2**
  - COVID-19: Přehled epidemiologické situace dle hlášení krajských hygienických stanic podle okresu
  - COVID-19: Celkový (kumulativní) počet provedených testů podle krajů a okresů ČR

- Obyvatelstvo podle pětiletých věkových skupin a pohlaví v krajích a okresech
- **5. celok - dotaz zo skupiny C**
  - COVID-19: Přehled vykázaných očkování podle krajů ČR
  - COVID-19: Přehled osob s prokázanou nákazou dle hlášení krajských hygienických stanic

## 3 Implementácia

Projekt je rozdelený do docker kontajnerov zadaných v docker-compose.yaml

### **influxdb**

Slúži na vytvorenie a beh inštalácie nerelačnej databázy Influx DB. Do tejto databázy sa ukladajú všetky dáta cez servisu scraper. Nad dátami sa robia vizualizácie znázorňujúce jednotlivé dotazy zo zadania. A to pomocou služby visualizer.

### **scraper**

Je to systém, ktorý má na starosti stiahnutie, predspracovanie a uloženie dát do Influx DB.

### **visualizer**

Vytvára vizualizácie na uloženými dátami v Influx DB.

### **Dotazy a vizualizácie nad dátami TODO**

## 4 Lokálne spustenie projektu

### Požiadavky na software

- Docker engine 20.10 a vyššie
- Docker compose 1.29 a vyššie

### Vytvorenie všetkých služieb

- `docker-compose build`

### Spustenie jednotlivých služieb

- Najskôr je potrebné spustiť službu pre vytvorenie databázy a používateľa a jej spustenie:  
`docker-compose up influxdb`
- Následne je potrebné spustiť službu pre naplnenie databázy:  
`docker-compose up scraper`
- Pre vygenerovanie potrebných grafov z influx DB je potrebné spustiť (ZATIAH TO NIE JE IMPLEMENTOVANÉ, BUDE TO AŽ SÚČASŤOU 2. ČASTI PROJEKTU):  
`docker-compose up visualizer`

### Zastavenie všetkých bežiacich služieb

- `docker-compose down`

### Odedelený mód

Všetky služby je možné spustiť na pozadí v takzvanom oddelenom móde. Pomocou `-d` prepínaču pre príkaze `docker-compose`.



## Priamy prístup do databáze

Do influx databáze bežiacej v službe sa dá pristupovať priamo a to pomocou príkazov:

- **docker exec -it influxdb bash**
- **influx**

Následne môže robiť používateľ rôzne dotazy nad uloženými dátami.

# Literatúra

- [1] Zadanie projektu [online]. Dostupné z <https://wis.fit.vutbr.cz/FIT/st/cwk.php.cs?title=Projects-topics&csid=768274&id=14826>
- [2] InfluxDB oficiálna dokumentácia [online]. Dostupné z <https://docs.influxdata.com/influxdb/v2.0/>