**Projekt SQL – popis řešení a postup**

V rámci projektu bylo vytvořeno několik pomocných tabulek. Jako výchozí tabulka byla braná covid19\_basic\_differences. Obsahuje 189 států, u kterých jsou uvedeny počty nakažených osob. Tato tabulka se spojila s tabulkou covid\_test, která vznikla z tabulky covid19\_test updatem názvů 5ti zemí (z důvodu sjednocení názvů) a vznikla nová tabulka s názvem zakladni. Poté vznikaly další pomocné tabulky countries\_2, religions\_2, life\_exp\_proj, staty, pocasi, weather\_2. Všechny tyto tabulky sloužily k vytvoření finální tabulky t\_jiri\_drahotsky\_projekt\_SQL\_final.

**Popis jednotlivých tabulek:**

## countries\_2

Aby bylo možné spojovat všechny tabulky, nejprve se upravily hodnoty v tabulce countries.

* Vytvoření kopie tabulky countries
* Oprava dat v tabulce countries\_2 pomocí fce UPDATE
  + Srovnání dat s tabulkou covid19\_basic\_differences tak, aby v obou tabulkách byly stejně značeny země (country)
  + Srovnání dat s tabulkou weather tak, aby v obou tabulkách byly stejně značeny hlavní města

## covid\_test

* Tabulka vytvořená z tabulky covid19\_tests
* V tabulce bylo pomocí fce UPDATE upraveny některé hodnoty v country, aby byly stejné jako v tabulce covid19\_basic\_differences
* Slouží pro sloučení s tabulkou covid19\_basic\_differences a bere se z ní hodnota tests\_performed (počet udělaných testů během jednoho dne)

## religions\_2

* slouží pro výpočet hodnot k procentuálnímu rozložení jednotlivých náboženství v dané zemi
* nejprve se převedou hodnoty všech náboženství z jednoho sloupce do dedikovaných sloupců pro jednotlivá náboženství a vytvoří se dočasná tabulka relig
* poté se spočítá kolik procent dané náboženství vyznává lidí v procentech
  + použije se CASE WHEN a sledují se nulové hodnoty, u těchto hodnot se nepočítají procenta a rovnou se uvádí 0

## life\_exp\_proj

* slouží pro výpočet hodnoty - rozdíl mezi očekávanou dobou dožití v roce 1965 a v roce 2015 - státy, ve kterých proběhl rychlý rozvoj mohou reagovat jinak než země, které jsou vyspělé už delší dobu
* z tabulky life\_expectancy se vezmou hodnoty ve sloupci life\_expectancy pro rok 1965 a 2015 a pomocí LEFT JOIN se tyto hodnoty dají do dedikovaných sloupců a pomocí rozdílu se vypočítá finální hodnota

## zakladni

* pomocná tabulka, která vznikne sloučením tabulek covid19\_basic\_differences a covid\_test přes country a date
* slouží jako základná tabulka, na kterou se poté budou připojovat další tabulky

## staty

* pomocná tabulka pro vytvoření všech požadovaných proměnných v oblasti časových proměnných a specifických proměnných pro daný stát
* vznikne spojením tabulek zakladni a dalších pomocných tabulek

## počasí

* vznikne spojením tabulek weather a countries\_2 přes hlavní města
* upravují se zde hodnoty ve sloupci gust, ze kterého se poté bude brát maximální denní hodnota každý den pro jednotlivý stát
  + nejprve se z hodnoty odstranily znaky km/h pomocí fce TRIM
  + z hodnoty zůstalo jen číslo, které se převedlo na INTEGER, aby bylo možné s tím pracovat jako s číslem a bylo možné později uplatnit fce MAX. Pokud by zůstalo jako text, nefungovalo by správně řazení..
* upravují se hodnoty ve sloupci temp stejným způsobem jako výše – bude sloužit k výčtu průměrné teploty

## weather\_2

* slouží ke zjištění všech proměnných v kategorii počasí

## t\_jiri\_drahotsky\_projekt\_SQL\_final

* finální tabulka
* vznikne spojením tabulek staty a weather\_2

DATA:

Klíčové hodnoty pro spojování tabulek byly country a date. Country se v některých tabulkách lišily a byly tedy pomocí UPDATE narovnány.

Finální tabulka obsahuje informace o počtu nakažených osob pro všech 189 států, které byly uvedeny v tabulce covid19\_basic\_differences. Nejsou zde tedy všechny země, o kterých jsou informace o provedených testech z tabulky covid19\_test (jednalo se o 6 zemí), k těm však nebyly informace o počtu nakažených (nejsou v tabulce covid19\_basic\_differences) – toto jsem však zjistil až před odevzdáním projektu a už jsem nestihl vymyslet, jak spojit tabulky covid19\_basic\_differences a covid19\_test tak, aby zde byly všechny potřebný údaje. Výsledná tabulka je tedy ošizena o vysvětlené proměnné po tyto země.

V tabulce weather jsou informace jen pro 34 zemí. Při spojení se zakladni tabulkou jsou tedy uvedeny hodnoty jen u těchto 34 zemí a u zbytku jsou NULL hodnoty.