**CREATE DATABASE:**

CREATE KEYSPACE [IF NOT EXISTS] population WITH replication = {‘class’: ‘SimpleStrategy’, ‘replication\_factor’: 3};

USE population;

CREATE TABLE vaccinations (data\_somministrazione date, fornitore text, area text, fascia\_anagrafica text, sesso\_maschile int, sesso\_femminile int, prima\_dose int, seconda\_dose int, pregressa\_infezione int, dose\_addizionale\_booster int, codice\_NUTS1 text, codice\_NUTS2 text, codice\_regione\_ISTAT int, nome\_area text, PRIMARY KEY((data\_somministrazione, fornitore, area, fascia\_anagrafica)) );

// Import data from a csv file located in the specified folder (e.g: 'C:\OneDrive\Desktop\prova.csv')

// The same dataset used for ElasticSearch can be used here

COPY vaccinations (data\_somministrazione, fornitore, area, fascia\_anagrafica, sesso\_maschile, sesso\_femminile, prima\_dose, seconda\_dose, pregressa\_infezione, dose\_addizionale\_booster, codice\_NUTS1, codice\_NUTS2, codice\_regione\_ISTAT, nome\_area) FROM 'path-to-csv-file\file.csv' WITH DELIMITER=',' AND HEADER=TRUE;

// for performance and avoiding ALLOW FILTERING in 2° and 3° query

CREATE INDEX region\_area ON vaccinations(area);

CREATE INDEX date\_index ON vaccinations(data\_somministrazione);

// User defined functions must be enabled in “path-to-cassandra-folder\cassandra\conf\cassandra.yaml”, setting the variable enable\_user\_defined\_functions to true

CREATE FUNCTION sum\_different\_columns ( arg1 int, arg2 int )

    RETURNS NULL ON NULL INPUT

    RETURNS int

    LANGUAGE java

    AS $$ return (arg1+arg2); $$;

**QUERIES:**

1) Query data about vaccinations on 28 december 2021 using Moderna on people aged between 40-49 years

SELECT \* FROM vaccinations WHERE data\_somministrazione='2021-12-28' AND fornitore='Moderna' AND fascia\_anagrafica='40-49' ALLOW FILTERING;

2A) Find the total of doses inoculated in Lombardia

SELECT area, sum\_different\_columns(SUM(sesso\_maschile), SUM(sesso\_femminile)) AS total\_vaccination FROM vaccinations WHERE area='LOM' ALLOW FILTERING;

2B) Find the total of doses inoculated in Lombardia on men and women

SELECT SUM(sesso\_maschile) AS man\_total, SUM(sesso\_femminile) AS women\_total FROM vaccinations WHERE area='LOM' ALLOW FILTERING;

3) Find for which vaccine type and on which region and age group were inoculated the most doses of booster on 28 december 2021

SELECT MAX(dose\_addizionale\_booster) AS max\_booster, area, fascia\_anagrafica, fornitore FROM vaccinations WHERE data\_somministrazione = '2021-12-28';

**COMMANDS:**

1) Insert a new record on the database

INSERT INTO vaccinations(data\_somministrazione, fornitore, area, fascia\_anagrafica, sesso\_maschile, sesso\_femminile, prima\_dose, seconda\_dose, pregressa\_infezione, dose\_addizionale\_booster, codice\_NUTS1, codice\_NUTS2, codice\_regione\_ISTAT, nome\_area)

VALUES ('2021-12-29','Pfizer/BioNTech','EMR','20-29',1,0,0,0,0,1,'ITH','ITH5',8,'Emilia-Romagna');

2) Update a record on the database using data\_somministrazione, fornitore, area and fascia\_anagrafica

UPDATE vaccinations

 SET sesso\_maschile= 2, dose\_addizionale\_booster=2

 WHERE data\_somministrazione = '2021-12-29' AND fornitore='Pfizer/BioNTech' AND area ='EMR' AND fascia\_anagrafica='20-29';