COMP1204: Data Management
Coursework Two: COVID-19 Coronavirus Data

Wei_Guo
33331626

May/9/2022

1 The Relational Model

1.1 EX1

In this subsection, the datatype will be displayed for every column title according to the file of the dataset.csv.

$$
\begin{aligned}
dataset.csv(&TEXT : dateRep, && (1)\\
&INTEGER : day, && (2)\\
&INTEGER : month, && (3)\\
&INTEGER : year, && (4)\\
&INTEGER : cases, && (5)\\
&INTEGER : deaths, && (6)\\
&TEXT : countriesAndTerritories, && (7)\\
&TEXT : geoId, && (8)\\
&TEXT : countryterritoryCode, && (9)\\
&INTEGER : popData2020, && (10)\\
&TEXT : continentExp && (11)\\
&) && (12)
\end{aligned}
$$

1.2 EX2

At here, the minimal set of Functional Dependencies will be shown.
{day, month, year} → {dateRep}
{geoId, dateRep} → {cases}
{countriesAndTerritories, dateRep} → {deaths}
{countryterritoryCode, dateRep} → {deaths}
{geoId} → {countriesAndTerritories}
{countriesAndTerritories} → {continentExp}
{countryterritoryCoder} → {continentExp}
{countriesAndTerritories, dateRep} → {cases}
{countryterritoryCode, dateRep} → {cases}
{geoId, dateRep} → {deaths}
{countryterritoryCode} → {popData2020}
{countryterritoryCode} → {countriesAndTerritories}
{geoId} → {continentExp}

1.3 EX3

This subsection show the all potential candidate keys, these are:
{dateRep, countriesAndTerritories},
{dateRep, geoId},
{dateRep, countryterritoryCode}

1.4 EX4

Here, the primary key will be set, that is {dateRep, countriesAndTerritories}

2 Normalisation

2.1 EX5

The original table was broken and be set table one {dateRepresentation}, table two {countriesAndTerritories}, and retain the original undecomposed data.
To be specific, the context of the dateRepresentation is composed of the column of day, month, year. Beside, the context of the countriesAndTerritories is composed of the column of geoId, countryterritoryCode, popData2020 and continentExp. And finally, the retain context of the original undecomposed data are composed of the column dateRep, cases, deaths countriesAndTerritories .
Then i drawed a buleprint to show:
Date(dateRep, day, month, year)
Country(countriesAndTerritories, geoId, countryterritoryCode, popData2020 and continentExp)
COVID-19 Coronavirus Data(dateRep, cases, deaths countriesAndTerritories).

2.2 EX6

The relation of the 2FN will be displayed in EX6.

$$
\begin{aligned}
Date(\quad & TEXT : dateRep, && (13)\\
& INTEGER : day, && (14)\\
& INTEGER : month, && (15)\\
& INTEGER : year && (16)\\
) && (17)\\
Country(\quad & TEXT : countriesAndTerritories, && (18)\\
& TEXT : geoId, && (19)\\
& TEXT : countryterritoryCode, && (20)\\
& TEXT : continentExp && (21)\\
& INTEGER : popData2020 && (22)\\
) && (23)\\
CoronavirusData(\quad & TEXT : dateRep, && (24)\\
& INTEGER : cases,\ \&INTEGER : deaths, && (25)\\
& TEXT : countriesAndTerritories && (26)\\
) && (27)\\
& && (28)
\end{aligned}
$$

I set dateRep and countriesAndTerritories are foreign key to together build CoronavirusData, and also we can see the column of the dateRep and the countriesAndTerritories are Compound primary key in CoronavirusData.

2.3 EX7

According the existing table, the transitive dependencies will be taken, these are:
countryterritoryCode → popData2020
Explain: the field of the popData2020 are not dependent on the key(countriesAndTerritories), so the transitive dependency should be countryterritoryCode → popData2020

2.4 EX8

$$
\begin{aligned}
Country(\quad & TEXT : countriesAndTerritories, && (29)\\
& TEXT : geoId, && (30)\\
& TEXT : countryterritoryCode && (31)\\
) & && (32)\\
The_Number_Of_People(\quad & TEXT : countryterritoryCode, && (33)\\
& INTEGER : popData2020 && (34)\\
) & && (35)\\
Date(\quad & TEXT : dateRep, && (36)\\
& INTEGER : day, && (37)\\
& INTEGER : month, && (38)\\
& INTEGER : year && (39)\\
) & && (40)\\
CoronavirusData(\quad & TEXT : dateRep, && (41)\\
& INTEGER : cases, && (42)\\
& INTEGER : deaths, && (43)\\
& TEXT : countriesAndTerritories && (44)\\
) & && (45)\\
& && (46)
\end{aligned}
$$

Explain: I set countryterritoryCode as a foreign key to dependent popData2020.

2.5 EX9

My relations suit Boyce-Codd Normal Form.
The construction is Country, $The_Number_Of_People, Date, CoronavirusData$.

3 Modelling

3.1 EX10

In this section, i import the raw dataset into SQLite into a single table, and it named coronavirus.db, then i attached to the specific code in below

$$
\begin{aligned}
CREATETABLEdataset(\quad & "dateRep"TEXT, && (47)\\
& "day"INTEGER, && (48)\\
& "month"INTEGER, && (49)\\
& "year"INTEGER, && (50)\\
& "cases"INTEGER, && (51)\\
& "deaths"INTEGER, && (52)\\
& "countriesAndTerritories"TEXT, && (53)\\
& "geoId"TEXT, && (54)\\
& "countryterritoryCode"TEXT, && (55)\\
& "popData2020"INTEGER, && (56)\\
& "continentExp"TEXT && (57)\\
); && (58)\\
& && (59)
\end{aligned}
$$

Next step, i dumped the entire database as dataset.sql.

3.2 EX11

The main task in here is to create the full normalised representation, including all additional tables (with correct types) with no data and excluding the dataset table. In here, i set dateRep, countriesAndTerritories as a Foreign Key, and they are represented the Primary Key as well. And, the column of the countryterritoryCode and popData2020 are allocated to countryterritoryCode, a Foreign Key in the table of Country.

$$CREATETABLECountry(\ countriesAndTerritoriesTEXTPRIMARYKEYNOTNULL, \tag{60}$$
$$geoIdTEXTNOTNULL, \tag{61}$$
$$countryterritoryCodeTEXTNOTNULL, \tag{62}$$
$$continentExpTEXTNOTNULL, \tag{63}$$
$$FOREIGNKEY(countryterritoryCode)REFERENCES \tag{64}$$
$$@TheNumberOfPeople(countryterritoryCode) \tag{65}$$
$$@\ ON\ UPDATE\ CASCADE\ ON\ DELETE\ SET\ NULL \tag{66}$$
$$); \tag{67}$$
$$CREATETABLETheNumberOfPeople(\ countryterritoryCodeTEXTPRIMARYKEYNOTNULL, \tag{68}$$
$$popData2020INTEGERNOTNULL \tag{69}$$
$$); \tag{70}$$
$$CREATETABLEDate(\ dateRepTEXTPRIMARYKEYNOTNULL, \tag{71}$$
$$dayINTEGERNOTNULL, \tag{72}$$
$$monthINTEGERNOTNULL, \tag{73}$$
$$yearINTEGERNOTNULL \tag{74}$$
$$); \tag{75}$$
$$CREATETABLECoronavirusData(\ dateRepTEXTNOTNULL, \tag{76}$$
$$casesINTEGERNOTNULL, \tag{77}$$
$$deathsINTEGERNOTNULL, \tag{78}$$
$$countriesAndTerritoriesTEXTNOTNULL, \tag{79}$$
$$FOREIGNKEY(dateRep)REFERENCESDate(dateRep) \tag{80}$$
$$@ONUPDATECASCADEONDELETESETNULL, \tag{81}$$
$$FOREIGNKEY(countriesAndTerritories) \tag{82}$$
$$@REFERENCESCountry(countriesAndTerritories) \tag{83}$$
$$@ONUPDATECASCADEONDELETESETNULL, \tag{84}$$
$$PRIMARY\quad KEY(dateRep, countriesAndTerritories) \tag{85}$$
$$); \tag{86}$$
$$\tag{87}$$

The sign "@" represents to pick up the above content.

3.3 EX12

Here, in order to insert the context, i use INSERT statements and using SELECT to populate the new tables from the 'dataset' table.

(1)
INSERT INTO CoronavirusData (dateRep, cases, deaths, countriesAndTerritories)
SELECT dateRep,cases,deaths,countriesAndTerritories
FROM dataset;

(2)
INSERT INTO Country (countriesAndTerritories ,geoId, countryterritoryCode, continentExp)
SELECT DISTINCT countriesAndTerritories, geoId, countryterritoryCode, continentExp
FROM dataset;

(3)
INSERT INTO Date (dateRep, day, month, year)
SELECT DISTINCT dateRep, day, month, year
FROM dataset;

(4)
INSERT INTO TheNumberOfPeople (countryterritoryCode,popData2020)
SELECT DISTINCT countryterritoryCode,popData2020
FROM dataset;

(5)
DELETE FROM CoronavirusData where countriesAndTerritories = 'countriesAndTerritories';
DELETE FROM Country where countriesAndTerritories = 'countriesAndTerritories';
DELETE FROM Date where dateRep = 'dateRep';
DELETE FROM TheNumberOfPeople where
countryterritoryCode = 'countryterritoryCode';

3.4 EX13

In here, the main task is to test and ensure that on a clean SQLite database.

4 Querying

4.1 EX14

The worldwide total number of cases and deaths
SELECT sum(cases) AS TotalCases FROM CoronavirusData;
SELECT sum(deaths) AS TotalDeaths FROM CoronavirusData;

4.2 EX15

The number of cases and the date, by increasing date order, for the United Kingdom
SELECT cases, dateRep, day, month, year,countriesAndTerritories
FROM dataset
WHERE countriesAndTerritories LIKE 'United$_K ingdom'$
$ORDERBY year ASC, month ASC, day ASC$

4.3 EX16

The number of cases, deaths and the date, by increasing date order, for each continent
SELECT Country . continentExp AS Continent,cases, deaths,Date.dateRep, year, month, day

FROM Country, CoronavirusData, Date

4.4 EX17

The number of cases and deaths as a percentage of the population, for each country
SELECT countriesAndTerritories as country,
CAST(sum(cases)AS double) *100 / CAST(sum(popData2020)AS double)*100 as thePercentCases,
CAST(sum(deaths)AS double) *100 / CAST(sum(popData2020)AS double)*100 as thePercentDeaths
FROM dataset
GROUP BY countriesAndTerritories;

4.5 EX18

A descending list of the the top 10 countries, by percentage deaths out of cases
COMMAND:
SELECT countriesAndTerritories AS country,
CAST(sum(cases)AS double) *100 / CAST(sum(popData2020)AS double)*100 as thePercentCases,
CAST(sum(deaths)AS double) *100 / CAST(sum(popData2020)AS double)*100 as thePercentDeaths
FROM dataset
GROUP BY countriesAndTerritories LIMIT 10;

4.6 EX19

The date against a cumulative running total of the number of deaths by day and cases by day for the united kingdom
COMMAND:
SELECT sum(day) AS Date,sum(deaths) AS TotalDeaths,sum(cases) AS TotalCases
FROM dataset
WHERE countriesAndTerritories LIKE 'United$_K ingdom'$

5 Extension

5.1 EX20