CSC8017 | Database Systems

Nazli Aldogan | 230404296

**Coursework 2**

i. (preparatory work)

ii.

a.

**SELECT** ID

**FROM** Stand

**WHERE** ID **NOT** **IN** (**SELECT** **DISTINCT** stand **FROM** BusRoute);

A screenshot of a computer

Description automatically generated

b.

**SELECT** **DISTINCT** b.**type**

**FROM** Bus b

**WHERE** b.**length** > (**SELECT** maxLength **FROM** Stand **WHERE** ID = 'F')

**ORDER** **BY** **type**;

A screenshot of a computer

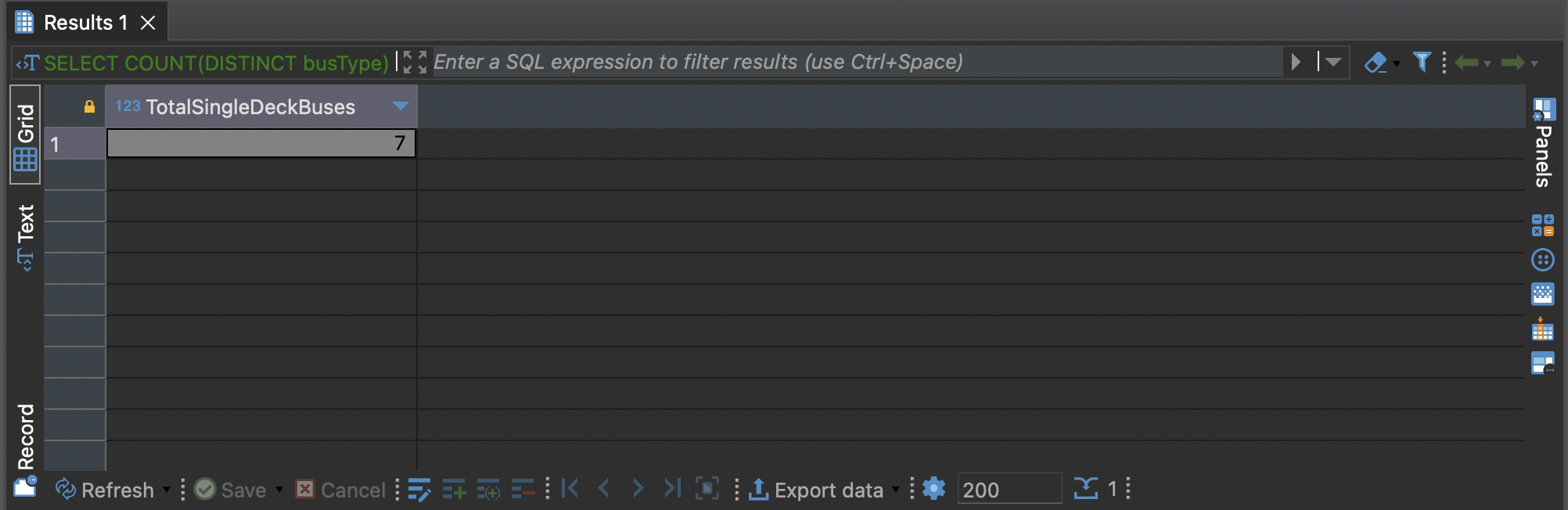
Description automatically generated

c.

**SELECT** **COUNT**(**DISTINCT** busType) TotalSingleDeckBuses

**FROM** Uses

**WHERE** busType **LIKE** '%single-deck%';



d.

**SELECT** **SUM**(R.frequency)

**FROM** BusRoute R

**JOIN** OperatesRoute ORT **ON** R.**number** = ORT.routeNumber

**WHERE** ORT.operatorName = 'United Services';

A screenshot of a computer

Description automatically generated

e.

**SELECT** **DISTINCT** B.**type**

**FROM** Bus B

**JOIN** Uses U **ON** B.**Type** = U.busType

**JOIN** OperatesRoute ORT **ON** U.operatorName = ORT.operatorName

**WHERE** ORT.routeNumber = '602';

A screenshot of a computer

Description automatically generated

f.

**SELECT** **DISTINCT** R.destination

**FROM** BusRoute R

**JOIN** Bus B **ON** R.EV ='Yes' **AND** B.**type** ='IE';

A screenshot of a video

Description automatically generated

iii.

a. This approach would cause an issue for the buses with multiple destinations/stops, as we wouldn’t be able to store more than one destination for a bus. This results in an inaccuracy for the data representation. In a real-world scenario, especially for say, buses in public transportation, buses would have more than one stop and people would leave at any of the stops, not only the destination (last) stop. So, as we don’t know the other stops that lead to the destination stop, this is a problem. In addition, managing data with the route changes would also get difficult.

b.

IDsOfStopsServed and NamesOfStopsServed would have more than one value on their columns which would violate the atomic value requirement of the 1NF rule. We need to ensure the 1NF rule is preserved first in order to check whether the 2NF rule is preserved, as it wouldn’t be achievable without meeting the 1NF rule; and we need to ensure the 2NF rule is preserved first in order to check whether the 3NF rule is preserved, as it wouldn’t be achievable without meeting the 2NF rule. Because 1NF rule is not met in our condition, we cannot move further to check whether 2NF rule is met, and then 3NF rule is met: as 1NF rule is not met, none of the rules are met for the data.

c.

**CREATE** **TABLE** BusRouteDetails (

number **INT** **PRIMARY** **KEY**,

frequency **INT**,

stand **INT**,

EV **VARCHAR**(3),

**FOREIGN** **KEY** (stand) **REFERENCES** Stand(ID)

);

**CREATE** **TABLE** BusRouteStops (

number **INT**,

stop\_number **INT**,

stop\_name **VARCHAR**(50),

isLastDestination **VARCHAR**(3),

**PRIMARY** **KEY** (number, stop\_number),

**FOREIGN** **KEY** (number) **REFERENCES** BusRouteDetails(number)

);

BusRouteDetails: (number (PK), frequency, stand(FK), EV)

BusRouteStops: (number(PK)(FK), stop number(PK), stop name, isLastDestination)

Here, I make two independent tables instead of one BusRoute table. We can call these two tables BusRouteDetails and BusRouteStops. They will have atomic values therefore 1NF is preserved. Non-key attributes depend on the primary key therefore 2NF is preserved. Finally, as these attributes depend solely on the primary key 3NF is also preserved.

iv.

a. Using a relational database is the most appropriate choice compared to the others because these databases can work with structured data really well. As sensitive data, personal details such as name, surname, financial data, account data, etc. is stored, in terms of data integrity, security, accuracy and independence, the relational database would be the most viable option. As this model also supports complex queries, we would be able to keep our records and make our financial calculations properly.

b. Using a document database would be the most appropriate choice for this situation as document databases can deal with unstructured data well as well as scale it. So, the database would be able to work with a dataset as large as 1 million parts, and deal with bus part details & specifications as the unstructured data for our case.