Project Documentation
Data Warehouses
Data Engineering
Fourth Semester 2023/2024

Mikołaj Woźniak 190473,

Jakub Szurmak 193095

1.1 Relational database topic

Our relational database's goal is to store data about bus courses in a bus transport organization. Database allows operators to check what drivers are or were hired in the company, get details about type of buses and their variants, check bus stations and get insights about repairs of buses and type of services used in repair. In summary, a brief overview of the company’s work.

1.2. Database description

Crucial aspect is to create a reliable source of information about courses held by the organization. The most valuable information which should be included in the database are:

- What driver and bus were responsible for particular course
- Information if the course was completed or not.
- Information if a course was a continuation of an unfinished course.
- What buses were repaired and what services were needed to fix a mechanical issue.
- Which stations were included in the particular course.
- Information about organization’s fleet (buses and their types)

1.3 Simplifications and conditions in the database

a) Course is not assigned to one particular bus number (three cruises of bus line 122 for one day is equal to three different courses)

b) One driver can be responsible for many courses, but one course can be done only by one driver

c) Courses which are supposed to finish an uncompleted one, can have a different driver than in original course

d) Two courses with the same bus number(bus line) can have different number of stops, it depends on type of the course (uncompleted course, continuity or normal one)

e) Drivers can be only of polish nationality

f) Course, which is a continuation of the uncompleted one, cannot have a technical malfunction

g) Bus line can have maximally four different courses on one day

h) Single course is held during a single day (Course started at 26.04 must end on 26.04)

1.4 Users of the database

a) Bus transportation company

b) National Agencies (Government, Department of Infrastructure)

c) Bus producing companies

d) GUS

1.5 Example of Queries

a) Check which driver has held the most courses

b) Check which bus cruise(particular bus line) was uncompleted most oftenly

c) Check which stations are used most often

d) Check which bus cruise (particular bus line) was completed the most often (no mechanical or any other failures)

e) Check which bus variant was repaired most often

f) Check what service was needed the most

g) Check which bus variant has not completed the course most often

1.6 Example scenario of adding a new course.

1) Beginning of the course, data like bus number, name of the line (direction), driver's PESEL, and VIN of the bus are saved

2) After stopping on the station, a tuple of (courseID, stationID) is added to the stops entity set

3) If any unexpected situation does not occur during the course, we confirm that particular course is completed

3.1) If technical malfunction occurs, we confirm that the course requires assistance

3.2) New course must be created, data like bus number, name of the line (direction), driver's PESEL, and VIN of the bus are saved, and we confirm that this course is a continuation

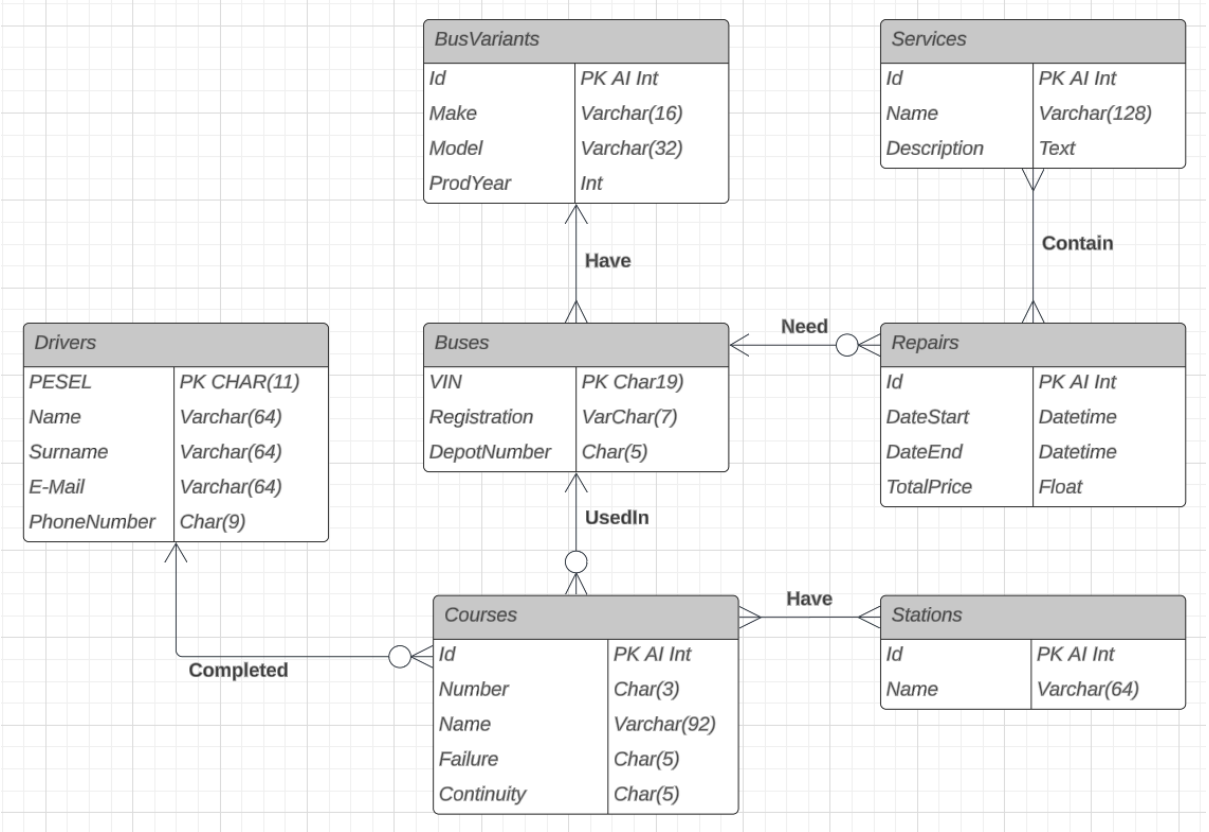
3.3) Initial station of the new course is the last station of the uncompleted course

4) Data is loaded into the database

1.7 Example of usage of the database

A. Company wants to check the correlation between driver and number of failures
B. Company wants to check the correlation between the bus's variant and number of failures
C. Company wants to minimize repair costs:
    a. Focus on acquiring bus variants that fail least often
    b. Focus on acquiring bus variants that are the cheapest to repair
    c. Focus on minimizing repair times:
        i. See which bus variant is taking the longest to repair
D. Company wants to decompress it's traffic by looking at the number of usages per station

2. ERD diagram



3. Description of entity sets

Table	Attribute	Attribute Type	Description
BusVariants	ID	PK AI Int	Auto increment from 1 by 1.
	Make	VarChar(96)	Name of buses manufacturer
	Model	VarChar(164)	
	ProdYear	Int	Year of buses production
Buses	VIN	PK Char(19)	
	Registration	VarChar(11)	Registration Number
	DepotNumber	Char(5)	Distinct number given to the bus by the bus transportation company.
Courses	Id	PK AI Int	Auto increment from 1 by 1.
	Number	Char(3)	Bus line number
	Name	VarChar(92)	Name of the line, direction.
	Failure	VarChar(5)	“True” there was a failure, “False” no failures.

	Continuity	VarChar(5)	“True” if the current course is a continuation of a course on which a failure occurred.
Stations	Id	PK AI Int	Identification number of station auto incremented from 1 by 1.
	Name	VarChar(64)	Unique name of a bus stop.
Drivers	PESEL	Char(11)	
	Name	VarChar(64)	First name of driver
	Surname	VarChar(64)	Last name of driver
	E-Mail	VarChar(64)	E-mail contact to driver
	PhoneNumber	Char(9)	Driver’s phone number
Repairs	Id	PK AI Int	Auto increment from 1 by 1.
	DateStart	Datetime	Date of beginning repair services.
	DateEnd	Datetime	Date of ending repair services.
	TotalPrice	Float	Cost of repair, accounting purposes.
Services	Id	PK AI Int	Auto increment from 1 by 1.

	Name	Varchar(128)	Simple name of performed service.
	Description	Text	Description of observed issue and how it got fixed.

4. Description of relationships:

Name of relationship	Type of relationship	Entity set nr 1	Entity set nr 2	Description
Have	1:1...n	BusVariants	Buses	One bus can have only one bus variant, but many buses in the fleet can be the same type(variant)
Need	1:0...n	Buses	Repairs	One bus can have zero or many repairs, a singular repair is the fact of a bus arriving at the workshop for fixing.
Contain	n:n	Repairs	Services	Many services can be a part of one repair and many repairs can have the same sets of services.
UsedIn	1:0...n	Buses	Courses	Single course is always executed by one bus, but one bus can be used in many

				different courses
Completed	1:0...n	Drivers	Courses	Single course is held by one driver, but one driver can do many courses (in total) or zero (if it is a recently hired driver)
Have	n:n	Courses	Stations	A station can be a part of multiple courses and a course always has more than one station.

5. RDB Diagram

Drivers (PESEL PRIMARY KEY, Name, Surname, E-Mail, PhoneNumber)
BusVariants(Id PRIMARY KEY, Make, Model, ProdYear)
Buses (VIN PRIMARY KEY, Registration, DepotNumber, BusVariantID REF BusVariants.Id)
Repairs (Id PRIMARY KEY, DateStart, DateEnd, TotalPrice, BusVIN REF Buses.VIN)
Services (Id PRIMARY KEY, Name, Description)
Usages ((RepairsID REF Repairs.Id, ServicesID REF Services.Id) PRIMARY KEY)
Courses (Id PRIMARY KEY, Number, Name, Failure, DriverPesel REF Drivers.PESEL, BusVIN REF Buses.VIN)
Stations (Id PRIMARY KEY, Name)
Stops ((CourseID REF Courses.Id, StationID REF Stations.Id) PRIMARY KEY)

Stops is the implementation of n-n relationship in between Courses and Stations,
Usages is the implementation of n-n relationship in between Repairs and Services