Bus Line Cruise - Data warehouse design

Business process

Data warehouse is designed to store information about made stops during a particular bus line course. The process is described in the document ‘Specification of business processes’.

Relational database schema

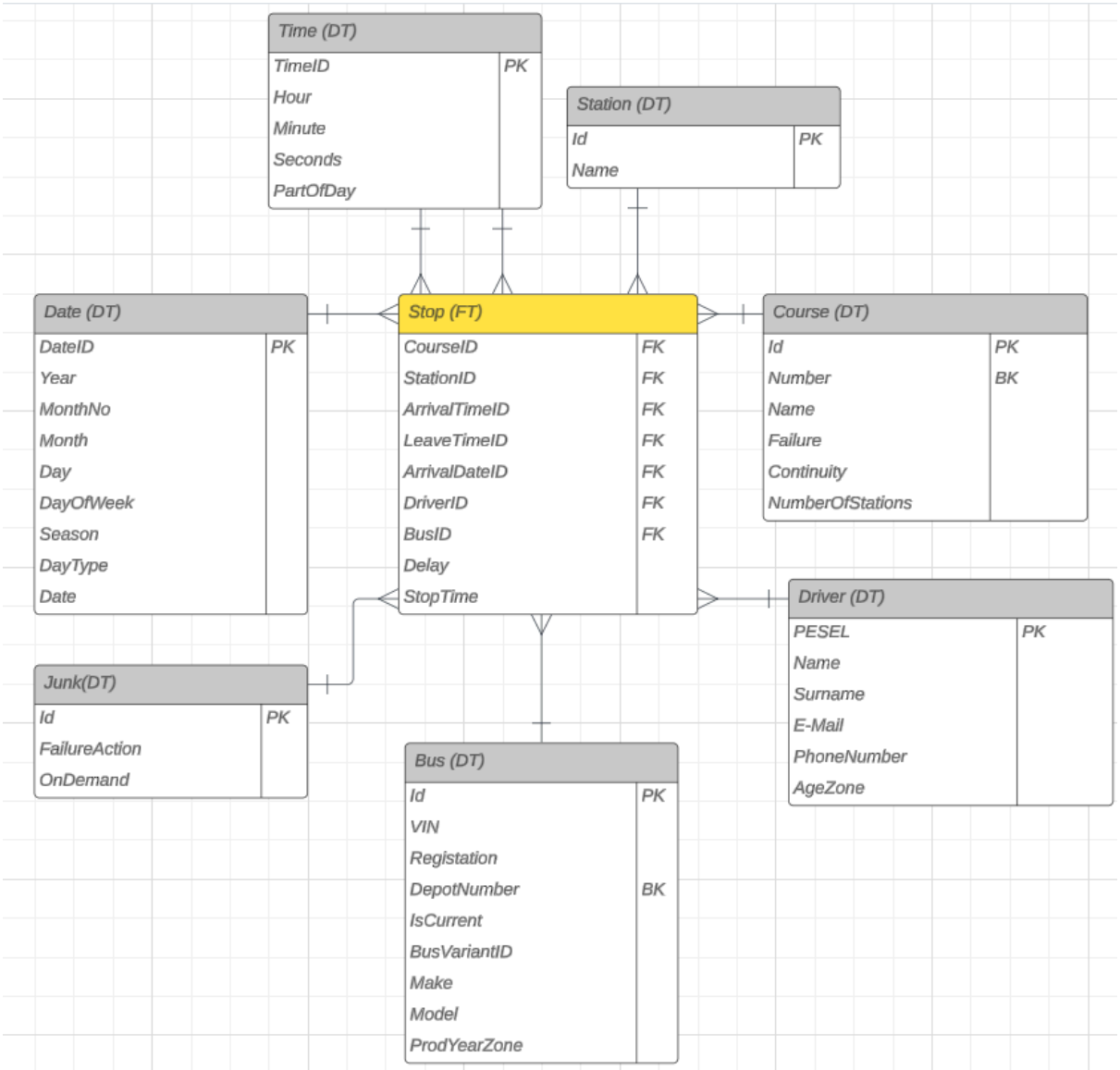

TABLE NAME	ATTRIBUTE	ATTRIBUTE TYPE	DESCRIPTION
STOP	One tuple describes one fact of bus stopping at the station		

Data warehouse: Bus line cruise data warehouse design

(FACT TABLE)			
	CourseID	Numeric	FK, Courses Id of the course
	StationID	Numeric	FK Stations Id of the station
	ArrivalTimeID	Numeric	FK Time Schedule time of arrival
	LeaveTimeID	Numeric	FK Time Time of actual departure
	ArrivalDateID	Numeric	FK Date Date of the course
	DriverID	Numeric	FK Driver Id of the driver who is responsible for the course
	BusID	Numeric	FK Bus Id of the bus that is riding the course.
	Delay	Numeric	Delay(deviation from planned arrival time in seconds) on the particular station (not sum of the delays)
	StopTime	Numeric	How many seconds bus has been on the station
COURSE (DIMENSION TABLE)	One tuple describes a single course with the specified bus number and specified name of the course. A singular course always takes place within 1 day. Night courses don't go through midnight into the next day. (three courses made by bus number 122 = 3 different tuples)		
	ID	Numeric	PK (surrogate key)
	Number	Char(3)	Bus 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, “False” otherwise.
	NumberOfStations	VarChar(20)	Three options: - less than 10 - from 10 to 20 - more than 20
DRIVER (DIMENSION TABLE)	One tuple describes one driver in the specified age category		
	PESEL	Char(11)	PK, Pesel number of the driver
	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
	AgeZone	VarChar(20)	Four options: - 24-30 - 31-45 - 46-60 - >60
BUS (DIMENSION TABLE)	One tuple describes one bus in the specified production year category, with the specified bus variant, depot number, registration number(license plate), make and model (Slowly changing dimension - 2 SCD)		
	Id	Numeric	PK (surrogate key)
	VIN	Char(19)	Identification number of the bus
	Registration	VarChar(11)	Registration Number
	DepotNumber	Char(5)	BK, Distinct number given to the bus by the bus transportation

			company.
	IsCurrent	Char(3)	Attribute which shows, if particular tuple active or not
	Make	VarChar(96)	Name of buses manufacturer
	Model	VarChar(164)	Model of the bus
	ProdYearZone	VarChar(20)	Production year of the bus. Allowed values: - <1980 - 1981-1990 - 1991-2000 - 2001-2010 - 2011-2020 - >2021
STATION (DIMENSION TABLE)	One tuple describes one station		
	Id	Numeric	PK (Surrogate key)
	Name	VarChar(64)	Unique name of a bus stop.
JUNK (DIMENSION TABLE)	The tuples correspond to “all” possible combinations of values for FailureAction and OnDemand		
	IdJunk	Numeric	PK (Surrogate key)
	FailureAction	Char(4)	Three possible options: **“None”** - everything is okay **“Take”** - course is continuity of unfinished one **“Give”** - course is unfinished
	OnDemand	Char(15)	Information if stop is on demand or not
DATE (DIMENSION TABLE)	One tuple describes one day		
	DateID	Numeric	PK
	Year	Numeric	Year

	MonthNo	Numeric	Month's numeric value
	Month	Numeric	Month's name "January", "February",…
	Day	Numeric	Day in the month
	DayOfWeek	VarChar(9)	Day of week. Allowed values: Monday, Tuesday, Wednesday, Thursday, Friday, Saturday and Sunday
	Season	VarChar(6)	Season during a particular date. Allowed values: Winter, Spring, Summer and Autumn
	DayType	VarChar(7)	Working day. Allowed values: "Working", "Free"
	Date	VarChar(10)	Whole date in a format - dd.mm.yyyy
TIME (DIMENSION TABLE)	One tuple describes one hour (independently on date)		
	TimeID	Numeric	PK
	Hour	Numeric	Hour, Allowed values from 0 – 23.
	Minute	Numeric	Minute, Allowed values from 0-59
	Second	Numeric	Second, Allowed values from 0-59
	PartOfDay	Varchar(5)	Time of day. Allowed values: between 0 and 6, between 6 and 9, between 10 and 14, between 14 and 18, between 18 and 23).

Dimensional model

Fact definitions

Fact 1 Making a Stop fact: Bus stopping at the given bus station at the given course, in the given bus, at the given time, at the given station by the given driver.

Fact table: Stop

Granularity:

- a specific course with the specified bus number and specified name of the course
- a specific planned arrival hour
- a specific date of bus stopping at the bus station
- a specific departure time
- a specific driver in the specified age category
- a specific bus in the specific production year category, with the specified bus variant, depot number, registration number, make and model
- a specific station

Measures and aggregation function:

Number of stop facts - COUNT (1)
Delay of the bus - SUM(delay)
Average delay of the bus - SUM(delay)/COUNT(1)
Time spent on the station during one course - SUM(StopTime)

Dimension definitions

Dimensions for Fact 1 Stop fact:

DIMENSION/DIMENSION ATTRIBUTE	TABLE/COLUMN	TYPE
DATE HIERARCHY	-Date.Year –Date.Season —Date.Month —-Date.Day —--Date.DayOfWeek —---Date.DayType	Hierarchical dimension
DATE	Date	Dimension
DATE.YEAR	Date.Year	Dimension attribute

Data warehouse: Bus line cruise data warehouse design

DATE.MONTH	Date.Month	Dimension attribute
DATE.DAY	Date.Day	Dimension attribute
DATE.DAYOFWEEK	Date.DayOfWeek	Dimension attribute
DATE.SEASON	Date.Season	Dimension attribute
DATE.DAYTYPE	Date.DayType	Dimension attribute
TIME	-Time.PartOfDay –Time.Hour —Time.Minute -—Time.Seconds	Hierarchical dimension
ArrivalTime	Time	Dimension
ARRIVALTIME.TIMEID	Time.TimeID	Dimension attribute
ARRIVALTIME.HOUR	Time.Hour	Dimension attribute
ARRIVALTIME.MINUTE	Time.Minute	Dimension attribute
ARRIVALTIME.SECONDS	Time.Seconds	Dimension attribute
ARRIVALTIME.PARTOFDAY	Time.PartOfDay	Dimension attribute
LeaveTime	Time	Dimension
LEAVETIME.TIMEID	Time.TimeID	Dimension attribute
LEAVETIME.HOUR	Time.Hour	Dimension attribute
LEAVETIME.MINUTE	Time.Minute	Dimension attribute
LEAVETIME.SECONDS	Time.Seconds	Dimension attribute
LEAVETIME.PARTOFDAY	Time.PartOfDay	Dimension attribute
STATION	Station	Dimension
STATION.Id	Station.Id	Dimension attribute
STATION.NAME	Station.Name	Dimension attribute
COURSE	Course	Dimension
COURSE.ID	Course.Id	Dimension attribute
COURSE.NUMBER	Course.Number	Dimension attribute
COURSE.NAME	Course.Name	Dimension attribute
COURSE.FAILURE	Course.Failure	Dimension attribute
COURSE.CONTINUITY	Course.Continuity	Dimension attribute

COURSE.NUMBEROFSTATIONS	Course.NumberOfStations	Dimension attribute
DRIVER	Driver	Dimension
DRIVER.PESEL	Driver.PESEL	Dimension attribute
DRIVER.NAME	Driver.Name	Dimension attribute
DRIVER.SURNAME	Driver.Surname	Dimension attribute
DRIVER.E-MAIL	Driver.E-Mail	Dimension attribute
DRIVER.PHONENUMBER	Driver.PhoneNumber	Dimension attribute
DRIVER.AGEZONE	Driver.AgeZone	Dimension attribute
BUS	Bus	Dimension
BUS	-Bus.Make –Bus.Model —Bus.ProdYearZone	Hierarchical Dimension
BUS.VIN	Bus.VIN	Dimension attribute
BUS.REGISTRATION	Bus.Registration	Dimension attribute
BUS.DEPOTNUMBER	Bus.DepotNumber	Dimension attribute
BUS.ISACTIVE	Bus.IsActive	Dimension attribute
BUS.BUSVARIANTID	Bus.BusVariantID	Dimension attribute
BUS.MAKE	Bus.Make	Dimension attribute
BUS.MODEL	Bus.Model	Dimension attribute
BUS.PRODYEAR	Bus.ProdYearZone	Dimension attribute

Checking the feasibility of queries based on the multidimensional model

1. Does the type of the bus have an impact on delays?
    - Measure: Delay of the bus,
    - Dimension: Bus (dimension attribute: busVariant)
2. Compare the duration of delays on working days and non working days
    - Measure: Delay of the bus,
    - Dimension: Date (dimension attribute: DayType)
3. Compare the average delay time between daytime and night courses
    - Measure: Average delay of the bus,
    - Dimension: Courses (dimension attribute: Number)
4. Measure the average delay among all bus drivers.

   - Measure: Average delay of the bus
   - Dimension: Driver (dimension attribute: PESEL)
5. Compare the duration of delays during rush and normal hours.
   - Measure: Delay of the bus,
   - Dimension: Time(dimension attribute: Hour)
6. Does the age of the bus have an impact on delays?
   - Measure: Delay of the bus,
   - Dimension: Bus (dimension attribute: ProdYearZone)
7. Do longer courses have bigger average delays?
   - Measure: Average delay of the bus,
   - Dimension: Courses (NumberOfStations)
8. Does the age of the driver have an impact on avg delays?
   - Measure: Average delay of the bus,
   - Dimension: Drivers (AgeZone)
9. Does the type of the bus have an impact on failures?
   - Measure: Number of facts
   - Dimension: Junk Table (FailureAction)
   - Dimension: Bus (BusVariant)
10. Does the age of the bus have an impact on failures?
   - Measure: Number of facts
   - Dimension: Junk Table (FailureAction)
   - Dimension: Bus (ProdYearZone)
11. Do longer courses spend more time on stations?
   - Measure: Time spent on the station during one course
   - Dimension: Course (NumberOfStations)
12.

Checking if there are Data in the Data sources needed to fill the Data warehouse

TABLE NAME	**COLUMN**	**SOURCE**
STOP	One tuple describes one fact of bus stopping at the station	
	CourseID	Course ID. Foreign key from dimension table. Based on Column A (Identification number of course in which the stop is contained) in ZTMmotion file

	StationID	Station ID. Foreign key from dimension table. Based on column B (Identification number of a bus stop) in the ZTMotion file.
	DriverID	Driver ID. Foreign key from dimension table. Based on the driver's PESEL number and specified age range. Based on driver entity set in BasedZTM source
	BusID	Bus ID. Foreign key from dimension table. Based on bus depot number, VIN number, variant and specified range of production year. Based on Buses entity set in BasedZTM source
	ArrivalTimeID	Arrival time Id. Foreign key from dimension table. Based on Column D (Planned time arrival) in ZTMmotion file
	ArrivalDateID	Arrival date ID. Foreign key from dimension table. Based on Column D (Planned time arrival) in ZTMmotion file
	LeaveTimeID	Arrival time Id. Foreign key from dimension table. Calculated from Column D (Planned time arriva), Column E (Delay) and Column F (Stop time) in the ZTMotion file
	Delay	Delay of the bus on the particular stop (not sum of the delays) taken from Column E (Delay) in the ZTMotion file
	StopTime	Time spent on the stop taken from Column F (Stop time) in the ZTMotion file
COURSE	One tuple describes a single course with the specified bus number and specified name of the course. (three courses made by bus number 122 = 3 different tuples)	

	Id	Id of course. Surrogate key - generated by database
	Number	Bus number - taken from Number in Courses table in BasedZTM source
	Name	Name of the line, direction - taken from Name in Courses table in BasedZTM source
	Failure	“True” there was a failure, “False” no failures - taken from Failure in Courses table in BasedZTM source
	Continuity	“True” if the current course is a continuation of a course on which a failure occurred - taken from Continuity in Courses table in BasedZTM source
	NumberOfStations	Number of stations in one course. Three options: - less than 10 - from 10 to 20 - more than 20 Values based on numberOfStations in Courses table in BasedZTM source
DRIVER	One tuple describes a one driver in the specified age category	
	Pesel	PK of the driver - taken from PESEL in Driver table in BasedZTM source
	Name	First name of driver - taken from Name in Driver table in BasedZTM source
	Surname	Last name of driver - taken from Surname in Driver table in BasedZTM source
	E-Mail	E-mail contact to driver - taken from E-Mail in Driver table in BasedZTM source

	PhoneNumber	Driver's phone number - taken from PhoneNumber in Driver table in BasedZTM source
	AgeZone	Four options: - 18-30 - 30-45 - 45-60 - >60 Calculated from age in Driver table in BasedZTM source
BUS	One tuple describes one bus in the specified production year category, with the specified bus variant, depot number, registration number, make and model (Implementation of SCD 2)	
	Id	Id of the bus (surrogate key) - generated by the database
	VIN	Identification number of the bus - taken from VIN in Bus table in BasedZTM
	Registration	Registration Number - taken from Registration in Bus table in BasedZTM
	DepotNumber	Business key, Distinct number given to the bus by the bus transportation company - taken from Depot Number in Bus table in BasedZTM
	IsActive	"1" if information is current, otherwise "0" (SCD2 implementation).
	BusVariantID	Number identifying the variant of a bus, describes the combination of (Make, Model, ProdYearZone)
	Make	Name of buses manufacturer - taken from make in BusVariants table in BasedZTM

	Model	Model of the bus - taken from model in BusVariants table in BasedZTM
	ProdYearZone	Production year of the bus. Allowed values: - 1980> - 1981-1990 - 1991-2000 - 2001-2010 - 2011-2020 - 2021< Calculated from ProdYear in BusVariants table in BasedZTM
STATION	One tuple describes one station	
	Id	Id of the station (surrogate key) generated by the database
	Name	Unique name of a bus stop - taken from name in Stations table in BasedZTM
DATE	One tuple describes one day. All the data in this table is generated tuple by tuple based on any calendar, before the ETL process.	
TIME	One tuple describes one hour (independently of date). All the data in this table is generated tuple by tuple based on the clock, before the ETL process.	
JUNK	The tuples correspond to “all” possible combinations of values for FailureAction and OnDemand and are generated before ETL process	
	Id	Junk Id. Surrogate key - generated by database.