Insurance company – Data warehouse design

Business process

The data warehouse is designed for the claim processing. This process is described in the document Specification of business processes.

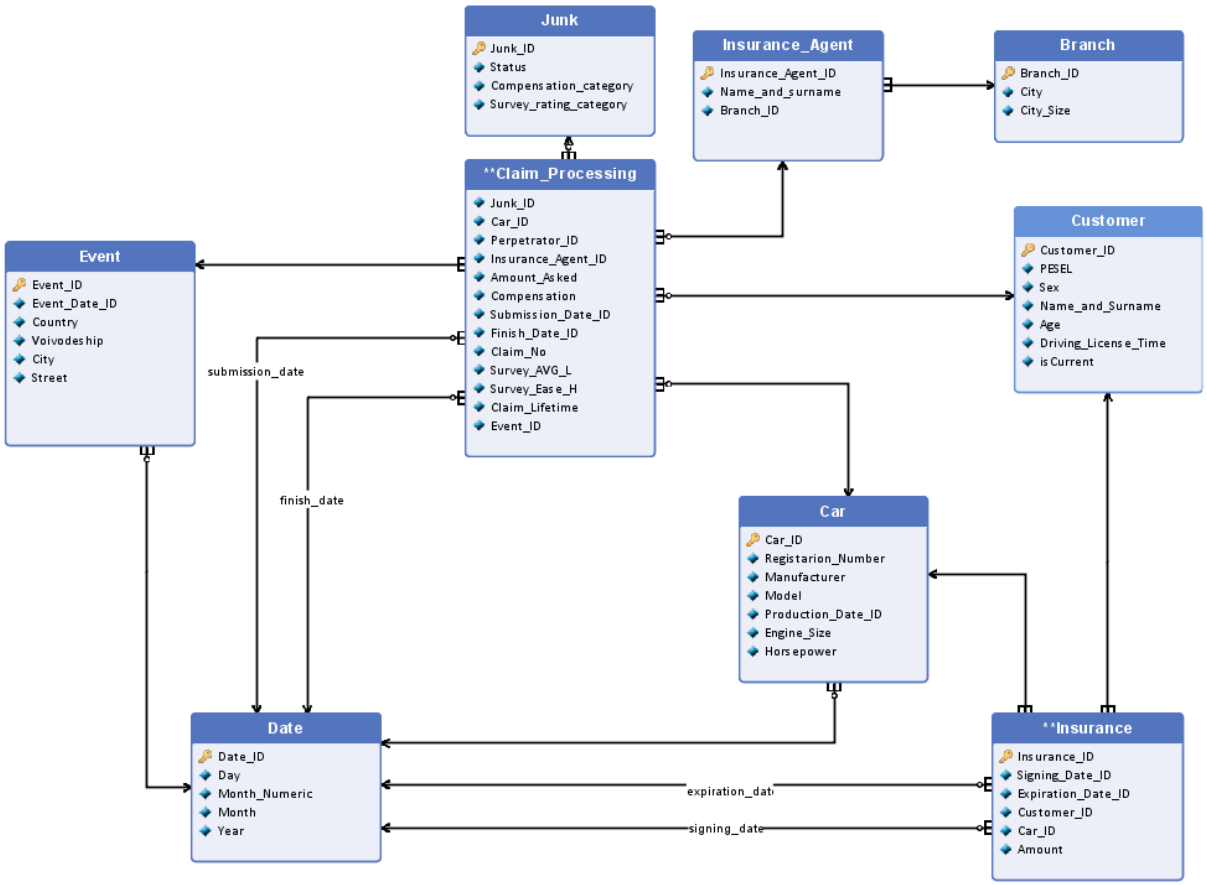

TABLE NAME	ATTRIBUTE	ATTRIBUTE TYPE	DESCRIPTION
Claim_Processing (FACT TABLE)	**One tuple describes one fact of the claim process.**		
	Junk_ID	Numeric	FK referencing to Junk table
	Car_ID	Numeric	FK referencing the Car
	Perpretator_ID	Numeric	FK referencing to Customer table, which is the perpetrator of the accident
	Inusrance_Agent_ID	Numeric	FK referencing to an Insurance Agent
	Amount_Asked	Int	Amount requested by the

			victim when submitting the claim
	Compensation	Int	Exact amount that was paid out
	Submission_Date_ID	Numeric	FK, date of submission
	Finish_Date_ID	Numeric	FK, date of finishing the claim
	Claim_No	Numeric	Process Number
	Survey_AVG_L	Numeric	The overall satisfaction of the processed claim (From 0 to 10)
	Survey_Ease_H	Numeric	The ease of filling the claim (From 0 to 10)
	Claim_Lifetime	Numeric	The lifetime of the claim
	Event_ID	Numeric	FK referencing to the Event table
Car (DIMENSION TABLE)	**One tuple describes one car owned by some customer.**		
	Car_ID	Numeric	PK
	Registration_Number	Varchar(10)	BK, license plate number
	Manufacturer	Varchar(15)	Manufacturer of the car
	Model	Varchar(25)	Model of the car.
	Prod_Date_ID	Numeric	FK date of production
	Engine_Size	Varchar(20)	Capacity of the engine in cm^3; Allowed values: ">900 and <=1200", ">1200 and <=1600", ">1600 and <=1800", ">1800 and <=2000", ">2000 and <=2200", ">2200 and <=3000", ">3000 and <=3500", "3500 or more",
	Horsepower	Varchar(20)	Power of the engine in horsepower unit; Allowed values: ">80 and <=110", ">110 and <=140", ">140 and <=170", ">170 and <=200", ">200 and <=300", ">300 and <=400", "400 or more"
Junk (DIMENSION TABLE)	**The tuples correspond to "all" possible combinations of values for Position and TypeOfPayment and are generated before ETL process.**		
	Junk_ID	Numeric	PK

	Status	Varchar(10)	Allowed Values ["canceled", "accepted", "rejected", "completed"]
	Compensation_Category	Varchar(20)	Allowed Values: "<= 1000", ">1000 and <=5000", ">5000 and <=10000", ">10000 and <=20000", ">20000 and <=50000", ">50000 and <=100000", ">100000"
	Survey_rating_category	Varchar(20)	Allowed Values: "Poor", "Bad", "Medium", "Good", "Perfect"
Insurance_Agent (DIMENSION TABLE)	**One tuple describes one insurance agent.**		
	Insurance_Agent_ID	Numeric	PK
	Name_and_Surname	Varchar(60)	Name and surname
	Branch_ID	Numeric	FK referencing to the branch in which the agent works
Branch (DIMENSION TABLE)	**One tuple describes one branch placed in some city.**		
	Branch_ID	Numeric	PK
	City	Varchar(20)	City in which the Branch is
	City_Size	Varchar(15)	Allowed values: "< 20000", ">= 20000 and =< 100000", "> 100000"
Event (DIMENSION TABLE)	**One tuple describes one event that occurred.**		
	Event_ID	Numeric	PK
	Event_Date_ID	Date	FK, key to the date of the accident
	Country	Varchar(50)	Country where the accident took place
	Voivodeship	Varchar(50)	Voivodeship where the accident took place
	City	Varchar(50)	City where the accident took place
	Street	Varchar(50)	Street where the accident took place
Date (DIMENSION TABLE)	**One tuple describes one day.**		
	Date_id	Numeric	PK

	Day	Numeric	Day
	Month_Numeric	Numeric	Month of the year in numeric format
	Month	Varchar(20)	Allowed values: "January", "February", "March", "April", "May", "June", "July, August", "September", "October", "November", "December"
	Year	Numeric	Year
Insurance (FACT TABLE)	**One tuple describes one signed insurance contract.**		
	Insurance_ID	Numeric	PK
	Signing_Date_ID	Numeric	FK, date of signing the contract
	Expiration_Date_ID	Numeric	FK, expiration date of the contract
	Customer_ID	Numeric	FK referencing to the insured Customer
	Car_ID	Numeric	FK referencing to the insured Car
	Amount	Numeric	Amount for which the insurance was bought
Customer (DIMENSION TABLE)	**One tuple describes one customer.**		
	Customer_ID	Numeric	PK
	PESEL	Numeic	BK
	Sex	Varchar(6)	Allowed values: "Male", "Female"
	Name_and_Surname	Varchar(60)	Name and surname
	Age	Varchar(20)	Age category. Allowed values: "18 to 21", "22 to 29", "30 to 49", "50 to 64", "65 or more"
	Driving_License_Time	Varchar(20)	Experience in driving in years. Allowed values: "0 to 5", "6 to 15", "16 to 35", "36 to 55", "56 or more"
	isCurrent	Boolean	1 if information is current, otherwise 0. (SCD2 implementation)

Dimensional model

Fact definitions

Fact 1 Claim Processing fact: Process of handling a claim submitted by the victim on a specified day regarding a specific event that took place on a specified day in a specified place. The claim corresponds to a specific car that took part in the accident. The process has specified the status of the claim.

Fact table: *Claim_Processing*

Granularity:

- a specified satisfaction
- a specified compensation
- a specified event
- a specified customer
- a specified date of submission of the claim
- a specified date of finish of the claim processing
- a specified car

Measures and aggregation functions:

- Number of claims processed - DISTINCT COUNT(Claim_No)
- Average amount paid out - AVG(Compensation)
- Amount paid out - SUM(Compensation)
- Average claim lifetime - AVG(Claim_Lifetime)
- Amount to be paid out - SUM(Amount_Asked)
- Average survey rating - AVG(Survey_AVG_L)
- Average ease of filling the claim - AVG(Survey_Ease_H)
- Amount of money saved - (Amount to be paid out - Amount paid out)

Fact 2 Insuring a car: A fact of insuring a specified car by a specified customer with a specified signing date and expiration date (when policy was signed and when it expires).

Fact table: *Insurance*

Granularity:

- a specified customer
- a specified car
- a specified signing date
- a specified expiration date

Measures and aggregation function:

- Number of all policies - COUNT(*)
- Company income - SUM(Amount)

Dimension definitions

Dimensions for Fact 1 Claim Process fact:

DIMENSION/DIMENSION ATTRIBUTE	TABLE / COLUMN	TYPE
ClaimNo	Claim_Processing.ClaimNo	Degenerate dimension
JUNK	Junk	Dimension
STATUS	Junk.Status	Dimension attribute
COMPENSATION	Junk.Compensation_category	Dimension attribute
SURVEY RATING CATEGORY	Junk.Survey_rating_category	Dimension attribute
EVENT	Event	Dimension
EVENT HIERARCHY	• Event.Country • Event.Voivodeship • Event.City • Event.Street	Hierarchical dimension
COUNTRY	Event.Country	Dimension attribute
VOIVODESHIP	Event.Voivodeship	Dimension attribute
CITY	Event.City	Dimension attribute
STREET	Event.Street	Dimension attribute
CUSTOMER	Customer	Dimension
PESEL	Customer.PESEL	Dimension attribute
SEX	Customer.Sex	Dimension attribute
NAME AND SURNAME	Customer.Name_And_Surname	Dimension attribute
AGE	Customer.Age	Dimension attribute, SCD 2

DRIVING LICENSE TIME	Customer.Driving_License_Time	Dimension attribute, SCD 2
CAR	Car	Dimension
REGISTRATION NUMBER	Car.Registration_Number	Dimension attribute
MANUFACTURER	Car.Manufacturer	Dimension attribute
MODEL	Car.Model	Dimension attribute
ENGINE SIZE	Car.Engine_Size	Dimension attribute
HORSEPOWER	Car.Horsepower	Dimension attribute
BRANCH	Branch	Dimension
CITY	Branch.City	Dimension attribute
CITY SIZE	Branch.City_Size	Dimension attribute
INSURANCE AGENT	Insurance_Agent	Dimension
NAME AND SURNAME	Insurance_Agent.Name_And_Surname	Dimension attribute
SUBMISSION DATE	Date	Dimension
SUBMISSION DATE HIERARCHY	● Date.Year ● Date.Month ● Date.Day	Hierarchical dimension
SUBMISSION YEAR	Date.Year	Dimension attribute
SUBMISSION MONTH	Date.Month	Dimension attribute
SUBMISSION DAY	Date.Day	Dimension attribute
FINISH DATE	Date	Dimension
FINISH DATE HIERARCHY	● Date.Year ● Date.Month ● Date.Day	Hierarchical dimension
FINISH YEAR	Date.Year	Dimension attribute
FINISH MONTH	Date.Month	Dimension attribute
FINISH DAY	Date.Day	Dimension attribute

Dimensions for Fact 2 Insuring a car:

DIMENSION/DIMENSION ATTRIBUTE	TABLE / COLUMN	TYPE
CUSTOMER	Customer	Dimension
PESEL	Customer.PESEL	Dimension attribute
SEX	Customer.Sex	Hierarchical dimension
NAME AND SURNAME	Customer.Name_And_Surname	Dimension attribute
AGE	Customer.Age	Dimension attribute
DRIVING LICENSE TIME	Customer.Driving_License_Time	Dimension attribute
CAR	Car	Dimension
REGISTRATION NUMBER	Car.Registration_Number	Dimension attribute
MANUFACTURER	Car.Manufacturer	Dimension attribute
MODEL	Car.Model	Dimension attribute
ENGINE SIZE	Car.Engine_Size	Dimension attribute
HORSEPOWER	Car.Horsepower	Dimension attribute
SIGNING DATE	Date	Dimension
SIGNING DATE HIERARCHY	• Date.Year • Date.Month • Date.Day	Hierarchical Dimension
SIGNING YEAR	Date.Year	Dimension attribute
SIGNING MONTH	Date.Month	Dimension attribute
SIGNING DAY	Date.Day	Dimension attribute
EXPIRATION DATE	Date	Dimension
EXPIRATION DATE HIERARCHY	• Date.Year • Date.Month • Date.Day	Hierarchical Dimension
EXPIRATION YEAR	Date.Year	Dimension attribute
EXPIRATION MONTH	Date.Month	Dimension attribute
EXPIRATION DAY	Date.Day	Dimension attribute

Checking the feasibility of queries based on the multidimensional model

1. Find and compare the average claim lifetime among all branches - from submitting till final decision.
   Measure: Average claim lifetime,
   Dimension: Branch (dimension attributes: City):

2. In which month did the most accidents occur?
   Measure: Number of claims processed
   Dimension: Date (dimension attributes: month)
   Dimension: Junk (dimension attributes: Status)

3. What is the processing time in relation to the amount of compensation?
   Measure: Average claim Lifetime
   Dimension: Junk (dimension attributes: Compensation_category)

4. Find the worker who handled the most and who the least claims in each month.
   Measure: Number of Claims processed
   Dimension: Insurance_Agent (dimension attributes: Name_and_Surname)

5. How many claims were submitted last year in comparison to the previous year (increase/decrease)?
   Measure: Number of Claims processed
   Dimension: Date (dimension attributes: year)

6. Find workers whose claim processing was rated above average (in terms of customer satisfaction).
   Measure: Average survey rating
   Dimension: Insurance_Agent (dimension attributes: Name_and_surname)

7. What was the average processing time for the highest rated claims?
   Measure: Average claim lifetime
   Dimension: Junk (dimension attributes: Survey_rating_category)

8. What was the average compensation for claims rated with a high satisfaction level?
   Measure: Average amount paid out
   Dimension: Junk (dimension attributes: Survey_rating_category)

9. Compare customer satisfaction levels among different cities.
   Measure: Average survey rating
   Dimension: Branch (dimension attributes: City)

10. Is there a relation between customers' perception of ease of filling the claim and his/hers satisfaction?
    Measure: Average ease of filling the claim

Dimension: Junk (dimension attribute: Survey_rating_category)

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

TABLE NAME	ATTRIBUTE	DESCRIPTION
Claim_Processing	**One tuple describes one fact of the claim process.**	
	Junk_ID	Junk ID. Surrogate key - generated by database.
	Car_ID	Car ID. Foreign key from dimension table. Based on registration number. Based on primary key reg_number in table Car in Database source.
	Perpetrator_ID	Perpetrator ID. Foreign key from dimension table. Based on Pesel number. Based on the primary key in Pesel in table Customer from Database source.
	Inusrance_Agent_ID	Insurance agent ID. Foreign key from dimension table. Based on the primary key Worker_ID in table Insurance Agent from Database source.
	Amount_Asked	The minimum repair coverage taken from Amount_Asked column from Claim table in Database source.
	Compensation	Exact amount that was paid out taken from the Compensation table in Database source.
	Submission_Date_ID	Submission date ID. Foreign key from dimension table. Based on Sub_Date stored in Claim table in Database source.
	Finish_Date_ID	Finish date ID. Foreign key from dimension table. Based on Finish_Date stored in Claim table in Database source.
	Claim_No	Process number taken from Claim_ID column in Claim table in Database source.
	Survey_AVG_L	Average satisfaction. Based on column L from the Satisfaction Survey.
	Survey_Ease_H	The ease of filling the claim. Based on column H from the Satisfaction Survey.
	Claim_Lifetime	The lifetime of the claim. Based on Sub_Date and Finish_Date in Claim table in Database source.

	Event_ID	Event ID. Foreign key from dimension table. Based on Event_ID from stored in the Event table in Database source.
Car	**One tuple describes one car.**	
	Car_ID	Car ID. Surrogate key generated by database.
	Registration_Number	Licence plate number. Taken from Reg_number from Car table in Database source.
	Manufacturer	Manufacturer of the car. Taken from Manufacturer from Car table in Database source.
	Model	Model of the car. Taken from Model from Car table in Database source.
	Prod_Date_ID	Date of production ID.Foreign key taken from dimension table Date. Based on Production_date from Car table in Database source.
	Engine_Size	Capacity of the engine in cm^3. Taken from Engine_size from Car table in Database source.
	Horsepower	Power of the engine in the horsepower unit. Taken from Horsepower from Car table in Database source.
Junk	**The tuples correspond to "all" possible combinations of values for Status, Compensation and Survey rating and are generated before ETL process.**	
	Junk_ID	Junk ID. Surrogate key - generated by database.
Insurance_Agent	**One tuple describes one insurance agent.**	
	Insurance_Agent_ID	Insurance Agent ID. Surrogate key - generated by the database.
	Name_and_Surname	Name and surname of the Insurance Agent. Taken from Name and Surname from Insurance_Agent table in Database source
	Branch_ID	Branch ID.Foreign key taken from dimension table Branch. Based on Branch_ID from Insurance_Agent table in Database source.
Branch	**One tuple describes one branch in some city.**	
	Branch_ID	Branch ID. Surrogate key - generated by the database.
	City	City in which the branch is. Taken

		from City from Branch table in Database source.
	City_Size	Size of the city. Taken from City_size from Branch table in Database source.
Event	**One tuple describes one event that occurred.**	
	Event_ID	Event ID, surrogate key based on Event_ID column in Event table in Database source
	Event_Date_ID	Date of the accident ID.Foreign key taken from dimension table Date. Based on Date from Event table in Database source.
	Country	Country where the accident took place. Taken from Country column from Event table in Database source
	Voivodeship	Voivodeship where the accident took place. Taken from Voivodeship column from Event table in Database source
	City	City where the accident took place. Taken from City column from Event table in Database source
	Street	Street where the accident took place. Taken from Street column from Event table in Database source
Date	**One tuple describes one day. All the data in this table are generated tuple by tuple based on any calendar, before ETL process.**	
	Date_id	Date ID. Surrogate key generated by the database
Insurance	**One tuple describes one signed insurance contract.**	
	Insurance_ID	Insurance ID, surrogate key based on Inurance_ID column in Insurance table in Database source
	Signing_Date_ID	Date of signing the insurance ID.Foreign key taken from dimension table Date. Based on Sign_Date from Insurance table in Database source.
	Expiration_Date_ID	Date of expiration of the insurance ID.Foreign key taken from dimension table Date. Based on Exp_Date from Insurance table in Database source.
	Customer_ID	Customer ID. Foreign key from dimension table. Based on Pesel

		number. Based on the primary key in Pesel in table Customer from Database source.
	Car_ID	Car ID. Foreign key from dimension table. Based on Registration number. Based on the primary key in Registration Number in table Car from Database source.
	Amount	Exact amount that was paid. Taken from Annual_Charge column in Insurance table in Database source.
Customer	**One tuple describes one customer.**	
	Customer_ID	Surrogate key generated by the database
	PESEL	PESEL number. Taken from PESEL from Customer table in Database source.
	Sex	Gender of the customer. Taken from Sex from Customer table in Database source
	Name_and_Surname	Name and surname of the customer. Taken from Name and Surname from Customer table in Database source.
	Age	Age of the customer. Based on the Birth_Date in Customer table in Database source.
	Driving_License_Time	Experience of the driver in years. Based on Driving_License_Date from Customer table in Database source.
	isCurrent	"1" if information is current, otherwise "0" (SCD2 implementation).