DATA WAREHOUSE DESIGN

1. Business process

The business process for which the data warehouse is designed is *choosing the new destinations of a trip.* This business process is described in detail in the document *Specification of requirements for the process of choosing the new destinations of a trip.*

2. Relational database schema

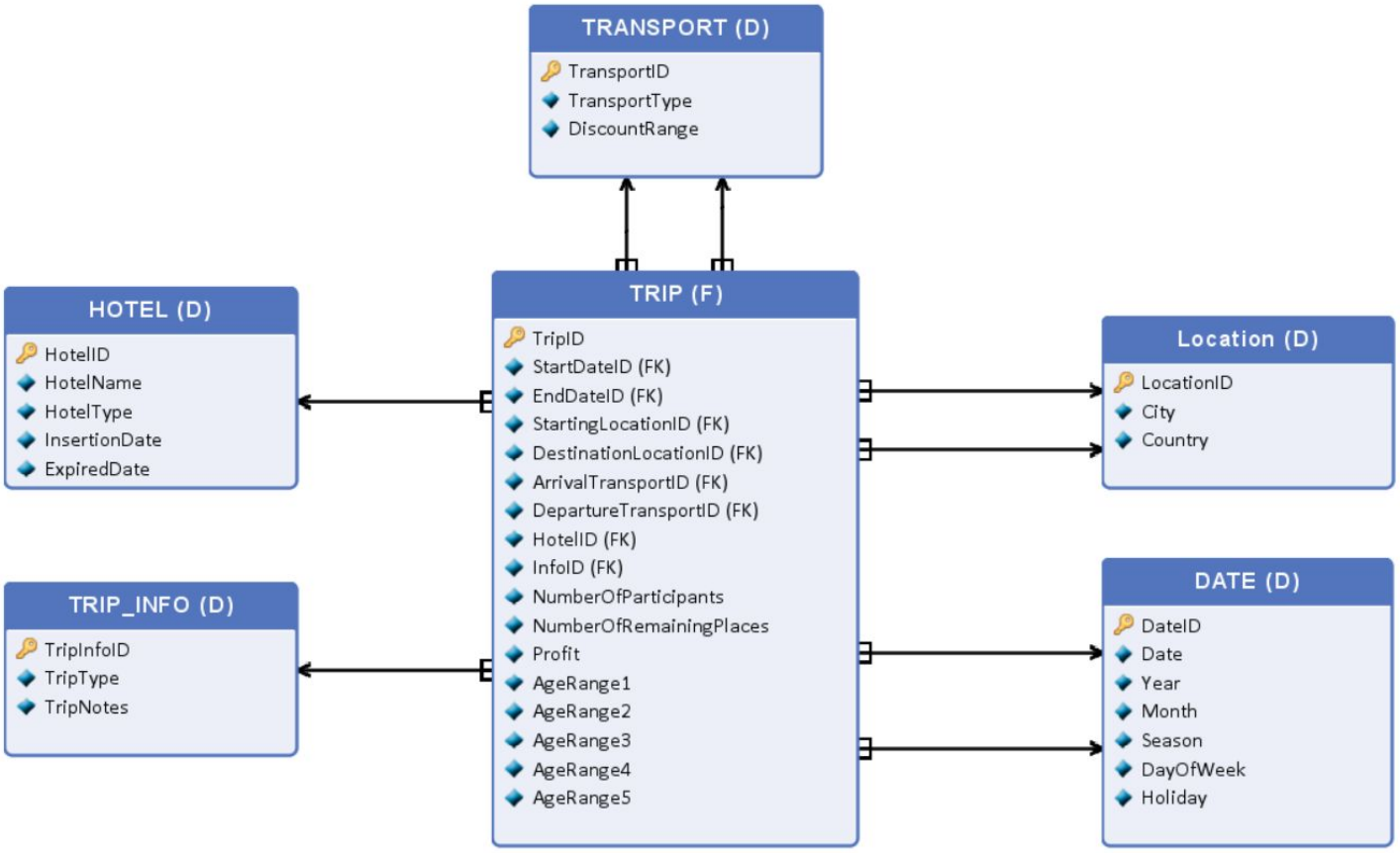

TABLE NAME	ATTRIBUTE NAME	ATTRIBUTE TYPE	DESCRIPTION
TRIP (fact table)	Tuples refer to the fact of trips that took place		
	TripID	Numerical	PK
	StartDateID	Numerical	FK Date

	EndDateID	Numerical	FK Date
	StartingLocationID	Numerical	FK Location
	DestinationLocationID	Numerical	FK Location
	ArrivalTransportID	Numerical	FK Transport
	DepartureTransportID	Numerical	FK Transport
	HotelID	Numerical	FK Hotel
	InfoID	Numerical	FK Trip_Info
	NumberOfParticipants	Numerical	Number of participants of the trip
	NumberOfRemainingPlaces	Numerical	The number of unsold places equal to *45* - *NumberOfParticipants*
	Profit	Float	The travel agency's profit from the entire trip
	AgeRange1	Numerical	The number of trip participants under the age of 15
	AgeRange2	Numerical	The number of trip participants at the age of 16-27
	AgeRange3	Numerical	The number of trip participants at the age of 28-39
	AgeRange4	Numerical	The number of trip participants at the age of 40-55
	AgeRange5	Numerical	The number of trip participants aged 56 or more

TRIP_INFO (dimension table)	Tuples refer to informations related to the trip		
	TripInfoID	Numerical	PK
	TripType	Character	The type of the trip (*round trip*, *recreation trip*, *leisure trip*, *individual trip*)
	TripNotes	Character	Travel agency's notes about the trip (problems encountered: '*transport delay*', '*natural catastrophe*', '*problems with the hotel*', '*none*')
HOTEL (dimension table)	Tuples refer to hotels where accommodation takes place		
	HotelID	Numerical	PK
	HotelName	Character	The name of the hotel
	HotelType	Character	The type of the hotel ('*', '**', '***', '****', '*****', '*motel*')
	InsertionDate	Date	Date of insertion information about the hotel (SCD implementation)
	ExpiredDate	Date	Date of information expiration (SCD implementation)

TRANSPORT (dimension table)	Tuples refers to the transports that took place		
	TransportID	Numerical	PK
	TransportType	Character	The type of the transport (*plane*, *bus*, *train*, *ship*)
	DiscountRange	Character	The range of the number of people needed to get a discount; transporting the number of participants within the range we can get a 10% discount on transport
DATE (dimension table)	Tuples correspond to a specific date		
	DateID	Numerical	PK
	Date	Date	Date
	Year	4 Digits	Year
	Month	Character	Month (January, February, March, April, May, June, July, August, September, October, November, December)
	Season	Character	Season (Spring, Summer, Autumn Winter)
	DayOfWeek	Character	Day of the week (Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday)

	Holiday	Character	Holiday (Christmas, Easter etc.)
LOCATION (dimension table)	Tuples correspond to the specified location		
	LocationID	Numerical	PK
	City	Character	City of the location
	Country	Character	Country of the location

3. Multidimensional model

3.1. Definition of facts

Fact 1: *Trip*

The fact of one specific trip taking place with given *start* and *end date*, from and to a specific location, using a specific *transport* to the *destination* and *back*, staying overnight in a given hotel, with informations about this specific trip, with a specific number of participants and remaining places, generating a specific profit, serving a specific number of people in specific age ranges.

Fact table: Trip

Granularity:

- *The fact of one specific trip taking place*
- Specific *date* of *departure* and *return* from the trip
- Specific *start location* and *destination*
- Specific *transport* to the *destination* and *back*
- Specific *hotel* where the accommodation took place
- Specific *informations* about *the whole trip*

Measures and aggregation functions:

- Number of facts of trip - *COUNT(1)*
- Company profit - *SUM(PROFIT)*
- Number of participants - *SUM(numberOFparticipants)*
- Number of unsold trips - *SUM(numberOfRemainingPlaces)*
- Average number of trip participants - *AVG(numberOFparticipants)*

- Total Number of participants - *SUM(numberOfParticipants)*
- Number of participants in Age Range 1 - *SUM(AgeRange1)*
- Number of participants in Age Range 2 - *SUM(AgeRange2)*
- Number of participants in Age Range 3 - *SUM(AgeRange3),*
- Number of participants in Age Range 4 - *SUM(AgeRange4)*
- Number of participants in Age Range 5 - *SUM(AgeRange5)*

3.2. Definition of dimensions

DIMENSION/DIMENSION ATTRIBUTE	TABLE/FIELD IN THE TABLE	TYPE
TRIP_INFO	TRIP_INFO	Dimension
Trip Info ID	TRIP_INFO.TripInfoID	Dimension attribute
Type of the trip	TRIP_INFO.TripType	Dimension attribute
Notes about the trip	TRIP_INFO.TripNotes	Dimension attribute
HOTEL	HOTEL	Dimension
Hotel ID	HOTEL.HotelID	Dimension attribute
Name of the hotel	HOTEL.HotelName	Dimension attribute
Type of the hotel	HOTEL.HotelType	Dimension attribute
TRANSPORT	TRANSPORT	Dimension
Transport ID	TRANSPORT.TransportID	Dimension attribute
Type of the transport	TRANSPORT.TransportType	Dimension attribute
Range of people needed to get discount	TRANSPORT.DiscountRange	Dimension attribute
LOCATION	LOCATION	Dimension
Location ID	LOCATION.LocationID	Dimension attribute

City of the location	LOCATION.City	Dimension attribute
Country of the location	LOCATION.Country	Dimension attribute
Location- hierarchy	LOCATION.Country LOCATION.City	Hierarchical dimension
DATE	DATE	Dimension
Date ID	DATE.DateID	Dimension attribute
Year	DATE.Year	Dimension attribute
Month	DATE.Month	Dimension attribute
Day	DATE.Day	Dimension attribute
Season	DATE.Season	Dimension attribute
Day of the week	DATE.DayOfWeek	Dimension attribute
Holiday	DATE.Holiday	Dimension attribute
Start Date- hierarchy	DATE.Year DATE.Season DATE.Month DATE.Date	Hierarchical dimension
End Date- hierarchy	DATE.Year DATE.Season DATE.Month DATE.Date	Hierarchical dimension

4. Checking the feasibility of queries based on the multidimensional model

Analytical problem: **What influences the willingness to buy a trip?**

1. **Compare destinations countries chosen by our clients month by month.**

   Measure: the total number of people traveling to each destination in each month *SUM(numberOfParticipants)*

   Dimension: (start) DATE (dimension attribute: *Month*)

   Dimension: (destination) LOCATION (dimension attributes: *City*, *Country*)

2. **Compare the number of participants grouped by the destination.**

   Measure: the total number of people travelling to every destination *SUM(numberOfParticipants)*

   Dimension: (destination) LOCATION (dimension attributes: *City*, *Country*)

3. **Compare no. of sold trips according to the type of transport**

   Measure: number of sold trips *SUM(TripID)*

   Dimension: TRANSPORT (dimension attribute: *TransportType*)

4. **How the number of sold trip is influenced by time of the year?**

   Measure: number of sold trips *SUM(TripID)*

   Dimension: (start) DATE (dimension attributes: *Season*)

5. **How the age of clients influence willingness to buy a trip for a given country?**

   Measure: age of clients *SUM(AgeRange1), SUM(AgeRange2), SUM(AgeRange3), SUM(AgeRange4), SUM(AgeRange5)*

   Dimension: LOCATION (dimension attribute: *Country*)

Analytical problem: **Which of the trip is the most profitable?**

1. **What is the total profit from trips according to the type of hotel?**

   Measure: total profit from trips *SUM(Profit)*

   Dimension: Hotel (dimension attributes: *HotelType*)

2. **Using what type of transport will we get a discount when transporting a small group of people (less than 14 participants)?**

   Measure: transport discount for a *small* group of people

   - In Our Data Warehouse the value of DiscountRange is calculated on the basis of the *Discount* value stored in the company's database - for more info read the description in the table below

   Dimension: TRANSPORT(dimension attributes: *TransportType*, *DiscountRange*)

3. **What are the groups of age of the customers and what is the number of people traveling in each group?**

   Measure: number of customers belonging to specific age ranges *SUM(AgeRange1), SUM(AgeRange2), SUM(AgeRange3), SUM(AgeRange4), SUM(AgeRange5)*

   - Our Data Warehouse is designed in a way that to answer this question we don't need to calculate anything more

4. **Which of the trips destinations created the biggest percent of unsold trips?**

   Measure: biggest percent of unsold trips given by the formula:

$$max\left(\frac{SUM(NumberOfRemainingPlaces)}{SUM(NumberOfParticipants)+SUM(NumberOfRemainingPlaces)}\right)$$

   Dimension: (destination) LOCATION (dimension attributes: City, Country)

5. How the type of hotel affects the amount of profit?

Measure: the amount of profit from the trip
Dimension: HOTEL (dimension attribute: *HotelType*)

5. Checking if data sources contain data that fill the data warehouse

TABLE NAME	ATTRIBUTE	DATA SOURCE
TRIP	Tuples refer to the fact of trips that took place	
	TripID	Surrogate key- generated by the database
	StartDateID	A foreign key is taken from the dimension table. Its value results from the *StartDate* stored in the *TRIP* table in the company's database
	EndDateID	A foreign key is taken from the dimension table. Its value results from the *FinishDate* stored in the *TRIP* table in the company's database
	StartingLocationID	A foreign key is taken from the dimension table. Its value results from the *StartCountryName and StartCityName* stored in company's Excel file
	DestinationLocationID	A foreign key is taken from the dimension table. Its value results from the *DestinationCountryName and DestinationCityName* stored in company's Excel file
	ArrivalTransportID	A foreign key is taken from the dimension table. Its value results from the *CompanyID* and *Arrival* attribute stored in the *TRANSPORTS* table in the company's database (*Arrival* = 1 when arrival transport) It is the

		transport that brings people on a trip.
	DepartureTransportID	A foreign key is taken from the dimension table. Its value results from the *CompanyID* and *Arrival* attribute stored in the *TRANSPORTS* table in the company's database (*Arrival = 0* when departure transport) It is the transport that brings people back from the trip.
	HotelID	A foreign key is taken from the dimension table. Its value results from the HotelID stored in the *RESIDENCES* table in the company's database
	InfoID	A foreign key is taken from the dimension table. Its value results from the *TripType* and *TripNotes* stored in the *TRIP* table in the company's database
	NumberOfParticipants	The number is taken from company's Excel file from column *TotalNumberOfParticipants*
	NumberOfRemainingPlaces	The *NumberOfRemainingPlaces* is equal to *45-NumberOfParticipants*. The *NumberOfParticipants* is taken from table *TRIP* in the data warehouse
	Profit	The number is taken from company's Excel file from column *Profit*
	AgeRange1	The value is calculated on the basis of *BirthDate* stored in the *PERSON* table in the company's database. It counts only people under the age of 15

	AgeRange2	The value is calculated on the basis of *BirthDate* stored in the *PERSON* table in the company's database. It counts only people at the age of 16 - 27
	AgeRange3	The value is calculated on the basis of *BirthDate* stored in the *PERSON* table in the company's database. It counts only people at the age of 28 - 39
	AgeRange4	The value is calculated on the basis of *BirthDate* stored in the *PERSON* table in the company's database. It counts only people at the age of 40 - 55
	AgeRange5	The value is calculated on the basis of *BirthDate* stored in the *PERSON* table in the company's database. It counts only people aged 56 or more
HOTEL	Tuples refer to hotels where accommodation takes place	
	HotelID	Surrogate key- generated by the database
	HotelName	Hotel name taken from *HotelName* stored in the *HOTEL* table in the company's database
	HotelType	Hotel type taken from *HotelType* stored in the *HOTEL* table in the company's database
	InsertionDate	Date of insertion information about the hotel (SCD implementation)
	ExpiredDate	Date of information expiration (SCD implementation)

TRANSPORT	Tuples refer to hotels where accommodation takes place	
	TransportID	Surrogate key- generated by the database
	TransportType	Transport type taken from *TransportType* stored in the *TRANSPORT_COMPANY* table in the company's database
	DiscountRange	The range of the number of people needed to get a discount. It is calculated on the basis of the *Discount* value (minimum number of people needed to get discount) stored in the *TRANSPORT_COMPANY* table in the company's database. Accepts values: - if *Discount* value is less than 14 then DiscountRange is *'small'* (a small group of people) - if *Discount* value is between 15 - 28 then DiscountRange is *'medium'* - if *Discount* value is above 29 then DiscountRange is *'big'*
LOCATION	Tuples correspond to the specified location	
	LocationID	Surrogate key- generated by the database
	City	City of address taken from *City* stored in the *ADDRESS* table in the company's database
	Country	Country of address taken from *Country* stored in the *ADDRESS* table in the company's database

DATE	The tuples correspond to a specific date. All data in this table is generated one by one based on any calendar before the ETL process.	
	DateID	Surrogate key- generated by the database
TRIP_INFO	The tuples correspond to all possible combinations of values for *TripType* and *TripNotes* columns. They are generated before the ETL process.	
	TripInfoID	Surrogate key- generated by the database