YourTour – Data warehouse design

Business process

The data warehouse is designed for tour sale business process. This process is described in the document Specification of business processes.
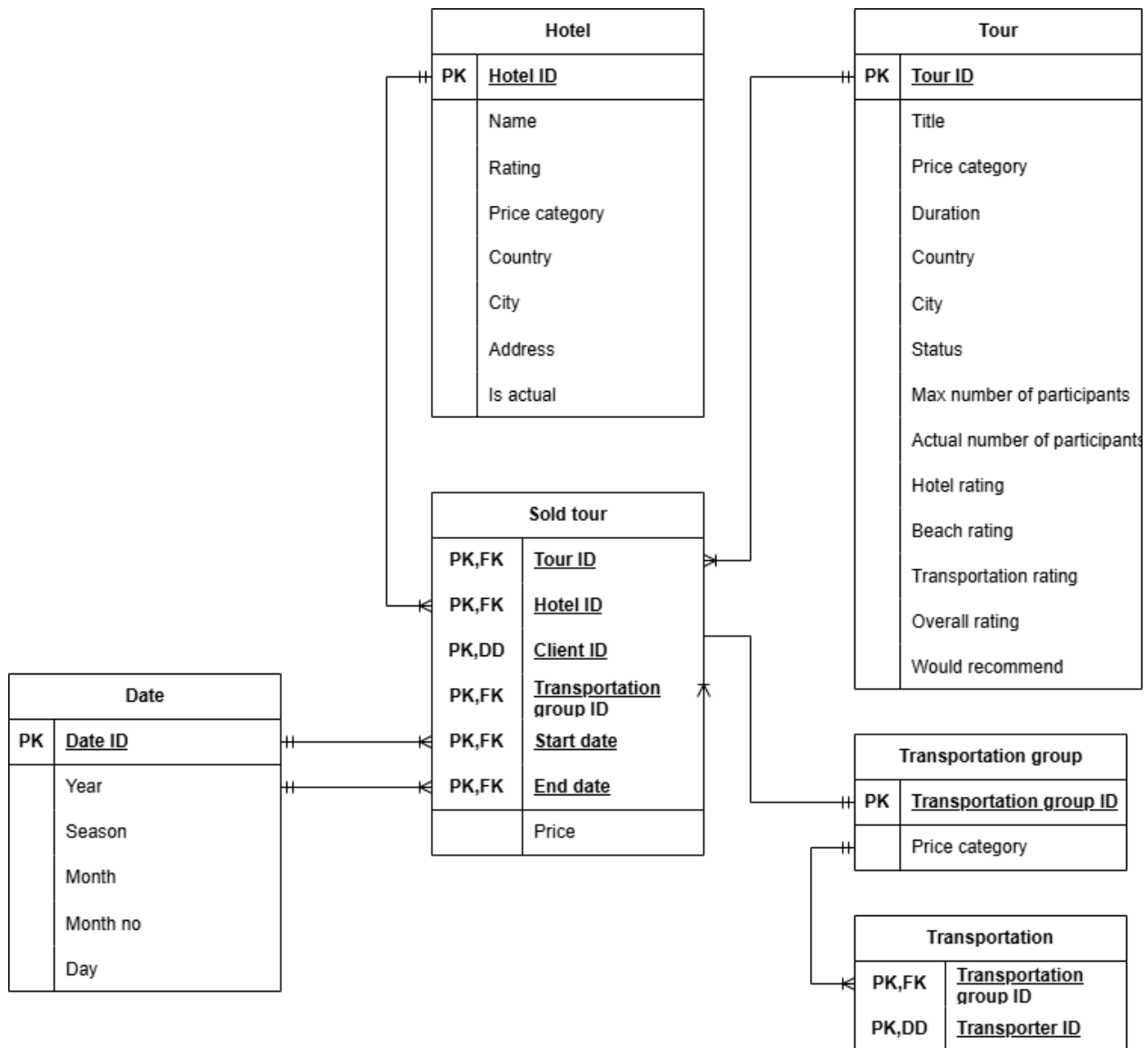
Relational database schema

TABLE NAME	ATTRIBUTE	ATTRIBUTE TYPE	DESCRIPTION
Date (Dimension table)	One tuple describes one day		
	Date ID	Numeric	Primary key
	Year	4 digits	Year
	Season	Varchar(10)	Season. Allowed values: Winter, Spring, Summer, Autumn
	Month	Varchar(10)	Month. Allowed values: January, February, March, April, May, June, July, August, September, October, November, December
	Month no	2 digits	Month's numeric value
	Day	2 digits	Day's numeric value
Hotel (Dimension table)	One tuple describes one hotel		
	Hotel ID	Numeric	Primary key
	Name	Varchar(70)	Name of the hotel
	Rating	1 digit	Rating of the hotel. Allowed values: 1-star, 2-star, 3-star, 4-star, 5-star
	Price category	Varchar(10)	Price category of the hotel. Allowed values: Low, Medium, High
	Country	Varchar(20)	Country in which the hotel is located
	City	Varchar(20)	City in which the hotel is located
	Address	Varchar(20)	Address of the hotel
	Is actual	Boolean	"1" if information is actual, otherwise "0" (SCD 2 implementation).

Sold tour (Fact table)	One tuple describes one fact of tour sale		
	Tour ID	Numeric	Foreign key. Tour ID
	Hotel ID	Numeric	Foreign key. Hotel ID
	Client ID	Numeric	Client ID
	Transportation group	Numeric	Foreign key. Transportation group ID
	Start date	Numeric	Foreign key. Start date of the tour
	End date	Numeric	Foreign key. End date of the tour
	Price	Numeric	Price of the tour
Tour (Dimension table)	One tuple describes one tour		
	Tour ID	Numeric	Primary key
	Title	Varchar(100)	Title of the tour
	Price category	Varchar(10)	Price category of the tour. Allowed values: Low, Medium, High
	Duration	Varchar(10)	Duration of the tour. Allowed values: Short, Long
	Country	Varchar(20)	Country in which the tour took place
	City	Varchar(20)	City in which the tour took place
	Status	Varchar(10)	Status of the tour. Allowed values: Finished, Cancelled
	Max number of participants	Numeric	Maximum number of people who can partake in the tour
	Actual number of participants	Numeric	Number of people who partook in the tour
	Hotel rating	Varchar(10)	Rating of the hotel, based on the questionnaires. Allowed values: Very low, Low, Mediocre, High

	Beach rating	Varchar(10)	Rating of the beach, based on questionnaires. Allowed values: Very low, Low, Mediocre, High
	Transportation rating	Varchar(10)	Rating of the transportation, based on the questionnaires. Allowed values: Very low, Low, Mediocre, High
	Overall rating	Varchar(10)	Overall rating, based on the questionnaires. Allowed values: Very low, Low, Mediocre, High
	Would recommend	Numeric	How many people would recommend the tour, based on the questionnaires
Transportation group (Dimension table)	One tuple describes one transportation group		
	Transportation group ID	Numeric	Primary key
	Price category	Varchar(10)	Price category of the transportation. Allowed values: Low, Medium, High
Transportation (Fact table)	One tuple describes one fact of transportation		
	Transportation group ID	Numeric	Foreign key. Transportation group ID
	Transporter ID	Numeric	Transporter ID

Dimensional model

Fact definitions

Fact 1 – Tour sale fact: sale of the specified tour, with the specified hotel, to the specified client, in the specified transportation group, with the specified start and end dates.

Fact table: Sold tour.

Granularity:
- A specified tour in the specified price category, to the specified destination
- A specified hotel in the specified price category
- A specified transportation group in the specified price category
- A specified client in the specified age category
- A specified start date
- A specified end date

Measures and aggregation functions:
- Number of places sold – COUNT(1)
- Number of tours – DISTINCT COUNT(Tour ID)
- Profit – SUM(Price)
- Average price – Profit / Number of places sold

Fact 2 – Transportation fact: providing transportation services. Fact resulting from the "many to many" relationship.

Fact table: Transportation

Granularity:
- A specified transportation group in the specified price category
- A specified transporter

Measures and aggregation functions:

- Number of transportations – COUNT(1)
- Number of transportations by transporter – COUNT(Transporter ID)

Dimension definitions

Dimensions for Fact 1 – Tour sale fact:

DIMENSION/DIMENSION ATTRIBUTE	TABLE/COLUMN	TYPE
Tour	Tour	Dimension
Tour title	Tour.Title	Dimension attribute
Tour price category	Tour.Price category	Dimension attribute
Tour destination hierarchy	•Tour.Country ••Tour.City	Hierarchical dimension
Tour country	Tour.Country	Dimension attribute

Tour city	Tour.City	Dimension attribute
Tour status	Tour.Status	Dimension attribute
Tour duration	Tour.Duration	Dimension attribute
Max number of participants	Tour.Max number of participants	Dimension attribute
Actual number of participants	Tour.Actual number of participants	Dimension attribute
Hotel rating	Tour.Hotel rating	Dimension attribute
Beach rating	Tour.Beach rating	Dimension attribute
Transportation rating	Tour.Transportation rating	Dimension attribute
Overall rating	Tour.Overall rating	Dimension attribute
Would recommend	Tour.Would recommend	Dimension attribute
Hotel	Hotel	Dimension
Hotel name	Hotel.Name	Dimension attribute
Hotel rating	Hotel.Rating	Dimension attribute
Hotel price category	Hotel.Price category	Dimension attribute
Hotel localization hierarchy	•Hotel.Country ••Hotel.City •••Hotel.Address	Hierarchical dimension
Hotel country	Hotel.Country	Dimension attribute
Hotel city	Hotel.City	Dimension attribute
Hotel address	Hotel.Address	Dimension attribute
Client ID	Sold tour.Client ID	Degenerate dimension
Transportation group	Transportation group	Dimension
Transportation group price category	Transportation group.Price category	Dimension attribute
Start date	Date	Dimension
Start date hierarchy	•Date.Year ••Date.Month no •••Date.Day	Hierarchical dimension
Start date year	Date.Year	Dimension attribute
Start date season	Date.Season	Dimension attribute
Start date month	Date.Month	Dimension attribute
Start date day	Date.Day	Dimension attribute
End date	Date	Dimension
End date hierarchy	•Date.Year ••Date.Month no •••Date.Day	Hierarchical dimension
End date year	Date.Year	Dimension attribute
End date season	Date.Season	Dimension attribute

End date month	Date.Month	Dimension attribute
End date day	Date.Day	Dimension attribute

Dimensions for Fact 2 – Transportation fact:

DIMENSION/DIMENSION ATTRIBUTE	TABLE/COLUMN	TYPE
Transportation group	Transportation group	Dimension
Price category	Transportation group.Price category	Dimension attribute
Transporter	Transportation.Transporter	Degenerate dimension

Checking the feasibility of queries based on the multidimensional model

1. What is the most popular destination in a given season?
   Measure: Number of places sold
   Dimension: Tour (dimension attribute: Tour city)
   Dimension: Start date (dimension attribute: Start date season)
2. How many tours received a low overall rating and low transportation rating?
   Measure: Number of tours
   Dimension: Tour (dimension attributes: Overall rating, Transportation rating)
3. How many people recommended the tours with a high price compared to the tours with a low price in the last year?
   Measure: Number of places sold
   Dimension: Tour (dimension attribute: Tour price category)
   Dimension: Tour (dimension attribute: Would recommend)
   Dimension: Start date (dimension attribute: Start date year)
4. Is there a trend of placing high ratings for hotels with high ratings?
   Measure: Number of places sold
   Dimension: Tour (dimension attribute: Hotel rating)
   Dimension: Hotel (dimension attribute: Hotel rating)
5. The average price of the tours with the highest overall rating in the last year.
   Measure: Average price
   Dimension: Tour (dimension attribute: Overall rating)
   Dimension: Start date (dimension attribute: Start date year)

6. In how many tours was the number of participants not maximal?
   Measure: Number of tours
   Dimension: Tour (dimension attribute: Max number of participants, Actual number of participants)

7. What tours had a long duration and high hotel prices in the last year?
   Measure: Number of tours
   Dimension: Tour (dimension attribute: Tour duration)
   Dimension: Hotel (dimension attribute: Hotel price category)
   Dimension: Start date (dimension attribute: Start date year)

8. What tours had a bad overall rating but cost highly in the last year?
   Measure: Number of tours
   Dimension: Tour (dimension attributes: Overall rating, Tour price category)
   Dimension: Start date (dimension attribute: Start date year)

9. In how many cases was there a big "negative" difference between the hotel rating and the room price in the last year?
   Measure: Number of tours
   Dimension: Tour (dimension attribute: Hotel rating)
   Dimension: Hotel (dimension attribute: Hotel price category)
   Dimension: Start date (dimension attribute: Start date year)

10. In how many cases was there a big "negative" difference between the transportation rating and the transportation price in the last year?
    Measure: Number of tours
    Dimension: Tour (dimension attribute: Transportation rating)
    Dimension: Transportation group (dimension attribute: Transportation group price category)
    Dimension: Start date (dimension attribute: Start date year)

Checking if there is data in the data sources needed to fill the data warehouse

TABLE NAME	COLUMN	SOURCE
Sold tour	One tuple describes one fact of book sale	
	Tour ID	Sold tour ID. Foreign key from Tour dimension table. Taken from foreign key Tour ID in Tourist table in TourMaster source

	Hotel ID	Hotel ID. Foreign key from Hotel dimension table. Taken from foreign key Hotel ID in Room reservation table in TourMaster source
	Client ID	Client ID. Degenerated Client dimension. Taken from foreign key Client ID in Tourist table in TourMaster source
	Transportation group ID	Transportation group ID. Foreign key from Transportation group dimension table. Taken from surrogate key generated by the database
	Start date	Start date of the tour. Foreign key from Date dimension table. Taken from Start date column in Tour table in TourMaster source
	End date	End date of the tour. Foreign key from Date dimension table. Taken from End date column in Tour table in TourMaster source
	Price	The price of sold tour. Taken from Price column in Tour table in TourMaster source
Tour	One tuple describes one tour in specified price category, with specified duration, to the specified country and city, with specified status, with specified maximum and actual number of participants, with specified ratings	
	Tour ID	Tour ID. Taken from Tour ID column in Tour table in TourMaster source
	Title	Title of the tour. Taken from Title column in Tour table in TourMaster source
	Price category	Price category of the tour.

		Values calculated from Price column in Tour table in TourMaster source. Allowed values[1]: Low, Medium, High
	Duration	Duration of the tour. Calculated from Start date and End date columns in Tour table in TourMaster source. Allowed values[2]: Short, Long
	Country	Country in which the tour took place. Taken from Country column in Tour table in TourMaster source
	City	City in which the tour took place. Taken from City column in Tour table in TourMaster source
	Status	Status of the tour. Taken from Status column in Tour table in TourMaster source. Allowed values: Finished, Cancelled
	Max number of participants	Maximum number of people who can partake in the tour. Taken from Max number of participants column in Tour table in TourMaster source.
	Actual number of participants	Actual number of people who partook in the tour. Calculated from counting the number of Tourist rows for given Tour table in TourMaster source
	Hotel rating	Rating of the hotel. Calculated from Column B on Sheet 1 in Excel for given Tour. Allowed values[3]: Very low, Low, Mediocre, High

	Beach rating	Rating of the beach. Calculated from Column C on Sheet 1 in Excel for given Tour. Allowed values[3]: Very low, Low, Mediocre, High
	Transportation rating	Rating of the transportation. Calculated from Column D on Sheet 1 in Excel for given Tour. Allowed values[3]: Very low, Low, Mediocre, High
	Overall rating	Overall rating. Calculated from Column E on Sheet 1 in Excel for given Tour. Allowed values[3]: Very low, Low, Mediocre, High
	Would recommend	How many people would recommend the tour. Calculated from Column F on Sheet 1 in Excel for given Tour.
Hotel	One tuple describes one hotel (implementation of SCD 2)	
	Hotel ID	Hotel ID. Taken from Hotel ID column in Hotel table in TourMaster source
	Name	Name of the hotel. Taken from Name column in Hotel table in TourMaster source
	Rating	Rating of the hotel. Taken from Rating column in Hotel table in TourMaster source. Allowed values: 1-star, 2-star, 3-star, 4-star, 5-star
	Price category	Price category of the hotel. Calculated from Price column in Room reservation table in TourMaster source. Allowed values[4]: Low, Medium, High
	Country	Country in which the hotel is located. Taken from

		Country column in Hotel table in TourMaster source
	City	City in which the hotel is located. Taken from Country column in Hotel table in TourMaster source
	Address	Address of the hotel. Taken from Address column in Hotel table in TourMaster source
	Is actual	"1" if information is actual, otherwise "0" (SCD 2 implementation).
Transportation	One tuple describes one fact of transportation	
	Transportation group ID	Transportation group ID. Foreign key from dimension table Transportation group. Taken from surrogate key generated by the database
	Transporter ID	Transporter ID. Degenerated Transporter dimension. Taken from Transporter ID column in Transporter table in TourMaster source
Transportation group	One tuple describes one transportation group	
	Transportation group ID	Transportation group ID. Surrogate key – generated by the database
	Price category	Price category of the transportation. Calculated from Price column in Transportation table in TourMaster source. Allowed values[5]: Low, Medium, High
Date	One tuple describes one day. All the data in this table is generated tuple by tuple based on any calendar, before ETL process	

[1] Low – the price of this tour is 50% more than (or equals to) the average price of the tours for this country.

Medium – the price of this tour is 20% more but 50% less than the average price of the tours for this country.

High – the price of this tour is 20% less than (or equals to) the average price of the tours for this country.

[2] Short – the duration of this tour is less than 11 days.

Long – the duration of this tour is more than 10 days.

[3] Very low – the average rating for this tour is from the range [0-2].

Low – the average rating for this tour is from the range [2.1-2.5]

Mediocre – the average rating for this tour is from the range [2.6-3.9]

High – the average rating for this tour is from the range [4-5]

[4] Low – the average price of the rooms in this hotel is 20% less than (or equals to) the average price of the hotel rooms in this country.

Medium – the average price of the rooms in this hotel is 20% more but 50% less than the average price of the hotel rooms in this country.

High – the average price of the rooms in this hotel is 50% more than (or equals to) the average price of the hotel rooms in this country.

[5] Low – the average price of transportation is 20% less than (or equals to) the average price of transportation in this country.

Medium – the average price of transportation is 20% more but 50% less than the average price of transportation in this country.

High – the average price of transportation is 50% more than (or equals to) the average price of transportation in this country.