Book's Bookstore – Date warehouse design

Business process

The Date warehouse is designed for book sale business process. This process is described in the document *Specification of business processes in Book's Bookshop network*.

Relational Database schema

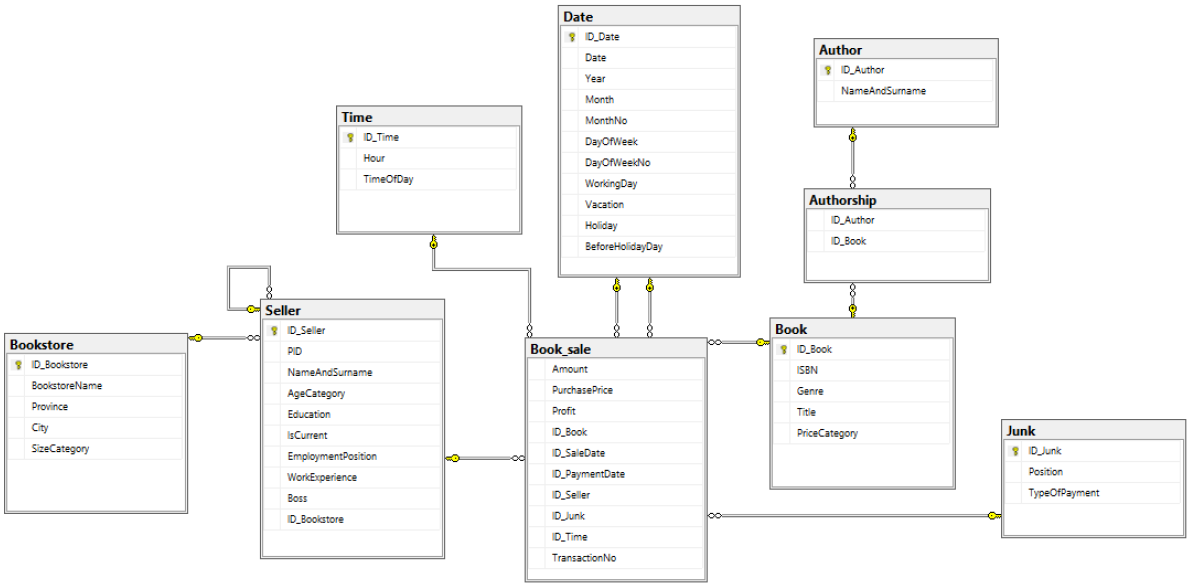

TABLE NAME	ATTRIBUTE	ATTRIBUTE TYPE	DESCRIPTION
BOOK_SALE (FACT TABLE)	**One tuple describes one fact of book sale.**		
	Id_SaleDate	Numeric	FK Date Sale date.
	Id_PaymentDate	Numeric	FK Date Payment date
	Id_Book	Numeric	FK Book Sold book
	Id_Seller	Numeric	FK Seller Seller
	Id_Junk	Numeric	FK Junk Junk attributes
	Id_Time	Numeric	FK Time Sale time
	TransactionNo	15 digits	Transaction no.
	Amount	Numeric	Amount of sold books
	Profit	Money	Profit equals to *PurchasePrice*0.07*
	PurchasePrice	Money	Price paid to the publisher for all copies sold within one transaction
AUTHORSHIP (FACT TABLE)	**One tuple describes one fact of being an author of the book.**		
	Id_Book	Numeric	FK_Book The book written by the author.
	Id_Author	Numeric	FK_Author The book's author.
BOOK (DIMENSION TABLE)	**One tuple describes one title purchased within one price category. (Note: book title not copy of the book)**		
	Id_Book	Numeric	PK (surrogate key)
	ISBN	varchar(17)	BK
	Genre	varchar (15)	Defines the book genre. Allowed values: encyclopedia, album, fantasy, other, informatics, magazine, history, language, cooking, drama, poetry, thriller
	Title	Varchar (30)	Book title.
	PriceCategory	Varchar (15)	Price category. Allowed values: occasion, cheap, typical, expensive, absurd

BOOKSTORE (DIMENSION TABLE)	**One tuple describes one bookstore**		
	Id_Bookstore	Numeric	PK
	BookstoreName	Varchar(20)	Bookstore name
	Province	Varchar(20)	Province, where the bookstore is located.
	City	Varchar(20)	City, where the bookstore is located.
	SizeCategory	Varchar(15)	Bookstore size. Allowed values: small, average, big
SELLER	**One tuple describes one seller, in the specified age category, with the specified education, position, work experience and boss.**		
	Id_Seller	Numeric	PK
	Id_Bookstore	Numeric	FK Bookstore Bookstore, the seller is employed in.
	PID	11 digits	Personal Identification Number.
	NameAndSurname	Varchar (50)	Name and surname.
	AgeCategory	Varchar (20)	Age category. Allowed values: between 15 and 17, from 18 to 21, from 22 to 29, from 30 to 49, from 50 to 64, more than 64
	Education	Varchar (20)	Education. Allowed values: vocational, incomplete secondary, secondary, incomplete higher, higher, doctorate
	Boss	Numeric	FK Seller Seller's boss.
	EmploymentPosition	Varchar (15)	Seller's employment position. Allowed values: seller, director
	WorkExperience	Varchar(30)	Work experience. Allowed values: up to one year, between one and five years, more than five years.
	IsCurrent	Boolean	1 if information is current, otherwise 0. (SCD2 implementation)
AUTHOR	**One tuple describe one author.**		

(DIMENSION TABLE)			
	Id_Author	Numeric	PK
	NameAndSurname	Varchar(75)	Name and surname
DATE (DIMENSION TABLE)	**One tuple describe one day.**		
	Id_Date	Numeric	PK
	Date	Date	Date
	Year	4 digits	Year
	Month	Varchar(10)	Month. Allowed values: January, February, March, April, May, June, July, August, September, October, November and December.
	MonthNo	Numeric	Month's numeric value
	DayOfWeek	Varchar(10)	Day of week. Allowed values: Monday, Tuesday, Wednesday, Thursday, Friday, Saturday and Sunday
	DayOfWeekNo	Numeric	Weekday's numeric value
	WorkingDay	Varchar(15)	Working day. Allowed values: day off and working day
	Vacation	Varchar(20)	Vacation time characteristics. Allowed values: non-holiday, winter holiday and summer holiday.
	Holiday	Varchar(50)	Type of holiday. Allowed values: Christmas, Grandmother's day, Grandfather's day …
	BeforeHolidayDay	Varchar(62)	Before holiday day. Allowed values: tomorrow is Grandmother's day, tomorrow is Grandfather's day, …
TIME (DIMENSION	**One tuple describes one hour (independently on date)**		

TABLE)			
	Id_Time	Numeric	PK
	Hour	Numeric	Hour. Allowed values from 0 – 23.
	TimeOfDay	Varchar(20)	Time of day. Allowed values: between 0 and 8, between 9 and 12, between 13 and 15, between 16 and 20, between 21 and 23).
JUNK (DIMENSION TABLE)	**The tuples correspond to "all" possible combinations of values for Stanowisko and FormaPlatnosci.**		
	Id_Junk	Numeric	PK
	Position	Varchar(20)	Position. Allowed values: information, main door, side door, …
	TypeOfPayment	Varchar(10)	Type of payment. Allowed values: check, card, cash.

Dimensional model

Fact definitions

Fact 1 Sale fact: Sale of the book (the title is considered not the copy), purchased on a specified day, at a specified time. Sold by a specified seller working on a specified position and employed as a seller or director in a specified bookstore within a single transaction paid with a certain type of payment on a specified day.

Fact table: Book_sale

Granularity:
- a specified transaction,
- a specified position,
- a specified type of payment,
- a specified hour of sale,
- a specified date of purchase,
- a specified seller in the specified age category, with the specified education, position, work experience and boss,
- a specified book title purchased within one price category.

Measures and aggregation functions:
Number of sale facts – COUNT (1)
Number of book copies sold - SUM (Amount)
Purchase price - SUM (PurchasePrice)

Profit - SUM (Profit)
Number of transactions - DISTINCT COUNT (TransactionNo)

Fact 2 Being the author fact: Being the author of a book.

Fact resulting from the "many to many" relationship.
Fact table: Authorship

Granularity:
- a specified book title purchased within one price category,
- a specified author.

Measures and aggregation functions:
Number of being the author facts – COUNT(1)

Dimension definitions

Dimensions for Fact 1 Sale fact:

DIMENSION/DIMENSION ATTRIBUTE	TABLE/COLUMN	TYPE
TRANSACTION NUMBER	Book_sale.TransactionNo	Degenerate dimension
BOOK	Book	Dimension
ISBN	Book.ISBN	Dimension attribute
BOOK GENRE	Book.Genre	Dimension attribute
BOOK TITLE	Book.Title	Dimension attribute
PRICE CATEGORY	Book.PriceCategory	Dimension attribute
SELLER	Seller	Dimension
PID	Seller.PID	Dimension attribute
SELLER NAME	Seller.NameAndSurname	Dimension attribute
SELLER WORK EXPERIENCE	Seller.WorkExperience	Dimension attribute
SELLER POSITION	Seller.PositionEmployment	Dimension attribute
SELLER AGE CATEGORY	Seller.AgeCategory	Dimension attribute
SELLER EDUCATION	Seller.Education	Dimension attribute
SUBORDINATE HIERARCHY	• Seller.Boss •• Seller.Id_Seller	Hierarchical dimension
SELLER HIERARCHY	• Seller.Education •• Seller.AgeCategory ••• Seller.PID	Hierarchical dimension
SALE DATE HIERARCHY	• Date.Year •• Date.Month ••• Date.Date	Hierarchical dimension
PAYMENT DATE HIERARCHY	• Date.Year •• Date.Month ••• Date.Date	Hierarchical dimension
SALE DATE	Date	Dimension
SALE YEAR	Date.Year	Dimension attribute
SALE MONTH	Date.Month	Dimension attribute
SALE DAY	Date.Date	Dimension attribute

VACATION DAY	Date.Vacation	Dimension attribute
HOLIDAY SALE DAY	Date.Holiday	Dimension attribute
BEFORE HOLIDAY SALE DAY	Date.BeforeHolidayDay	Dimension attribute
PAYMENT DATE	Date	Dimension
PAYMENT YEAR	Date.Year	Dimension attribute
PAYMENT MONTH	Date.Month	Dimension attribute
PAYMENT DAY	Date.Day	Dimension attribute
VACATION PAYMENT DAY	Date.Vacation	Dimension attribute
HOLIDAY PAYMENT DAY	Date.Holiday	Dimension attribute
BEFORE HOLIDAY PAYMENT DAY	Date.BeforeHolidayDay	Dimension attribute
SALE HOLIDAY TIME HIERARCHY	• Date.Year •• Date.Vacation	Hierarchical dimension
PAYMENT HOLIDAY TIME HIERARCHY	• Date.Year •• Date.Vacation	Hierarchical dimension
SALE WORKING DATE HIERARCHY	• Date.Year •• Date.Month ••• Date.WorkingDay	Hierarchical dimension
PAYMENT WORKING DATE HIERARCHY	• Date.Year •• Date.Month ••• Date.WorkingDay	Hierarchical dimension
SALE DAY OF WEEK HIERARCHY	• Date.Year •• Date.Month ••• Date.DayOfWeek	Hierarchical dimension
PAYMENT DAY OF WEEK HIERARCHY	• Date.Year •• Date.Month ••• Date.DayOfWeek	Hierarchical dimension
SALE TIME HIERARCHY	• Time.TimeOfDay •• Time.Hour	Hierarchical dimension
JUNK	Junk	Dimension
SALE POSITION	Junk.Position	Dimension attribute
TYPE OF PAYMENT	Junk.TypeOfPayment	Dimension attribute
AUTHOR	Author	Dimension
AUTHOR NAME	Author.NameAndSurname	Dimension attribute
BOOKSTORE	Bookstore	Dimension
BOOKSTORE NAME	Bookstore.BookstoreName	Dimension attribute
STATE	Bookstore.Province	Dimension attribute
CITY	Bookstore.City	Dimension attribute
BOOKSTORE SIZE	Bookstore.SizeCategory	Dimension attribute
BOOKSTORE LOCALIZATION HIERARCHY	• Bookstore. Province •• Bookstore.City ••• Bookstore. BookstoreName	Hierarchical dimension

Dimensions for Fact 2 Being the author fact:

DIMENSION/DIMENSION ATTRIBUTE	**TABLE/COLUMN**	**TYPE**
BOOK	Book	Dimension
ISBN	Book.ISBN	Dimension attribute
AUTHOR	Author	Dimension

AUTHOR NAME	Author.NameAndSurname	Dimension attribute

Checking the feasibility of queries based on the multidimensional model

1. Compare the number of copies of sold books of various genres in the analyzed versus previous months.
   Measure: Number of book copies sold,
   Dimension: Book (dimension attributes: Book genre)
   Dimension: Sale date (dimension attributes: Sales month)
2. Specify what is the book sale with respect to before-holiday days in this versus previous month.
   Measure: Number of book copies sold,
   Dimension: Sale date (dimension attributes: Sales month, Before holiday sale day)
3. List the best-selling books in this and the previous month?
   Measure: Number of book copies sold,
   Dimension: Sale date (dimension attributes: Sales month)
4. Compare the profits from book sales for individual sellers including current and previous month.
   Measure: Profit,
   Dimension: Sale date (dimension attributes: Sales month)
   Dimension: Seller (dimension attributes: Seller name)
5. What were the most popular authors in this and the previous month?
   Measure: Number of book copies sold,
   Dimension: Author (dimension attributes: Author name)
   Dimension: Sale date (dimension attributes: Sales month)
6. How profits are shaped in relation to individual bookstores in this and the previous months?
   Measure: Profit,
   Dimension: Sale date (dimension attributes: Sales month)
   Dimension: Bookstore (dimension attributes: Bookstore name)
7. Specify the sales volume in relation to the sellers' work experience at a specified position.
   Measure: Number of book copies sold,
   Dimension: Seller (dimension attributes: Seller work experience, Seller position)
8. Provide the sales volume in relation to the size of the bookstore, understood as a number of employees, in this and the previous month.
   Measure: Number of book copies sold,
   Dimension: Bookstore (dimension attributes: Bookstore size)
   Dimension: Sale date (dimension attributes: Sales month)

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

TABLE NAME	COLUMN	SOURCE
BOOK_SALE	**One tuple describes one fact of book sale.**	

	Id_ SaleDate	Sale date Id. Foreign key from dimension table. Based on IssueDate stored in Bill table in BillMaster source.
	Id_PaymentDate	Payment date Id. Foreign key from dimension table. Based on PaymentDate stored in Bill table in BillMaster source.
	Id_Book	Sold book Id. Foreign key from dimension table. Based on ISBN number of a book and a fee paid for it. Based on foreign key FK_Book in table Booksale in BillMaster source and PurchasePrice in fact table Book_sale.
	Id_Seller	Seller Id. Foreign key from dimension table. Based on employee PID number and the age, education, position and work experience at the moment of the sale. Based on Sheet 2 and IssueDate in Bill table in BillMaster source.
	Id_Junk	Junk Id. Foreign key from dimension table. Based on Place and Payment from Bill table in BillMaster source.
	Id_Time	Time Id. Foreign key from dimension table. Based on IssueDate in Bill table in BillMaster source.
	TransactionNo	Transaction number taken from BillNumber column in Bill table in BillMaster source.
	Amount	Number of books sold taken from NumberOfCopies column from Booksale table in BillMaster source.
	Profit	Profit from the fact of book sale equal to Purchase price*1.07. Purchase price taken from PurachasePrice column in Book_sale table in date warehouse.
	PurchasePrice	Purchace price for all items in a bill. Taken from Price column in Booksale table in BillMaster source and multiplied by NumberOfCopies from Booksale table in BillMaster source.
AUTHORSHIP	**One tuple describes one fact of being an author of the book.**	
	Id_Book	Book Id. Foreign key from dimension table sourced in Price column in Booksale table and book's ISBN number sourced in reference FK_Book in Authorship table in BillMaster source.
	Id_Author	Foreign key from dimension table sourced in reference FK_Author in Authorship table in BillMaster source.
BOOK	**One tuple describes one title purchased within one price category.**	

	Id_Book	Book Id. Surrogate key - generated by database
	ISBN	Business key taken from ISBN from Book table in BillMaster source.
	Genre	Book genre. Allowed values: Encyclopedia, album, fantasy, other, informatics, magazine, history, language, cooking, drama, poetry, thriller – taken from Genre from Book table in BillMaster source.
	Title	Book title. Taken from Title from Book table in BillMaster source
	PriceCategory	Price category. Allowed values: Price < 20 - occasion, Price < 50 - cheap, Price < 90 - typical, Price < 150 - expensive, Price) >= 150 absurd. Values bases on Price from Booksale table in BillMaster source.
BOOKSTORE	**One tuple describes one bookstore.**	
	Id_Book	Bookstore Id. Surrogate key – generated by database.
	BookstoreName	Bookstore name. Business key taken from Name column from Bookstore table in BillMaster source.
	Province	Province, where the bookstore is located. Based on columns C, D, E of Sheet 1, and the value can be calculated from https://kody.pocztapolska.pl/index.php.
	City	City, where the bookstore is located. Taken from Sheet 1, column E
	SizeCategory	Bookstore size defined in terms of number of employees. Allowed values: Employee count <=10 - small, Employee count from 11 to 30 - average, Employee count > 30 - big. Values calculated from data in Sheet 2.
SELLER	**One tuple describes one seller, in the specified age category, with the specified education, position, work experience and boss. (Implementation of SCD 2)**	
	Id_Seller	Seller Id. Surrogate key - generated by the database.
	Id_Bookstore	Bookstore Id. Foreign key from dimension table. Value based on data from Sheet 2.
	PID	Employee Personal Identification Number. Taken from PESEL from Salesman table in BillMaster source.
	NameAndSurname	Seller's name and surname. Taken from

		Name and Surname columns in Salesman table in BillMaster source.
	AgeCategory	Age category. Allowed values: from 15 to 17, from 18 to 21, from 22 to 29, from 30 to 49, from 50 to 64), more than 64. Calculated from birthday stored in column E of Sheet 2.
	Education	Seller's education. Allowed values: vocational, incomplete secondary, secondary, incomplete higher, higher, doctorate. Based on value stored in column F of Sheet 2.
	Boss	Employee's boss. Foreign key from dimension table. Value based on column G from Sheet 2. If an employee is not a manager, the information about manager of the given bookstore is stored.
	PositionEmployment	Seller's employment position. Allowed values: seller and manager. Value taken from column G or Sheet 2 and based on IssueDate stored in Bill table in BillMaster source.
	WorkExperience	Work experience. Allowed values: up to one year, from one year to five years, more than five years). Calculated from column H of Sheet 2.
	IsCurrent	"1" if information is current, otherwise "0" (SCD2 implementation).
AUTHOR	**One tuple describes one author.**	
	Id_Author	Author Id. Surrogate key - generated by database.
	NameAndSurname	Author's Name and Surname taken from Name1, Name2, Surname from Author table in BillMaster 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.	
TIME	**One tuple describes one hour (independently of date).** All the data in this table are generated tuple by tuple based on clock, before ETL process.	
JUNK	The tuples correspond to "all" possible combinations of values for Position and TypeOfPayment and are generated before ETL process.	
	Id_Junk	Junk Id. Surrogate key - generated by database.