

## Real estate database: Date warehouse design

### Business process:

The Date warehouse is designed for book sale business process.

### Relational Database schema:

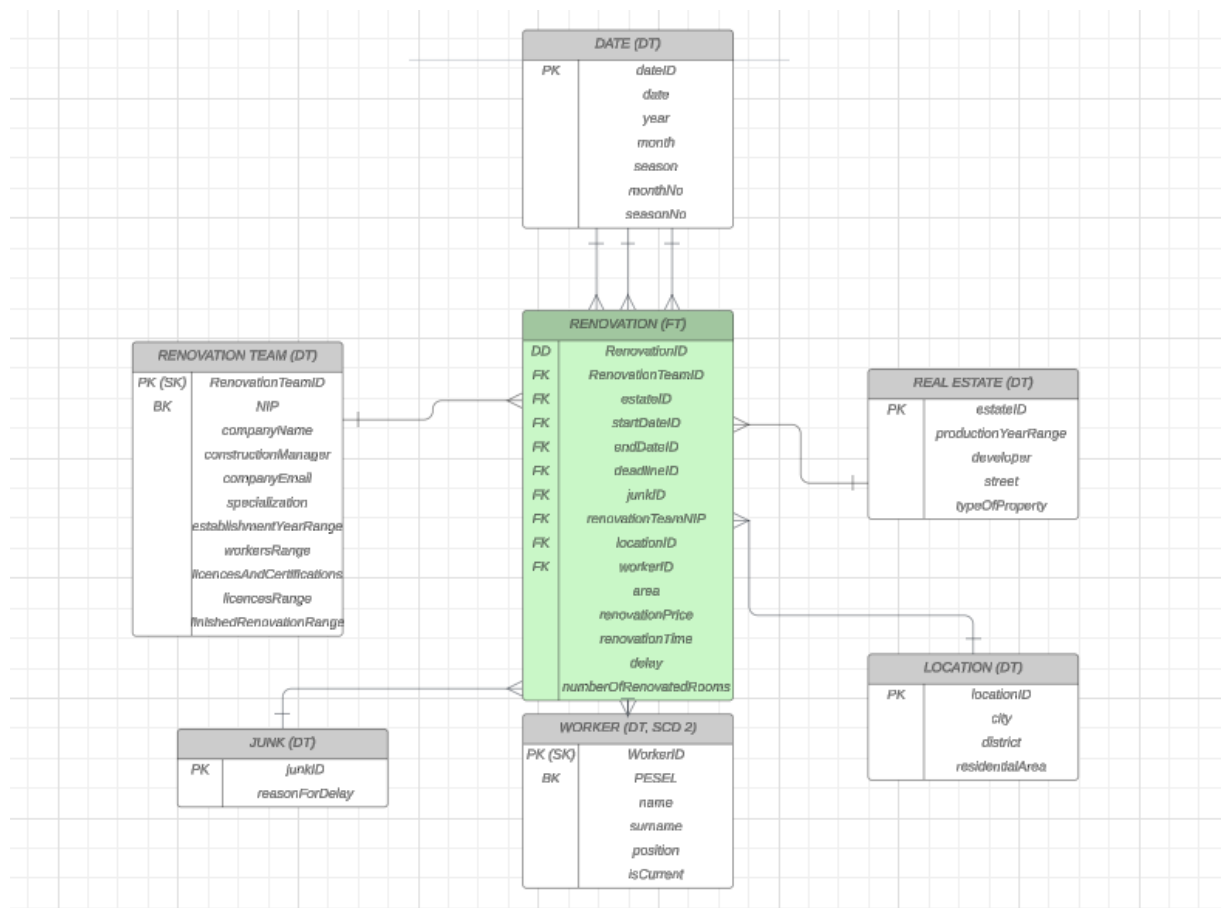


TABLE NAME	ATTRIBUTE	ATTRIBUTE TYPE	DESCRIPTION
Renovation (Fact table)	One tuple describes one fact of renovation		
	RenovationID	Numeric	
	NIP	Numeric	FK renovation team, company nip
	estateID	Numeric	FK Real estate, estate's id

	startDateID	Numeric	FK Date, date of starting the renovation
	endDateID	Numeric	FK Date, date of finishing the renovation
	deadlineID	numeric	FK Date, date of deadline the renovation
	JunkID	Numeric	FK Junk, junk attributes
	LocationID	Numeric	FK, table Location
	workerID	Numeric	FK, table Worker
	Area	Numeric	Area of the renovated property
	renovationPrice	numeric	sum price of the renovation
	renovationTime	Numeric	Quantity of days how long the renovation lasts
	Delay	numeric	Quantity of days how long the delay was
	NumberOfRenovatedRooms	Numeric	Quantity of renovated rooms during given renovation
Renovation Team ( dimension table)	One tuple describes one renovation company		
	NIP	Numeric	PK
	companyName	Varchar(30)	Name of the renovation team
	constructionManager	Varchar(30)	Name and surname of the renovation team's manager

	companyEmail	Varchar(30)	Renovation team's email
	Specializations	Varchar(300)	All specialization of company, possible values:
	establishYearRange	Varchar(30)	year category. Allowed values: between 1900-1950 between 1951-1980 between 1981-2000 between 2001-2024
	workersRange	Varchar(30)	worker category. Allowed values: between 1-5 between 6-15 between 16-50 between 50-200 over 201
	licensesAndCertifications	Varchar(300)	All certifications and licenses that company posses
	licensesRange	Varchar(30)	licenses category. Allowed values: between 0-5 between 6-15 between 16-50 over 51
	finishedRenovationRange	Varchar(30)	Finished Renovation category. Allowed values: between 0-5, between 6-20, between 21-100, between 101-500, more than 500.
Real estate (dimension table)	One tuple describes one real estate, with the specific street, type of property, developer, and in specific production year range		
	estateID	Numeric	PK
	Street	Varchar(30)	the real estate's street

	Typeofproperty	Varchar(30)	Category to choose: flat, apartment, house, twin
	developer	Varchar(30)	Name of the developer, that built the real estate
	productionYearRange	Varchar(30)	year category. Allowed values: between 1900-1950 between 1951-1980 between 1981-2000 between 2001-2010 between 2011-2020 between 2021 - 2024
Location (dimension table)	One tuple describes one location		
	LocationID	numeric	PK
	City	Varchar(30)	Location's city of property
	District	Varchar(30)	Location's district of property
	ResidentialArea	Varchar(30)	Location's residential area of the real estate
Worker(dimension table, scd 2)	One tuple describes one worker		
	Pesel	numeric	PK
	Name	Varchar(30)	Worker's name
	Surname	Varchar(30)	Worker's surname
	isCurrent	Boolean	Checking if the worker is still working in a company, possible values : 'yes', 'no'
	position	Varchar(30)	Position of the worker. Possible values: researcher, analyst, CEO, supervisor

Date( dimension table)	One tuple describes one day		
	dateID	Numeric	PK
	Date	Datetime	Exact date of the date ID in format: dd.mm.yyyy
	Year	Numeric	Year of the date
	Month	Varchar(30)	Possible categories: January, February, march, April, may, June, July, august, October, November, December
	season	Varchar(30)	Possible categories: winter, spring, summer, august
	monthNo	numeric	Number of the month as a integer, range( 1-12)
	seasonNo	numeric	Number of the season as a integer range(1-4)
Junk( dimension table)	One tuple describes one “junk” information		
	JunkID	numeric	PK
	reasonforDelay	Varchar(30)	Category to choose: Material shortages, Unexpected structural issues , Weather-related setbacks, Changes in project scope, Labor shortages, Budgetary constraints. The explanation why the delay occurred

## Dimensional model

### Dimension model

Fact 1 Renovation fact: Renovation of the specific real estate (whole or some part of it), started and ended on the specific date with a specific deadline. Conducted by a specific renovation team with a specific NIP, specific name and a specific construction manager. The specific real estate has a specific location, specific street, specific developer and specific type of property.

Fact\_table: Renovation

Granularity:

- A specific renovation
- A specific start date
- A specific end date
- A specific deadline
- A specific real estate, with specific street, location, developer, and type of property
- A specific renovation team, with specific NIP, name and specific construction manager

Measures and aggregation functions:

Number of renovations - COUNT (1)

Number of real estate - DISTINCT COUNT (realEstateID)

Number of renovation teams - DISTINCT COUNT (renovationTeamID)

SumOfDelay = SUM(Delay)

DelayCount = DISTINCT COUNT(Delay)

Sum of renovation time – SUM(Renovation time)

SumOfRenovatedRooms – SUM(NumerRenovatedRooms)

RenovatedRoomsCount – DISTINCT COUNT(numberRenovatedRooms)

SumOfArea = SUM(Area)

Average renovation time – Sum of renovation time / number of renovations

Average delay – SumOfDelay / DelayCount

Average area = SumOfArea / Number of Renovations

AverageRenovatedRooms – SumOfRenovatedRooms/ RenovatedRoomsCount

Dimension definition

Dimension for fact 1:

DIMENSION/DIMENSION ATTRIBUTE	TABLE/COLUMN	TYPE
Renovation number	Renovation.RenovationID	Degenerate dimension
Real estate	RealEstate	dimension
Street	RealEstate.street	Dimension attribute
Type of property	RealEstate.typeOfProperty	Dimension attribute
Developer of building	RealEstate.Developer	Dimension attribute
Production year range	RealEstate.productionYearRange	Dimension attribute
Property location	Location	Dimension
Property City	Location.city	Dimension attribute
Property district	Location.district	Dimension attribute
Property Residential area	Location.ResidentialArea	Dimension attribute
LOCATION HIERARCHY	<ul style="list-style-type: none"> <li>● Location.city</li> <li>● ● Location.district</li> <li>● ● ● Location.ResidentialArea</li> </ul>	Hierarchical dimension
Renovation team	Renovation Team	Dimension
Name of team	RenovationTeam.CompanyName	Dimension attribute
Name&surname of leader	RenovationTeam.ConstructionManager	Dimension attribute
Company email	RenovationTeam.CompanyEmail	Dimension attribute
Company's specialization	RenovationTeam.specialization	Dimension attribute
Company establish year in a range	RenovationTeam.establishYearRange	Dimension attribute

Range of renovation team workers	RenovationTeam.workersRange	Dimension attribute
All certifications and licenses obtained by renovation team	RenovationTeam.CertificationAndLicenses	Dimension attribute
Number of licenses in a range	RenovationTeam.LicensesRange	Dimension attribute
Number of finished renovations in a range	RenovationTeam.FinishedRenovationRange	Dimension attribute
Worker ID	Worker	Dimension
Worker's name	Worker.name	Dimension attribute
Worker's surname	Worker.surname	Dimension attribute
Worker's position	Worker.position	Dimension attribute
If still working	Worker.isCurrent	Dimension attribute
StartDateID	Date	Dimension
Startdate	Date.date	Dimension attribute
StartYear	Date.year	Dimension attribute
StartMonth	Date.month	Dimension attribute
Startseason	Date.season	Dimension attribute
Number of the month	Date.monthNo	Dimension attribute
Number of the season	Date.seasonNo	Dimension attribute
STARTDATE HIERARCHY	<ul style="list-style-type: none"> <li>● Date.Year</li> <li>● ● Date.season</li> <li>● ● ● Date.month</li> </ul>	Hierarchical dimension
EndDateID	Date	Dimension
Enddate	Date.date	Dimension attribute
EndYear	Date.year	Dimension attribute
EndMonth	Date.month	Dimension attribute
Endseason	Date.season	Dimension attribute
Number of the month	Date.monthNo	Dimension attribute
Number of the season	Date.seasonNo	Dimension attribute
ENDDATE HIERARCHY	<ul style="list-style-type: none"> <li>● Date.Year</li> <li>● ● Date.season</li> <li>● ● ● Date.month</li> </ul>	Hierarchical dimension



DeadlineDateID	Date	Dimension
Deadlinedate	Date.date	Dimension attribute
DeadlineYear	Date.year	Dimension attribute
DeadlineMonth	Date.month	Dimension attribute
Deadlineseason	Date.season	Dimension attribute
Number of the month	Date.monthNo	Dimension attribute
Number of the season	Date.seasonNo	Dimension attribute
DEADLINE HIERARCHY	● Date.Year ● ● Date.season ● ● ● Date.month	Hierarchical dimension
ID of the “junk” information	Junk	Dimension
Reason, why the finish of renovation was delayed	Junk.reasonForDelay	Dimension attribute

## Checking the feasibility of queries based on the multidimensional model

1. Which renovation team compared to previous months have the highest price change of renovation.

*Measure:* Renovation price,

*Dimension:* Renovation team (*dimension attributes:* NIP)

*Dimension:* endDateID (*dimension attributes:* endMonth)

2. What is the most often reason for delay in current and previous month.

*Measure:* Reason for delay,

*Dimension:* Junk (*dimension attributes:* reasonForDelay)

*Dimension:* endDateID (*dimension attributes:* endMonth)

3. What is the most common certification for each type of building?

*Measure:* certification,

*Dimension:* RenovationTeam (*dimension attributes:* licencesRange)

*Dimension:* Real Estate (*dimension attributes:* typeOfProperty)

4. Compare average delay days of renovation team in current and previous month

*Measure:* average delay days of renovation,

*Dimension:* Renovation (*dimension attributes:* delay)

*Dimension: endDateID (dimension attributes: endMonth)*

5. Compare average days of delay for each supervisor

*Measure: average delay days,*

*Dimension: Renovation (dimension attributes: delay)*

*Dimension: Renovation team (dimension attributes: constructionManager)*

6. How does the type of property affect the average number of renovated rooms?

*Measure: average number of renovated rooms,*

*Dimension: Renovation (dimension attributes: numberOfRenovatedRooms)*

*Dimension: Real estate (dimension attributes: typeOfProperty)*

7. In which cities and districts were there the most different renovation teams and in which the least?

*Measure: distinct renovation teams,*

*Dimension: Renovation team (dimension attributes: NIP)*

*Dimension: Location (dimension attributes: City, District)*

8. What is the average number of workers in each city?

*Measure: average number of workers,*

*Dimension: Renovation team (dimension attributes: NIP)*

*Dimension: Location (dimension attributes: City, District)*

9. How does the number of specializations of a renovation team affect the average price of the renovation?

*Measure: average price of the renovation,*

*Dimension: Renovation (dimension attributes: renovationPrice)*

*Dimension: Renovation team (dimension attributes: licencesRange)*

10. What is the average number of workers of a team for each area of renovated property?

*Measure: area of renovated property,*

*Dimension: Renovation (dimension attributes: area)*

*Dimension: Renovation team (dimension attributes: workersRange)*

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

Table name	Column	Source
<b>Renovation</b>	<b>One tuple describes one fact of renovation</b>	
	Renovation ID	
	NIP	Company's NIP. Foreign key from dimension table. Based on Renovation_team table in PropertyMaster source.
	estateID	Estate ID. Foreign key from dimension table. Based on estateID stored in Real_estate table in PropertyMaster source.
	startDateID	Renovation start date ID. Foreign key from dimension table. Based on start_date stored in Renovation table in PropertyMaster source.
	endDateID	Renovation end date ID. Foreign key from dimension table. Based on end_date stored in Renovation table in PropertyMaster source.
	deadlineID	Renovation deadline ID. Foreign key from dimension table. Based on deadline stored in Renovation table in PropertyMaster source.
	junkID	Junk Id. Foreign key from dimension table. Based on column F from RenovationInfo table in PropertyMaster CSV.
	locationID	Renovation location ID. Foreign key from dimension table. Based on the city, district and residential area of the renovated real estate. Based on Location table in PropertyMaster source.
	workerID	Worker's PESEL. Foreign key from dimension table. Based on Worker table in PropertyMaster source.
	area	Area of the renovated property.

		Based on the Area column in Real_Estate table in PropertyMaster source.
	renovationPrice	Total price of all renovations. Based on the Price column in Renovation table in PropertyMaster source.
	renovationTime	The total time of the renovation equal to end date – start date. End date and start date both taken from respective columns in Renovation table in date warehouse.
	delay	The total delay time of the renovation equal to 0 when deadline is later than end date, in the other case it is equal to deadline – end date. End date and deadline both taken from respective columns in Renovation table in date warehouse.
	numberOfRenovatedRooms	Number of rooms that were renovated during renovation taken from column G from renovationInfo table from PropertyMaster csv.
<b>RenovationTeam</b>	<b>One tuple describes one renovation company</b>	
	NIP	Company's NIP Number. Taken from NIP column from Renovation_Team table in PropertyMaster source.
	companyName	Company's name. Taken from companyName column from Renovation_Team table in PropertyMaster source.
	constructionManager	Company's construction manager. Taken from constructionManager column from Renovation_Team table in PropertyMaster source.
	companyEmail	Company's email. Taken from column D from RenovationTeams table in PropertyMaster CSV.
	specializations	Company's specializations. Taken from column F from RenovationTeams table in

		PropertyMaster CSV.
	establishYearRange	Establishment year category. Allowed values: between 1900-1950, between 1951-1980, between 1981-2000, between 2001-2024.
	workersRange	Number of workers category. Allowed values: between 1-5, between 6-15, between 16-50, between 50-200, over 200.
	licencesAndCertifications	Company's licences and certifications. Taken from column I from RenovationTeams table in PropertyMaster CSV.
	licencesRange	Licenses category. Allowed values: between 0-5, between 6-15, between 16-50, over 50.  Calculated from column J from RenovationTeams sheet from PropertyMaster CSV.
	finishedRenovationRange	Finished Renovation category. Allowed values: between 0-5, between 6-20, between 21-100, between 101-500, more than 500.  Calculated from column K from RenovationTeams sheet from PropertyMaster CSV.
<b>Real estate</b>	<b>One tuple describes one real estate, with the specific street, type of property, developer, and in specific production year range</b>	
	estateID	Estate ID. Surrogate key – generated by the database.
	street	Street, on which the real estate is located taken from Street column from Real_Estate table in PropertyMaster source.

	typeOfProperty	Type of property of the real estate taken from typeOfProperty column from Real_Estate table in PropertyMaster source.
	developer	Developer of the real estate taken from developer column from Real_Estate table in PropertyMaster source.
	productionYearRange	<p>Production year category. Allowed values:  before 1900,  between 1900-1950,  between 1951-1980,  between 1981-2000,  between 2001-2010,  between 2011-2020,  between 2021-2024.</p> <p>Calculated from productionYear column from Real_Estate table in PropertyMaster source.</p>
<b>Location</b>	<b>One tuple describes one location</b>	
	locationID	Location ID. Surrogate key – generated by the database.
	city	City, where the real estate is located taken from City column from Location table in PropertyMaster source.
	district	District, where the real estate is located taken from District column from Location table in PropertyMaster source.
	residentialArea	Residential area, where the real estate is located taken from residentialArea column from Location table in PropertyMaster source.
<b>Worker</b>	<b>One tuple describes one worker</b>	
	PESEL	Worker's Personal Identification Number. Taken from PESEL column from Worker table in PropertyMaster source.
	name	Worker's name taken from name column from Worker table in

		PropertyMaster source.
	surname	Worker's surname taken from surname column from Worker table in PropertyMaster source.
	isCurrent	"1" if information is current, otherwise "0" (SCD2 implementation).
	position	Worker's employment position.
<b>Date</b>	<b>One tuple describes one day</b>  All the data in this table are generated tuple by tuple based on any calendar, before ETL process.	
<b>Junk</b>	<b>The tuple corresponds to "all" possible reasons for delay provided by renovation companies</b>	
	junkID	Junk Id. Surrogate key - generated by database.