

School – Date warehouse design

Business process

The Date warehouse is designed for teaching business process which includes getting marks and attendance during lessons. This process is described in the document Specification for Teaching Students business process.

Relational Database schema

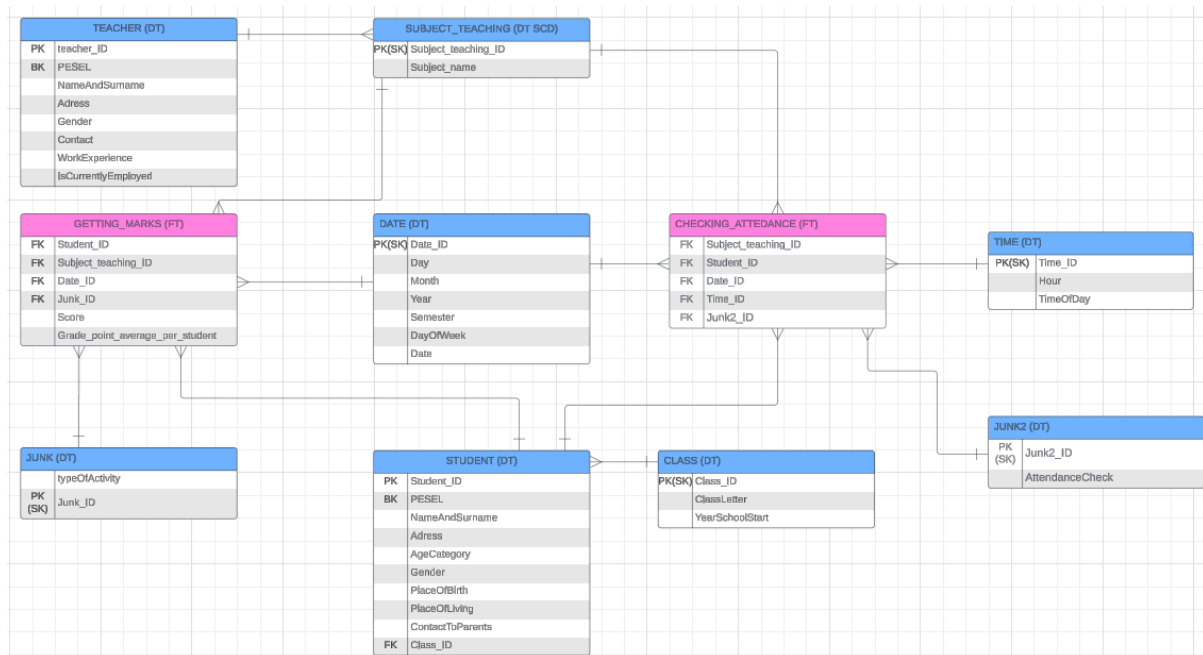


TABLE NAME	ATTRIBUTE	ATTRIBUTE TYPE	DESCRIPTION
GETTING_MARKS	One tuple describes a fact of getting one mark by one student.		

(FACT TABLE)			
	Score	Numeric	Score gained Allowed values from 1 to 6
	Student_ID	Numeric	FK Student Student who gained the mark
	Subject_teaching_ID	Numeric	FK Subject_teaching The subject in which the grade was obtained
	MarkDate_ID	Numeric	FK Date Date of assessment
	Junk_ID	Numeric	FK Junk Information about activity in which the mark was obtained
	Mark_num	Numeric	Number of the mark
STUDENT (DIMENSION TABLE)	One tuple describes one student		
	Student_ID	Varchar(6)	PK
	PESEL	Varchar(11)	Student's PESEL Number
	NameAndSurname	Varchar (50)	Name and surname.
	Address	Varchar (130)	Street, number of the house/flat, name of the city/countryside
	AgeCategory	Varchar(20)	Age category of student, allowed values: between 9 and 10, from 11 to 12, from 13 to 14
	PlaceOfBirth	Varchar(15)	Name of the city where the student was born
	Gender	Varchar(6)	Gender of the student Allowed values: male, female
	PlaceOfLiving	Varchar(11)	Category of the place of living. Allowed values: City, countryside
	ContactToParent	Varchar(30)	Phone number and email address of a parent
	Class_ID	Numeric	FK Class

			Class to which student is assigned
CLASS (DIMENSION TABLE)	One tuple describes one class		
	Class_ID	Numeric	PK
	ClassLetter	Char	Letter of a class
	YearSchoolStarts	Varchar(4)	Year
TEACHER (DIMENSION TABLE)	One tuple describes one teacher		
	Teacher_ID	Numeric	PK
	PESEL	Varchar(11)	Teacher's PESEL Number
	NameAndSurname	Varchar (50)	Name and surname.
	Address	Varchar (80)	Street, number of the house/flat, name of the city/countryside
	Gender	Varchar(5)	Gender Allowed values: Woman, Man
	Contact	Varchar(30)	Phone number and email address
	WorkExperience	Varchar(30)	Work experience. Allowed values: up to two years, more than two years.
	IsCurrentlyEmployed	Varchar(3)	"YES" if information is current, otherwise "NO".
SUBJECT_TEACHING (DIMENSION TABLE, SCD)	One tuple describes one subject.		
	Subject_teaching_ID	Numeric	PK
	Subject_name	Varchar(11)	Subject Allowed values: Mathematics, Polish, English, Science, WOS, History, Music, Art, Religion, PE, IT, German.
	Teacher_ID	Numeric	FK Teacher who performed teaching.
	isCurrent	Varchar(3)	"YES" if information is current, otherwise "NO".
CHECKING_ATTENDANCE (FACT TABLE)	One tuple describes one fact about checking attendance type (present, absent) for particular lesson		
	Subject_teaching_ID	Numeric	FK

	Student_ID	Numeric	FK Student Student who got the attendance marked
	AssessmentDate_ID	Numeric	FK Date Date of issue of attendance
	Time_ID	Numeric	FK Time Time of issue of attendance
	Junk2_ID	Numeric	FK Junk2 Information about attendance check: present or absent
JUNK (DIMENSION TABLE)	One tuple describes one mark received for particular assignment.		
	Junk_ID	Numeric	PK
	TypeOfActivity	Varchar (8)	Type of assessment Allowed values: Exam, homework, test
JUNK2 (DIMENSION TABLE)	One tuple describes one mark received for particular assignment.		
	Junk2_ID	Numeric	PK
	AttendanceCheck	Varchar(10)	Recorded attendance Allowed values: "absent", "present"
DATE (DIMENSION TABLE)	One tuple describe one day.		
	Date	Date	Date
	Day	Numeric	Day numeric value
	Month	Varchar(10)	Month. Allowed values: January, February, March, April, May, June, July, August, September, October, November and December.
	Year	Varchar(4)	Year
	Semester	Varchar(15)	Semester Allowed values: "Winter semester", "Summer semester"
	DayOfWeek	Varchar(10)	Day of week. Allowed values: Monday, Tuesday, Wednesday,

			Thursday, Friday, Saturday and Sunday
	Date_ID	Numeric	PK
TIME (DIMENSION TABLE)	One tuple describes one hour (independently on date).		
	Time_ID	Numeric	PK
	Hour	Numeric	Hour. Allowed values from 8 – 15.
	TimeOfDay	Varchar(10)	8-9 „ Lesson 1” 9-10 „ Lesson 2” 10-11 „ Lesson 3” 11-12 „ Lesson 4” 12-13 „ Lesson 5” 13-14 „ Lesson 6” 14-15 „ Lesson 7”

Dimensional model

Fact definitions

Fact 1 – Getting marks

Getting marks (grades from 1 to 6, 1-lowest grade, 6-highest grade), achieved from specified subject performed by specific teacher, on specified day, from specified activity. Assigned to the specified student who is attending a specific class.

Fact table – getting marks

Granularity:

- A specified mark
- A specified student who attends a specified class
- A specified subject within which the assessment was made performed by specific teacher
- A specified date of assessment
- A specified type of activity

Measures and aggregation functions:

Number of students – COUNT(Students)

Number of grades obtained – COUNT (Score)

Sum of grades – SUM(Score)

Grade point average of student - Sum of grades/ Number of grades obtained

Grade point average of class – SUM(Grade point average of student)/ Number of students

People graduate with honour certificate – COUNT(Grade point average of student>4,5)

DIMENSION/DIMENSION ATTRIBUTE	TABLE/COLUMN	TYPE
MARK NUMBER	Getting_marks.mark_num	Degenerate dimension
MARK DATE HIERARCHY	<ul style="list-style-type: none"> • Date.Year •• Date.Semester ••• Date.Month •••• Date.Date 	Hierarchical dimension
MARK DATE	Date	Dimension
MARK DAY	Date.Day	Dimension attribute
MARK MONTH	Date.Month	Dimension attribute
MARK YEAR	Date.Year	Dimension attribute
MARK SEMESTER	Date.Semester	Dimension attribute
MARK DAY OF WEEK HIERARCHY	<ul style="list-style-type: none"> • Date.Year •• Date.Semester ••• Date.Month •••• Date.DayOfWeek 	Hierarchical dimension
SUBJECT_TEACHING	SubjectTeaching	Dimension
SUBJECT_NAME	SubjectTeaching.SubjectName	Dimension attribute
STUDENT	Student	Dimension
STUDENT_ID	Student.StudentID	Dimension attribute
PESEL	Student.PESEL	Dimension attribute
NAME AND SURNAME	Student.NameAndSurname	Dimension attribute
ADRESS	Student.Address	Dimension attribute
AGE CATEGORY	Student.AgeCategory	Dimension attribute
PLACE OF BIRTH	Student.PlaceOfBirth	Dimension attribute
GENDER	Student.Gender	Dimension attribute
PLACE OF LIVING	Student.PlaceOfLiving	Dimension attribute
CONTACT TO PARENT	Student.ContactToParent	Dimension attribute
LOCALISATION HIERARCHY	<ul style="list-style-type: none"> • Student.PlaceOfLiving •• Student.Address 	Hierarchical dimension
CLASS	Class	Dimension
CLASS LETTER	Class.ClassLetter	Dimension attribute
YEAR SCHOOL STARTS	Class.YearSchoolStarts	Dimension attribute
JUNK	Junk	Dimension
ACTIVITY TYPE	Junk.typeOfActivity	Dimension attribute
TEACHER	Teacher	Dimension
TEACHER_ID	Teacher.Teacher_ID	Dimension attribute
PESEL	Teacher.PESEL	Dimension attribute
NAME AND SURNAME	Teacher.NameAndSurname	Dimension attribute
ADRESS	Teacher.Address	Dimension attribute
GENDER	Teacher.Gender	Dimension attribute
CONTACT	Teacher.Contact	Dimension attribute
WORK EXPERIENCE	Teacher.WorkExperience	Dimension attribute
IS CURRENTLY EMPLOYED	Teacher.IsCurrentlyEmployed	Dimension attribute

Fact 2 – Checking attendance

Checking attendance (0 – present, 1 – absent), checked on specified lesson (subject), on specified day, at a specified time. Assigned to the specified student who is attending a specified class.

Fact table: Checking_attendance

Granularity:

- A specified attendance (present or absent)
- A specified student who attends a specified class
- A specified subject within which the attendance was checked performed by specified teacher
- A specified date of checking attendance
- A specified time of checking attendance

Measures and aggregation functions:

Number of Checking Attendance facts – COUNT(1)

DIMENSION/DIMENSION ATTRIBUTE	TABLE/COLUMN	TYPE
ATTENDANCE DATE HIERARCHY	<ul style="list-style-type: none">• Date.Year•• Date.Semester••• Date.Month•••• Date.Date	Hierarchical dimension
ATTENDANCE DATE	Date	Dimension
ATTENDANCE DAY	Date.Day	Dimension attribute
ATTENDANCE MONTH	Date.Month	Dimension attribute
ATTENDANCE YEAR	Date.Year	Dimension attribute
ATTENDANCE SEMESTER	Date.Semester	Dimension attribute
ATTENDANCE DAY OF WEEK HIERARCHY	<ul style="list-style-type: none">• Date.Year•• Date.Semester••• Date.Month•••• Date.DayOfWeek	Hierarchical dimension
TIME HIERARCHY	<ul style="list-style-type: none">• Time.TimeOfDay•• Time.Hour	Hierarchical dimension
SUBJECT_TEACHING	SubjectTeaching	Dimension
SUBJECT_NAME	SubjectTeaching.SubjectName	Dimension attribute
TEACHER	Dimension	Dimension
TEACHER_ID	Teacher.Teacher_ID	Dimension attribute
PESEL	Teacher.PESEL	Dimension attribute
NAME AND SURNAME	Teacher.NameAndSurname	Dimension attribute
ADRESS	Teacher.Address	Dimension attribute
GENDER	Teacher.Gender	Dimension attribute
CONTACT	Teacher.Contact	Dimension attribute

WORK EXPERIENCE	Teacher.WorkExperience	Dimension attribute
IS CURRENTLY EMPLOYED	Teacher.IsCurrentlyEmployed	Dimension attribute
STUDENT	Student	Dimension
STUDENT_ID	Student.StudentID	Dimension attribute
PESEL	Student.PESEL	Dimension attribute
NAME AND SURNAME	Student.NameAndSurname	Dimension attribute
ADDRESS	Student.Address	Dimension attribute
AGE CATEGORY	Student.AgeCategory	Dimension attribute
PLACE OF BIRTH	Student.PlaceOfBirth	Dimension attribute
GENDER	Student.Gender	Dimension attribute
PLACE OF LIVING	Student.PlaceOfLiving	Dimension attribute
CONTACT TO PARENT	Student.ContactToParent	Dimension attribute
LOCALISATION HIERARCHY	<ul style="list-style-type: none"> • Student.PlaceOfLiving • Student.Address 	Hierarchical dimension
CLASS	Class	Dimension
CLASS LETTER	Class.ClassLetter	Dimension attribute
YEAR SCHOOL STARTS	Class.YearSchoolStarts	Dimension attribute
JUNK 2	Junk2	Dimension
ATTENDANCE CHECK	Junk2.AttendanceCheck	Dimension attribute

Checking the feasibility of queries based on the multidimensional model

1. Compare the number of people (on the same level for each class) who graduated with an honours certificate in the previous year and 2 years before. (Grade point average above 4.5).

Measure: People graduate with honour certificate

Dimension: Class (*dimension attributes:* Class YearSchoolStarts)

Dimension: Date (*dimension attributes:* Date Year)

2. At which level of learning (in which grade 4-8) children receive the highest grades?
Based on previous year.

Measure: Grade point average of class

Dimension: Class (*dimension attributes:* Class YearSchoolStarts)

Dimension: Date (*dimension attributes:* Date Year)

3. Compare the academic grade point average of students based on tutoring of teachers who work for this school less than 2 years or more (teacher who are currently employed). Based on previous year.

Measure: Grade point average of students

Dimension: Teacher (*dimension attributes:* Teacher WorkExperience, IsCurrentlyEmployed)

Dimension: Date (dimension attributes: Date Year)

4. Whether the place of living (countryside/city) influences the students' grade point average.

Measure: Grade point average of students

Dimension: Student (dimension attributes: Student PlaceOfLiving)

5. Does the number of absences affect a student's final grade point average? Based on the previous year.

Measure: Grade point average per students

Measure: Number of Checking Attendance facts

Dimension: Date (dimension attributes: Date Year)

Dimension: Junk2 (dimension attribute: AttendanceCheck)

6. Which class has the highest absence rate in the last semester?

Measure: Number of Checking Attendance facts

Dimension: Class (dimension attribute: Class_ID, YearSchoolStarts, ClassLetter)

Dimension: Date (dimension attribute: attendance semester, attendance year)

Dimension: Junk2 (dimension attribute: AttendanceCheck)

7. Which subject has the highest number of absences?

Measure: Number of Checking Attendance facts

Dimension: Subject_teaching (dimension attribute: Subject_name)

Dimension: Junk2 (dimension attribute: AttendanceCheck)

8. During which semester (winter/summer) there were more absences (last academics year)?

Measure: Number of Checking Attendance facts

Dimension: Date (dimension attribute: attendance semester, attendance year)

Dimension: Junk2 (dimension attribute: AttendanceCheck)

9. Compare the number of absences among genders (boys and girls).

Measure: Number of Checking Attendance facts

Dimension: Student (dimension attribute: Gender)

Dimension: Junk2 (dimension attribute: AttendanceCheck)

10. Compare the number of absences of children taking into account their age. For the previous year.

Measure: Number of Checking Attendance facts

Dimension: Student (dimension attribute: AgeCategory)

Dimension: Date (dimension attribute: attendance Year)

Dimension: Junk2 (dimension attribute: AttendanceCheck)

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

TABLE NAME	COLUMN	SOURCE
GETTING_MARKS (FACT TABLE)	One tuple describes a fact of getting one mark by one student.	
	Student_ID	Student ID. Foreign key from dimension table. Based on the Student ID, PESEL, place of living, gender, dateOfBirth and ClassId to which the student is assigned. Based on the Sheet 2 and AssessmentDate in the Marks table in EduTracker source.
	Subject_teaching_ID	Subject teaching ID. Foreign key from dimension table. Based on the Name stored in Subjects table in EduTracker source.
	Junk_ID	Junk ID. Foreign key from dimension table. Based on the TypeOfActivity from Marks table from EduTracker source.
	MarkDate_ID	Mark date ID. Foreign key from dimension table. Based on the AssesmentDate from Marks table from EduTracker source.
	Score	Score of a student based on the Score column in Marks table from EduTracker source. Allowed values from 1 to 6.
	Mark_num	Mark number based on the Mark_id column in Marks table from EduTracker source.
STUDENT (DIMENSION TABLE)	One tuple describes one student.	
	Student_ID	Student ID, taken from StudentID column from Students table from EduTracker source.
	PESEL	Student PESEL, business key taken from PESEL column from Students table from EduTracker source.
	NameAndSurname	Student Name and Surname, taken from Name column and Surname column from Students table from EduTracker source.
	Address	Student address, taken from columns H, I, J from Sheet 2 from Director source.
	AgeCategory	Student age category, calculated based on column E from Sheet 2 from Director source. Allowed values between 9 and 10, from 11 to 12, from 13 to 14
	PlaceOfBirth	Student place of birth, taken from column F from Sheet 2 from Director source.
	Gender	Student gender, taken from Column G from Sheet 2 from Director data source.
	PlaceOfLiving	Student place of living based on the column K from Sheet 2 from Director source. Allowed values: City, Countryside

	ContactToParent	Student contact to parents, email and phone number based on the columns L, M from Sheet 2 from Director source.
	Class_ID	Class ID. Foreign key from dimension table. Value based on data from Class table from EduTracker source.
	CLASS (DIMENSION TABLE)	
	Class_ID	Class ID, Surrogate key - generated by the database.
	ClassLetter	Class letter of a class based on the ClassLetter column in Class table from EduTracker source.
	YearSchoolStarts	Year when class started school, data taken from YearSchoolStarts column from Classes table from EduTracker source.
TEACHER (DIMENSION TABLE)		One tuple describes one teacher.
	Teacher_ID	Employee ID, taken from EmployeeID column from Teachers table from EduTracker source.
	PESEL	Teacher PESEL, business key taken from PESEL column from Teachers table from EduTracker source.
	NameAndSurname	Teacher Name and Surname, taken from Name column and Surname column from Teachers table from EduTracker source.
	Address	Teacher address, taken from columns G, H, I from Sheet 1 from Director source.
	Gender	Teacher gender based on the column E from Sheet 1 from Director source.
	Contact	Teacher contact, email and phone number based on the columns J, K from Sheet 1 from Director source.
	WorkExperience	Teacher work experience, based on the years working, calculated from Column L from Sheet 1 from Director source. Allowed values: up to two years, more than two years.
	IsCurrentlyEmployed	Is teacher currently employed, based on the column M from Sheet 1 from Director source. Allowed values: "YES", "NO".
SUBJECT_TEACHING (DIMENSION TABLE, SCD)		One tuple describes one subject.
	Subject_teaching_ID	Subject_teaching ID, Surrogate key - generated by the database.
	Subject_name	Subject name, taken from Name column from Subjects table from EduTracker source. Allowed values: Mathematics, Polish, English, Science, WOS, History, Music, Art, Religion, PE, IT, German.

	Teacher_ID	Teacher ID. Foreign key from dimension table. Based on the employee ID, PESEL, gender, work experience in the moment of assessment of mark. Based on the Sheet 1.
	isCurrent	"YES" if information is current, otherwise "NO".
CHECKING_ATTENDANCE (FACT TABLE)	One tuple describes one fact about checking attendance type (present, absent) for particular lesson	
	Subject_teaching_ID	Subject teaching ID. Foreign key from dimension table. Based on the Name stored in Subjects table in EduTracker source.
	Student_ID	Student ID. Foreign key from dimension table. Based on the Student ID, PESEL, place of living, gender, dateOfBirth and ClassID of the class to which the student is assigned. Based on the Sheet 2 and AssessmentDate in the Marks table in EduTracker source.
	AttendanceDate_ID	Attendance Date ID. Foreign key from dimension table. Based on the DateOfLesson from Lessons table from EduTracker source.
	Time_ID	Time ID. Foreign key from dimension table. Based on the StartTime from ScheduleLessons table from EduTracker source.
	Junk2_ID	Junk2 ID. Foreign key from dimension table. Based on the Checker from Attendances table from EduTracker source.
JUNK (DIMENSION TABLE)	The tuples correspond to "all" possible combinations of values for TypeOfActivity are generated before ETL process.	
	Junk_ID	Junk ID, Surrogate key - generated by the database.
DATE (DIMENSION TABLE)	One tuple describes one day. All the data in this table are generated tuple by tuple based on school calendar containing only school days, before ETL process.	
TIME (DIMENSION TABLE)	One tuple describes one hour (independently of date). All the data in this table are generated tuple by tuple based on clock from 8 to 15, before ETL process.	
JUNK2 (DIMENSION TABLE)	The tuples correspond to "all" possible combinations of values for AttendanceCheck are generated before ETL process.	
	Junk2_ID	