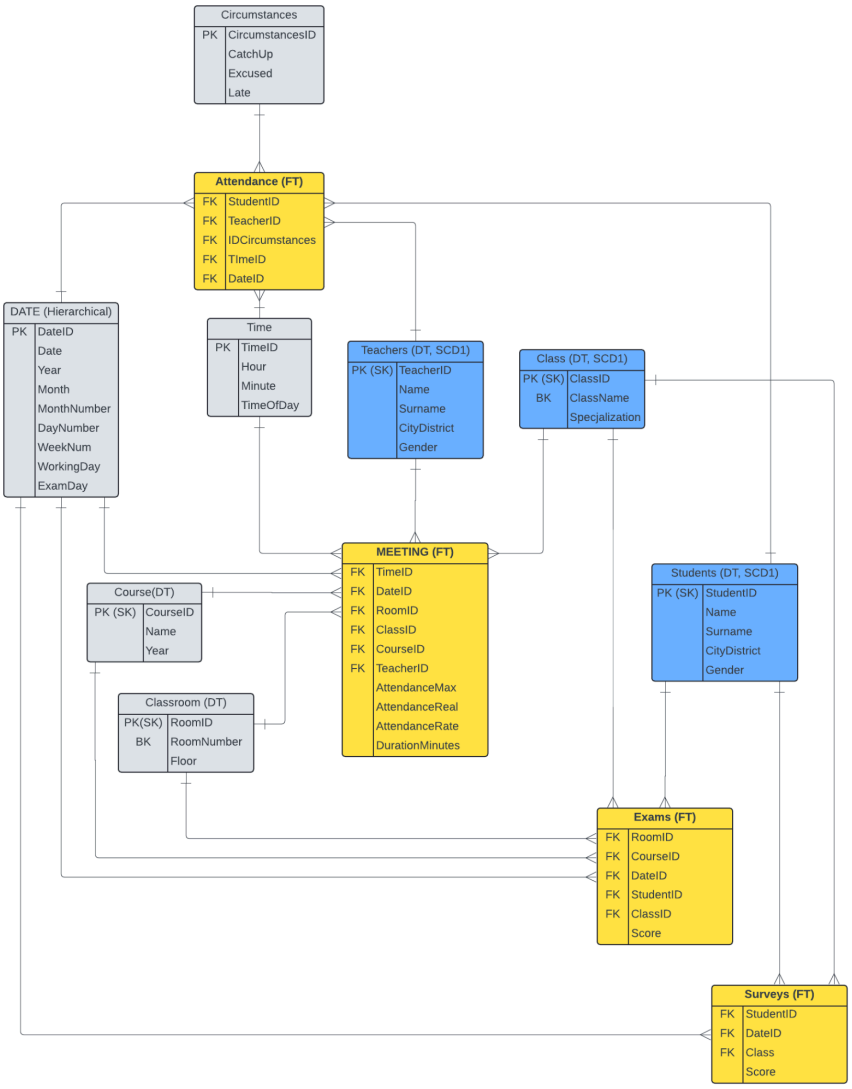
School - Data warehouse Design

Business Processes

The Data warehouse is designed for school business processes. This process is described in the document *Specification of business processes in the School archive network*.

Relational Database schema



Dimensional model

Table Name	Attribute	Attribute type	Description
Classroom (Dimension Table)	One tuple describes one physical room.		
	RoomID	Numeric	PK(SK), Unique id of room
	RoomNumber	Numeric	Business Key(BK), Unique number of the room, written on it's door
	Floor	String - 10 characters	Specifies on which floor the room is (first, second …)
Course (Dimension Table)	One tuple describes one course.		
	CourseID	Numeric	PK(SK), Unique id of a course
	Name	String - 20 characters	Name of the course (e.g. Mathematics, Polish, English)
	Year	Numeric	Year of school, when this course is taught
Teachers (Dimension Table)	One tuples describes one Teacher. (Implementation of SCD1)		
	TeacherID	Numeric	PK(SK), Unique id of a teacher
	Name	String - 20 characters	First name of the teacher
	Surname	String - 20 characters	Surname of the teacher
	CityDistrict	String - 20 characters	City from which the teacher comes (SCD1 Implementation)
	Gender	String - 1 characters	Gender of the teacher

Students (Dimension Table)	One tuple describes one Student. (Implementation of SCD1)		
	StudentID	Numeric	PK(SK), Unique id of a student
	Name	String - 20 characters	First name of the student
	Surname	String - 20 characters	Surname of the student
	CityDistrict	String - 20 characters	City from which the student comes (SCD1 Implementation)
	Gender	String - 1 characters	Gender of the student
Exams (Fact Table)	One tuple describes one exam written by one specific student on a specific day from a specific course and specific meeting.		
	FK_RoomID	Numeric	FK - id of a room on which the exam was written
	FK_CourseID	Numeric	FK - id of a course - from which exam was taken
	FK_DateID	Numeric	FK - id of date, when the exam was written
	FK_StudentID	Numeric	FK - id of a student that wrote the exam
	FK_ClassID	Numeric	Fk - id of a class, which was writing the exam
	Score	decimal - 2 digits precision	total score from the exam, falling in between 0 and 50

Meeting (Fact Table)	One tuple describes one meeting on a specific date and time with a specific teacher, teaching a specific course, to a specific class in a specific classroom.		
	FK_TimeID	Numeric	id of time when the meeting happened
	FK_DateID	Numeric	id of date when the meeting happened
	FK_RoomID	Numeric	id of a classroom where the meeting took place
	FK_ClassID	Numeric	id of the class which will be attending the meeting
	FK_CourseID	Numeric	id of a course that the meeting will be about
	FK_TeacherID	Numeric	id of the teacher that was conducting the meeting
	AttendanceMax	Numeric	Maximum number of people that can be present (1 attendance)
	AttendanceReal	Numeric	The actual number of people present at the meeting
	AttendanceRate	Numeric	Rate of attendance on the meeting (How many students present as a fraction). Based on count of AttendanceReal/AttendanceMax.
	DurationMinutes	Numeric	The number of minutes the meeting lasted

Class (Dimension Table)	One tuple describes one specific class (group of students attending meetings together) (Implementation of SCD1)		
	ClassID	Numeric	PK(SK), Unique id number of class
	ClassName	String - 3 characters	Business Key(BK), Unique name of the class (e.g. b23), consisting of a letter representing specialization, and the last two digits of the year when the class was created.
	Specialization	String - 20 characters	What class is supposed to learn more about (e.g.Mathematical, Geographical, Humanistic) (SCD1 Implementation)
Surveys (Fact Table)	One tuple describes one Survey written by one specific student on a specific meeting.		
	FK_StudentID	Numeric	FK, id of a student that filled in the survey
	FK_DateID	Numeric	FK - id of date, when the survey was written
	FK_ClassID	Numeric	FK - id of class to which the student taking survey belongs
	Score	Decimal - 2 digits precision	total score, falling in between 0 and 20

Time	One tuple describes one hour (independently on date)		
	TimeID	Numeric	PK, Id number of time
	Hour	Numeric	Hours of time: allowed numbers 0-23
	Minute	Numeric	Minute of time: allowed numbers 0-59
	TimeOfDay	String - 20 characters	Time of day. Allowed values: “Early Morning” (between 7 and 9), “Morning” (between 9 and 12), ”Noon”(between 12 and 16), “Evening” (between 16 and 21), “Night” (between 21 and 7)
Date (DIMENSION TABLE)	One tuple describes one day.		
	DateID	Numeric	PK, Id number of date
	Date	Date	Date
	Year	4 digits	Year
	Month	String - 10 characters	Month allowed values: January,February, March, April, May, June, July, August, September, October, November and December.
	MonthNumber	Numeric	Allowed values: 1-12 (1- means January, 12 - means December)
	DayNumber	Numeric	Allowed values: 1-31

	WeekNumber	Numeric	Allowed values: 1-7 (1- Monday, 7 - Sunday)
	WorkingDay	String - 13 characters	Working day. Allowed values: day off and working day
	ExamDay	String - 3 characters	Two options : Yes/No
Attendance (Fact Table)	One tuple describes one absence with specific: student, day, time, Circumstances, teacher		
	FK_StudentID	Numeric	FK, Id of student that wasn't present on specific meeting
	FK_TeacherID	Numeric	FK, Id of Teacher who checked attendance
	FK_CircumstancesID	Numeric	FK, Id of the Circumstances
	FK_TimeID	Numeric	FK, Id time of missing attendance
	FK_DateID	Numeric	FK, Id of Date when student was absent

Circumstances	The tuples correspond to: Catch up for an absence of the student, was there an excuse, was the student late.		
	CircumstancesID	Numeric	PK, id number of Circumstances
	CatchUp	Varchar(30)	Tasks to do to catch up for absence
	Excused	String - 3 characters	Two options: Yes/No
	Late	String - 3 characters	Two options: Yes/No

Fact definitions

1. Fact 1 - Meeting fact:

One instance of specific class (group of students) meeting in a specific room at specific date and time, to learn about a specific course (topic i.e. Math, Polish, English) from a specific teacher.

Granularity:

- a specified date and time of meeting
- a specified course being taught
- a specified room where it takes place
- a specified class learning
- a specified teacher

Measures and aggregation functions:

- count of meetings COUNT(MeetingID)
- number of students that attended SUM(AttendanceReal)
- number of students that were supposed to attend SUM(AttendanceMax)
- total duration of all meetings SUM(DurationMinutes)

2. Fact 2 - Survey fact:

Result of one survey (aggregated score from all questions) of one specific student from a specific class, on a specific date

Granularity:

- a specified student
- a specified class to which student belongs
- a specified date when the survey was written

Measures and aggregation functions:

- average survey score: SUM(Score)/COUNT(1)
- count of surveys answered: COUNT(1)
- count of survey-taking meetings: DISTINCT COUNT(MeetingID)

3. Fact 3 - Exam fact:

Result of one exam (aggregated score from all questions) of one specific student from a specific class, on a specific date, from a specific course, written in a specific room

Granularity:

- a specified student
- a specified class to which student belongs
- a specified date when the exam was written
- a specified course on which the exam was based
- a specified physical room where the exam was written

Measures and aggregation functions:

- average exam score: SUM(Score)/COUNT(1)
- count of exams written: COUNT(1)
- count of exam-taking meetings DISTINCT COUNT (MeetingID)

4. Fact 4 - Attendance fact:
   Record of a specific student **not** attending some subject lead by specific teacher, at specific date and time with specific Circumstances(CatchUp task, presence and lateness records)

   Granularity:
   - a specified student
   - a specified teacher
   - a specified date and time
   - a specified Circumstances(CatchUp task, presence and lateness records),

   Measures and aggregation functions
   - count of all attendances/being late: COUNT(1)

Dimension definition

Dimensions for Fact 1 Meeting fact:

DIMENSION/DIMENSION ATTRIBUTE	TABLE/COMUN	TYPE
Class	Class	Dimension
Class Name	Class.ClassName	Dimension attribute
Class specialization	Class.Specialization	Dimension attribute
Meeting Date	Date	Dimension
Meeting Year	Date.Year	Dimension attribute
Meeting Month	Date.Month	Dimension attribute
Meeting Month Number	Date.MonthNumber	Dimension attribute
Meeting Day Number	Date.DayNumber	Dimension attribute
Meeting Week Number	Date.WeekNun	Dimension attribute
Meeting Working Day	Date.WorkingDay	Dimension attribute
Meeting Exam Day	Date.ExamDay	Dimension attribute
Classroom	Classroom	Dimension
Room Number	Classroom.RoomNumber	Dimension attribute
Room Floor	Classroom.Floor	Dimension attribute

Course	Course	Dimension
Course Name	Course.Name	Dimension attribute
Course Year	Course.Year	Dimension attribute
Time	Time	Dimension
Time Hour	Time.Hour	Dimension attribute
Time Minute	Time.Minute	Dimension attribute
Time of the day	Time.TimeOfDay	Dimension attribute
Meeting Time Hierarchy	*Time.TimeOfDay **Time.Hour ***Time.Minute	Hierarchical dimension
Meeting Date Hierarchy	*Date.Year **Date.MonthNumber ***Date.DayNumber	Hierarchical dimension
Teacher	Teacher	Dimension
Teacher Name	Teacher.Name	Dimension attribute
Teacher Surname	Teacher.Surname	Dimension attribute
Teacher City/District	Teacher.CityDistrict	Dimension attribute
Teacher Gender	Teacher.Gender	Dimension attribute

Dimensions for Fact 2 Survey fact:

DIMENSION/DIMENSION ATTRIBUTE	**TABLE/COMUN**	**TYPE**
Student	Student	Dimension
Student Name	Student.Name	Dimension attribute
Student Surname	Student.Surname	Dimension attribute
Student CityDistrict	Student.CityDistrict	Dimension attribute
Student Gender	Student.Gender	Dimension attribute

Class	Class	Dimension
Class Name	Class.ClassName	Dimension attribute
Class specialization	Class.Specialization	Dimension attribute
Survey Date	Date	Dimension
Survey Year	Date.Year	Dimension attribute
Survey Month	Date.Month	Dimension attribute
Survey Month Number	Date.MonthNumber	Dimension attribute
Survey Day Number	Date.DayNumber	Dimension attribute
Survey Week Number	Date.WeekNun	Dimension attribute
Survey Working Day	Date.WorkingDay	Dimension attribute
Survey Exam Day	Date.ExamDay	Dimension attribute
Survey Date Hierarchy	*Date.Year **Date.Month ***Date.DayNumber	Hierarchical dimension

Dimensions for Fact 3 Exam fact:

DIMENSION/DIMENSION ATTRIBUTE	**TABLE/COMUN**	**TYPE**
Student	Student	Dimension
Student Name	Student.Name	Dimension attribute
Student Surname	Student.Surname	Dimension attribute
Student CityDistrict	Student.CityDistrict	Dimension attribute
Student Gender	Student.Gender	Dimension attribute

Class	Class	Dimension
Class Name	Class.ClassName	Dimension attribute
Class specialization	Class.Specialization	Dimension attribute
Course	Course	Dimension
Course Name	Course.Name	Dimension attribute
Course Year	Course.Year	Dimension attribute
Classroom	Classroom	Dimension
Room Number	Classroom.RoomNumber	Dimension attribute
Room Floor	Classroom.Floor	Dimension attribute
Exam Date	Date	Dimension
Exam Year	Date.Year	Dimension attribute
Exam Month	Date.Month	Dimension attribute
Exam Month Number	Date.MonthNumber	Dimension attribute
Exam Day Number	Date.DayNumber	Dimension attribute
Exam Week Number	Date.WeekNun	Dimension attribute
Exam Working Day	Date.WorkingDay	Dimension attribute
Exam Exam Day	Date.ExamDay	Dimension attribute
Exam Date Hierarchy	*Date.Year **Date.Month ***Date.DayNumber	Hierarchical dimension

Dimensions for Fact 4 Attendance fact:

DIMENSION/DIMENSION ATTRIBUTE	TABLE/COMUN	TYPE
Student	Student	Dimension
Student Name	Student.Name	Dimension attribute
Student Surname	Student.Surname	Dimension attribute
Student CityDistrict	Student.CityDistrict	Dimension attribute
Student Gender	Student.Gender	Dimension attribute
Teacher	Teacher	Dimension
Teacher Name	Teacher.Name	Dimension attribute
Teacher Surname	Teacher.Surname	Dimension attribute
Teacher CityDistrict	Teacher.CityDistrict	Dimension attribute
Teacher Gender	Teacher.Gender	Dimension attribute
Time	Time	Dimension
Time Hour	Time.Hour	Dimension attribute
Time Minute	Time.Minute	Dimension attribute
Time of the day	Time.TimeOfDay	Dimension attribute
Attendance Time Hierarchy	*Time.TimeOfDay **Time.Hour ***Time.Minute	Hierarchical dimension
Circumstances	Circumstances	Dimension
Circumstances CatchUp	Circumstances.CatchUp	Dimension attribute
Circumstances Excused	Circumstances.Excused	Dimension attribute
Circumstances Late	Circumstances.Late	Dimension attribute
Attendance Date	Date	Dimension
Attendance Year	Date.Year	Dimension attribute
Attendance Month	Date.Month	Dimension attribute
Attendance Month Number	Date.MonthNumber	Dimension attribute

Attendance Day Number	Date.DayNumber	Dimension attribute
Attendance Week Number	Date.WeekNun	Dimension attribute
Attendance Working Day	Date.WorkingDay	Dimension attribute
Attendance Exam Day	Date.ExamDay	Dimension attribute
Attendance Date Hierarchy	*Date.Year **Date.Month ***Date.DayNumber	Hierarchical dimension

Checking the feasibility of queries based on the multidimensional model

1. Compare average exam scores between classes.
   Measure : Exam Score
   Dimension: Class(dimension attributes: ClassName)
2. Compare average survey scores between specializations.
   Measure: Exam Score
   Dimension: Class(dimension attributes: specialization)
3. From which city districts the students with the highest scores are?
   Measure: Exam Scores
   Dimension: Student(dimension attributes: CityDistrict)
4. Does the classroom where the exam has been taken influence the average results?
   Measure: Exam Scores
   Dimension: Classroom(dimension attributes: room number)
5. Compare average scores of each gender category.
   Measure: Exam Scores
   Dimension: Students(dimension attributes: gender)
6. Give an average grade from each course.
   Measure: Exam scores
   Dimension: Courses(dimension attributes: courseID)
7. Compare average attendance rates (excused absence still counts as an absence) of the best 20 students and worst 20 students.
   Measure: Attendance rates for each student
   Dimension: Students(dimension attribute to be computed: average exam scores)
8. Compare the average lateness rate of the best 20 students and the worst 20 students.
   Measure: Lateness for each student
   Dimension: Students(dimension attribute to be computed: average exam scores)
9. Compare average lateness rate between cities.
   Measure: Lateness for each student

Dimension: Students(dimension attribute

10. Does the average attendance rate (excused absence still counts as absence) depend on a teacher?

    Measure: Meeting attendance rate

    Dimension Teachers(dimension attributes: teacher name, teacher surname)

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

Table Name	Column	Source
Classroom (Dimension Table)	One tuple describes one physical room.	
	RoomID	Classroom Id. Surrogate key - generated by database.
	RoomNumber	Classroom number. Business key. Taken from the RoomNumber column in the Classroom table from SchoolMaster database
	Floor	Specifies on which floor the room is (first, second …) Taken from the Floor column from the Classroom table in SchoolMaster source.
Course (Dimension Table)	One tuple describes one course.	
	CourseID	Course Id. Surrogate key - generated by database.
	Name	Name of the course (e.g. Mathematics, Polish, English). Taken from column Name from Course table in SchoolMaster source.
	Year	Year of school, on which this course is taught. Taken from column Year from Course table in SchoolMaster source.
Teachers	One tuples describes one Teacher.	

(Dimension Table)	(Implementation of SCD1)	
	TeacherID	Id of a teacher. Surrogate key - generated by database.
	Name	First name of the teacher. Taken from column Name from Teachers table in SchoolMaster source.
	Surname	Surname of the teacher. Taken from column SurName from Teachers table in SchoolMaster source.
	CityDistrict	City from which the teacher comes. Taken from column CityDistric from Teachers table in SchoolMaster source. (SCD1 Implementation)
	Gender	Gender of the teacher. Taken from column Gender from Teachers table in SchoolMaster source.

Students (Dimension Table)	One tuple describes one Student. (Implementation of SCD1)	
	StudentID	Student ID. Surrogate key - generated by database.
	Name	First name of the student. Taken from column Name from Students table in SchoolMaster source.
	Surname	Surname of the student. Taken from column Surname from Students table in SchoolMaster source.
	CityDistrict	City from which the student comes. Taken from column CityDistrict from

		Students table in SchoolMaster source. (SCD1 Implementation)
	Gender	Gender of the student. Taken from column Gender from Students table in SchoolMaster source.
Exams (Fact Table)	One tuple describes one exam written by one specific student on a specific day from a specific course and specific meeting.	
	FK_RoomID	FK - id of a room on which the exam was written Taken from column RoomID from Exams table in Schoolmaster source.
	FK_CourseID	FK - id of a course - from which exam was taken. Taken from column CourseID from Exams table in Schoolmaster source.
	FK_DateID	FK - id of date, when the exam was written. Taken from column DateID from Exams table in Schoolmaster source.
	FK_StudentID	Id of a student that wrote the exam. Foreign key Taken from column StudentID from Exams table in Schoolmaster source.
	FK_ClassID	Fk - id of a class, which was writing the exam. Taken from column ClassID from Exams table in Schoolmaster source.
	Score	Total score from the exam, falling in between 0 and 50. Taken from column Score from Exams table in Schoolmaster source.
Meeting (Fact Table)	One tuple describes one meeting on a specific date and time with a specific teacher, teaching a specific course, to a specific class in a specific classroom.	
	FK_TimeID	Id of time when the meeting happened. Based on column TimeID from the Meeting table in the Schoolmaster source.
	FK_DateID	Id of date when the meeting happened.

		Based on column DateID from the Meeting table in the Schoolmaster source.
	FK_RoomID	Id of a classroom where the meeting took place. Taken from column RoomID from Meeting table in Schoolmaster source.
	FK_ClassID	Unique ID of the class. Taken from column ClassID from the Meeting table in the Schoolmaster source.
	FK_CourseID	Id of a class that will attend the meeting. Taken from column CourseID from Meeting table in Schoolmaster source.
	Fk_TeacherID	id of the teacher that was conducting the meeting, Taken from column TeacherID from Meeting table in Schoolmaster source.
	AttendanceMax	Maximum number of people that can be present (100% attendance) Based on from count of students in the class that has the meeting from class, student and meeting tables from schoolmaster source
	AttendanceReal	The actual number of people present at the meeting. Based on AttendanceMax and count of relevant rows in Attendance.csv.
	AttendanceRate	Rate of attendance on the meeting (How many students were present). Based on count of AttendanceReal/AttendanceMax.
	DurationMinutes	The number of minutes the meeting lasted. Taken from Column Duration Minutes from the Meeting table in the Schoolmaster source.

Class (Dimension Table)	One tuple describes one specific class (group of students attending meetings together) (Implementation of SCD1)	
	ClassID	Class ID. Surrogate key - generated by database
	ClassName	Unique name of the class (e.g. b23), consisting of a letter representing specialization, and the last two digits of the year when the class was created. Business Key(BK). Taken from column ClassName from Class table in Schoolmaster source.
	Specialization	What class is supposed to learn more about (e.g.Mathematical, Geographical, Humanistic), taken from column Specialization from Class table in Schoolmaster source. (SCD1 Implementation)
Surveys (Fact Table)	One tuple describes one Survey written by one specific student on a specific meeting.	
	FK_StudentID	Id of a student taking survey Foreign key from Student Based on table Surveys column FK_StudentID from SchoolMaster database
	FK_DateID	Id of date when the survey happened. Based on column DateID from the Meeting table in the Schoolmaster source.
	FK_ClassID	id of class to which the student taking survey belongs Forgey key from Class, Based on table Surveys column ClassID from SchoolMaster database.
	Score	total score, falling in between 0 and 20, taken from column Score from table Surveys in the SchoolMaster database.

Attendance (Fact Table)	One tuple describes one absence with specific: student, day, time, Circumstances, teacher	
	FK_StudentID	Id of a student that didn't attend the meeting. Foreign key from Student dimension table. Based on StudentID column from Attendance.csv
	FK_DateID	Id of Date when student was absent. Foreign key from Date dimension table. Based on Date column from Attendance.csv
	FK_TeacherID	Id of Teacher who checked attendance when a student was absent. Foreign key from Student dimension table. Based on TeacherID column from Attendance.csv
	FK_CircumstancesID	Id of the Circumstances Foreign key from Circumstances dimension table Based on CatchUp, Late and Excused columns from Attendance.csv
	FK_Time	Time of missing attendance Foreign key from Date dimension table Based on Time column from Attendance.csv

Circumstances	One tuple describes one absence with specific: student, day, time, Circumstances, teacher	
	CircumstancesID	Id of circumstances of lack of attendance Generated by the database
	CatchUp	Decription of needs to be done to catch up to the student's who attended Taken from CatchUp column of Attendance.csv
	Excused	Whether the absence has been excused. Taken from Excused column of the Attendance.csv
	Late	Whether the student showed up late. Taken from Late column of the Attendance.csv

Date	**One tuple describes one day.** Every record in this table is generated based on any calendar, before ETL process
Time	**One tuple describes one hour (independent of date)** Every record in this table is generated based on any calendar, before ETL process