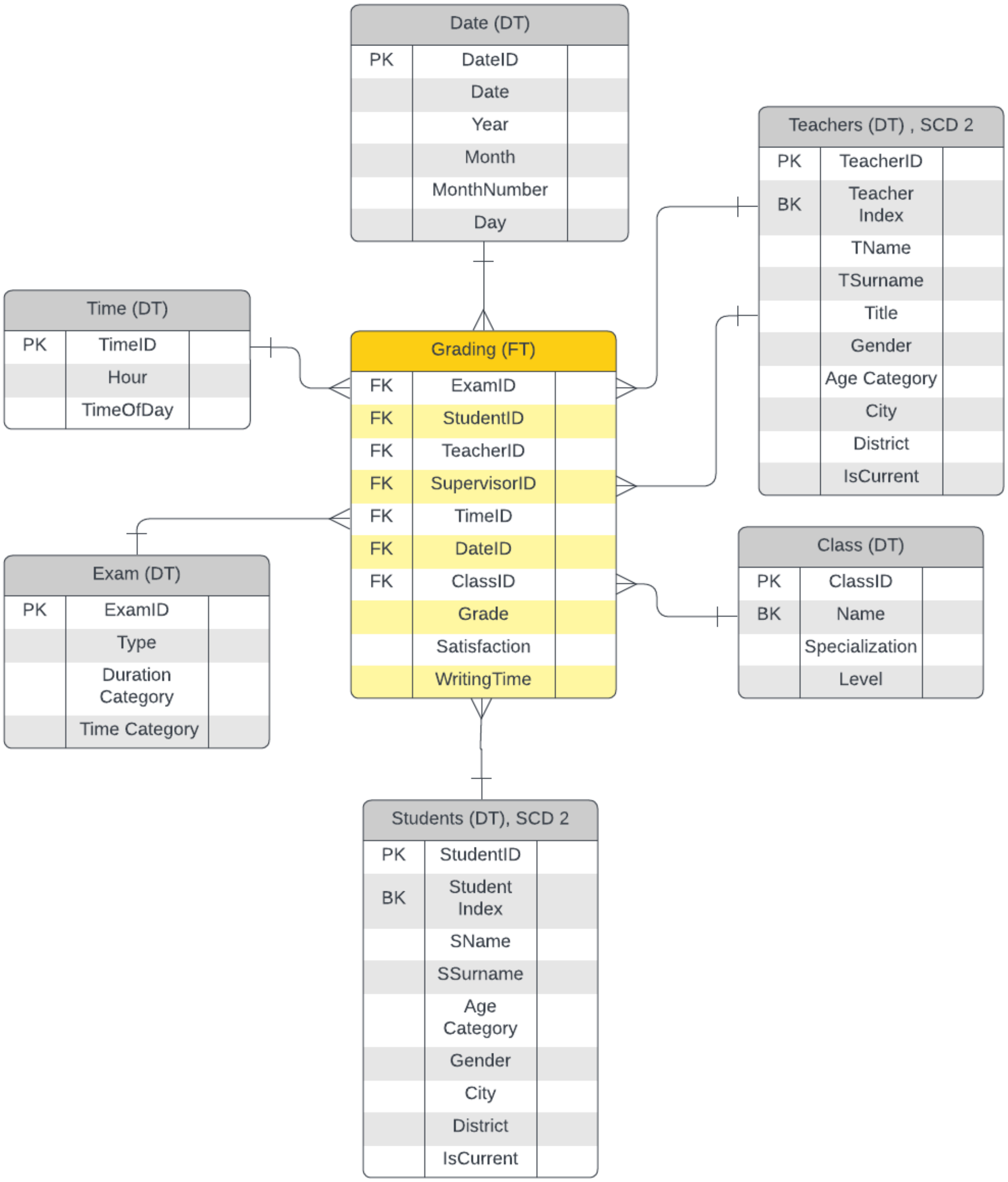
xAcademy of Success – Data Warehouse design

Business process

The data warehouse is designed for grading process in the school. The process is described in the document *RequirementsProcessSpecification*.

Data warehouse schema visualization



Date (DT)	
PK	DateID
	Date
	Year
	Month
	MonthNumber
	Day

Teachers (DT) , SCD 2	
PK	TeacherID
BK	Teacher Index
	TName
	TSurname
	Title
	Gender
	Age Category
	City
	District
	IsCurrent

Time (DT)	
PK	TimeID
	Hour
	TimeOfDay

Grading (FT)	
FK	ExamID
FK	StudentID
FK	TeacherID
FK	SupervisorID
FK	TimeID
FK	DateID
FK	ClassID
	Grade
	Satisfaction
	WritingTime

Exam (DT)	
PK	ExamID
	Type
	Duration Category
	Time Category

Class (DT)	
PK	ClassID
BK	Name
	Specialization
	Level

Students (DT), SCD 2	
PK	StudentID
BK	Student Index
	SName
	SSurname
	Age Category
	Gender
	City
	District
	IsCurrent

Description of the attributes

GRADING **One tuple describes one grade given.**

(FACT TABLE)

Attribute	Attribute Type	Description
ExamID	Numeric	FK Exam
StudentID	Numeric	FK Student Student
TeacherID	Numeric	FK Teacher Teacher grading exam
SupervisorID	Numeric	FK Supervisor Teacher supervising exam
TimeID	Numeric	FK Time Time of putting the grade
DateID	Numeric	FK Date Date of putting the grade
ClassID	Numeric	FK Class Class of the student
Grade	Numeric	Value of grade
Satisfaction	Numeric	Student’s satisfaction from the grade
WritingTime	Numeric	Student’s writing time

Dimensions

Dimension

Teachers (DT) One **tuple describes one teacher working in the school.**

TeacherID	Numeric	PK
Teacher Index	6 digits	BK
TName	Varchar(30)	Teacher’s first name
TSurname	Varchar(30)	Teacher’s surname
Title	Varchar(20)	Teacher’s academic title
Gender	Varchar(10)	Gender that was assigned to the teacher after birth
Age Category	Varchar(5)	Teacher’s age. Allowed values: 32-35, 36-39, 40-45, 46-49, 50-55, 56-59, 60-65, 65+
City	Varchar(30)	City, in which the teacher is living
District	Varchar(30)	District of the city, in which the teacher is living
IsCurrent	Boolean	1 if information is current, otherwise 0. (SCD2 implementation)

Students (DT) **One tuple describes one student.**

StudentID	Numeric	PK (Primary Key)
Student Index	6 digits	BK (Business Key)
SName	Varchar(30)	Student’s first name
SSurname	Varchar(30)	Student’s surname
Age Category	Varchar(5)	Student’s age. Allowed values: 14-15, 16-17, 18-19, 20+ (Implementation of SCD 2)
Gender	Varchar(10)	Gender that was assigned to the student after birth
City	Varchar(30)	City, in which the student is living
District	Varchar(30)	District of the city, in which the student is living
isCurrent	Boolean	1 if information is current, otherwise 0. (SCD2 implementation)

Exam (DT) **One tuple describes one exam.**

ExamID	Numeric	PK (Primary Key)

Type	Varchar(15)	Type of Exam. Allowed values: Humanistic, Scientific, Linguistic
Duration Category	Varchar(10)	Duration of Exam. Allowed values: short, medium, long
Exam Time Category	Varchar(10)	Time of day of exam. Allowed values: Morning, noon, evening, night

Class (DT) One **tuple describes one class in the school.**

ClassID	Numeric	PK (Primary Key)
Name	Varchar(5)	BK (Business Key)
Specialization	Varchar(25)	Specialization of the class.
Level	Varchar(25)	Level of the class

Time (DT) One **tuple describes one hour (independently of date).**

TimeID	Numeric	PK (Primary Key)
Hour	Numeric	Hour (allowed values 0-23)
TimeOfDay	Varchar(20)	Time of day. Allowed values: morning, noon, afternoon, evening, night

Date (DT) **One tuple describes one day.**

DateID	Numeric	PK (Primary Key)
Date	Date	Date
Year	4 digits	Year
Month	Varchar(20)	Month. Allowed values: January, February, March, April, May, June, July, August, September, October, November and December
MonthNumber	Numeric	Month's numeric value
Day	Numeric	Day number of current month

Dimensional model

Fact definitions:
Fact of grading the exam: The grading procedure involves a teacher assigning a grade to a particular student enrolled in a specific class. This grade is derived from an examination supervised by an appointed instructor. The grade is posted on a specified date and time.

Fact table: Grading

Granularity:

- A specified grade
- A specified teacher in a specified age category, specified gender, with a specified academic title, that lives in a specified city and a specified district of this city
- A specified student in a specified age category, specified gender, that lives in a specified city and a specified district of this city
- A specified exam with a specified type, of a specified duration, in a specified time category
- A specified hour of entering the grade
- A specified date of entering the grade
- A specified class with a specified name, specified specialization, of a specified educational level

Measures and aggregation functions:

Number of grades – COUNT(1)

Number of positive grades – COUNT(Grade >= 2)

Average writing time – SUM(WritingTime)/ Number of grades

Average satisfaction from the grade – SUM(Satisfaction)/ Number of grades

Average grade – SUM(Grade)/ Number of grades

Dimension definitions

Dimensions for Grading Fact Table:

Dimension	Dimension attribute	Table/column	Type
Exam	Type	Exam.Type	Dimension attribute
	DurationCategory	Exam.DurationCategory	Dimension attribute
	TimeCategory	Exam.TimeCategory	Dimension attribute
Teachers	Index	Teachers.Index	Dimension attribute
	TName	Teachers.TName	Dimension attribute
	TSurname	Teachers.TSurname	Dimension attribute
	Title	Teachers.Title	Dimension attribute
	Gender	Teachers.Gender	Dimension attribute
	AgeCategory	Teachers.AgeCategory	Dimension attribute
	City	Teachers.City	Dimension attribute
	District	Teachers.District	Dimension attribute

Teachers Hierarchy	• Teachers.Title •• Teachers.AgeCategory ••• Teachers.Index		Hierarchical dimension
Grading Date Hierarchy	• Date.Year •• Date.Month ••• Date.Day •••• Date.Date		Hierarchical dimension
Grading Time Hierarchy	• Time.TimeOfDay •• Time.Hour		Hierarchical dimension
Students	Index	Students.Index	Dimension attribute
	SName	Students.SName	Dimension attribute
	SSurname	Students.SSurname	Dimension attribute
	AgeCategory	Students.AgeCategory	Dimension attribute
	Gender	Students.Gender	Dimension attribute
	City	Students.City	Dimension attribute
	District	Students.District	Dimension attribute
Class	CName	Class.CName	Dimension attribute
	Specialization	Class.Specialization	Dimension attribute
	Level	Class.Level	Dimension attribute
Date	Date	Date.Date	Dimension attribute
	Year	Date.Year	Dimension attribute
	Month	Date.Month	Dimension attribute
	Day	Date.Day	Dimension attribute
Time	Hour	Time.Hour	Dimension attribute
	TimeOfDay	Time.TimeOfDay	Dimension attribute

Checking the feasibility of queries based on the smultidimensional model

1.**Compare the grades from exams conducted in the morning to the exams conducted in the noon**

Measure : Average grade

Dimension: Exam (dimension attributes: Exam time category)

2. **Which class achieves the lowest scoring?**

Measure: Average grade

Dimension: Class (dimension attributes: Class name)

3. **Compare the scoring of the students that are not living the same city as the school is located to those who are living the same city**

Measure: Average grade

Dimension: Students (dimension attributes: City)

4. **Teachers from which districts grade the highest?**

Measure: Average grade

Dimension: Teachers (dimension attributes: District)

5. **Which of the teachers grading the exams from the previous month managed to give students the highest grades?**

Measure: Average grade

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

Dimension: Grade Date (dimension attributes: month)

6. **What type of exam achieves the highest average satisfaction?**

Measure: Average satisfaction

Dimension: Exam (dimension attribute: type)

7. **Which class specialization has the highest average satisfaction?**

Measure: Average satisfaction

Dimension: Class (dimension attribute: specialization)

8. **Does the average satisfaction from the exam differ when the supervisor was a Female or Male?**

Measure: Average satisfaction

Dimension: Teachers (dimension attributes: gender)

9. **Does the average satisfaction differentiate between genders?**

Measure: Average satisfaction

Dimension: Students (dimension attributes: gender)

10. **Compare the average satisfaction from diagnostic exams from the previous month compared to the current month, grouped by title of the teacher**

Measure: Average satisfaction

Dimension: Teacher (dimension attributes: title)

Dimension: Grade Date (dimension attributes: month)

11. Compare the students' average satisfaction from the grade based on their age

Measure: Average satisfaction

Dimension: Students (dimension attributes: age category)

12. **Compare the average writing time based on the type of the exam**

Measure: Average writing time

Dimension: Exam (dimension attributes: type)

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

Table Name	Column	Source
Grading	**One tuple describes one grade given**	
	ExamID	ExamID. Foreign key from dimension table. Based on ExamID from *SchoolSystem* Source.
	StudentID	StudentID. Foreign Key from dimension table. Based on Students' index, age, city, district, name, surname. Based on School_Students.csv and *SchoolSystem* source.
	TeacherID	TeacherID. Foreign Key from dimension table. Based on Teachers' index,age,city,district,title,name,surname. Based on School_Teachers.csv and SchoolSystem source.
	TimeID	Time id. Foreign Key from dimension table. Based on Date from Grade Table in SchoolSystem source.

	DateID	Date id. Foreign Key from dimension table. Based on Date from Grade table in SchoolSystem source.
	ClassID	Class id. Foreign Key from dimension table. Based on name, specialization and level of the class. Based on SchoolSystem source.
	SupervisorID	Supervisor id. Foreign Key from dimension table. Based on Teachers' index,age,city,district,title,name,surname. Based on School_Teachers.csv and SchoolSystem source.
	Grade	Value of the grade posted in the system. Based on Value from Grade table in SchoolSystem source.
	Satisfaction	Student satisfaction from the grade is taken from Satisfaction column from Grade table in SchoolSystem source.
	WritingTime	The time student was writing the exam is taken from WritingTime column from Grade table in SchoolSystem source.
Exam	**One tuple describes one exam.**	
	ExamID	ExamID. Surrogate key – generated by database
	Type	Type of exam. Based on Type from Diagnosting Exam table in SchoolSystem source.
	Duration Category	Duration of the exam. Based on Duration from Diagnosting Exam table in SchoolSystem source.

	Time Category	Time when the exam took place. Based on Date from Diagnosting Exam table in SchoolSystem source.
Teacher	**One tuple describes one teacher working in the school.**	
	TeacherID	TeacherID. Surrogate key – generated by database
	Index	Teacher Index. Business key taken from TeacherID column from Teachers from SchoolSystem
	TName	Teacher's first name. Taken from column TName from table Teachers from SchoolSystem.
	TSurname	Teacher's surname. Taken from column TSurname from table Teachers from SchoolSystem.
	Title	Teacher's title. Taken from column D from School_Teachers csv file.
	Gender	Teacher's sex. Taken from column H from School_Teachers csv file.
	Age Category	Teacher's age in form of numerical ranges. Information about age can be extracted from column I from School_Teachers csv file.
	City	City where the teacher lives. Based on column F from School_Teachers.csv file.
	District	District of the city where the teacher lives. Based on column G from School_Teachers.csv file.
	IsCurrent	"1" if information is current, otherwise "0" (SCD2 implementation).

Student	**One tuple describes one student.**	
	StudentID	StudentID. Surrogate key – generated by database
	Index	Student Index. Business key taken from StudentID column from Teachers from SchoolSystem
	SName	Student’s first name. Taken from column SName from table Students from SchoolSystem.
	SSurname	Student’s surname. Taken from column SSurname from table Students from SchoolSystem.
	Age Category	Student’s age in form of numerical ranges. Information about age can be extracted from column E from School_Students csv file.
	Gender	Student’s sex. Taken from column D from School_Students csv file.
	City	City where the student lives. Based on column H from School_Teachers.csv file.
	District	District of the city where the student lives. Based on column I from School_Teachers.csv file.
	IsCurrent	“1” if information is current, otherwise “0” (SCD2 implementation).
Class	**One tuple describes one class in the school.**	

	ClassID	ClassID. Surrogate key – generated by database
	Name	Name of the class. Business Key taken from column Name from Class table in SchoolSystem source.
	Specialization	Specialization of the class. Based on Specialization from Class table in SchoolSystem source.
	Level	Level of the class. Based on Level column from Class table in SchoolSystem 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.	