UniLearn – Data warehouse design

Business process

Relational Database schema

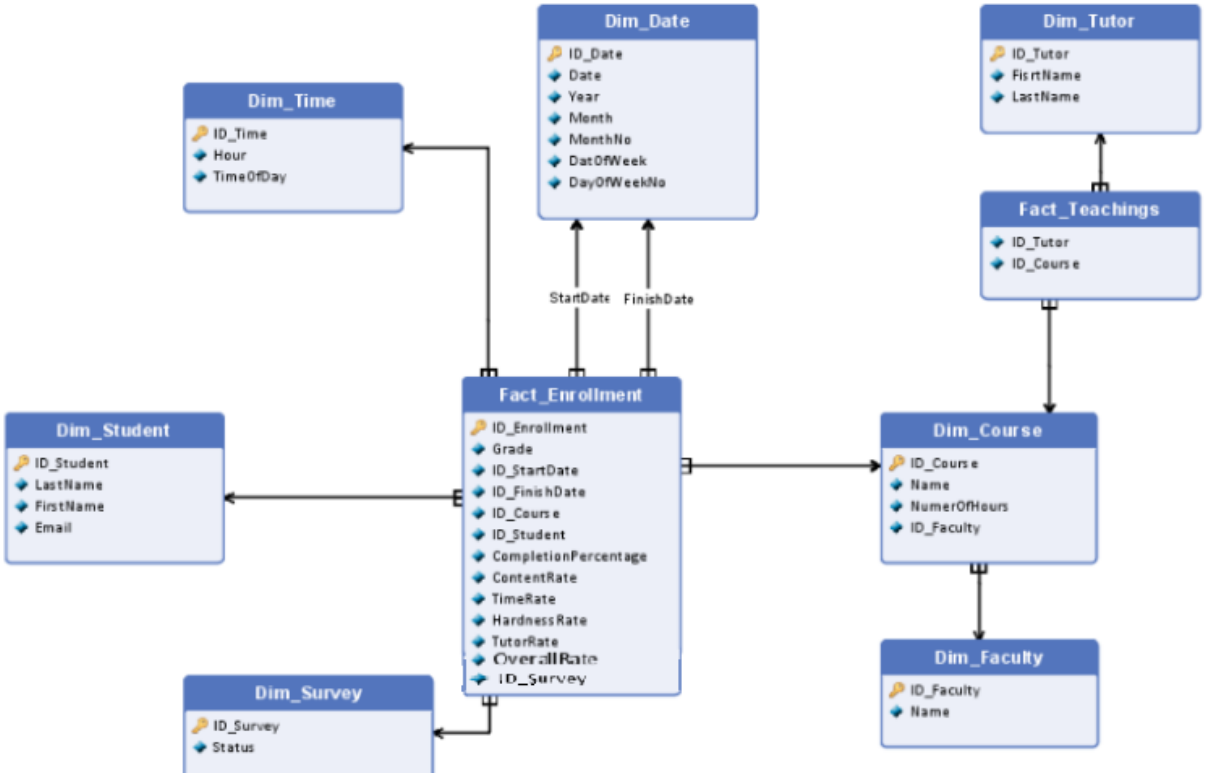

Table name	Attribute	Attribute type	Description
Enrollment (Fact table)	One tuple describes one fact of enrollment to the course.		
	Grade	Decimal	Grade obtained from the course. Domain is from two to five
	ID_StartDate	Numeric	FK Dim_Date Start Date
	ID_FinishDate	Numeric	FK Dim_Date Finish Date
	ID_Course	Numeric	FK Dim_Course Course assigned

	ID_Student	Numeric	FK Dim_Students Student assigned
	ID_Survey	Numeriv	FK Dim_Survey Survey fulfillment flag
	CompletitionPercentage	Numeric	The percentage of completion of course (e.g. How many % of slides have been seen by a particular user). Value from zero to one hundred.
	ContentRate	Numeric	The rate of the course. Value between 1 and 10
	TimeRate	Numeric	The rate of the number of hours of course. Value between 1 and 10
	HardnessRate	Numeric	The rate of the hardness of the course. Value between 1 and 10
	ToutorRate	Numeric	The rate of the tutors assigned to the course. Value between 1 and 10
	OverallRate	Numeric	The overall rate of the survey. Calculated as a mean of 4 rates((Content+tutor +hardness+time)/4)
Teachings (Fact table)	One tuple describes one fact of being a tutor of the course.		
	ID_Tutor	Numeric	FK Dim_Tutors. Tutor assigned to course
	ID_Course	Numeric	FK Dim_Course. Which course tutor is assigned
Survey (Dimension Table)	One tuple describes one survey		
	ID_Survey	Numeric	PK

	Status	Varchar(14)	Status of the survey if it was “fulfilled” or “not fulfilled”
Student (Dimension Table)	One tuple describes one student		
	ID_Student	Numeric	PK
	LastName	Varchar(55)	Last name of the student
	FirstName	Varchar(55)	First name of the student
	Email	Varchar(60)	Email of the student
Time (DImension Table)	One tuple describes one hour (independently on date)		
	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).
Date (Dimension Table)	One tuple describes 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, …
Tutor (Dimension Table)	ID_Tutor	Numeric	PK
	FirstName	Varchar(50)	First name
	LastName	Varchar(50)	Last name
Course (Dimension Table)	One tuple describes one course		
	ID_Course	Numeric	PK
	Name	Varchar(55)	Name of the course
	NumOfHours	Varchar(20)	Number of hours of the course. Allowed values: between 15 and 30, between 31 and 45, between 46 and 60, between 61 and 75, between 76 and 90, between 91 and 105, between 106 and 120).

	ID_Faculty	Numeric	FK Dim_Faculty On which faculty course is conducted
Faculty (Dimension Table)	One tuple describes one faculty		
	ID_Faculty	Numeric	PK
	Name	Varchar(55)	Name of the faculty

Dimensional model

Fact definitions

Fact 1 Enrollment Fact:
Enrollment of students to the course, enrolled on a specified day, at a specified time. Enrolled to a specific course on specific faculty, the course is graded, if grade is positive finished at specified day, at a specified time, evaluated completion percentage of the course. May be rated by survey for specific statistics : content rate, time rate ,hardness rate, tutor rate.

Fact table: Fact_Enrollment

Granularity:
- a specified student,
- a specified course with specified faculty,
- a specified enrollment date,
- a specified hour of enrollment,
- a specified date of finishing course,
- a specified hour of finishing course,
- a specified survey status about the course,

Measures and aggregation functions:
Number of enrollments facts – COUNT (1)
Grade average- AVG(Grade)
Completion percentage average - AVG(CompletionPercentage)
Content rate average- AVG(ContentRate)
Time rate- AVG(TimeRate)
Hardness rate- AVG(HardnessRate)
Tutor rate- AVG(TutorRate)
Overall course rating - AVG(Content rate, TimeRate,HardnessRate,TutorRate)

Number of transactions - DISTINCT COUNT (TransactionNo)

Fact 2 Teaching Fact: Being a tutor of a given course.

Fact resulting from “many to many” relationship.
Fact table: Fact_Teachings

Granularity:

- a specified tutor,
- a specified course on the platform.

Measures and aggregation functions:
Number of being the tutor facts– COUNT(1)
Fact 2: Teachings

Dimension definitions

Dimension for Fact 1 Enrollment fact:

Dimension / Dimension attribute	Table/Column	Type
Survey	**Survey**	**Dimension**
Status	**Survey.Status**	**Dimension attribute**
Student	**Student**	**Dimension**
FirstName	**Student.FirstName**	**Dimension attribute**
LastName	**Student.LastName**	**Dimension attribute**
Email	**Student.Email**	**DImension attribute**
Time	**Time**	**Dimension**
Hour	**Time.Hour**	**Dimension attribute**
TimeOfDay	**Time.TimeOfDay**	**Dimension attribute**
Start Date Hierarchy	**• Date.Year •• Date.Month ••• Date.Date**	**Hierarchical dimension**
Finish Date Hierarchy	**• Date.Year •• Date.Month ••• Date.Date**	**Hierarchical dimension**
Start Date	**Date**	**Dimension**
Start Year	**Date.Year**	**Dimension attribute**
Start Month	**Date.Month**	**Dimension attribute**
Start Date	**Date.Date**	**Dimension attribute**
Finish Date	**Date**	**Dimension**

Finish Year	**Date.Year**	**Dimension attribute**
Finish Month	**Date.Month**	**Dimension attribute**
Finish Date	**Date.Date**	**Dimension attribute**
Course	**Course**	**Dimension**
Course Name	**Course.Name**	**Dimension attribute**
NumOfHours	**Course.NumOfHours**	**Dimension attribute**
Faculty	**Faculty**	**Dimension**
Name	**Faculty.Name**	**Dimension attribute**
Tutor	**Tutor**	**Dimension**
First Name	**Tutor.FirstName**	**Dimension attribute**
Last Name	**Tutor.LastName**	**Dimension attribute**

Dimensions for Fact 2, Teachings fact:

Dimension/ Dimension attribute	**Table/ Column**	**Type**
Course	**Course**	**Dimension**
Course Name	**Course.Name**	**Dimension attribute**
Tutor	**Tutor**	**Dimension**
First Name	**Tutor.FirstName**	**Dimension attribute**
Last Name	**Tutor.LastName**	**Dimension attribute**

Checking the feasibility of queries based on the multidimensional model

1. Compare the average grade of courses with the biggest/lowest number of hours and the smallest.
   Measure: Grade obtained from the course
   Dimension: Course(dimension attributes: Number of hours)
2. Find the average completion percentage for the courses with the biggest/lowest number of hours.
   Measure: Completion percentage
   Dimension: Course(dimension attributes: Number of hours)
3. Find the number of hours of courses with the best average rating from the survey.
   Measure: Average rating from survey

Dimension: Survey (dimension attributes: status),
Dimension: Course(dimension attributes: Number of hours)

4. Find how many hours from courses has the best time rate according to survey?
   Measure: Time rate (number of hours rating)
   Dimension: Survey (dimension attributes: status)
   Dimension: Course (dimension attribute: NumOfHours**)**
5. What was the number of hours for the courses with the highest hardness rate?
   Measure: Hardness rate
   Dimension: Course(dimension attribute: Number of hours)
6. Compare average grades from previous years.
   Measure: Grade obtained from the course
   Dimension: Start Date*(dimension attributes: Year, month)*
7. Which faculty has the lowest/highest average grades?
   Measure: Grade obtained from the course
   Dimension: Faculty (dimension attribute: name)
8. How the number of hours during courses are shaped in relation to average grade?
   Measure:Grade obtained from the course ;
   Dimension: Course (dimension attribute: NumOfHours**)**
9. Which courses have the highest content rate level?
   Measure: Content rate (content rating obtained from survey)
   Dimension:Survey (dimension attributes: status)
   Dimension: Course(dimension attribute: Name*)*
10. Find the courses names with the highest tutor level.
    Measure: Tutor rate (tutor teaching the course)
    Dimension:Survey (dimension attributes: status)
    Dimension: Course(dimension attribute: Name*)*

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

TABLE NAME	**COLUMN**	**SOURCE**
Enrollments	**One tuple describes one fact of student enrolling to the course**	
	ID_Enrollment	Enrollment Id. Surrogate key – generated by database
	Grade	Grade obtained from the course. Calculated as

		average grade obtained from tasks and assignments Stored in UniLearn database table Enrollment
	ID_StartDate	Start date Id. Date of enrolling to the course Foreign key from dimension table. Based on StartDate stored in Enrollment table in UniLearn database source
	ID_FinishDate	Finish date Id. Foreign key from dimension table. Based on FinishDate stored in Enrollment table in UniLearn database source
	ID_Course	Sold book Id. Foreign key from dimension table. Based on foreign key FK_Course in table Enrollment in UniLearn database source.
	ID_Survey	Survey Id. Foreign key from dimension table. Based on the semestral survey source.
	ID_Student	Student Id. Foreign key from dimension table. Based on the Enrollment table in UniLearn database source.
	CompletionPercentage	Completion percentage achieved at given course. Stored in UniLearn database table Enrollment
	ContentRate	Content rate obtained from survey. Based on the semestral survey source.Column F. Numerical Values from 0-10.
	TimeRate	Time rate obtained from survey. Based on the semestral survey source.Column H. Numerical Values from 0-10.
	HardnessRate	Hardness rate obtained from survey. Based on the semestral survey source. Column E.Numerical Values

		from 0-10.
	TutorRate	Tutor rate obtained from survey. Based on the semestral survey source. Column G.Numerical Values from 0-10.
CALCULATED MEASURE	OverallRate	Overall rate obtained from survey. Based on the four ratings. Calculated as a mean of 4 ratings ((ContentRate+tutorRate+hardnessRate+timeRate)/4)
Course	**One tuple describes one course.**	
	ID_Course	Course Id. Surrogate key - generated by the database.
	Name	Name of the course. Value from Course table UniLearn database.
	NumberOfHours	Number of hours of the course. Allowed values: between 15 and 30, between 31 and 45, between 46 and 60, between 61 and 75, between 76 and 90, between 91 and 105, between 106 and 120). Values based on Number of hours from Course table in UniLearn database source.
	ID_Faculty	Faculty Id. Foreign key from dimension table. Value based on data from UniLearn database table Courses
Faculty	**One tuple describes one faculty.**	
	ID_Faculty	Faculty Id. Surrogate key - generated by the database.
	Name	Name of the faculty. From UniLearn database table Faculty
Teachings	**One tuple describes one fact of being a tutor of the course**	

	ID_Tutor	Foreign key from dimension table sourced in reference FK_Course in Courses table in UniLearn database source
	ID_Course	Foreign key from dimension table sourced in reference FK_Tutor in Tutors table in UniLearn database source
Tutor	**One tuple describes one tutor**	
	ID_Tutor	Tutor Id. Surrogate key - generated by the database.
	FirstName	Tutor's Name taken from name from Tutor table in UniLearn database source.
	LastName	Tutor's surname taken from name from Tutor table in UniLearn database source.
Survey	**One tuple describes the status of survey fulfillment.**	
	ID_Survey	Survey id.Surrogate key - generated by the database.
	Status	Flag for semestral survey fulfillment
Student	**One tuple describes one student (Implementation of SCD 2)**	
	ID_Student	Student Id. Surrogate key - generated by the database.
	LastName	Last name of the student Students table in UniLearn database source.
	FirstName	First name of the student taken from Students table in UniLearn database source.
	Email	Email of the student can be updated. (Implementation of SCD 2) Stored in UniLearn database table Students
Date	**One tuple describes one day.** All the data in this table is generated tuple by tuple based on any calendar, before the 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