

JULY 1, 2016



IST 722- DATA WAREHOUSING GROUP 5-PROJECT REPORT

RAJITH JAYADEVAN

SAURABH JAPE

ANANYA CHAKRABORTY

SUKHMEET SINGH

ACKNOWLEDGEMENT

An endeavor over a period of time can be successful only with the advice and guidance of all the well-wishers, some directly through their technical assistance and some through their encouragement and help. We take this opportunity to owe our thanks and deep sense of gratitude to all who encouraged us and stood by our side in completing this project and making it a successful one. We are highly indebted to Prof. Michael Fudge, for his active help and guidance all throughout the course. We would also like to thank all our peers for their comments and feedbacks, which helped us improve on our flaws thus enhancing our work.

OVERVIEW

Problem Statement

There has been tremendous growth in the education structure at Syracuse University because of the fact that they provide classes on campus, distance learning programs and online classes. Hence a need arises for the assessment for course catalog. This assessment requires the university to conduct a trend analysis for the enrolment activity for the courses. This trend analysis involves the number of enrolments for courses and the number of students that drop the class.

Background

Students can drop classes due to many factors that can range from being obvious reasons to unknown/indeterminable reasons. The factors can be like manual errors, opting for substitute classes, time clashes, course-work, reputation of the faculty. Also students drop classes if get a chance to join the same course under a different faculty who is a generous grader.

Project Objective

The objective of our project is to have a rough idea of the trends for enrolment activity and provide a count of the number of enrollments and drops for a course to the University Staff. The end-product of this project is to provide valuable information to the University in order to make a business decision to improve course services, course technology, course modification, or even creating or developing a new course according to the industry requirement, dropping courses from the catalog or continuing with the same course offerings based on the assessment of the course catalog.

Project Scope & Mission

The scope of the project is to provide the University with facts and figures which will enable the University and staff to analyze data to provide Key Performance Indicators, KRI- Key Risk Indicators and provide trends and opportunities for Business intelligence, Business Development and Business Analyst Professionals. Since it is a trend analysis, there must be certain numbers or indicating factors like ratings or scorecards which will enable the professionals and University officials to make decisions regarding the courses.

Team Member & Roles



Rajith Jayadevan- ETL Architect and Data Architect

Saurabh Jape– Business Analyst and Data Architect

Sukhmeet Singh–BI Designer

Ananya Chakraborty– Decision Support Analyst

Stakeholders

Business Owner- University

Business Managers – University Staff

Customer – Faculty and Students

University Provost

Business Analyst

Business Development Professionals

Program Manager

Project Manager

Project Group 5

ANALYSIS

Business Process

Name: Student Registration Activity

The business process that we have selected for our project is to determine the class registration activity in order to get information and trends about the number of enrollments and drops for a particular course and the class section during an academic semester.

Databases Used

ist722_group5_dw

ist722_group5_stage

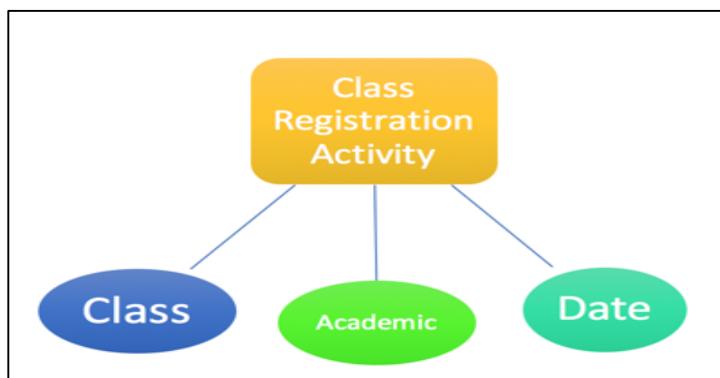
Bus Matrix

Instructions!		Fact Grain Type	Granularity	Facts	Student	Academic	Term	Course	Date
Business Process Name	Fact Table				X	X	X	X	X
Student Registration Activity	FactCourseRegistration	Accumulative Snapshot	Each row represents number of courses registered and dropped per student per term	Student Course Enrollment count per term, Student Courses dropped count per term					

The following is a screenshot for the bus matrix that we created for the project.

Bubble Chart

A bubble chart showcases the relationship between various fact and dimensions. Following is the bubble chart that we created for our project.



Attribute List

Following is the list which is called the Attribute List. The attribute list includes attributes for the various dimensions and facts.

1 Instructions /	2 Dimension / Fact Table	3 Attribute / Fact Name	4 Description	5 Alternate Names	6 Sample Values
3 FactCourseRegistration	StudentKey	university			1,2,3,....
4 FactCourseRegistration	TermKey	Term key representing each term in the university			1,2,3,...
5 FactCourseRegistration	DateKey	Date key representing the date on which registration was done			6/6/07
6 FactCourseRegistration	CoursesReg	Total number of courses registered by each student per term			35
7 FactCourseRegistration	CoursesDropped	Total number of courses dropped by each student per term			12
8 Student	StudentKey	Surrogate Key			1,2,3,...
9 Student	StudentsUID	Student University ID			631888224
10 StudentGender	StudentGender	Student Gender			Male
11 StudentGPA	StudentGPA	Student GPA			3.2
12 Student	StudentAcadCareer	Defines whether student is Graduate or Undergraduate			UGRAD
13 Student	StudentProgramDesc	Defines each student's major			Internship in Info Studies
14 Academic Terms	AcademicKey	Surrogate key			1,2,3,...
15 Academic Terms	AcademicTermid	Defines a particular ID for each Academic Term			12
16 Academic Terms	AcademicTermDesc	Definice term associated to particular term ID			123431
17 Course	CourseKey	Surrogate Key			1,2,3,...
18 Course	CourseTerm	Term in which the course is offered			11232
19 Course	CourseTitle	Title of each course			Introduction to Info. Sciences
20 Course	CourseCode	Specifc ID for each course			321123
21 Date	DateKey	Uniqye Key for Date			1,2,3,...
22 Date	Date	Full date as a SQL date			1/1/44
23 Date	FullDateUSA	Full Date in USA format			1/1/44
24 Date	DayOfWeek	Day Number in the week			7
25 Date	DayName	Name of the Day			Saturday
26 Date	DayOfMonth	Day Number in the month			1
27 Date	DayOfYear	Day Number in the year			1
28 Date	WeekOfYear	Week Number in the year			1
29 Date	MonthName	Name of the Month			January
30 Date	MonthOfYear	Month Number of the year			1
31 Date	Year	Year			1944

Issue List

Following is the list of issues that came up while working on the project. We were able to solve all of the issues faced.

1	Topic	Issue Details	Date Identified	Identified by	Solved Date	Solved by	Status
2	1 Project scope	Which external source will be analyzed?	6/23/16	All	6/23/16	All	Solved
3	2 Determination of Dimensions	Which dimension table should be involved with the fact?	6/24/16	Rajith	6/26/16	Rajith	Solved
4	3 Conversion of Data	The data type of destination table and source database are different	6/28/16	Saurabh	6/29/16	Saurabh	Solved
5	4 Fact generation	Generate fact table and upload data from source table	6/28/16	Sukhmeet	6/29/16	Sukhmeet	Solved

DESIGN

Detailed Bus Matrix

Instructions!		Fact								
Business Process Name	Fact Table	Fact Grain Type	Granularity	Facts	Student	Academic Term	Course	Date		
Student Registration Activity	FactCourseRegistration	Accumulative Snapshot	Each row represents number of courses registered and dropped per student per term	Student Course Enrollment count per term, Student Courses dropped count per term	x	x	x	x		

Fact/ Derived Fact Worksheets:

➤ FACT TABLE

- FactCourseRegistration- Description, Example Value, Display Folder

Table Name		FactCourseRegistration	Home Page			
Table Type		Fact				
Display Name		Course Registration Fact				
Database Schema		ExternalSources				
Table Description		Course Registration Fact				
Comment		Course Registration Fact				
Biz Filter Logic						
Size						
Generate Script?		Y				
Column Name		Display Name	Description	Example Values	Display Folder	
StudentKey	StudentKey	Student key representing each student in the university	1, 2, 3	key		
TermKey	TermKey	Term key representing each term in the university	1	key		
CourseKey	CourseKey			key		
DateKey	DateKey	Date key representing the date on which registration was done	20180501	key		
CoursesReg	CoursesReg	Total number of courses registered by each student per term	5	amounts		
CoursesDropped	CoursesDropped	Total number of courses dropped by each student per term	4	amounts		

- FactCourseRegistration- Target, Source

Table Name	FactCourseRegistration											
Column Name	Display Name	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
StudentKey	StudentKey				PK,FK DimStudent.StudentKey	N		Derived				
TermKey	TermKey				PK,FK DimAcademicTerm.AcademicTermKey	N		Derived				
CourseKey	CourseKey				PK,FK DimClassSection.CourseKey	N		Derived				
DateKey	DateKey				PK,FK DimDate.DateKey	N		Derived				
CoursesReg	CoursesReg						N	Derived				nvarchar
CoursesDropped	CoursesDropped				FK		N	Derived				nvarchar

- FactCourseRegistration- ETL Rules, Comments

Table Name	FactCourseRegistration											
Column Name	Display Name	ETL Rules				Comments						
StudentKey	StudentKey					Key look up from DimStudent.StudentKey						
TermKey	TermKey					Key look up from DimAcademicTerm.AcademicTermKey						
CourseKey	CourseKey					Key look up from DimClassSection.CourseKey						
DateKey	DateKey					Key look up from DimDate.DateKey						
CoursesReg	CoursesReg											
CoursesDropped	CoursesDropped											

➤ Detailed Dimensional Design Worksheet

A) Dimension Tables:

DimStudent

Table Name	DimStudent	Home Page																		
Column Name	Display Name	Description	Unknown Member	Example Values	SCD Type	Display Folder	ETL Rules	Comments	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
StudentKey	StudentKey	Surrogate primary key	-1	1, 2, 3...	key				int			PK ID	N			v_it_students				
StudentSID	StudentSID	Business key from source system (aka natural key)	-1	12345	key				varchar	10		N				ExternalSource	dbo	v_it_students	student_id	varchar
StudentGPA	StudentGPA	Student GPA	-1	3.48	2				varchar	10		N				ExternalSource	dbo	v_it_students	stu_gpa	numeric
StudentGender	StudentGender	Student Gender	None	Male	2				varchar	10		N				ExternalSource	dbo	v_it_students	gender	varchar
StudentAcadCareer	StudentAcadCareer	Defines whether student is Graduate or Undergraduate	None	UGRAD	2				varchar	50		N				ExternalSource	dbo	v_it_students	acad_career	varchar
StudentProgramDesc	StudentProgramDesc	Defines each student's major	None	Internship in Info Studies	2				varchar	50		N				ExternalSource	dbo	v_it_students	acad_prog_primary_desc	varchar
RowIsCurrent	Row Is Current	Is this the current row for this member (Y/N)?	Y	Y, N	n/a	Exclude from cube	Standard SCD-2		char	1		N				Derived				
RowStartDate	Row Start Date	When did this row become valid for this member?	1/1/00	1/24/11	n/a	Exclude from cube	Standard SCD-2		datetime			N				Derived				
RowEndDate	Row End Date	When did this row become invalid? (12/31/9999 if current row)	12/31/99	1/14/1998, 12/31/9999	n/a	Exclude from cube	Standard SCD-2		datetime			N	12/31/99			Derived				
RowChangeReason	Row Change Reason	Why did the row change last?	N/A		n/a	Exclude from cube	Standard SCD-2		varchar	200		N				Derived				

DimAcademicTerm

Table Name	DimAcademicTerm	Home Page																		
Column Name	Display Name	Description	Unknown Member	Example Values	SCD Type	Display Folder	ETL Rules	Comments	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
AcademicTermKey	AcademicTermKey	Surrogate primary key	-1	1, 2, 3...	key				int			PK ID	N			Derived				
TermID	TermID	Business key from source system (aka natural key)	-1	1102	key				varchar	4		N				ExternalSource	dbo	v_academic_terms	TERM	varchar
Term Desc	Term Desc	Description of the academic term	None	Spring 2018	2				varchar	50		N				ExternalSource	dbo	v_academic_terms	TERM_DESC	varchar
TermStartDateKey	TermStartDateKey	Unique key of the academic start date	-1	20180901	2				int			N				ExternalSource	dbo	v_academic_terms	TermStartDateKey	int
TermEndDateKey	TermEndDateKey	Unique key of the academic end date	-1	20180905	2				int			N				ExternalSource	dbo	v_academic_terms	TermEndDateKey	int
TermAcademicYear	TermAcademicYear	Academic term's academic Year	None	2017-2018	2				varchar	12		N				ExternalSource	dbo	v_academic_terms	TermAcademicYear	varchar
RowIsCurrent	Row Is Current	Is this the current row for this member (Y/N)?	Y	Y, N	n/a	Exclude from cube	Standard SCD-2		char	1		N				Derived				
RowStartDate	Row Start Date	When did this row become valid for this member?	1/1/00	1/24/11	n/a	Exclude from cube	Standard SCD-2		datetime			N				Derived				
RowEndDate	Row End Date	When did this row become invalid? (12/31/9999 if current row)	12/31/99	1/14/1998, 12/31/9999	n/a	Exclude from cube	Standard SCD-2		datetime			N	12/31/99			Derived				
RowChangeReason	Row Change Reason	Why did the row change last?	N/A		n/a	Exclude from cube	Standard SCD-2		varchar	200		N				Derived				

DimClassSection

Table Name	DimClassSection																			
Table Type	Dimension																			
Display Name	DimClassSection																			
Database Schema	ExternalSources																			
Table Description	Class Section Dimension																			
Comment	Consists list of class sect																			
Biz Filter Logic																				
Size																				
Generate Script?	Y																			
Column Name	Display Name	ETL Rules	Com	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype					
CourseKey	CourseKey			int				PK ID	N		Derived									
ClassNumber	ClassNumber			nvarchar	10				N		ExternalSources dbo	v_lst_class_section	classNumber	nvarchar						
CourseTitle	CourseTitle			nvarchar	50				N		ExternalSources dbo	v_lst_class_section	courseTitle	nvarchar						
CourseCode	CourseCode			nvarchar	50				N		ExternalSources dbo	v_lst_class_section	courseNum	nvarchar						
RowIsCurrent	Row Is Current	Standard SCD-2		nchar	1				N		Derived									
RowStartDate	Row Start Date	Standard SCD-2		datetime					N		Derived									
RowEndDate	Row End Date	Standard SCD-2		datetime					N	12/31/99	Derived									
RowChangeReason	Row Change Reason	Standard SCD-2		nvarchar	200				N		Derived									

DimDate

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
Table Name	DimDate																			
Table Type	Dimension																			
Display Name	DimDate																			
Database Schema	ExternalSources																			
Table Description	Date dimension contains one row for every day, beginning at 1/1/2005. There may also be rows for "hasn't happened yet."																			
Comment	The Date dimension is derived; it is not populated from any transaction system.																			
Biz Filter Logic																				
Size	365/year																			
Generate Script?	Y																			
Column Name	Display Name	Description	Unknown Member	Example Values	SCD Type	Display Folder	ETL Rules	Com	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
FullDateCSA	FullDateCSA	String expression of the full date, eg MM/DD/YYYY	Unk date	23-Nov-04	1 Day				nchar	10			N		ExternalSources dbo	v_date_dimension	FullDateCSA	nchar		
DayOfWeek	DayOfWeek	Number of the day of week, Sunday = 1	0..7	1..7	1 Day				tinyint				N		ExternalSources dbo	v_date_dimension	DayOfWeek	tinyint		
DayName	DayName	Day name of week, eg Monday	Unk date	Sunday	1 Day				nvarchar	10			N		ExternalSources dbo	v_date_dimension	DayName	nvarchar		
DayOfMonth	DayOfMonth	Number of the day in the month	0..31	1..31	1 Day				tinyint				N		ExternalSources dbo	v_date_dimension	DayOfMonth	tinyint		
DayOfYear	DayOfYear	Number of the day in the year	0..365	0..365	1 Day				int				N		ExternalSources dbo	v_date_dimension	DayOfYear	int		
WeekOfYear	WeekOfYear	Week of year, 1..53	0..52 or 53	0..52 or 53	1 Calendar				tinyint				N		ExternalSources dbo	v_date_dimension	WeekOfYear	tinyint		
MonthName	MonthName	Month name, eg January	Unk month	November	1 Calendar				nvarchar	10			N		ExternalSources dbo	v_date_dimension	MonthName	nvarchar		
MonthOfYear	MonthOfYear	Month of year, 1..12	0..12..12	1..12..12	1 Calendar				tinyint				N		ExternalSources dbo	v_date_dimension	Month	tinyint		
Year	Year	Calendar year, eg 2010	0	2004	1 Calendar				int				N		ExternalSources dbo	v_date_dimension	Year	int		

DIMENSIONAL HIERARCHIES

- Dimensional Hierarchy in DimAcademicTerm

The screenshot shows the 'Hierarchies' section of the Dimensional Hierarchy interface. A hierarchy for 'Academic Year-Academic Term-Academic Term ID-Academic Term Key' is displayed, with levels: Academic Year, Academic Term, Academic Term ID, and Academic Term Key. A tooltip 'To create a new hierarchy, drag an attribute here.' is visible below the hierarchy tree. The 'Attributes' pane on the left lists attributes for 'Academic Term' such as Academic Term, Academic Term ID, Academic Term Key, and Academic Year. The 'Data Source View' pane on the right shows the schema for 'DimAcademicTerm' with columns: AcademicTermKey, TermID, Term_Desc, TermStartDateKey, TermEndDateKey, TermAcademicYear, RowIsCurrent, RowStartDate, RowEndDate, and RowChangeReason.

- Dimensional Hierarchy in DimClassSection

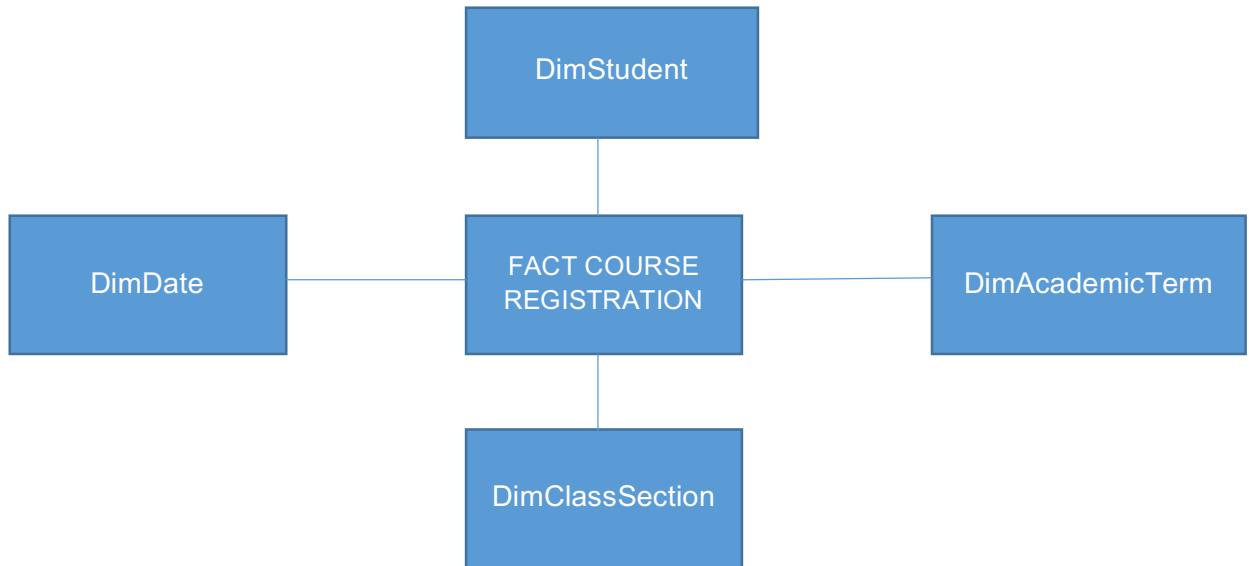
The screenshot shows the 'Hierarchies' section of the Dimensional Hierarchy interface. A hierarchy for 'Term-Academic Level-Location-Course Title-Course Subject-Class ID-Class Key' is displayed, with levels: Term, Academic Level, Location, Course Title, Course Subject, Class ID, and Class Key. A tooltip 'To create a new hierarchy, drag an attribute here.' is visible below the hierarchy tree. The 'Attributes' pane on the left lists attributes for 'Class Section' such as Acad Career, Class ID, Class Key, Class Section, Course Key, Course Subject, Course Title, Location, and Term. The 'Data Source View' pane on the right shows the schema for 'DimClassSection' with columns: ClassSectionKey, ClassID, TermID, ClassNumber, CourseSubject, CourseNumber, CourseTitle, ClassSection, ClassStartDateKey, and ClassEndDateKey.

- Dimensional Hierarchy in DimDate

The screenshot shows the 'Hierarchies' section of the Dimensional Hierarchy interface. A hierarchy for 'Year-Month-Date' is displayed, with levels: Year, Month, and Date. A tooltip 'To create a new hierarchy, drag an attribute here.' is visible below the hierarchy tree. The 'Attributes' pane on the left lists attributes for 'Date' such as Date, Day, Day Of Week, Month, and Year. The 'Data Source View' pane on the right shows the schema for 'DimDate' with columns: DateKey, Date, FullDateUSA, DayOfWeek, DayName, DayOfMonth, DayOfYear, WeekOfYear, MonthName, MonthOrder, and MonthYear.

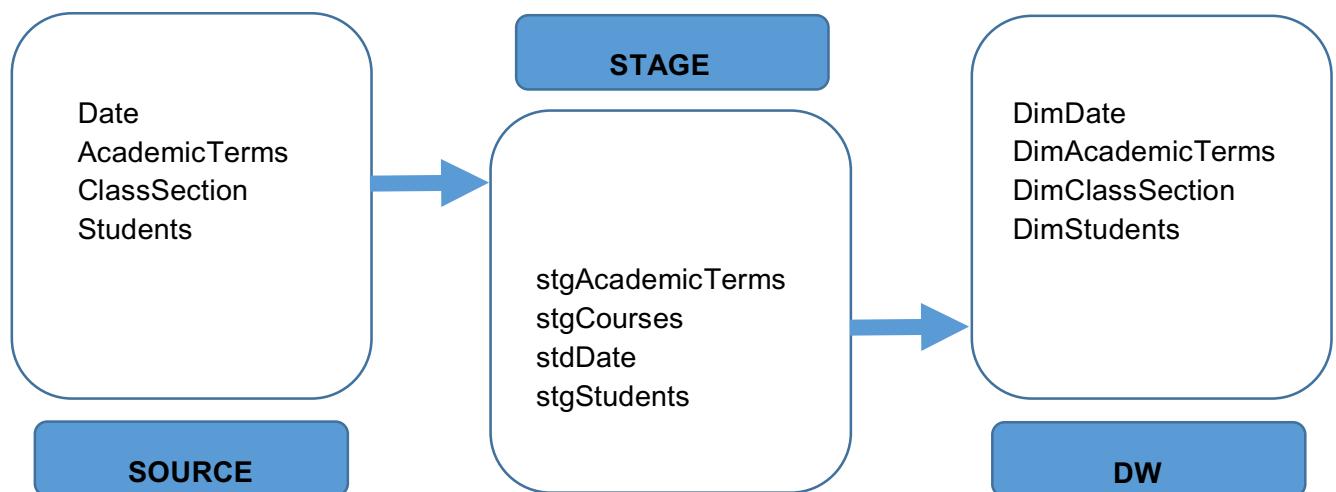
STAR SCHEMA MODEL DIAGRAM

In this case, we have used the Star Schema model as each dimension is represented by a single table in the ROLAP schema.



ETL SPECIFICATIONS

High - Level Source to Target Map



Peer Evaluation

Task	Rajith Jayadevan	Saurabh Jape	Sukhmeet Singh	Ananya Chakraborty
Creating deliverables	5	5	5	5
Integrating deliverables	3	3	5	5
High Level Dimensional Model	5	5	4	5
Detailed Level Dimensional Model	5	5	5	4
Creating ROLAP Schema	4	4	4	3
ETL code Package	4	5	3	4
MOLAP Database	5	4	5	5
Final Review	5	5	5	5

- 1- No Contribution
- 2- Barely made a contribution
- 3- Medium contribution

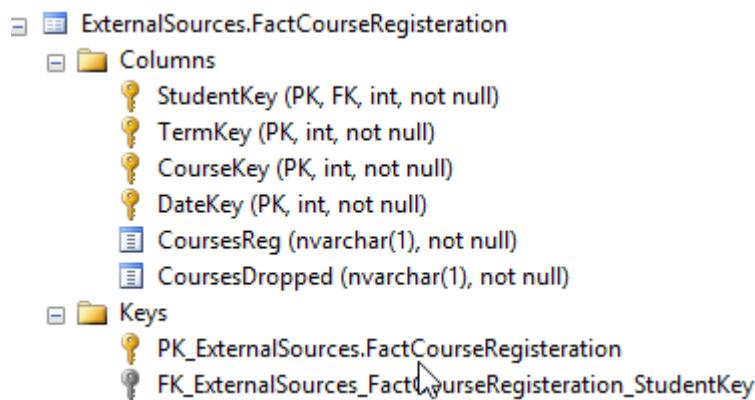
- 4- Good Contribution
- 5- Excellent Contribution

IMPLEMENTATION

➤ ROLAP SCHEMA IN SQL SERVER

a) Fact Table

- Fact Registration Activity:



b) Dimension Tables

DimAcademicTerm:

ExternalSources.DimAcademicTerm
Columns
AcademicTermKey (PK, int, not null)
TermID (nvarchar(4), not null)
Term_Desc (nvarchar(50), null)
TermStartDateKey (int, not null)
EndDateKey (int, null)
TermAcademicYear (nvarchar(12), not null)
RowIsCurrent (nchar(1), null)
RowStartDate (datetime, null)
RowEndDate (datetime, null)
RowChangeReason (nvarchar(200), null)
Keys
PK_ExternalSources.DimAcademicTerm
Constraints

DimClassSection:

ExternalSources.DIMClassSection
Columns
CourseKey (PK, int, not null)
ClassNumber (nvarchar(400), null)
CourseTitle (nvarchar(50), not null)
CourseCode (nvarchar(50), not null)
RowIsCurrent (nchar(1), null)
RowStartDate (datetime, null)
RowEndDate (datetime, null)
RowChangeReason (nvarchar(200), null)
Keys
PK_ExternalSources.DimClassSection

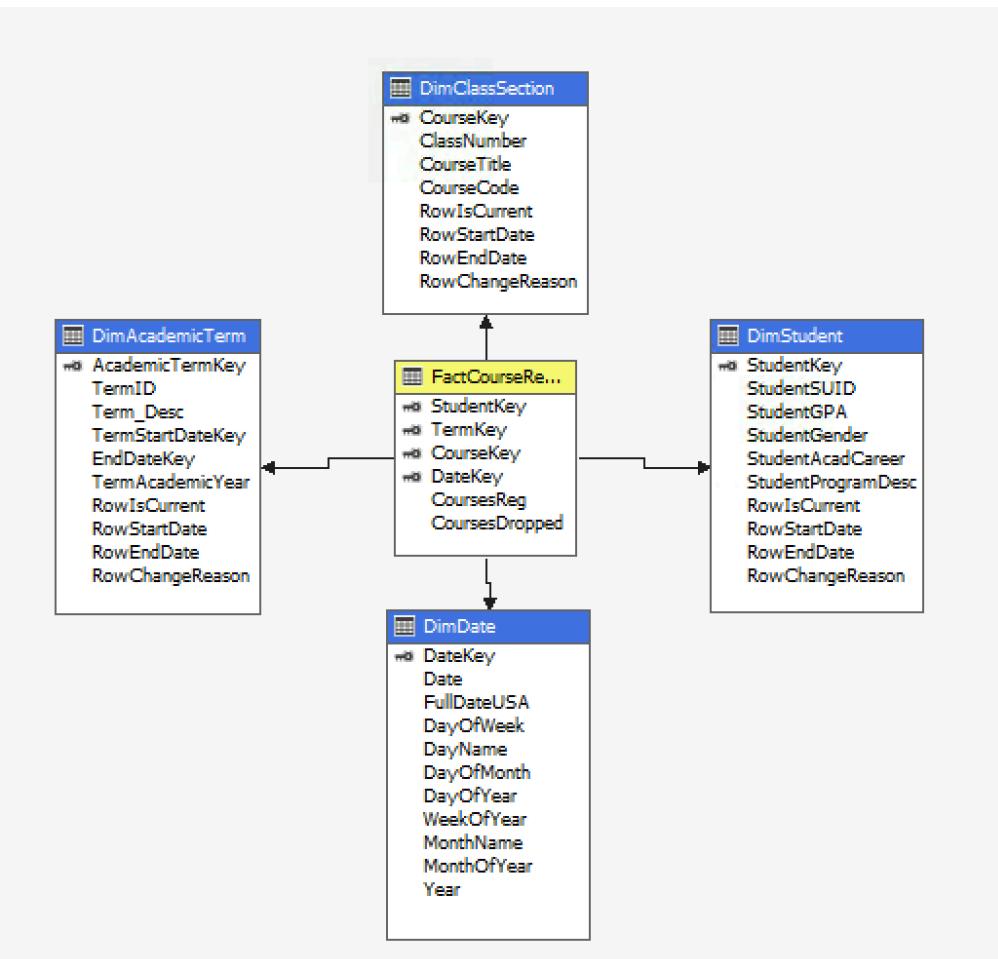
DimDate:

ExternalSources.DimDate
Columns
DateKey (PK, int, not null)
Date (datetime, null)
FullDateUSA (nchar(10), not null)
DayOfWeek (tinyint, not null)
DayName (nvarchar(10), not null)
DayOfMonth (tinyint, not null)
DayOfYear (int, not null)
WeekOfYear (tinyint, not null)
MonthName (nvarchar(10), not null)
MonthOfYear (tinyint, not null)
Year (int, not null)
Keys
PK_ExternalSources.DimDate

DimStudent:

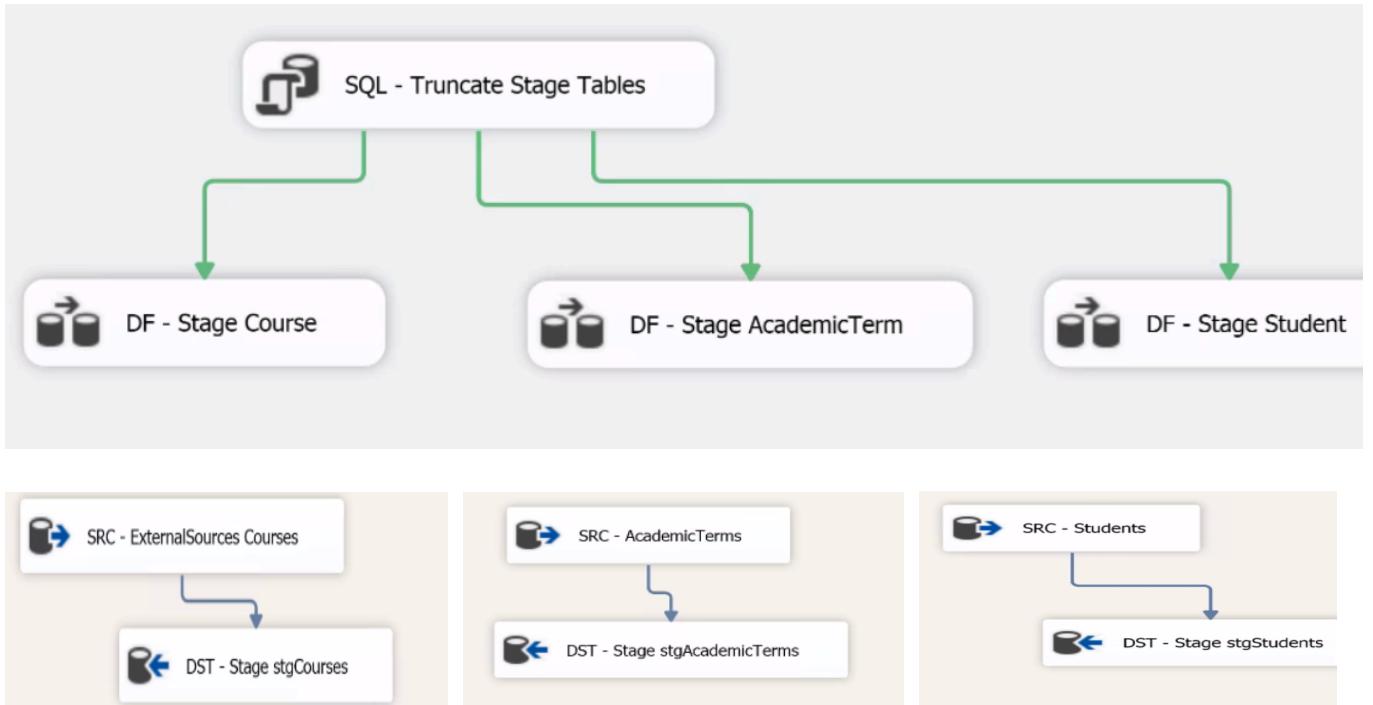
ExternalSources.DimStudent
Columns
StudentKey (PK, int, not null)
StudentSID (nvarchar(100), null)
StudentGPA (decimal(18,0), null)
StudentGender (nvarchar(20), null)
StudentAcadCareer (nvarchar(50), not null)
StudentProgramDesc (nvarchar(50), not null)
RowIsCurrent (nchar(1), null)
RowStartDate (datetime, null)
RowEndDate (datetime, null)
RowChangeReason (nvarchar(200), null)
Keys
PK_ExternalSources.DimStudent
Comments

Star Schema Diagram

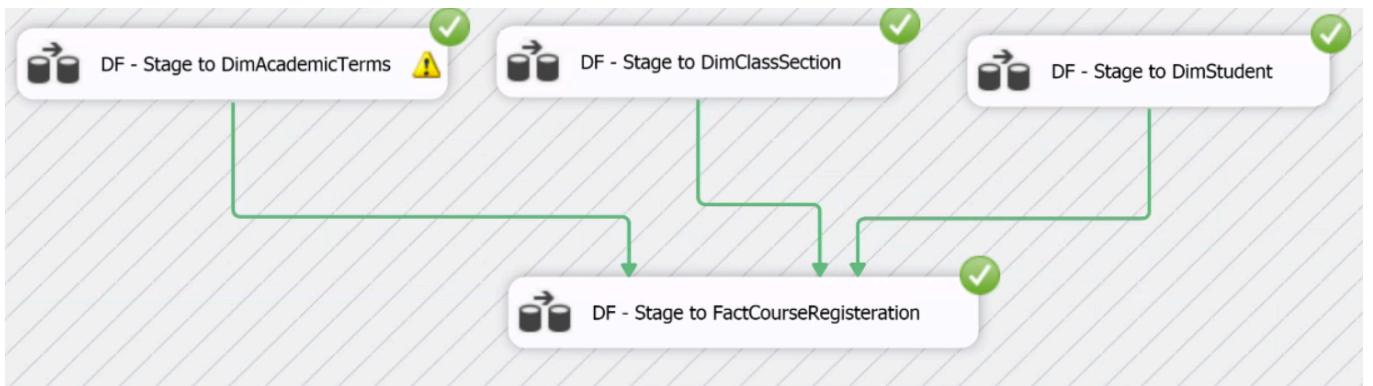


SSIS ETL Code/Packages

Populate ROLAP Schema from Sources



Load from Stage table to Target tables



Analysis Service Molap Database

➤ Molap Cube Structure

