Enrollment Analysis Group Project

IST 722

Spring Semester 2015

Group B

Sahil Baxi, Sam Edelstein, Garnette Pereira, Gaurav Shinde, Qiong Wu

Table of Contents

Overview Section 3
- **Project Charter** 3
  - Define project background 3
  - Project Scope 3
  - Success Criteria 3
  - Business Justification 4
- **Team Member Roles and Responsibilities** 5
- **Key Stakeholders** 6

Analysis Section 7
- **Business processes** 7
- **Bus matrix** 7
- **Bubble Chart** 7
- **Attribute List** 8
- **Issues list** 10

Design Section 11
- **Detailed bus matrix** 11
- **Fact / Derived fact worksheet** 11
- **Detailed dimensional design worksheet** 11
- **Dimensional Hierarchies** 11
- **Snowflake model diagrams** 13
- **ETL Specifications** 13
- **Detailed ETL flow for each source to target** 14

Team Contribution Report 21

Implementation 22
- **Rolap Schema in SQL Server** 22
- **SSIS ETL Code / Packages** 22
- **Analysis Services MOLAP Database** 22
- **BI Front End** 22

Poster 23

Appendix – Attached Files 24

Overview Section

Project Charter

Define project background

The School of Information Studies at Syracuse University has a large amount of data that needs to be analyzed. Based on feedback from both end-users and staff in the school, people are disappointed with the enrollment process. Oftentimes, classes are filled immediately after the enrollment period opens, but by the time the first day of class comes, and again after the first couple of weeks of classes have passed, many students have dropped their respective classes, leaving empty spots where others might have enrolled, but now it is too late.

This project will use data from the School of Information Studies' database to increase the efficiency of the student enrollment processes. The project team will create dimensional models within a data warehouse to aid in making the final recommendations.

Project Scope

This project's scope will focus on several areas, including:

- Recommending, designing, building, and implementing a data warehouse for the School of Information Studies to help better plan their enrollment processes
- Analyze enrollment data for both online and in-person classes along with the respective drop rates and waitlist numbers for each class
- Create findings that include an overall documentation of each class that gets dropped by term (transactional), how many classes are dropped in online classes versus in person classes (periodic snapshot), and how many people that sit on a waitlist end up enrolling in a class (accumulating snapshot).
- No new attributes will be added to existing data. Only the real-time data that is currently collected will be used.
- Answer whether the School of Information Studies should purposely over-enroll classes to protect against empty seats after people drop the class. Answer if online courses experience higher drop rates than in person courses.
- Only looking at solutions for the School of Information Studies
- Deliver the Business Intelligence tool using Excel or similar program
- Create a poster for presentation to be delivered to Professor Fudge

Success Criteria

A successful project will answer the following questions adequately:

- Based on waitlist data, which iSchool courses warrant additional sections?

- Based on enrollment data, which iSchool course sections should be dropped in future semesters?

Business Justification

This project will benefit the School of Information Studies in several ways. When students are unable to enroll in a course because it is full, they may be frustrated and end up signing up for a class that was not their top choice. This dissatisfaction could result in lower grades due to a lack of motivation. It could also lead to a student feeling that the experience at the School of Information Studies was not good, potentially negatively affecting future recruitment efforts.

The School of Information Studies also wants to fill its classes as much as possible. If classes are experiencing high drop rates, seats that would have otherwise been filled may stay empty. This presents an inefficiency for the School of Information Studies. Even those students who choose to sit on a waitlist in hopes that other students drop, may end up enrolling in a different class, and ultimately will not accept their opportunity to join the class they originally wanted.

Team Member Roles and Responsibilities

Roles and responsibilities for the project were split up amongst the group members. Though each group member had specific tasks with which they were responsible, all group members weighed in on every aspect of the project and had a hand in the project's ultimate success. Work was well distributed, and each group member met deadlines. The work responsibilities were distributed as follows:

Business Lead: Sam Edelstein and Garnette Pereira
Project Manager: Sahil Baxi and Qiong Wu
Data Architect: All
ETL: Sahil Baxi and Gaurav Shinde
Business Analyst: Gaurav Shinde and Qiong Wu
Business Intelligence: Sam Edelstein and Garnette Pereira

Key Stakeholders

This project focused on enrollment analysis for the School of Information Studies at Syracuse University and was completed for Professor Michael Fudge. The key stakeholders for this project are:

- **School of Information Studies Dean and other leadership** – Analysis helps to answer how to dedicate limited faculty resources to class sections and which courses are over or under-enrolled. The leadership needs to know long term trends to help better plan for future courses and faculty workload.
- **School of Information Studies faculty** – The faculty can track how their class enrollment numbers compare to other similar class sections and can see why they would be asked to teach different courses because of student demand.
- **School of Information Studies enrollment management and counselors –** The analysis will give a better sense of the types of courses other students are taking and will result in better advice about the types of courses that will be available.
- **School of Information Studies students** – Students are eager to enroll in the class sections they need to fill requirements and that are essential to their education to help with a future career. More course sections in classes that have consistently high demand and fewer course sections of course that have low demand will enable every student to enroll in the classes they need to graduate.
- **Syracuse University administration –** Though this exact model would not apply to schools and colleges outside of the iSchool, similar models could be built across Syracuse University to ensure efficient allocation of resources when it comes to class sections. Particularly for undergraduate students who tend to take classes in many different schools, ensuring the appropriate number of seats, and professors, are available is critical to a well-functioning university.

Analysis Section

Business processes

The business process for this project focuses on enrollment analysis. All associated details are listed below. The files can also be found in the attached documents:

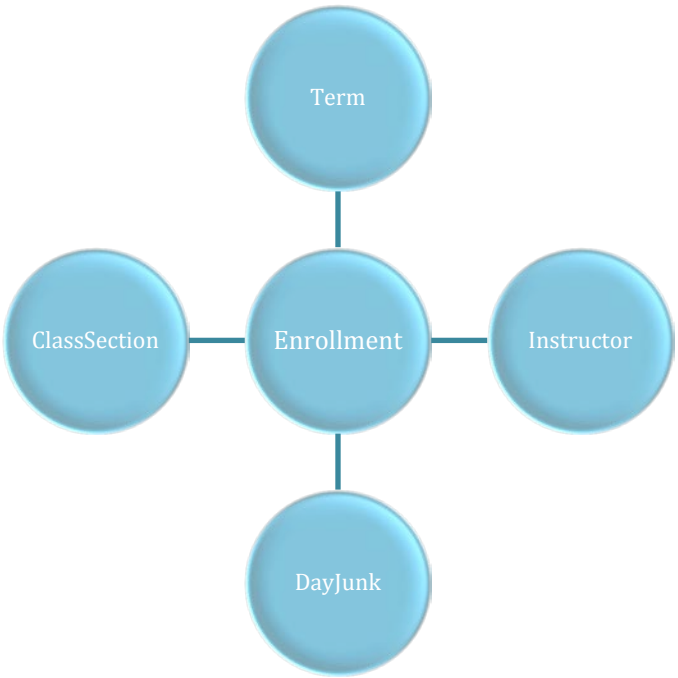
- Project-High-Level-Worksheet.xlsx

Bus matrix

The matrix shown above represents iSchool Enrollment business processes and its associated dimensions. This matrix is a blueprint for our Data Warehouse/Business Intelligence data organization. Each row in the matrix represents a business process, while each column corresponds to a dimension of the business. The matrix cells marked with an 'X' indicate which columns are associated with each row.

Business Process Name	Fact Table	Fact Grain Type	Granulairty	Facts	Term	Instructor	ClassSection	DayJunk
EnrollmentAnalysis	FactEnrollment	Transaction	One row means per class section at one term	CountCapacity, CountEnrolled, CountWaitlist, CountAvailable, CpuntKPIClass	X	X	X	X

Bubble Chart



Above is an example of a high-level model diagram shown through the graphical design of a bubble chart. It shows the dimensional model relationship with fact table in the middle and other five dimensions surrounding. This particular

bubble chart shows the fact and dimension table applicable to the enrollment analysis business process.

Attribute List

The Attributes and Metrics List break down each dimension and fact into further detail. The dimensions contain descriptions of the attributes of each dimension to include: examples of the data found in those fields of information, and alternative names and/or descriptions in which the data may be seen. The fact tables contain attributes and foreign keys, which reference the primary keys in other dimensions. These primary and foreign keys are important to establish, as they will be referenced through the dimensional model. This list also allows the design team to further understand how the dimensions will work with one another in support of the business processes.

Dimension / Fact Table	Attribute / Fact Name	Description	Alternate Names	Sample Values
DimTerm	termKey	Primary Key		1
DimTerm	termId	Natural Key		1141
DimTerm	termName	Term Description		Spring 2015
DimTerm	TermAcademicYear	Term Acedemic Year		2014-2015
Dim_Instructor	InstructorKey	Primary Key		
Dim_Instructor	instId	instructor id /hashed_Suid		F201936345F8C3C2C39
Dim_Instructor	Name	firstName + ' ' + lastName		Michael A Fudge Jr
Dim_Instructor	instAffiliation	instructor primary affiliation		Adjunct Faculty
DimDayJunk	dayJunkKey	Primary Key		
DimDayJunk	MeetingDays	Class Meeting days		MW
DimDayJunk	OnMonday	Class on Monday or not		1
DimDayJunk	OnTuesday	Class on Tuesday or not		0
DimDayJunk	OnWednesday	Class on Wednesday or not		1
DimDayJunk	OnThursday	Class on Thursday or not		0
DimDayJunk	OnFriday	Class on Friday or not		0
DimDayJunk	OnSaturday	Class on Saturday or not		0
DimDayJunk	OnSunday	Class on Sunday or not		0
DimClassSections	classSectionKey	Primary Key		1
DimClassSections	id	Natural Key		1112.39757
DimClassSections	classNumber	5 digit class number		39757
DimClassSections	courseSubj	Course Subject		IST
DimClassSections	courseNum	Course Number		659
DimClassSections	courseTitle	Course Title		Data Admin Concepts
DimClassSections	classEnrollmentTotal	Total number of enrollment		15
DimClassSections	classAcadCareer	Undergrad or grad		GRAD
DimClassSections	COURSE_KEY	course key		10709
DimClassSections	classEnrollmentCapacity	class capacity		40
DimClassSections	classSection	Class Section		M001
DimClassSections	TERM	Term		1112
FactEnrollment	classSectionKey	PK,FK		1
FactEnrollment	dayJunkKey	FK		1
FactEnrollment	instructorsKey	FK		1
FactEnrollment	termKey	PK, FK		1
FactEnrollment	CountCapacity	Capacity of a partcular class section		40
FactEnrollment	CountEnrolled	number of enrolled students		25
FactEnrollment	CountWaitlist	number of students on the waitlist		10
FactEnrollment	CountAvailable	the number of seats available		15
FactEnrollment	CountKPIClass	Capacity - Enrolled + Waitlist		

Issues list

Issue #	Task / Topic	Issue	Identified Date	Reported By	Respon-sible	Status	Priority	Date Closed
1	Topic	Bus Matrix Design	27-Feb	Team	Team	Resolved		
2	Task	verify data available from External Source in five dimensions	8-Apr	Team	Team	Resolved		
3	Task	Detailed Dimensional Modeling * Revised with day junk	10-Apr	Team	Team	Resolved		
4	Task	Load data from stage to datawarehouse	15-Apr	Team	Team	Resolved		
5	Topic	Identify expected findings and data visualization	20-Apr	Team	Team	Resolved		

Design Section

Detailed bus matrix

Find full file attached: Project-High-Level-Worksheet.xlsx

Business Process Name	Fact Table	Fact Grain Type	Granulairty	Facts	Term	Instructor	ClassSection	Daylunk
EnrollmentAnalysis	FactEnrollment	Transaction	One row means per class section at one term	CountCapacity, CountEnrolled, CountWaitlist, CountAvailable, CpuntKPIClass	X	X	X	X

Fact / Derived fact worksheet

Table Name	FactEnrollment
Table Type	Fact
Display Name	Enrollment
Database Schema	
Table Description	Enrollment Fact
Comment	Comes from the business intelligence initative of external Sources
Biz Filter Logic	
Size	one row for class
Generate Script?	Y

Home Page

Column Name	Display Name	Description	Unknown Member	Example Values	Display Folder	ETL Rules	Target Datatype	Size	Precision	Key?	FK To	NULL?
ClassSectionKey	ClassSectionKey	Key to Class Sections Dimension		2			int			PK,FK	DimClassSection.ClassSectionKey	N
dayJunkKey	Day Junk Key	Key to Day Junk Dimension		1,2,3			int			FK		N
instructorsKey	instructorsId	key to Instructors Dimension		1,2,3			int			FK		N
termKey	Term ID	key to Term Dimension		1,2,3			int			PK,FK	DimTerm.termKey	N
CountCapacity	Count Capacity	Capacity of the Class		13,14,15			int					N
CountEnrolled	Count Enrolled	Number of students enrolled in a class		20,25,30			int					N
CountWaitlist	Count Waitlist	Number of students Waitlisted for a class		1,5,10,50			int					N
CountAvailable	Count Available	Count of Capacity - Count of Enrolled		1,2,4,5			int					N
CountKPIClass	Count KPI Class	Count Available + Count Waitlist		1,2,4,5,6			int					N
InsertAuditKey	Insert Audit Key	Key to Audit dimension for row insertion		ก, 2, 3	key	Standard auditing	int			FK	DimAudit.AuditKey	N
UpdateAuditKey	Update Audit Key	Key to Audit dimension for row update	-2	ก, 2, 3	key	Standard auditing	int			FK	DimAudit.AuditKey	N
BKFactTable	BK Fact Table	The natural key for the fact table, if any (eg order number)			Exclude from cube		int					Y

Detailed dimensional design worksheet

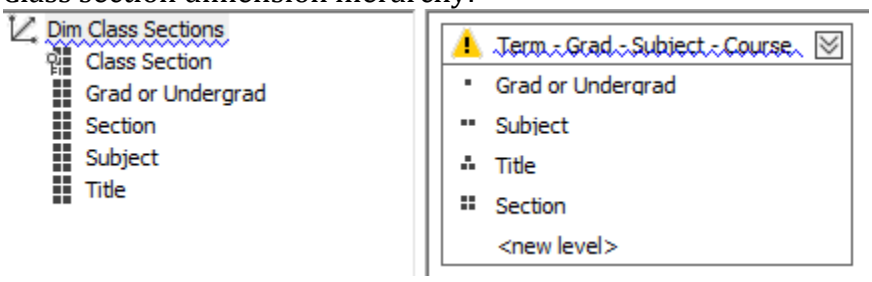
Table Name	DimClassSection
Table Type	Dimension
Display Name	ClassSection
Database Schema	
Table Description	Class Section dimension
Comment	comes from Class Section table in External Sources
Biz Filter Logic	
Size	one for each class number
Generate Script?	Y

Home Page

Column Name	Display Name	Description	Unknown Member	Example Values	SCD Type	Display Folder	ETL Rules
classSectionKey	ClassSectionKey	Surrogate primary key	-1	1, 2, 3...	key		
COURSE_KEY	CourseKey	Business key from source system (aka natural key)		7342	key		
id	ClassID	ID of the class		1112,3978	2		
TERM	Term	ID of the term		1112	2		
classAcadCareer	classAcademicCareer	Graduate or Undergraduate class		UGRD	2		
classEnrollmentCapacity	ClassCapacity	Enrollment capacity of the class		30	2		
classNumber	ClassNumber	Number of the class		39779	2		
classSection	ClassSection	Section number of the class		M001	2		
courseSubj	CourseSubject	Subject code of the course		IST	2		
courseNum	CourseNumber	Number of the course		639	2		
courseTitle	CourseTitle	Name of the course		Enterprise Technologies	2		
classEnrollmentTotal	TotalEnrollment	Total Enrollment of the class		40	2		
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
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
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
RowChangeReason	Row Change Reason	Why did the row change last?	N/A		n/a	Exclude from cube	Standard SCD-2
InsertAuditKey	InsertAuditKey	What process loaded this row?	-1		n/a	Exclude from cube	Standard Audit dim
UpdateAuditKey	UpdateAuditKey	What process most recently updated this row?	-1		n/a	Exclude from cube	Standard Audit dim

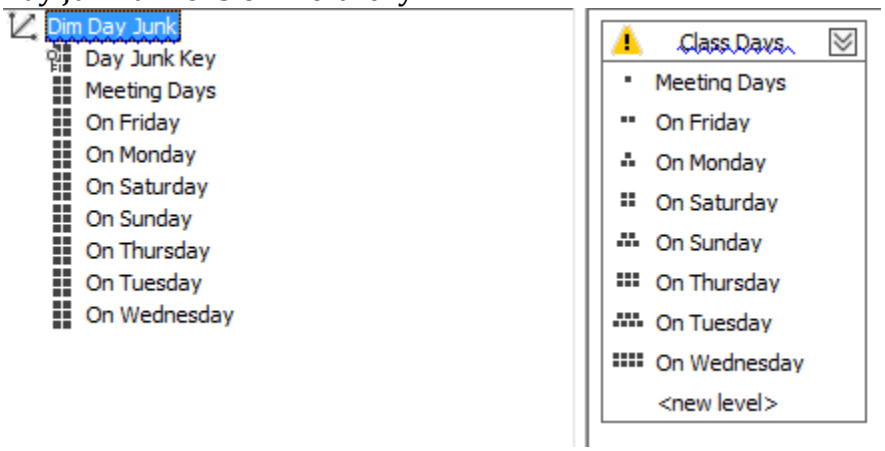
Dimensional Hierarchies

Class section dimension hierarchy:

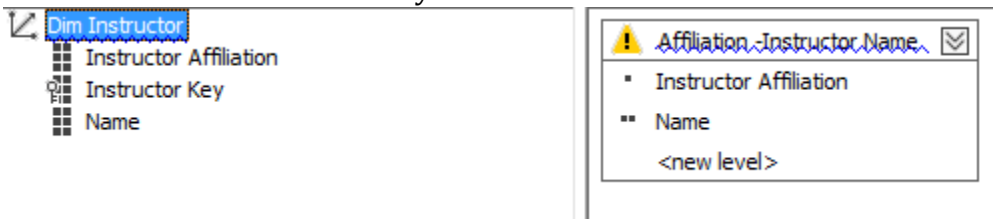


Term dimension hierarchy:

Day Junk dimension hierarchy:

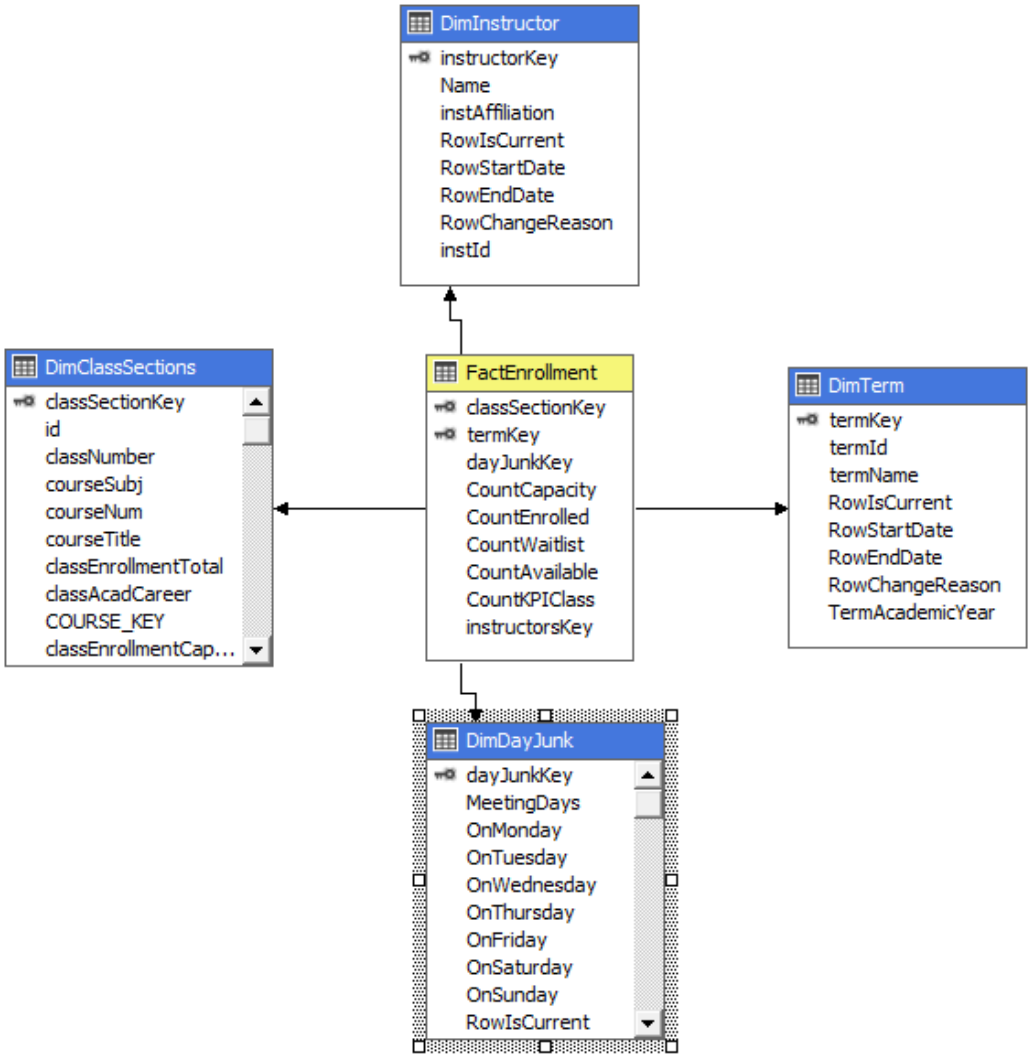


Instructor dimension hierarchy:



Snowflake model diagrams

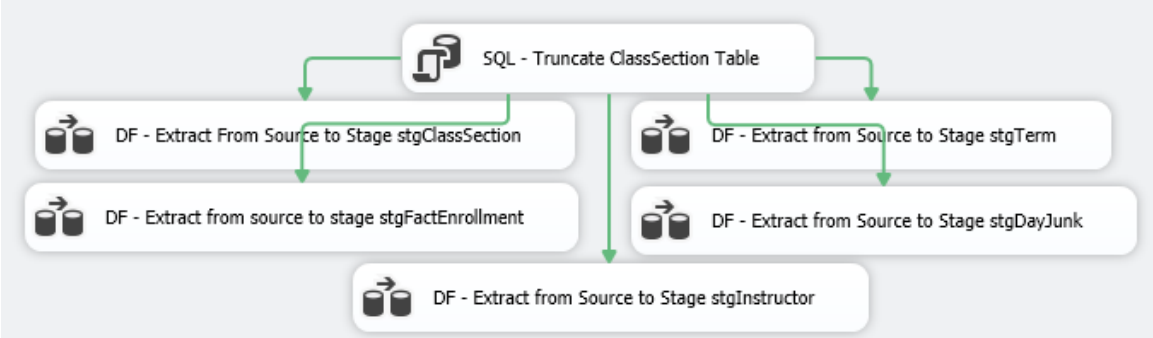
See SSAS Attachment for more details



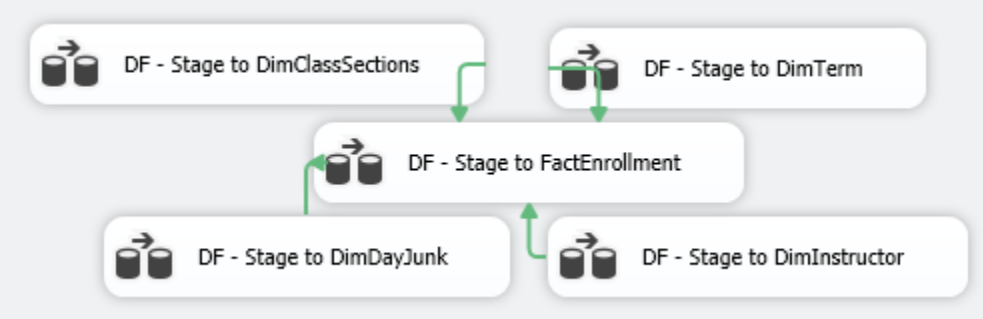
ETL Specifications

See SSIS Attachment for more details

From Source to Stage overview:



From Stage to Data Warehouse Overview:

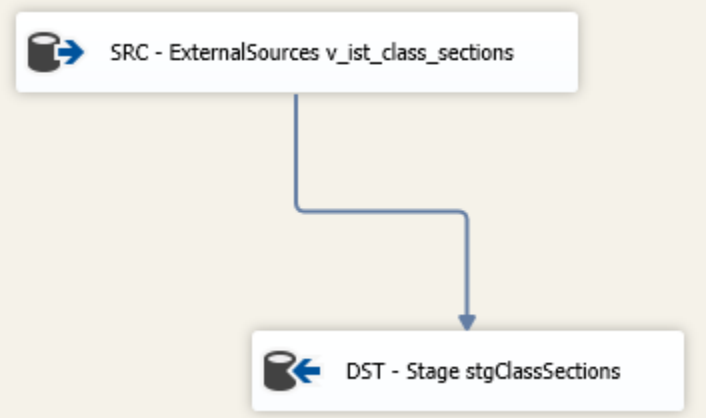


Detailed ETL flow for each source to target

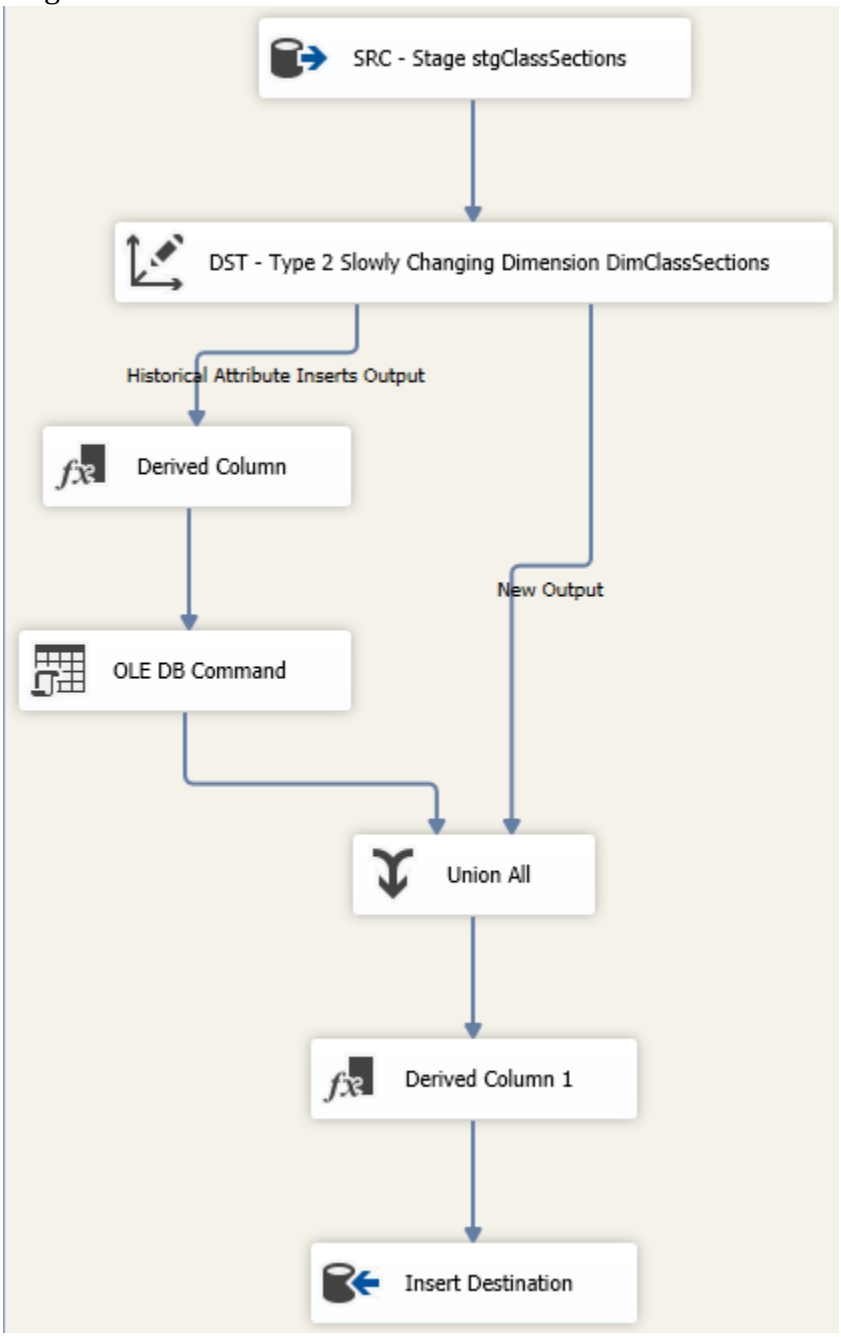
See SSIS Attachment for more details

		Target							Source				
Column Name	Display Name	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
ClassSectionKey	ClassSectionKey	int			PK,FK	DimClassSection.ClassSectionK	N	ETL Process	Dimension	dbo	DimClassSection	ClassSectionKey	int
dayJunkKey	Day Junk Key	int			FK		N	ETL Process	Dimension	dbo	DimDayJunk	dayJunkKey	int
instructorsKey	instructorsId	int			FK		N		Dimension	dbo	DimInstructor	instructorsKey	int
termKey	Term ID	int			PK,FK	DimTerm.termKey	N		Dimension	dbo	DimTerm	termKey	int
CountCapacity	Count Capacity	int					N						
CountEnrolled	Count Enrolled	int					N						
CountWaitlist	Count Waitlist	int					N						
CountAvailable	Count Available	int					N						
CountKPIClass	Count KPI Class	int					N						
InsertAuditKey	Insert Audit Key	int			FK	DimAudit.AuditKey	N	ETL Process					
UpdateAuditKey	Update Audit Key	int			FK	DimAudit.AuditKey	N	ETL Process					
BKFactTable	BK Fact Table	int					Y						

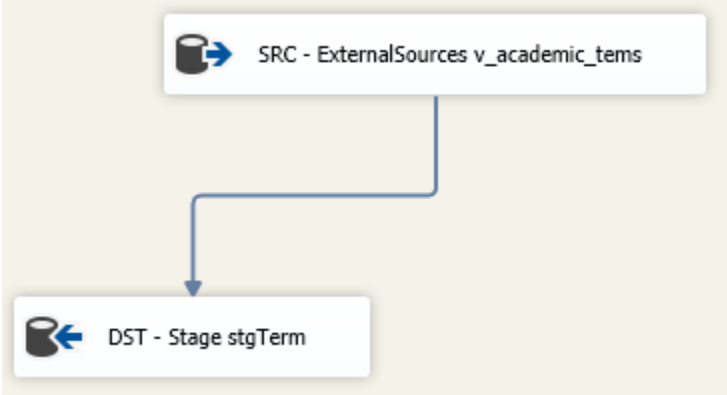
Class Section Dimension:
Source to Stage



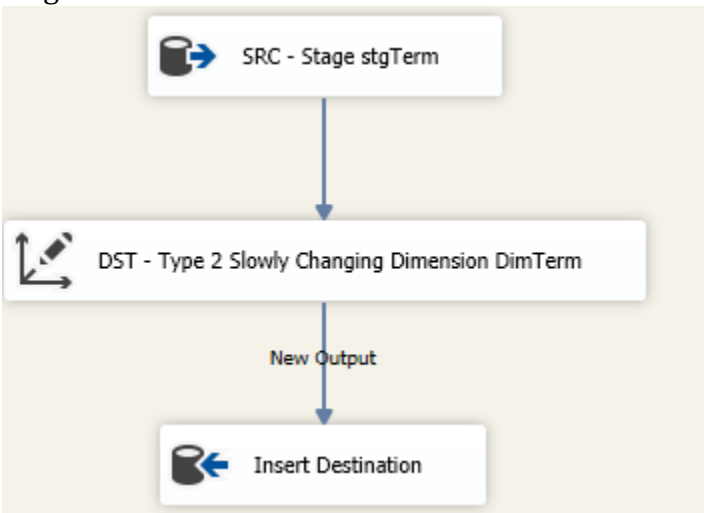
Stage to Data Warehouse



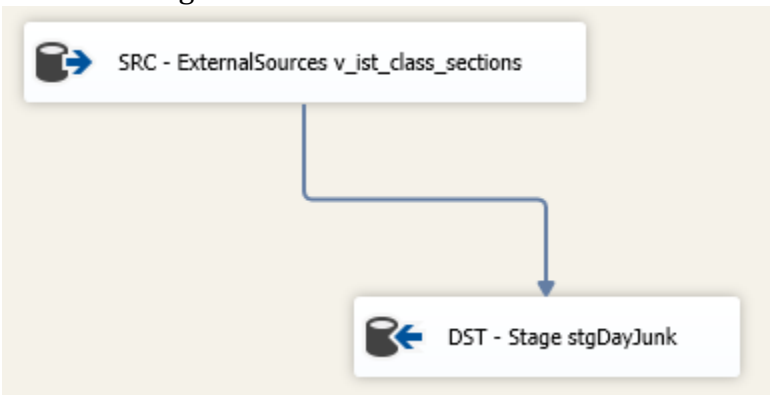
Term Dimension:
Source to Stage



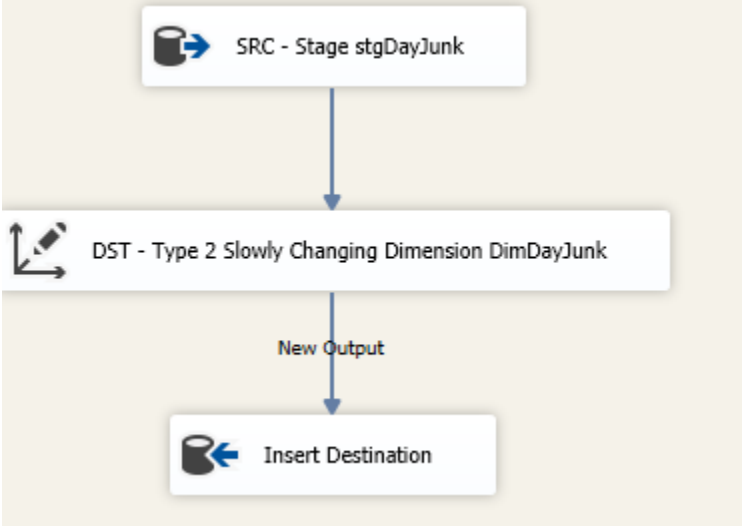
Stage to Data Warehouse



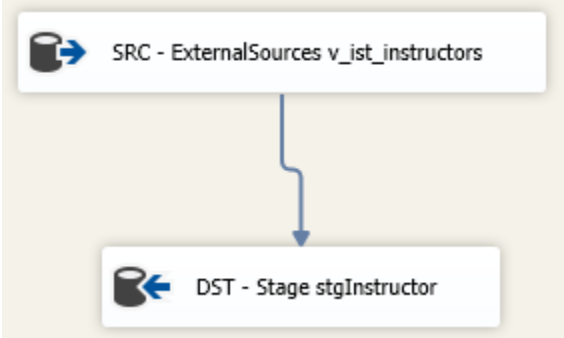
Day Junk Dimension:
Source to Stage



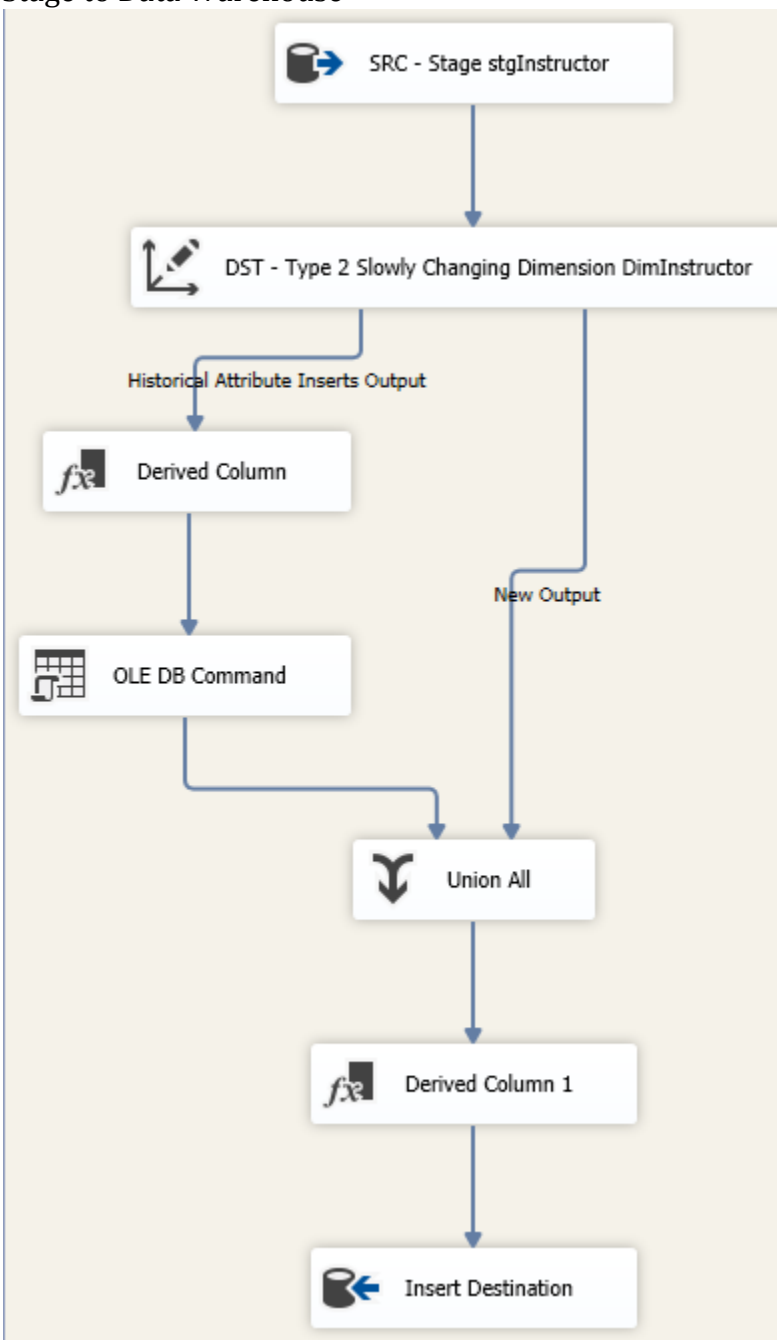
Stage to Data Warehouse



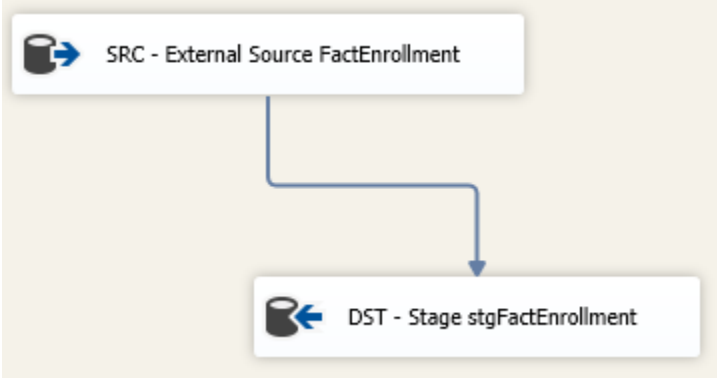
Instructor Dimension:
Source to Stage



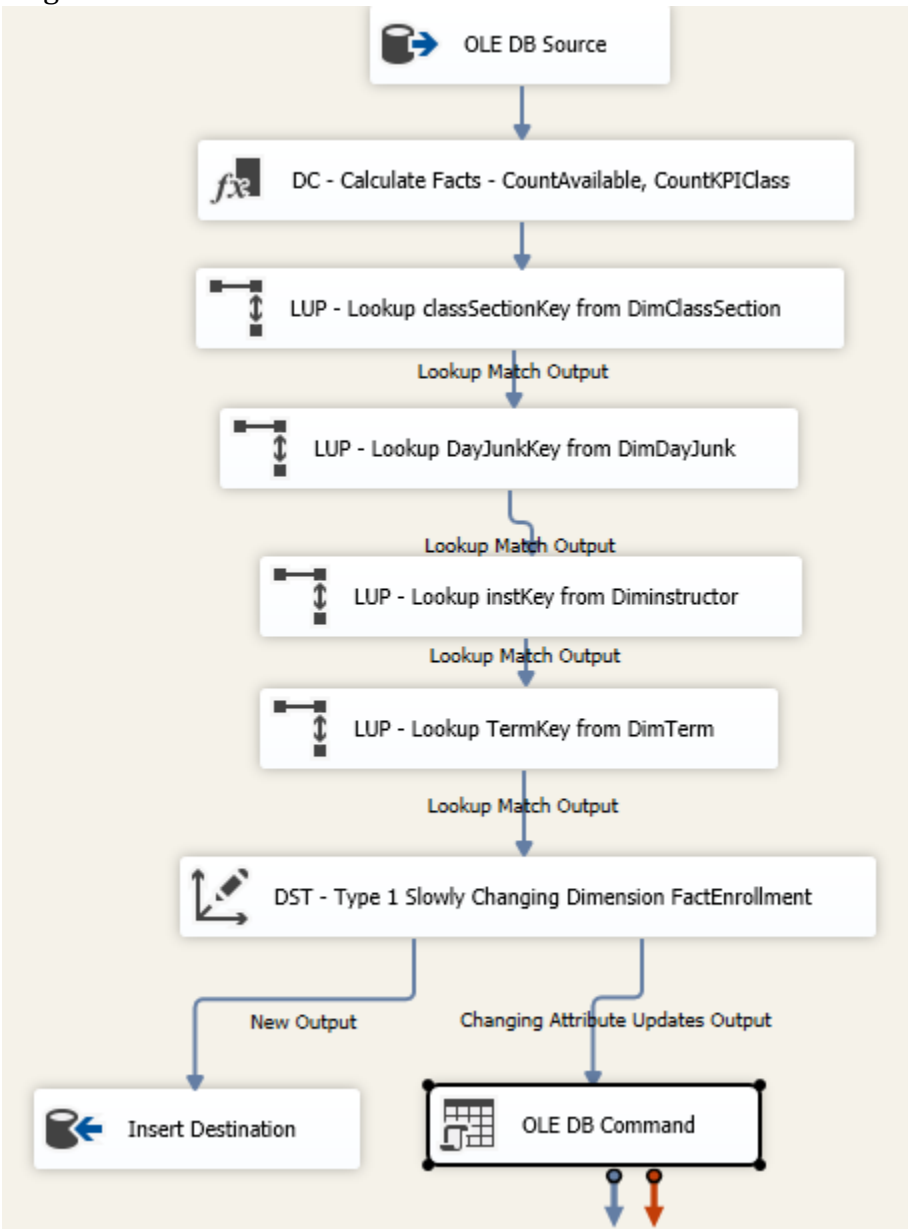
Stage to Data Warehouse



Fact Enrollment Dimension:
Source to Stage



Stage to Data Warehouse



Team Contribution Report

For this project, each team member had specific responsibilities, but all team members also worked collaboratively and stepped in to help in all aspects of the project when needed. The team held, at minimum, weekly meetings outside of class time to discuss the project and stay on task. At these meetings, work was completed and expectations were set for future meetings. All group members came to every meeting as long as they were available, and everyone played a part in ensuring project success.

Implementation

Rolap Schema in SQL Server

For this project, all data loaded into SQL Server can be found in the siedelst_stg and siedelst_dw databases.

SSIS ETL Code / Packages

All files related to SSIS ETL Code and packages can be found in the attached files: Group B Project.sln

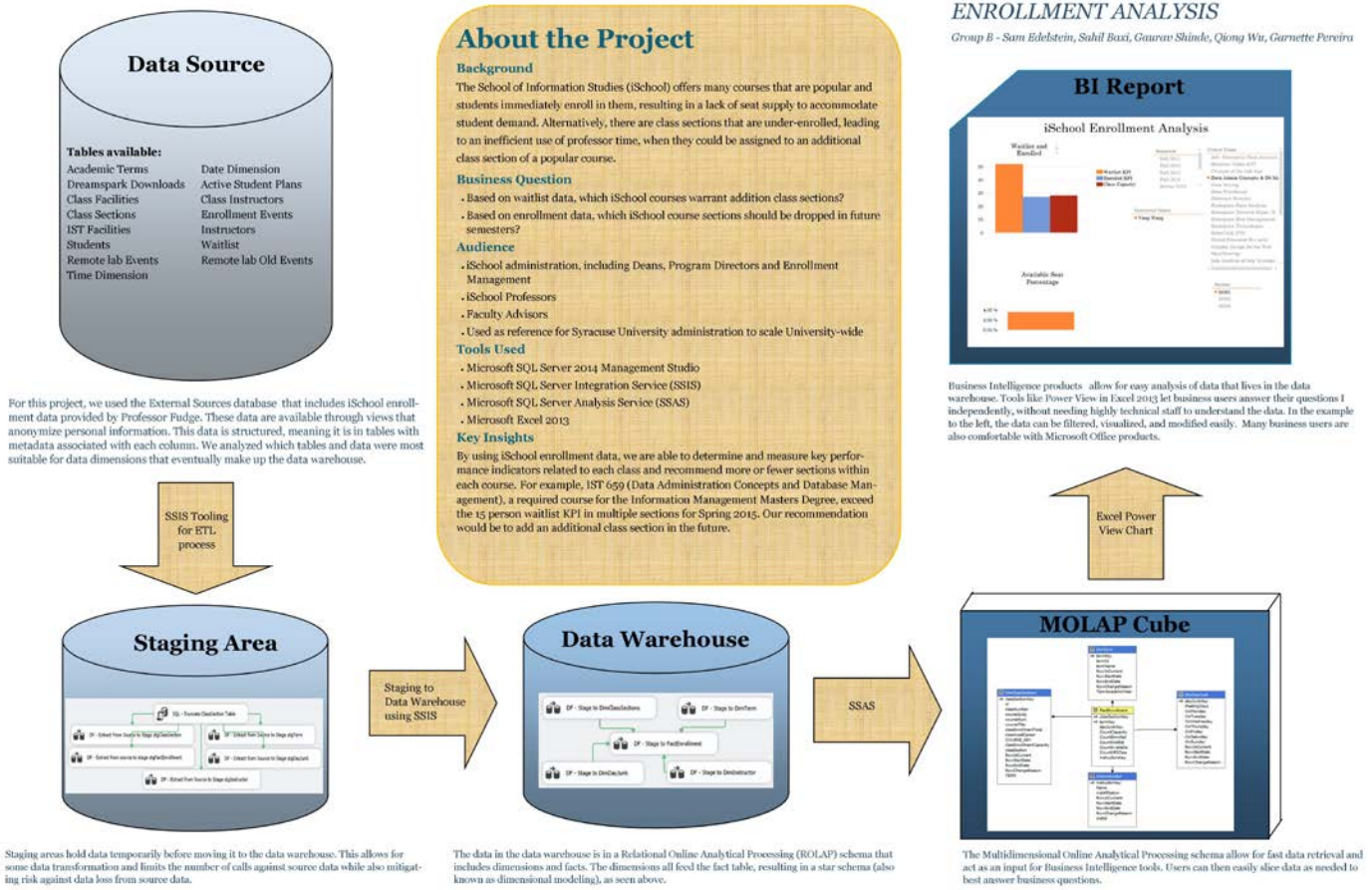
Analysis Services MOLAP Database

All files related to Analysis Service and the MOLAP database can be found us the ist722_siedelst database.

BI Front End

Excel 2013 was used, along with PowerPivot and Power View to serve as front end Business Intelligence. The Excel file is attached, and titled Visualize.xlsx.

Poster



See “Group B Poster-Sahil.pdf” for a more detailed version of the poster

Appendix – Attached Files

- Project-High-Level-Worksheet.xlsx
- Detailed-Dimensional-Modeling-V6.xlsm
- Visualize.xlsx
- Group B Poster.pdf
- DimensionQueries.sql
- Group B Project.sln
- SSAS – found in IST722_siedelst database