

Term Project

CST 4704

Professor Gonzalez

Date: 11/05/20

Steve Carchi, Phurba Sherpa

Contents

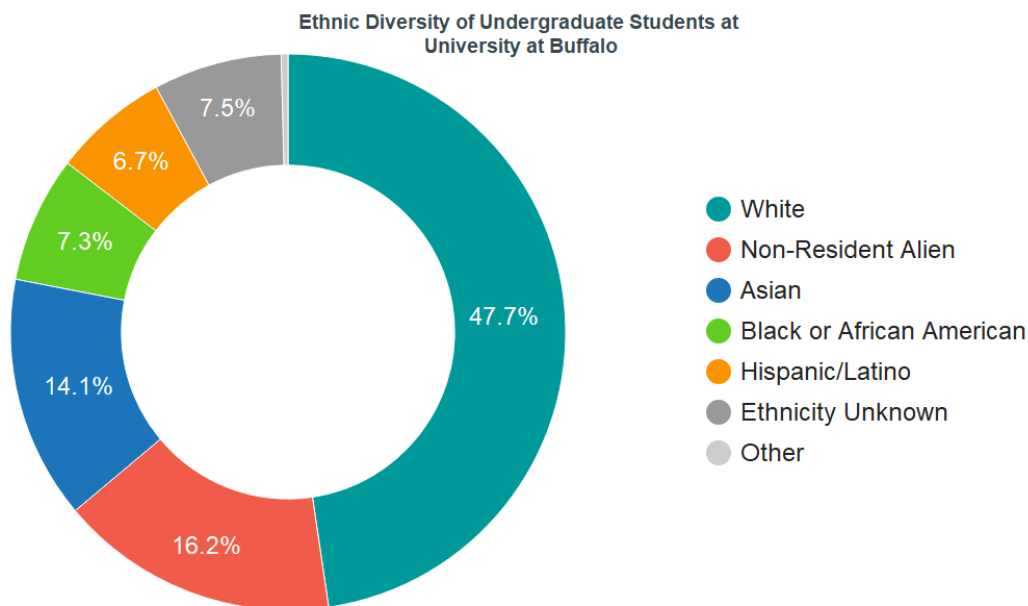
Introduction.....	2
Business Case Assessment.....	3-4
Project Scope.....	5
Database Analysis & Design.....	6-8
..... Enrollment Data Mart Schema.....	9
..... Data Software Selection.....	10
ETL Tool Selection.....	11
Data Modeling Tool Selection.....	12
Technical Evaluation.....	13
Cost Evaluation.....	14
Risk Assessment.....	15
Sources.....	16

Introduction

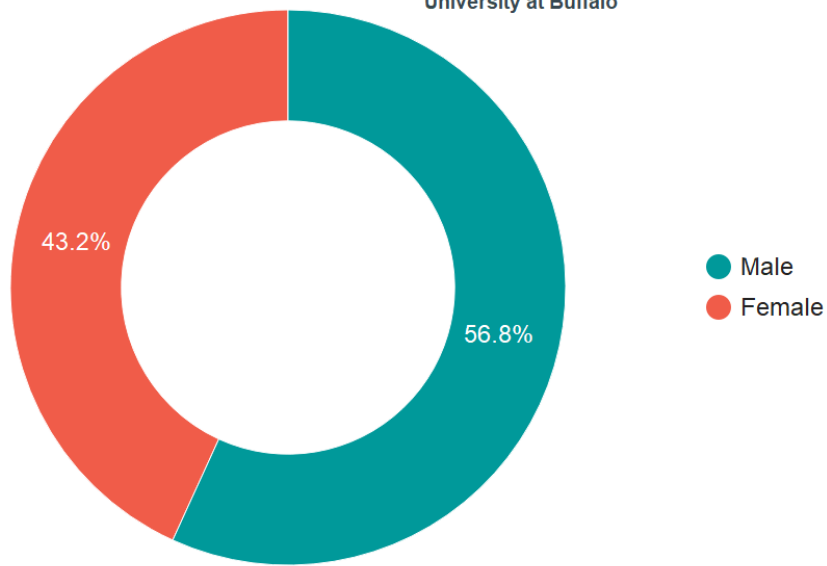
University at buffalo was referred to the ABC company to create an enterprise data warehouse. It is a public university located in buffalo New York. The university was founded in 1846. It was originally started as a private medical college. The University at Buffalo has 140+ undergraduate options and 70+ minors, and more than 300 graduate, professional and certificate degrees. It has an acceptance rate of 57% and has a graduation rate of 73%. The university is affordable compared to other universities with average cost after financial aid being around 17K. The university has a diverse group of population with the majority being white population, then Asian, Black and Hispanic. The university at buffalo has an endowment of 795 million US dollar and they are comfortable in spending towards creating a more efficient data warehouse.

Business Case Assessment

The University at Buffalo wants to implement an Enterprise Data Warehouse. The data warehouse would be used to see the information of all student, faculty and staff in a single view. The current database system they have in place is insufficient due to the growing staff, faculty, and student in the university and the problem to view the data needed. Their information is contained in a separate database by each university department, and are interested in creating a new data warehouse. The university has plans to use the new data warehouse to make better business decisions about the university and plans to use the reports to improve the quality of their campus and education. Below are some statistics about the university.

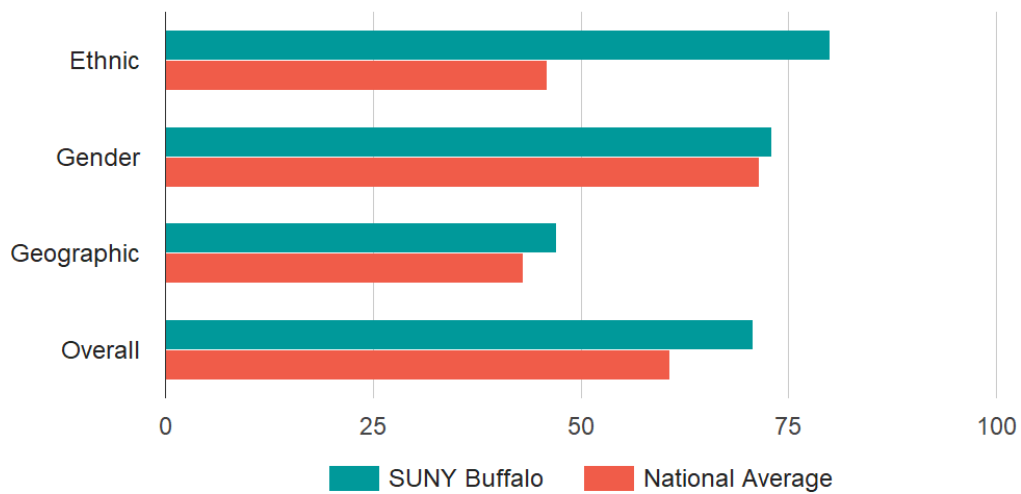


Male/Female Breakdown of Undergraduate Students at University at Buffalo



collegefactual.com

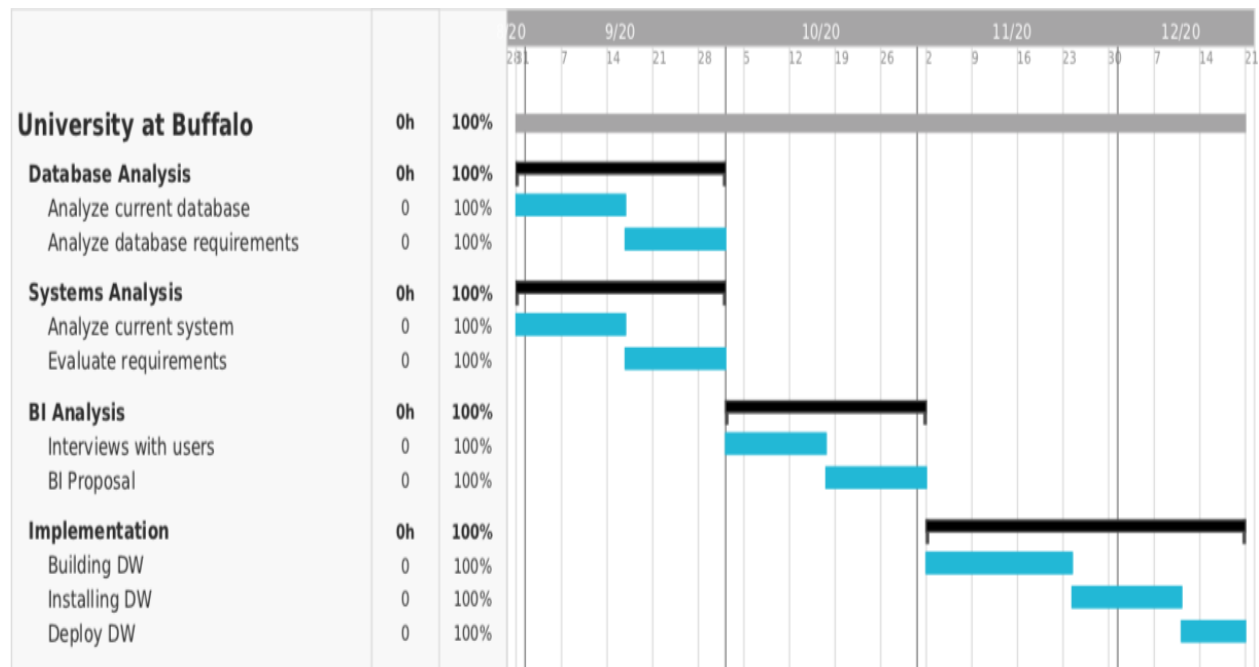
University at Buffalo
Overall Diversity Score



collegefactual.com

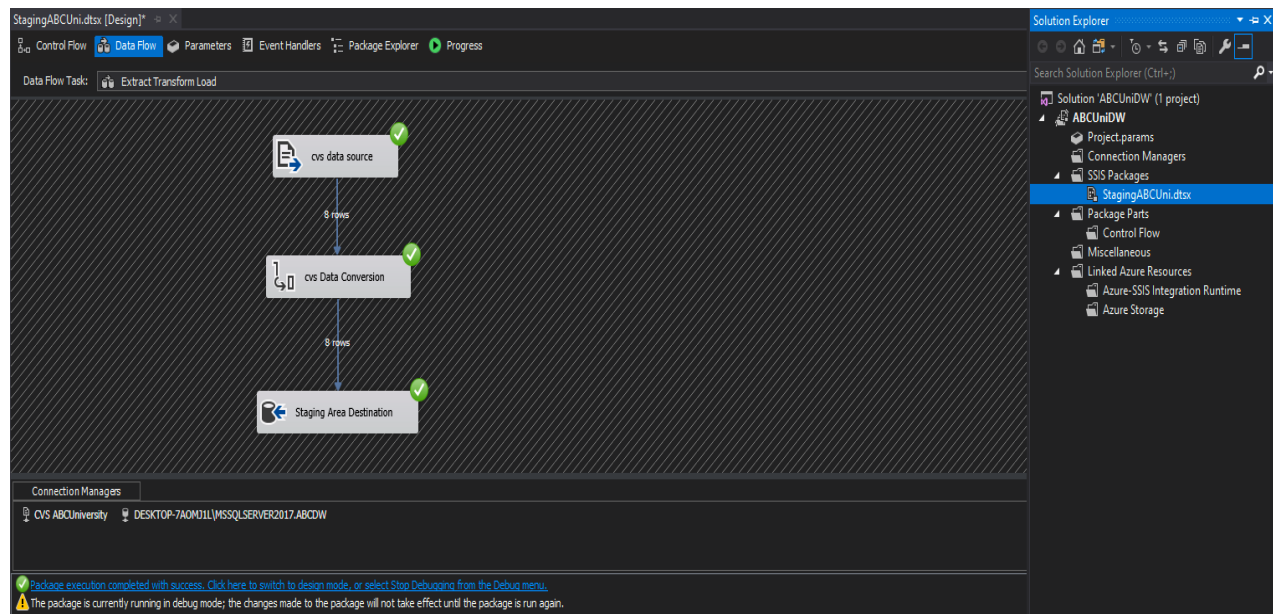
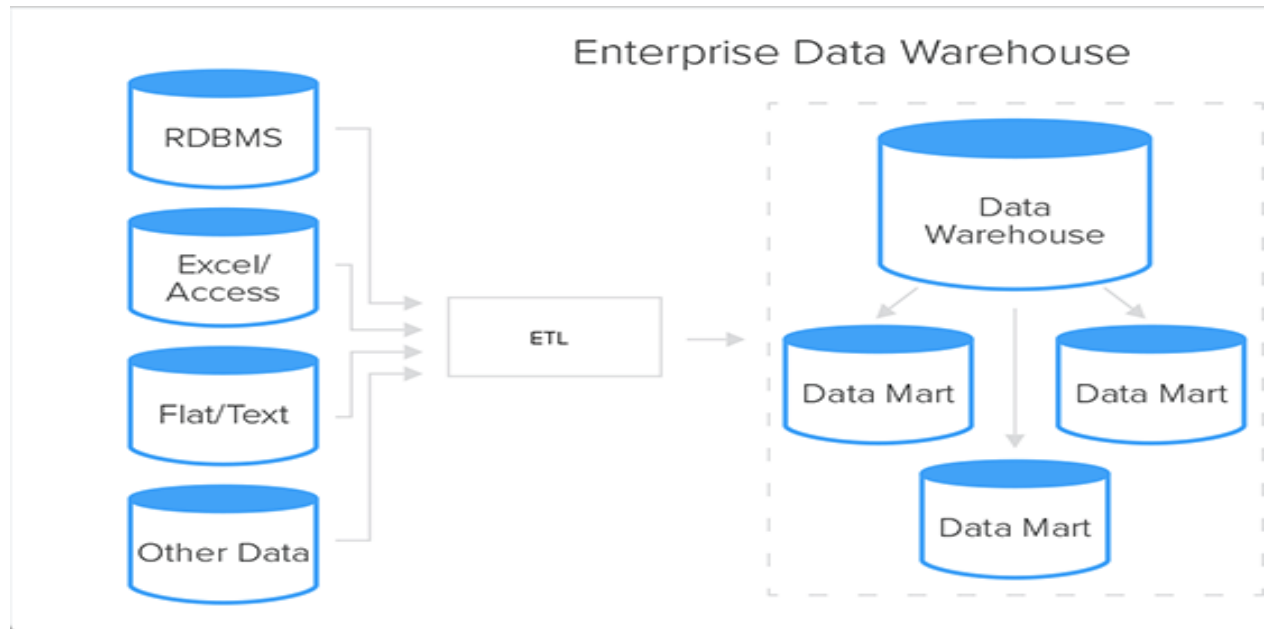
Project Scope

In order to create an Enterprise data warehouse ABC Company has recommended that the school would need a data analyst, project manager, system analyst and a BI Solution Analyst. With Project manager being the leader in the project, his overall responsibility is initiation, planning, design, executing and closure of the project. System analyst will gather information, requirement and determine software that is best fit for the school. Data analyst will help in managing data, number and other information about the school. After the data and information are collected BI Analyst will observe the information provide and will come up with better proposals and solution for the school. With the use of Gantt chart the team decide to create the timeline of the project from the start of the project to the end.



Database Analysis & Design

The analysis concluded that the OLTP data warehouse would be best if it was modified into an OLAP database warehouse by putting its data through the ETL process so it can transformed instead of having to create a new one. Currently the various departments have their own databases. Even though each individual database department uses the same software, each database is designed to meet its own specific needs for each department based on how it queried and shown. Once the new data warehouse is modified and operational the separate databases based on each department will no longer exist but now it will work together as one. This is will allow for a better analytical performance. For the general schema of the data warehouse it will use a star schema since it is easier to implement and for better performance. Our team will also implement several data marts for each department involved in the university such as faculty, bursar, enrollment, etc. A data mart is a simple selection of the data warehouse that delivers a single functional data set, it is basically a smaller subset of data. This will provide easier access for the university and data retrieval from different departments from the university.



Object Explorer

Connect + [Server Icon] [Database Icon] [Query Icon] [Tools Icon] [Help Icon]

DESKTOP-7AOM11L\MSSQLSERVER2017

- Databases
 - System Databases
 - Database Snapshots
 - ABCOW
 - Database Diagrams
 - Tables
 - System Tables
 - dbo.sysdiagrams
 - FileTables
 - External Tables
 - Graph Tables
 - dbo.DimCourse
 - dbo.DimDate
 - dbo.DimLocation
 - dbo.DimStudent
 - dbo.FactEnrollment
 - dbo.StageAreaStudent
 - Views
 - External Resources
 - Synonyms
 - Programmability
 - Service Broker
 - Storage
 - Security
- ABCUnDW
 - AdventureWorks2017
 - AdventureWorksDW2017
 - BankprocurementDW
 - CourseProj

SQL Query (3.sql - D:\AOM11L\stave (66))

```
/****** Script for SelectTopRows command from SSMS ******/
SELECT TOP (1000) [StageAreaKey]
, [Cost]
, [SessionID]
, [CourseID]
, [Course_Section_Num]
, [Start_Date_Time]
, [Course_Location]
, [Capacity]
, [Stud_enrollDate]
, [Grades]
, [StudentID]
, [Stud_First_Name]
, [Stud_Last_Name]
, [Stud_Address]
, [Stud_Registration_Date]
, [StudCountry]
, [Stud_Zip]
, [Stud_City]
, [Stud_State]
FROM [ABCOW].[dbo].[StageAreaStudent]
```

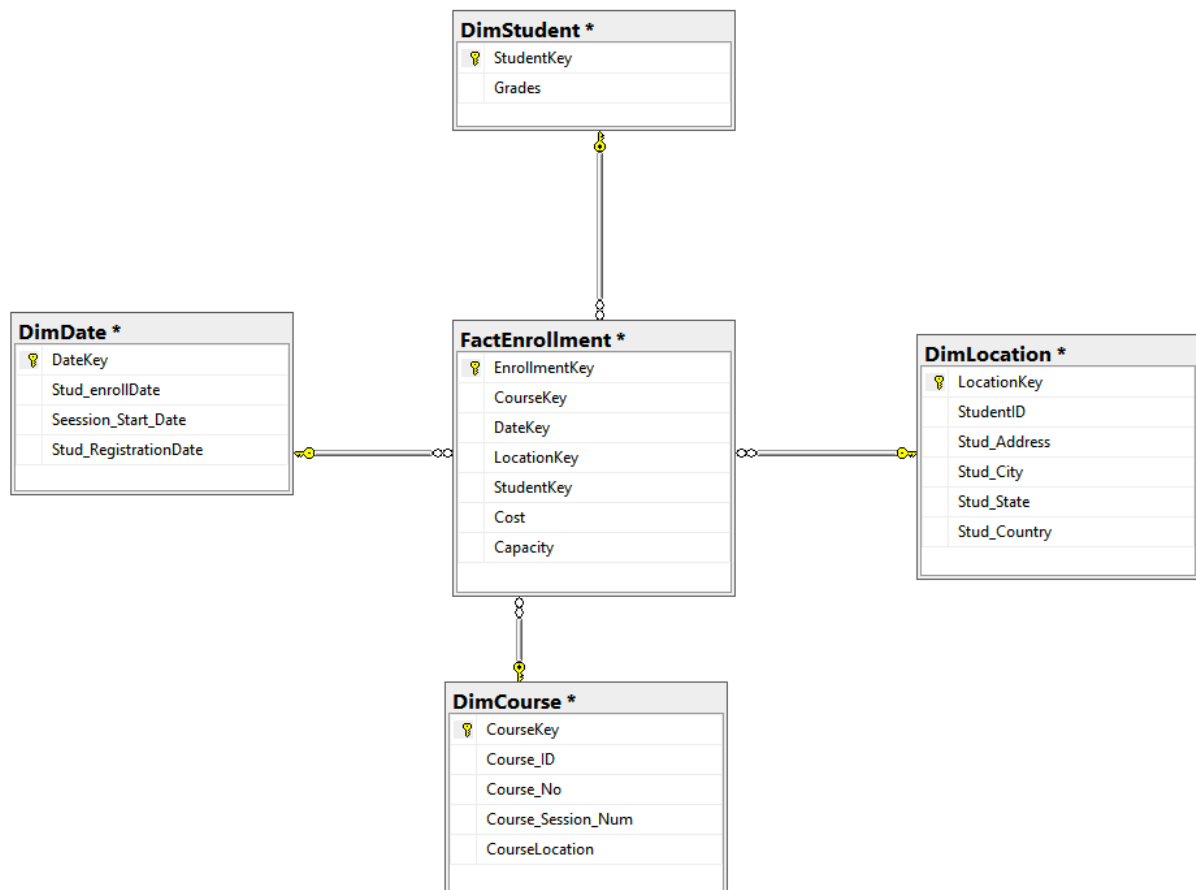
100 %

Results Messages

	StageAreaKey	Cost	SessionID	CourseID	Course_Section_Num	Start_Date_Time	Course_Location	Capacity	Stud_enrollDate	Grades	StudentID	Stud_First_Name	Stud_Last_Name	Stud_Address	Stud_Registration_Date	StudCountry	Stud_Zip	Stud_City	Stud_State
1	1	1195.00	81	20	2	2007-01-02	L210	15	2007-01-30	B	102	Fred	Crocto	101-09 120th St.	2007-01-22	US	70024	Rt Lee	NJ
2	2	1195.00	86	25	2	2007-01-02	L210	15	2007-01-30	A	103	J	Landy	7435 Boulevard East #45	2007-01-22	US	7047	North Bergen	NJ
3	3	1195.00	89	25	5	2007-01-02	L509	25	2007-01-30	NULL	104	Latia	Enson	144-61 87th Ave	2007-01-22	US	10005	New York	NY
4	4	1195.00	92	25	8	2007-01-02	L509	25	2007-01-30	A	105	Angel	Moskowitz	320 John St.	2007-01-22	US	10015	New York	NY
5	5	1195.00	104	330	1	2007-01-02	L511	25	2007-02-10	NULL	163	Nicole	Glen	4301 N Ocean #103	2007-02-02	US	10025	New York	NY
6	6	1195.00	119	142	1	2007-01-02	L211	25	2007-02-16	C	223	Frank	Pace	13 Burlington Dr.	2007-02-08	US	10035	New York	NY
7	7	1195.00	155	122	4	2007-01-01	L210	15	2007-02-16	NULL	399	Jeny	Abdou	460 15th St. #4	2007-02-23	US	11419	Richmond Hill	NY
8	8	0.00	0	0	0	1899-12-30		0	1899-12-30		0				1899-12-30		11435	Jamaica	NY

Database Analysis & Design Cont.

Enrollment Data Mart Schema



Database Analysis & Design Cont.

Data Software Selection

The current software that the university is running is the Microsoft SQL server 2008. Through analyzing the software, the ABC Company recommend the university to upgrade their software to Microsoft SQL server 2019 enterprise edition. With more high security, Fast Performance and advanced workload performance, it help's support big platform of data such as the University. With the core-based licensing, it helps perform more advanced computing solution weather it's used in on-premises or the cloud. It is versatile with providing options to run on any platform and language.

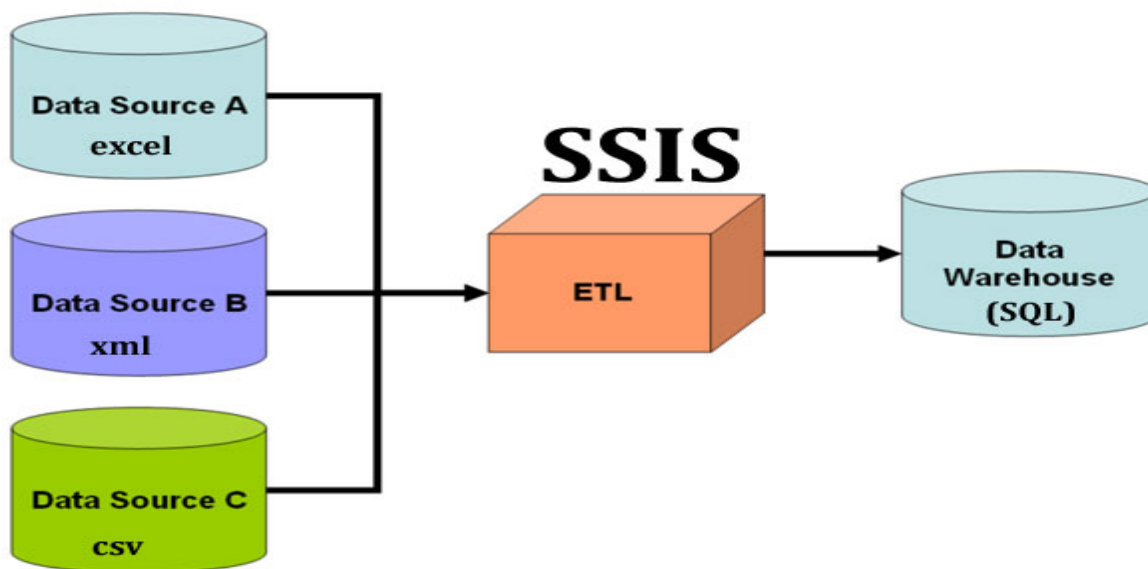
What you'll love about SQL Server 2019

SQL Server 2019 brings innovative security and compliance features, industry-leading performance, mission-critical availability, and advanced analytics to all your data workloads, now with support for big data built-in.

				
Intelligence over any data	Choice of language and platform	Industry-leading performance	Advanced security features	Make faster, better decisions
<p>SQL Server is a hub for data integration. Deliver transformational insights over structured and unstructured data with the power of SQL Server and Spark.</p>	<p>Build modern applications with innovative features using your choice of language and platform. Now on Windows, Linux, and containers.</p>	<p>Take advantage of breakthrough scalability, performance, and availability for mission-critical, intelligent applications, data warehouses, and data lakes.</p>	<p>Protect data at rest and in use. SQL Server has been the least vulnerable database over the last 8 years in the NIST vulnerabilities database.</p>	<p>Power BI Report Server gives your users access to rich, interactive Power BI reports, and the enterprise reporting capabilities of SQL Server Reporting Services.</p>

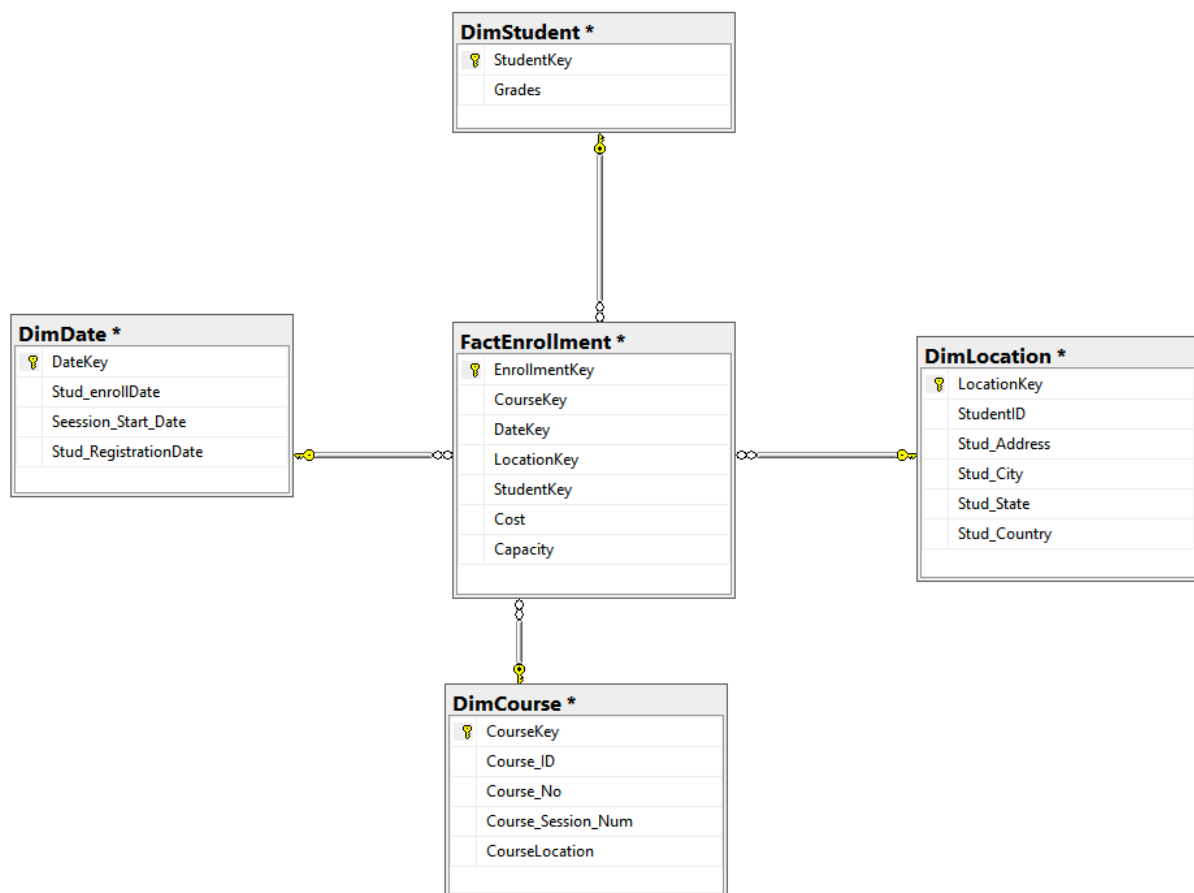
ETL Tool Selection

ETL Tool is used for extracting transferring and loading, it helps in managing raw data's and put them in order and use as information when needed. After doing some research, ABC Company recommended that the university should USE SSIS (SQL Server Integration Services) which is part of the Microsoft SQL Server as their ETL Tool. Since the university uses the Microsoft SQL Server Software it will be easier to use the SSIS because it is a part of their software. SSIS provides with important value in organization such as providing good support, connects to many different important sources, easy to customize the projects and packages, easy to configure if there is any problem.



Data Modeling Tool Selection

A proper data modeling tool needs to go along with the data warehouse and ETL software. The team has decided that the Data modeler that SQL server 2019 provides will be used for the university data warehouse. This data modeler is convenient and useful to use. It is already integrated with SQL server so there is no need to waste more money on another data modeling tool. Below an image is shown on how it works which is how the schema for the enrollment data mart was created.



Technical Evaluation

Microsoft SQL Server 2019 Enterprise

- Provides with advanced Security
- Gives an organization a high performance and scalability
- Versatility to run on any platform and many programming languages.
- Fast performance when dealing with advanced workload.
- Runs efficiently on both on-premises and cloud.

SSIS (SQL Server Integration Services)

- It can be connected to many different sources
- Project and packed can be easily modified
- Part of Microsoft SQL Server Software
- Has a built in data Source and Data Connector
- Easy to use and can be used through visual studio

Cost Evaluation

Microsoft SQL Server 2019 Enterprise - \$13,748 (2 core pack), 7,128 (per core)

- Enterprise Edition supports up to 524 PB (524,00 TB) of data
- Unlimited cores of CPU
- Software Assurance Benefits
- Unlimited memory
- Unlimited virtualization benefits
- Petabyte scale data warehousing
- Scalability performance

SSIS (SQL Server Integration Services) – comes with Microsoft SQL Server 2019 Enterprise

- Can handle data from heterogenous data sources at a same package
- Consumes data
- Provides transformation functionality
- Easier to maintain and package configuration
- Integrated with SQL Server
- Usage of the SQL Server Destination which allows to load data into SQL faster
- Better for complex transformations, multi-step operations, aggregating data from different data sources or types, and structured exception handling.
- Data can be loaded in parallel to many varied destinations

Microsoft SQL Server 2019 Data Modeler – comes with Microsoft SQL Server 2019 Enterprise

Risk Assessment

Level of Risk			
Variable	Green (Low)	Yellow (Medium)	Red (High)
Technology	Experienced with mature technology	Minimal experience with technology	New technology, little experience
Complexity	Simple, minimal workflow impact	Moderate, some workflow impact	Mission critical, will require extensive reengineering
Integration	Stand-alone, no integration	Limited integration required	Extensive integration required
Organization	Solid internal support	Supportive to a large extent	Little internal support
Project team	Business experience, business-driven, talented, great attitude	Some business experience, business-driven, talented, fair attitude	No business experience, only technology-driven, limited talent, bad attitude
Financial Investment	Possible ROI within a very short time	Possible ROI within a moderate time frame	Possible ROI after a few years

Sources

1. <https://www.sarjen.com/ssis-advantages-disadvantages/>
2. <https://www.microsoft.com/en-us/sql-server/sql-server-2019-pricing>
3. <https://study.com/academy/lesson/what-is-a-data-mart-design-types-example.html>
4. <https://www.collegefactual.com/colleges/university-at-buffalo/student-life/diversity/#characteristics-ethnic-diversity>
5. <https://www.xplenty.com/blog/ssis-etl-tool/>
6. file:///Users/phurbasherpa/Downloads/SQL_Server_2017_Licensing_guide.pdf
7. <https://www.xplenty.com/blog/ssis-etl-tool/>
8. <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiLrcbOtdXtAhVxRN8KHQpDDwIQFjAAegQIBhAC&url=https%3A%2F%2Fdownload.microsoft.com%2Fdownload%2F2%2F9%2F0%2F290f991b-5971-42e5-bb2c-81f700622b2e%2FSQL%2520Server%25202019%2520Editions%2520Datasheet.pdf&usg=AOvVaw3MbY9UZSpWR0vP6yj9X7m4>
9. <https://redmondmag.com/articles/2019/11/08/sql-server-2019-licensing.aspx>