Datawarehouse and Business Intelligence for Healthcare Provider Data

Objective/ Scope of Document

Medicare is the most important sector for any government to gauge their performance. However, it can be tedious for a government official to see all the metrics at a glance like cost, revenue, expenses etc. Our objective for this project would be to create a tool which will enable the user to slice and dice this data as per convenience. This includes creating dashboards and a cube(yet to decide based on further scope of project). Upon completion of the project, following objectives are expected to be achieved:

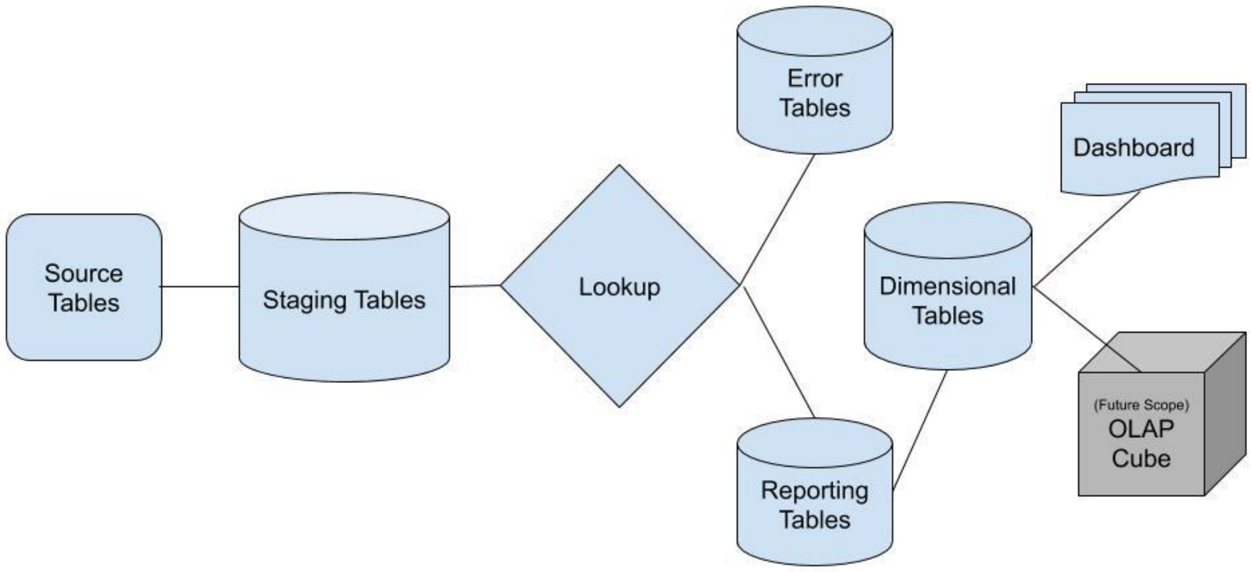
- To analyze the Provider's covered charges, Total payments to all Providers and Medicare payments at hospital level and also based on the type of clinical condition (diagnosis) and the procedure furnished by the hospital during the stay of the inpatients
- To analyze the Provider's submitted charges, Medicare and Beneficiaries' payments to the provider at hospital level and also based on Medicare's APC (Ambulatory Payment Classification) description for outpatients
- To connect the hospital dataset with census data based on City and State level and finding if there is sufficient availability of Hospitals for the people living in that area
- As an individual interested in knowing the cost of setting up a hospital. I want to know the cost break-up to set-up a new hospital in a given location in the US.
- As an investor I want to know if there is a difference in the cost and the break-up of cost for a given provider. I would be interested in knowing the difference in cost for the different healthcare provided by a given hopital.
- An exhaustive data warehouse, easily available, clean and understood
- Dashboard, with easily interpretable information and interactive
- An OLAP cube, with slicing and dicing by geography, type of hospitals, time, titles etc.
- Create a sentiment on hospital financials on their performance state wise to gauge performances

Stakeholders:
Anyone who wants to understand the Healthcare performance at a glance and make decisions regarding the Hospitals.

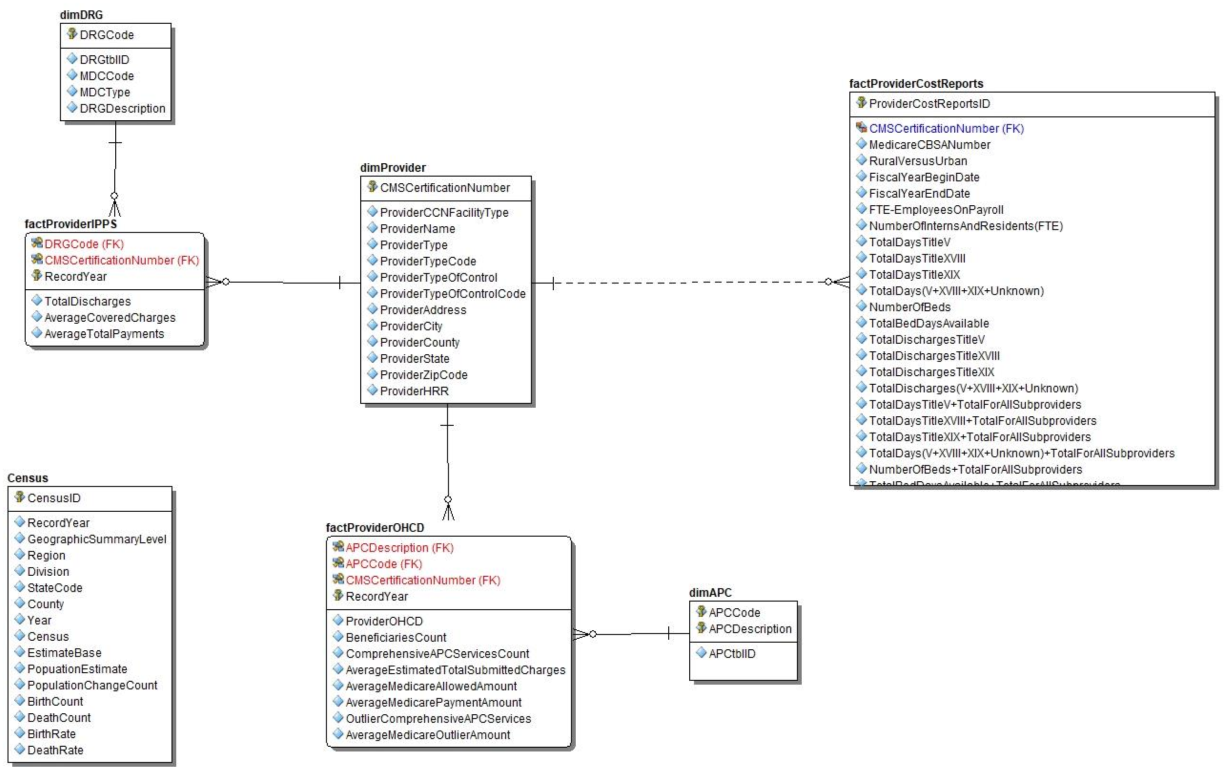
Tools

1. SQL Server Integration Services (SSIS)
2. SQL Server Management Studio (SSMS)
3. Azure SQL Server
4. Azure SQL Database
5. Azure DevOps
6. Power BI Desktop

Overview of the data model



ER diagram


dimDRG
DRGCode
DRGtblID
MDCCode
MDCType
DRGDescription
factProviderIPPS
DRGCode (FK)
CMSCertificationNumber (FK)
RecordYear
TotalDischarges
AverageCoveredCharges
AverageTotalPayments
dimProvider
CMSCertificationNumber
ProviderCCNFacilityType
ProviderName
ProviderType
ProviderTypeCode
ProviderTypeOfControl
ProviderTypeOfControlCode
ProviderAddress
ProviderCity
ProviderCounty
ProviderState
ProviderZipCode
ProviderHRR
factProviderCostReports
ProviderCostReportsID
CMSCertificationNumber (FK)
MedicareCBSANumber
RuralVersusUrban
FiscalYearBeginDate
FiscalYearEndDate
FTE-EmployeesOnPayroll
NumberOfInternsAndResidents(FTE)
TotalDaysTitleV
TotalDaysTitleXVIII
TotalDaysTitleXIX
TotalDays(V+XVIII+XIX+Unknown)
NumberOfBeds
TotalBedDaysAvailable
TotalDischargesTitleV
TotalDischargesTitleXVIII
TotalDischargesTitleXIX
TotalDischarges(V+XVIII+XIX+Unknown)
TotalDaysTitleV+TotalForAllSubproviders
TotalDaysTitleXVIII+TotalForAllSubproviders
TotalDaysTitleXIX+TotalForAllSubproviders
TotalDays(V+XVIII+XIX+Unknown)+TotalForAllSubproviders
NumberOfBeds+TotalForAllSubproviders
Census
CensusID
RecordYear
GeographicSummaryLevel
Region
Division
StateCode
County
Year
Census
EstimateBase
PopuationEstimate
PopulationChangeCount
BirthCount
DeathCount
BirthRate
DeathRate
factProviderOHCD
APCDescription (FK)
APCCode (FK)
CMSCertificationNumber (FK)
RecordYear
ProviderOHCD
BeneficiariesCount
ComprehensiveAPCServicesCount
AverageEstimatedTotalSubmittedCharges
AverageMedicareAllowedAmount
AverageMedicarePaymentAmount
OutlierComprehensiveAPCServices
AverageMedicareOutlierAmount
dimAPC
APCCode
APCDescription
APCtblID

Implementation

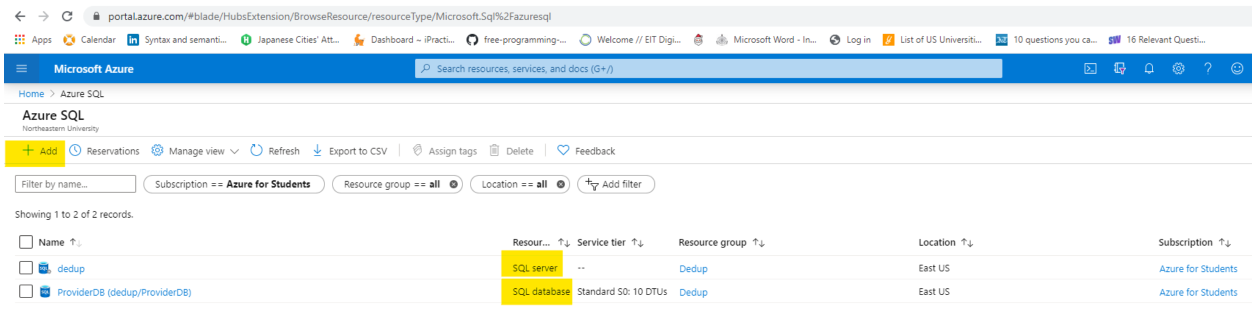
Azure Setup Instructions:

1. Create the Azure account.
   Reference link: https://azure.microsoft.com/en-us/free/students/

2. Go to https://portal.azure.com/ to access the resource creation.

3. To create the Azure Database

Go to Azure SQL server . Click Add



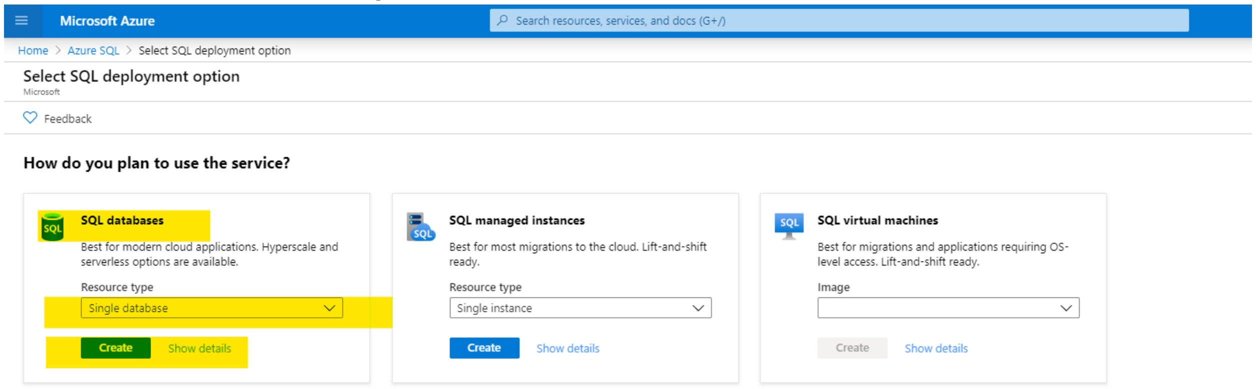
portal.azure.com/#blade/HubsExtension/BrowseResource/resourceType/Microsoft.Sql%2Fazuresql

Azure SQL
Northeastern University

Showing 1 to 2 of 2 records.

Name	Resour...	Service tier	Resource group	Location	Subscription
dedup	SQL server	--	Dedup	East US	Azure for Students
ProviderDB (dedup/ProviderDB)	SQL database	Standard S0: 10 DTUs	Dedup	East US	Azure for Students

Create a SQL database: Single database



Select SQL deployment option
Microsoft

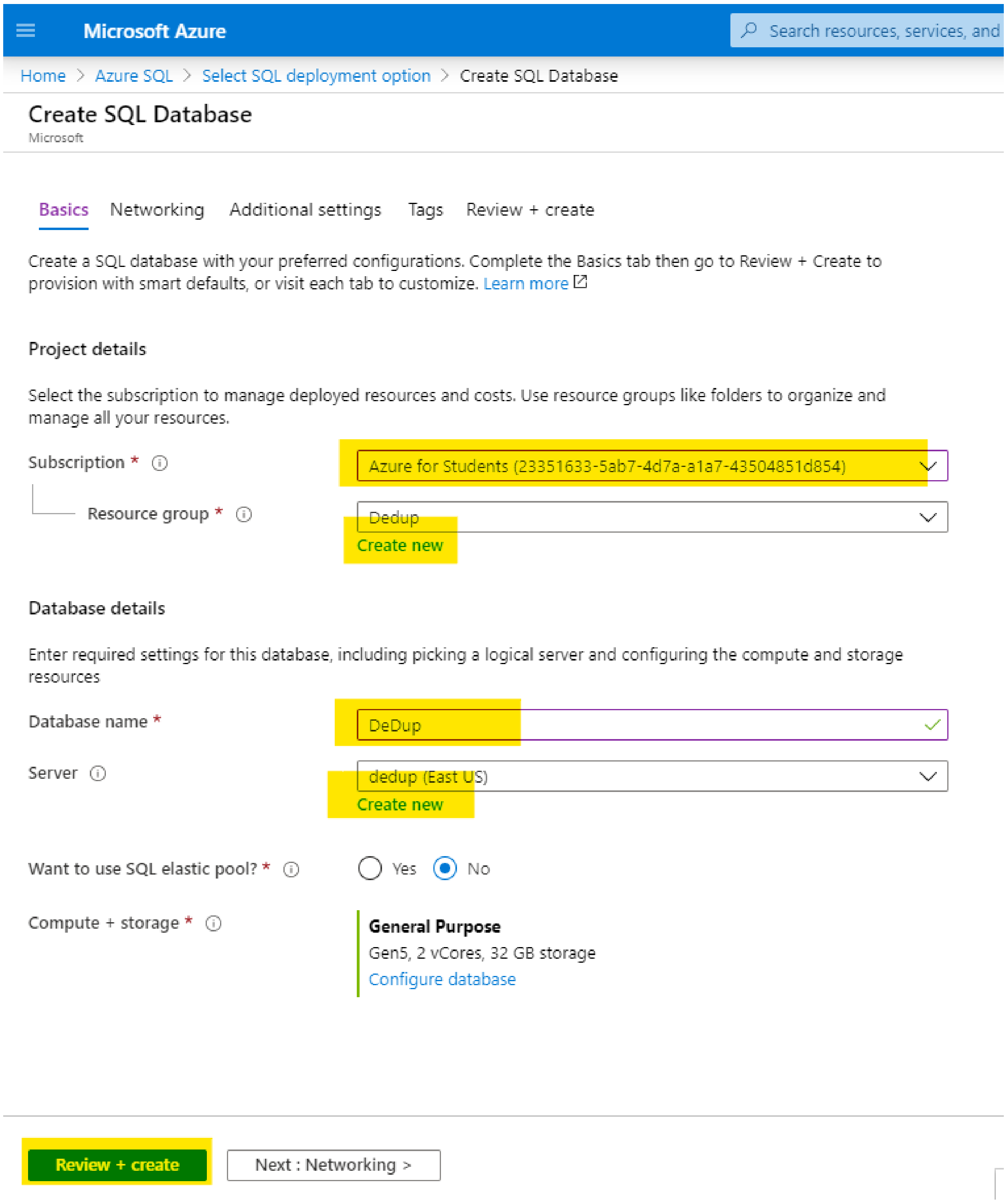
How do you plan to use the service?

SQL databases	SQL managed instances	SQL virtual machines
Best for modern cloud applications. Hyperscale and serverless options are available.	Best for most migrations to the cloud. Lift-and-shift ready.	Best for migrations and applications requiring OS-level access. Lift-and-shift ready.
Resource type: Single database	Resource type: Single instance	Image:

Microsoft Azure
Search resources, services, and docs (G+/)
Home > Azure SQL > Select SQL deployment option
Feedback
Create
Show details
Create
Show details
Create
Show details

Enter the required details to create the database:

Create the Server if it isn’t created:



Microsoft Azure

Search resources, services, and

Home > Azure SQL > Select SQL deployment option > Create SQL Database

Create SQL Database

Microsoft

Basics | Networking | Additional settings | Tags | Review + create

Create a SQL database with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults, or visit each tab to customize. Learn more

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

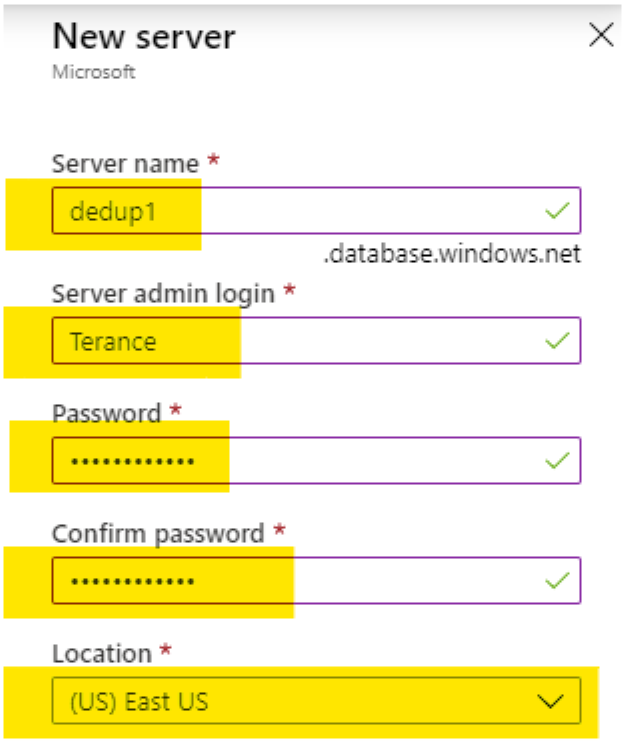
Subscription *	Azure for Students (23351633-5ab7-4d7a-a1a7-43504851d854)
Resource group *	Dedup Create new

Database details

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

Database name *	DeDup
Server	dedup (East US) Create new
Want to use SQL elastic pool? *	Yes [✓ No]
Compute + storage *	**General Purpose** Gen5, 2 vCores, 32 GB storage Configure database

Review + create | Next : Networking >



New server

Microsoft

Server name *
dedup1
.database.windows.net

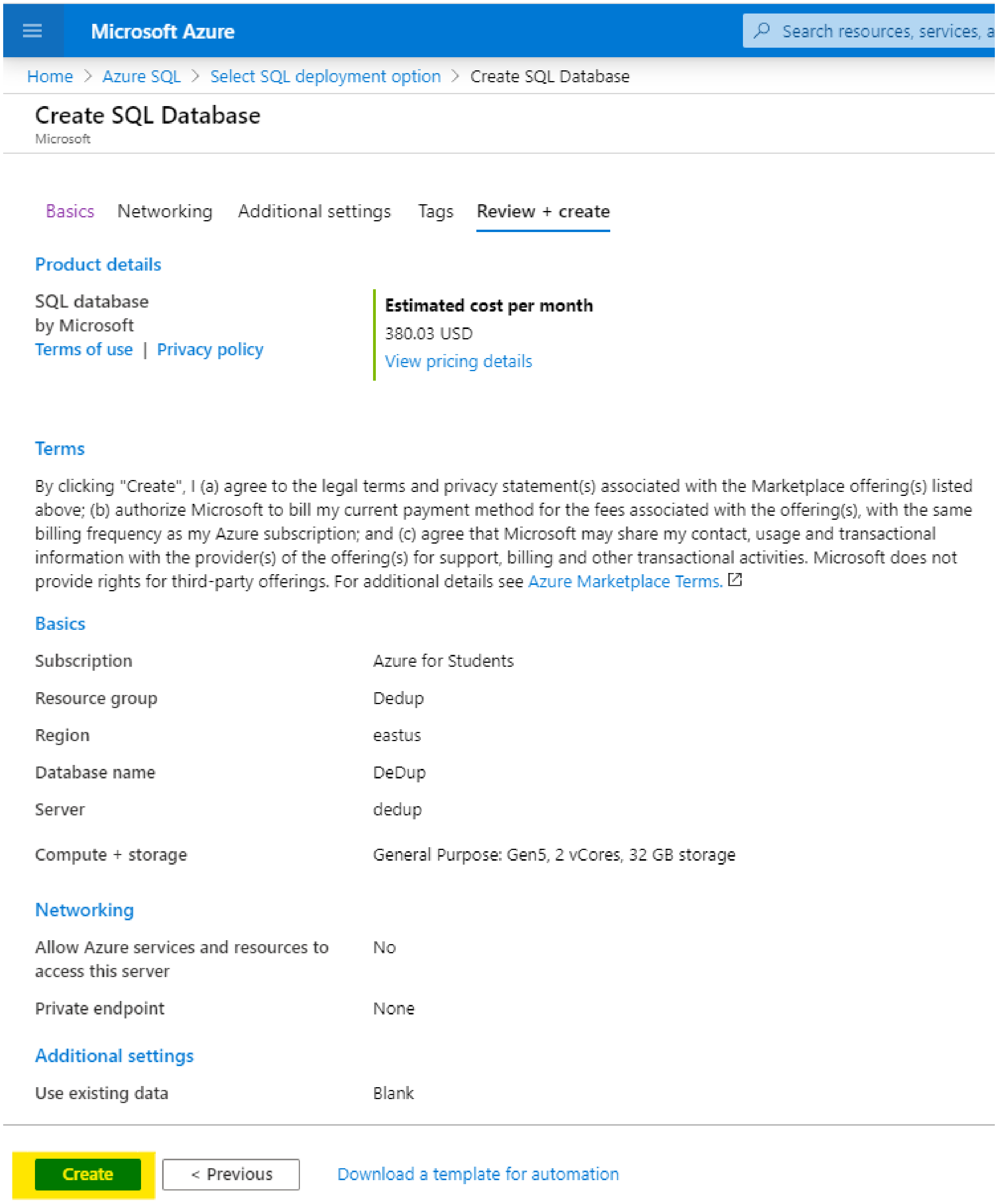
Server admin login *
Terance

Password *
•••••••••••••

Confirm password *
•••••••••••••

Location *
(US) East US

Check the resources and additional resources and create the server.


Microsoft Azure
Search resources, services, a
Home > Azure SQL > Select SQL deployment option > Create SQL Database

Create SQL Database

Microsoft

Basics
Networking
Additional settings
Tags
Review + create

Product details

SQL database
by Microsoft
Terms of use | Privacy policy

Estimated cost per month
380.03 USD
View pricing details

Terms

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. For additional details see Azure Marketplace Terms.

Basics

Subscription	Azure for Students
Resource group	Dedup
Region	eastus
Database name	DeDup
Server	dedup
Compute + storage	General Purpose: Gen5, 2 vCores, 32 GB storage

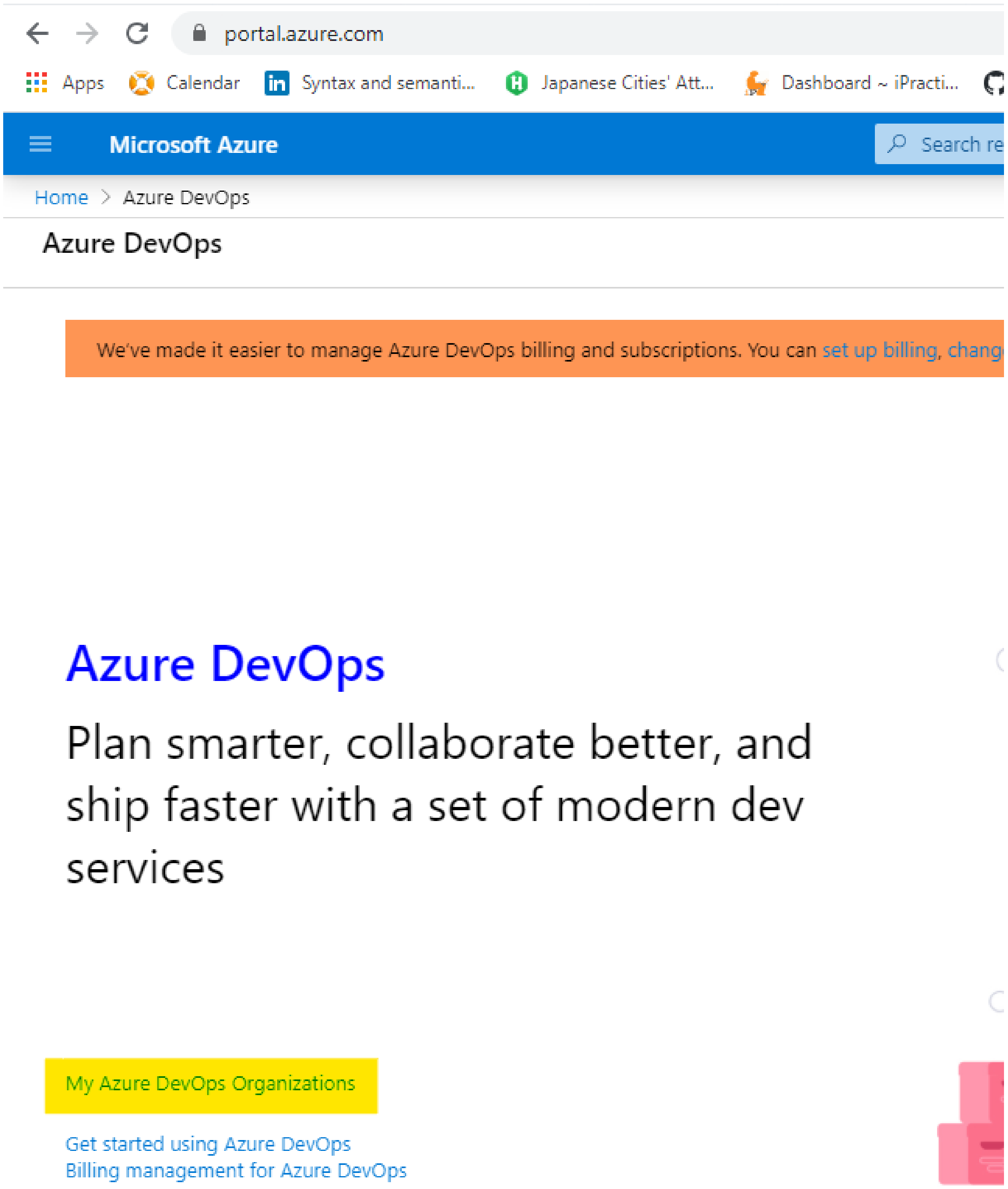
Networking

Allow Azure services and resources to access this server	No
Private endpoint	None

Additional settings

Use existing data	Blank

4. Create Azure DevOps



Azure DevOps

We've made it easier to manage Azure DevOps billing and subscriptions. You can set up billing, chang

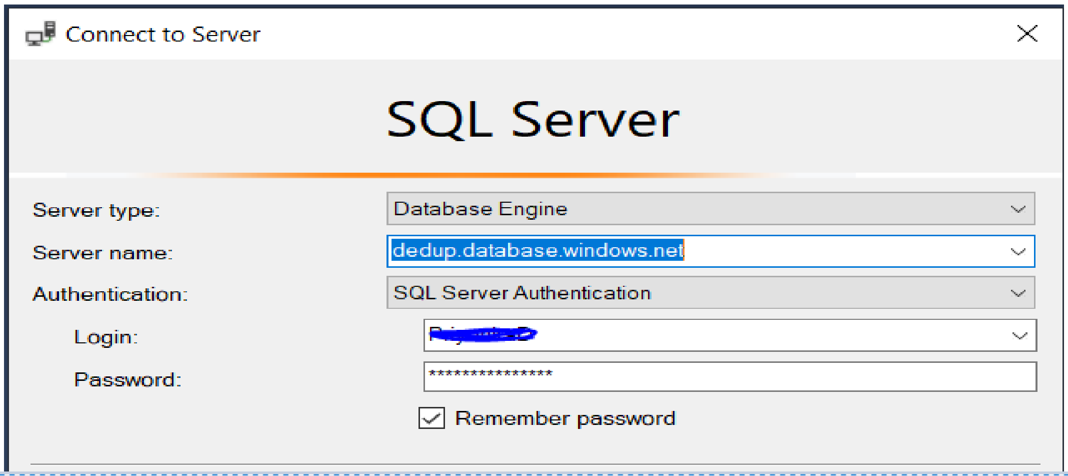
Azure DevOps

Plan smarter, collaborate better, and ship faster with a set of modern dev services

My Azure DevOps Organizations

Get started using Azure DevOps
Billing management for Azure DevOps

1. Download the files
2. Connect to Azure SQL server database



Connect to Server

SQL Server

Server type: Database Engine

Server name: dedup.database.windows.net

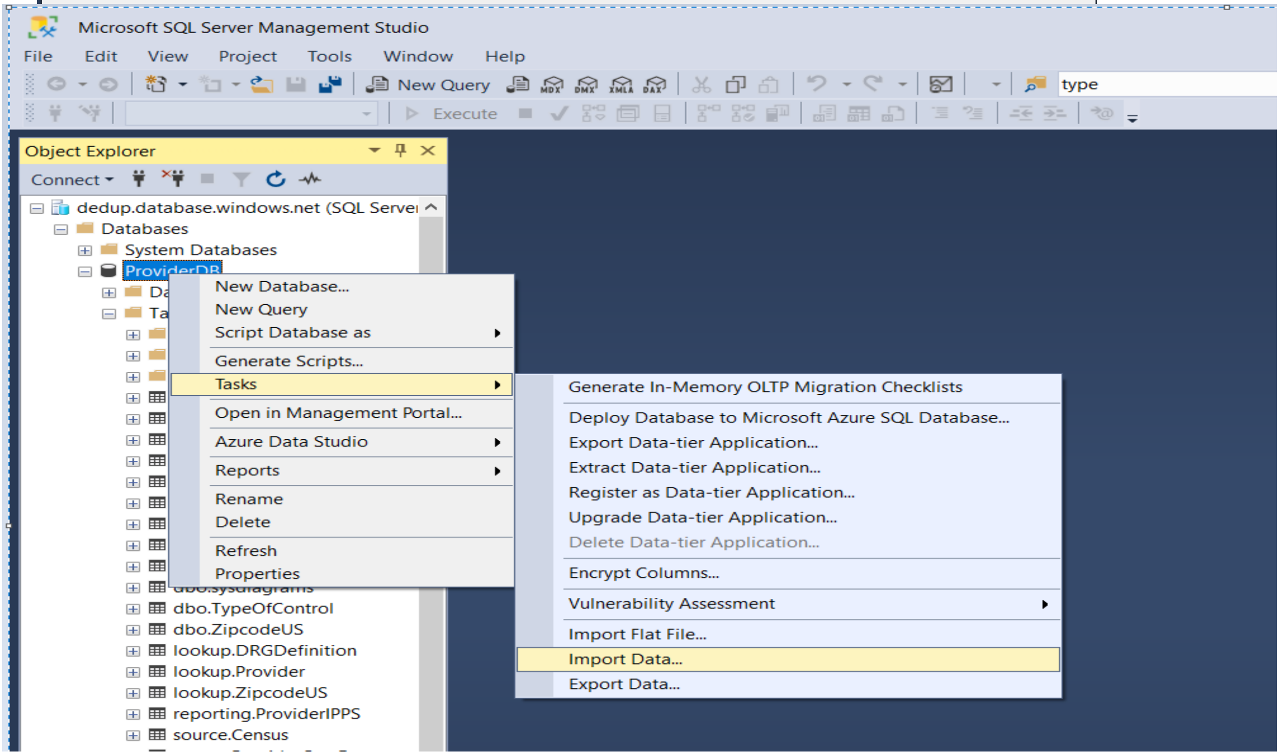
Authentication: SQL Server Authentication

Login: [redacted]

Password: ***************

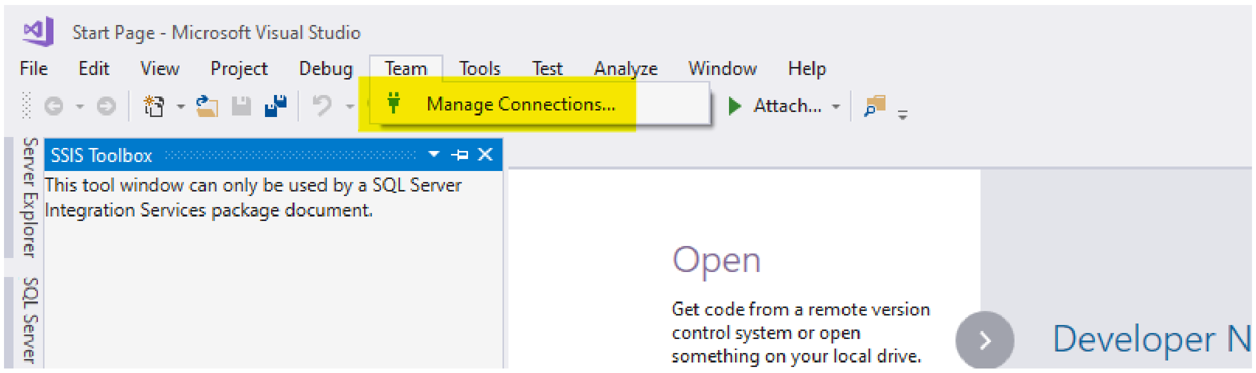
☑ Remember password

3. Import the file in SSMS into the Storage Schema



4. Connect to Azure SSIS

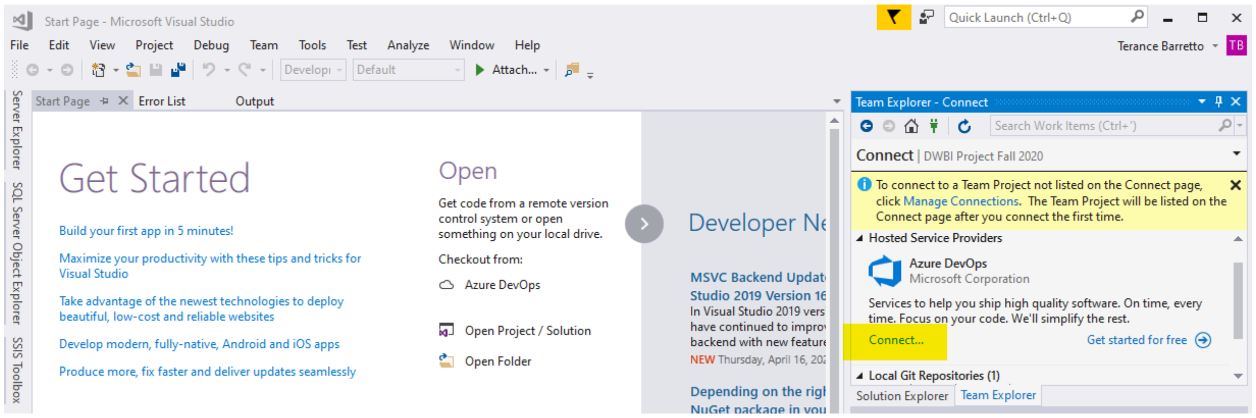
a) Make a connection to the Git Repo created via AzureDevOps



This tool window can only be used by a SQL Server Integration Services package document.

Get code from a remote version control system or open something on your local drive.

b) Click on manage connection to make a connection to the repo



Build your first app in 5 minutes!

Maximize your productivity with these tips and tricks for Visual Studio

Take advantage of the newest technologies to deploy beautiful, low-cost and reliable websites

Develop modern, fully-native, Android and iOS apps

Produce more, fix faster and deliver updates seamlessly

Get code from a remote version control system or open something on your local drive.

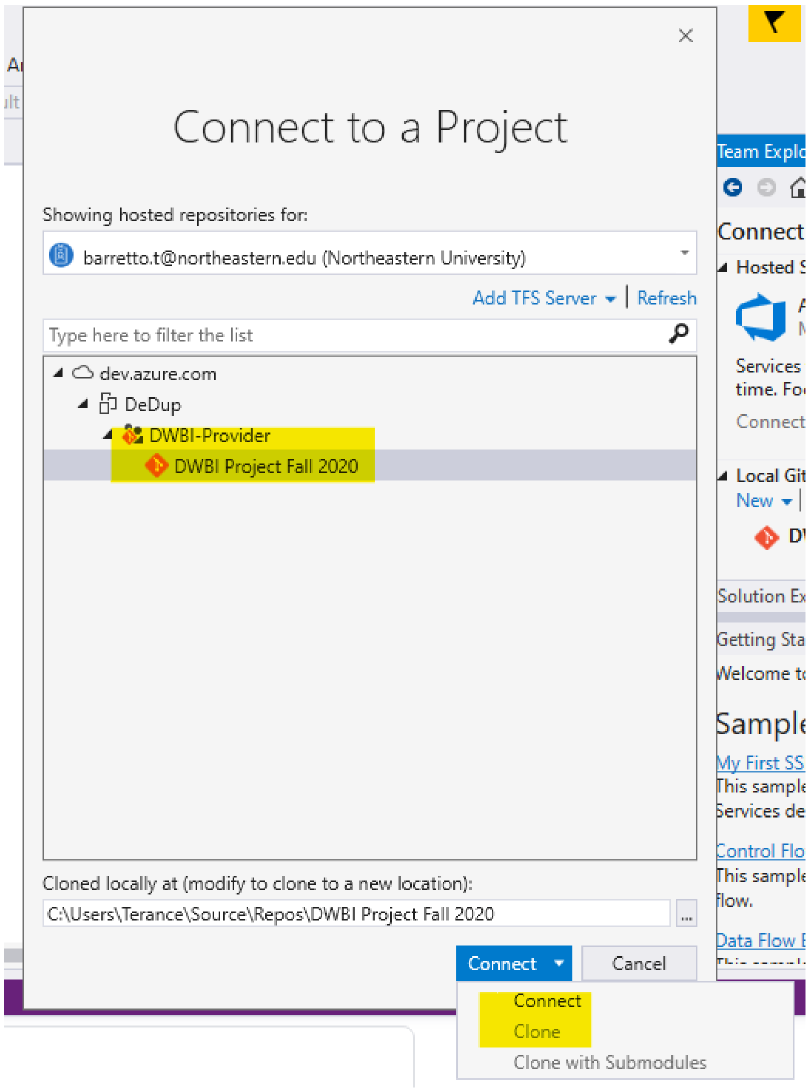
MSVC Backend Updat
Studio 2019 Version 16
In Visual Studio 2019 vers
have continued to impro
backend with new featur
NEW Thursday, April 16, 20

Depending on the rig
NuGet package in you

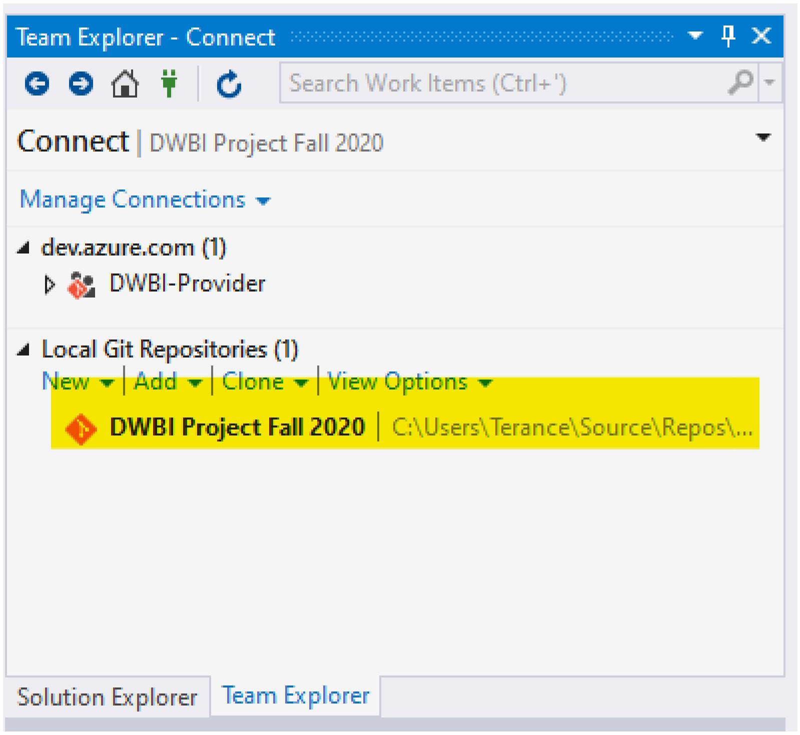
To connect to a Team Project not listed on the Connect page, click Manage Connections. The Team Project will be listed on the Connect page after you connect the first time.

Services to help you ship high quality software. On time, every time. Focus on your code. We'll simplify the rest.

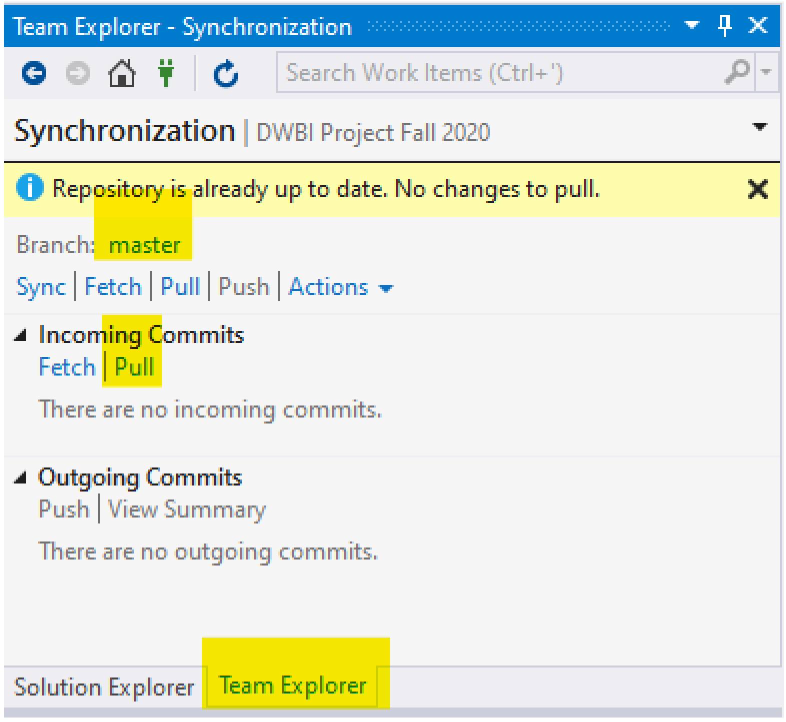
c) Connect to your microsoft account and clone the project repo


Connect to a Project
Showing hosted repositories for:
barretto.t@northeastern.edu (Northeastern University)
Add TFS Server
Refresh
Type here to filter the list
dev.azure.com
DeDup
DWBI-Provider
DWBI Project Fall 2020
Cloned locally at (modify to clone to a new location):
C:\Users\Terance\Source\Repos\DWBI Project Fall 2020
Connect
Cancel
Connect
Clone
Clone with Submodules

d) Go to the cloned repo location, and open the project file.


Team Explorer - Connect
Search Work Items (Ctrl+')
Connect | DWBI Project Fall 2020
Manage Connections
dev.azure.com (1)
DWBI-Provider
Local Git Repositories (1)
New | Add | Clone | View Options
DWBI Project Fall 2020 | C:\Users\Terance\Source\Repos\...
Solution Explorer
Team Explorer

e) Use the "Pull" command to sycn the online repo with the copy on your local machine.

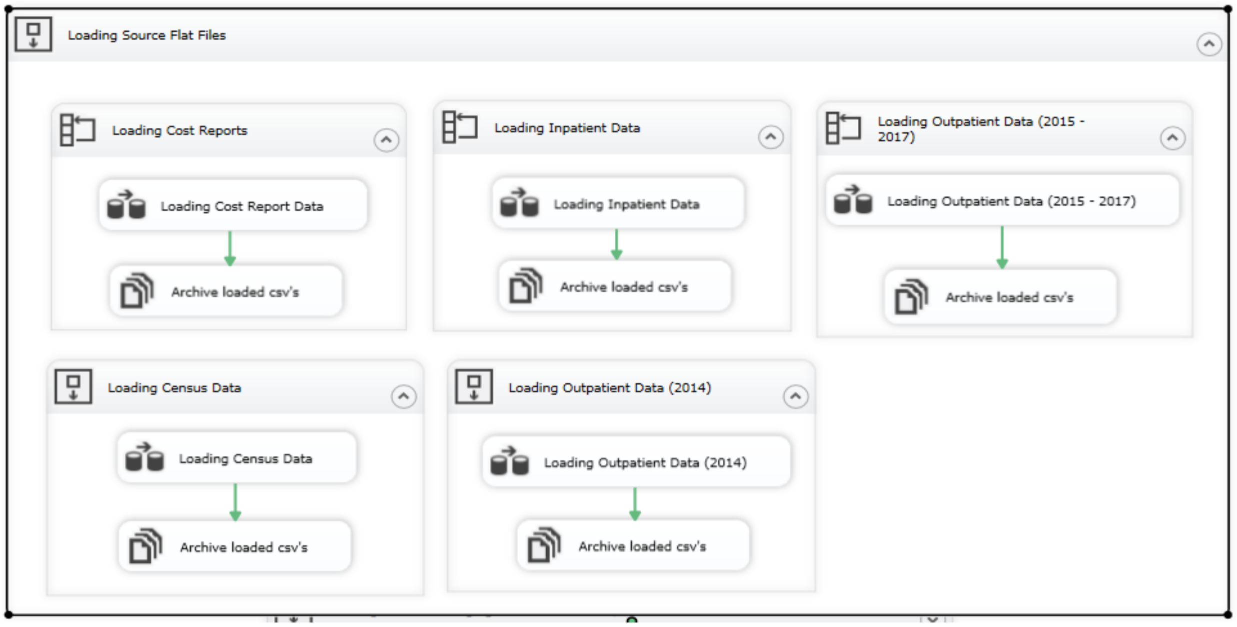

Team Explorer - Synchronization
Search Work Items (Ctrl+')
Synchronization | DWBI Project Fall 2020
Branch: master
Sync | Fetch | Pull | Push | Actions
Incoming Commits
Fetch | Pull
Outgoing Commits
Push | View Summary
Solution Explorer
Team Explorer

Repository is already up to date. No changes to pull.

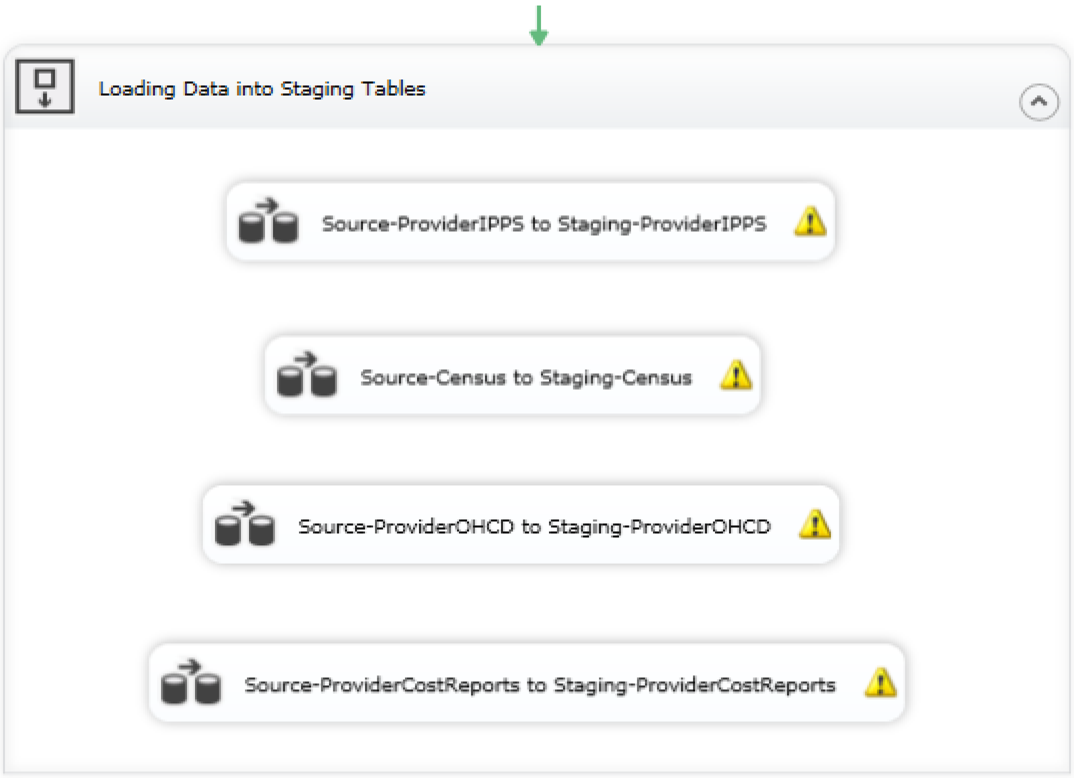
There are no incoming commits.

There are no outgoing commits.

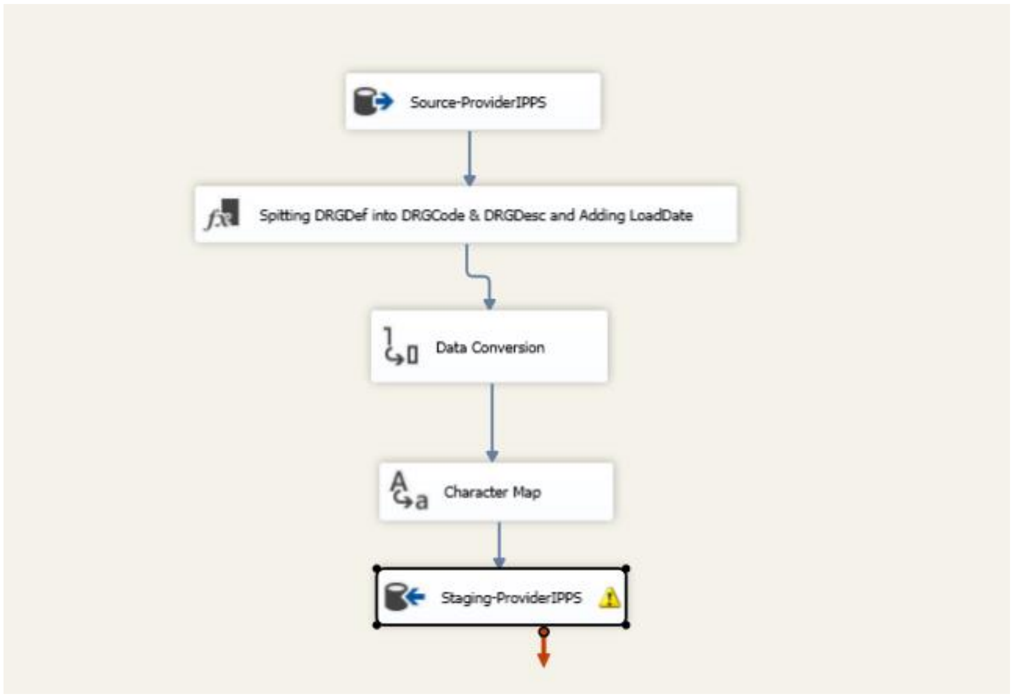
5. Run the container "Load Source Flat Files" in SSIS to load it to Source Table



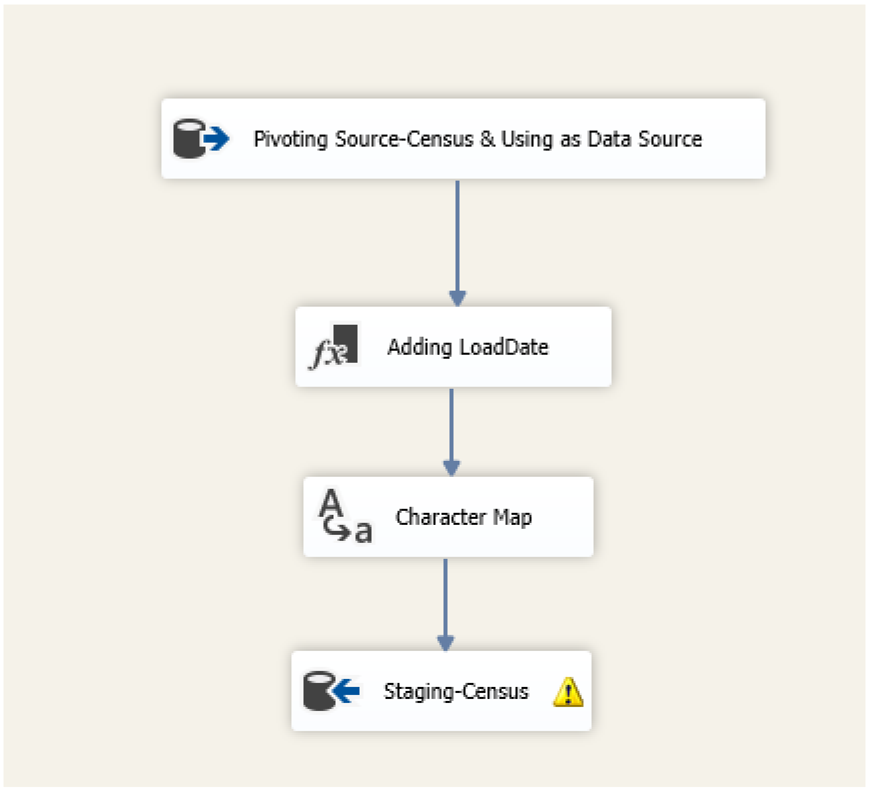
6. Run the container “Loading Data into Staging Tables” in SSIS to load it to Staging table


Loading Data into Staging Tables
Source-ProviderIPPS to Staging-ProviderIPPS
Source-Census to Staging-Census
Source-ProviderOHCD to Staging-ProviderOHCD
Source-ProviderCostReports to Staging-ProviderCostReports

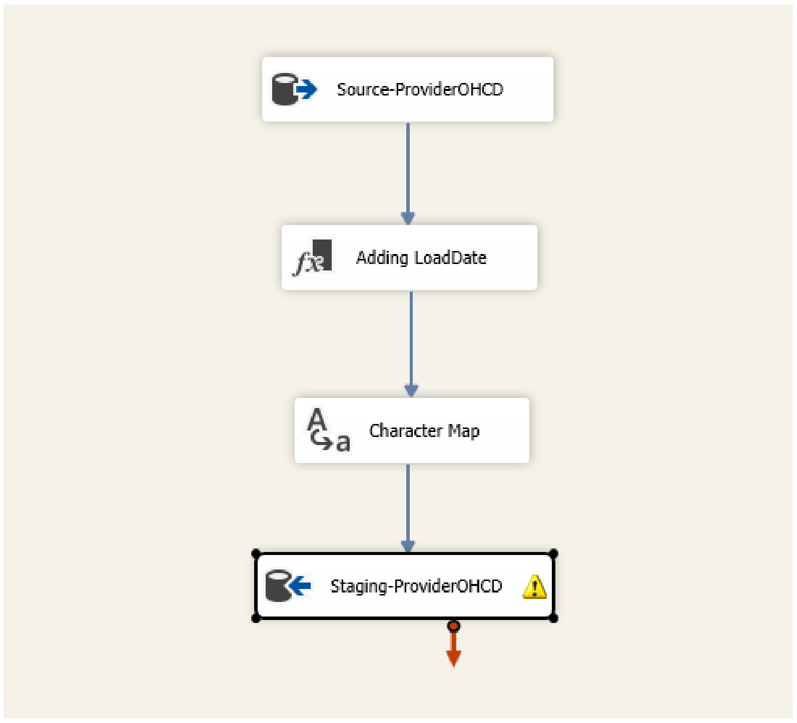
- Provider IPPS (Source to Staging)



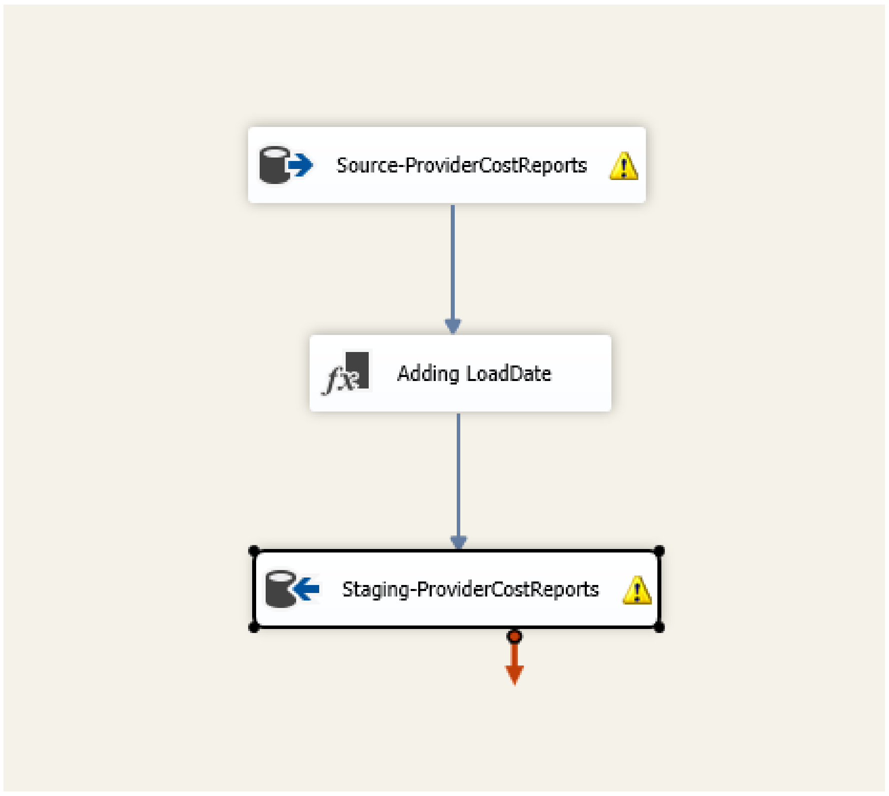
- Census (Source to Staging)



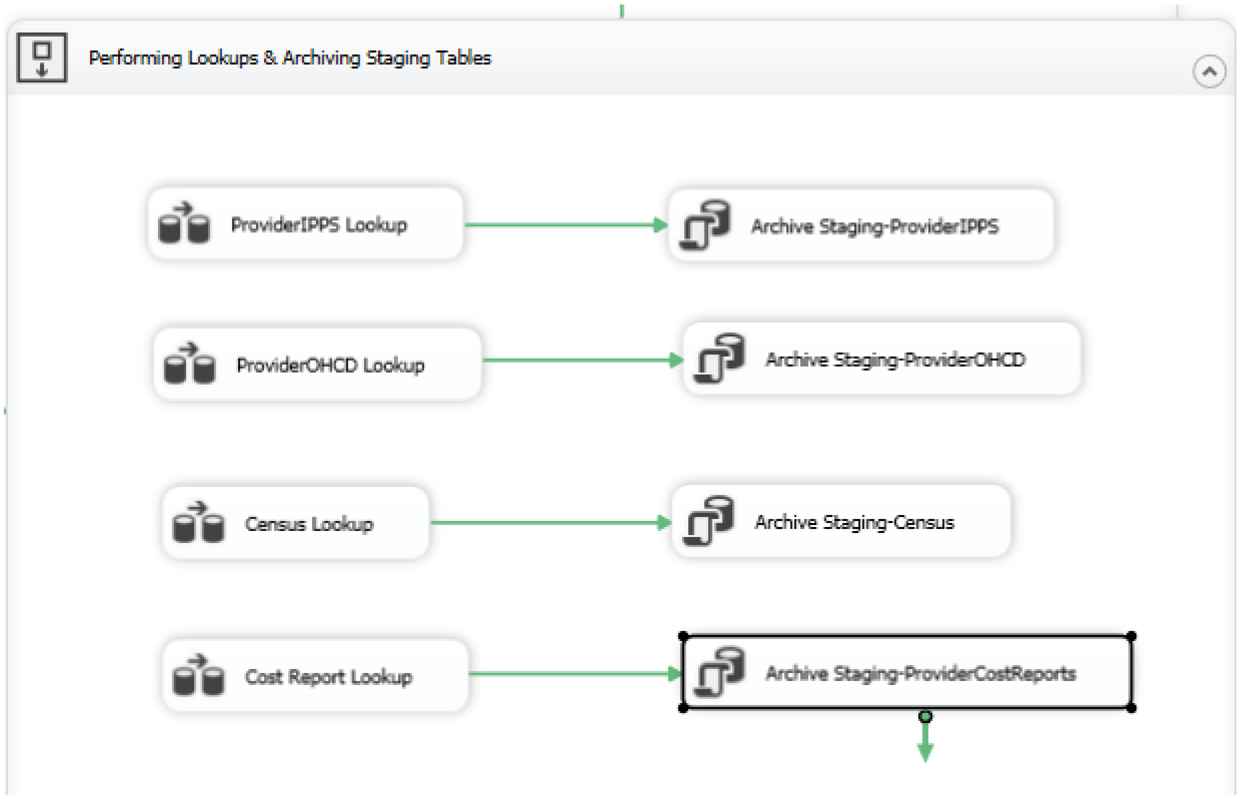
- Provider OHCD (Source to Staging)



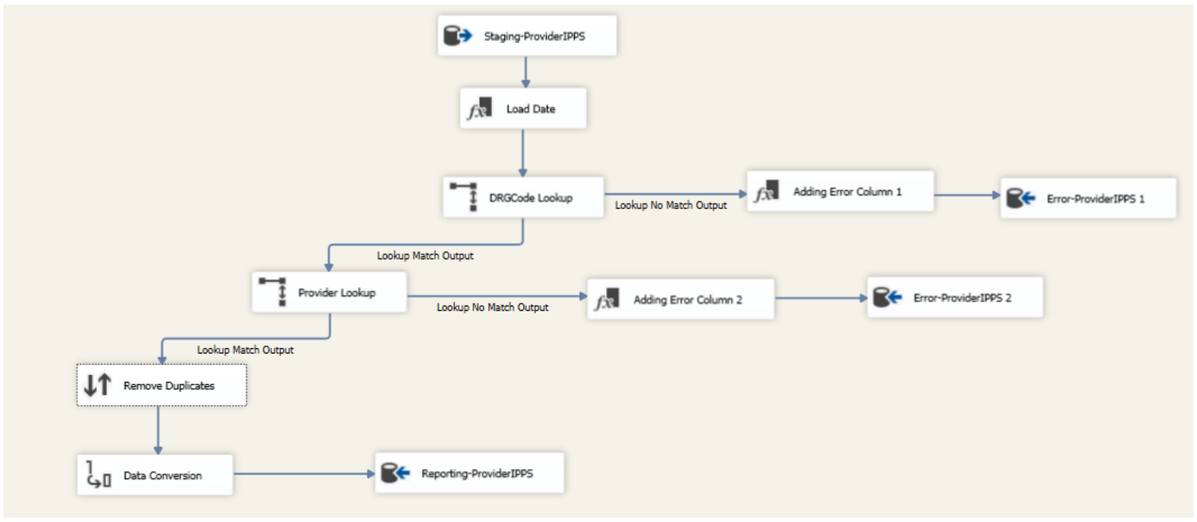
- Provider Cost Reports (Source to Staging)


Source-ProviderCostReports
Adding LoadDate
Staging-ProviderCostReports

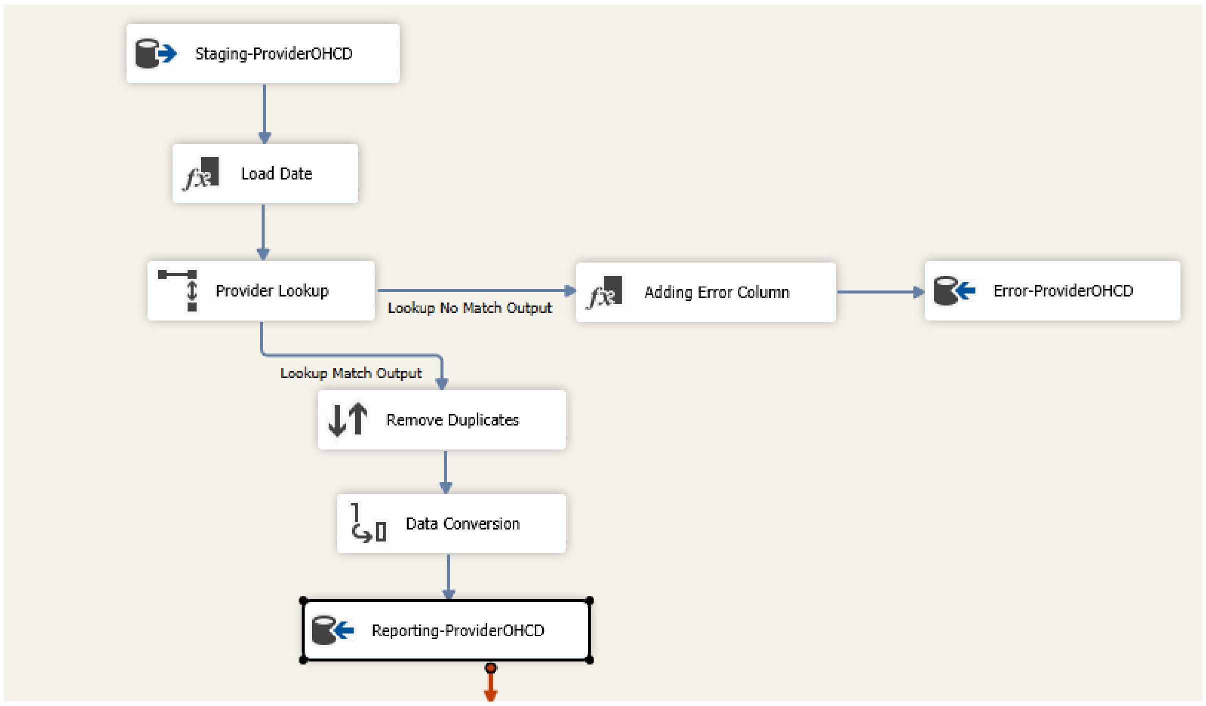
7. Lookup


Performing Lookups & Archiving Staging Tables
ProviderIPPS Lookup
Archive Staging-ProviderIPPS
ProviderOHCD Lookup
Archive Staging-ProviderOHCD
Census Lookup
Archive Staging-Census
Cost Report Lookup
Archive Staging-ProviderCostReports

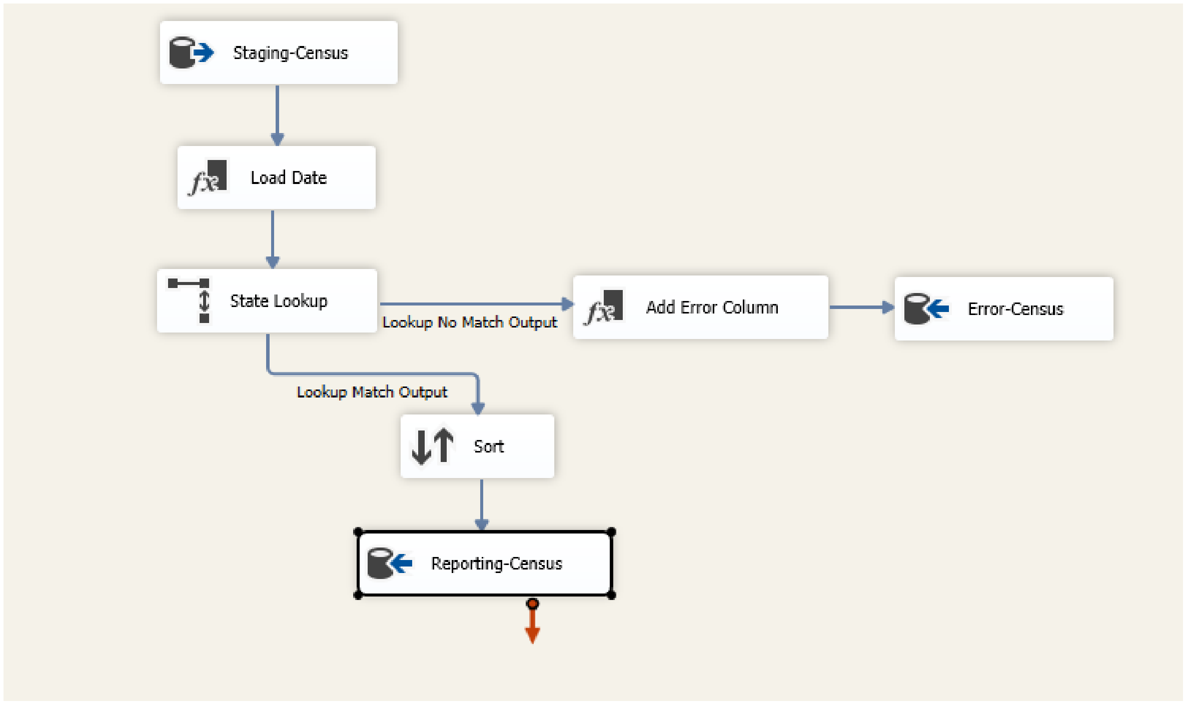
- Provider IPPS Lookup



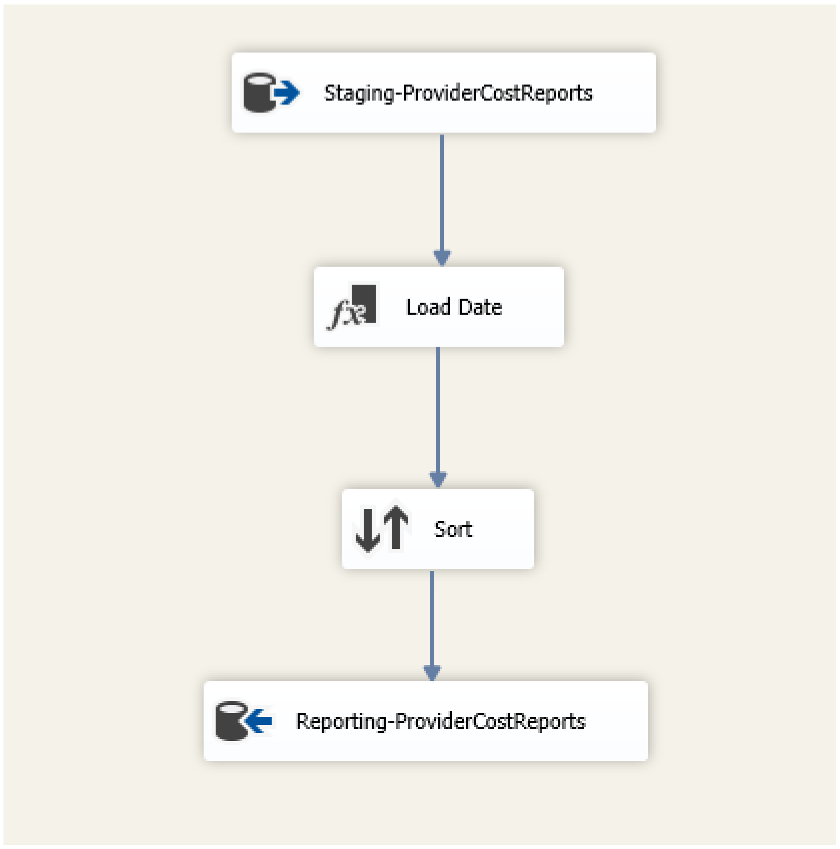
- Provider OHCD Lookup



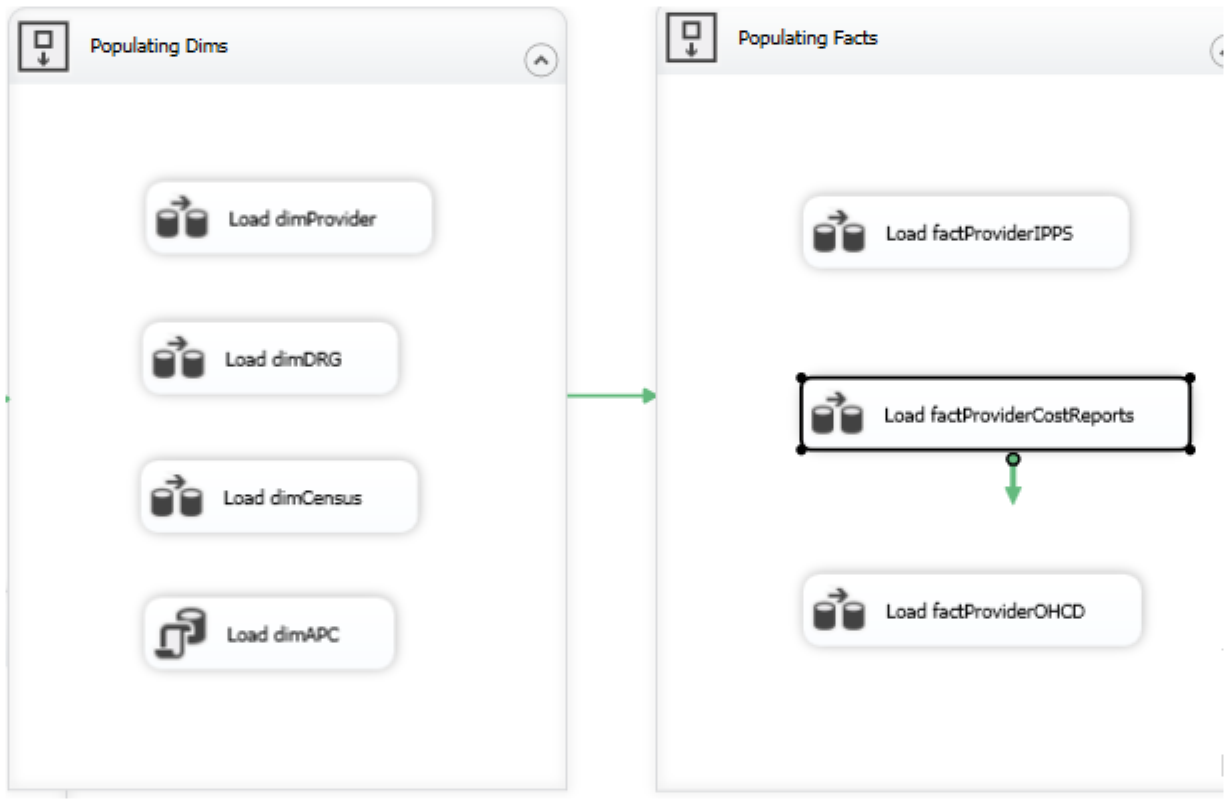
- Provider OHCD Lookup



- Provider OHCD Lookup



8. Populating Dims and Facts

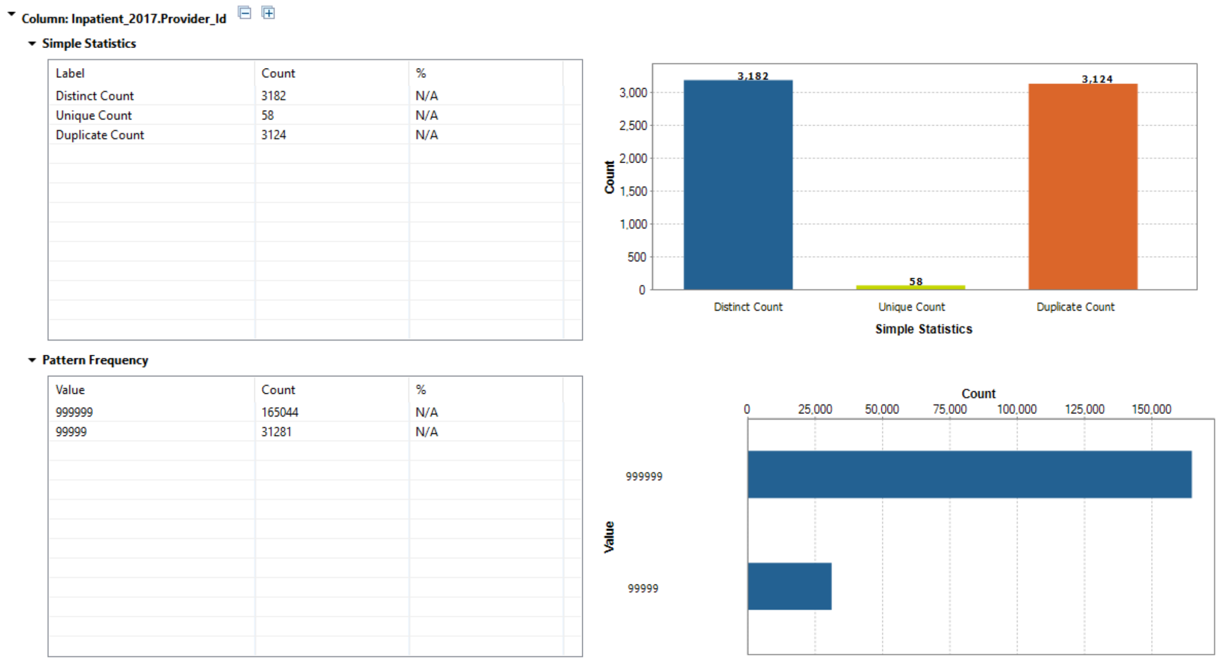


Observations

Data Cleaning

For data cleaning,we used Talend for data profiling to examine the available data. The key fields from each data source were examined using the Pattern Frequency and Simple Statistics feature of Talend.

- Provider ID



Column: Inpatient_2017.Provider_Id

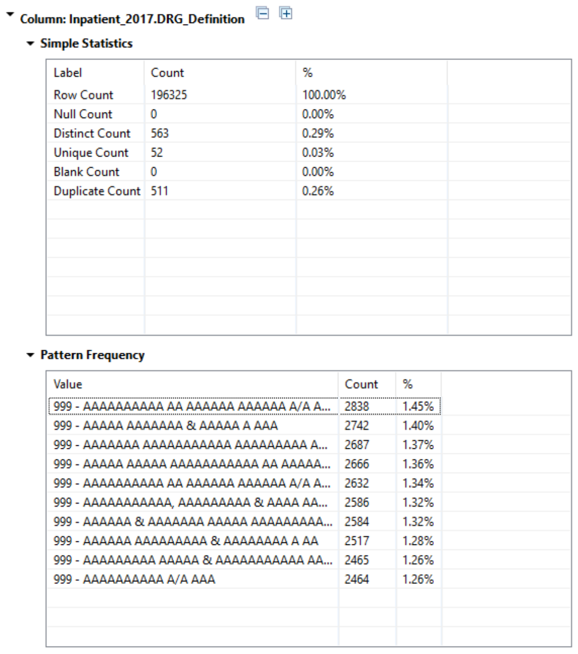
Simple Statistics

Label	Count	%
Distinct Count	3182	N/A
Unique Count	58	N/A
Duplicate Count	3124	N/A

Pattern Frequency

Value	Count	%
999999	165044	N/A
99999	31281	N/A

- DRG Definition



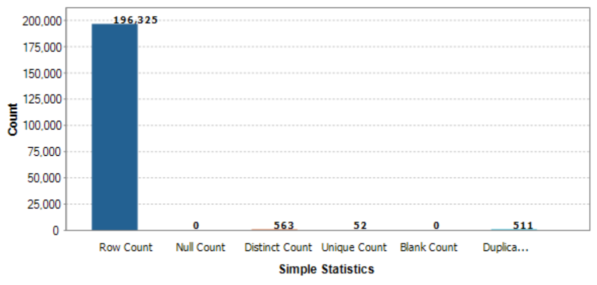
Column: Inpatient_2017.DRG_Definition

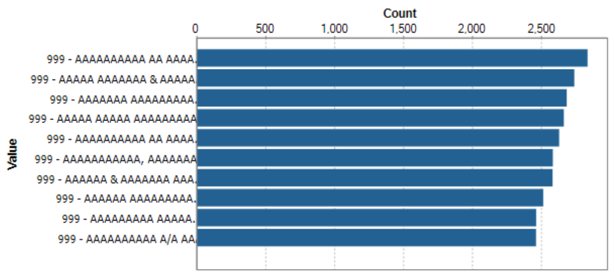
Simple Statistics

Label	Count	%
Row Count	196325	100.00%
Null Count	0	0.00%
Distinct Count	563	0.29%
Unique Count	52	0.03%
Blank Count	0	0.00%
Duplicate Count	511	0.26%

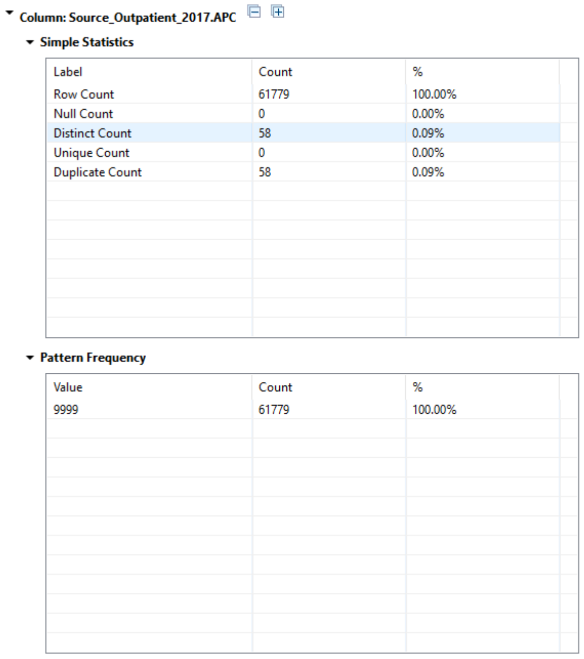
Pattern Frequency

Value	Count	%
999 - AAAAAAAAAA AA AAAAAA AAAAAA A/A A...	2838	1.45%
999 - AAAAA AAAAAAA & AAAAA A AAA	2742	1.40%
999 - AAAAAAA AAAAAAAAAAA AAAAAAAAA A...	2687	1.37%
999 - AAAAA AAAAA AAAAAAAAAAA AA AAAAA...	2666	1.36%
999 - AAAAAAAAAA AA AAAAAA AAAAAA A/A A...	2632	1.34%
999 - AAAAAAAAAAA, AAAAAAAAA & AAAA AA...	2586	1.32%
999 - AAAAAA & AAAAAAA AAAAA AAAAAAAAA...	2584	1.32%
999 - AAAAAA AAAAAAAAA & AAAAAAAA A AA	2517	1.28%
999 - AAAAAAAAA AAAAA & AAAAAAAAAAA AA...	2465	1.26%
999 - AAAAAAAAAA A/A AAA	2464	1.26%





- APC Code



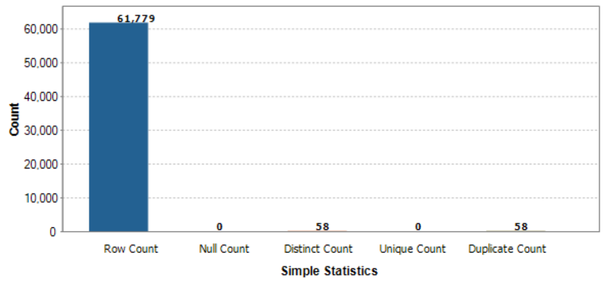
Column: Source_Outpatient_2017.APC

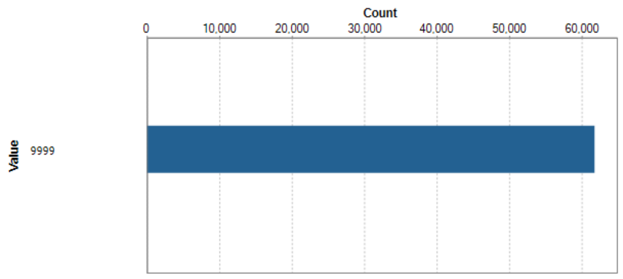
Simple Statistics

Label	Count	%
Row Count	61779	100.00%
Null Count	0	0.00%
Distinct Count	58	0.09%
Unique Count	0	0.00%
Duplicate Count	58	0.09%

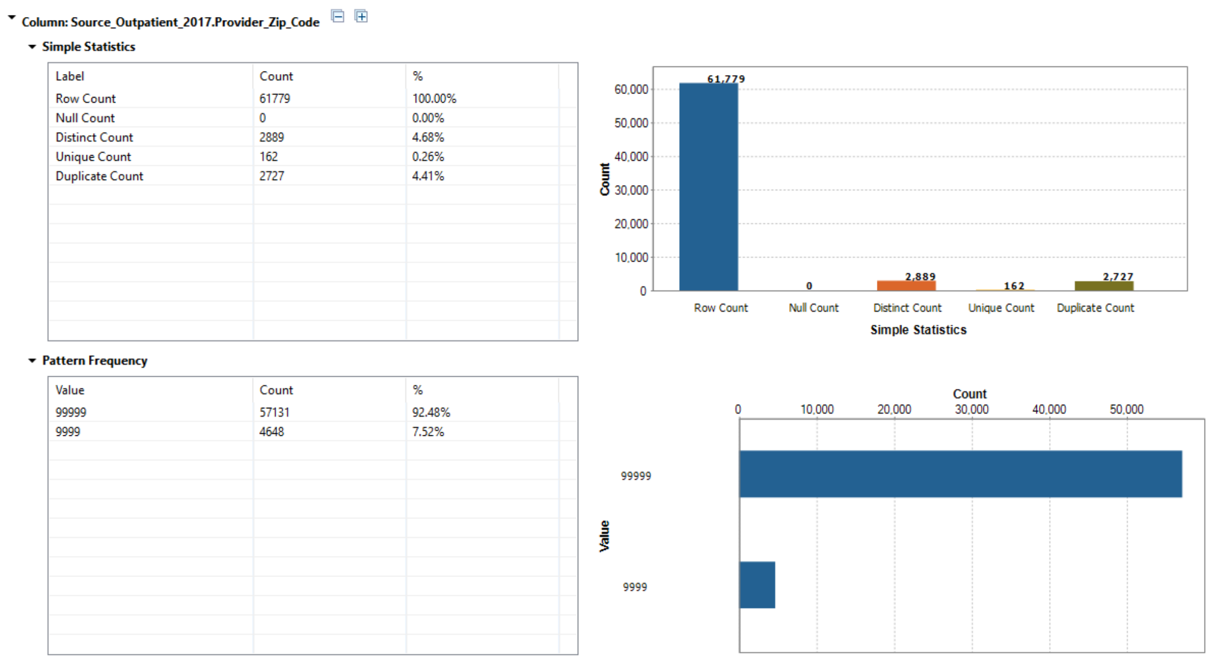
Pattern Frequency

Value	Count	%
9999	61779	100.00%





- Zip Code



Column: Source_Outpatient_2017.Provider_Zip_Code

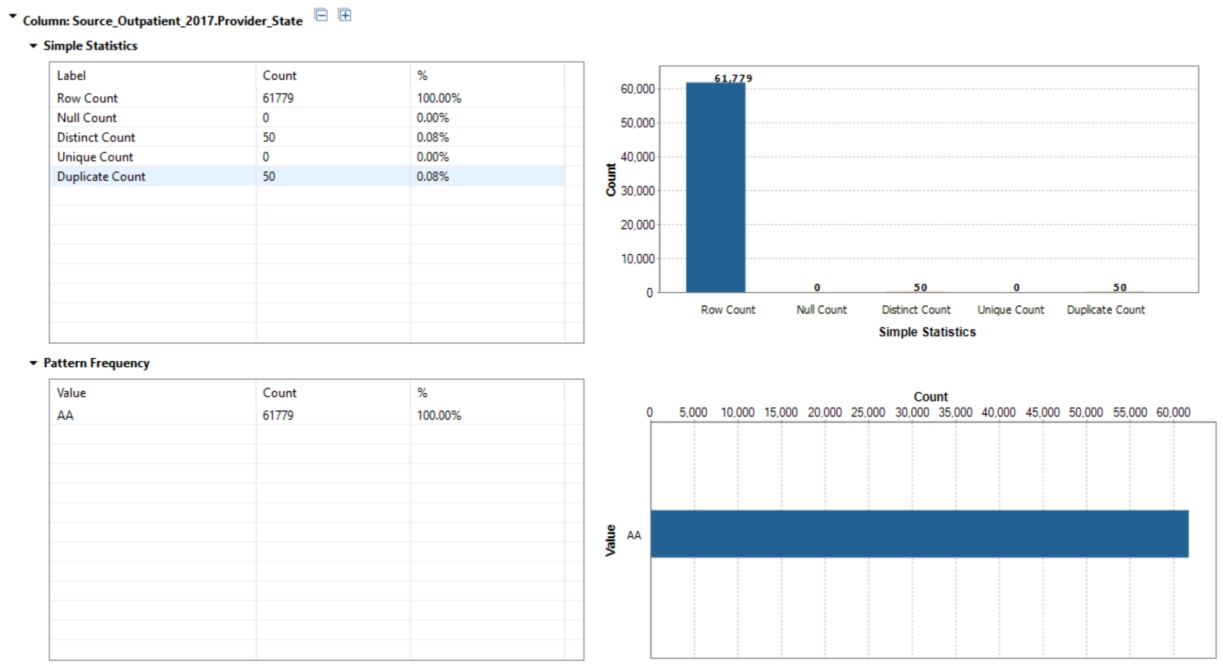
Simple Statistics

Label	Count	%
Row Count	61779	100.00%
Null Count	0	0.00%
Distinct Count	2889	4.68%
Unique Count	162	0.26%
Duplicate Count	2727	4.41%

Pattern Frequency

Value	Count	%
99999	57131	92.48%
9999	4648	7.52%

- State Code



Column: Source_Outpatient_2017.Provider_State

Simple Statistics

Label	Count	%
Row Count	61779	100.00%
Null Count	0	0.00%
Distinct Count	50	0.08%
Unique Count	0	0.00%
Duplicate Count	50	0.08%

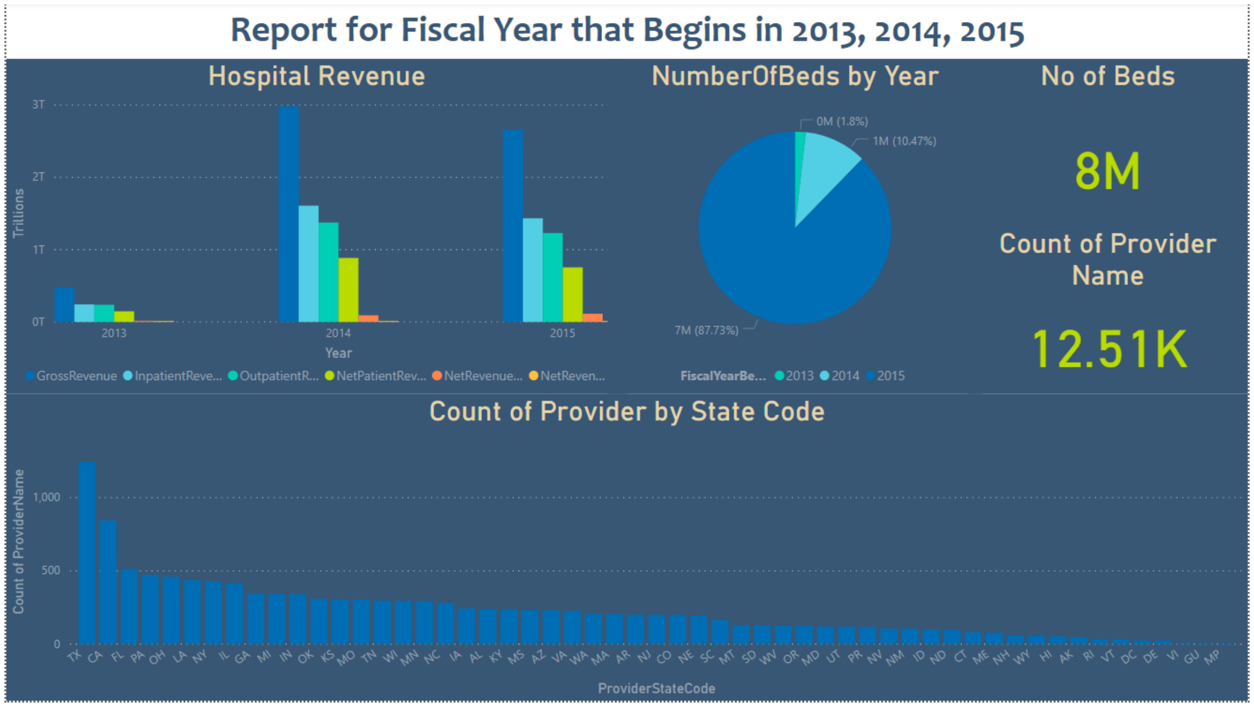
Pattern Frequency

Value	Count	%
AA	61779	100.00%

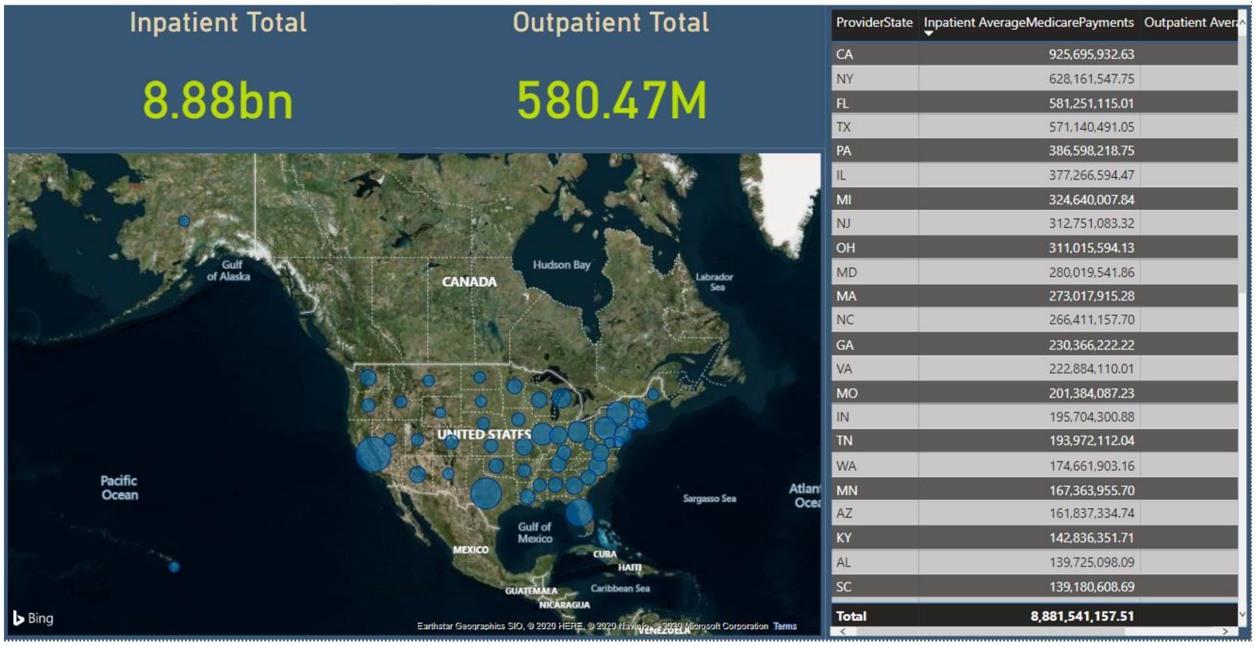
- Apart from this, fields in Provider Cost report required to exclude certain characters such as “$” and “,” to make the data uniform.
- All the character fields were converted to UpperCase

- CamelCase was used as naming convention
- Zip Code, Provider CCN, APCCode and DRGCodes were kept as varchar to maintain data integrity of the codes.

Visualizations



Inpatient and Outpatient Average Medicare Amount



Inpatient Total

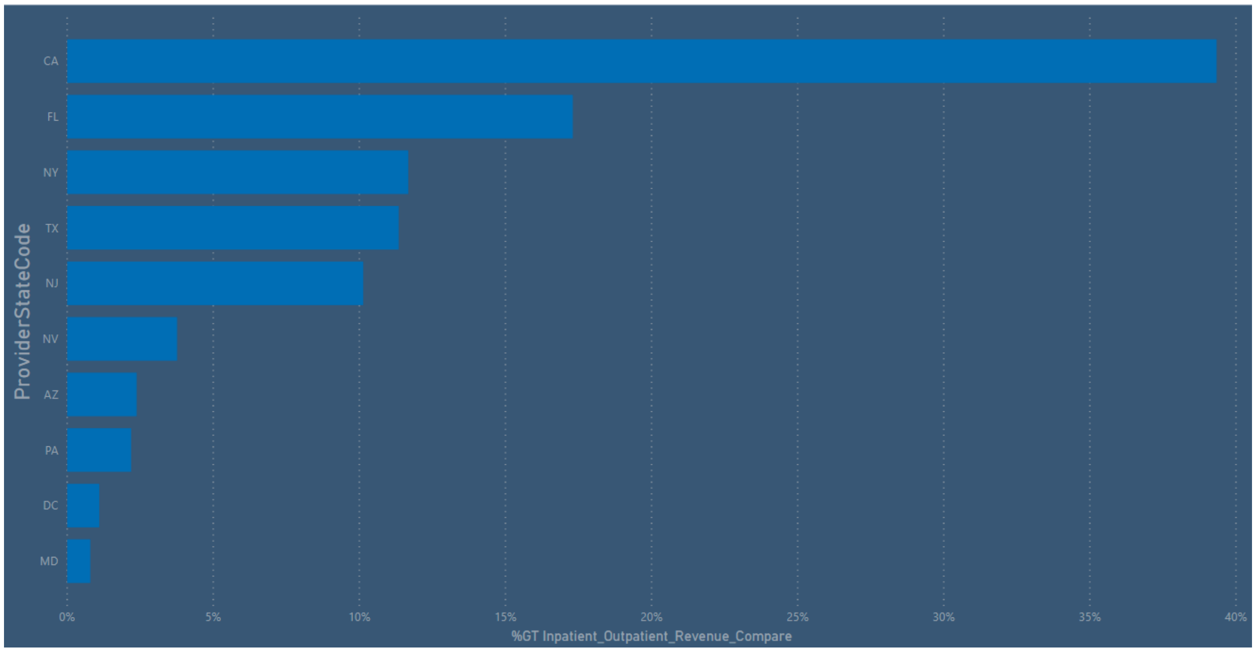
8.88bn

Outpatient Total

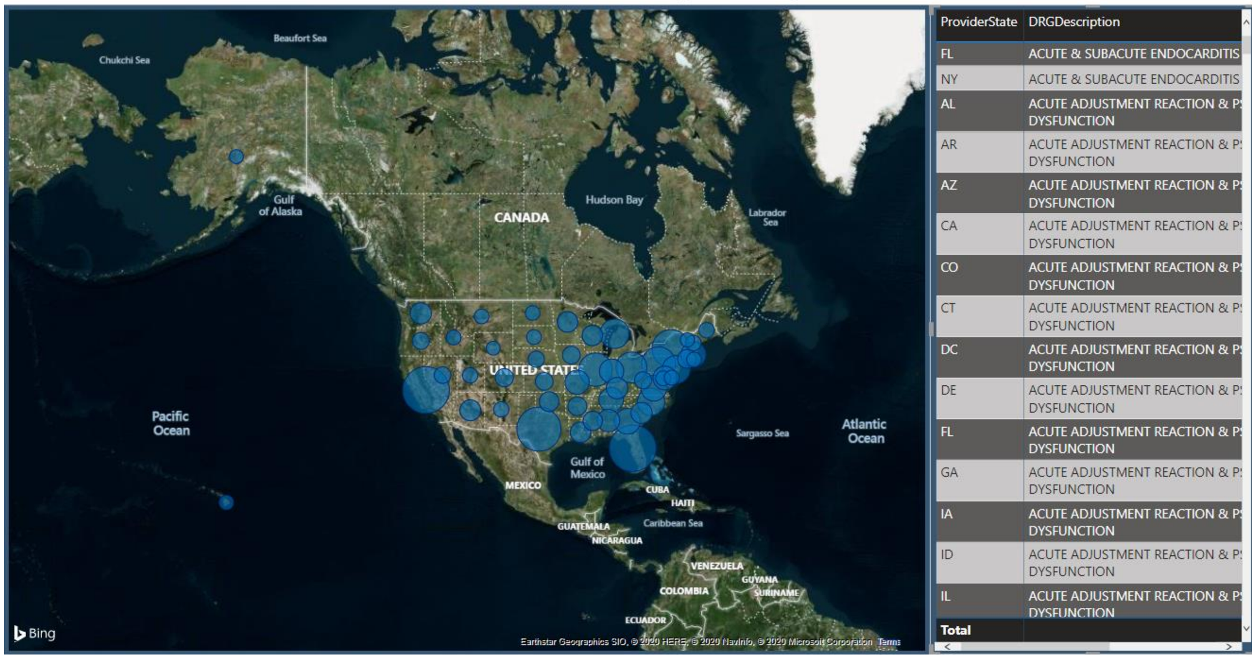
580.47M

ProviderState	Inpatient AverageMedicarePayments	Outpatient Aver
CA	925,695,932.63	
NY	628,161,547.75	
FL	581,251,115.01	
TX	571,140,491.05	
PA	386,598,218.75	
IL	377,266,594.47	
MI	324,640,007.84	
NJ	312,751,083.32	
OH	311,015,594.13	
MD	280,019,541.86	
MA	273,017,915.28	
NC	266,411,157.70	
GA	230,366,222.22	
VA	222,884,110.01	
MO	201,384,087.23	
IN	195,704,300.88	
TN	193,972,112.04	
WA	174,661,903.16	
MN	167,363,955.70	
AZ	161,837,334.74	
KY	142,836,351.71	
AL	139,725,098.09	
SC	139,180,608.69	
Total	**8,881,541,157.51**	

Top 10 Provider States with Inpatient greater than Outpatient Revenue

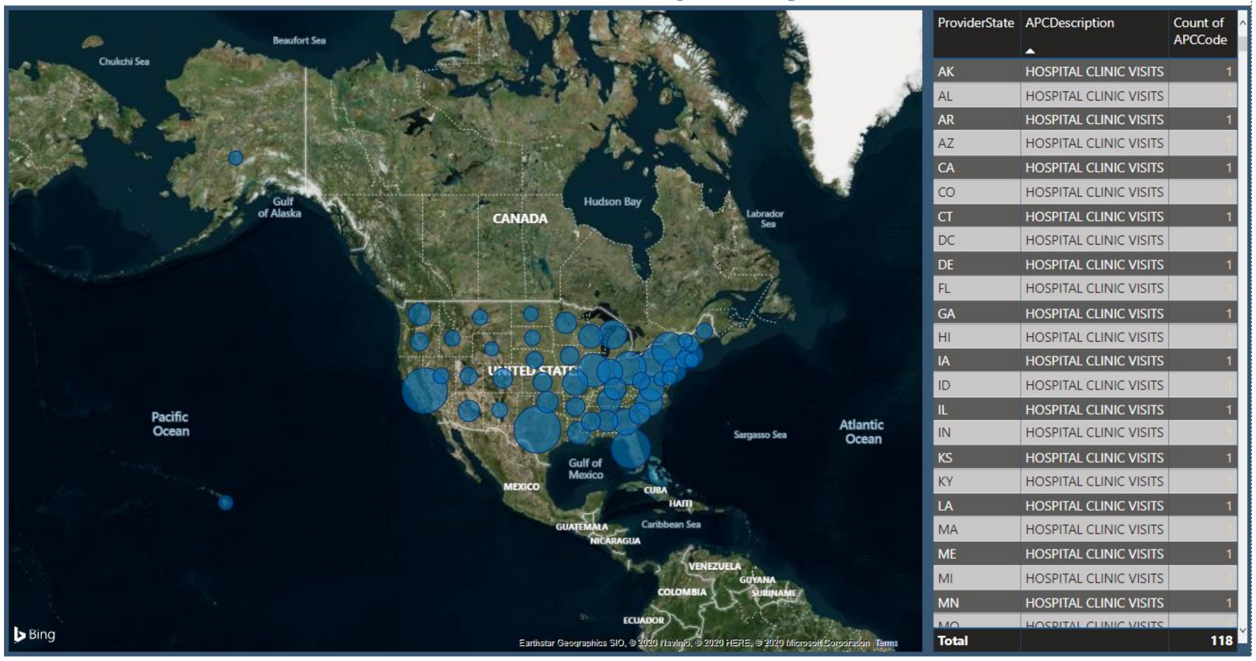


Count of DRG Code and Description by Provider State



ProviderState	DRGDescription
FL	ACUTE & SUBACUTE ENDOCARDITIS
NY	ACUTE & SUBACUTE ENDOCARDITIS
AL	ACUTE ADJUSTMENT REACTION & P DYSFUNCTION
AR	ACUTE ADJUSTMENT REACTION & P DYSFUNCTION
AZ	ACUTE ADJUSTMENT REACTION & P DYSFUNCTION
CA	ACUTE ADJUSTMENT REACTION & P DYSFUNCTION
CO	ACUTE ADJUSTMENT REACTION & P DYSFUNCTION
CT	ACUTE ADJUSTMENT REACTION & P DYSFUNCTION
DC	ACUTE ADJUSTMENT REACTION & P DYSFUNCTION
DE	ACUTE ADJUSTMENT REACTION & P DYSFUNCTION
FL	ACUTE ADJUSTMENT REACTION & P DYSFUNCTION
GA	ACUTE ADJUSTMENT REACTION & P DYSFUNCTION
IA	ACUTE ADJUSTMENT REACTION & P DYSFUNCTION
ID	ACUTE ADJUSTMENT REACTION & P DYSFUNCTION
IL	ACUTE ADJUSTMENT REACTION & P DYSFUNCTION
Total	

Count of APC Code and Description by Provider State



ProviderState	APCDescription	Count of APCCode
AK	HOSPITAL CLINIC VISITS	1
AL	HOSPITAL CLINIC VISITS	1
AR	HOSPITAL CLINIC VISITS	1
AZ	HOSPITAL CLINIC VISITS	1
CA	HOSPITAL CLINIC VISITS	1
CO	HOSPITAL CLINIC VISITS	1
CT	HOSPITAL CLINIC VISITS	1
DC	HOSPITAL CLINIC VISITS	1
DE	HOSPITAL CLINIC VISITS	1
FL	HOSPITAL CLINIC VISITS	1
GA	HOSPITAL CLINIC VISITS	1
HI	HOSPITAL CLINIC VISITS	1
IA	HOSPITAL CLINIC VISITS	1
ID	HOSPITAL CLINIC VISITS	1
IL	HOSPITAL CLINIC VISITS	1
IN	HOSPITAL CLINIC VISITS	1
KS	HOSPITAL CLINIC VISITS	1
KY	HOSPITAL CLINIC VISITS	1
LA	HOSPITAL CLINIC VISITS	1
MA	HOSPITAL CLINIC VISITS	1
ME	HOSPITAL CLINIC VISITS	1
MI	HOSPITAL CLINIC VISITS	1
MN	HOSPITAL CLINIC VISITS	1
MO	HOSPITAL CLINIC VISITS	
Total		**118**

Future Scope

- Move flat file loading completely to cloud to implement all cloud based infrastructure
- Improve dimensional model by adding dimTime and dimGeography