Northeastern University

Project IE7290-

Data Warehouse and Business Intelligence.

Analyzing Opioid Drug Evolution from
2013 to 2017

Team Name   :   Husky Coders

Team Number: 09

Professor Name: Vincent Lattuada

Subject Name: Data Warehouse and Business Intelligence.

Team Members

Sl no	Name	Contact	Team Designation
1	Anirudh Sundaresan	Sundaresan.a@northeastern.edu	Team Leader
2	Archisha Dama	Dama.a@husky.neu.edu	Member
3	Akshay Rao	Udupiragavendrarao.a@northeastern.edu	Member

Document Revision History

Change Number	Change Made By	Color / Font Identification and Purpose	Version
CH0001	All Members	Project Proposal	1.0
CH0002	Archisha Dama	Change in Datasets, Methodology and Loading Process	1.1
CH003	Akshay Rao	Changes in number of years in Analysis.	1.2

CH004	Anirudh Sundaresan	ERD – Star Schema	1.3
CH005	Akshay Rao	Data flow diagram	1.4
CH006	Anirudh Sundaresan	Logical model	1.4
CH007	Anirudh Sundaresan	ETL Mapping	1.4
CH008	Archisha Dama	Stakeholders	1.4
CH009	Archisha Dama	References	1.4

Contents

Project IE7290 – Data Warehouse and Business Intelligence.

Abstract

Examination of Opioid Drugs in United States of America from 2013 to 2017, Detailed analysis on the prescription of Opioid Drugs by medical facilities during the year 2017.

Background:

Prescribing Opioids is controversial with evidence of both significant underutilization and over utilization. There is some evidence to support the efficacy for chronic non- malignant pain, but community and individual harms are increasing and are respectively reported.

Aim:

To review the number of opioids drugs their availability in different drug names over the years 2013 and 2017, counts and trends of Opioid Drugs available from the year 2013 to 2017.

Their usage trends by initial drug prescription and taking the refill counts for 30 Days, total claims form Drug usages, Drug Costs, Drug usage in patients over 65 Years for the year 2017.

Geographic visualization of patterns in Opioid Drug Usage in the year 2017 in United States of America.

Analysis on the rate of drug prescriptions/ Consumption with respect to the year 2013 in United States of America, Their Urban and Rural Trends broken down to each State and County.

Stakeholders:

Physicians, Healthcare employers, Insurance companies and Pharmaceutical firms who want to understand Medicare Drug prescription and Utilization at a high level and make decisions.

Data Sets

- **Base_Provider_Enrollment_File :** The source of this data is the Provider Enrollment, Chain, and Ownership System. PECOS is an electronic Medicare enrollment system and national data repository for individual and organizational providers who have enrolled or are enrolling in Medicare . Records Enrollment level data for providers actively enrolled in Medicare.

  NPI- National Provider Identifier, Unique number issued to Health Care providers

  PECOS_ASCT_CNTL_ID- PECOS Associate Control ID (PAC ID): Unique number assigned by PECOS to the enrolling entity.

  ENRLMT_ID -Unique number assigned by PECOS to the provider enrollment application. The first character encodes whether the enrollment is for an individual provider (first character is 'I') or for an organizational provider (first character is 'O').

  STATE_CD – Enrollment State code of Provider

- **Medicare_Provider_Utilization_and_Payment_Data_Part_D_Prescriber** : Data set contains information on prescription drugs prescribed by individual physicians and other health care providers and paid for under the Medicare Part D Prescription Drug Program. The link provides the data for the year 2016 and 2017.

  BENE_COUNT– The total number of unique Medicare Part D beneficiaries with at least one claim for the drug. Counts fewer than 11 are suppressed and are indicated by a blank.

  TOTAL_CLAIM_COUNT- The number of Medicare Part D claims. This   includes original prescriptions and refills.

  TOTAL_DAY_SUPPLY–The aggregate number of day's supply for which this drug was dispensed.

  TOTAL_DRUG_COST-The aggregate drug cost paid for all associated claims. This amount includes ingredient cost, dispensing fee, sales tax, and any applicable vaccine administration fees and is based on the amounts paid by the Part D plan, Medicare beneficiary, government subsidies, and any other third-party payers.

  BENE_COUNT_GE65– The total number of unique Medicare Part D beneficiaries age 65 and older with at least one claim for the drug.

- **Opioid_Drug_Lists_CY2013_CY2017** : This contains the list of Opioid drugs from the year 2013 to 2017. The link leads to Medicare Part D Opioid Prescribing Mapping Tool where there is a zip file at the bottom which can be downloaded.

- **Part D Prescriber National Summary Report**-Provides information on prescription drugs prescribed by individual physicians and other health care providers and paid for under the Medicare Part D Prescription Drug Program.

- **Medicare_Speaciality_Code** : Links the Supplier type description to its Medicare specialty code.

  MEDICARE SPECIALTY CODE -Describes the type of physicians, non-physicians and health care providers
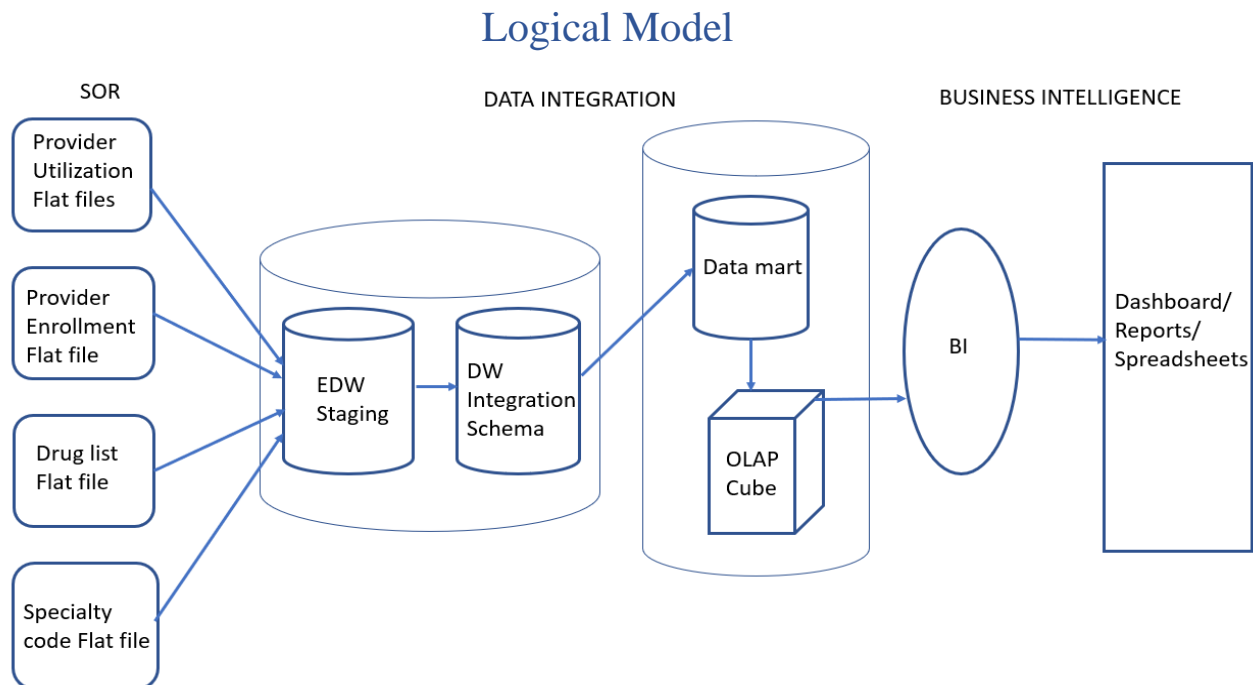
  TAXONOMY CODE-Taxonomy codes are 10-digit administrative codes set for identifying the provider type and area of specialization for health care providers

Methodology

- **Select the business process:** The business process is to study the Prescriber-D Provider Enrollment, drugs, usages, claim patterns. The source data is taken from CMS.gov which are in different formats such as CSV, Excel, text etc
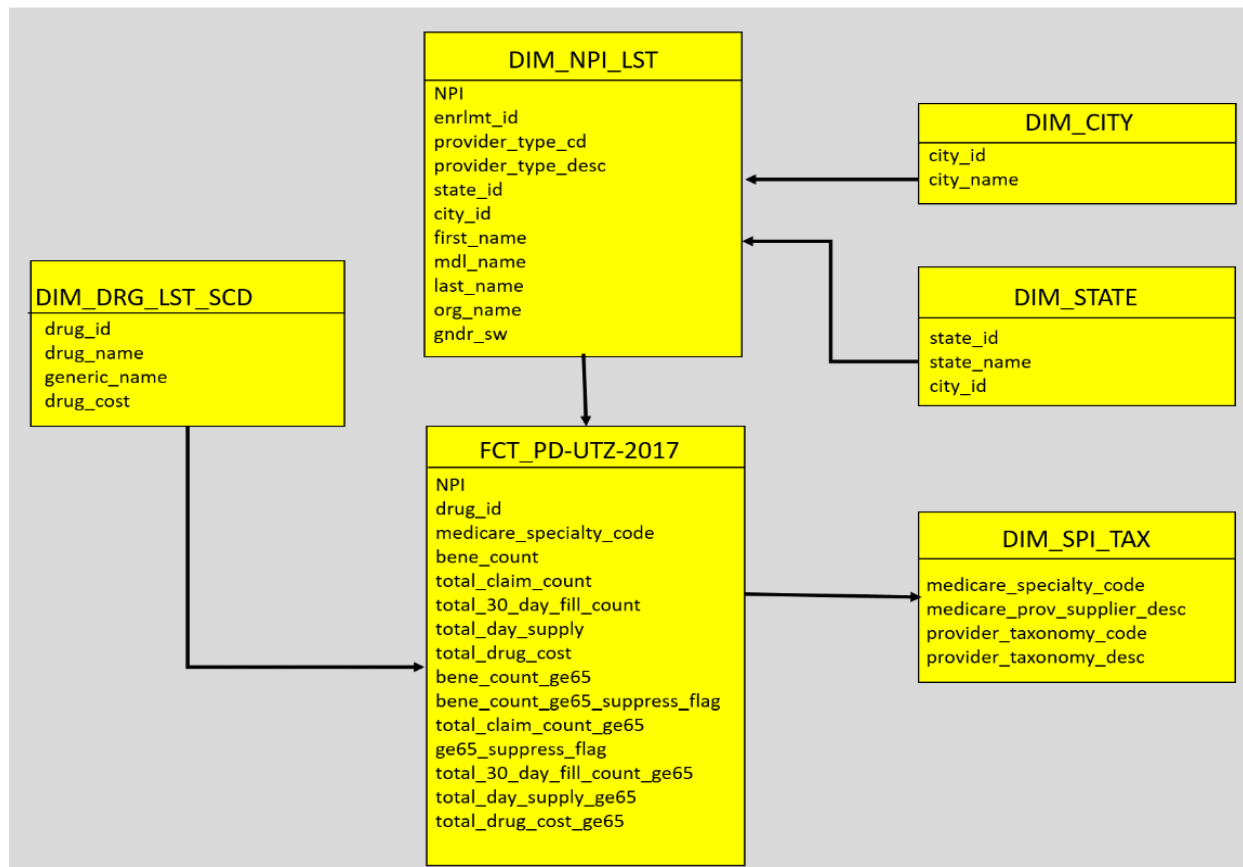
- **Determining the granularity and keys:** Create a high-level conceptual Entity Relationship model with entities on which the project is based on. Identify the primary keys, foreign keys, datatypes and connections.
- **ETL Mapping document:** An ETL Mapping document with source and Target tables is defined with data transfer logic such as direct mapping, extracting data with logical conditions, datatypes definitions.
- **Data Cleaning:** Create a temporary stage table to Clean the Data sets and transform them into proper format using "Execute SQL Task" Tool of SSIS. The data in the staging table is stored in an archive table as a backup.
- **Designing the Datawarehouse:** Create a Database in SQL Server in de-normalized form for easy read along with Facts and Dimensions. Indexes are created to access data faster.
- **Populate the Datawarehouse:** Create a Workflow on SQL Server Integrated Service Tool to populate the Tables in the Warehouse and cater to Slowly changing dimensions using Lookup tables.
- **Analyze the data:** SQL Server Analysis Service can be used to create Data sources, Data views and cubes which help to identify the relationship and hierarchies between attributes and perform aggregations.
- **Visualization:** SSRS/Tableau/PowerBI can be used to answer our business questions.

Logical Model

Project IE7290 – Data Warehouse and Business Intelligence.

Star Schema

DIM_NPI_LST
- NPI
- enrlmt_id
- provider_type_cd
- provider_type_desc
- state_id
- city_id
- first_name
- mdl_name
- last_name
- org_name
- gndr_sw

DIM_CITY
- city_id
- city_name

DIM_STATE
- state_id
- state_name
- city_id

DIM_DRG_LST_SCD
- drug_id
- drug_name
- generic_name
- drug_cost

FCT_PD-UTZ-2017
- NPI
- drug_id
- medicare_specialty_code
- bene_count
- total_claim_count
- total_30_day_fill_count
- total_day_supply
- total_drug_cost
- bene_count_ge65
- bene_count_ge65_suppress_flag
- total_claim_count_ge65
- ge65_suppress_flag
- total_30_day_fill_count_ge65
- total_day_supply_ge65
- total_drug_cost_ge65

DIM_SPI_TAX
- medicare_specialty_code
- medicare_prov_supplier_desc
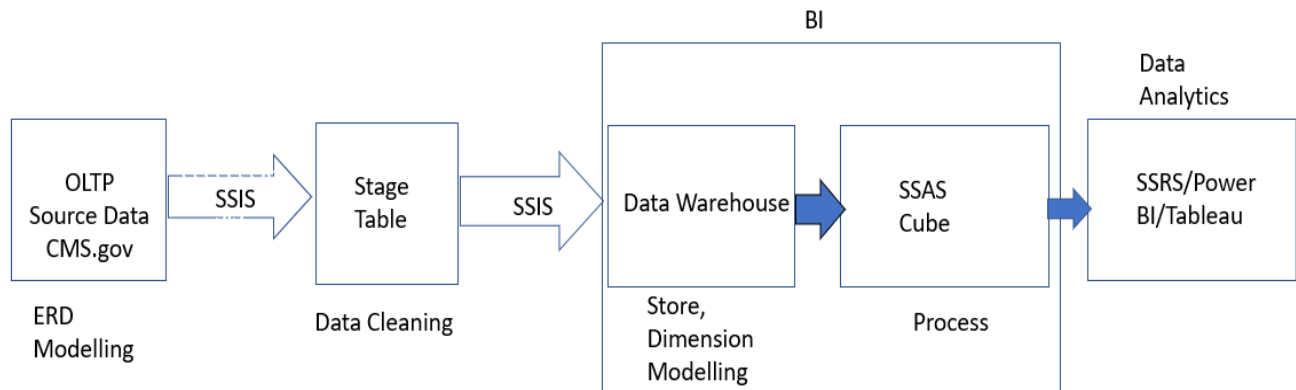- provider_taxonomy_code
- provider_taxonomy_desc

Data Definitions and ETL Mapping document

Source, Staging detailed definitions, ETL mapping

Project IE7290 – Data Warehouse and Business Intelligence.

Load Process

- Creating Projects in MS SQL Integration Services to get the loading process which can be used to periodically load and refresh the data based on a timely instance.
- Creating the Error Table, Archive Table and Staging table for each data sets.
- Drug list of Opioid Drugs is master data starting from the list of drugs from 2013 to 2017, This master data can be looked up and used for error handling and to create Dimension Table.
- Drug List will also have Slowly Changing Dimension set up throughout the years 2013 to 2017.
- Appropriate National Provider Identifier (NPI), Drug Name, Generic Names to have its respective.
- Dimension table in the schema and the data flows and loading process will be created in SSIS packages.
- Measures such as count of Drugs, Prescription rates are all stored in respective fact tables in the star schema and its loading process will be developed in an SSIS packages.
- SQL Logic to extract the row numbers and creating Lookup/ Dimension table will be made and tested in SSMS and then Integrated in a standard loading process in SSIS package.
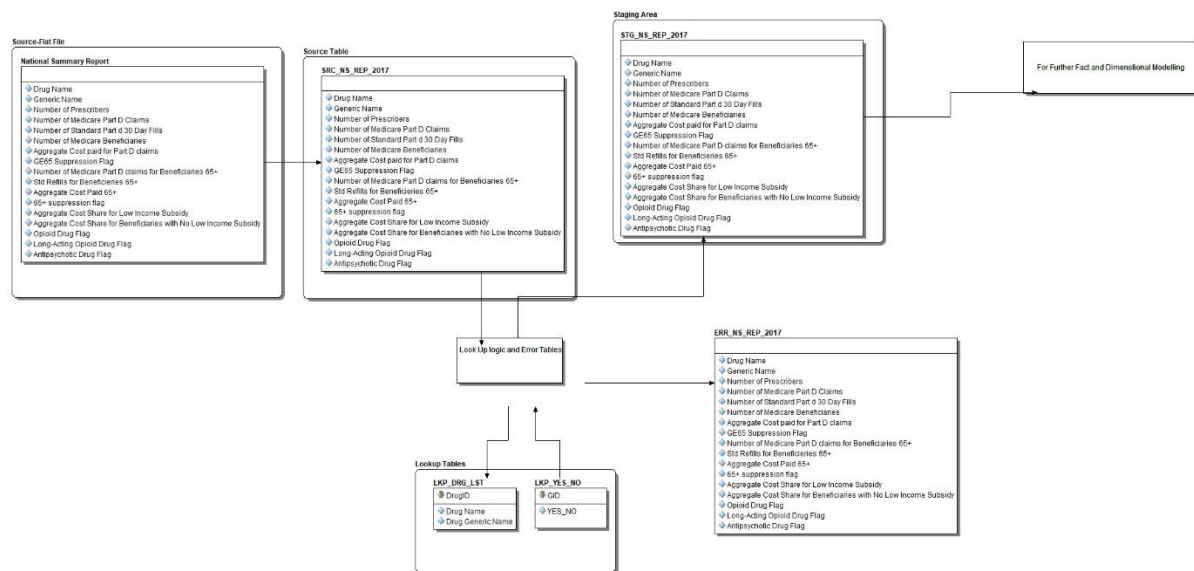
Detailed data flow diagram

Utilization payment:
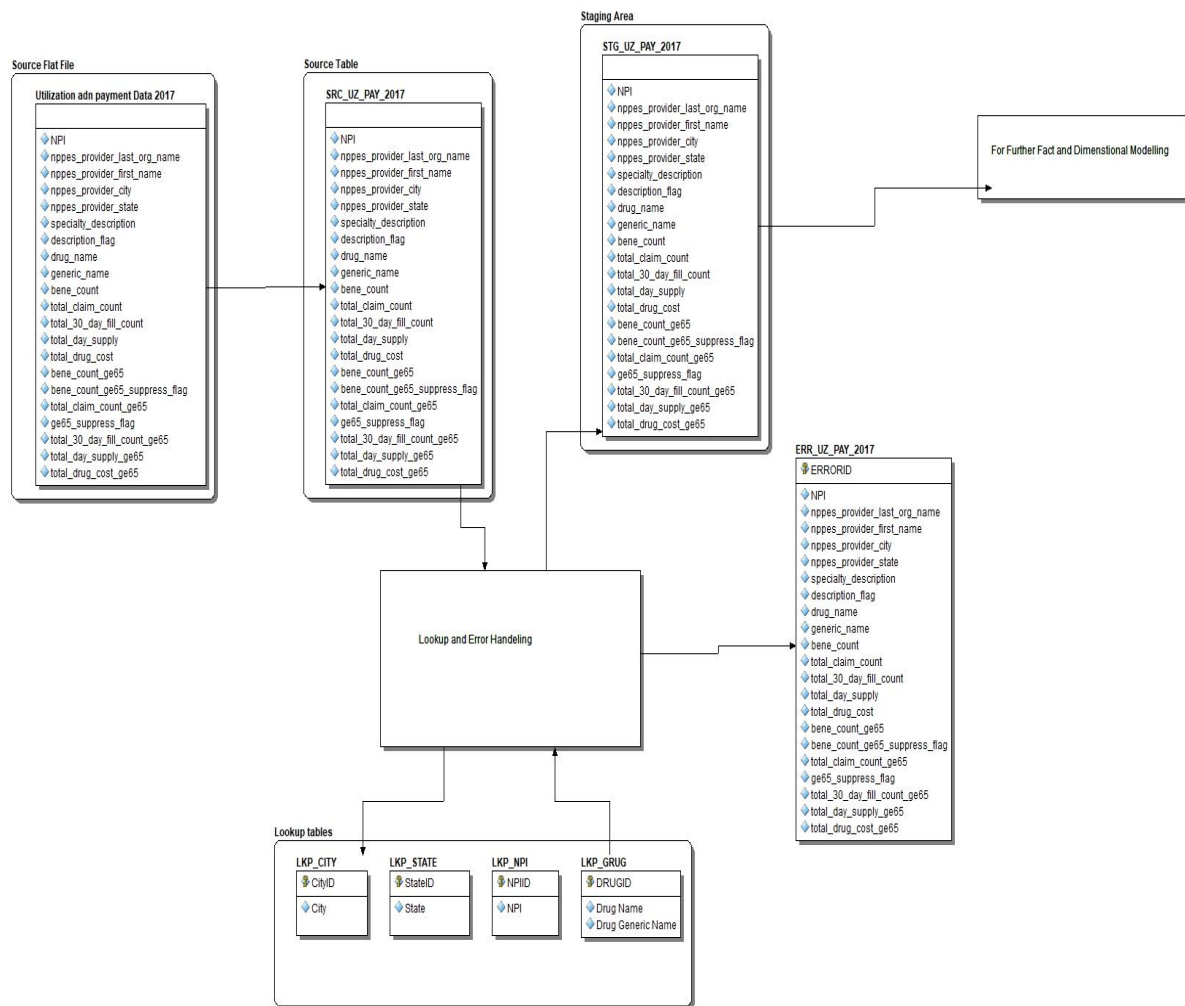
Summary report on Prescribed drugs:

Drug List:
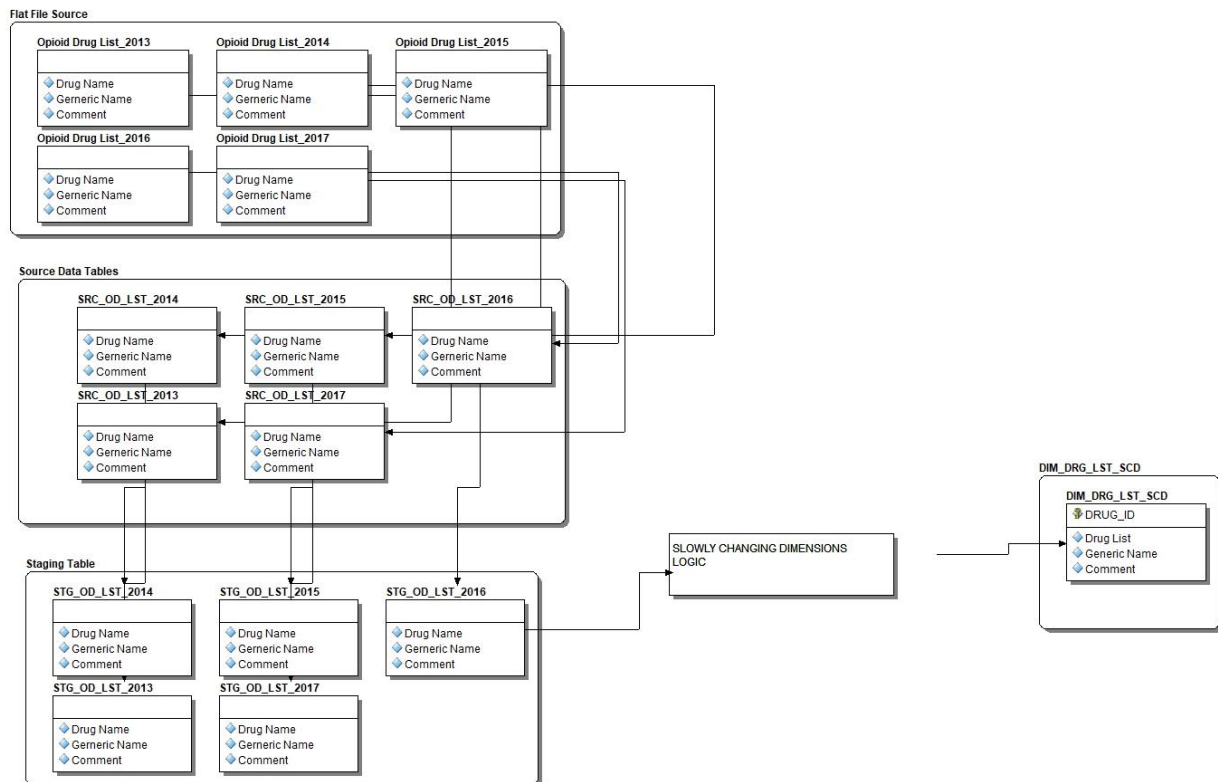
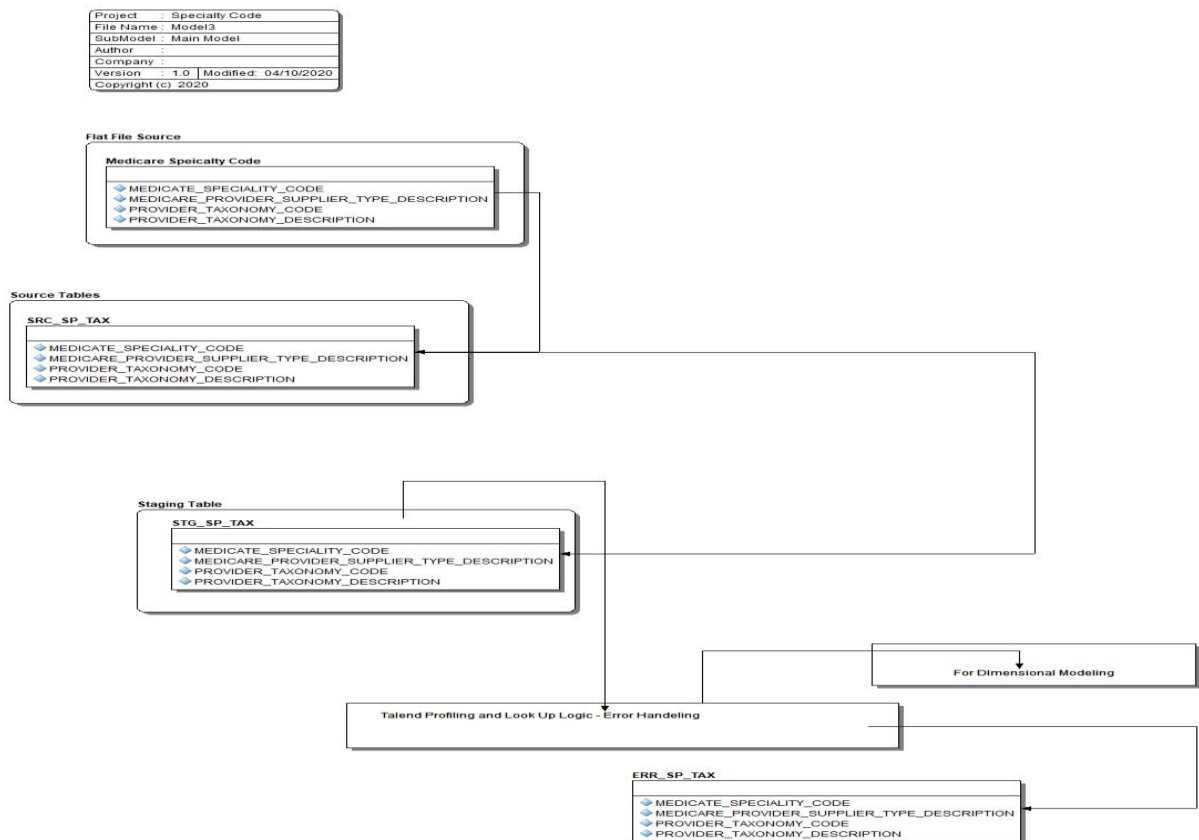# Project IE7290 – Data Warehouse and Business Intelligence.

Base Provider Data Flow:
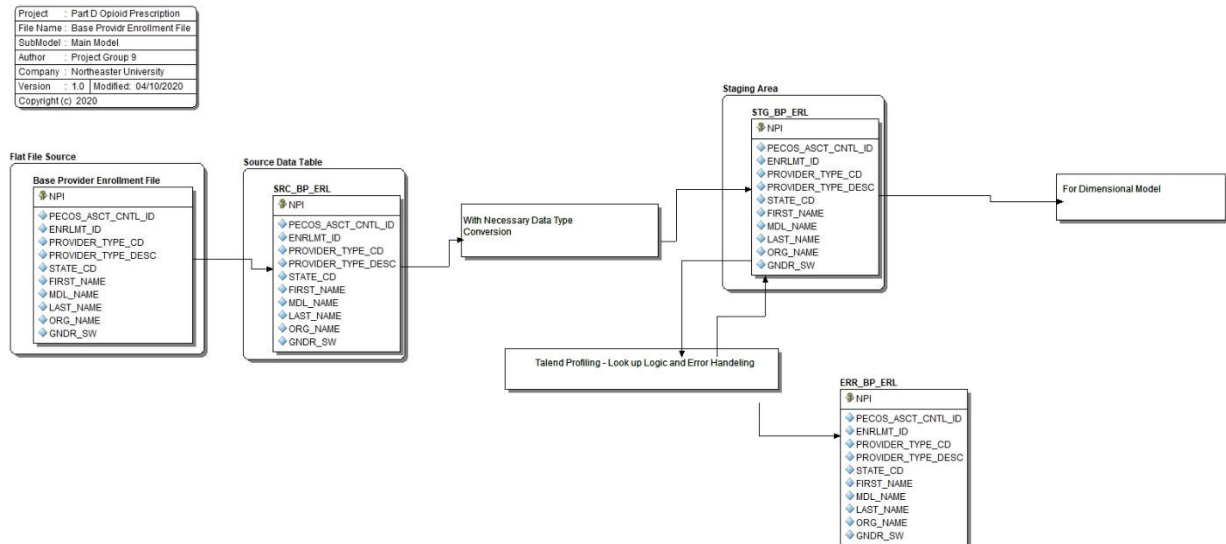
Medicare Specialty Code Data Flow:

Project IE7290 – Data Warehouse and Business Intelligence.

Tools Intended to use

Seral Number	Tool	Use Purpose
1	MS SQL Server Management Studio.	For having tables and storing Data
2	MS SQL Server Integration Services	For Having Perioding Loading, Error Handling, Staging Logic, Archive Logic, Slowly Changing Dimensions.
3	Talend Data Quality	Analyzing Data.
4	MS SQL Analysis Services	For Creating Multidimensional Warehouse (Star Schema)
5	MS SQL Reporting Services	Periodic data refresh for Visualization
6	Tableau/Power BI	Visualization

Visualization Techniques

- **Charts:**

   Relevant charts are displayed such as maps for geographical information, number charts for   Key metrics, Line charts to display change over time etc.

   Example: Prescriber count increase/decrease over the years or drug usage over the years

- **Summary reports:**

   Crucial high-level overview of Key important metrics is displayed here.

   Example: Total Average Submitted charges, Total Average claim amount etc

- **Custom reports:**

   Reports are customized to suit the needs of the respective audience by show showing different views based on selection and granularity the customer chooses. Legends, filters are used to achieve the same.

- **Story view:**

   Logical narrative of the outcomes of the Dataset in the form of a Story.

Project IE7290 – Data Warehouse and Business Intelligence.

Business Goals (Set of Questions)

Main KPI

- Which provider type is prescribing the most opioid?
- Which gender is prescribed the most opioid?
- Which state has the highest and least usage of opioid?
- What is the highest prescribed opioid name?

Analysis and Trends

- Analyzing the drug quantity prescribed over different years to help stakeholders limit drug prescription based on quantity limits
- Number of unique drugs added each year
- Drug cost analysis, Understanding the least and most expensive drugs over different years and help providers decide if least expensive drug can be used an alternative option.
- Aggregate cost paid by beneficiaries for different drugs based on status (Beneficiaries 65+, Low income subsidy, No Low-income subsidy)
- Number of beneficiaries, claims and prescribers for different drugs
- Analyzing the suppressed data

References

1. https://www.cms.gov/Medicare/Medicare
2. https://www.fda.gov/drugs/drug-approvals-and-databases/national-drug-code-directory