Efficient & Accurate Reporting with a Dedicated Telehealth Data Mart

MUSC Health
Medical University of South Carolina

Katie Kirchoff, MSHI; Ryan Kruis, PhD, MSW, LISW-CP; Katharine Larson, MSHI

ABSTRACT

Given the nature of telehealth programs, data is collected and stored in disparate systems owned by numerous telehealth vendors, separate institutions participating in services and various institutional departments who own different reporting databases and software systems. Several challenges result from this architecture including inconsistent methodologies used to merge and clean data sets, variation in metrics reported depending on the source of the data, and redundant work required to build different reporting outputs.

At the Medical University of South Carolina, (MUSC) Center for Telehealth, a dedicated data mart was designed for our telehealth programs to create a more efficient, consistent, and accurate reporting process to address these challenges. Data are obtained numerous ways including: vendors sharing flat files and transferring them to MUSC via Secure File Transfer Protocol (SFTP); data are transferred from a direct database connection from a software system's database; or files are manually downloaded from software system. Data are then loaded and stored in a SQL Server database and cleaned through a set of Extract, Transform, Load (ETL) processes built in SQL Server Integration Services (SSIS) components through Microsoft Visual Studio. Merged and cleaned data sets are built and stored in the database from the source data loaded in the data mart. The same cleaned data are then used for all subsequently built reports in different analytical tools such as Tableau or Microsoft BI.

METHODS

For each new data source, analysts must complete the following steps

1. Obtain access to the data source
2. Define data elements of interest
3. Create DDL (Data Definition Language) to store data
4. Create transformation process from data at source to target data format
5. Load historical data as one time test of processes
6. Validate data was loaded correctly
7. Determine refresh frequency and type (incremental or full refresh)
8. Create and test refresh process
9. Automate process
10. Add error handling and notifications to automation process

Naming Conventions

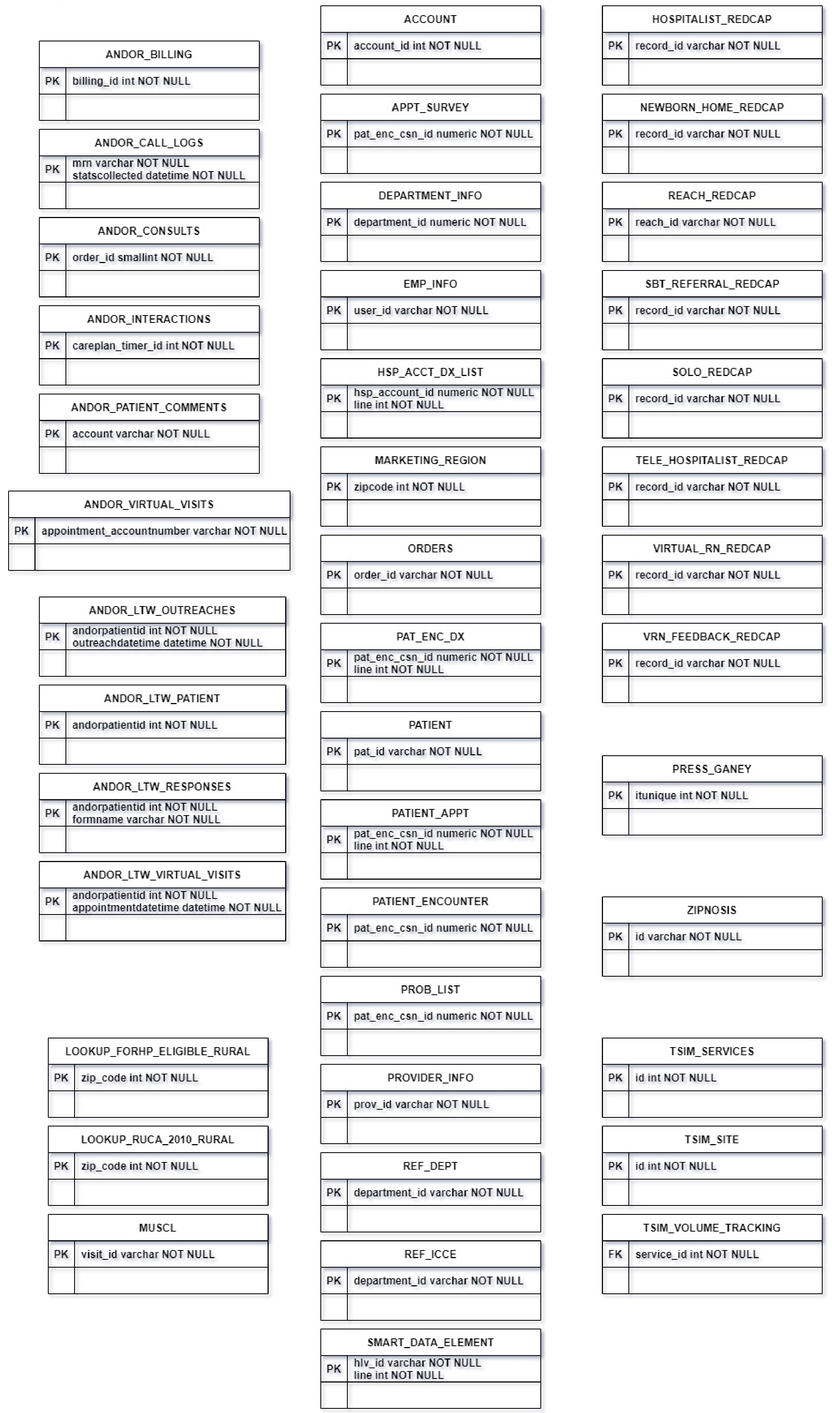
All table names include the source system of the data when feasible to know which software system the data is from. Tables created during the ETL process to clean or merge data from multiple source systems together do not have a source related table name. When possible, we create column names that are consistent with how they are named in the source data.

Security and Access

Only database administrators and reporting analysts have user account access to the data mart. SQL knowledge is required to obtain this access. For managers, leaders, vendors and other stakeholders we build and provide data feeds or business intelligence reporting dashboards that connect to the database for self-service access to the data. The database is backed up weekly and user accounts are reviewed yearly. A reporting user was created for business intelligence tools to connect to the data mart.

Entity Relationship Diagram (ERD)

Entity relationship diagrams show the structure of the objects that make up a database. Each rectangle below represents a table in our telehealth data mart with PK specifying the primary key or columns that form a unique identifier for each row of data within the table.



Future Plans

Data Marts require ongoing maintenance from both a technical and workflow prospective. As vendors change or new vendors are utilized, new feeds to and from the data mart will be required to enhance reporting capabilities. Backups, storage requirements and ETL processes will need to be monitored and updated periodically. We also anticipate moving the data mart from an on-premise server to cloud based technology on Microsoft Azure in the next year or two in line with enterprise initiatives.

More immediate enhancements include:

1. Adding in additional billing data from EPIC to help monitor billing compliance and improve operational processes and ROI reporting
2. Adding Vizient benchmarking data
3. Adding Neuroflow vendor data
4. Adding new Andor data as functionality continues to be developed on the platform
5. Increasing frequency of data refreshes
6. Building tableau self-service dashboards for stakeholders to have consistent and immediate access to the data

CONCLUSIONS

This work reduces the amount of effort needed from analysts to build reports and ensures all reports utilize consistent definitions and methodologies to present the data to stakeholders. Validated and consistent reporting allows stakeholder focus to shift from investigating and explaining data discrepancies between data sources to utilizing reports for data driven decision making with confidence.