**Topic**: Measure the number of Influenza cases in U.S using Data management and masking on Azure

Problem Statement: Although most organizations have stringent security controls in place to protect production data in storage or in business use, sometimes that same data has been used for operations that are less secure. The issue is often compounded if these operations are outsourced and the organization has less control over the environment. In the wake of [compliance](http://searchdatamanagement.techtarget.com/definition/compliance) legislation, most organizations are no longer comfortable exposing real data unnecessarily. In data masking, the format of data remains the same; only the values are changed.

In this example we are going to demonstrate how to mask the patient data and make it available for the analytics. We are going to measure the number of Influenza cases in 2018 in all the United States without comprising any of the patient PHI.

Overview of technology: Azure dynamic data masking is Microsoft’s cloud based platform that limits sensitive data exposure by masking it to the non-privileged users.

Overview of steps:

1. Prepare a clean Patient data set with Influenza cases and save a CSV
2. Create a SQL server and a sql database
3. Create tables and multiple users with access to the newly created database
4. Create azure storage account and container to hold your source data files
5. Create Azure data factory to load the data from flat file into the SQL database
6. Mask the data on the sql server for the required users using Azure Data Masking
7. Connect to Sql Server DB using Tableau and import all the required tables to measure the Influenza cases
8. Build a map report to show the number of Influenza cases in different states of U.S.

Data set obtained at: <https://www.challenge.gov/challenge/patient-matching-algorithm-challenge/>

Hardware/OS: Intel Core i5-5300U CPU 2.30 GHZ, 12 GB RAM, 64 bit Windows 7 operating system

Software used: Azure, Visual Studio 2017, C#, SQL Server Management studio 2017, Tableau 10.5

Lessons learnt: Only dynamic data masking is possible/available. This doesn’t change the base data but just changes the view of the data based on the user access. This might be an issue in some cases where you need to mask the base data and not change the view of the data

Pros: Ease of design and implementation.

* No external tool/algorithm required for dynamic data masking.
* No constraint violation since underlying data is not changed and only the data view is changed
* Simplicity of Microsoft GUI
* Data Masking is not reversible for unauthorized users.

Cons:

* Data cannot be masked using custom dictionaries
* Limited customization available out of the box for dynamic data masking

YouTube Links:

* 2 Min: <https://youtu.be/8OWP0vaV6bc>
* 15 Min: <https://youtu.be/2ye9KoqC4Xo>
* GitHub Repository with all artifacts: <https://github.com/syenneti/FinalProject.git>