**Final Project**

**Data Management and Masking on Azure**

**Problem Statement**:

Cyber security is a big challenge in the current world. Most of the companies has a need to mask enterprise data to protect from outside intrusions and accidental loss of protected data. So, Data Masking is an important technology used to maintain data security even in case of external threats. Data exposure to the right user is essential to properly organize and maintain data processes to meet ongoing information lifecycle needs.

**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. You enable DDM by defining masking rules on designated database fields, which determine how you want the data in these fields to appear in query results. You can define a partial mask which exposes some of the data in the selected field such as the first and last few characters, while masking out the rest. Or, you can define a full mask which doesn't leave any of the data exposed and always replaces the field's data with a constant value.

**High Level Steps:**

1. Create a SQL server and a sql database
2. Create tables and multiple users with access to the newly created database
3. Create azure storage account and container to hold your source data flat files
4. Create Azure data factory to load the data from flat file into the SQL database
5. Mask the data using Azure Data Masking service on the sql server for the required users
6. View the data on Tableau or any other visualization tool

**Data set obtained at**:

<https://www.challenge.gov/challenge/patient-matching-algorithm-challenge/>

**Hardware/OS**:

Intel Core i5-5300U CPU 2.30 GHZ, 16 GB RAM, 64 bit Windows 7 operating system

**Software used**:

Visual Studio 2017, C#, SQL Server Management studio 2017, Tableau

**Links:**

* <https://youtu.be/6K-Yg_f0LiI>(2 min)
* <https://youtu.be/3jxA-Lr5IQU> (15min)
* <https://github.com/dilreddy143/AzureFinalProject>