Sujatha Kestur

02/06/2018

**Data management and masking on Azure**

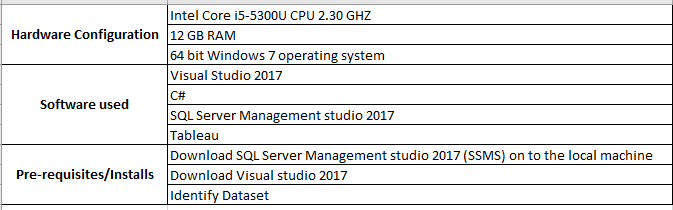
**Introduction**

Data masking is a method of creating a structurally similar but inauthentic version of an organization's data. The purpose is to protect the actual data while having a functional substitute for occasions when the real data is not required and to simultaneously maintain the usability of the data. It is used to reduce the unnecessary spread and exposure of sensitive data within or outside an organization.

**Problem Statement**

With umpteen attempts being done in the real world to hack the data and publicize it, it becomes even more important to protect the patient personal information such as Names, date of birth, SSN, address etc. Patient health information and insurance information. The data masking that is offered out of box by Azure only aims at changing the view of the data based on user access rights. The project demonstrates this fact by only showing masked data to unauthorized user. The data view for the authorized user proves that the actual underlying data still remains the same and only the data view changes

**Prerequisites**



**Implementation**

* Create a SQL server and a sql database
* Create tables and multiple users with access to the newly created database
* Create azure storage account and container to hold your source data flat files
* Create Azure data factory to load the data from flat file into the SQL database
* Mask the data on the sql server for the required users
* Data viewed by user on Tableau

**Data set**

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

Provider and Patient Metastasis – mocked up data

**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, simplicity of Microsoft GUI

**Cons** -- Data cannot be masked using custom dictionaries, limited customization available out of box