**De-Identify sensitive data in SQL**

**Summary:**

De-Identify sensitive data is a simple SQL logic which is built to satisfy the purpose of hiding sensitive information (in this case hospital patient PHI) but still keep the original link between data fields.

For example:

1. The middle 8 characters/digits of patient control number within a Medicare claim contains the patient hospital account number.

**Patient Control Number:** HAR-32555888-12HCNMCD

**Account Number:** 32555888

1. HIC number is a combination of Prefix + SSN + Suffix

**SSN:** 999889999

**HIC:** A999889999

Because of how the logic de-identifies the sensitive data, there is not a way to reverse the action.

**How it works.**

1. Creating a random map between characters and digits.
2. Gathering all unique values from fields (Identified by the user) which need to be de-identified into a temp table.
3. Going thru each characters/digits one by one and replacing the old character with the new character from the character map.
4. Updating the old values with the new de-identified values.
5. Dropping all temp table to avoid the ability to reverse logic.

**The logic characteristics:**

1. This characters and digits map changes every time the logic is refresh.
2. This characters and digits map can be expanded to all characters or simplified to only digits (or characters).
3. The logic can handle up to 255 in field length. This is not a hard limit.