

Die Struktur einer Tabelle - und wie diese erhalten bleibt!

Eine SAS-Tabelle besteht nicht nur aus Spalten und Zeilen.

Sie kann zum Beispiel sortiert, indiziert oder komprimiert sein und ein Label enthalten.

Bei der SPDE-Engine wird angegeben, wie groß einzelne Dateien im Dateisystem werden sollen.

Wenn nun Tabellen per Data-Step überschrieben werden ("in place"), gehen die Metainformationen verloren.

Im konkreten Anwendungsbeispiel (Berücksichtigung der Datenschutz-Grundverordnung (DSGVO)) müssen viele Tabellen eines Data Warehouse betrachtet werden.

Anhand von Schlüsselinformationen (z.B. Vertragsnummern) sollen Datensätze gelöscht oder anonymisiert werden.

Damit jede Tabelle (aus Gründen der Performance) nur genau einmal gelesen und geschrieben wird, empfiehlt sich die Nutzung des Data-Steps.

Dafür müssen die Metainformationen vorher gespeichert und beim Schreiben der Tabelle berücksichtigt werden.

Für die Nachvollziehbarkeit ist eine Protokollierung unabdingbar, was ebenfalls für den Data-Step spricht (und gegen z.B. PROC SQL).

In dem Vortrag werden die wichtigsten Metainformationen vorgestellt und gezeigt, wie diese ermittelt und erhalten werden können.

Aufgrund der hohen Anzahl Tabellen muss dieses generisch unter der Verwendung von SAS Makros passieren.

Verwendet werden die BASE- und die SPDE-Engine.