LBS
Bausparkasse der Sparkassen

Die Struktur einer Tabelle – und wie diese erhalten bleibt!

Christian Kothenschulte

Wiesbaden | September 2022 | KSFE

Die LBS West in Zahlen

Eine der größten deutschen Bausparkassen stellt sich vor

- Spezialist im Sparkassenverbund für Bau, Kauf, Umschuldung und Modernisierung von Immobilien
- 1,6 Mio. Kunden mit 2,0 Mio. Verträgen über eine Bausparsumme von 64 Mrd. Euro.
- Marktanteil von rund 39 % im Geschäftsgebiet NRW und Bremen
- Spareinlagen > 13 Mrd. Euro
- Bilanzsumme 14,8 Mrd. Euro = eine der größten deutschen Bausparkassen
- Eigentümer zu je 50 % Sparkassenverband Westfalen-Lippe und Rheinischer Sparkassen- und Giroverband

1. **Die Struktur einer Tabelle –**
2. **und wie diese erhalten bleibt!**

Meine Ziele für heute!

Agenda

1 Motivation

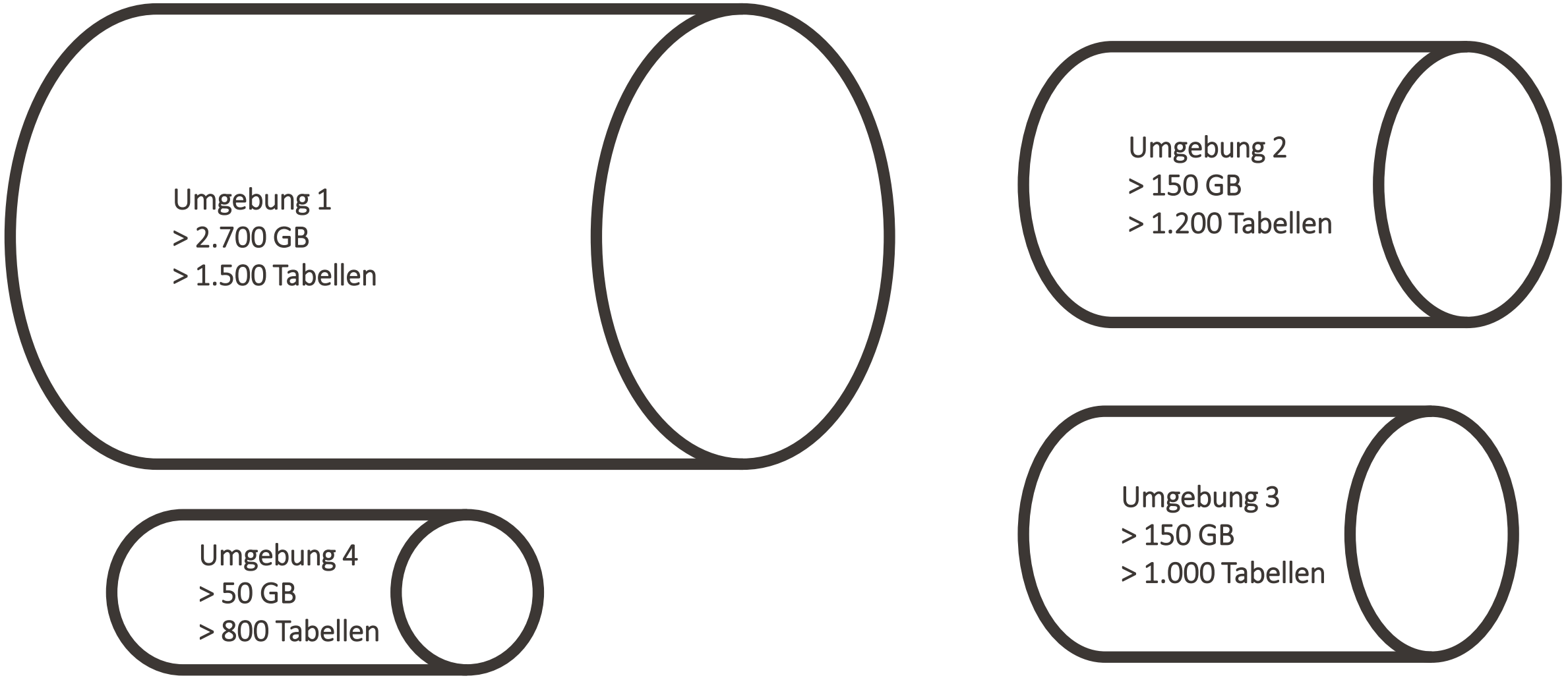
2 Struktur einer Tabelle

3 Beibehaltung der Struktur

4 Anwendungsfall als Beispiel

5 Fazit

Warum sollte man darüber nachdenken?



Eine Tabelle in SAS

Was man sieht:

Obs	SchuelerID	Name	Sex	Age	Height	Weight
1	1	Alfred	M	14	69.0	112.5
2	2	Alice	F	13	56.5	84.0
3	3	Barbara	F	13	65.3	98.0
4	4	Carol	F	14	62.8	102.5
5	5	Henry	M	14	63.5	102.5
6	6	James	M	12	57.3	83.0
7	7	Jane	F	12	59.8	84.5
8	8	Janet	F	15	62.5	112.5
9	9	Jeffrey	M	13	62.5	84.0
10	10	John	M	12	59.0	99.5
11	11	Joyce	F	11	51.3	50.5
12	12	Judy	F	14	64.3	90.0
13	13	Louise	F	12	56.3	77.0
14	14	Mary	F	15	66.5	112.0
15	15	Philip	M	16	72.0	150.0
16	16	Robert	M	12	64.8	128.0
17	17	Ronald	M	15	67.0	133.0
18	18	Thomas	M	11	57.5	85.0
19	19	William	M	15	66.5	112.0

```
data WORK.CLASS;
   SchuelerID = _n_;
   set SASHELP.CLASS;
run;
```


Datenquelle: SASHELP.CLASS

Anwendungsfälle mit Auswirkungen auf die Tabellenstruktur

Mögliche Anpassungen aus fachlicher Sicht (Auszug)

Obs	SchuelerID	Name	Sex	Age	Height	Weight
1	1	Alfred	M	14	69.0	112.5
2	2	Alice	F	13	56.5	84.0
3	3	Barbara	F	13	65.3	98.0
4	4	Carol	F	14	62.8	102.5
5	5	Henry	M	14	63.5	102.5
6	6	James	M	12	57.3	83.0
7	7	Jane	F	12	59.8	84.5
8	8	Janet	F	15	62.5	112.5
9	9	Jeffrey	M	13	62.5	84.0
10	10	John	M	12	59.0	99.5
11	11	Joyce	F	11	51.3	50.5
12	12	Judy	F	14	64.3	90.0
13	13	Louise	F	12	56.3	77.0
14	14	Mary	F	15	66.5	112.0
15	15	Philip	M	16	72.0	150.0
16	16	Robert	M	12	64.8	128.0
17	17	Ronald	M	15	67.0	133.0
18	18	Thomas	M	11	57.5	85.0
19	19	William	M	15	66.5	112.0

Migrationen

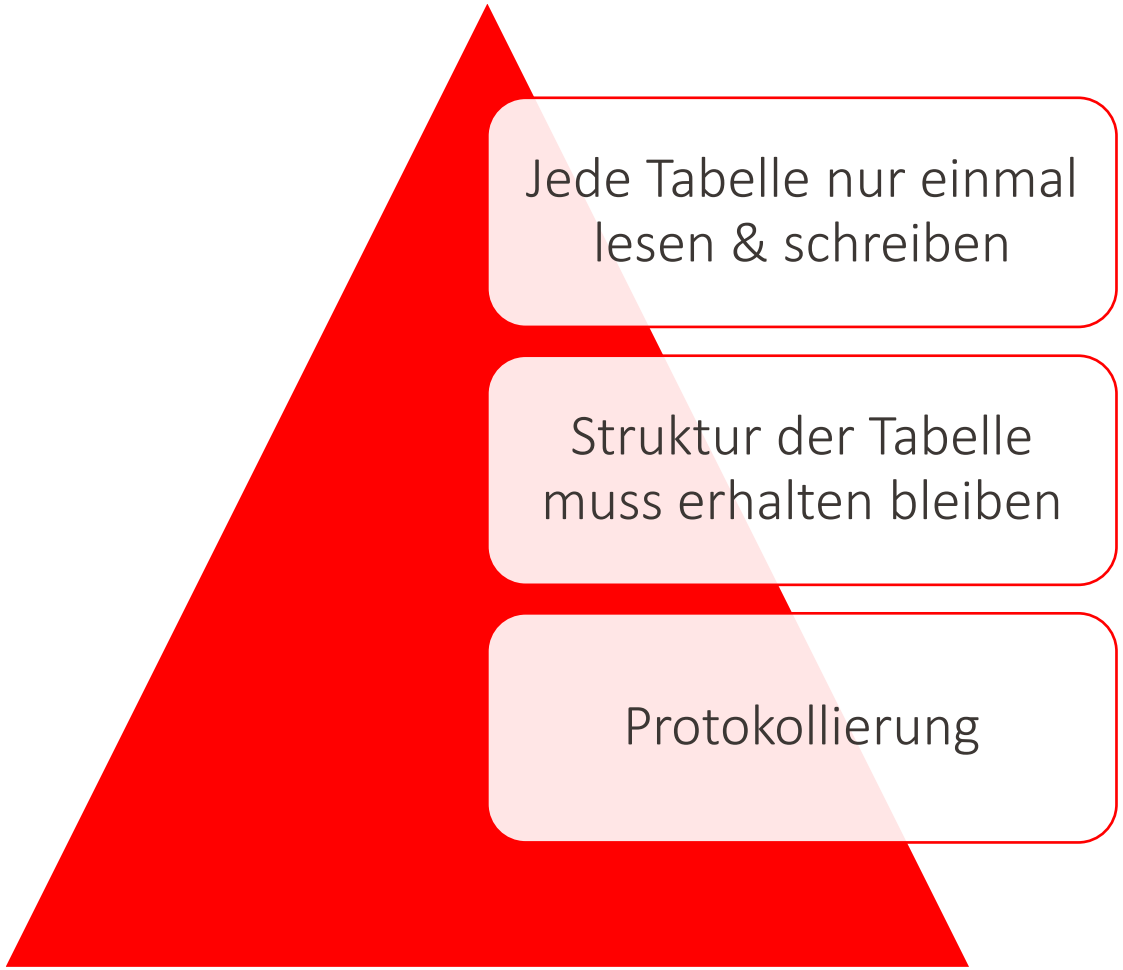
- Löschen von Spalten
- Hinzufügen von Zeilen
- Hinzufügen von Spalten (mit/ohne Inhalt)
- Spalte mit Logik aus der gleichen Zeile
- Spalte mit komplexer Logik
- Ermittlung aus weiteren Datenquellen

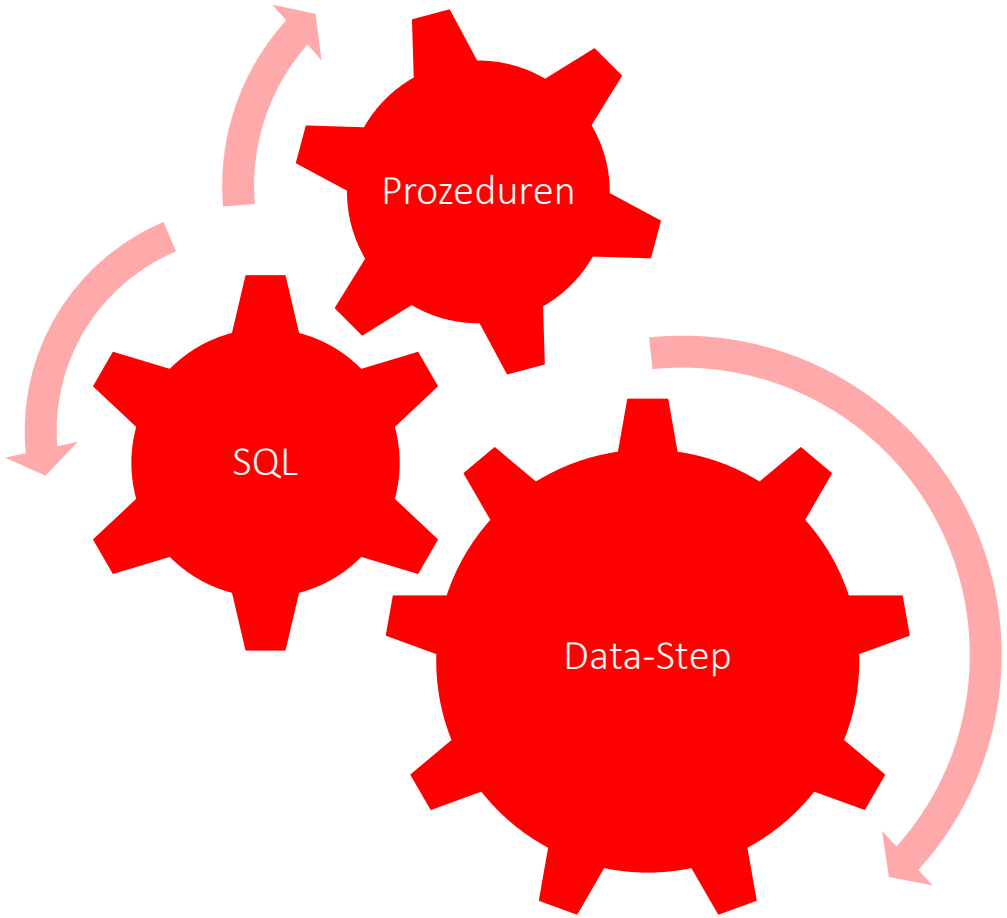
Berücksichtigung DSGVO

- Löschen von Zeilen
- Löschen von einzelnen Feldinhalten
- **Protokollierung**

Datenquelle: SASHELP.CLASS

Rahmenbedingungen / Anforderungen





Eine Tabelle in SAS

Was man nicht auf den ersten Blick sieht:

Obs	SchuelerID	Name	Sex	Age	Height	Weight
1	1	Alfred	M	14	69.0	112.5
2	2	Alice	F	13	56.5	84.0
3	3	Barbara	F	13	65.3	98.0
4	4	Carol	F	14	62.8	102.5
5	5	Henry	M	14	63.5	102.5
6	6	James	M	12	57.3	83.0
7	7	Jane	F	12	59.8	84.5
8	8	Janet	F	15	62.5	112.5
9	9	Jeffrey	M	13	62.5	84.0
10	10	John	M	12	59.0	99.5
11	11	Joyce	F	11	51.3	50.5
12	12	Judy	F	14	64.3	90.0
13	13	Louise	F	12	56.3	77.0
14	14	Mary	F	15	66.5	112.0
15	15	Philip	M	16	72.0	150.0
16	16	Robert	M	12	64.8	128.0
17	17	Ronald	M	15	67.0	133.0
18	18	Thomas	M	11	57.5	85.0
19	19	William	M	15	66.5	112.0

```
data WORK.CLASS
 (label="Klasse 8c" index=(AGE) sortedby=SchuelerID);
   SchuelerID = _n_;
   set SASHELP.CLASS;
run;
```


Datenquelle: SASHELP.CLASS

Strukturinformationen einer Tabelle
Grafische Ausgabe

PROC CONTENTS 1/3

The CONTENTS Procedure

Data Set Name	WORK.CLASS	**Observations**	19
Member Type	DATA	**Variables**	6
Engine	V9	**Indexes**	1
Created	04.06.2022 12:16:48	**Observation Length**	48
Last Modified	04.06.2022 12:16:48	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	YES
Label	Klasse 8c		
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

Strukturinformationen einer Tabelle
Grafische Ausgabe

PROC CONTENTS 2/3

Engine/Host Dependent Information	
Data Set Page Size	65536
Number of Data Set Pages	2
First Data Page	1
Max Obs per Page	1361
Obs in First Data Page	19
Index File Page Size	4096
Number of Index File Pages	2
Number of Data Set Repairs	0
ExtendObsCounter	YES
Filename	D:\WORK\ _TD14932_ _\Prc2\class.sas7bdat
Release Created	9.0401M7
Host Created	X64_SRV16
Owner Name	WIN\
File Size	192KB
File Size (bytes)	196608

Datenquelle: SASHELP.CLASS

Strukturinformationen einer Tabelle Grafische Ausgabe

PROC CONTENTS 3/3

Alphabetic List of Variables and Attributes

#	Variable	Type	Len
4	Age	Num	8
5	Height	Num	8
2	Name	Char	8
1	SchuelerID	Num	8
3	Sex	Char	1
6	Weight	Num	8

Alphabetic List of Indexes and Attributes

#	Index	# of Unique Values
1	Age	6

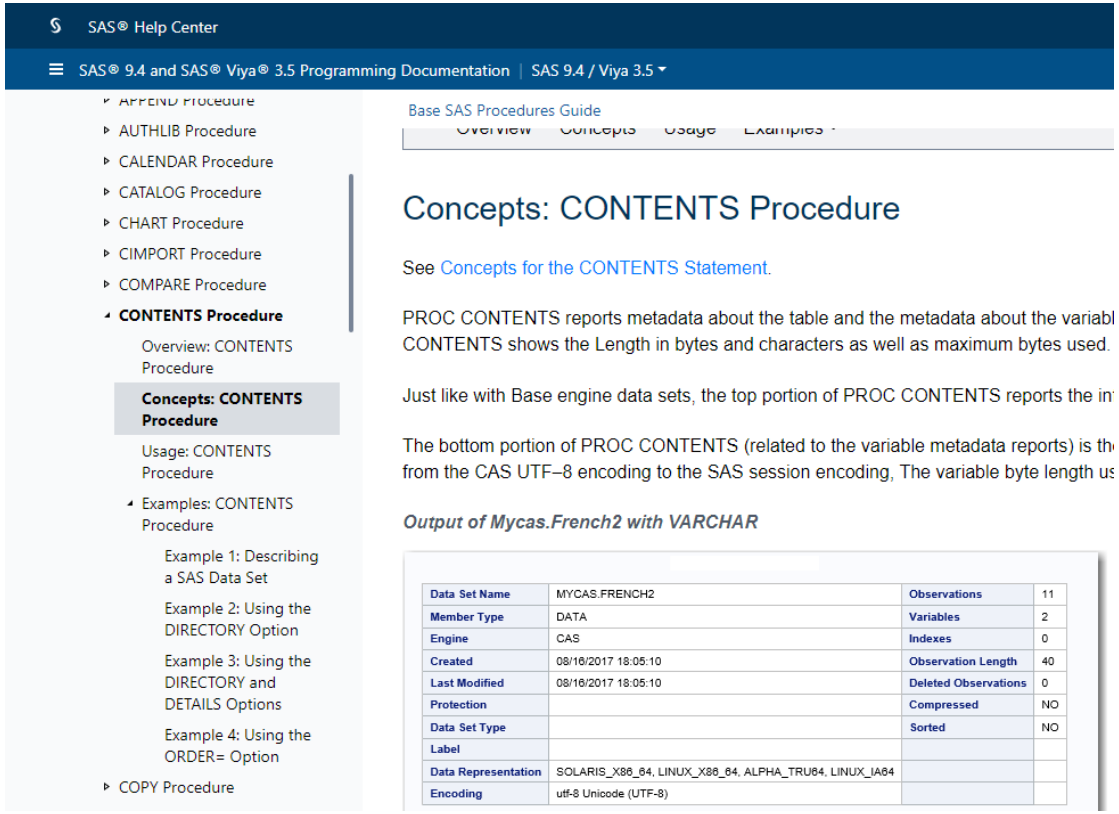
Sort Information

Sortedby	SchuelerID
Validated	NO
Character Set	ANSI

Strukturinformationen einer Tabelle
Ausgabe in Tabellen

```
proc contents data=WORK.CLASS noprint
              out=WORK.COLUMN
              out2=WORK.INDEXE;
run;
```

```
NOTE: The data set WORK.COLUMN has 6 observations and 41 variables.
NOTE: The data set WORK.INDEXE has 1 observations and 20 variables.
NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds
```



Strukturinformationen einer Tabelle
Ausgabe in Tabellen

```
proc contents data=WORK.CLASS NOPRINT
                  out=WORK.COLUMN
                      (keep=COMPRESS NAME SORTEDBY SORTED MEMLABEL ENGINE)
                  out2=WORK.INDEXE
                      (keep=TYPE RECREATE);
run;
```

COLUMN

Filtern und sortieren | Abfrage erstellen | Where | Daten | Beschreiben | Grafiken | Analy

	MEMLABEL	NAME	ENGINE	COMPRESS	SORTED	SORTEDBY
1	Klasse 8c	Age	V9	NO	0	.
2	Klasse 8c	Height	V9	NO	0	.
3	Klasse 8c	Name	V9	NO	0	.
4	Klasse 8c	SchuelerID	V9	NO	0	1
5	Klasse 8c	Sex	V9	NO	0	.
6	Klasse 8c	Weight	V9	NO	0	.

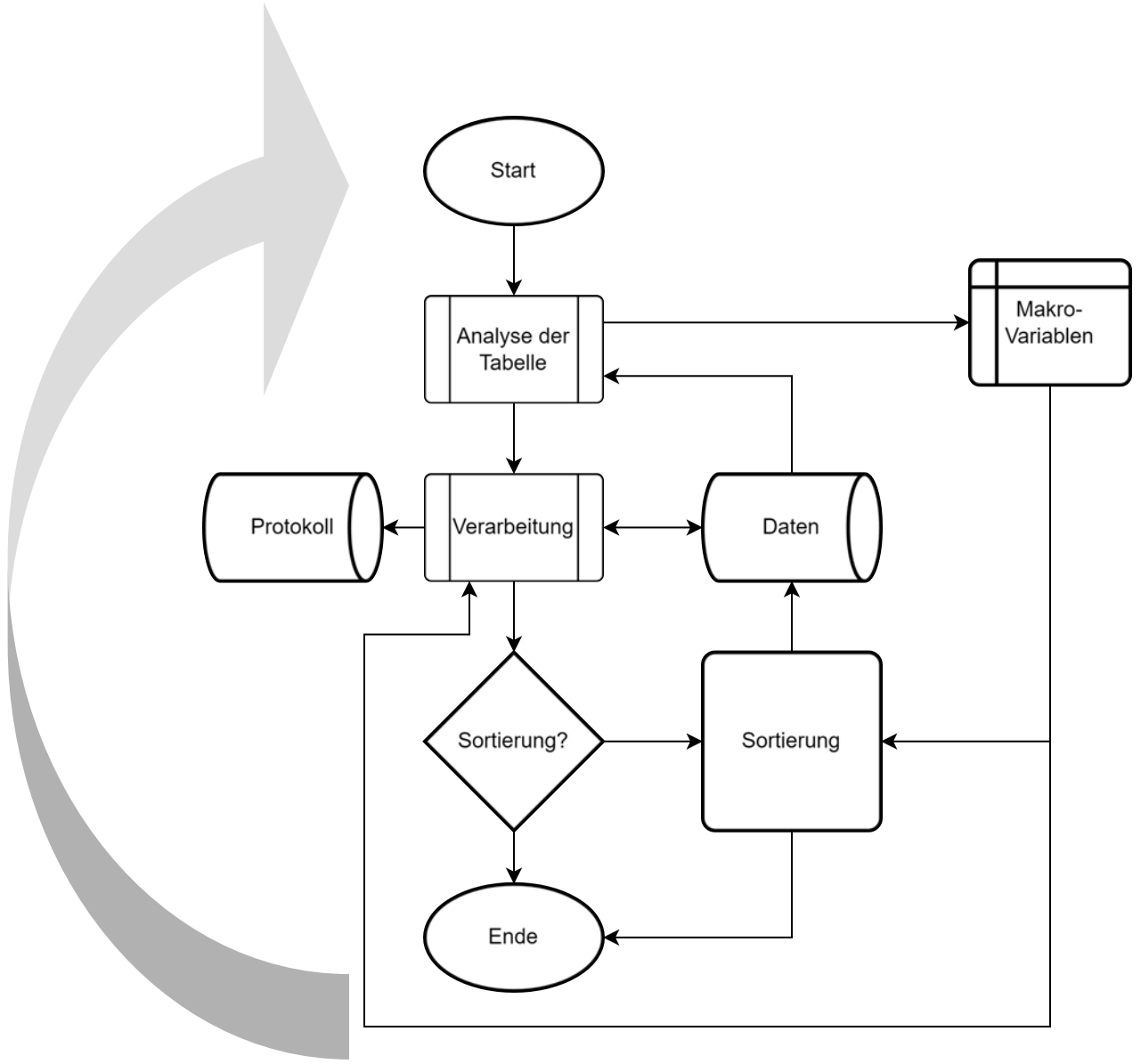
INDEXE

Filtern und sortieren | Abfrage erstellen | Whe

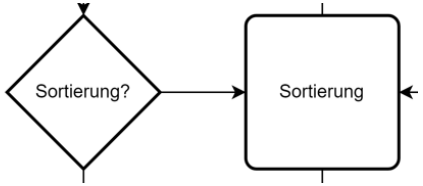
	Type	Recreate
1	Index	Index create Age / Updatecentiles=5;

Verarbeitung unter Beibehaltung der Struktur

Schematische Darstellung



Warum ist das Sortieren eine Besonderheit?



Sortierung sollte nachgelagert erfolgen!

Für Validierungskennzeichen (Validated=YES)

Sicherstellung der Sortierung
=> Auswirkung von Datenmanipulation

Berücksichtigung Index

```
ERROR: Indexed data set cannot be sorted in place unless the FORCE option is used.
NOTE: The SAS System stopped processing this step because of errors.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds
```

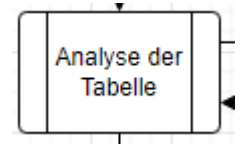
Analyse und Verarbeitung am konkreten Beispiel

Inhalt

Strukturmerkmal	Im Beispiel enthalten
Tabellenname	X
Bibliothek	X
Spalten	X
Zeilen	X
Protokollierung	X
Index	X
Sortierung	
Unterscheidung Engine (Base oder SPDE (inklusive Partsize))	
Label	X
Komprimierung	X

Analyse und Verarbeitung am konkreten Beispiel

Analyse




```
/* Komprimierung und Label in globale Makrovariablen */
data _null_;
   set WORK.COLUMN (obs=1);
   call symputx("COM", strip(COMPRESS),'G');
   call symputx("LABEL", strip(MEMLABEL),'G');
run;
```

COLUMN

Filtern und sortieren | Abfrage erstellen

	MEMLABEL	NAME	COMPRESS
1	Klasse 8c	Age	NO
2	Klasse 8c	Height	NO
3	Klasse 8c	Name	NO
4	Klasse 8c	SchuelerID	NO
5	Klasse 8c	Sex	NO
6	Klasse 8c	Weight	NO

```
/* Indizes sichern */
data _null_;
    set WORK.INDEXE(where=(TYPE="Index")) end=EOF;

    index = scan(scan(substr(recreate,14),1,'/'),1,";");

    call symputx(compress("INDEX_" !! put(_N_,8.)), index, 'G' );

    if EOF then do; /* Anzahl merken */
      call symputx("ANZ_INDEX", strip(put(_N_,8.)),'G');
    end;
run;
```

INDEXE

Filtern und sortieren | Abfrage erstellen | Whe

	Type	Recreate
1	Index	Index create Age / Updatecentiles=5;

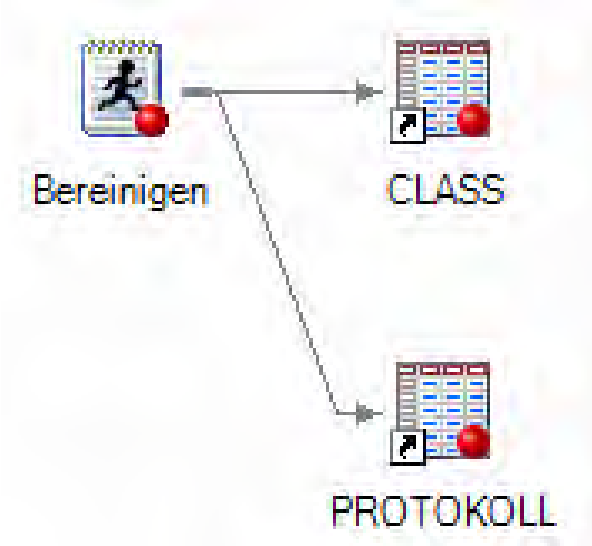
Analyse und Verarbeitung am konkreten Beispiel

Verarbeitung 1/3 / Kopf

```
data WORK.CLASS   (compress=&COM.   /* Komprimierung beibehalten */
                   label="&LABEL." /* Label beibehalten */
                   drop=Lauf Tabelle Gesamt Geloescht Nicht_Geloescht

 /* Indexe aufbauen... (falls keine Sortierung!) */
 index=(
           %do K=1 %to &ANZ_INDEX.;
            &&INDEX_&K.
           %end;
       )
 )

 WORK.PROTOKOLL (keep=Lauf Tabelle Gesamt Geloescht Nicht_Geloescht );
 ;
     set WORK.CLASS end=EOF;
```



Analyse und Verarbeitung am konkreten Beispiel

Verarbeitung 2/3 / Inhalt löschen


```
/* Einen Satz löschen */
if SCHUELERID = 9 then do;
   NAME = "***";
   DWH_GELOESCHT_GEM_DSGVO_MM = "X";
end;
```

	FELD	FRM	LABEL	DSGVO_PERSONENBEZUG_KZ
1	SCHUELER_ID	8.	ID des Schülers / der Schülerin zur eindeutigen Identifikation	S
2	NAME	$8.	Name	J
3	SEX	$1.	Geschlecht	J
4	AGE	8.	Alter	J
5	HEIGHT	8.	Größe	J
6	WEIGHT	8.	Gewicht	J

Dynamisch
per Hash-Objekt und Makro

- Schlüsselfelder (hier SchuelerID)
- Schlüsselwerte (hier 9)
- Relevantes Datenfeld (hier Name)
- Löschregel (hier = „***“)

Analyse und Verarbeitung am konkreten Beispiel

Verarbeitung 3/3 / Protokollierung


```
  /* Protokollfelder */
  attrib Lauf format=DDMMYYP10.;
  attrib Tabelle length=$32.;
  Lauf = today();
  Tabelle = "CLASS";
  retain Gesamt Nicht_Geloescht Geloescht 0;

  /* Zählen für das Protokoll */
  Gesamt+1;
  if DWH_GELOESCHT_GEM_DSGVO_MM eq 'X' then
    Geloescht+1;
  else Nicht_Geloescht+1;

  /* Datensatz ausgeben */
  output WORK.CLASS;

  if EOF then
  /* Zum Abschluss eine Protokollzeile rausschreiben */
    output WORK.PROTOKOLL;
run;
```

	SchuelerID	Name	DWH_GELOESCHT_GEM_DSGVO_MM
1	1	Alfred	
2	2	Alice	
3	3	Barbara	
4	4	Carol	
5	5	Henry	
6	6	James	
7	7	Jane	
8	8	Janet	
9	9	***	X
10	10	John	
11	11	Joyce	

Auszug ohne die Spalten *Sex*, *Age*, *Height* und *Weight*

PROTOKOLL

Filtern und sortieren | Abfrage erstellen | Where | Daten | Beschreiben | Gra

	Lauf	Tabelle	Gesamt	Nicht_Geloescht	Geloescht
1	24.06.2022	CLASS	19	18	1

Unterschied BASE-Engine und SPDE

Engine	V9

class	27.06.2022 13:19	SAS Data Set	192 KB
class	27.06.2022 13:19	SAS Data Set Index	9 KB

Engine/Host Dependent Information	
Blocking Factor (obs/block)	365
Data Partsize	4294949632bytes

Name	Größe
alle_konten.dpf.94b29281.0.1.spds9	4.194.287 KB
alle_konten.dpf.94b29281.1.1.spds9	4.194.287 KB
alle_konten.dpf.94b29281.2.1.spds9	4.194.287 KB
alle_konten.dpf.94b29281.3.1.spds9	4.194.287 KB
alle_konten.dpf.94b29281.4.1.spds9	4.194.287 KB
alle_konten.dpf.94b29281.5.1.spds9	4.194.287 KB
alle_konten.dpf.94b29281.6.1.spds9	4.194.287 KB
alle_konten.dpf.94b29281.7.1.spds9	1.863.477 KB
alle_konten.mdf.0.0.0.spds9	78 KB

```
data KONTEN.ALLE_KONTEN (PARTSIZE=4G);
```

Engine/Host Dependent Information	
Blocking Factor (obs/block)	365
Data Partsize	134179840

Name	Größe
alle_konten.dpf.94b29281.0.1.spds9	131.035 KB
alle_konten.dpf.94b29281.1.1.spds9	131.035 KB
alle_konten.dpf.94b29281.2.1.spds9	131.035 KB
alle_konten.dpf.94b29281.3.1.spds9	131.035 KB
alle_konten.dpf.94b29281.4.1.spds9	131.035 KB
alle_konten.dpf.94b29281.5.1.spds9	131.035 KB
alle_konten.dpf.94b29281.6.1.spds9	131.035 KB
alle_konten.dpf.94b29281.7.1.spds9	131.035 KB
alle_konten.dpf.94b29281.8.1.spds9	131.035 KB
alle_konten.dpf.94b29281.9.1.spds9	131.035 KB
alle_konten.dpf.94b29281.10.1.spds9	131.035 KB

Sortierung

```
proc sort data=SASHELP.CLASS noduprecs
   out=WORK.CLASS;
by AGE descending WEIGHT NAME;
run;
```

Sort Information	
Sortedby	Age DESCENDING Weight Name
Validated	YES
Character Set	ANSI
Sort Option	NODUPREC

	NAME	SORTED	SORTEDBY	NODUPKEY	NODUPREC
1	Age	1	1	NO	YES
2	Height	1	.	NO	YES
3	Name	1	3	NO	YES
4	Sex	1	.	NO	YES
5	Weight	1	-2	NO	YES

Fazit

Die Struktur einer Tabelle -

- SAS-Prozedur *CONTENTS* liefert alle Informationen
- Heterogene Umgebungen erfordern generischen Ansatz
- Hohe Anzahl Tabellen ebenfalls

und wie diese erhalten bleibt!

- Parameter in Makrovariablen speichern
- Im Data-Step berücksichtigen (einmal lesen und schreiben!)
- Sonderbehandlung Sortierung

Auf Wiedersehen.

Kontakt

Christian Kothenschulte

LBS Westdeutsche Landesbausparkasse
Anstalt des Öffentlichen Rechts
52-50430 DWH-/BI- und OMS-Systeme
Himmelreichallee 40, 48149 Münster

Telefon
0251 – 412 – 3528

E-Mail
christian.kothenschulte@lbswest.de