SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Fang Meng, SAS Institute Inc.

Fang Meng is Test at SAS Institute Inc. Beijing R&D. I work in SAS solution validation and implementation support in CI products.

Elizabeth Bales, SAS Institute Inc.

Elizabeth is a Senior Software Manager in the Platform Research and Development Division at SAS Institute. Elizabeth manages the Software Globalization R&D team which provides Internationalization (I18N) features for SAS products and tools that find I18N violations in software.

Using CVP for the Best Migration Experience

Fang Meng, SAS Institute Inc.

USERS PROGRAM

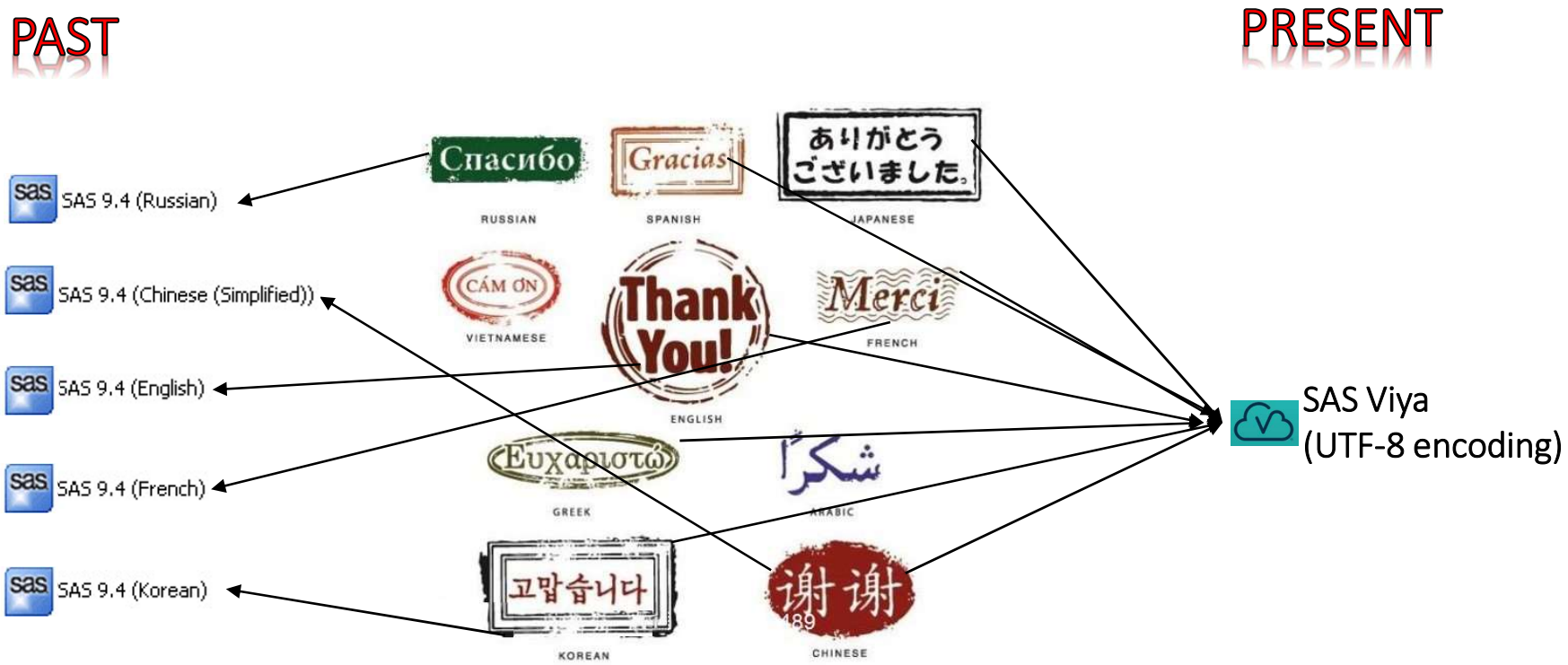
SAS® GLOBAL FORUM 2020

Outline

- Transcode character variables to SAS Viya with CEDA
- Expand the character variables space with CVP engine in SAS Viya
- Migrate format catalogs to SAS Viya
- Migrate indexes to SAS Viya
- Demo

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Background



SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Transcode character variables to UTF-8 encoding with CEDA(1/2)

- Cross Environment Data Access (CEDA)
  - Transcode the characters in input data from its original encoding to the current session encoding if the encodings are not the same.
  - Run automatically and transparently.

```
 ENCODING=UTF-8      Specifies the default character-set encoding for the SAS session.
NOTE: PROCEDURE OPTIONS used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


75         /* Print observations of a SAS data set with euc-cn encoding in SAS Viya with UTF8 session */
76         proc print data = euclib.class_cn;
NOTE: Data file EUCLIB.CLASS_CN.DATA is in a format that is native to another host, or the file encoding does not match the session
      encoding. Cross Environment Data Access will be used, which might require additional CPU resources and might reduce
      performance.
77         run;
```


SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Transcode character variables to UTF-8 encoding with CEDA(2/2)

- Truncation warnings or errors occur when the length of character variable is not enough to store the UTF-8 characters.

Euro symbol	WLATIN1	UTF-8
€	80	E282AC
	1 byte	3 bytes

Using CVP engine to expand the space of character variables.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Using CVP Engine to expand the space (1/2)

- Character Variables Padding (CVP) engine expands the length of character variables. By default, the expansion amount is 1.5 times in 9.4M5, and 0(AUTO) in SAS Viya.

#	Variable	Type	Len	Format
5	Cylinders	Num	8	BEST12.
8	DriveTrain	Char	9	$DRIVEN.
4	EngineSize	Num	8	BEST12.
6	Horsepower	Num	8	BEST12.
3	MSRP	Num	8	NLMNY16.
1	Make	Char	37	$15.
2	Model	Char	40	$39.
7	Origin	Char	4	$4.

CVP LIBNAME

```
libname euclib2 CVP "/home/sasdemo/demoscript/data";
proc contents data=euclib2.carsinfo;
run;
```

#	Variable	Type	Len	Format
5	Cylinders	Num	8	BEST12.
8	DriveTrain	Char	14	$DRIVEN.
4	EngineSize	Num	8	BEST12.
6	Horsepower	Num	8	BEST12.
3	MSRP	Num	8	NLMNY16.
1	Make	Char	56	$23.
2	Model	Char	60	$59.
7	Origin	Char	6	$6.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Using CVP Engine to expand the space (2/2)

- CVPMULTIPLIER can specify the expansion amount.
- CVPBYTES can specify the number of bytes.
- CVPFORMATWIDTH can specify whether to expand the character format width.

WLATIN1 1 byte	Character	Column Length
	€	1

UTF-8 3 byte	Character	Column Length
	€	3

```
/* Expand the character variables length 3 times*/
libname euclib2 CVP "/home/sasdemo/demoscript/data" CVPMULTIPLIER=3;
```


SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Migrate Format Catalog to UTF-8 (1/2)

- CEDA does not transcode characters in format catalog.
- If you created user-defined formats in another encoding, you will need to migrate those to the SAS Viya manually.

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
5	Cylinders	Num	8	BEST12.	BEST32.	气缸
8	DriveTrain	Char	9	$DRIVEN.	$9.	传动系
4	EngineSize	Num	8	BEST12.	BEST32.	发动机尺寸（L）
6	Horsepower	Num	8	BEST12.	BEST32.	马力
3	MSRP	Num	8	NLMNY16.	BEST32.	建议零售价
1	Make	Char	37	$15.	$15.	制造商
2	Model	Char	40	$39.	$39.	型号
7	Origin	Char	4	$4.	$4.	原产

```
/*
 * The definition of
 * the user-defined format.
*/

proc format lib=euclib;
value $driven 'D'='双驱动'
              'F'='前驱动'
              'R'='后驱动';
run;
```


SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Migrate Format Catalog to UTF-8 (2/2)

- Steps to migrate the format catalog to UTF-8 session.

1. Output format catalog to SAS data set in the SAS session where you created the format.

```
/* Create format data set in EUC-CN Session */
libname euclib "C:\SGF2020\use_cvp_to_migrate_to_viya";
proc format cntlout=euclib.cnformat lib=euclib;
run;

data euclib.cnformat;
  set euclib.cnformat;
  keep FMTNAME START LABEL TYPE;
run;
```

2. Re-create the format using the outputted SAS data set in SAS Viya

```
/* Run the code in SAS Viya with UTF-8 encoding */
libname euclib2 CVP "/home/sasdemo/demoscript/data";
libname u8lib "/home/sasdemo/demoscript/u8_data";

proc format cntlin=euclib2.cnformat lib=u8lib;
run;
```


SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Migrate indexes (1/2)

- CEDA does not always transcode characters in indexes.
- To migrate indexes, it is required to re-create them in UTF-8.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Migrate indexes (2/2)

- Steps to migrate the indexes to UTF-8 session.

1. Output SAS index to a SAS data set in the SAS session where you created the SAS index.

```
/* Output the index in EUC-CN session */
proc datasets lib=euclib nolist;
   contents data=carsinfo out2= euclib.index;
quit;
```

2. Get the code from Recreate column.

Obs	Libname	Member	Numvars	Name	Type	Recreate
1	EUCLIB	CARSINFO	1	Make	Index	Index create Make / Updatecentiles=5;

3. Re-create the index for migrated UTF-8 SAS data set.

```
/* Recreate the index in UTF-8 session*/
proc datasets lib=u8lib nolist;
modify carsinfo;
Index create Make / Updatecentiles=5;
quit;
```


SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Demo

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Take Away

- Cross Environment Data Access
- CVP can expand the character variable length
  - CVPMULTIPLIER
  - CVPBYTES
  - CVPFORMATWIDTH
- SAS Format and index
  - Migrate manually (format procedure)
  - Migrate manually (datasets procedure)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Thank you!

Contact Information
fang.meng@sas.com

Reminder:

Complete your session survey in the conference mobile app.

MARCH 29 – APRIL 1 | WASHINGTON, DC