

DB2 for z/OS Unicode & CCSID



IBM Information
ON Demand 2009

DB2 z/OS 2009 Technical Conference

Marcelo Tamberlini “Tamber”

DB2 for z/OS Specialist

tamber@br.ibm.com

Versão 1.0

Disclaimer

- **O material deste Workshop foi criado com base em documentação oficial IBM.**
 - **O material deste Workshop não foi submetido a nenhuma validação formal ou revisão pela IBM.**
 - **Este material encontra-se “no estado” (“as-is”) e não tem garantias formais da IBM.**
- THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY.
 - WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED.
 - IN ADDITION, THIS INFORMATION IS BASED ON IBM’S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE.
 - IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION.
 - NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, OR SHALL HAVE THE EFFECT OF:
 - CREATING ANY WARRANTY OR REPRESENTATION FROM IBM (OR ITS AFFILIATES OR ITS OR THEIR SUPPLIERS AND/OR LICENSORS); OR
 - ALTERING THE TERMS AND CONDITIONS OF THE APPLICABLE LICENSE AGREEMENT GOVERNING THE USE OF IBM SOFTWARE.

Tópicos

- **Conceitos**
 - **CCSIDs**
 - **Unicode**
- **DB2 e CCSID**
- **DB2 e Unicode**
- **Referências**

Conceitos

Computadores só armazenam números.

Letras e outros caracteres são armazenados atribuindo-se números a eles.

Diferentes conjuntos de caracteres foram criados para padronização em diferentes plataformas computacionais.

A representação de um caractere em bytes é chamada Code Point.

Um conjunto de caracteres (Code Points) é chamado de Code Page.

Conceitos

Code Page

Conjunto de atribuições de caracteres a Code Points.

Code Points

Offsets (em hexadecimal) em um Code Page, representando um caractere específico.

CCSID

Número que identifica um particular Code Page.

HEX DIGITS	4-	5-	6-	7-	8-	9-	A-	B-	C-	D-	E-	F-
1ST →	4-	5-	6-	7-	8-	9-	A-	B-	C-	D-	E-	F-
2ND ↓	4-	5-	6-	7-	8-	9-	A-	B-	C-	D-	E-	F-
-0	(SP) SP010000	& SM030000	- SP100000	ø LO810000	Ø LO820000	° SM190000	μ SM170000	^ SD150000	{ SM110000	}	\ SM070000	0 ND100000
-1	(RSP) SP300000	é LE110000	/ SP120000	É LE120000	a LA010000	j LJ010000	~ SD190000	£ SG020000	A LA020000	J LJ020000	÷ SA080000	1 ND010000
-2	â LA160000	ê LE150000	Â LA180000	Ê LE160000	b LB010000	k Lk010000	s LS010000	¥ SC060000	B LB020000	K Lk020000	S LS020000	2 ND020000
-3	ä LA170000	ë LE170000	Ä LA190000	Ë LE180000	c LC010000	l LL010000	t LT010000	· SD830000	C LC020000	L LL020000	T LT020000	3 ND030000
-4	à LA130000	è LE130000	À LA140000	È LE140000	d LD010000	m LM010000	u LU010000	© SM520000	D LD020000	M LM020000	U LU020000	4 ND040000
-5	á LA110000	í LI110000	Á LA120000	Í LI120000	e LE010000	n LN010000	v LV010000	§ SM240000	E LE020000	N LN020000	V LV020000	5 ND050000
-6	ã LA190000	î LI150000	Ã LA200000	Ï LI160000	f LF010000	o LO010000	w LW010000	¶ SM250000	F LF020000	O LO020000	W LW020000	6 ND060000
-7	å LA270000	ï LI170000	Å LA280000	Ï LI180000	g LG010000	p LP010000	x LX010000	¼ NF040000	G LG020000	P LP020000	X LX020000	7 ND070000
-8	ç LC410000	ì LI130000	Ç LC420000	Ï LI140000	h LH010000	q LQ010000	y LY010000	½ NF010000	H LH020000	Q LQ020000	Y LY020000	8 ND080000
-9	ñ LN190000	ß LS810000	Ñ LN200000	` SD130000	i LI010000	r LR010000	z LZ010000	¾ NF050000	I LI020000	R LR020000	Z LZ020000	9 ND090000
-A	ø SC040000	! SP020000	 SM860000	: SP130000	« SP170000	ª SM210000	ï SP030000	[SM080000	(SHY) SP320000	1 ND011000	2 ND021000	3 ND031000
-B	· SP110000	\$ SC030000	, SP080000	# SM010000	» SP180000	º SM200000	¿ SP180000]	ô LO150000	û LU150000	ô LO160000	û LU180000
-C	< SA030000	* SM040000	% SM020000	@ SM050000	ð LD830000	æ LA510000	Ð LD620000	- SM150000	ö LO170000	ü LU170000	Ö LO180000	Ü LU180000
-D	(SP060000) SP070000	_ SP090000	' SP050000	ý LY110000	· SD410000	Ý LY120000	¨ SD170000	ò LO130000	ù LU130000	Ò LO140000	Ù LU140000
-E	+ SA010000	; SP140000	> SA050000	= SA040000	þ LT830000	Æ LA520000	Þ LT840000	' SD110000	ó LO110000	ú LU110000	Ó LO120000	Ú LU120000
-F	 SM130000	¬ SM880000	? SP150000	" SP040000	± SA020000	∩ SC010000	® SM530000	× SA070000	õ LO190000	ÿ LY170000	Ï LO200000	(EO)

Code Page 37

Conceitos

Encoding Scheme

Coleção de Code Pages para várias línguas usadas em uma determinada plataforma computacional.

ASCII (American Standard Code for Information Interchange)

Usado em sistemas Intel e Unix

EBCDIC (Extended Binary Coded Decimal Information Code)

Usado em z/OS e iSeries (AS/400)

0-9	A	B	C	D	E	F
0	1	2	3	4	5	6
7	8	9	A	B	C	D
E	F	G	H	I	J	K
L	M	N	O	P	Q	R
S	T	U	V	W	X	Y
Z	[]	^	_	`	{
	~	?	!	@	#	\$
%	&	'	()	*	+,-,=,/,>,<,~
~	~	~	~	~	~	~

Code Page 03700

EBCDIC

0-9	A	B	C	D	E	F
0	1	2	3	4	5	6
7	8	9	A	B	C	D
E	F	G	H	I	J	K
L	M	N	O	P	Q	R
S	T	U	V	W	X	Y
Z	[]	^	_	`	{
	~	?	!	@	#	\$
%	&	'	()	*	+,-,=,/,>,<,~
~	~	~	~	~	~	~

Code Page 03700

ASCII

Conceitos

Tipos de Code Pages / CCSIDs

Single Byte Character Set (SBCS CCSIDs)

1-byte Code Page, representam até 256 caracteres (Code Points)

Double Byte Character Set (DBCS CCSIDs)

2-byte Code Page, representam até 65536 caracteres (Code Points)

Mixed Code Pages

Usam mais de um Code Page para representar caracteres (Code Points)

Usados somente para representar caracteres Orientais, devido à quantidade e complexidade de seus símbolos.

(Japonês, Chinês e Coreano)

Representam caracteres Ocidentais

Code Page: 37
(US Latin English)

0	...	7	...	C	...	E	F
0		o	{	\	0		
1		É	A	+	1		
2		È	B	S	2		
3		É	C	T	3		
...							
B		#					
...							
E		=	ó	Ó	U		
F		"	õ	Ö			

Code point 7E

Code Page: 284
(Latin America Spanish)

0	...	7	...	C	...	E	F
0		o	{	\	0		
1		É	A	+	1		
2		È	B	S	2		
3		É	C	T	3		
...							
B		N					
...							
E		=	ó	Ó	U		
F		"	õ	Ö			

Code point 7B

Code Page: 930
(DBCS Japanese)

	00	98	E0	FF
00				
...				
41	A			
42	B			
...				
D0		解	解	
D1		英	猪	
...				
FF				

Code point 98D1

Code Page: 1399
(MBCS Japanese)

		C	F	
0		{			0	
1		A	00	98	E0	FF
2		B	00			
3		C	...			
...						
B		41		道	道	
		42		理	理	
...						
E		D0		信	英	
F		D1		保	保	
...						
FF						

MBCS code page 1

MBCS code page 2

Conceitos

Diferentes Encoding Schemes, diferentes CCSIDs

Mesmos caracteres, mesmos Code Points?

Encoding Scheme: EBCDIC

Code Page: CCSID 500

Caractere "A": Code Point x'C1'

EBCDIC Code Page – CCSID 500

HEX DIGITS 1ST → 2ND ↓	4-	5-	6-	7-	8-	9-	A-	B-	C-	D
-0	SP SP010000	& SM030000	- SP100000	ø LC610000	Ø LC620000	° SM190000	μ SM170000	¢ SC040000	{ SM140000	}
-1	(RSP) SP300000	é LE110000	/ SP120000	Ê LE120000	a LA010000	j LJ010000	~ SD190000	£ SC020000	A LA020000	J
-2	â LA150000	ê LE150000	Â LA160000	Ê LE160000	b LB010000	k LK010000	s LS010000	¥ SC050000	B LB020000	K
-3	ã LA170000	ë LE170000	Ã LA180000	Ë LE180000	c LC010000	l LL010000	t LT010000	· SD530000	C LC020000	I
-4	â LA130000	è LE130000	À LA140000	È LE140000	d LD010000	m LM010000	u LU010000	© SM200000	D LD020000	M
-5	á LA110000	í LI110000	Á LA120000	Í LI120000	e LE010000	n LN010000	v LV010000	§ SM240000	E LE020000	N
-6	ã LA190000	î LI190000	Ã LA200000	Ï LI160000	f LF010000	o LO010000	w LW010000	¶ SM250000	F LF020000	C
7	â LA100000	ï LI100000	Ä LA110000	Ï LI170000	σ LC010000	π LL010000	ϣ LV010000	¼ SM260000	G LF030000	F

Encoding Scheme:ASCII

Code Page: CCSID 850

Caractere "A": Code Point x'41'

ASCII Code Page – CCSID 850

HEX DIGITS 1ST → 2ND ↓	0-	1-	2-	3-	4-	5-	6-	7-	8
-0		▶ SM590000	(SP) SP010000	0 ND100000	@ SM150000	P LP020000	` SD130000	p LP010000	Ç LC42
-1	☺ SS000000	◀ SM630000	! SP020000	1 ND010000	A LA020000	Q LP030000	a LA010000	q LQ010000	Û LU17
-2	☹ SS010000	↑ SM760000	" SP040000	2 ND020000	B LB020000	R LR020000	b LB010000	r LR010000	É LE11
-3	♥ SS020000	!! SP330000	# SM010000	3 ND030000	C LC020000	S LS020000	c LC010000	s LS010000	Ê LA15
-4	♦ SS030000	¶ SM250000	\$ SC030000	4 ND040000	D LD020000	T LT020000	d LD010000	t LT010000	Ë LA17
-5	♣ SS040000	§ SM240000	% SM020000	5 ND050000	E LE020000	U LU020000	e LE010000	u LU010000	Ë LA13
-6	♠ SS050000	— SM700000	& SM030000	6 ND060000	F LF020000	V LV020000	f LF010000	v LV010000	Ë LA27
7	• SM700000	↑ SM700000	· SM030000	7 ND060000	G LG020000	W LW020000	g LG010000	w LW010000	Ë LA27

Conceitos

Mesmo Encoding Scheme, diferentes CCSIDs

Mesmos caracteres, mesmos Code Points?

HEX DIGITS 1st → 2nd ↓	4-	5-	6-	7-	8-	9-	A-	B-	C-	D-	E-	F-
-0	SP	&	-	ø	Ø	°	μ	^	{	}	\	0
-1	RS	é	/	Ê	a	j	~	£	A	J	÷	1
-2	ã	ê	Ã	Ê	b	k	s	¥	B	K	S	2
-3	ä	ë	Ä	Ë	c	l	t	·	C	L	T	3
-4	å	è	Å	È	d	m	u	©	D	M	U	4
-5	á	í	Á	Í	e	n	v	§	E	N	V	5
-6	â	î	Â	Î	f	o	w	¶	F	O	W	6
-7	ã	ï	Ã	Ï	g	p	x	¼	G	P	X	7
-8	ç	ì	Ç	Ì	h	q	y	½	H	Q	Y	8
-9	ñ	ó	Ñ	Ó	i	r	z	¾	I	R	Z	9
-A	é	!	!	:	«	#	i	[ñ	1	2	3
-B	.	\$,	#	»	º	¿]	ô	û	ô	û
-C	<	*	%	@	ð	æ	Ð	-	ö	ü	Ö	Ü
-D	()	'	´	ý	Ý	"	ò	ù	ò	ù	Ò
-E	+	;	>	=	þ	Æ	Þ	'	ó	ú	Ó	Ú
-F		~	?	"	±	±	®	×	ð	ÿ	Ö	oo

Code Page 00057

HEX DIGITS 1st → 2nd ↓	4-	5-	6-	7-	8-	9-	A-	B-	C-	D-	E-	F-
-0	SP	&	-	ø	Ø	°	μ	é	{	}	\	0
-1	RS	é	/	Ê	a	j	~	£	A	J	÷	1
-2	ã	ê	Ã	Ê	b	k	s	¥	B	K	S	2
-3	ä	ë	Ä	Ë	c	l	t	·	C	L	T	3
-4	å	è	Å	È	d	m	u	©	D	M	U	4
-5	á	í	Á	Í	e	n	v	§	E	N	V	5
-6	â	î	Â	Î	f	o	w	¶	F	O	W	6
-7	ã	ï	Ã	Ï	g	p	x	¼	G	P	X	7
-8	ç	ì	Ç	Ì	h	q	y	½	H	Q	Y	8
-9	ñ	ó	Ñ	Ó	i	r	z	¾	I	R	Z	9
-A	[]	!	:	«	#	i	~	ñ	1	2	3
-B	.	\$,	#	»	º	¿	!	ô	û	ô	û
-C	<	*	%	@	ð	æ	Ð	-	ö	ü	Ö	Ü
-D	()	'	´	ý	Ý	"	ò	ù	ò	ù	Ò
-E	+	;	>	=	þ	Æ	Þ	'	ó	ú	Ó	Ú
-F	!	^	?	"	±	±	®	×	ð	ÿ	Ö	oo

Code Page 00500

Conceitos

Collating Sequence

Sequência de ordenação dos caracteres.

- Em ASCII : espaço, números, maiúsculas, minúsculas.
- Em EBCDIC : espaço, minúsculas, maiúsculas, números.

Exemplos

```
SELECT NAME FROM TAMBER.EBCDIC_TB  
ORDER BY NAME;
```

```
NAME  
-----  
abcde  
TESTA  
TESTB  
TESTC  
TEST1  
TEST2  
TEST3
```

```
SELECT NAME FROM TAMBER.ASC_TB  
ORDER BY NAME;
```

```
NAME  
-----  
TEST1  
TEST2  
TEST3  
TESTA  
TESTB  
TESTC  
abcde
```

Conceitos

Unicode – Uma solução para a “Torre de Babel”



- Diferentes Encoding Schemes, diferentes Code Points, mesmos caracteres...
- Diferentes Encoding Schemes, diferentes caracteres, mesmos Code Points...
- Mesmos Encoding Schemes, mesmos caracteres, diferentes Code Points....

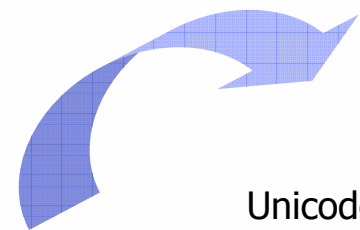
ᠠ	ᠢ
1874	1884
ᠡ	ᠣ
1875	1885
ᠣ	ᠣᠣᠣ
1876	1886

А	О	Й	К
046A	047A	049A	04BA
а	о	й	к
046B	047B	049B	04BB

路	磊	綾
F937	F941	F957
露	賂	菱
F939	F946	F95A
魯	雷	陵
F93A	F945	F953

1	A	Q
0031	0041	0051
2	B	R
0032	0042	0052
3	C	S
0033	0043	0053
4	D	T
0034	0044	0054

✂	☪	✪
2701	2711	2721
✂	☪	✪
2702	2712	2722



Unicode provê padrão único de mapeamento de Code Points para todos os caracteres conhecidos pela Raça Humana.

Independente de plataforma computacional, fornecedor, país etc.

Caracteres existentes não são removidos ou alterados, novos são sempre adicionados.



www.unicode.org

Conceitos

Unicode em System z

Unicode é um Encoding Scheme
Code Pages / CCSIDs...

Baseado em ASCII

Mesmo Collating Sequence

Primeiros 128 Code Points iguais

APAR **OA04069** provê suporte para configuração automática do Unicode.

Carrega conjunto de conversões default, o que requer aproximadamente 39 MB (9862 pages) de page-fixed central storage

A configuração personalizada normalmente requer menos que 3 MB (clientes americanos)

```

Session A - [24 x 80]
File Edit View Communication Actions Window Help
-----
Display Filter View Print Options Help
-----
SDSF STATUS DISPLAY ALL CLASSES          B1 RESPONSES NOT SHOWN
RESPONSE=SA23
CUN3000I 12.09.54 UNI DISPLAY 768
ENVIRONMENT:  CREATED      05/08/2005 AT 03.37.54
                MODIFIED    05/08/2005 AT 03.37.57
                IMAGE CREA 11/07/2003 AT 15.29.54
SERVICE: CHARACTER CASE NORMALIZATION COLLATION
STORAGE: ACTIVE 1274 PAGES
                LIMIT 524287 PAGES
CASECONV: SPECIAL
NORMALIZE: DISABLED
COLLATE:  DISABLED
CONVERSION: 01047-00037-ER 00037-01047-ER
              01047-00367-ER 00367-01047-ER
              00819-01208-ER 01208-00819-ER
              01047-01208-ER 01208-01047-ER
              01252-01208-ER 01208-01252-ER
              00037-00367-ER 00367-00037-ER
              01200-00300-ER 00300-01200-ER
COMMAND INPUT ==> /d uni,all          SCROLL ==> PAGE
F1=HELP      F2=SPLIT  F3=END      F4=RETURN  F5=IFIND   F6=BOOK
F7=UP        F8=DOWN   F9=SWAP    F10=LEFT  F11=RIGHT  F12=RETRIEVE
MB          22/031
-----
Connected to remote server/host.Tlba23me.torolab.ibm.com using lu/pool SA23036 and port 23
  
```

Serviço de conversão disponível no z/OS.

Conceitos

Unicode Consortium UTFs (Unicode Transformation Formats)

UTF-8 Desenhado para Internet

Cada caractere codificado de 1 a 4 bytes.

Primeiros 128 Code Points (x'00'a x'7F') são idênticos ao ASCII (1 byte)

UTF-16 Baseado em 16-bit code units

Alguns caracteres 1-byte code unit em UTF-8 são 2-byte code units UTF-16

Caracteres de 4 bytes, nenhuma diferença ao UTF-8

UTF-32 Todos caracteres representados em 4 bytes

Examples:

UTF - Unicode Transformation Format

Unicode	UTF-8
U+00000000 - U+0000007F	0xxxxxxx
U+00000080 - U+000007FF	110xxxxx 10xxxxxx
U+00000800 - U+0000FFFF	1110xxxx 10xxxxxx 10xxxxxx
U+00010000 - U+001FFFFF	11110xxx 10xxxxxx 10xxxxxx 10xxxxxx

Character	Unicode	Bit Pattern (binary)	UTF-8(binary) Half bytes in hexadecimal
'A'	U-41	1000001	01000001 4 1 => x'41'
'S'	U-A7	10100111	11000010 10100111 C 2 A 7 => x'C2A7'

Conceitos

Em Unicode, alguns caracteres são representados em 2+ bytes...

	000	001	002	003	004	005	006	007
0	NUL	DLE	SP	0	@	P	^	p
1	SOH	DC1	!	1	A	Q	a	q
2	STX	DC2	"	2	B	R	b	r
3	ETX	DC3	#	3	C	S	c	s
4	EDT	DC4	\$	4	D	T	d	t
5	ENO	NAK	%	5	E	U	e	u
6	ACK	SYN	&	6	F	V	f	v
7	BEL	ETB	'	7	G	W	g	w
8	BS	CAN	(8	H	X	h	x
9	HT	EM)	9	I	Y	i	y
A	LF	SUB	*	:	J	Z	j	z
B	VT	ESC	+	;	K	[k	{
C	FF	FS	,	<	L	\	l	
D	CR	GS	-	=	M]	m	}
E	SO	RS	.	>	N	^	n	~
F	SI	US	/	?	O	_	o	DEL

1 Byte

	008	009	00A	00B	00C	00D	00E	00F
0	XXX	DC5	NR SP	°	À	Đ	à	đ
1	XXX	PU1	ı	±	Á	Ñ	á	ñ
2	BPH	PU2	ç	²	Â	Ò	â	ò
3	NBH	STB	£	³	Ã	Ó	ã	ó
4	IND	CGH	¤	´	Ä	Ô	ä	ô
5	NEL	MW	¥	µ	Å	Õ	å	õ
6	SSA	SPA	ı	¶	Æ	Ö	æ	ö
7	ESA	EPA	§	·	Ç	×	ç	÷
8	HTS	DOB	¨	,	È	Ø	è	ø
9	HTZ	XXX	©	¹	É	Ù	é	ù
A	VTS	DCI	ª	º	Ê	Ú	ê	ú
B	PLD	CGI	«	»	Ë	Û	ë	û
C	PLU	ST	¼	¼	Ì	Ü	ì	ü
D	RI	DCB	½	½	Í	Ý	í	ý
E	SS2	PM	¾	¾	Î	Þ	î	þ
F	SS3	APC	¿	¿	Ï	ß	ï	ÿ

2 Bytes

Incluindo diversos caracteres anteriormente representados em 1 byte em EBCDIC...

Conceitos

Conversão de Caracteres

Round-Trip (RT) conversions (Ex: ASCII ou EBCDIC -> Unicode)

Integridade do dado mantida no CCSID fonte

- Potencial representação incorreta dos caracteres no CCSID destino
- Representação do caractere recuperada ao converter novamente para CCSID fonte

Pode ser um problema para aplicações que lêem os dados no destino

Enforced Subset (ES) conversions (Ex: Unicode -> ASCII/EBCDIC)

Caracteres existentes no CCSID fonte e destino – ok

Caracteres não existentes no CCSID destino são substituídos por “curinga”

Não recupera representação ao voltar para CCSID fonte



Provido no z/OS pelo suporte a Unicode e serviços de conversão.

OK, Tamber...

mas o que o DB2 tem a ver
com tudo isso?

DB2 e CCSIDs

DB2 e CCSIDs

DB2 armazena dados baseado em CCSIDs

- Para cada diferente Encoding Scheme suportado (EBCDIC, ASCII e Unicode)
- Informados no painel DSNTIPF (DSNTINST), e macro DSNHDECM (DSNHDECP)

```

DSNTIPF          INSTALL DB2 - APPLICATION PROGRAMMING DEFAULTS PANEL 1
====> _

Enter data below:

 1 LANGUAGE DEFAULT      ==> IBMCOB   ASM,C,CPP,IBMCOB,FORTRAN,PLI
 2 DECIMAL POINT IS     ==> .         . or ,
 3 STRING DELIMITER     ==> DEFAULT  DEFAULT, " or ' (COBOL or COB2 only)
 4 SQL STRING DELIMITER ==> DEFAULT  DEFAULT, " or '
 5 DIST SQL STR DELIMTR ==> '         ' or "
 6 MIXED DATA          ==> NO        NO or YES for mixed DBCS data
 7 EBCDIC CCSID         ==>          CCSID of SBCS or mixed data. 1-65533.
 8 ASCII CCSID         ==>          CCSID of SBCS or mixed data. 1-65533.
 9 UNICODE CCSID       ==> 1208      CCSID of UNICODE UTF-8 data.
10 DEF ENCODING SCHEME ==> EBCDIC   EBCDIC, ASCII, or UNICODE
11 APPLICATION ENCODING ==> EBCDIC   EBCDIC, ASCII, UNICODE, ccsid (1-65533)
12 LOCALE LC_CTYPE     ==>
13 DECFLOAT ROUNDING MODE==> ROUND_HALF_EVEN

PRESS:  ENTER to continue  RETURN to exit  HELP for more information

```

```

DSNHDECM          CHARSET=ALPHANUM,
                  ASCCSID=1252,
                  AMCCSID=65534,
                  AGCCSID=65534,
                  SCCSID=37,
                  MCCSID=65534,
                  GCCSID=65534,
                  USCCSID=367,
                  UMCCSID=1208,
                  UGCCSID=1200,
                  ENSCHHEME=EBCDIC,
                  APPENSCH=EBCDIC,
                  DATE=ISO,
                  DATELEN=0,
                  DECARTH=DEC31,
                  DECIMAL=COMMA,
                  DEFLANG=IBMCOB,
                  DELIM=DEFAULT,
                  MIXED=NO,
                  NEWFUN=YES,
                  PADNTSTR=NO,
                  SQLDELI=DEFAULT,
                  DSQLDELI=APOST,
                  SSID=DSN,
                  STDSQL=NO,
                  TIME=ISO,
                  TIMELEN=0,
                  DYNRULS=NO,
                  LC_CTYPE=

```

END

DB2 e CCSIDs

DB2 armazena dados baseado em CCSIDs

- Para cada diferente Encoding Scheme suportado
- Informados no painel DSNTIPF (DSNTINST), e macro DSNHDECM (DSNHDECP)

```

DSNHDECM  CHARSET=ALPHANUM,
          ASCCSID=1252,
          AMCCSID=65534,
          AGCCSID=65534,
          SCCSID=37,
          MCCSID=65534,
          GCCSID=65534,
          USCCSID=367,
          UMCCSID=1208,
          UGCCSID=1200,
          ENSCHEME=EBCDIC,
          APPENSCH=EBCDIC,
          DATE=ISO,
          DATELEN=0,
          DECARTH=DEC31,
          DECIMAL=COMMA,
          DEFLANG=IBMCOB,
          DELIM=DEFAULT,
          MIXED=NO,
          NEWFUN=YES,
          PADNTSTR=NO,
          SQLDELI=DEFAULT,
          DSQLDELI=APOST,
          SSID=DSN,
          STDSQL=NO,
          TIME=ISO,
          TIMELEN=0,
          DYNRULS=NO,
          LC_CTYPE=
END

```

Especifica o CCSID default a ser usado em função do Encoding Scheme escolhido (ENSCHHEME/APPENSCH).

Especifica o Encoding Scheme (e CCSID) default a ser usado em:

- CREATE DATABASE
- CREATE DISTINCT TYPE
- CREATE FUNCTION
- CREATE GLOBAL TEMPORARY TABLE
- DECLARE GLOBAL TEMPORARY TABLE
- CREATE TABLESPACE (in DSNDB04 database)

Especifica o Encoding Scheme (e CCSID) default a ser usado pela aplicação:

- BIND PLAN/PACKAGE
- SET CURRENT PACKAGE SET/PATH

DB2 e CCSIDs

DB2 armazena informação sobre CCSID em diversos locais

CCSIDs are stored in the following places

SYSIBM.SYSDATABASE (V5)

SYSIBM.SYSCOLUMNS

- FOREIGNKEY (V2.3) - subtype information

- CCSID (V8)

SYSIBM.SYSPACKAGE (V7) – Application Encoding

SYSIBM.SYSPARMS (V6)

SYSIBM.SYSPLAN (V7) – Application Encoding

SYSIBM.SYSROUTINES (V8)

SYSIBM.SYSTABLES (V8)

SYSIBM.SYSTABLESPACE (V5)

SYSIBM.SYSVTREE (V5)

Plans and Packages (SCT02 and SPT01) – No external

Directory (DSNDB01) (V5) – No external

DECP (V2.3)

BSDS (V8)

In ENCODING_SCHEME column of - Stored as 'A', 'E', 'U', or blank (default)


SYSIBM.SYSDATATYPES

SYSIBM.SYSDATABASE

SYSIBM.SYSPARMS

SYSIBM.SYSTABLESPACE

SYSIBM.SYSTABLES



Se valor 0
no Catálogo,
assume-se
EBCDIC.

DB2 e CCSIDs

DB2 pode precisar converter caracteres de diferentes CCSIDs...

Mas quando isso ocorre?

Local

Generally, conversion does not occur for local applications

When dealing with ASCII/Unicode tables

When specified by application

- CCSID Override in SQLDA (V2.3)

- Declare Variable (V7)

- Application Encoding Bind Option (V7)

- Current Application Encoding Special Register (V7)

ODBC/JDBC/SQLJ

Remote

Automatically when needed

- DRDA Receiver Makes Right

DB2 e CCSIDs

Como se dá a conversão de caracteres?

Conversão de Caracteres

- Native DB2 - SYSIBM.SYSSTRINGS
- z/OS support for Unicode

Round Trip - vs - Enforced Subset

Round Trip (RT) Conversions

Preserves codepoints that are not representable in both codepages
Work well in a two-tier environment

Enforced Subset (ES) Conversions

Codepoints that are not representable are converted to SUB character

Works well in an heterogeneous environment

DB2 Uses a combination of RT and ES conversions

Trend is toward ES conversions

Continue to use RT conversions in some cases for compatibility reasons

Unicode and RT/ES Conversions

ASCII/EBCDIC -> Unicode conversions are RT

Unicode -> ASCII/EBCDIC conversions are ES

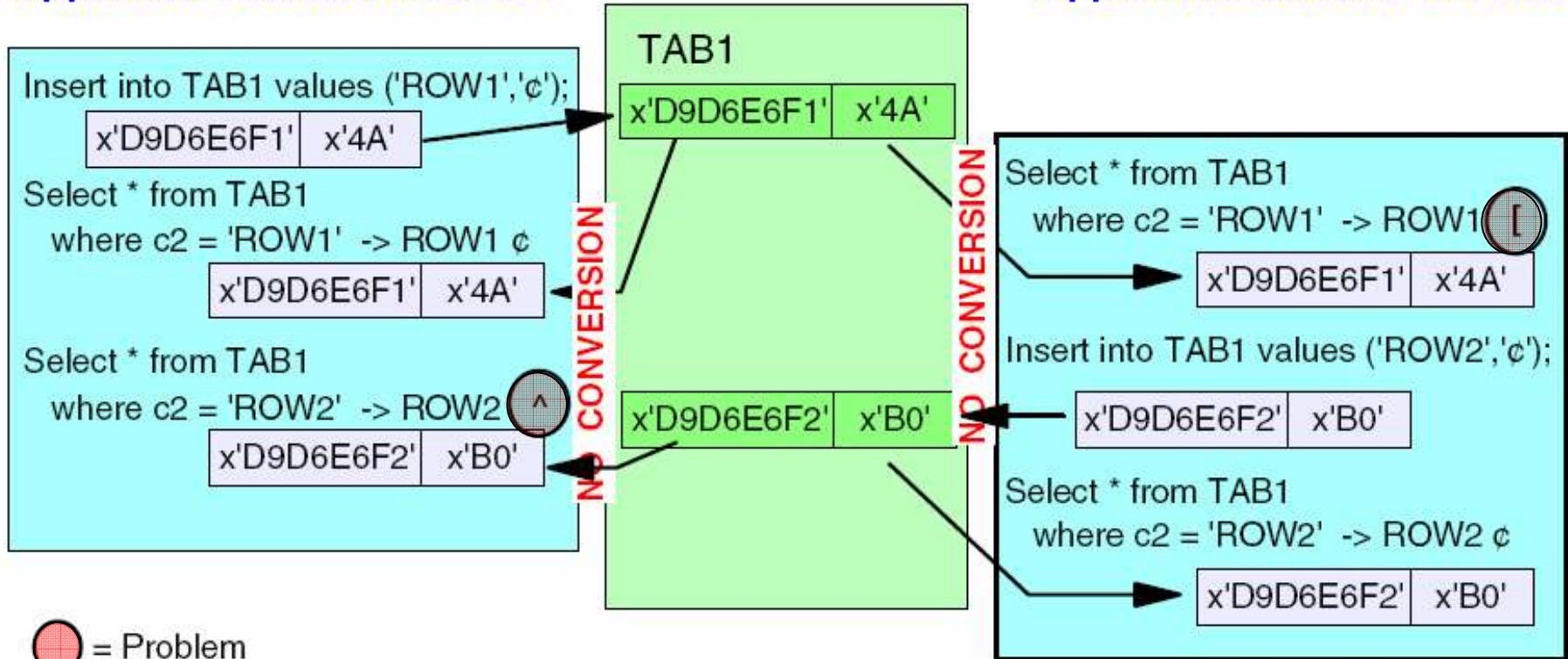
DB2 e CCSIDs

Exemplo: Acesso local EBCDIC – nenhuma conversão

3270 Emulation CCSID 37
 Appl. Enc. Scheme: EBCDIC

DB2 - SCCSID 500

3270 Emulation CCSID 500
 Appl. Enc. Scheme: EBCDIC



Múltiplos CCSIDs em uma única tabela DB2

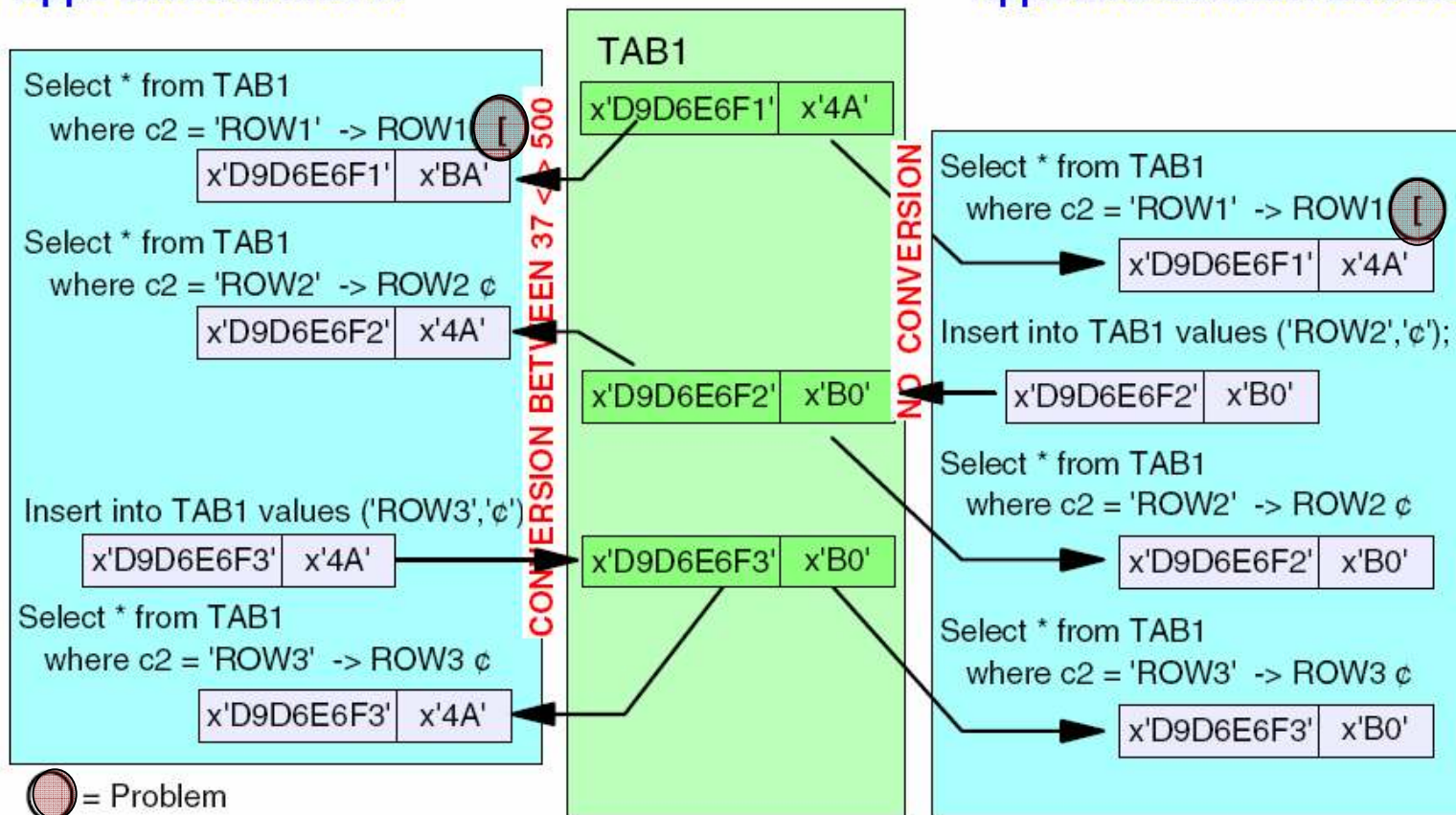
DB2 e CCSIDs

Possível solução: Acesso local EBCDIC – com conversão

3270 Emulation CCSID 37
 Appl. Enc. Scheme: 37

DB2 - SCCSID 500

3270 Emulation CCSID 500
 Appl. Enc. Scheme: EBCDIC

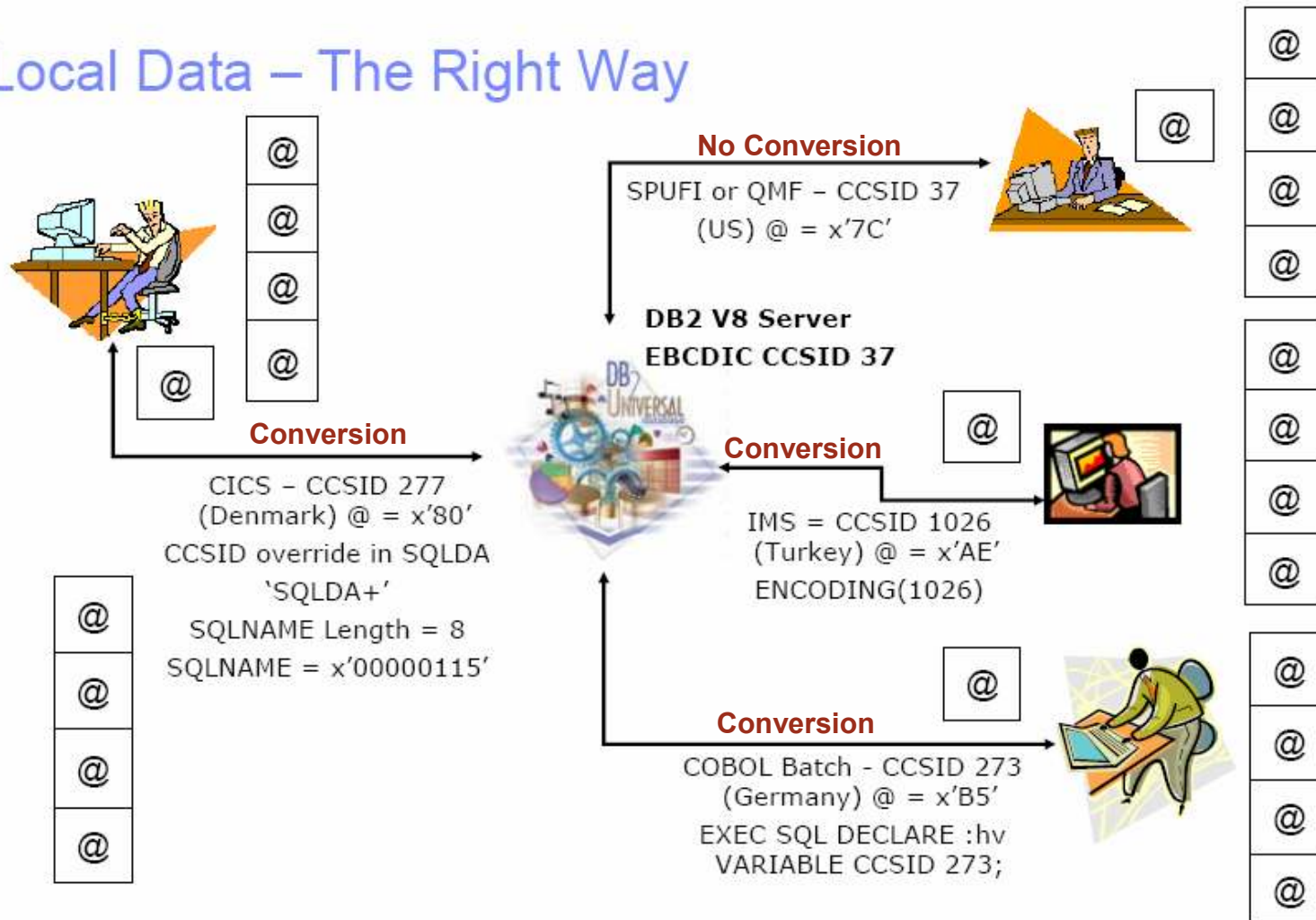


Dado inserido anteriormente deve ser “corrigido”

DB2 e CCSIDs

Outras possibilidades de conversão - Acesso local EBCDIC

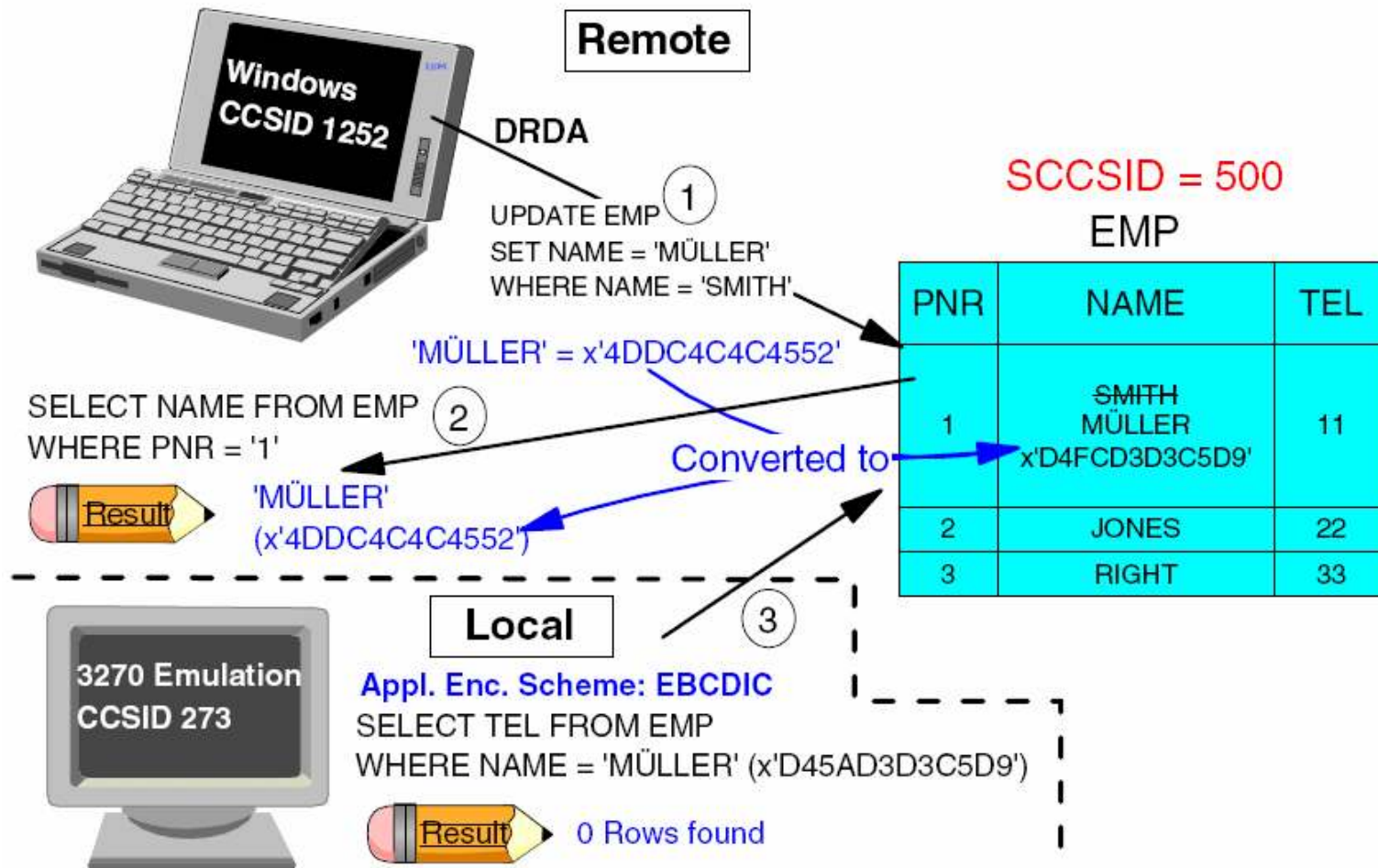
Local Data – The Right Way



DB2 e CCSIDs

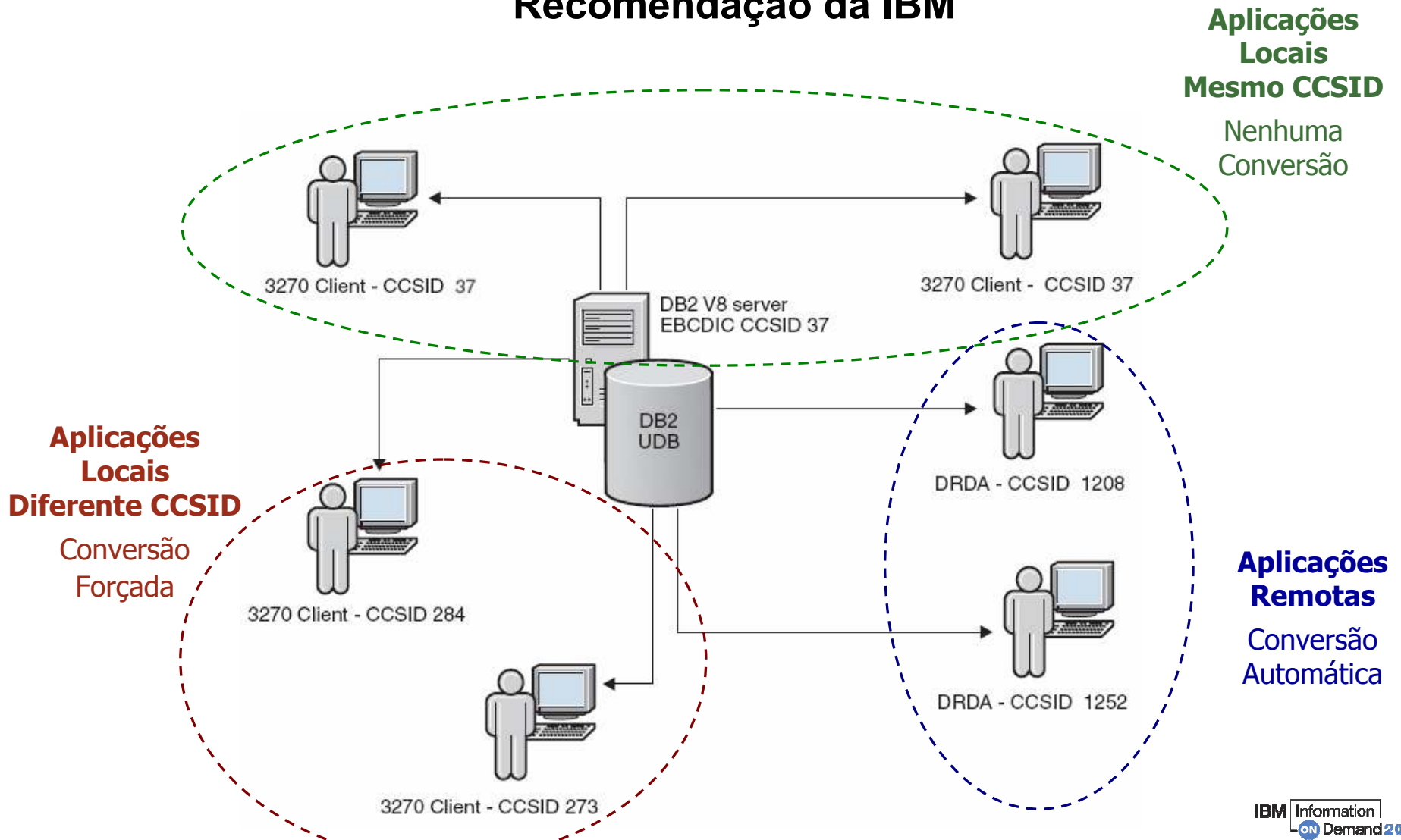
Exemplo: Acesso remoto

Diferentes Encoding Schemes Conversão Automática



DB2 e CCSIDs

Recomendação da IBM



DB2 e CCSID

Dicas para obter DB2 CCSID

Listando BSDS, DSNHDECP...

Query – diferentes CCSIDs no DB2:

```
SELECT SBCS_CCSID, COUNT(*) as "Quantidade"
  FROM SYSIBM.SYSTABLESPACE
 WHERE ENCODING_SCHEME = 'E'
 GROUP BY SBCS_CCSID;
```

Obs: Ignore os "0s"

Sample REXX para obter o terminal CCSID:

```
/* REXX */
ADDRESS ISPEXEC "VGET (ZTERMCID)"
SAY "ZTERMCID=<"ZTERMCID"> (EBCDIC), <"C2X(ZTERMCID)"> (HEXIDECIMAL) "
```

How Can I Tell - GETVARIABLE (V8)

- V8 Function to retrieve SESSION Variables.
 - ▶ SYSIBM Variables contain system information – in particular:
 - SYSIBM.SYSTEM_ASCII_CCSID
 - SYSIBM.SYSTEM_EBCDIC_CCSID
 - SYSIBM.SYSTEM_UNICODE_CCSID
- Example:


```
SET :hv3 = GETVARIABLE('SYSIBM.SYSTEM_EBCDIC_CCSID');
: hv3 = '37,65534,65534'
```

DB2 e CCSID



Descobri que meu DB2 é CCSID=500 e o ambiente é CCSID 37

A IBM recomenda adequar DB2 CCSID, e possui um procedimento especial.

Em geral os dados estão ok, porém é necessário ajustar CCSID em vários lugares no DB2

Visão Geral do processo*

Aplicar usermod para ALTER CCSID
ALTER DATABASES
DROP VIEWS
ALTER TABLESPACES
UPDATE LOBS
REPAIR DBDs (LOBs)
RECREATE VIEWs e AUTHS
UPDATE SYSPARMS
REBIND PACKAGES
REBIND PLANS
Atualizar DSNZPARMs, DSNHDECPS todos membros

Muito Importante: Jamais execute este procedimento sem auxílio da IBM.

Este procedimento é suportado até DB2 v8.

* A intenção aqui é apenas dar uma idéia do esforço requerido no procedimento.

DB2 e CCSID

Descobri que meu DB2 é CCSID=500 e o ambiente é CCSID 37

Procedimento especial – Alteração do CCSID: Chave para o sucesso

Etapa	Tempo	Detalhe
Planejamento	60%	Criação de um plano de alteração em conjunto com a IBM.
Testes	25%	Aplicação do plano em ambiente não produtivo e ajustes no plano original.
Implementação	15%	Alteração em ambientes produtivos. Eventual ajuste fino do plano.



Final feliz! DB2 CCSID coerente com CCSID do ambiente!

DB2 e Unicode

DB2 e Unicode

Unicode CCSIDs within DB2



DSNHDECP - USCCSID/UMCCSID/UGCCSID

CREATE DATABASE/TABLESPACE/TABLE/stored procedure/UDF -
CCSID UNICODE option

CCSIDs used for Unicode encoding in tables depends on the column type and subtype (FOR xxx) option in your DDL:

- CHAR / VARCHAR
 - **FOR SBCS DATA**
 - 7-bit ASCII **CCSID 367**
 - Is a subset of UTF-8
 - **FOR MIXED DATA**
 - UTF-8 **CCSID 1208**
 - This CCSID is the **default for Unicode tables**
 - Allowed, even with MIXED=NO specifies on DSNTIPF
- FOR GRAPHIC / VARGRAPHIC
 - UTF-16 **CCSID 1200**

```

DSNHDECM  CHARSET=ALPHANUM,
           ASCCSID=1252,
           AMCCSID=65534,
           AGCCSID=65534,
           SCCSID=37,
           MCCSID=65534,
           GCCSID=65534,
           USCCSID=367,
           UMCCSID=1208,
           UGCCSID=1200,
           ENSHEME=EBCDIC,
           APPENSCH=EBCDIC,
           DATE=ISO,
           DATELEN=0,
           DECARTH=DEC31,
           DECIMAL=COMMA,
           DEFLANG=IBMCOB,
           DELIM=DEFAULT,
           MIXED=NO,
           NEWFUN=YES,
           PADNTSTR=NO,
           SQLDELI=DEFAULT,
           DSQLELI=APOST,
           SSID=DSN,
           STDSQL=NO,
           TIME=ISO,
           TIMELEN=0,
           DYNRULS=NO,
           LC_CTYPE=
END
  
```


DB2 e Unicode

Conversão de Caracteres

- Native DB2 - SYSIBM.SYSSTRINGS
- z/OS support for Unicode

```

COMMAND INPUT ==> /d uni,all

RESPONSE=SC63
CUN3000I 02.37.19 UNI DISPLAY 752
  ENVIRONMENT:  CREATED      04/17/2004 AT 10.08.19
                  MODIFIED   04/17/2004 AT 10.08.20
                  IMAGE CREATED 12/08/2003 AT 17.02.33
  SERVICE: CHARACTER  CASE           NORMALIZATION
COLLATION
  STORAGE: ACTIVE      424 PAGES
          LIMIT       51200 PAGES
  CASECONV: NORMAL
  NORMALIZE: DISABLED
  COLLATE:  DISABLED
  CONVERSION: 00850-01047-ER           01047-00850-ER
              00037-01200(13488)-ER   01200(13488)-00037-ER
              00037-01208-ER          01208-00037-ER
              00437-01208-ER          01208-00437-ER
              00037-00367-ER          01252-00037-ER
              00037-01252-ER          00367-00037-ER
              00500-01200(13488)-ER   01200(13488)-00500-ER
              01047-01200(13488)-ER   01200(13488)-01047-ER
              01047-01208-ER          01208-01047-ER
              01208-01200-ER          01200-01208-ER
              01383-01200-ER          01200-01383-ER
              00932-01200-ER          01200-00932-ER

```

DB2 v8 z/OS
usa dois
métodos para
conversão de
caracteres

DB2 e Unicode

Collating Sequence



Unicode Table

```

SELECT SUBSTR(NAME,1,8) , HEX(NAME)
FROM SYSIBM.SYSTOGROUP ORDER BY NAME ;
-----+-----+-----+-----+
-----+-----+-----+-----+
!          21
DSN8G810  44534E3847383130
SABI      53414249
SABIGRP1  5341424947525031
SHOW11    53484F573131
SHOW22    53484F573232
SHOW33    53484F573333
SHOW44    53484F573434
SHOW55    53484F573535
SHOWAA    53484F574141
SHOWBB    53484F574242
SHOWCC    53484F574343
SHOWDD    53484F574444
SYSDEFLT  5359534445464C54
baba      62616261
    
```

EBCDIC Table

```

SELECT SUBSTR(NAME,1,8) , HEX(NAME)
FROM EBCDICVOL ORDER BY NAME
-----+-----+-----+-----+
-----+-----+-----+-----+
|          4F
!          5A
baba      82818281
DSN8G810  C4E2D5F8C7F8F1F0
SABI      E2C1C2C9
SABIGRP1  E2C1C2C9C7D9D7F1
SHOWAA    E2C8D6E6C1C1
SHOWBB    E2C8D6E6C2C2
SHOWCC    E2C8D6E6C3C3
SHOWDD    E2C8D6E6C4C4
SHOW11    E2C8D6E6F1F1
SHOW22    E2C8D6E6F2F2
SHOW33    E2C8D6E6F3F3
SHOW44    E2C8D6E6F4F4
SHOW55    E2C8D6E6F5F5
SYSDEFLT  E2E8E2C4C5C6D3E3
    
```

Unicode tem Collating Sequence similar ao ASCII !

DB2 e Unicode

Functions and Routines



Functions

- LENGTH, SUBSTR, POSSTR, LOCATE
 - For SBCS and MIXED (UTF-8) they are byte-oriented
 - For DBCS (UTF-16) they are double-byte-oriented
- New character based functions (and enhanced existing functions)
 - New CHARACTER_LENGTH, SUBSTRING, POSITION
 - New parm indicating how to count (CODEUNITS32, CODEUNITS16, OCTETS)
- Cast functions
 - Unicode generally accepted where CHAR is accepted
 - For CHAR functions, UTF-8 is result data type

Routines

- UDFs, UDTs, and stored procedures allow Unicode parameters
- Parameters converted as necessary between CHAR(UTF-8) and GRAPHIC(UTF-16)
- Date, time, timestamp passed as UTF-8 (ISO format)

DB2 e Unicode

Em Unicode, alguns caracteres são representados em 2+ bytes...

TRANSLATE, SUBSTR, LENGTH etc.

Exemplo:

```

select character_length(name, octets) as bytes,
       substr(name, 1, 6) as substr,
       substring(name, 1, 6, codeunits32) as substring,
       hex(name) as hex,
       translate(name, 'e', 'é') as translate,
       length(translate(name, 'e', 'é')) as translate_bytes
from sysibm.systables
where creator = 'dnet459' and tsname = 'TEST05';

```

BYTES	SUBSTR	SUBSTRING	HEX
8	zéman	zémané	7AC3A96D616EC3A9
		TRANSLATE	
		ze mane	
		TRANSLATE_BYTES	
		8	

DB2 e Unicode

Em Unicode, a representação hexadecimal é diferente de EBCDIC...

Algumas funções podem sofrer mudanças no resultado.

Exemplo – função HEX:

Unicode Table

```
SELECT SUBSTR(NAME,1,8) , HEX(NAME)
FROM SYSIBM.SYSTOGROUP ORDER BY NAME ;
```

NAME	HEX
DSN8G810	44534E3847383130
SABI	53414249
SABIGRP1	5341424947525031

EBCDIC Table

```
SELECT SUBSTR(NAME,1,8) , HEX(NAME)
FROM EBCDICVOL ORDER BY NAME
```

NAME	HEX
!	5A
baba	82818281
DSN8G810	C4E2D5F8C7F8F1F0



DB2 e Unicode

DB2 v8 New Function Mode, Catálogo é convertido para Unicode...

<i>NAME</i>	<i>ENCODING_SCHEME</i>
<i>SYSALTER</i>	<i>U</i>
<i>SYSCOPY</i>	<i>E</i>
<i>SYSDBASE</i>	<i>U</i>
<i>SYSDBAUT</i>	<i>U</i>
<i>SYSDDF</i>	<i>U</i>
<i>SYSEBCDC</i>	<i>E</i>
<i>SYSGPAUT</i>	<i>U</i>
<i>SYSGROUP</i>	<i>U</i>
<i>SYSGRTNS</i>	<i>U</i>
<i>SYSHIST</i>	<i>U</i>
<i>SYSJAUXA</i>	<i>U</i>
<i>SYSJAUXB</i>	<i>U</i>
<i>SYSJAVA</i>	<i>U</i>
<i>SYSOBJ</i>	<i>U</i>
<i>SYSPKAGE</i>	<i>U</i>
<i>SYSPLAN</i>	<i>U</i>
<i>SYSSEQ</i>	<i>U</i>
<i>SYSSEQ2</i>	<i>U</i>
<i>SYSSTATS</i>	<i>U</i>
<i>SYSSTR</i>	<i>U</i>
<i>SYSUSER</i>	<i>U</i>
<i>SYSVIEWS</i>	<i>U</i>

DSNE610I NUMBER OF ROWS DISPLAYED IS 22

Exceções
são
SYSEBCD e
SYSCOPY

DB2 e Unicode

TEXT in SYSIBM.SYSSTMT



Extract of column TEXT

```

.....RELEASE ALL ....
.....RELEASE CURRENT ....
.....SET CONNECTION : H .....Ø.....D.
.....CONNECT TO : H .....Ø.....D.....
.....CONNECT RESET ....
.....CONNECT ....
.....WHENEVER NOT FOUND GO TO EXECWRN
.....WHENEVER SQLWARNING GO TO EXECWRN
.....WHENEVER SQLERROR GO TO EXECERR
..3.º.....1..ØL
.....àää< êá.î&ç|+á.è â<á...< ëè+ (á.î êääç ê..
.....ïääää< êá.îá(&&.è â<á...á (&<|ßää+í(âáê.ääç
.....ääää< êá.èá<á..ääíêê|ê.ã|ê.ëá<ääè...ãê| (.î&
.....v...)ääää< êá.èá<á..ääíêê|ê.ã|ê.ëá<ääè...ãê| (.î&
.....u...!ääää< êá.èá<á..ääíêê|ê.ã|ê.ëá<ääè...ãê| (.î&
    
```

V7 package or
V8 package bound
with NEWFUN=NO
EBCDIC



V8 package
bound with
NEWFUN=YES
Unicode

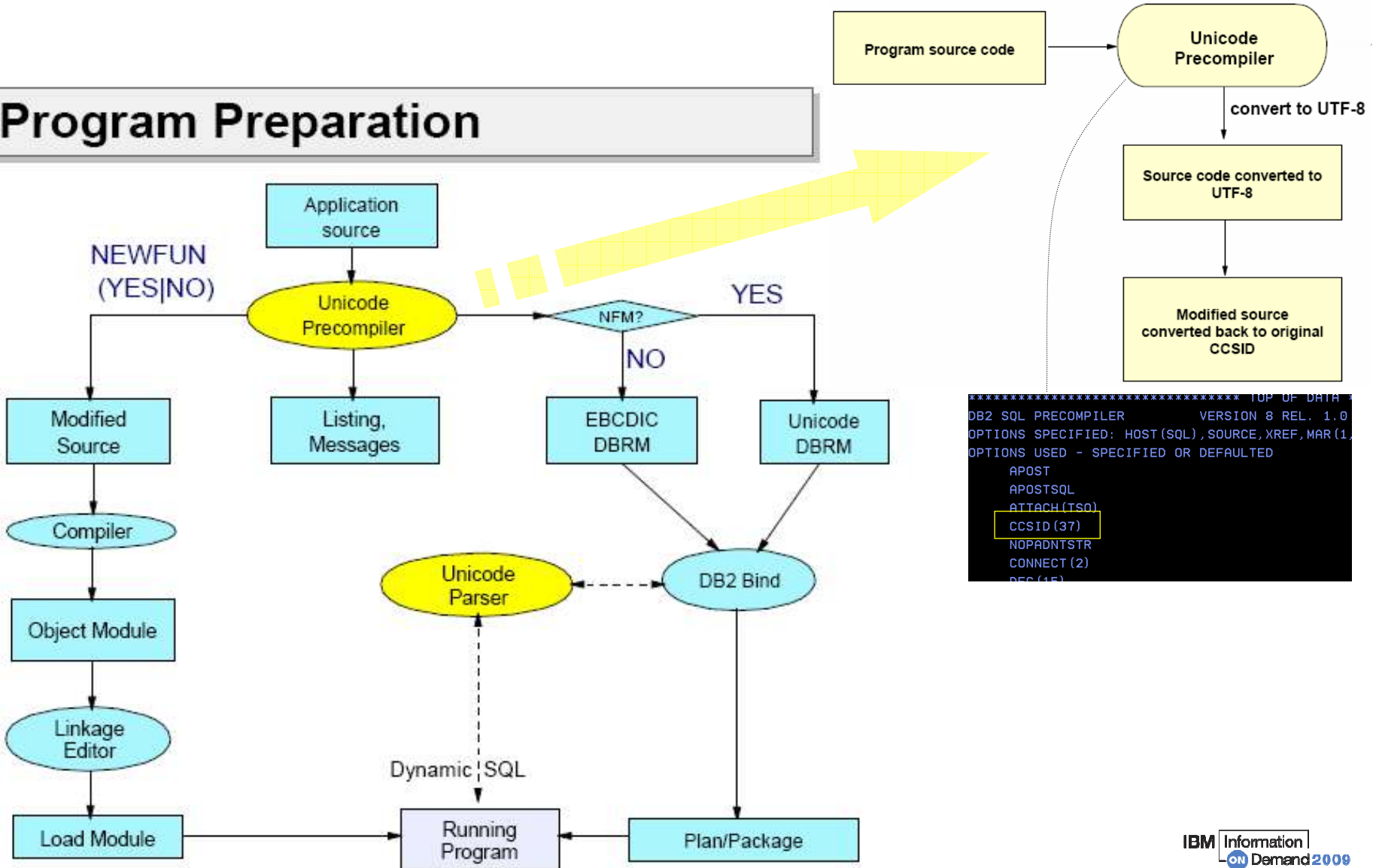


The same applies to STMT column in SYSIBM.SYSPACKSTMT

Essas colunas são "FOR BIT DATA"

DB2 e Unicode

Program Preparation



DB2 e Unicode

DB2 v8 (NEWFUN=YES) gera DBRMs em Unicode

```

BROWSE      DNET459.DBRMLIB.DATA(COBPGM)                Line 00000000 Col 001 080
Command ==>                                         Scroll ==> CSR
***** Top of Data *****
DBRM...µDNET459 COBPGM  .<.=.Ýídx.3.....1..ØLL
..
DBRM.....Ûàää< êá.à+áè.....êääè.-â éá.è â<á...ä|àê ñà|ä.äç ê.....
..+|è.+í<<...ääêíääà|ä.äç ê.....+|è.+í<<...+áíèí <.äç ê.....+|è.+í<<...ä|àà
|ãñàè.äç ê.....+|è.+í<<...àèää+àè&ä.äç ê.....+|è.+í<<...+í(á+àè&ä.àääñ( <.
.....+|è.+í<<...+|(â ñëä&.äç ê.....+|è.+í<<...+|(<|äè&ä.äç ê.....+|è.
+í<<...ëå<í+ã.äç ê.....+|è.+í<<...ä|àää&.äç ê.....+|è.+í<<...ä|àààà.äç ê....
..+|è.+í<<...+í(èá<.äç ê.....+|è.+í<<...à è+ è.à éá.+|è.+í<<...+|(( áè&ä.äç ê
.....+|è.+í<<...+|(&áèè&ä.äç ê.....+|è.+í<<.....
DBRM.....#.....içá+áiáé.ëé<i ê+ñ+â.ä|èñ+íá.
DBRM.....@.....içá+áiáé.ëé<áèè|ê.á|è|.àâ.áèè|ê.
DBRM...È.....p.....>éá<áèè..+áíèí <...ä|àà|ãñàè...+|(â ñëä&.ñ+è|...ç.....ç.
...ç.ãè|(.à+áè.....êääè.-â éá.içáéá.ä|àê ñà|ä.....ç.....D.....
ää<éääè...â éá.ä à èèè <.+ää.+áíèí <.....D.....ää<éääè...â éá.ä à èèè <
.+ää.ä|àà|ãñàè.....D.....ää<éääè...â éá.ä à èèè <.+ää.+|(â ñëä&.....,Ø
.....D.....ää<éääè...â éá.ä à èèè <.+ää.ä|àê ñà|ä
DBRM.....x.....ä| (ñè.....
***** Bottom of Data *****
    
```

DB2 e Unicode DB2 v8 (NEWFUN=YES) gera DBRMs em Unicode

```

BROWSE      DNET459.DBRMLIB.DATA(COBPGM)                Converted data shown
Command ==> display utf8                               Scroll ==> CSR
***** Top of Data *****
.....Ö..@@.L8~..V.....%.....
..@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
.....DECLARE DNET459 . RCT01_BASE TABLE ( CODRAIDOC CHAR ( 9
) NOT NULL , DGTVFCDOC CHAR ( 2 ) NOT NULL , NEXTVAL CHAR ( 10 ) NOT NULL , CODD
OCIDT CHAR ( 20 ) NOT NULL , DSCENDSPC CHAR ( 50 ) NOT NULL , NUMENDSPC DECIMAL
( 5 , 0 ) NOT NULL , NOMBAISCP CHAR ( 25 ) NOT NULL , NOMLOCSPC CHAR ( 30 ) NOT
NULL , SGLUNF CHAR ( 2 ) NOT NULL , CODCEP CHAR ( 8 ) NOT NULL , CODDDD CHAR ( 2
) NOT NULL , NUMTEL CHAR ( 8 ) NOT NULL , DATNAS DATE NOT NULL , NOMMAESPC CHAR
( 45 ) NOT NULL , NOMPESPC CHAR ( 45 ) NOT NULL ) .....@@@@@@@@@@@@@@@@@@@@
.....5.....{.....WHENEVER SQLWARNING CONTINUE @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
.....8.....|.....WHENEVER SQLERROR GOTO DB2ERROR @@@@@@@@@@@@@@@@@@@@@@@@@@
.....t.....nSELECT "NEXTVAL" , CODDOCIDT , NOMBAISCP INTO : H , : H
, : H FROM DNET459 . RCT01_BASE WHERE CODRAIDOC = : H .....9.../@.....#
DCLRCT01-BASE-CADAstral-NBC.NEXTVAL.;...5@.....%DCLRCT01-BASE-CADAstral
-NBC.CODDOCIDT.;...;@.....%DCLRCT01-BASE-CADAstral-NBC.NOMBAISCP.;...k.
.....%DCLRCT01-BASE-CADAstral-NBC.CODRAIDOC@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
.....#.....COMMIT .....@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
***** Bottom of Data *****

-----
| Data specified in a DISPLAY command has been converted from CCSID '1208' |
-----
    
```

Apar OA07685 (ISPF) provê instrumentação para converter o DISPLAY: DISPLAY UTF8 | EBCDIC

DB2 e Unicode

SQL pode ser afetado quando mais de um Encoding Scheme envolvido

Exemplo:

```
SELECT a.name, a.creator, b.charcol, 'ABC',  
       :hvchar, X'C1C2C3'  
FROM sysibm.systables a,  
     ebcdictable b  
WHERE a.name = b.name AND  
      b.name > 'B' AND  
      a.creator = 'SYSADM'  
ORDER BY b.name;
```

Result or Evaluated:

EBCDIC

Unicode

Application Encoding Scheme

Assumindo catálogo DB2 em Unicode

DB2 e Unicode

Comparison and resulting data types for **multiple CCSID** sets...

If an expression or comparison involves two strings which contain columns with different CCSID sets,

Drive to Unicode if necessary

WHERE T1.C1 = T2.C1

If an expression or comparison involves two strings with different CCSID sets where only one of them contains a column,

Drive to the column's CCSID set

WHERE T1.C1 = X'C1C2'

If an expression or comparison involves two strings with different CCSID sets and neither contains a column,

Drive to Unicode

WHERE GX'42C142C2' = 'ABC' -- GX literal and 'ABC' are different CCSIDs

String constants and special registers in a context by themselves use the application encoding scheme

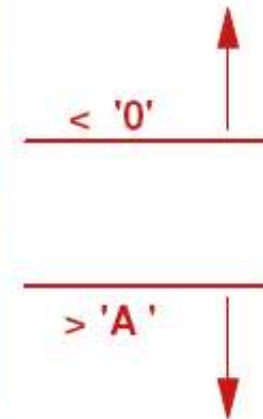
SELECT 'ABC' FROM T1 . . .

DB2 e Unicode

Range predicates

SELECT NAME FROM SYSIBM.SYSTABLES WHERE NAME > 'A' AND NAME < '0';

EBCDIC	hex value	Unicode (ASCII)	hex value
space	'40'x	space	'20'x
\$	'5B'x	#	'23'x
_ (underscore)	'6D'x	\$	'24'x
#	'7B'x	numerals	'30-39'x
@	'7C'x note 1	@	'40'x
lower case	'81-89'x '91-99'x 'A1-A9'x	upper case	'41-4F'x '50-5A'x
upper case	'C1-C9'x 'D1-D9'x 'E1-E9'x	_ (underscore)	'5F'x
numerals	'F0-F9'x	lower case	'61-6F'x '70-7A'x



note 1: '7D'x in CCSID 37; hexadecimal value varies by CCSID

DB2 e Unicode

```
SELECT * FROM SYSIBM.SYSTABLES
WHERE NAME BETWEEN 'TEST2' AND 'TESTB'
ORDER BY NAME;
```

EBCDIC

```
NAME
-----
TESTC
TEST1
```

Unicode

```
NAME
-----
TEST3
TESTA
```

SYSIBM.SYSTABLES

NAME (EBCDIC)	NAME (Unicode)
TESTA	TEST1
TESTB	TEST2
TESTC	TEST3
TEST1	TESTA
TEST2	TESTB
TEST3	TESTC

Tamber,
onde posso obter mais
informações?

URLs

Code Pages

<http://www.ibm.com/servers/eserver/series/software/globalization/codepages.html>

<http://www-01.ibm.com/support/docview.wss?uid=swg21197185>

Unicode

<http://www.unicode.org>

<http://www.unicode.org/unibook/>

Artigos

<http://www.ibm.com/developerworks/data/library/techarticle/dm-0506chong/>

<http://www.ibm.com/support/docview.wss?uid=swg21203359>

<http://www-01.ibm.com/support/docview.wss?rs=0&uid=swg21203360>

Downloads

<http://www.redbooks.ibm.com>

<http://www-01.ibm.com/support/docview.wss?rs=64&uid=swg27011656>

Livros

Redbooks

SG24-6079 DB2 UDB for z/OS Version 8: Everything You Ever Wanted to Know (Capítulo 6)

SG24-6465 DB2 UDB for z/OS Version 8 Performance Topics

SG24-6851 e-business Globalization Solution Design Guide Getting Started

Manuais

SC19-1161 DB2 Version 9 – Internationalization Guide

SA22-7649 z/OS Support for Unicode: Using Unicode Services



The Future Runs on System z