

TS-691: SAS[®] Encoding Values, IANA Preferred MIME Charset, Java[™] and Oracle[®] Encoding Names

The SAS[®] system uses the ENCODING= system option to specify the SAS session encoding, which establishes the environment to process SAS syntax and SAS data sets, and to read and write external files.

The following tables map names used in the ENCODING= system option to the names used in XML, HTML, MIME and other applications that mark encodings with unique names. This information is a sub-set of the character set names registered by the Internet Assigned Numbers Authority (IANA). Currently, the IANA list is at IANA CHARACTER SETS (<http://www.iana.org/assignments/character-sets>).

Whenever possible, the items listed in the IANA column of each table are the preferred MIME names. If a preferred MIME name is not available, these items represent the primary registered names.

Please note that encoding names appear in the exact case specified in the respective documentation since some uses of these encoding names may be case-sensitive. SAS encoding names are not case-sensitive.

Mainframe Single-Byte Encodings

ENCODING= VALUE	IANA PREFERRED MIME CHARSET	JAVA ENCODING[1]	ORACLE CHARSET[2]	DESCRIPTION
ebcdic037	IBM037	Cp037	WE8EBCDIC37	North America EBCDIC
ebcdic275	IBM275	n/a	n/a	Brazil EBCDIC
ebcdic424	IBM424	Cp424	IW8EBCDIC424	Hebrew EBCDIC
ebcdic425	n/a	n/a	n/a	Arabic EBCDIC
ebcdic500	IBM500	Cp500	WE8EBCDIC500	International EBCDIC
ebcdic838	IBM-Thai	Cp838	TH8TISEBCDIC	Thai EBCDIC
ebcdic870	IBM870	Cp870	EE8EBCDIC870	Central Europe EBCDIC
ebcdic875	n/a	Cp875	EL8EBCDIC875	Greek EBCDIC
ebcdic924	IBM00924	n/a	WE8EBCDIC924	European EBCDIC
ebcdic1025	n/a	Cp1025	CL8EBCDIC1025	Cyrillic EBCDIC
ebcdic1026	IBM1026	Cp1026	TR8EBCDIC1026	Turkish EBCDIC
ebcdic1047	IBM1047	n/a	WE8EBCDIC1047	Western EBCDIC
ebcdic1112	IBM1112	Cp1112	BLT8EBCDIC1112	Baltic EBCDIC
ebcdic1130	n/a	n/a	n/a	Vietnamese EBCDIC
ebcdic1140	IBM01140	Cp1140	WE8EBCDIC1140	North America EBCDIC
ebcdic1141	IBM01141	Cp1141	D8EBCDIC1141	Austria/Germany EBCDIC

ENCODING=VALUE	IANA PREFERRED MIME CHARSET	JAVA ENCODING[1]	ORACLE CHARSET[2]	DESCRIPTION
ebcdic1142	IBM01142	Cp1142	DK8EBCDIC1142	Denmark/Norway EBCDIC
ebcdic1143	IBM01143	Cp1143	S8EBCDIC1143	Finland/Sweden EBCDIC
ebcdic1144	IBM01144	Cp1144	I8EBCDIC1144	Italy EBCDIC
ebcdic1145	IBM01145	Cp1145	WE8EBCDIC1145	Spain EBCDIC
ebcdic1146	IBM01146	Cp1146	WE8EBCDIC1146	United Kingdom EBCDIC
ebcdic1147	IBM01147	Cp1147	F8EBCDIC1147	France EBCDIC
ebcdic1148	IBM01148	Cp1148	WE8EBCDIC1148	International EBCDIC

Mainframe Double-Byte Encodings

PLATFORM	LANGUAGE	ENCODING=VALUE	IANA PREFERRED MIME CHARSET	JAVA ENCODING[1]	ORACLE CHARSET[2]
IBM	Japanese	ibm-939	n/a	Cp939	JA16DBCS
IBM	Korean	ibm-933	n/a	Cp933	KO16DBCS
IBM	Simplified Chinese	ibm-935	n/a	Cp935	ZHS16DBCS
IBM	Traditional Chinese	ibm-937	n/a	Cp937	ZHT16DBCS

UNIX Single-Byte Encodings

ENCODING=VALUE	IANA PREFERRED MIME CHARSET	JAVA ENCODING[1]	ORACLE CHARSET[2]	DESCRIPTION
arabic	ISO-8859-6	ISO8859_6	AR8ISO8859P6	Arabic (ISO)
cyrillic	ISO-8859-5	ISO8859_5	CL8ISO8859P5	Cyrillic (ISO)
greek	ISO-8859-7	ISO8859_7	EL8ISO8859P7	Greek (ISO)
hebrew	ISO-8859-8	ISO8859_8	IW8ISO8859P8	Hebrew (ISO)
latin1	ISO-8859-1	ISO8859_1	WE8ISO8859P1	Western (ISO)
latin2	ISO-8859-2	ISO8859_2	EE8ISO8859P2	Central Europe (ISO)
latin5	ISO-8859-9	ISO8859_9	WE8ISO8859P9	Turkish (ISO)
latin6	ISO-8859-10	n/a	NE8ISO8859P10	Baltic (ISO)

ENCODING=VALUE	IANA PREFERRED MIME CHARSET	JAVA ENCODING[1]	ORACLE CHARSET[2]	DESCRIPTION
latin9	ISO-8859-15	ISO8859_15	WE8ISO8859P15	European (ISO)
thai	TIS-620 or ISO-8859-11[3]	TIS620	TH8TISASCII	Thai (ISO)

UNIX Double-Byte Encodings

PLATFORM	LANGUAGE	ENCODING=VALUE	IANA PREFERRED MIME CHARSET	JAVA ENCODING[1]	ORACLE CHARSET[2]
AIX	Japanese	euc-jp	EUC-JP	EUC_JP	JA16EUC
AIX	Japanese	shift-jis	Shift_JIS	SJIS	JA16SJIS
AIX	Korean	euc-kr	EUC-KR	EUC_KR	KO16KSC5601
AIX	Simplified Chinese	euc-cn	GBK	GBK	ZHS16GBK
AIX	Traditional Chinese	big5 or ibm-950	Big5	Big5	ZHT16BIG5
HP-UX	Japanese	euc-jp	EUC-JP	EUC_JP	JA16EUC
HP-UX	Japanese	shift-jis	Shift_JIS	SJIS	JA16SJIS
HP-UX	Korean	euc-kr	EUC-KR	EUC_KR	KO16KSC5601
HP-UX	Simplified Chinese	euc-cn	GBK	GBK	ZHS16GBK
HP-UX	Traditional Chinese	hp15-tw	n/a	n/a	ZHT16CCDC
HP-UX	Traditional Chinese	euc-tw	ISO-2022-CN or EUC-TW	EUC_TW	ZHT32EUC
Linux	Japanese	euc-jp	EUC-JP	EUC_JP	JA16EUC
Linux	Japanese	shift-jis	Shift_JIS	SJIS	JA16SJIS
Linux	Korean	euc-kr	EUC-KR	EUC_KR	KO16KSC5601
Linux	Simplified Chinese	euc-cn	GBK	GBK	ZHS16GBK
Linux	Traditional Chinese	big5 or ms-950	Big5	Big5	ZHT16BIG5
Solaris	Japanese	euc-jp	EUC-JP	EUC_JP	JA16EUC
Solaris	Japanese	shift-jis	Shift_JIS	SJIS	JA16SJIS
Solaris	Korean	euc-kr	EUC-KR	EUC_KR	KO16KSC5601

PLATFORM	LANGUAGE	ENCODING=VALUE	IANA PREFERRED MIME CHARSET	JAVA ENCODING[1]	ORACLE CHARSET[2]
Solaris	Simplified Chinese	euc-cn	GBK	GBK	ZHS16GBK
Solaris	Traditional Chinese	big5 or ms-950	Big5	Big5	ZHT16BIG5
Solaris	Traditional Chinese	euc-tw	ISO-2022-CN or EUC-TW	EUC_TW	ZHT32EUC
True64	Japanese	euc-jp	EUC-JP	EUC_JP	JA16EUC
True64	Japanese	shift-jis	Shift_JIS	SJIS	JA16SJIS
True64	Korean	euc-kr	EUC-KR	EUC_KR	KO16KSC5601
True64	Simplified Chinese	euc-cn	GBK	GBK	ZHS16GBK
True64	Traditional Chinese	big5, ms-950 or euc-tw	Big5	Big5	ZHT16BIG5

VMS Single-Byte Encodings

See UNIX Single-Byte Encodings

Windows Single-Byte Encodings

ENCODING=VALUE	IANA PREFERRED MIME CHARSET	JAVA ENCODING[1]	ORACLE CHARSET[2]	DESCRIPTION
msdos720	n/a	n/a	AR8ADOS720	Arabic MS-DOS
msdos737	n/a	Cp737	EL8PC737	Greek MS-DOS
msdos775	IBM775	Cp775	BLT8PC775	Baltic MS-DOS
pcoem437	IBM437	Cp437	US8PC437	USA IBM-PC
pcoem850	IBM850	Cp850	WE8PC850	Western IBM-PC
pcoem852	IBM852	Cp852	EE8PC852	Central Europe IBM-PC
pcoem857	IBM857	Cp857	TR8PC857	Turkish IBM-PC
pcoem858	IBM00858	Cp858	WE8PC858	European IBM-PC
pcoem860	IBM860	Cp860	WE8PC860	Portuguese IBM-PC
pcoem862	IBM862	Cp862	IW8PC1507	Hebrew IBM-PC
pcoem863	IBM863	Cp863	CDN8PC863	French Canadian IBM-PC
pcoem864	IBM864	Cp864	n/a	Arabic IBM-PC

ENCODING=VALUE	IANA PREFERRED MIME CHARSET	JAVA ENCODING[1]	ORACLE CHARSET[2]	DESCRIPTION
pcoem865	IBM865	Cp865	N8PC865	Nordic IBM-PC
pcoem866	IBM866	Cp866	RU8PC866	Cyrillic IBM-PC
pcoem869	IBM869	Cp869	EL8PC869	Greek IBM-PC
pcoem874	IBM874	Cp874	TH8TISASCII	Thai IBM-PC
pcoem921	n/a	Cp921	LT8MSWIN921	Baltic IBM-PC
pcoem922	n/a	Cp922	n/a	Estonia IBM-PC
pcoem1129	n/a	n/a	n/a	Vietnamese IBM-PC
warabic	windows-1256	Cp1256	AR8MSWIN1256	Arabic Windows
wbaltic	windows-1257	Cp1257	BLT8MSWIN1257	Baltic Windows
wcyrillic	windows-1251	Cp1251	CL8MSWIN1251	Cyrillic Windows
wgreek	windows-1253	Cp1253	EL8MSWIN1253	Greek Windows
whebrew	windows-1255	Cp1255	IW8MSWIN1255	Hebrew Windows
wlatin1	windows-1252	Cp1252	WE8MSWIN1252	Western Windows
wlatin2	windows-1250	Cp1250	EE8MSWIN1250	Central Europe Windows
wturkish	windows-1254	Cp1254	TR8MSWIN1254	Turkish Windows
wvietnamese	windows-1258	Cp1258	VN8MSWIN1258	Vietnamese Windows

Windows Double-Byte Encodings

PLATFORM	LANGUAGE	ENCODING=VALUE	IANA PREFERRED MIME CHARSET	JAVA ENCODING[1]	ORACLE CHARSET[2]
Windows	Japanese	shift-jis or ms-932	Shift_JIS	SJIS	JA16SJIS
Windows	Korean	euc-kr or ms-949	EUC-KR	MS949	KO16MSWIN949
Windows	Simplified Chinese	euc-cn or ms-936	GBK	GBK	ZHS16GBK
Windows	Traditional Chinese	big5 or ms-950	Big5 or Big5-HKSCS	Big5	ZHT16BIG5

Universal Encodings

ENCODING=VALUE	IANA PREFERRED MIME CHARSET	JAVA ENCODING[1]	ORACLE CHARSET[2]	DESCRIPTION
utf-8	UTF-8	UTF8	UTF8	UTF-8 Universal

[1] Source: <http://java.sun.com/j2se/1.4.2/docs/guide/intl/encoding.doc.html>

[2] Source: http://download.oracle.com/docs/cd/B10501_01/server.920/a96529/appa.htm#956722

[3] See: <http://www.nectec.or.th/it-standards/mlit99/mlit99-country.html>

SAS[®] and all other SAS Institute product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.

Oracle and all other Oracle Corporation product or service names are registered trademarks or trademarks of Oracle Corporation in the USA and other countries. Other brand and product names are registered trademarks or trademarks of their respective companies.

[®] indicates USA registration.