

PRESENTATION MANAGING HYBRID DATABASES

Uploading a CSV Bank Statement onto IBM DB2.

IBM Db2 on Cloud

Storage: 5%

LOAD

Source

Target

Define

You are loading the file **export.csv** into **NRM17357.EXPORT**

Code page (character encoding): 1208 (UTF-8) Separator: , Header in first row: ☒ Time & date format: ☐

Datum	Transaktion	Kategori	Belopp	Saldo
DATE	VARCHAR	VARCHAR	DECIMAL	DECIMAL
1 2019-05-06	Reservation Kortköp HEMKÖP STOCKHOLM G		-330,23	
2 2019-05-06	MARTIN.DOM		4.000,00	3.971,86
3 2019-05-01	Vardagspaket Tillval Betalningar		-12,00	-28,14
4 2019-05-01	Vardagspaket Månadspris kort		-24,00	-16,14
5 2019-04-23	Kortköp 190417 104 MEDBORGARPLATS		-128,00	7,86
6 2019-04-23	Kortköp 190417 HEMKÖP STOCKHOLM GUL		-185,65	135,86
7 2019-04-23	Kortköp 190418 HEMKÖP STOCKHOLM GUL		-99,85	321,51
8 2019-04-23	Kortköp 190418 HEMKÖP STOCKHOLM GUL		-116,94	421,36
9 2019-04-23	Kortköp 190417 REST KVARNEN		-136,00	538,30
10 2019-04-23	Kortköp 190417 REST KVARNEN		-78,00	674,30

Load details

Warnings 165 warnings

My computer export.csv Target NRM17357.EXPORT

View Table Load More Data

Status Settings

The data load job succeeded.
You can now work with your data.

84 84 0

Errors 0 Warnings 165

SQL3123W: The field value in row "F0-1" and column "4" cannot be converted to a PACKED DECIMAL value. A null was loaded. (More info)
Number of occurrences: 165

SQL3123W: The field value in row "F0-2" and column "4" cannot be converted to a PACKED DECIMAL value. A null was loaded. (More info)
Number of occurrences: 165

SQL3123W: The field value in row "F0-2" and column "5" cannot be converted to a PACKED DECIMAL value. A null was loaded. (More info)
Number of occurrences: 165

SQL3123W: The field value in row "F0-3" and column "4" cannot be converted to a PACKED DECIMAL value. A null was loaded. (More info)
Number of occurrences: 165

SQL3123W: The field value in row "F0-3" and column "5" cannot be converted to a PACKED DECIMAL value. A null was loaded. (More info)
Number of occurrences: 165

ENCOUNTERED PROBLEM: all negative values were loaded and returned as NULL.

NRM17357.EXPORT

Delete Table

Export to CSV

	DATUM	TRANSAKTION	KATEGORI	BELOPP	SALDO
	DATE	VARCHAR(39)	VARCHAR(5)	DECIMAL(7,2)	DECIMAL(8,2)
1	2019-05-06	Reservation Kortköp HEMKÖP STOCKHOLM G			
2	2019-05-06	MARTIN.DOM			
3	2019-05-01	Vardagspaket Tillval Betalningar			
4	2019-05-01	Vardagspaket Månadspris kort			
5	2019-04-23	Kortköp 190417 104 MEDBORGARPLATS			

PRESENTATION MANAGING HYBRID DATABASES

SOLVING THE ISSUE:

First step: In the original Bank Statement .xls file creating a new column which is a copy of transactions with absolute values.

F2		fx		=ABS(D2:D84)			
	A	B	C	D	E	F	G
1	Datum	Transaktion	Kategori	Belopp	Saldo	Belopp_pos	
2	2019-05-06	Reservation Kortköp HEMKOP STOCKHOLM G		-330,23		330,23	
3	2019-05-06	MARTIN.DOM		4.000,00	3.971,86	4000	
4	2019-05-01	Vardagspaket Tillval Betalningar		-12,00	-28,14	12	
5	2019-05-01	Vardagspaket Månadspris kort		-24,00	-16,14	24	
6	2019-04-23	Kortköp 190417 104 MEDBORGARPLATS		-128,00	7,86	128	
7	2019-04-23	Kortköp 190417 HEMKOP STOCKHOLM GUL		-185,65	135,86	185,65	
8	2019-04-23	Kortköp 190418 HEMKOP STOCKHOLM GUL		-99,85	321,51	99,85	
9	2019-04-23	Kortköp 190418 HEMKOP STOCKHOLM GUL		-116,94	421,36	116,94	
10	2019-04-23	Kortköp 190417 REST KVARNEN		-136,00	538,30	136	
11	2019-04-23	Kortköp 190419 BEERPOT		-78,00	674,30	78	
12	2019-04-23	Kortköp 190419 PIERIAS LIVS		-49,00	752,30	49	
13	2019-04-23	Kortköp 190420 HEMKOP STOCKHOLM GUL		-49,95	801,30	49,95	
14	2019-04-18	Kortköp 190416 HEMKOP STOCKHOLM GUL		-149,09	851,25	149,09	
15	2019-04-16	MARTINA.DOM		1.000,00	1.000,34	1000	
16	2019-04-16	Kortköp 190414 IZ *GOTT SMAPLOC		-40,00	0,34	40	
17	2019-04-15	Kortköp 190413 7 ELEVEN 4119113		-66,00	40,34	66	
18	2019-04-15	Kortköp 190412 7 ELEVEN 4119113		-58,00	106,34	58	
19	2019-04-15	Kortköp 190412 HEMKOP STOCKHOLM GUL		-19,95	164,34	19,95	

Step two: using an Open Source Tool (Jupyter notebook) to read the updated .xls and convert to a .csv file with Python commandline

```
df = pd.read_excel('bank.xls')  
df.head()
```

	Datum	Transaktion	Kategori	Belopp	Saldo	Belopp_pos
0	2019-05-06	Reservation Kortköp HEMKOP STOCKHOLM G	NaN	-330,23	NaN	330.23
1	2019-05-06	MARTIN.DOM	NaN	4.000,00	3.971,86	4000.00
2	2019-05-01	Vardagspaket Tillval Betalningar	NaN	-12,00	-28,14	12.00
3	2019-05-01	Vardagspaket Månadspris kort	NaN	-24,00	-16,14	24.00
4	2019-04-23	Kortköp 190417 104 MEDBORGARPLATS	NaN	-128,00	7,86	128.00

```
df.to_csv('bank.csv', encoding='utf-8')
```

PRESENTATION

MANAGING HYBRID DATABASES

Step Three: Uploading the updated .csv on DB2 now shows new columns with positive values.

	Column_0 SMALLINT	Datum DATE	Transaktion VARCHAR	Kategori VARCHAR	Belopp DECIMAL	Saldo DECIMAL	Belopp_pos DECIMAL
1	0	2019-05-06	Reservation Kortköp HEMKÖP STOCKHOLM G		-330,23		330.23
2	1	2019-05-06	MARTIN.DOM		4.000,00	3.971,86	4000.0
3	2	2019-05-01	Vardagspaket Tillval Betalningar		-12,00	-28,14	12.0
4	3	2019-05-01	Vardagspaket Månadspris kort		-24,00	-16,14	24.0

NRM17357.BANK

Delete Table Export to CSV

	COLUMN_0 SMALLINT	DATUM DATE	TRANSAKTION VARCHAR(39)	KATEGORI VARCHAR(6)	BELOPP DECIMAL(11, 4)	SALDO DECIMAL(11, 4)	BELOPP_POS DECIMAL(7, 2)
1		0 2019-05-06	Reservation Kortköp HEMKÖP STOCKHOLM G				330.23
2		1 2019-05-06	MARTIN.DOM				4000.00
3		2 2019-05-01	Vardagspaket Tillval Betalningar				12.00
4		3 2019-05-01	Vardagspaket Månadspris kort				24.00
5		4 2019-04-23	Kortköp 190417 104 MEDBORGARPLAT				128.00
6		5 2019-04-23	Kortköp 190417 HEMKÖP STOCKHOLM				185.65
7		6 2019-04-23	Kortköp 190418 HEMKÖP STOCKHOLM				99.85
8		7 2019-04-23	Kortköp 190418 HEMKÖP STOCKHOLM				116.94
9		8 2019-04-23	Kortköp 190417 REST KVARNEN				136.00
10		9 2019-04-23	Kortköp 190419 BEERPOT				78.00
11		10 2019-04-23	Kortköp 190419 PIERIAS LIVS				49.00
12		11 2019-04-23	Kortköp 190420 HEMKÖP STOCKHOLM				49.95

Step Four: Accessing the updated DataBase on Cloud and converting the searchquery into a DataFrame.

```
%load_ext sql

%sql ibm_db_sa://nrm17357:5gzf%2B8btmhwg1nv7@dashdb-txn-sbox-yp-lon02-01.services.eu-gb.ibmcloud.net:50000/BLUDB

'Connected: nrm17357@BLUDB'

T = %sql SELECT * FROM BANK

* ibm_db_sa://nrm17357:***@dashdb-txn-sbox-yp-lon02-01.services.eu-gb.ibmcloud.net:50000/BLUDB
Done.

dataframe = T.DataFrame()

dataframe
```

	Column_0	Datum	Transaktion	Kategori	Belopp	Saldo	Belopp_pos
0	0	2019-05-06	Reservation Kortköp HEMKÖP STOCKHOLM G	None	None	None	330.23
1	1	2019-05-06	MARTIN.DOM	None	None	None	4000.00
2	2	2019-05-01	Vardagspaket Tillval Betalningar	None	None	None	12.00
3	3	2019-05-01	Vardagspaket Månadspris kort	None	None	None	24.00
4	4	2019-04-23	Kortköp 190417 104 MEDBORGARPLATS	None	None	None	128.00
5	5	2019-04-23	Kortköp 190417 HEMKÖP STOCKHOLM GUL	None	None	None	185.65
6	6	2019-04-23	Kortköp 190418 HEMKÖP STOCKHOLM GUL	None	None	None	99.85
7	7	2019-04-23	Kortköp 190418 HEMKÖP STOCKHOLM GUL	None	None	None	116.94
8	8	2019-04-23	Kortköp 190417 REST KVARNEN	None	None	None	136.00
9	9	2019-04-23	Kortköp 190419 BEERPOT	None	None	None	78.00
10	10	2019-04-23	Kortköp 190419 PIERIAS LIVS	None	None	None	49.00
11	11	2019-04-23	Kortköp 190420 HEMKÖP STOCKHOLM GUL	None	None	None	49.95
12	12	2019-04-18	Kortköp 190416 HEMKÖP STOCKHOLM GUL	None	None	None	149.09
13	13	2019-04-16	MARTINA.DOM	None	None	None	1000.00

PRESENTATION MANAGING HYBRID DATABASES

Step Five: cleaning the Data, dropping irrelevant columns

```
dataframe.drop(['Kategori', 'Belopp', 'Saldo'], axis=1, inplace=True)
```

dataframe

	Datum	Transaktion	Belopp_pos
0	2019-05-06	Reservation Kortköp HEMKÖP STOCKHOLM G	330.23
1	2019-05-06	MARTIN.DOM	4000.00
2	2019-05-01	Vardagspaket Tillval Betalningar	12.00
3	2019-05-01	Vardagspaket Månadspris kort	24.00
4	2019-04-23	Kortköp 190417 104 MEDBORGARPLATS	128.00
5	2019-04-23	Kortköp 190417 HEMKÖP STOCKHOLM GUL	185.65
6	2019-04-23	Kortköp 190418 HEMKÖP STOCKHOLM GUL	99.85
7	2019-04-23	Kortköp 190418 HEMKÖP STOCKHOLM GUL	116.94

and performing the original search of 66SEK transactions:

```
dataframe.loc[dataframe['Belopp_pos'] == 66.00]
```

	Datum	Transaktion	Belopp_pos
15	2019-04-15	Kortköp 190413 7 ELEVEN 4119113	66.00
29	2019-04-08	Kortköp 190405 DIREKTEN TAREQS TOBA	66.00
46	2019-04-08	Kortköp 190405 SHML KIOSK	66.00
70	2019-03-11	Kortköp 190309 GOTGATSBACKEN	66.00