

# Data Set Transaction Processing Using UPDATE

John Fleming  
Senior Data Analyst

Edmonton SAS User Group  
November 13, 2009

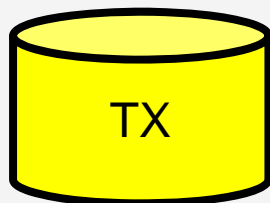


## The Task We Want To Accomplish

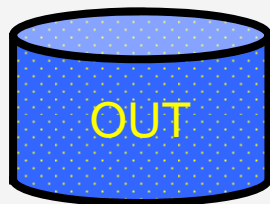
Replace Some Values Of Some Variables In A Master Data Set (MX) With New Values From A Second, Transaction, Data Set (TX)



KEY	VAR1	VAR2	VAR3
1	BLUE	2	Y
2	RED	2	Y



KEY	VAR1	VAR2	VAR3
1		<b>3</b>	
2			<b>N</b>



KEY	VAR1	VAR2	VAR3
1	BLUE	<b>3</b>	Y
2	RED	2	<b>N</b>

## Data Step With MERGE

```
DATA OUT (DROP=VAR1B . . . VARNB);
  MERGE MX (IN=A)
        TX (IN=B RENAME=(VAR1=VAR1B . . . VARN=VARNB));
  BY KEY;

  IF A;

  IF MISSING(VAR1B) = 0 THEN VAR1 = VAR1B;
    .
    .
    .
  IF MISSING(VARNB) = 0 THEN VARN = VARNB;

RUN;
```

## Data Step With UPDATE

```
DATA OUT;  
    UPDATE MX TX;  
    BY KEY;  
  
RUN;
```



**Alberta Health  
Services**

## **MERGE Allows Two Or More Data Sets**

```
DATA OUT;  
    MERGE MX TX AX BX CX DX . . . ;  
    BY KEY;  
  
RUN;
```

## **UPDATE Is Limited To Two Data Sets**

```
DATA OUT;  
    UPDATE MX TX;  
    BY KEY;  
  
RUN;
```



## **BY Group Processing**

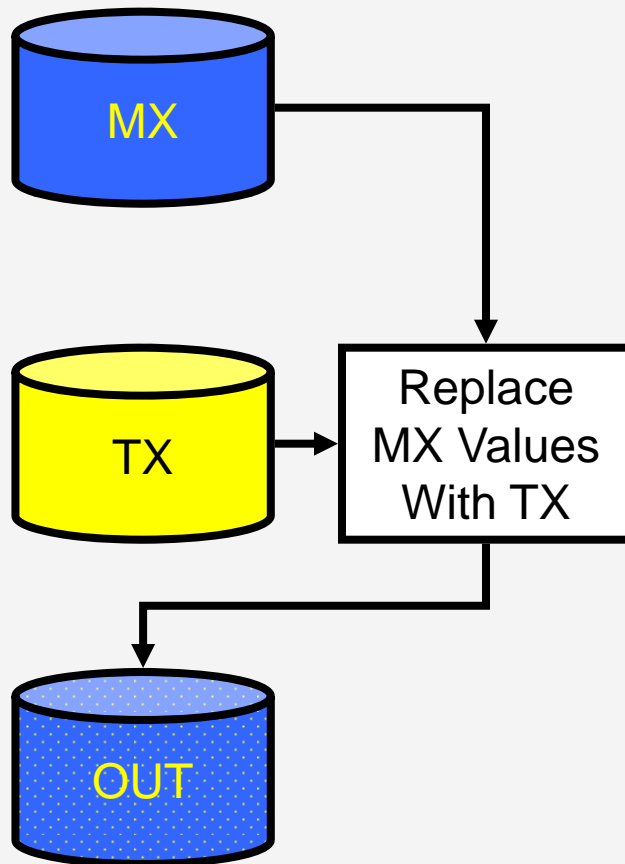
- BY Groups Optional With MERGE
- With UPDATE
  - Data Must Be Sorted With One Or More BY Variables
  - BY Variables In MX Should Be Unique
  - Records With Duplicate BY Variables Can Occur In TX



## **Records With Duplicate BY Variables In TX**

- MERGE Creates One Record In OUT For Each Pair Of Records With Same BY Variable Values In MX And TX
- UPDATE Creates One Updated Record In OUT For All Records With Same BY Variable Values In TX
  - Multiple Records In TX Combined Into A Single Transaction Record
  - Single Transaction Record Applied To MX
  - Record Written To OUT

## Processing Multiple Transaction Records With UPDATE



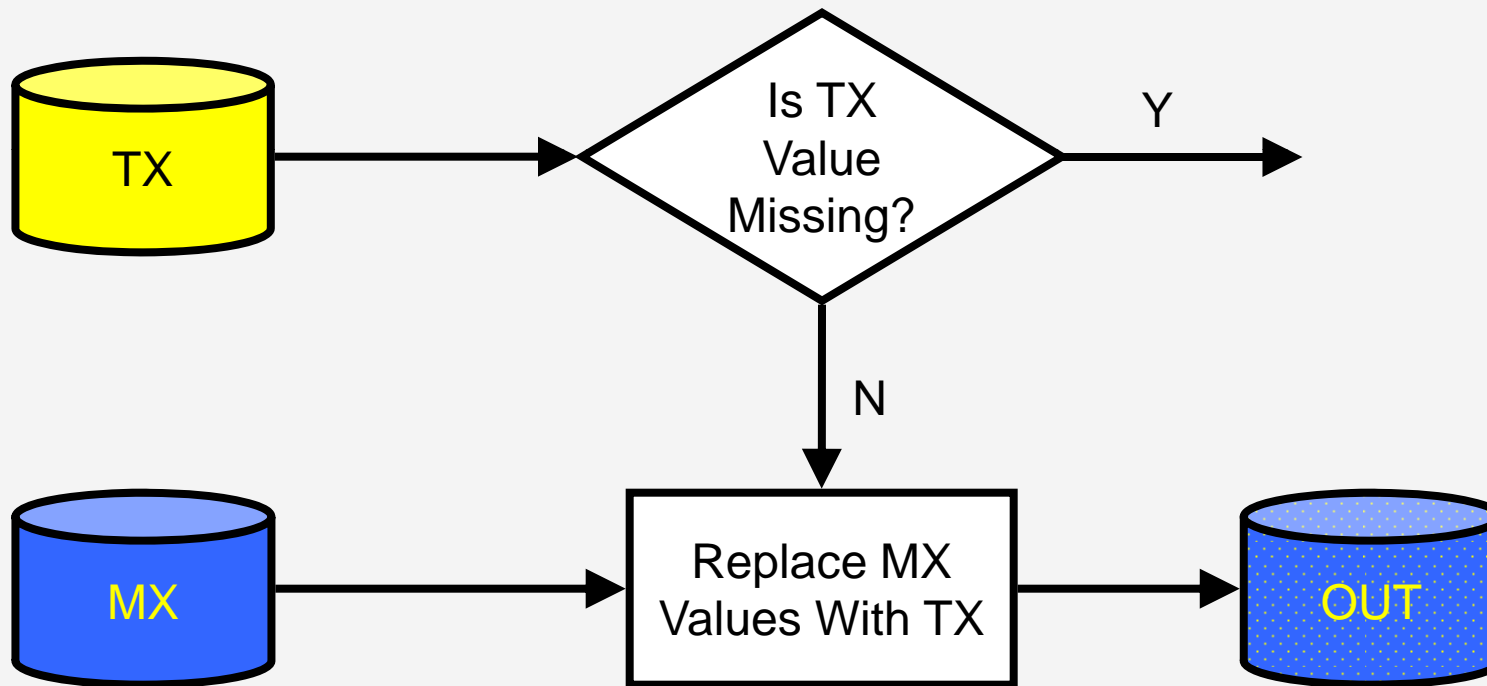
KEY	VAR1	VAR2	VAR3
1	BLUE	2	Y
2	RED	2	Y
3	PINK	4	Y
4	GREEN	1	Y

KEY	VAR1	VAR2	VAR3
4	<b>CYAN</b>	.	<b>X</b>
4	<b>AQUA</b>	<b>6</b>	

KEY	VAR1	VAR2	VAR3
1	BLUE	2	Y
2	RED	2	Y
3	PINK	4	Y
4	<b>AQUA</b>	<b>6</b>	<b>X</b>



## How UPDATE Treats Missing Values





## **Treatment Of Special Missing Values In UPDATE**

- SAS Allows Us To Define Special Missing Values (.A - .Z and .\_) For Numeric Data
- UPDATE Treats These Values As If They Were Non-missing Data

## Treatment Of Special Missing Values In UPDATE (2)

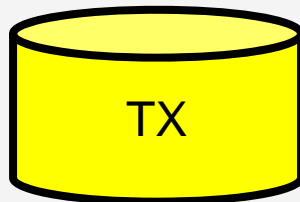
```
DATA OUT;  
  MISSING A;  
  UPDATE MX TX;  
  BY KEY;
```

```
RUN;
```

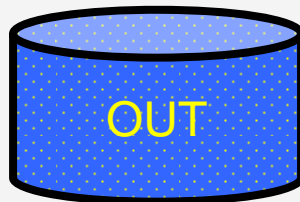
## Treatment Of Special Missing Values In UPDATE (3)



KEY	VAR1	VAR2	VAR3
1	BLUE	2	3



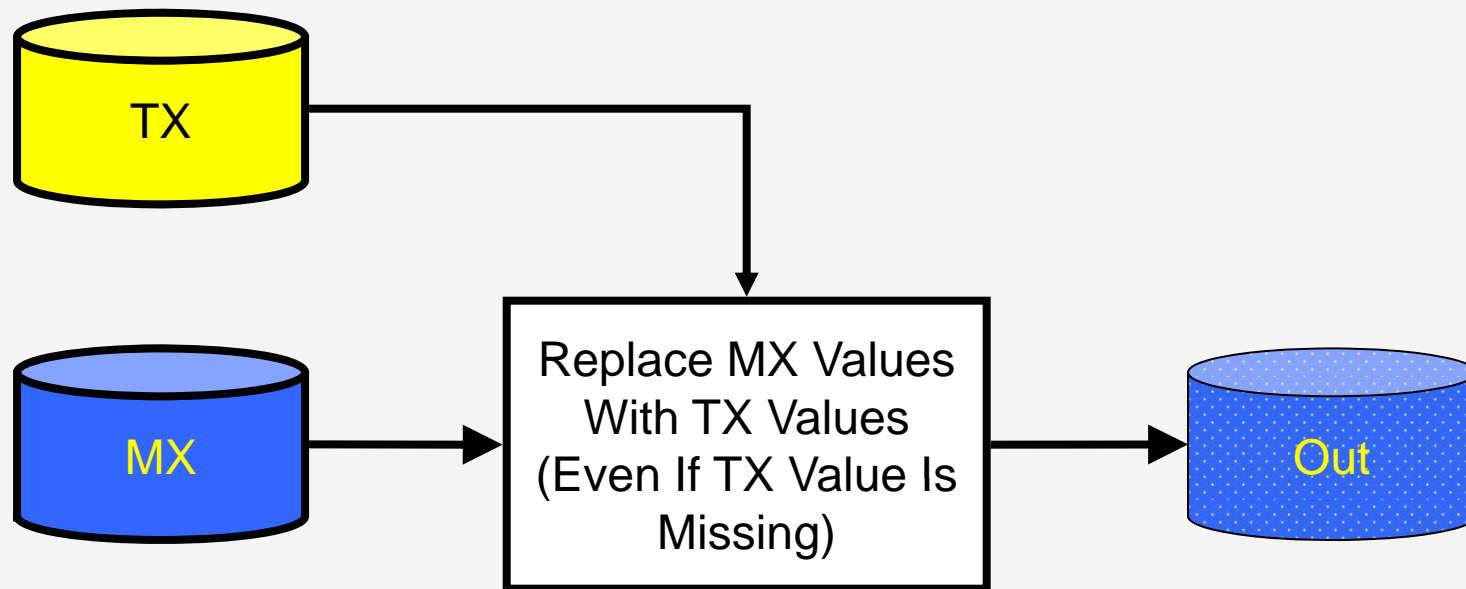
KEY	VAR1	VAR2	VAR3
1		.A	.



KEY	VAR1	VAR2	VAR3
1	BLUE	.A	3

## We Can Override UPDATE Default Behavior

```
DATA OUT;  
  UPDATE MX TX UPDATEMODE=NOMISSINGCHECK;  
  BY KEY;  
  
RUN;
```





**Alberta Health  
Services**

## **Our Regular Data Set Options Are Available**

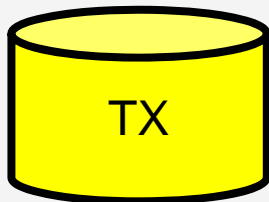
We Can Use KEEP, DROP, WHERE, RENAME And IN

```
DATA OUT;  
    UPDATE MX (RENAME=(SIZE=TUBBINESS)) TX;  
    BY KEY;  
  
RUN;
```

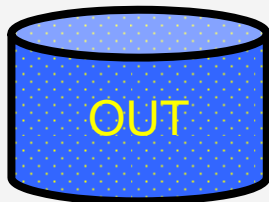
## Add New Variables To MX With UPDATE



KEY	VAR1	VAR2	VAR3
1	BLUE	2	Y
2	RED	2	Y



KEY	VAR1	VAR2	VAR3	VAR4
1		<b>3</b>		<b>BIG</b>
2			<b>N</b>	<b>SMALL</b>



KEY	VAR1	VAR2	VAR3	VAR4
1	BLUE	<b>3</b>	Y	<b>BIG</b>
2	RED	2	<b>N</b>	<b>SMALL</b>



**Alberta Health  
Services**

## Regular Data Step Processing Is Available

```
DATA OUT;  
  UPDATE MX TX;  
  BY KEY;  
  
  IF GST_EXEMPT = 'N' THEN GST = PRICE * 0.05;  
  ELSE GST = 0;  
  
  TOTAL_PRICE = PRICE + GST;  
  
RUN;
```





**Alberta Health  
Services**

## Questions?

John Fleming

John.Fleming@albertahealthservices.ca

780.643.4341