













Using Suprlink to Combine Data Files without Keys

- ◆ **A Robelle Tutorial
Interex 1995
Toronto, Canada
August 15 - 18, 1995**

Copyright 1995, Robelle Solutions Technology Inc.

What's Inside

	<u>Page</u>
 Linking information manually 	3
 Linking information using Suprlink 	9
 The linking order makes a big difference 	14
 What if the link value is missing? 	17
 How to select entries that fail to link? 	18
 Understanding self-describing files 	21

Let's print the customers in B.C.

```
>get m-customer
>if state-code="BC"
>extract cust-account,name-first,name-last
>list standard,title "Customers in BC"
>xeq
```

```
Apr 10, 1995 9:41 Customers in BC Page 1
```

CUST-ACCO	NAME-FIRST	NAME-LAST
10010	Wayne	Humphreys
10014	Elizabeth	Welton
10011	William	Kirk
10012	Percy	Ferguson
10020	Walley	Nisbet
10002	Gordon	Lackner
10001	Darlene	Hamilton
10003	John	Melander

Finding dispersed information


 insert page 4 from newlink2.ppt

Does d-sales contain purchase date information?

>form d-sales

Database: STORE.DEMO.ROBELLE

D-SALES Detail Set 5

Entry:		Offset	
CUST-ACCOUNT	Z8	1	(!M-CUSTOMER)
DELIV-DATE	J2	9	
PRODUCT-NO	Z8	13	(M-PRODUCT)
PRODUCT-PRICE	J2	21	
PURCH-DATE	J2	25	
 SALES-QTY	J1	29	
SALES-TAX	J2	31	
SALES-TOTAL	J2	35	

Capacity: 602 (14) Entries: 8 Bytes: 38

 Which field is also in the m-customer dataset?

Let's list the October purchases

```
>get d-sales  
>if purch-date >= 19931001 and purch-date <= 19931031  
>extract purch-date,cust-account  
>list standard,title "Purchases in October 93"  
>xeq
```

```
Apr 11, 1995 8:31 Purchases in October 93 Page 1
```

PURCH-DATE	CUST-ACCO
19931001	10020
19931015	10003
19931015	10003
19931015	10003
19931021	10016
19931021	10016
19931028	10020
19931020	10010

Here are the sorted listings

 Can you find the October purchases by B.C. customers?

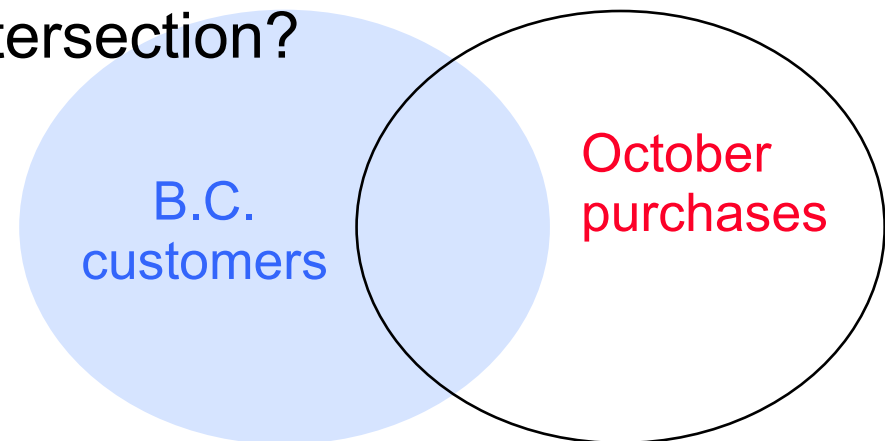
Purchases in October 93 Customers in BC

PURCH-DATE	CUST-ACCO	CUST-ACCO	NAME-FIRST	NAME-LAST	
19931015	10003---	10001	Darlene	Hamilton	
19931015	10003---	10002	Gordon	Lackner	
19931015	10003-----	>10003	John	Melander	
		19931020 10010---		10008	Thomas
		Serafin			
19931021	10016	10009	Gordon	Oxenbury	
19931021	10016 -----	>10010	Wayne	Humphreys	
		19931001 10020---		10011	William
		Kirk			
19931028	10020---	10012	Percy	Ferguson	
		10013	Colin	Andersen	
		10014	Elizabeth	Welton	
		10019	Rupert	Hillstrom	
	-----	>10020	Walley	Nisbet	

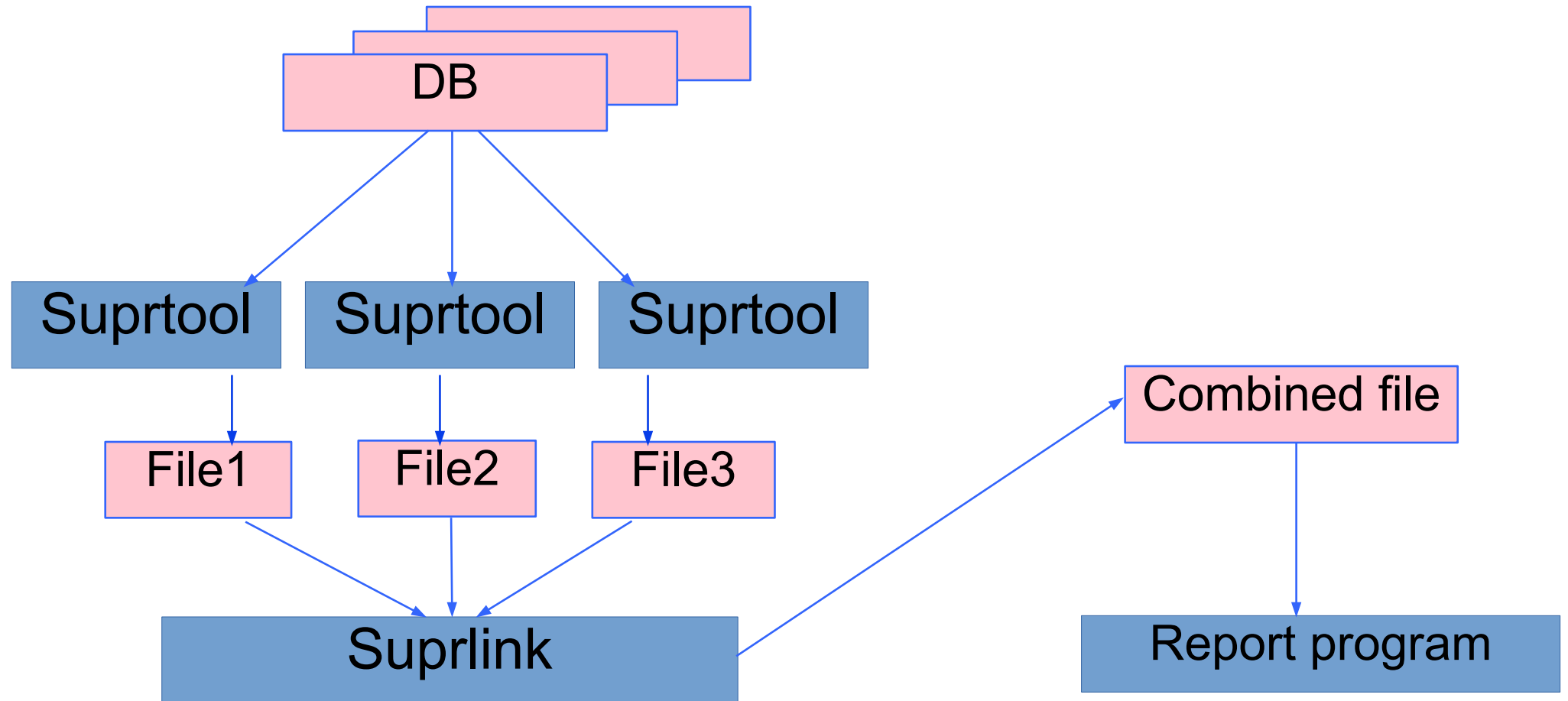
The intersection of your two reports

PURCH-DATE	CUST-ACCO	CUST-ACCO	NAME-FIRST	NAME-LAST
19931015	10003	10003	John	Melander
19931015	10003	10003	John	Melander
19931015	10003	10003	John	Melander
19931020	10010	10010	Wayne	Humphreys
19931001	10020	10020	Walley	Nisbet
19931028	10020	10020	Walley	Nisbet

 How can the computer select the intersection?



Suprlink ties your data together



Extract October purchases into an SD file

```
>get d-sales  
>if purch-date >= 19931001 and purch-date <= 19931031  
>extract purch-date,product-no,cust-account  
  
>sort cust-account  
>output tempsale, temp, LINK  
  
>xex
```

 Now create second SD file containing B.C. customers

Now extract the B.C. customers

```
>get m-customer  
>if state-code="BC"  
>extract cust-account,name-first,name-last  
>sort cust-account  
>output tempcust, temp, link  
>xeq
```



Final Step: Link the two SD files together

```
>link input tempsale  
>link link tempcust  
>link output tempfile, temp  
>link xeq
```

October Purchases by BC Customers

PURCH-DATE	PRODUCT-N	CUST-ACCO	NAME-FIRST	NAME-LAST
19931015	50511501	10003	John	Melander
19931015	50512501	10003	John	Melander
19931015	50513001	10003	John	Melander
19931020	50533001	10010	Wayne	Humphreys
19931001	50511501	10020	Walley	Nisbet
19931028	50512501	10020	Walley	Nisbet

A review of the entire Suprlink task

 insert page 13 from newlink2.ppt

What happens if we reverse the linking order?

❑ What if we link Tempsale to Tempcust?

```
>link input tempcust
```

```
>link link tempsale
```

```
>link output treverse, temp
```

```
>link xeq
```

```
Apr 11, 1995 14:43 October Purchases by BC Customers Page 1
```

CUST-ACCO	NAME-FIRST	NAME-LAST	PURCH-DATE	PRODUCT-N
10003	John	Melander	19931015	50511501
10010	Wayne	Humphreys	19931020	50533001
10020	Walley	Nisbet	19931001	50511501

❑ There are only 3 records (one per customer) instead of 6

Student Exercise

 Enhance Tempfile by adding the product description to each purchase

 Hint: The command FORM M-PRODUCT gives you this information:

```
Database: STORE.DEMO.ROBELLE
```

```
M-PRODUCT      Master      Set 2
```

```
Entry:          Offset
```

```
PRODUCT-DESC   X30      1
```

```
PRODUCT-MODEL  X10      31
```

```
PRODUCT-NO     Z8       41 <<Search Field>>
```

```
Capacity: 307 (12) Entries: 13 Bytes: 48
```

Solution to Student Exercise

 insert page 16 from newlink2.ppt

What if there is no product description?

- ❑ By default, Suprlink drops records that fail any match
- ❑ To override default, use the OPTIONAL keyword on the LINK command
- ❑ Exercise Solution using the OPTIONAL description
 - >link input tempfile
 - >link link tempprod, optional
 - >link output temptest,temp
 - >link xeq
- ❑ Suprlink provides zero or blank values if link record is missing

How do you find the entries that don't match?

ANSWER: Append a dummy field to each record

Purchases missing a standard product description

Define Flag,1,2

Extract product-no, product-desc, flag = "xx"

Link optional; missing links have Flag = " "

Re-input the output file, apply selection:

```
If Flag <> "xx"          {product-desc not found}
```

Finding information from several datasets

- ❑ Pattern-match on m-product dataset; find customers who bought any drill
- ❑ Sort d-sales by product-no and cust-account; remove duplicates
- ❑ Link drills to sales; only select matches
- ❑ Sort all customer names by account number
- ❑ Re-sort drill purchases by account number
- ❑ Link customers to drill sales

Link from m-product to m-customer through d-sales

 insert page 20 from newlink2.ppt

What is a self-describing file?

- ☐ Regular file plus a mini-dictionary
- ☐ Accepted in INPUT command
- ☐ Allows use of full field names in IF, EXTRACT, etc.
- ☐ Created by OUTPUT command in two flavors:
 - > `output archfile, link`
 - > `output extrfile, query`

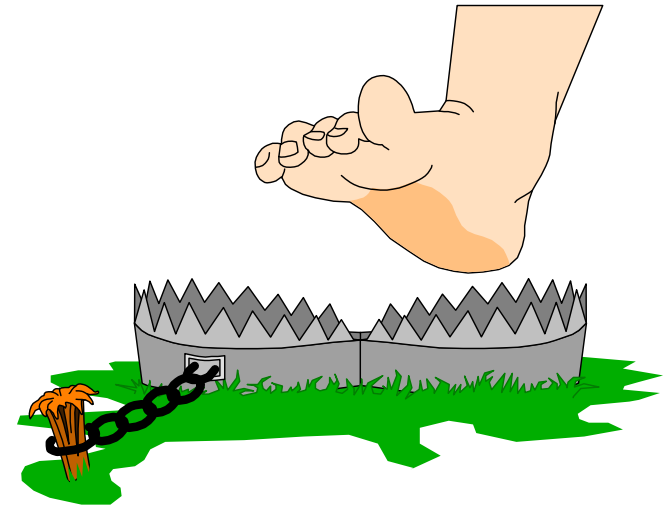
SD files have a record structure

```
>get m-customer
>if custstatus = 10,20
>extract custnumber,custname,custstatus
>output mysdfile,link
>xeq

>form mysdfile      {this file now has structure!}
>input mysdfile
>if custstatus = "10"
>sort custname
>extract custnumber
>list standard
>xeq
```

Suprlink warnings

- Consumes a lot of disc space
- Extract files must be sorted
- Link fields of any type except real or long real
- File restrictions

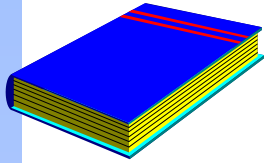


File restrictions

Maximum	Input file	Link files	Output file
No. of files	1	7	1
No. of fields	255	255	1023
Record size (words)	2048	2048	4096
Block size (words)	4096	2048	4096

Performance guidelines for Suprlink

- ❑ Don't sort and merge if you must re-sort many times
- ❑ Keep record sizes small by only extracting needed fields
- ❑ Keep files small by only writing selected records
- ❑ Use serial scan with table lookup for sets of key values
- ❑ Use keyed access when number of key values is small
- ❑ It can be faster to sort the entire dataset



Summary of Linking without indexing

- ❑ Links sorted data by a common key field
- ❑ Drops records that do not match
- ❑ Combines records and eliminates duplicate key fields
- ❑ Reads and creates self-describing files
- ❑ SD files are like little databases
- ❑ Suprlink is a standard part of Suprtool