

Introduction to Databases Worksheets

Peter M^cBrien

Sunday 8th October 2023

Bank Branch Database

Many examples in the lectures and worksheets will use the `bank_branch` database, illustrated below.

branch		
<u>sortcode</u>	bname	cash
56	'Wimbledon'	94340.45
34	'Goodge St'	8900.67
67	'Strand'	34005.00

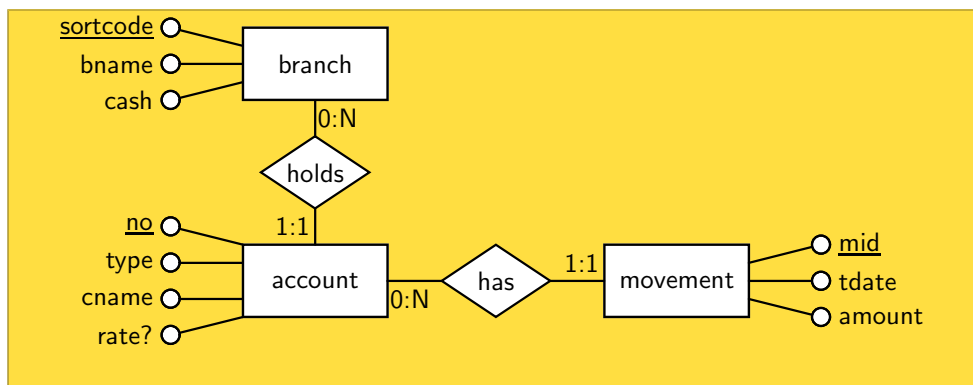
movement			
<u>mid</u>	no	amount	tdate
1000	100	2300.00	5/1/1999
1001	101	4000.00	5/1/1999
1002	100	-223.45	8/1/1999
1004	107	-100.00	11/1/1999
1005	103	145.50	12/1/1999
1006	100	10.23	15/1/1999
1007	107	345.56	15/1/1999
1008	101	1230.00	15/1/1999
1009	119	5600.00	18/1/1999

account				
<u>no</u>	type	cname	rate?	sortcode
100	'current'	'McBrien, P.'	NULL	67
101	'deposit'	'McBrien, P.'	5.25	67
103	'current'	'Boyd, M.'	NULL	34
107	'current'	'Poulovassilis, A.'	NULL	56
119	'deposit'	'Poulovassilis, A.'	5.50	56
125	'current'	'Bailey, J.'	NULL	56

key branch(sortcode)
 key branch(bname)
 key movement(mid)
 key account(no)
 $\text{movement}(\text{no}) \xrightarrow{fk} \text{account}(\text{no})$
 $\text{account}(\text{sortcode}) \xrightarrow{fk} \text{branch}(\text{sortcode})$

This database records branches of a bank, each one of which may be identified by their sort code (underlined to show it is the primary key) or by their name. Each branch may have a number of accounts, each identified by their account number (`no`), with details of the type of the account, customer name, and the interest `rate` for the account. The fact that `rate` is not a mandatory attribute (in SQL terms, it is nullable) is indicated by the use of a question mark after the attribute name. Each account may have a number of movements identified by their movement identifier (`mid`).

The relational database may be represented by an ER model shown below, which makes clear the fact each movement is associated to exactly one account, and each account is associated to exactly one branch.



The database is available to use on the departmental Postgres server. To connect to the Postgres version, execute on any Linux CSG lab machine:

```
psql -h db.doc.ic.ac.uk -U lab -d bank_branch -W
```

IDB Worksheet 1: Primitive Relational Algebra Operators

The following questions relate to queries over the bank_branch database.

1. What is the result of the RA query $\pi_{\text{cname}, \text{sortcode}} \sigma_{\text{type}='deposit'} \text{account}$

cname	sortcode
'McBrien, P.'	67
'Poulouvassilis, A.'	56

2. What is the result of the RA query

$\pi_{\text{account.no}, \text{cname}, \text{mid}} \sigma_{\text{account.no}=\text{movement.no} \wedge \text{movement.amount} < 0} (\text{account} \times \text{movement})$

no	cname	mid
100	'McBrien, P.'	1002
107	'Poulouvassilis, A.'	1004

3. Write an RA query to return the scheme (cname) that lists the cname of all customers that have made a withdrawal from an account.

$\pi_{\text{cname}} \sigma_{\text{account.no}=\text{movement.no} \wedge \text{movement.amount} < 0} (\text{account} \times \text{movement})$
can be amount < 0, as it's unambiguous

4. Write an RA query to return the scheme (cname, bname) that lists the cname of all deposit account holders, together with bname where the account is held.

$\pi_{\text{cname}, \text{bname}} \sigma_{\text{branch.sortcode}=\text{account.sortcode} \wedge \text{type}='deposit'} (\text{account} \times \text{branch})$

5. Write an RA query to return the scheme (sortcode) that lists the sortcodes of branches that either have less than £10,000 of cash, or that hold deposit accounts.

$\pi_{\text{sortcode}} \sigma_{\text{cash} < 10000} \text{branch} \cup \pi_{\text{branch.sortcode}} \sigma_{\text{branch.sortcode}=\text{account.sortcode} \wedge \text{type}='deposit'} \text{account}$
union operator
not necessary as duplicates automatically removed

6. Write an RA query to return the scheme (sortcode) that lists those branches where no deposit account is held

$\pi_{\text{sortcode}} \sigma_{\text{type}='current'} \text{account}$

$\pi_{\text{sortcode}} \text{branch} \ominus \pi_{\text{sortcode}} \sigma_{\text{type}='deposit'} \text{account}$
negation, so think about the DIFFERENCE operator

as relations can't have duplicates

IDB Worksheet 2: Derived Relational Algebra Operators

The following questions relate to queries over the `bank_branch` database.

1. What is the result of the RA query $\pi_{\text{bname}, \text{cname}}(\text{branch} \bowtie \text{account})$
2. What is the result of the RA query $\pi_{\text{bname}, \text{cname}}(\text{branch} \bowtie \text{account} \bowtie \text{movement})$
3. What is the result of the RA query $\pi_{\text{sortcode}, \text{type}} \text{account} \div \pi_{\text{sortcode}} \text{branch}$
4. Write an RA query returning the scheme $(\text{bname}, \text{mid})$ that lists branch names, together with mids that have occurred at that branch.
5. Write an RA query returning the scheme $(\text{sortcode}, \text{bname}, \text{cash})$ that lists rows of all branches that have at least one deposit account.
6. Write an RA query returning the scheme (sortcode) that lists the sortcodes of branches that have both have less than £10,000 of cash, and that hold deposit accounts.

IDB Worksheet 3: Equivalences Between RA Expressions

The following questions relate to queries over the `bank_branch` database.

- Put a tick or cross to indicate if each equivalence holds:

Equivalence	Correct?
$\pi_{no,type} \sigma_{type='deposit'} \text{account} \equiv \sigma_{type='deposit'} \pi_{no,type} \text{account}$	
$\pi_{no,type} \sigma_{type='deposit'} \text{account} \equiv \sigma_{type='deposit'} \pi_{type,no} \text{account}$	
$\pi_{type} \sigma_{type='deposit'} \text{account} \equiv \sigma_{type='deposit'} \pi_{no} \text{account}$	
$\sigma_{sortcode=56}(\text{account} \times \text{movement}) \equiv \sigma_{sortcode=56} \text{account} \times \text{movement}$	
$\pi_{sortcode}(\text{account} \times \text{movement}) \equiv \pi_{sortcode} \text{account} \times \text{movement}$	
$\pi_{no,type} \sigma_{type='deposit'} \text{account} \equiv \pi_{no,type} \sigma_{type <> 'current'} \text{account}$	

- Simplify the following RA query to contain as few RA operators as possible

$\pi_{no,type} \sigma_{sortcode=56} \pi_{no,type,sortcode} \sigma_{type='deposit'} \text{account}$

- Rewrite the following RA query into the form $\pi_{...} \sigma_{...} R \times S$, the general form of which we will call a **project select product (PSP)** query.

$\sigma_{\text{account.no}=\text{movement.no}}(\pi_{no,cname} \text{account} \times \pi_{mid,no} \sigma_{\text{amount}>1000} \text{movement})$

- Rewrite the following RA to be a union between two PSP queries

$\sigma_{\text{account.no}=\text{movement.no}}(\pi_{no,cname,rate} \text{account} \times (\pi_{mid,no} \sigma_{\text{amount}>1000} \text{movement} \cup \pi_{mid,no} \sigma_{\text{amount}<100} \text{movement}))$

- Rewrite the following query to an equivalent form that minimises the number of tuples, and the number of attributes within those tuples, that are handled by the \times operator.

$\pi_{no,cname,tdate} \sigma_{\text{amount}<0 \wedge \text{account.no}=\text{movement.no}}(\text{account} \times \text{movement})$

IDB Worksheet 4: Translating Between RA and SQL

The following questions relate to queries over the `bank_branch` database.

1. Write the RA expression using project, select and product operators equivalent to the SQL query below.

```
SELECT account.cname,  
       movement.amount  
FROM   account  
       JOIN movement  
ON     account.no=movement.no  
WHERE  account.rate>2.0  
AND    movement.amount>1000
```

2. Modify your RA expression to use any other RA operators you have been shown to write a minimal RA expression (*i.e.* one that uses as few operators as possible).
3. Write a minimal SQL query (*i.e.* as compact as possible) equivalent to the RA expression $\pi_{no} \text{ movement} - \pi_{no} \text{ account}$
4. Write a minimal SQL query equivalent to the RA expression $\pi_{no,type} \text{ account}$
5. Write a minimal SQL query equivalent to the RA expression $\pi_{type} \text{ account}$
6. Write a minimal SQL query equivalent to the RA expression $\pi_{type,cname} \text{ account}$
7. Write a minimal SQL query equivalent to the RA expression $\text{account} \times \text{movement}$

IDB Worksheet 5: SQL Set Operators

The following questions relate to queries over the `bank_branch` database.

1. Write an SQL query returning the scheme (mid,no,amount,tdate) listing all details of movements for accounts 100,101,103 and 107.
2. Write an SQL query returning the scheme (sortcode) listing the sortcode of all branches without any deposit accounts.
3. Write an SQL query without using any negation (*i.e.* without the use of **NOT** or **EXCEPT**) returning the scheme (no) listing accounts with no movements on or before the 11-Jan-1999.
4. Write an SQL query returning the scheme (cname) listing customers that have every type of account that appears in `account`.

IDB Worksheet 6: Null Values in SQL

In a modified version of the **bank_branch** database, called **bank_branch_null**, there are the following two tables:

movement			
mid	no?	amount?	tdate?
0999	119	45.00	null
1000	100	2300.00	5/1/1999
1001	101	4000.00	5/1/1999
1002	100	-223.45	8/1/1999
1004	107	-100.00	11/1/1999
1005	103	145.50	12/1/1999
1006	100	10.23	15/1/1999
1008	101	1230.00	15/1/1999
1009	119	5600.00	18/1/1999
1010	100	null	20/1/1999
1011	null	null	20/1/1999
1012	null	600.00	20/1/1999
1013	null	-46.00	20/1/1999

account				
no	type	cname	rate?	sortcode
100	'current'	'McBrien, P.'	null	67
101	'deposit'	'McBrien, P.'	5.25	67
119	'deposit'	'Poulovassilis, A.'	5.50	56
125	'current'	'Bailey, J.'	null	56

1. Write an SQL query returning the scheme (mid) to find movements known not to have occurred on 5/1/1999
2. Write an SQL query returning the scheme (mid) to find movements that have or might have occurred on 5/1/1999.
3. Write an SQL query returning the scheme (no,mid) that lists account numbers, and any movements mids that have or might have occurred on that account
4. What is the result of the following query:

```
SELECT  account.no
FROM    account
WHERE   account.no NOT IN (SELECT movement.no FROM movement)
```
5. Write an SQL query returning the scheme (no) that lists those **account** numbers that might not have any movements.

IDB Worksheet 7: Left, Right, Outer and Inner Joins

The following question uses the **bank_branch_null** database:

movement			
mid	no?	amount?	tdate?
0999	119	45.00	null
1000	100	2300.00	5/1/1999
1001	101	4000.00	5/1/1999
1002	100	-223.45	8/1/1999
1004	107	-100.00	11/1/1999
1005	103	145.50	12/1/1999
1006	100	10.23	15/1/1999
1008	101	1230.00	15/1/1999
1009	119	5600.00	18/1/1999
1010	100	null	20/1/1999
1011	null	null	20/1/1999
1012	null	600.00	20/1/1999
1013	null	-46.00	20/1/1999

account				
no	type	cname	rate?	sortcode
100	'current'	'McBrien, P.'	null	67
101	'deposit'	'McBrien, P.'	5.25	67
119	'deposit'	'Poulovassilis, A.'	5.50	56
125	'current'	'Bailey, J.'	null	56

1. What is the result of

```
SELECT account.no, movement.mid
FROM account NATURAL LEFT JOIN movement
```

2. What is the result of

```
SELECT account.no, movement.mid
FROM account NATURAL FULL OUTER JOIN movement
```

3. Write an SQL query returning the scheme (no,cname,mid) that lists all account numbers in the database (including just those in movement), and lists any known mid or cname for each account. The result for the current data should be:

no	cname	mid
100	McBrien, P.	1000
100	McBrien, P.	1006
100	McBrien, P.	1010
100	McBrien, P.	1002
101	McBrien, P.	1008
101	McBrien, P.	1001
103	null	1005
107	null	1004
119	Poulovassilis, A.	999
119	Poulovassilis, A.	1009
125	Bailey, J.	null

IDB Worksheet 8: OLAP Queries in SQL

The following questions should be written to run on the **bank_branch** database.

1. Write an SQL query returning the scheme (no,balance,avg_trans) that lists for each account that has transactions the account no, the balance (computed as the sum of the movements for the account), and the average value of transactions on that account.
2. Alter the query for (1) so that it includes the customer name of each account, and includes *all* accounts held at the bank in the listing of balances, even if the account has no movements.
3. Write an SQL query returning the scheme (cname,current_balance,deposit_balance) that lists one row for each customer (*i.e.* each distinct cname), with a column for the net balance of all current accounts held by the customer, and a column for the net balance of all deposit accounts held by the customer.
4. Write an SQL query returning the scheme (no,cname,type,pc_cust_funds,pc_type_funds) that lists one row for each account, and for each account, lists the no, cname and type of the account, and in pc_cust_funds the percentage of the customer funds held in the account, and in pc_type_funds the percentage of the total funds in this particular type of account. For the current data this should result in:

no	cname	type	pc_cust_funds	pc_type_funds
100	McBrien, P.	current	28.52	84.22
101	McBrien, P.	deposit	71.48	48.29
103	Boyd, M.	current	100.00	5.87
107	Poulovassilis, A.	current	4.20	9.91
119	Poulovassilis, A.	deposit	95.80	51.71
125	Bailey, J.	current	NULL	0.00

IDB Worksheet 9: Datalog

The following questions relate to queries over the `bank_branch` database.

1. Write a Datalog query returning the scheme `(cname,bname)` listing the `cname` of all deposit account holders, together with `bname` where the account is held.
2. Write a Datalog query returning the scheme `(no,bname)` listing the account number and branch name of all accounts that have no movements recorded for the account.
3. Write a Datalog query that returns the scheme `(no)` listing the 'target account' numbers, defined as those accounts that have made a withdrawal, or are of type deposit.

4. List what the following Datalog query returns, and explain its semantics.

```
query(CName) :-  
    account(_, _, CName, _, _),  
    ¬sub_query(CName).  
sub_query(CName) :-  
    account(_, _, CName, _, _),  
    account(_, Type, _, _, _),  
    ¬account(_, Type, CName, _, _).
```

IDB Worksheet 10: Constructing an ER^{KLMOS} Schema

The following text gives a description of a UoD, which you have been asked to build a relational database that stores data associated with the UoD.

The payroll system for *BIG Inc* records the salaries, status, joining date, name, and payroll number for all of the corporations 30,000 employees. Each employee works for one division, and each division has an account number for paying its staff. We identify divisions by their name, and record the address where the division's HQ is located.

For employees sent abroad by *BIG Inc*, we record the address, country and telephone number of the foreign tax office that will handle the employee. It is assumed that each country has one central tax office that we have to deal with. All other employees have their tax affairs dealt with by the Inland Revenue.

Draw an ER^{KLMOS} schema of the UoD

Map your ER schema into a relational schema

IDB Worksheet 11: Constructing an ER^{ADHKLMNOSVW} Schema

The following text gives a description of a UoD, which you have been asked to build a relational database that stores data associated with the UoD.

The customer and supplier database of *Big Inc* will hold all accounts of the company, divided into customer accounts and supplier accounts. All accounts have an account number, and one account manager assigned from the company's staff. *Big Inc* identifies staff by a *sid*, and records the staff member's name and room. The account managers have a limit on the number of accounts they can manage. Only certain staff members are permitted to be account managers.

For customer accounts we need to record a credit limit on the balance of the account, and the telephone number of the accounts department at the customer.

For supplier accounts we need to record which *Big Inc* products are supplied, and at what price.

Big Inc products are identified by the company standard **part_no** and all have a description. For some we record the colour. Some products have a record of the components, each component identified by a combination of **part_no** and component number, and again each has a description. Some products do not have a supplier.

Big Inc has purchased a copy of the Post Office address file, and associates every account to an address from this file. The address data includes street number, street name, town, county and post code, and uses a combination of street number and post code as a key.

Draw an ER^{ADHKLMNOSVW} schema of the UoD

Map your ER schema into a relational schema

IDB Worksheet 12: Minimal Cover of FDs and Candidate Keys

Suppose a relation $R(A, B, C, D, E, F, G, H)$ has the FDs

$$S = \{AB \rightarrow DEH, BEF \rightarrow A, FGH \rightarrow C, D \rightarrow EG, EG \rightarrow BF, F \rightarrow BH\}$$

1. Rewrite S to an equivalent set of FDs which only have a single attribute on the RHS of each FD.
2. Consider each FD $X \rightarrow A$, and for each $B \in X$, consider if $(X - B) \rightarrow B$ from the other FDs. If so, replace $X \rightarrow A$ by $(X - B) \rightarrow A$ in S .
3. Consider each FD $X \rightarrow A$, and compute X^+ without using $X \rightarrow A$. If $A \subseteq X^+$, delete $X \rightarrow A$ since it is redundant. This will give a minimal cover S_c of S .
4. Justify what are the minimal candidate keys of R constrained by S_c

IDB Worksheet 13: Lossless Decomposition of Relations

1. Suppose relation $R(A, B, C, D, E)$ has the FDs $S = \{AB \rightarrow C, C \rightarrow DE, E \rightarrow A\}$. For each of the following decompositions, determine if the decomposition is lossless, and if not lossless, illustrate a dataset for R that fails to decompose in a lossless manner over the relations.

(a) $R_1(A, B, C), R_2(C, D, E)$

(b) $R_1(A, B, C), R_2(C, D), R_3(D, E)$

2. Suppose relation $R(A, B, C, D, E, F)$ has the FDs $S = \{AB \rightarrow CD, C \rightarrow E, A \rightarrow F\}$. Give a lossless decomposition of R into three relations R_1 , R_2 , and R_3 .
3. Suppose relation $R(A, B, C, D, E, F)$ has the FDs $S = \{AB \rightarrow CD, C \rightarrow E, F \rightarrow A\}$. Give a lossless decomposition of R into three relations R_1 , R_2 , and R_3 .

IDB Worksheet 14: Normal Forms

Suppose a relation $R(A, B, C, D, E, F, G, H)$ has the FDs

$$S_c = \{AB \rightarrow D, EF \rightarrow A, FG \rightarrow C, D \rightarrow EG, EG \rightarrow F, F \rightarrow BH\}$$

1. Decompose the relation into 3NF
2. Decompose the relation into BCNF
3. Determine if your decompositions in (1) and (2) preserve FDs, and if they do not, suggest how to amend your schema to preserve FDs.

IDB Worksheet 15: Anomalies in Transactions

```
BEGIN TRANSACTION rental_charge
  UPDATE directory
  SET charge=charge+17
COMMIT TRANSACTION rental_charge
```

```
BEGIN TRANSACTION transfer_charge
  UPDATE directory
  SET charge=charge+100
  WHERE telephone=1000

  UPDATE directory
  SET charge=charge-100
  WHERE telephone=1002
COMMIT TRANSACTION transfer_charge
```

```
BEGIN TRANSACTION total_charge
  SELECT SUM(charge)
  FROM directory
COMMIT TRANSACTION total_charge
```

directory		
telephone	name	charge
1000	Adams	10.00
1001	Jones	120.25
1002	Black	344.00

rental_charge $H_1 = r_1[d_{1000}], w_1[d_{1000}], r_1[d_{1001}], w_1[d_{1001}], r_1[d_{1002}], w_1[d_{1002}]$

transfer_charge $H_2 = r_2[d_{1000}], w_2[d_{1000}], r_2[d_{1002}], w_2[d_{1002}]$

total_charge $H_3 = r_3[d_{1000}], r_3[d_{1001}], r_3[d_{1002}]$

For each of the following histories, identify if they are a concurrent execution of some pair of the above histories, and if so, then determine if any of the following three anomalies has occurred:

- A lost update
- B inconsistent analysis
- C dirty read

1. $r_3[d_{1000}], r_1[d_{1000}], w_1[d_{1000}], r_1[d_{1001}], w_1[d_{1001}], r_1[d_{1002}], w_1[d_{1002}], c_1, r_3[d_{1001}], r_3[d_{1002}], c_3$
2. $r_1[d_{1000}], r_1[d_{1001}], r_1[d_{1002}], r_2[d_{1000}], w_2[d_{1000}], r_2[d_{1002}], w_2[d_{1002}], w_1[d_{1000}], w_1[d_{1001}], w_1[d_{1002}], c_1, c_2$
3. $r_1[d_{1000}], r_2[d_{1000}], w_1[d_{1000}], r_1[d_{1001}], w_1[d_{1001}], r_1[d_{1002}], w_2[d_{1000}], r_2[d_{1002}], w_1[d_{1002}], w_2[d_{1002}], c_1, c_2$
4. $r_3[d_{1000}], r_3[d_{1001}], r_2[d_{1000}], w_2[d_{1000}], r_3[d_{1002}], r_2[d_{1002}], w_2[d_{1002}], c_2, a_3$
5. $r_2[d_{1000}], w_2[d_{1000}], r_1[d_{1000}], r_2[d_{1002}], w_2[d_{1002}], a_2, w_1[d_{1000}], r_1[d_{1001}], w_1[d_{1001}], r_1[d_{1002}], w_1[d_{1002}], c_1$

IDB Worksheet 16: Serialisability

$$H_1 = r_1[o_1], w_1[o_1], w_1[o_2], w_1[o_3], c_1$$

$$H_2 = r_2[o_2], w_2[o_2], w_2[o_1], c_2$$

$$H_3 = r_3[o_1], w_3[o_1], w_3[o_2], c_3$$

$$H_x = r_1[o_1], w_1[o_1], r_2[o_2], w_2[o_2], w_2[o_1], c_2, w_1[o_2], r_3[o_1], w_3[o_1], w_3[o_2], c_3, w_1[o_3], c_1$$

$$H_y = r_3[o_1], w_3[o_1], r_1[o_1], w_1[o_1], w_3[o_2], c_3, w_1[o_2], r_2[o_2], w_2[o_2], w_2[o_1], c_2, w_1[o_3], c_1$$

1. Determine what are the ordered conflicting pairs in H_x , and write them down in the form $rw_i[o] \rightarrow rw_j[o]$ (where the rw should be replaced by a r or w in your answer as appropriate). Use your answer to determine if H_x is CSR.

2. Determine the conflicting pairs in H_y , and use your answer to determine if H_y is CSR.

IDB Worksheet 17: Recoverability

$$H_w = r_2[o_1], r_2[o_2], w_2[o_2], r_1[o_2], w_2[o_1], r_2[o_3], c_2, c_1$$

$$H_x = r_2[o_1], r_2[o_2], w_2[o_1], w_2[o_2], w_1[o_1], w_1[o_2], c_1, r_2[o_3], c_2$$

$$H_y = r_2[o_1], r_2[o_2], w_2[o_2], r_1[o_2], w_2[o_1], c_1, r_2[o_3], c_2$$

$$H_z = r_2[o_1], w_1[o_1], r_2[o_2], w_2[o_2], r_2[o_3], c_2, r_1[o_2], w_1[o_2], w_1[o_3], c_1$$

1. Fill in for each history the list of pairs of operations where one transaction has read from another transaction ($w_i[o] \rightarrow r_j[o]$), and where one transaction has overwritten another transaction ($w_i[o] \rightarrow w_j[o]$).

H_w	H_x	H_y	H_z

2. Determine if H_w is RC, ACA, or ST.
3. Determine if H_x is RC, ACA, or ST.
4. Determine if H_y is RC, ACA, or ST.
5. Determine if H_z is RC, ACA, or ST.

IDB Worksheet 18: Deadlocks and Waits-For Graphs

You are told that three transactions H_1, H_2, H_3 perform the following sequence of operations:

$H_1 : w_1[o_1], r_1[o_2], r_1[o_4]$

$H_2 : r_2[o_3], r_2[o_2], r_2[o_1]$

$H_3 : r_3[o_4], w_3[o_4], r_3[o_3], w_3[o_3]$

1. Write down all the possible conflict pairs from H_1, H_2, H_3 in the form $rw_i \rightarrow rw_j$.
2. Write a concurrent execution H_d of H_1, H_2, H_3 which results in a deadlock involving all three transactions. [Hint, use the conflict pairs to determine which transaction might have to wait for another, and then work out possible cycles of such waits-for].
3. Draw the WFG for your answer in (2).
4. How would you resolve the deadlock?