How SQL Database Engines Work

D. Richard Hipp
OpenSQL
2008-11-15

How ~~SQL~~ SQLite Database Engines Work

with occasional comments on how the operation of SQLite compares to other SQL database engines.

D. Richard Hipp
OpenSQL
2008-11-15

Why Do You Care?

So you can better understand what your query is doing.

So that you can appreciate how much work the database is saving you.

So that you can fix things when they go wrong.

So that you can write better SQL that runs faster and uses less memory and disk space.


Fire Hydrant
Water Fountain, 1991

Key Concept

SQL is a peculiar programming language

- Each SQL statement is a separate program
- SQL describes *what* instead of *how*

An RDBMS consists of...

- Compiler to translate SQL into procedures
- Virtual Machine to evaluate the procedures

Example

```
SELECT * FROM table1;
```

Translates into:

```
Open database file containing table1
Rewind the file
while not at end-of-file
  read all columns out of current record
  return the columns to the caller
  advance file to the next record
end-while
close the file
```

Imagine what this translates into:

```
SELECT eqptid, enclosureid
  FROM eqpt
 WHERE typeid IN (
   SELECT typeid FROM typespec
    WHERE attrid=(
      SELECT attrid FROM attribute
       WHERE name='detect_autoactuate'
    )
    AND value=1
   INTERSECT
   SELECT typeid FROM typespec
    WHERE attrid=(
      SELECT attrid FROM attribute
       WHERE name='algorithm'
    )
    AND value IN ('sensor','wetbulb')
 )
```

The Whole Point Of SQL...

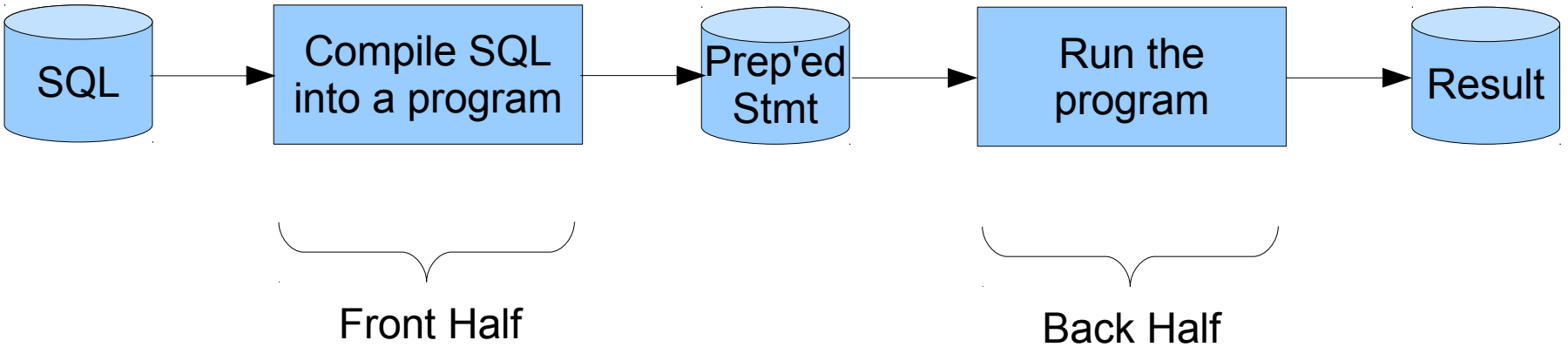
A few lines of SQL generates the equivalent of hundreds of lines of procedural code.

By adding an index, entirely new procedures are used without recoding.

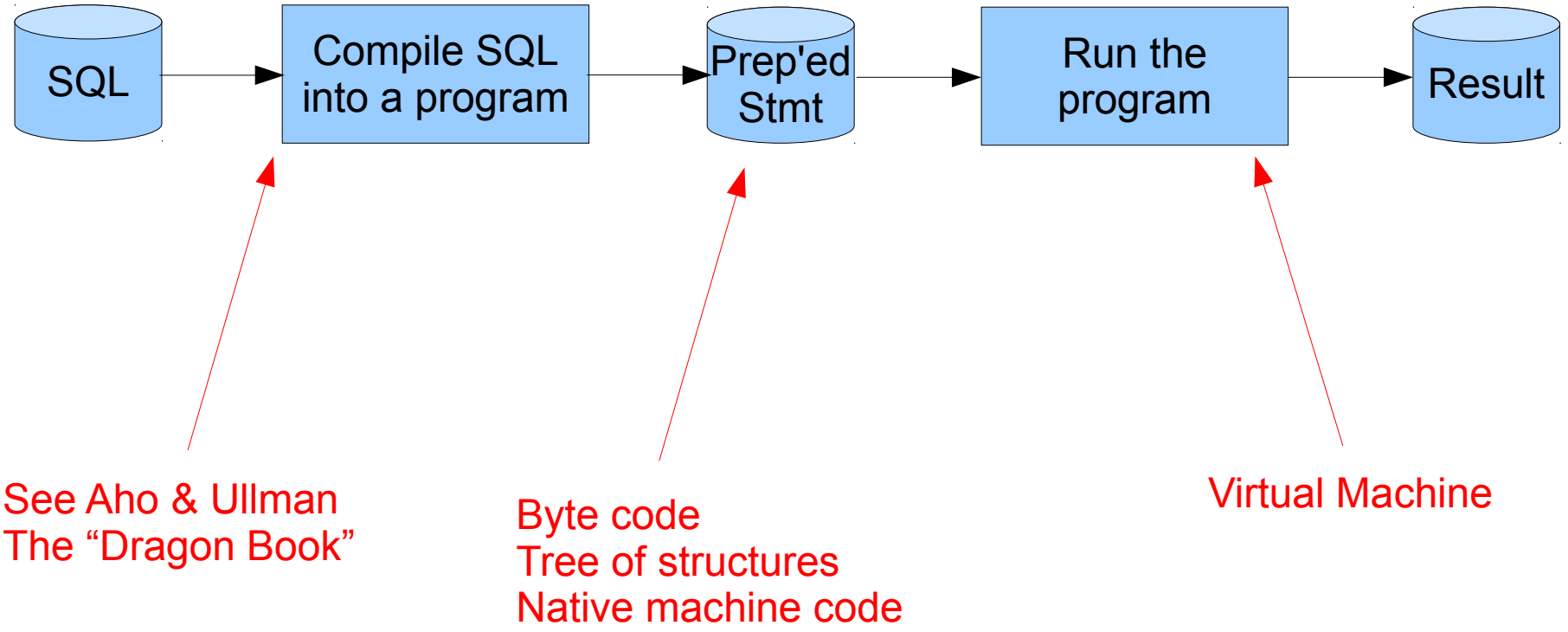
The SQL Query Optimizer is charged with picking the algorithm

so that the application developer doesn't have to

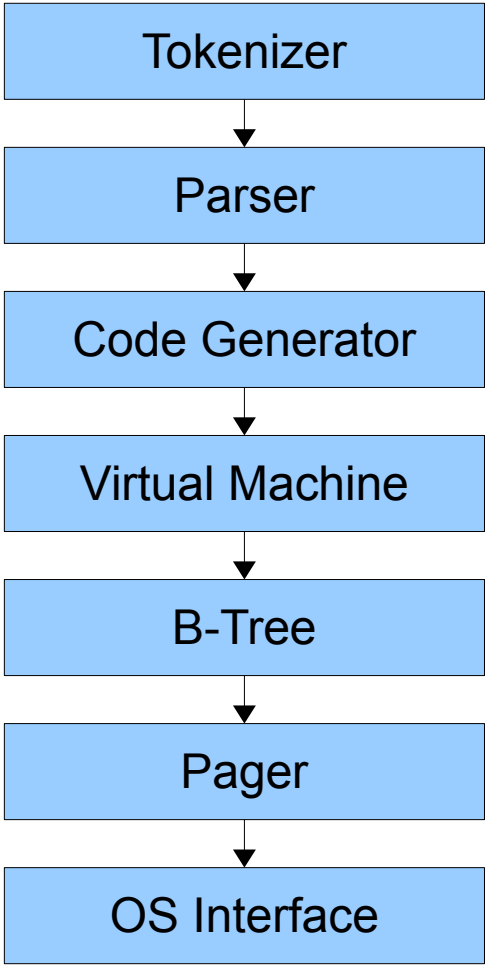
Ins and Outs of SQL



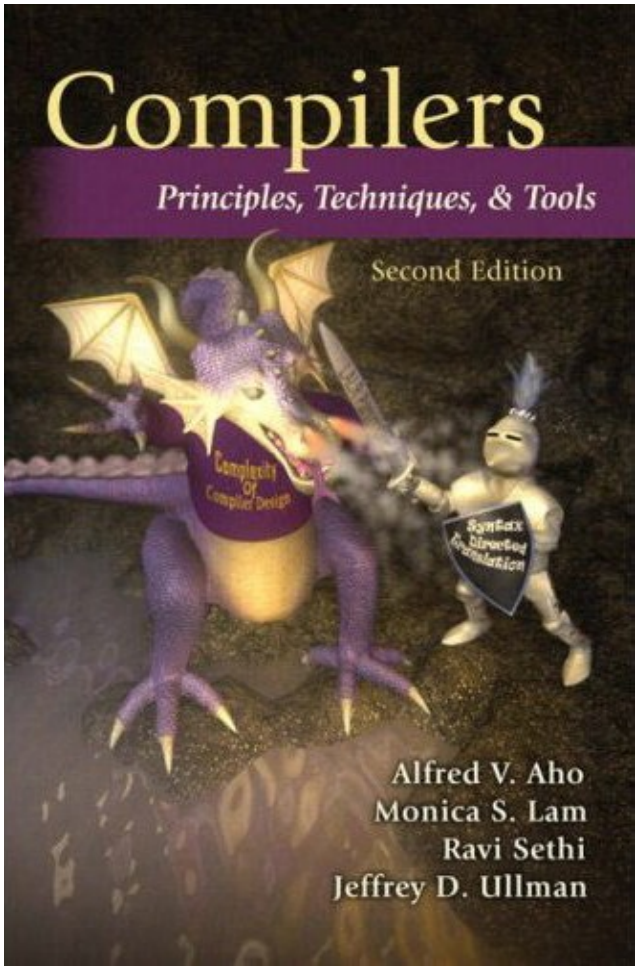
Ins and Outs of SQL

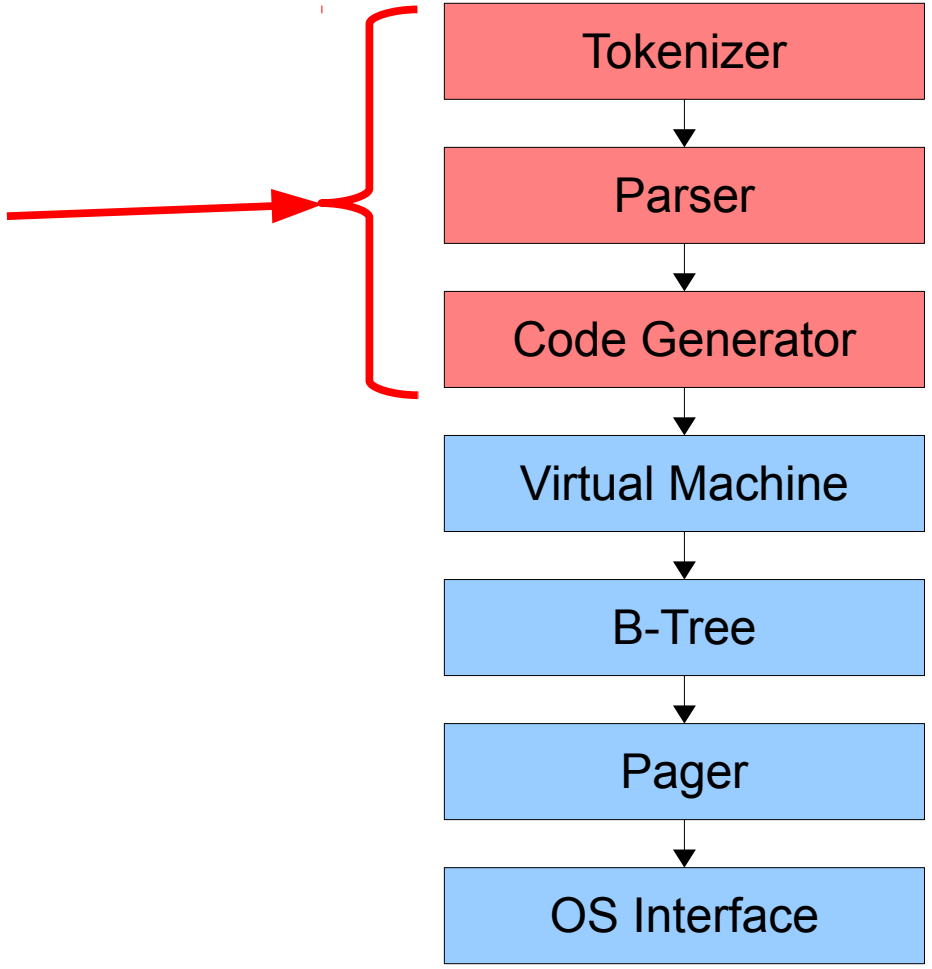


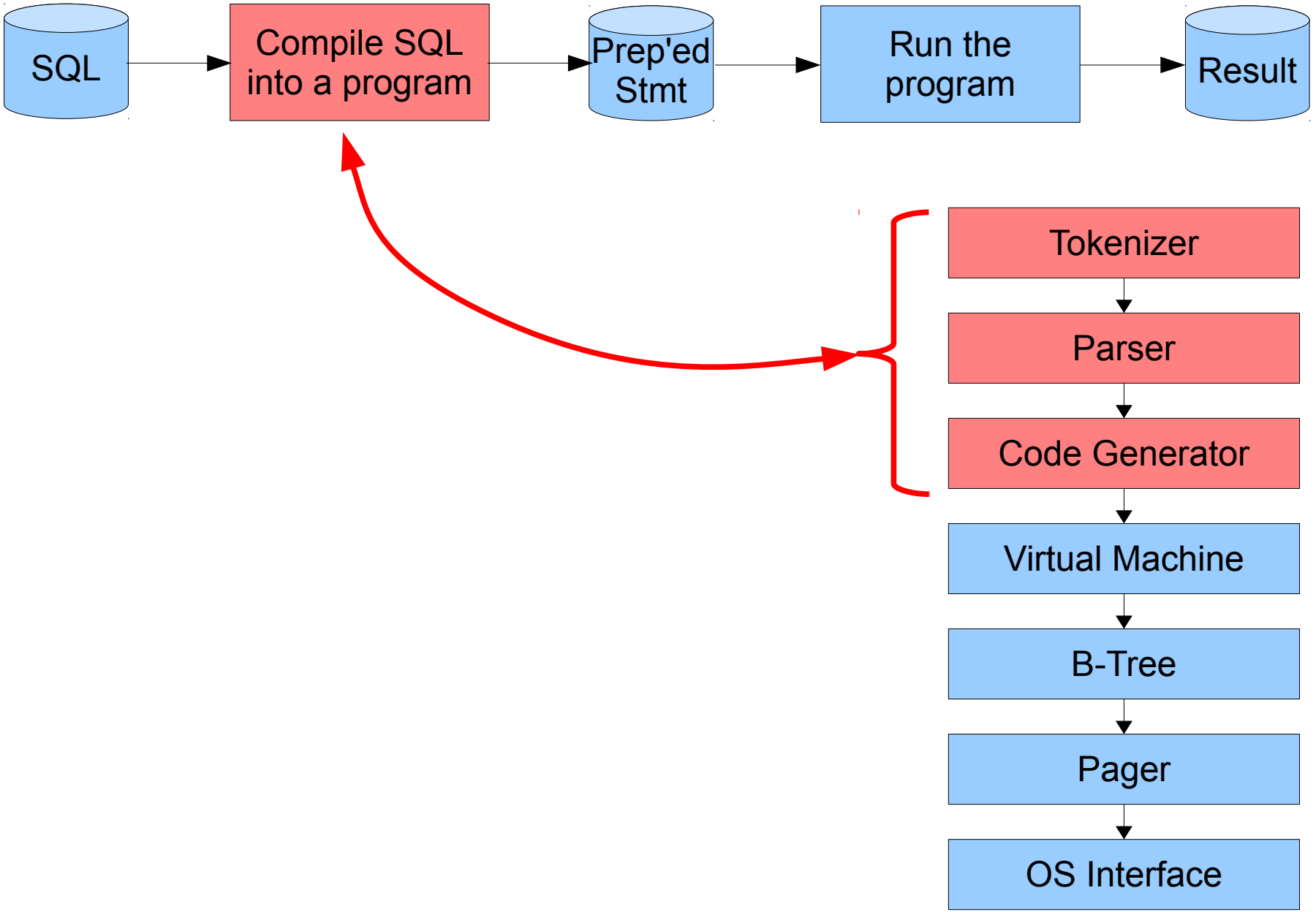
Architecture Of SQLite

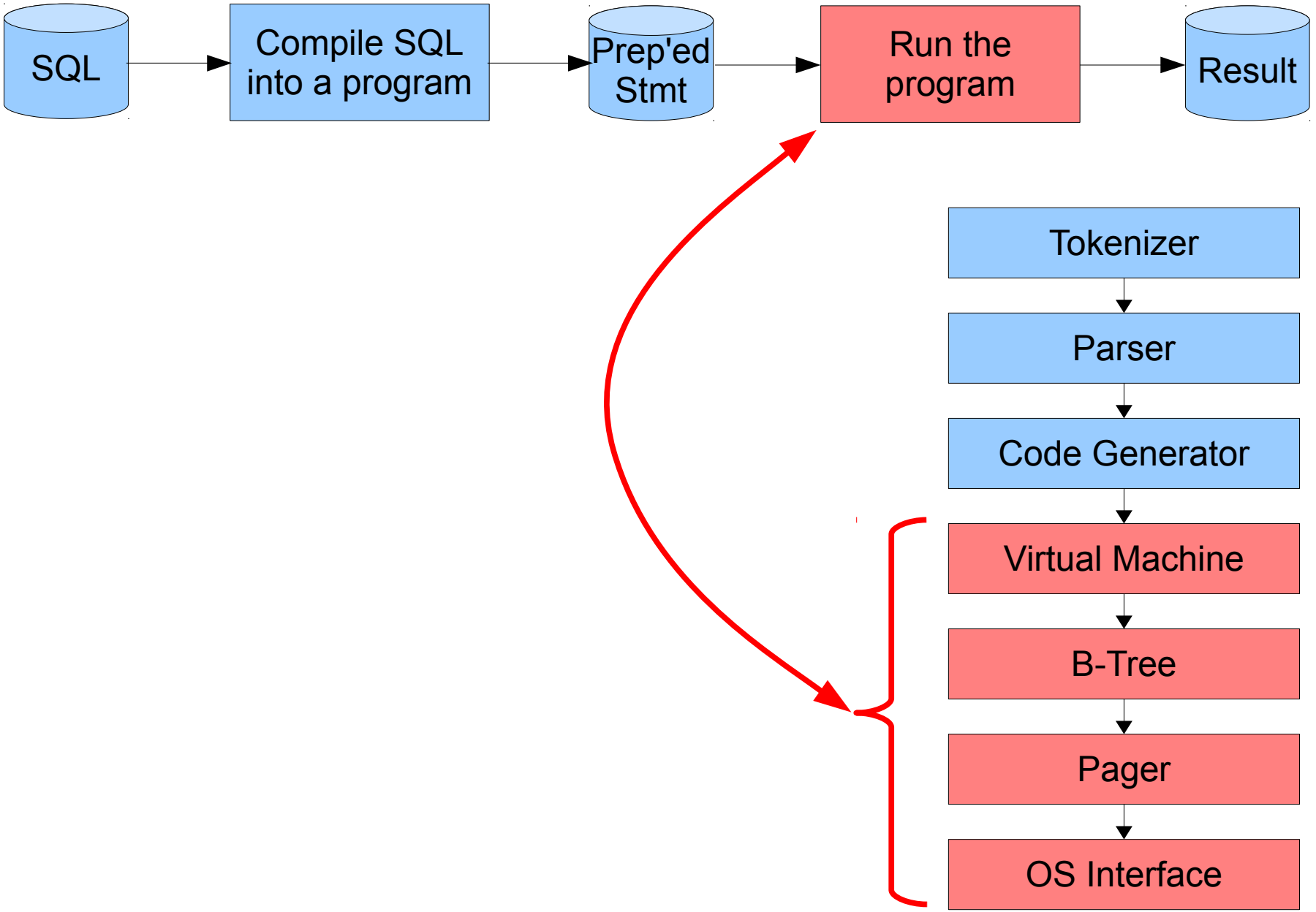


Architecture Of SQLite



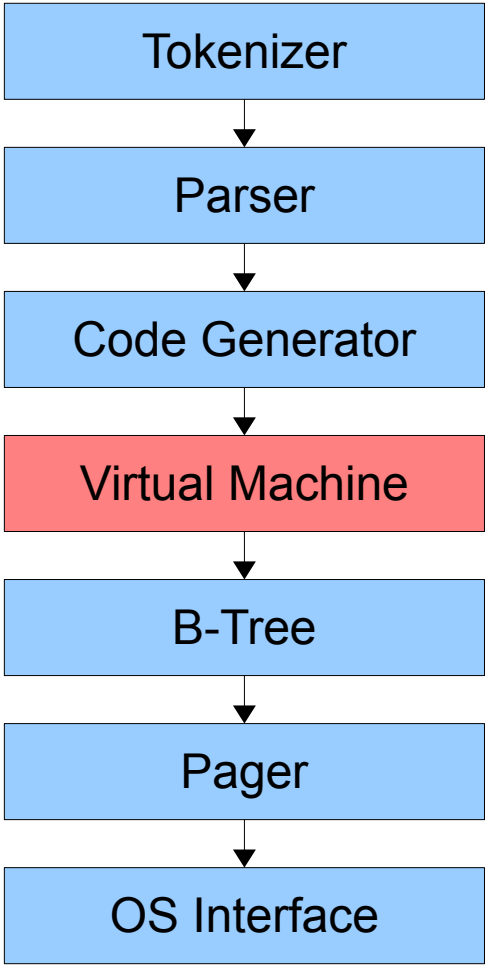



SQL
Compile SQL into a program
Prep'ed Stmt
Run the program
Result
Tokenizer
Parser
Code Generator
Virtual Machine
B-Tree
Pager
OS Interface


SQL
Compile SQL into a program
Prep'ed Stmt
Run the program
Result
Tokenizer
Parser
Code Generator
Virtual Machine
B-Tree
Pager
OS Interface

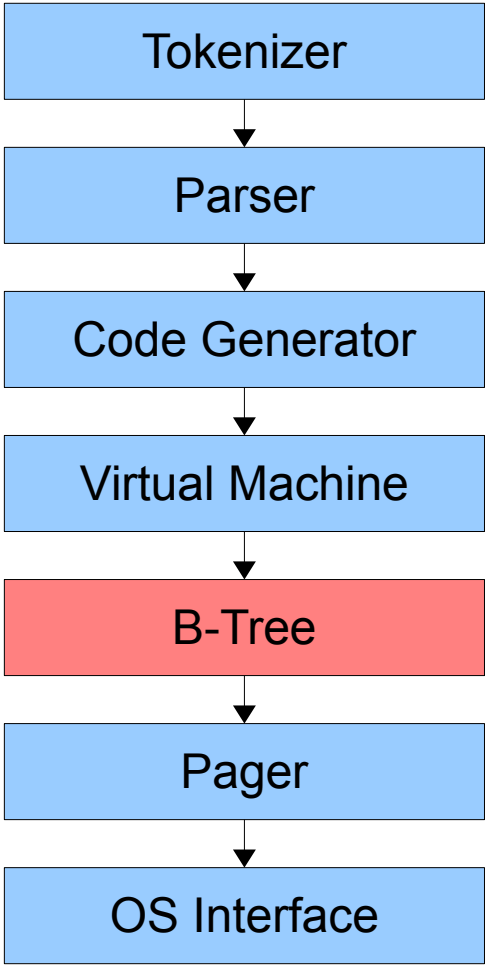
Architecture Of SQLite

- Byte code interpreter
- Big `switch` statement inside a `for` loop.
- Other engines walk a tree of structures
- Similar to JVM or Parrot or P-Code

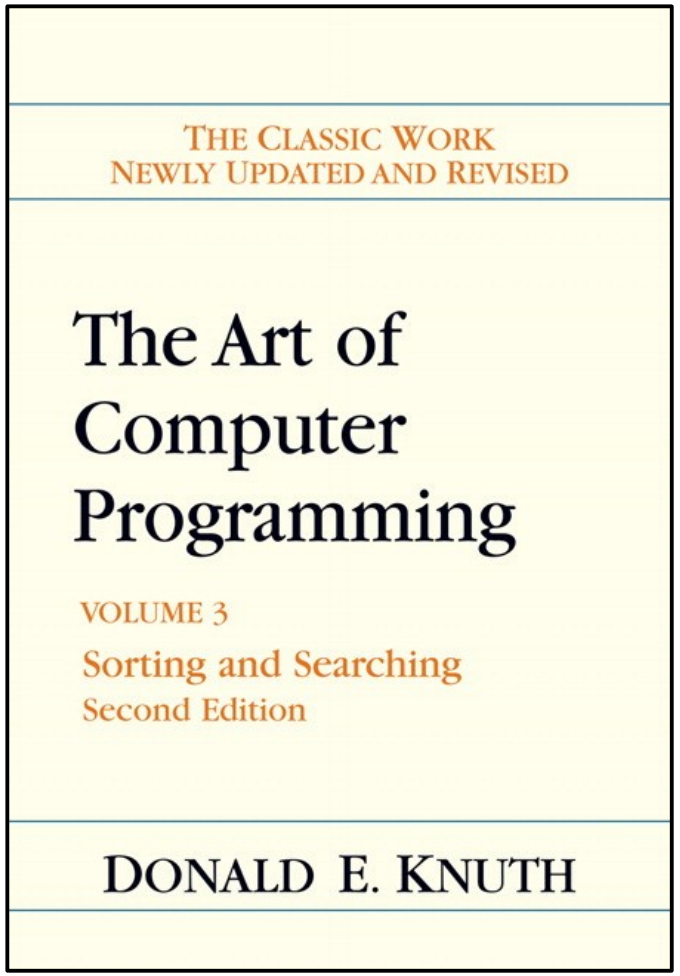


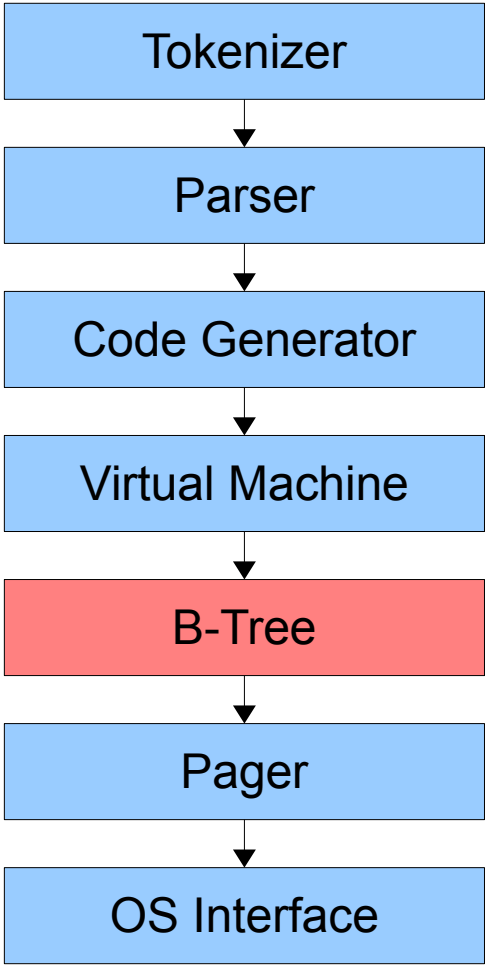
Architecture Of SQLite

- Ordered key/value pairs with unique keys
- O(logN) insert, seek, and delete
- O(1) next and previous



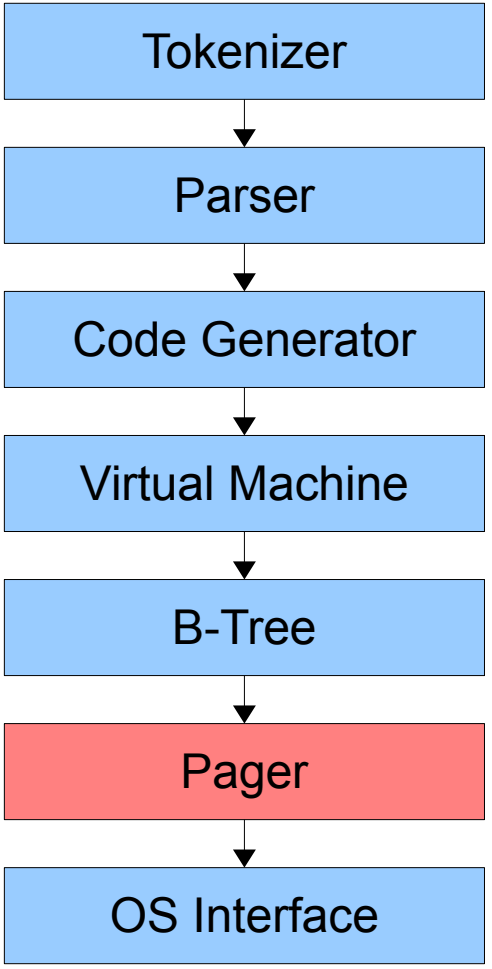
Architecture Of SQLite





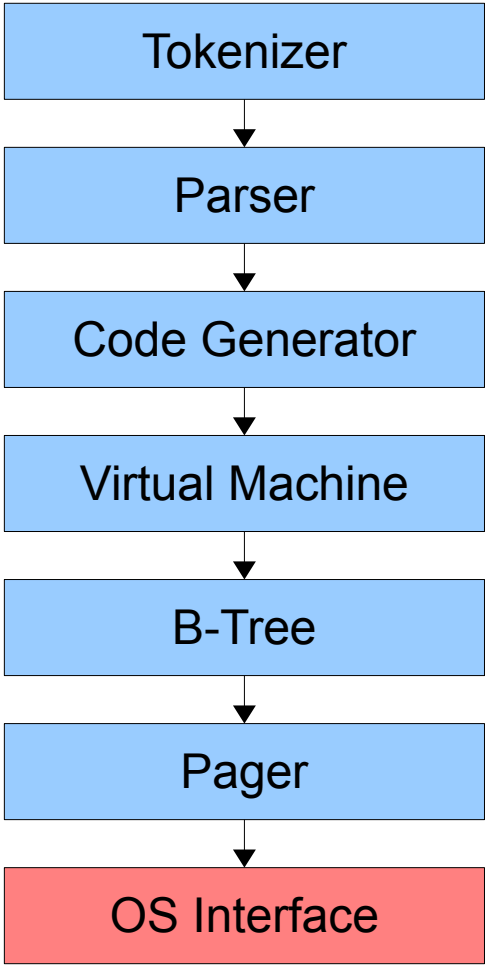
Architecture Of SQLite

- Atomic commit and rollback
- Uniform size pages numbered from 1
- No interpretation of page content
- Cache



Architecture Of SQLite

- Platform-specific interface to the OS
- Run-time changeable
- Portability layer



```
CREATE TABLE tab(
   Fruit TEXT,
   GrownIn TEXT,
   UnitPrice REAL
);
```

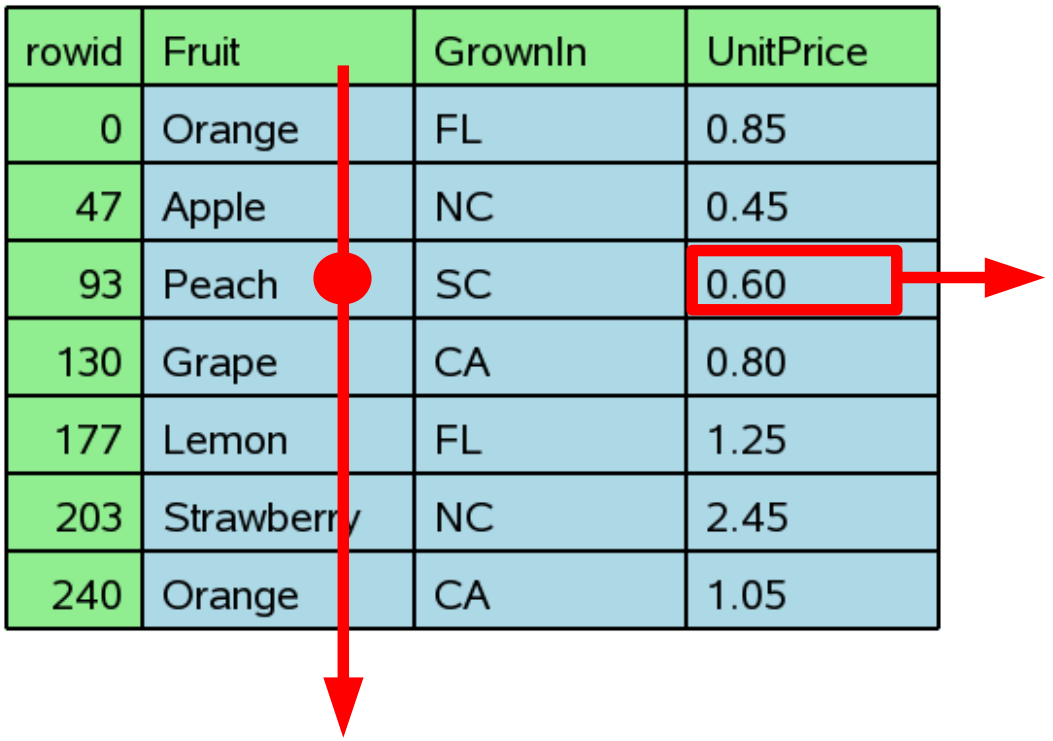
rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

```
CREATE TABLE tab(
   Fruit TEXT,
   GrownIn TEXT,
   UnitPrice REAL
);
```

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

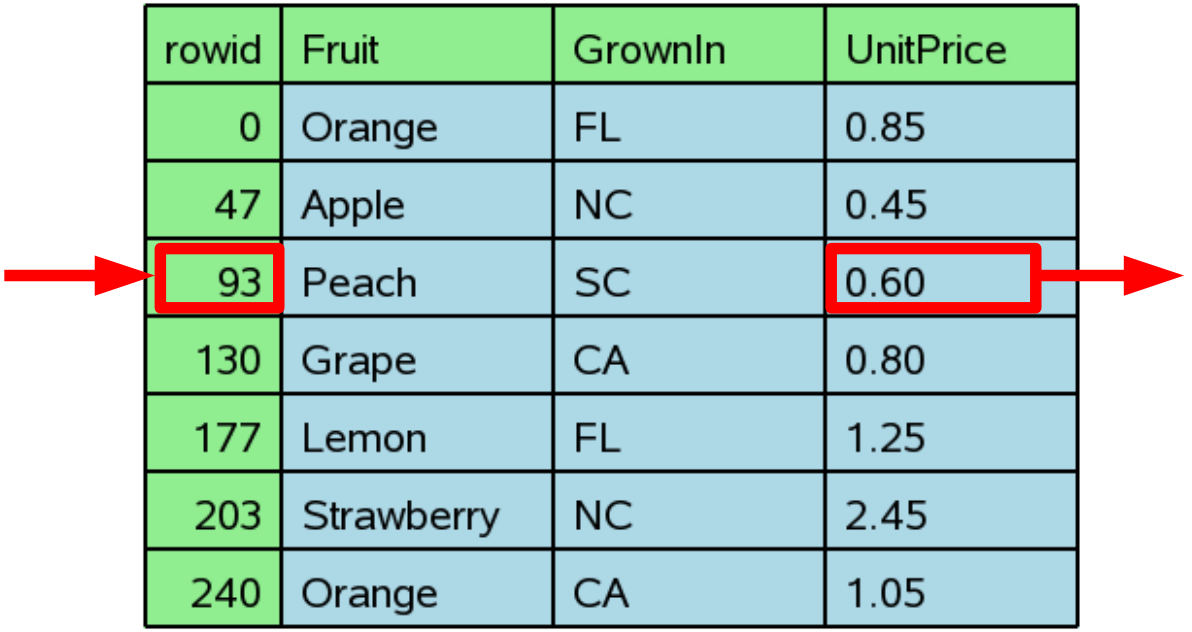
Key **Data**

SELECT unitprice FROM tab WHERE fruit='Peach'



rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

SELECT unitprice FROM tab WHERE rowid=93



rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

CREATE INDEX idx1 ON tab(fruit)

idx1

Fruit	rowid
Apple	47
Grape	130
Lemon	177
Orange	0
Orange	240
Peach	93
Strawberry	203

tab

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

CREATE INDEX idx1 ON tab(fruit)

Fruit	rowid
Apple	47
Grape	130
Lemon	177
Orange	0
Orange	240
Peach	93
Strawberry	203

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

CREATE INDEX idx1 ON tab(fruit)

Fruit	rowid
Apple	47
Grape	130
Lemon	177
Orange	0
Orange	240
Peach	93
Strawberry	203

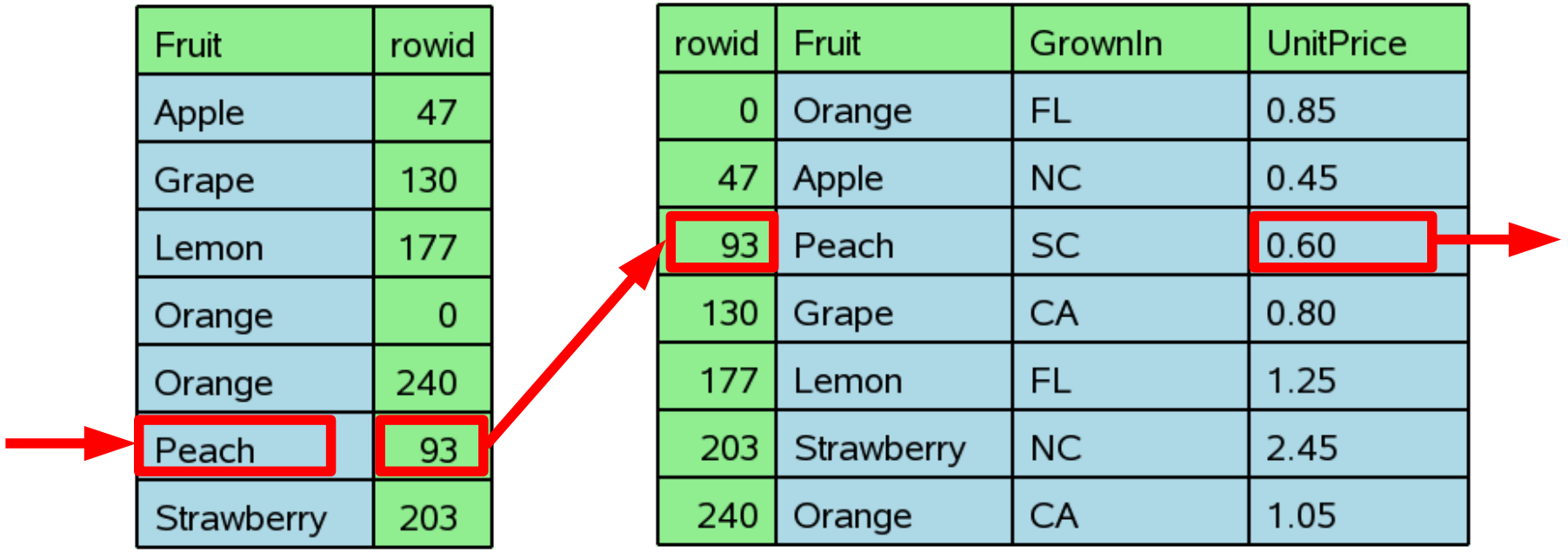
rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

CREATE INDEX idx1 ON tab(fruit)

Fruit	rowid
Apple	47
Grape	130
Lemon	177
Orange	0
Orange	240
Peach	93
Strawberry	203

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

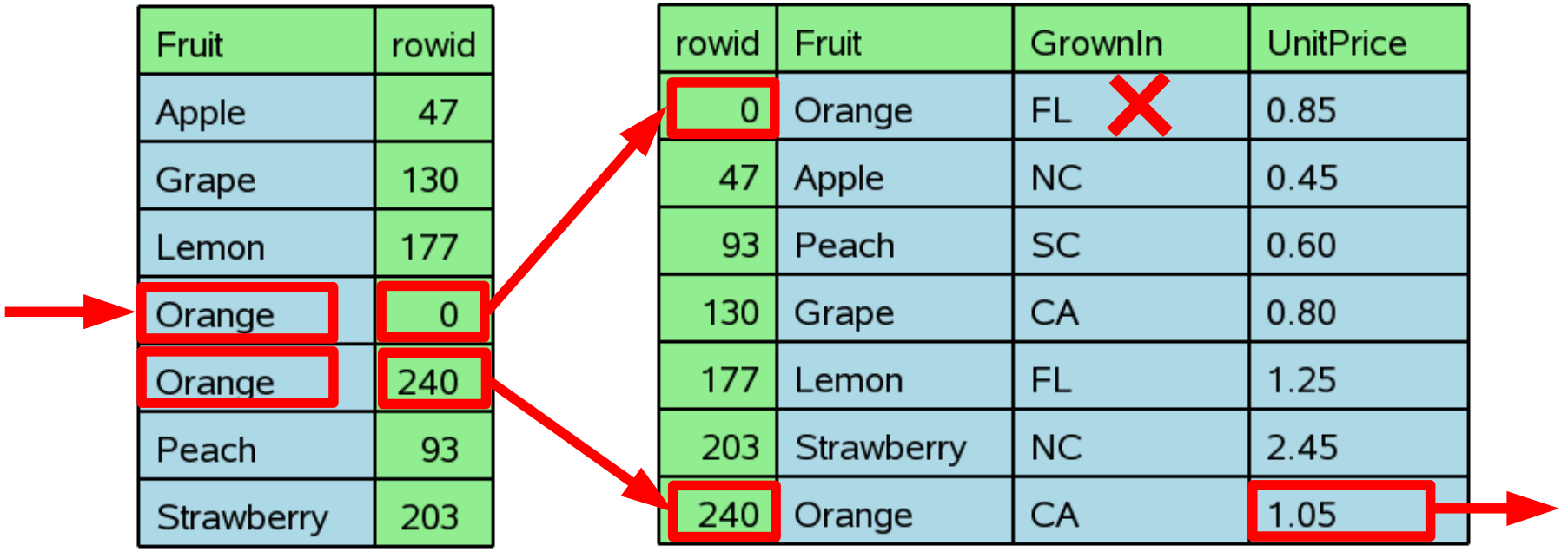
SELECT unitprice FROM tab WHERE fruit='Peach'



Fruit	rowid
Apple	47
Grape	130
Lemon	177
Orange	0
Orange	240
Peach	93
Strawberry	203

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
     AND  grownin='CA'
```



CREATE INDEX idx2 ON tab(grownin)

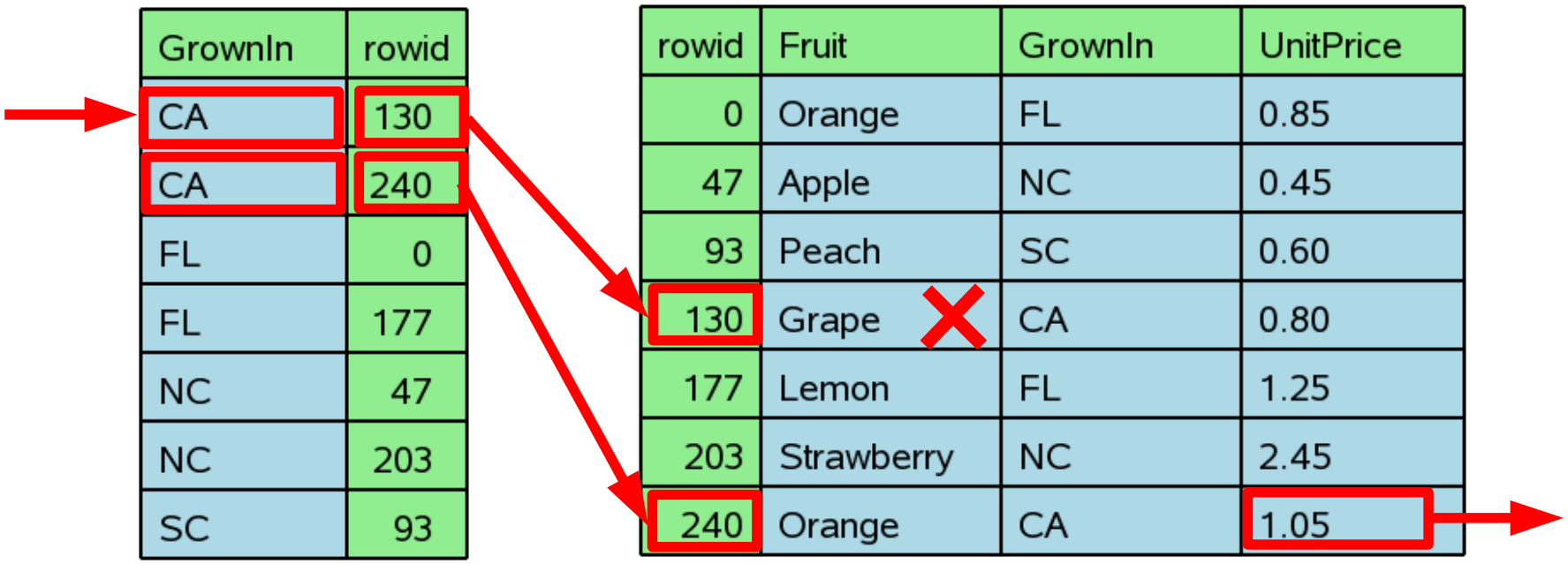
idx2

GrownIn	rowid
CA	130
CA	240
FL	0
FL	177
NC	47
NC	203
SC	93

tab

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
     AND  grownin='CA'
```



GrownIn	rowid
CA	130
CA	240
FL	0
FL	177
NC	47
NC	203
SC	93

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

CREATE INDEX idx3 ON tab(fruit, grownin)

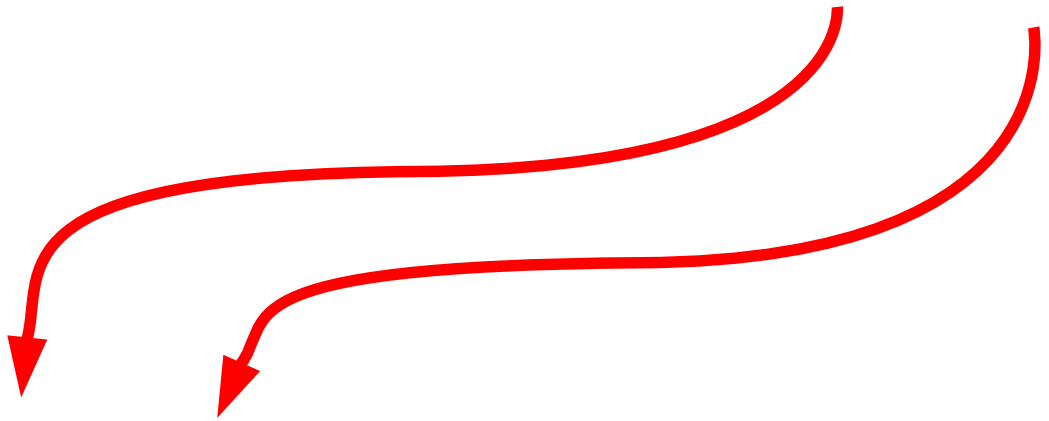
idx3

Fruit	GrownIn	rowid
Apple	NC	47
Grape	CA	130
Lemon	FL	177
Orange	CA	240
Orange	FL	0
Peach	SC	93
Strawberry	NC	203

tab

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

CREATE INDEX idx2 ON tab(fruit, grownin)



Fruit	GrownIn	rowid
Apple	NC	47
Grape	CA	130
Lemon	FL	177
Orange	CA	240
Orange	FL	0
Peach	SC	93
Strawberry	NC	203

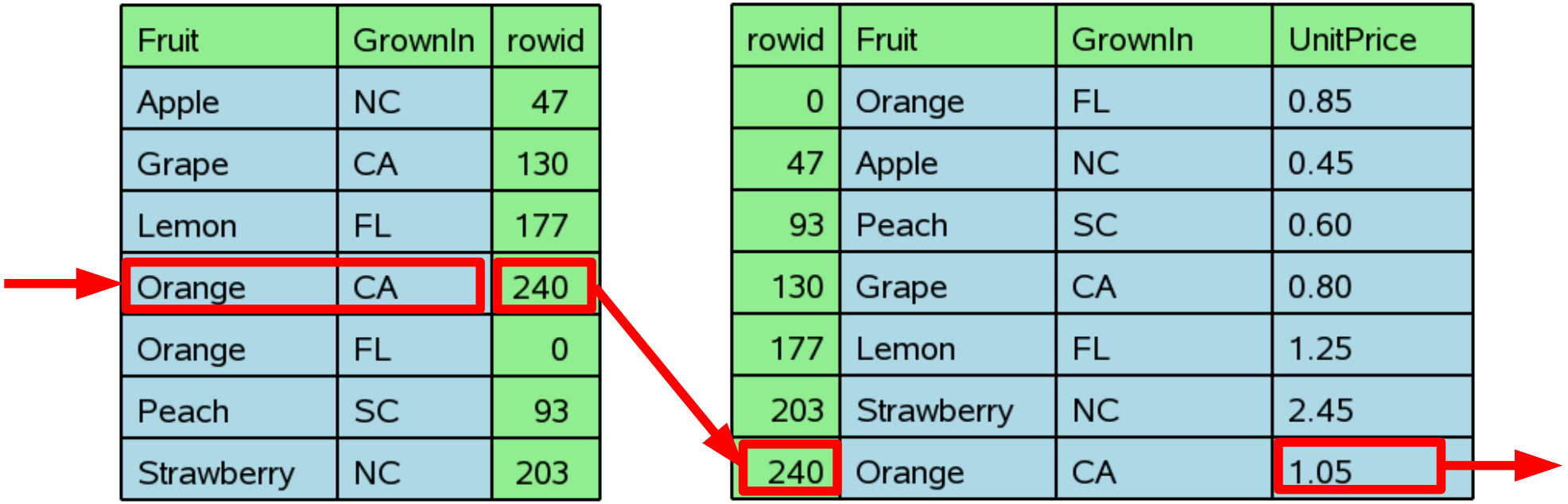
rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

CREATE INDEX idx2 ON tab(fruit, grownin)

Fruit	GrownIn	rowid
Apple	NC	47
Grape	CA	130
Lemon	FL	177
Orange	CA	240
Orange	FL	0
Peach	SC	93
Strawberry	NC	203

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
     AND  grownin='CA'
```



Fruit	GrownIn	rowid
Apple	NC	47
Grape	CA	130
Lemon	FL	177
Orange	CA	240
Orange	FL	0
Peach	SC	93
Strawberry	NC	203

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

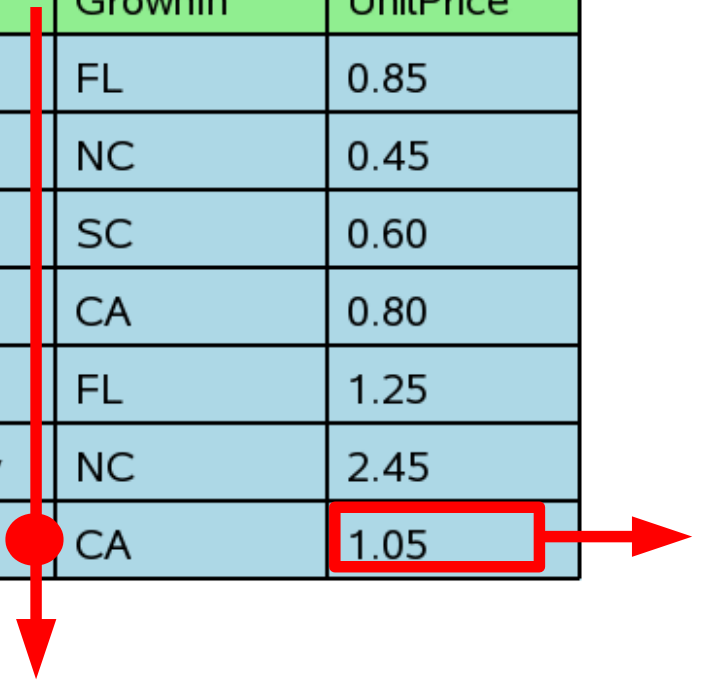
```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
     AND  grownin='CA'
```

4 ways to do this query
(so far...)

```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
     AND  grownin='CA'
```

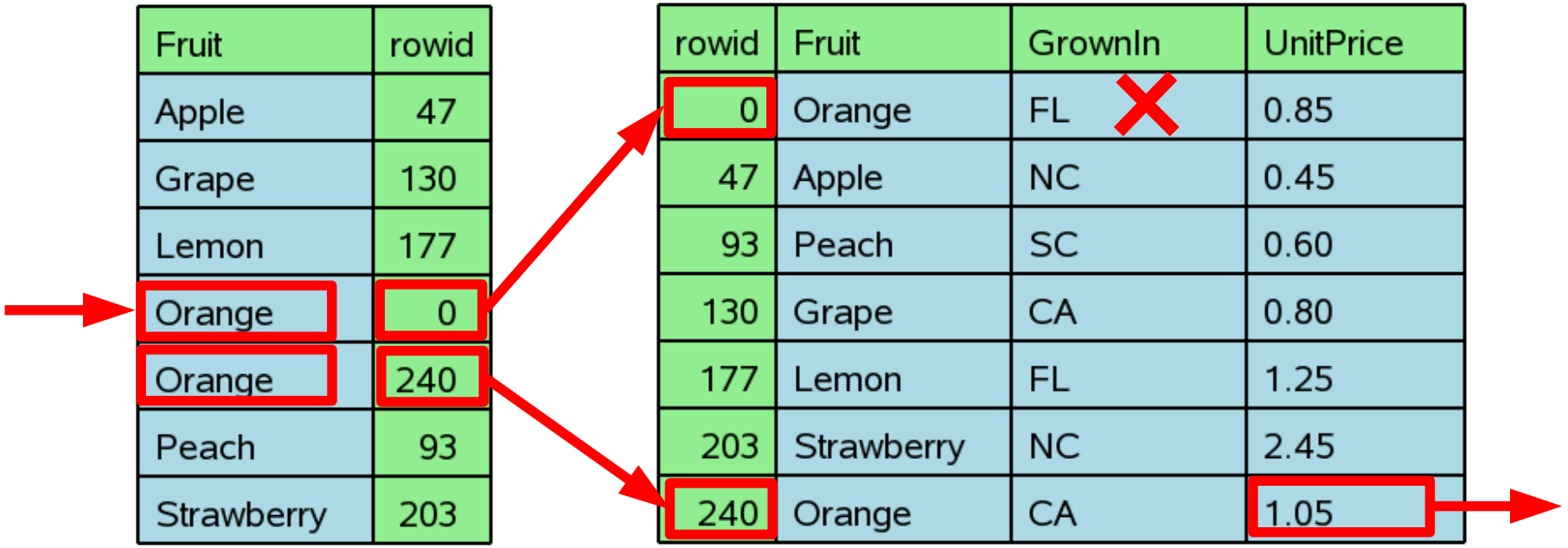
1

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05



```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
     AND  grownin='CA'
```

2

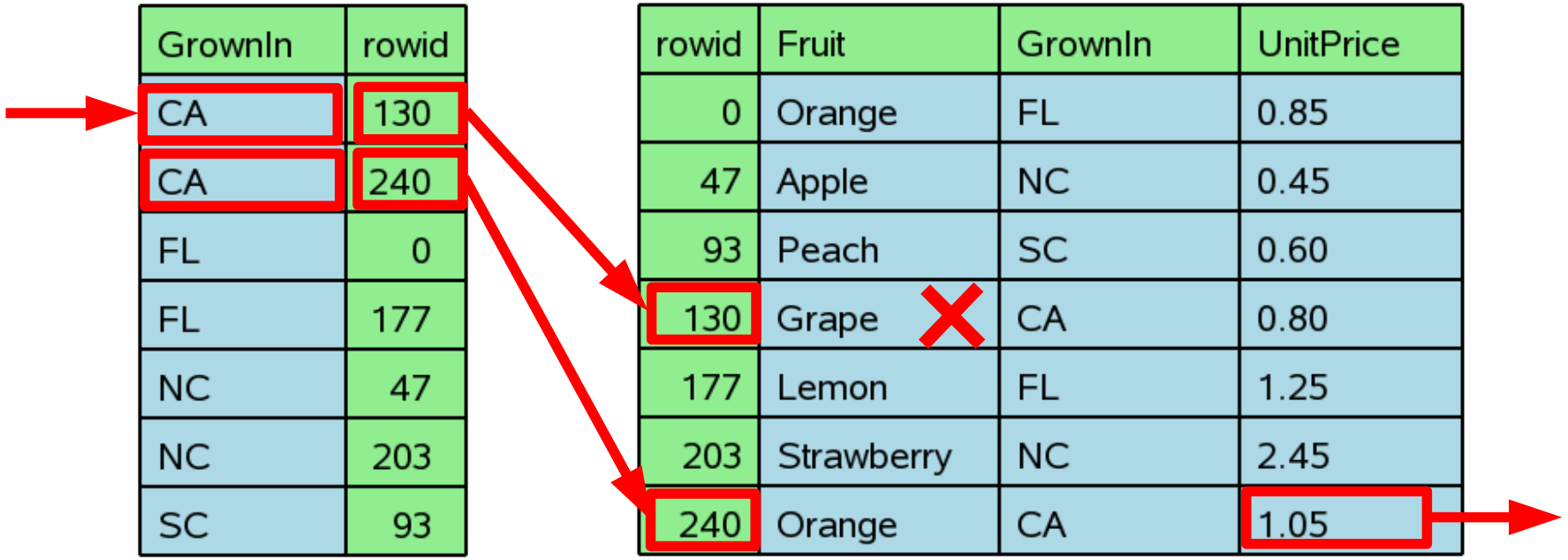


Fruit	rowid
Apple	47
Grape	130
Lemon	177
Orange	0
Orange	240
Peach	93
Strawberry	203

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
     AND  grownin='CA'
```

3

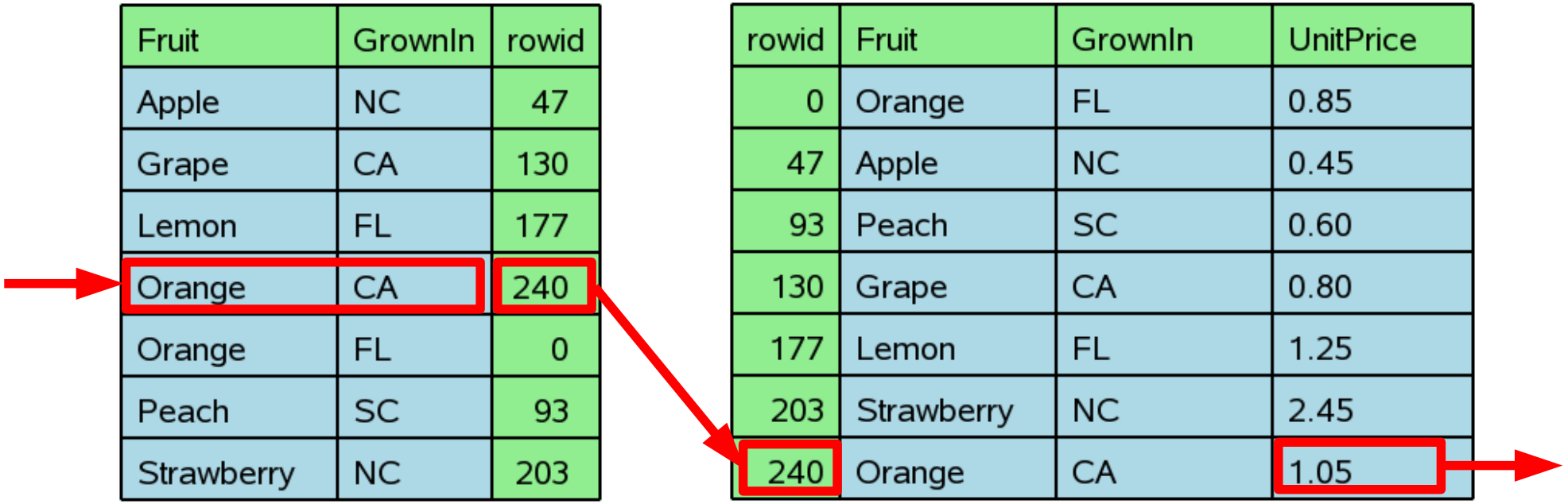


GrownIn	rowid
CA	130
CA	240
FL	0
FL	177
NC	47
NC	203
SC	93

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
     AND  grownin='CA'
```

4



Fruit	GrownIn	rowid
Apple	NC	47
Grape	CA	130
Lemon	FL	177
Orange	CA	240
Orange	FL	0
Peach	SC	93
Strawberry	NC	203

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

CREATE INDEX idx4 ON tab(fruit, grownin, unitprice)

idx4

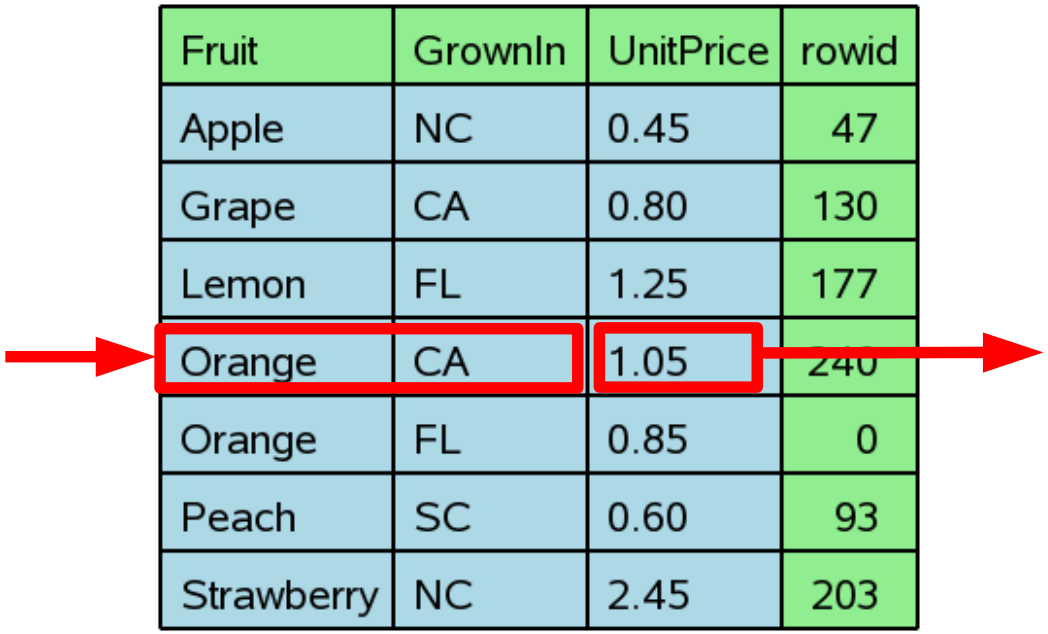
Fruit	GrownIn	UnitPrice	rowid
Apple	NC	0.45	47
Grape	CA	0.80	130
Lemon	FL	1.25	177
Orange	CA	1.05	240
Orange	FL	0.85	0
Peach	SC	0.60	93
Strawberry	NC	2.45	203

tab

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

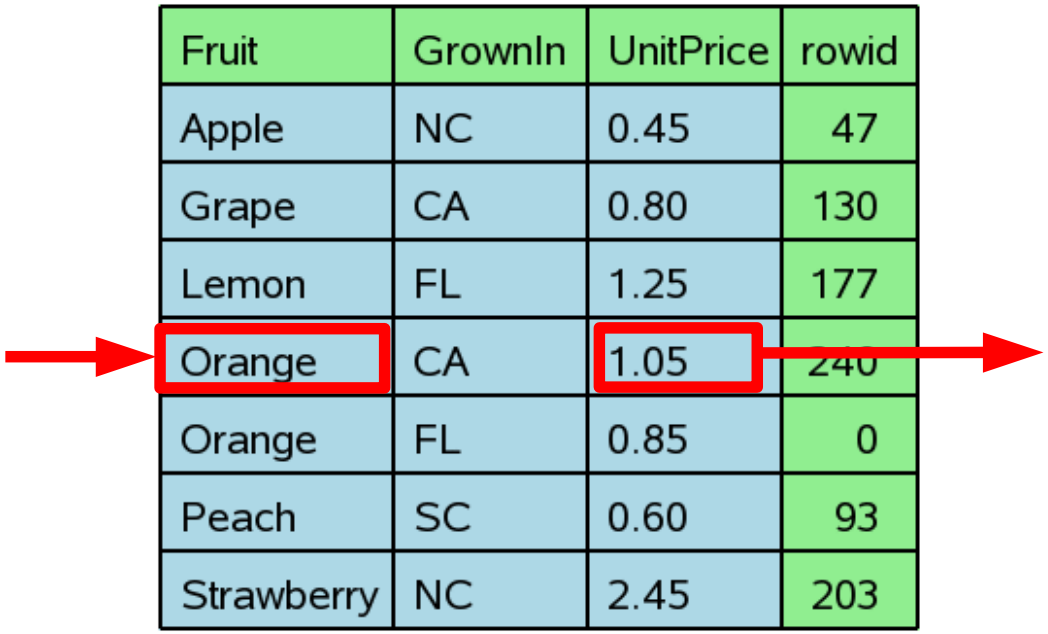
```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
     AND  grownin='CA'
```

5



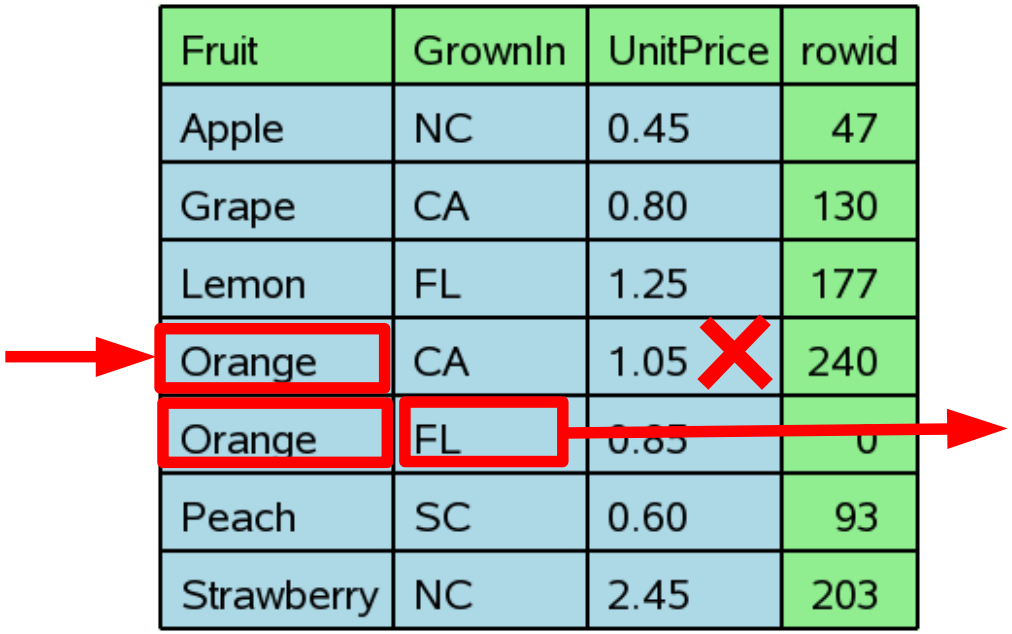
Fruit	GrownIn	UnitPrice	rowid
Apple	NC	0.45	47
Grape	CA	0.80	130
Lemon	FL	1.25	177
Orange	CA	1.05	240
Orange	FL	0.85	0
Peach	SC	0.60	93
Strawberry	NC	2.45	203

SELECT unitprice FROM tab
WHERE fruit='Orange'



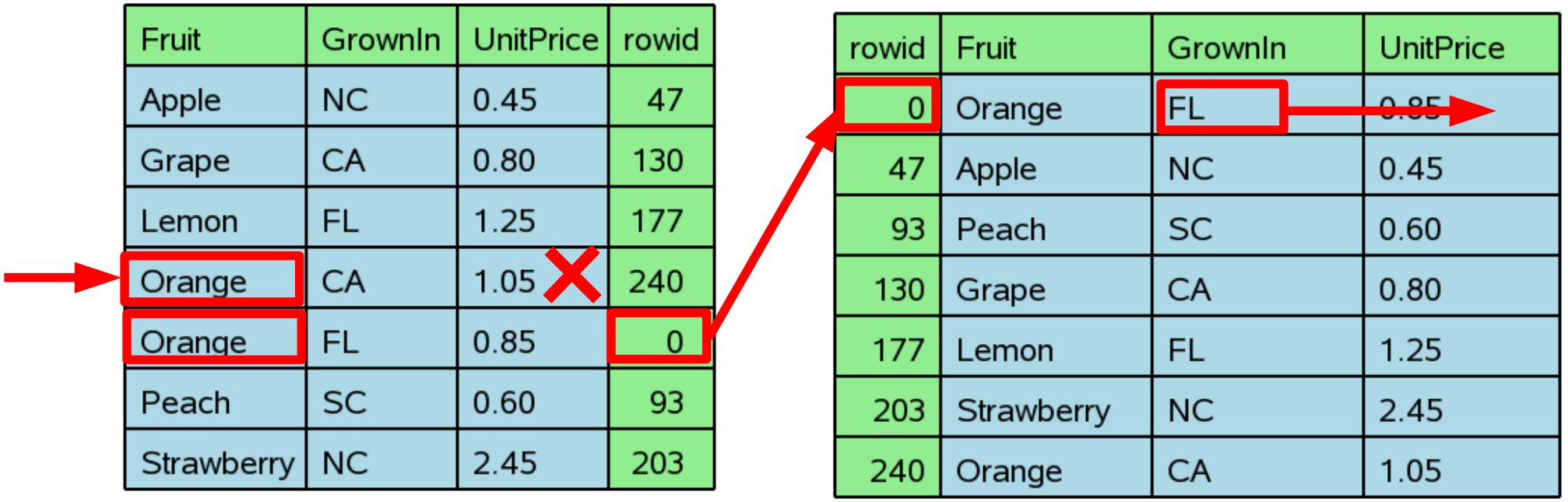
Fruit	GrownIn	UnitPrice	rowid
Apple	NC	0.45	47
Grape	CA	0.80	130
Lemon	FL	1.25	177
Orange	CA	1.05	240
Orange	FL	0.85	0
Peach	SC	0.60	93
Strawberry	NC	2.45	203

```
SELECT grownin FROM tab
 WHERE fruit='Orange'
       AND unitprice<1.00
```



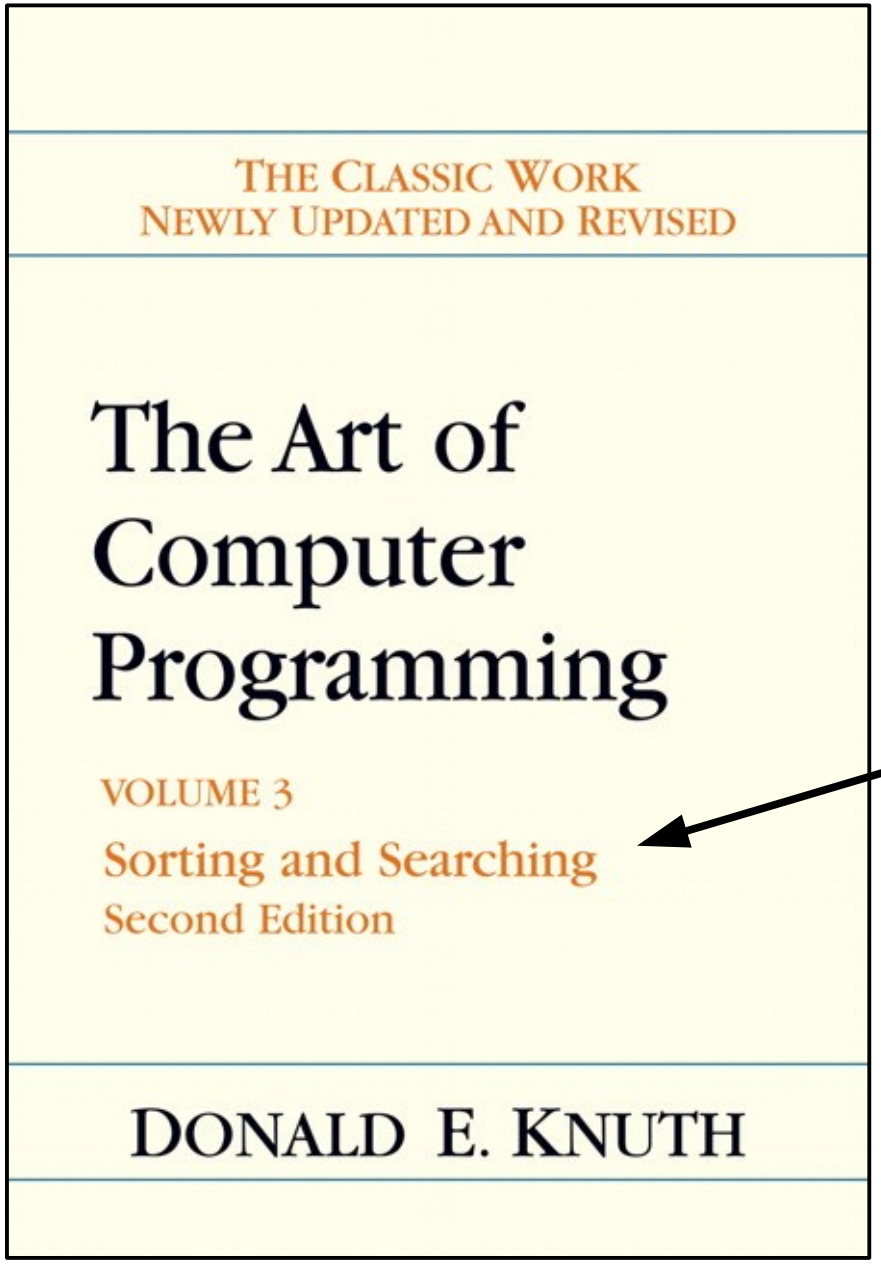
Fruit	GrownIn	UnitPrice	rowid
Apple	NC	0.45	47
Grape	CA	0.80	130
Lemon	FL	1.25	177
Orange	CA	1.05	240
Orange	FL	0.85	0
Peach	SC	0.60	93
Strawberry	NC	2.45	203

```
SELECT grownin FROM tab
 WHERE fruit='Orange'
       AND unitprice<1.00
```



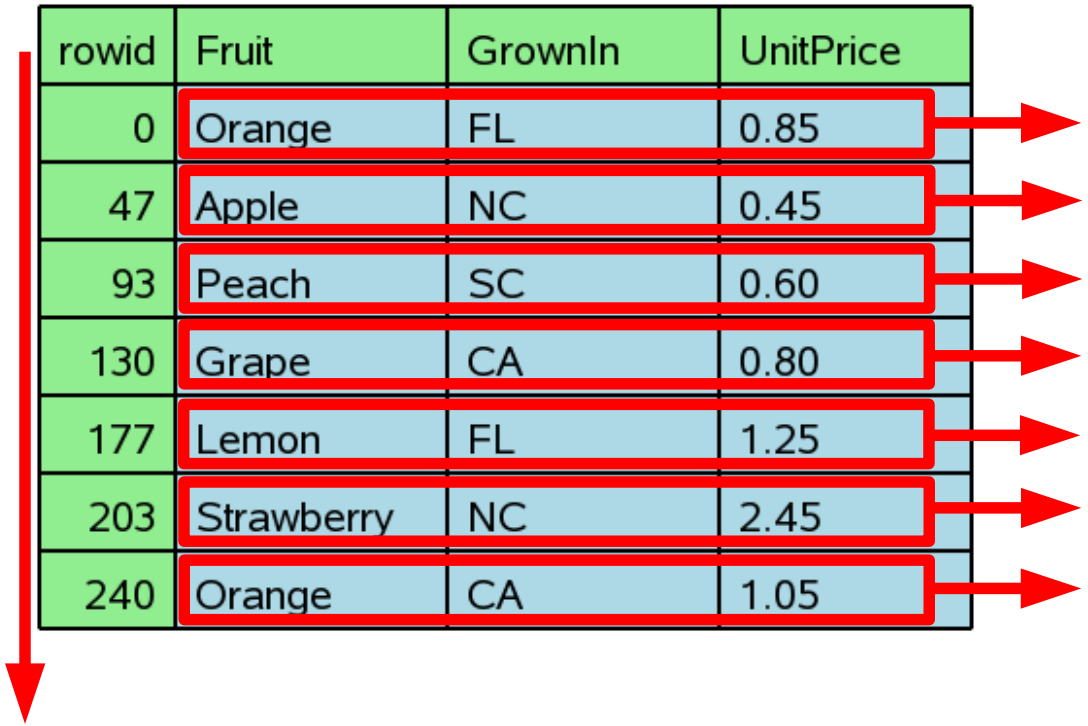
Fruit	GrownIn	UnitPrice	rowid
Apple	NC	0.45	47
Grape	CA	0.80	130
Lemon	FL	1.25	177
Orange	CA	1.05	240
Orange	FL	0.85	0
Peach	SC	0.60	93
Strawberry	NC	2.45	203

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05


THE CLASSIC WORK
NEWLY UPDATED AND REVISED
The Art of Computer Programming
VOLUME 3
Sorting and Searching
Second Edition
DONALD E. KNUTH

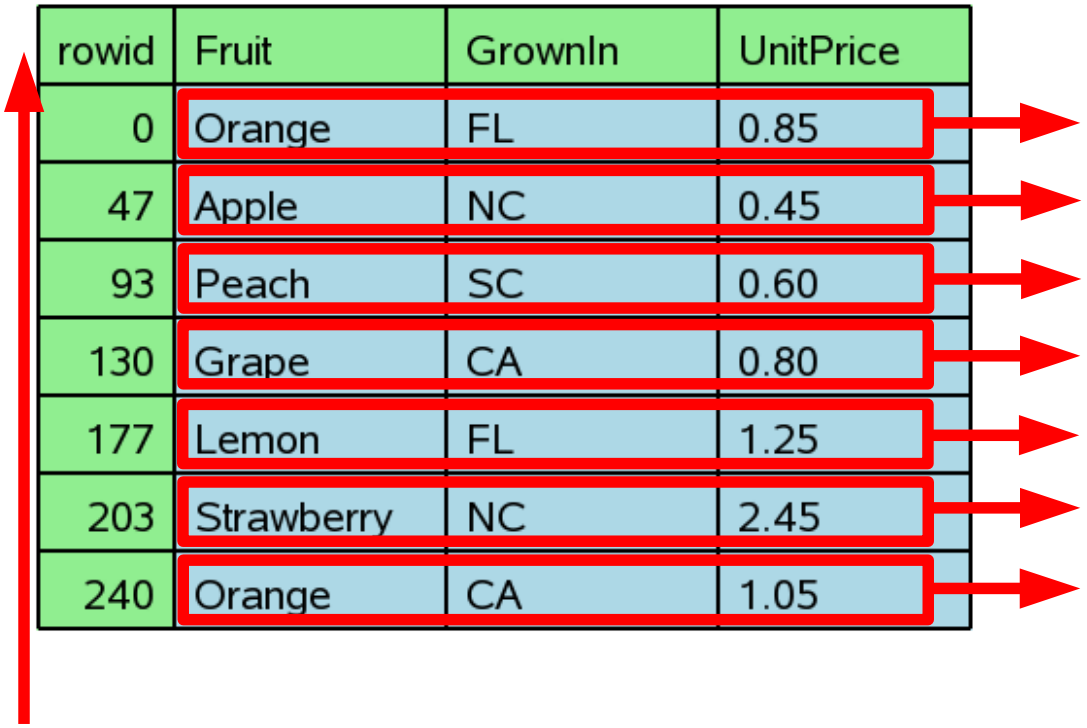
SELECT * FROM tab ORDER BY fruit

SELECT * FROM tab ORDER BY rowid



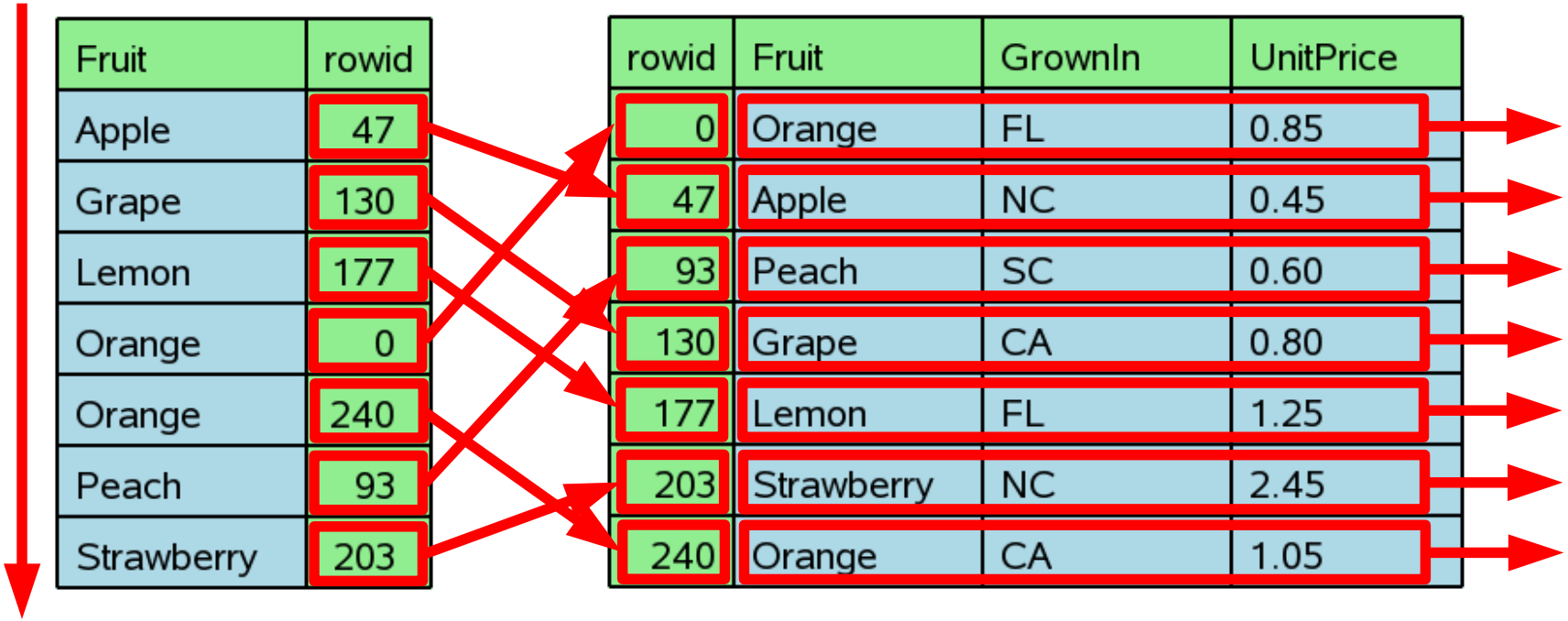
rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

SELECT * FROM tab ORDER BY rowid DESC

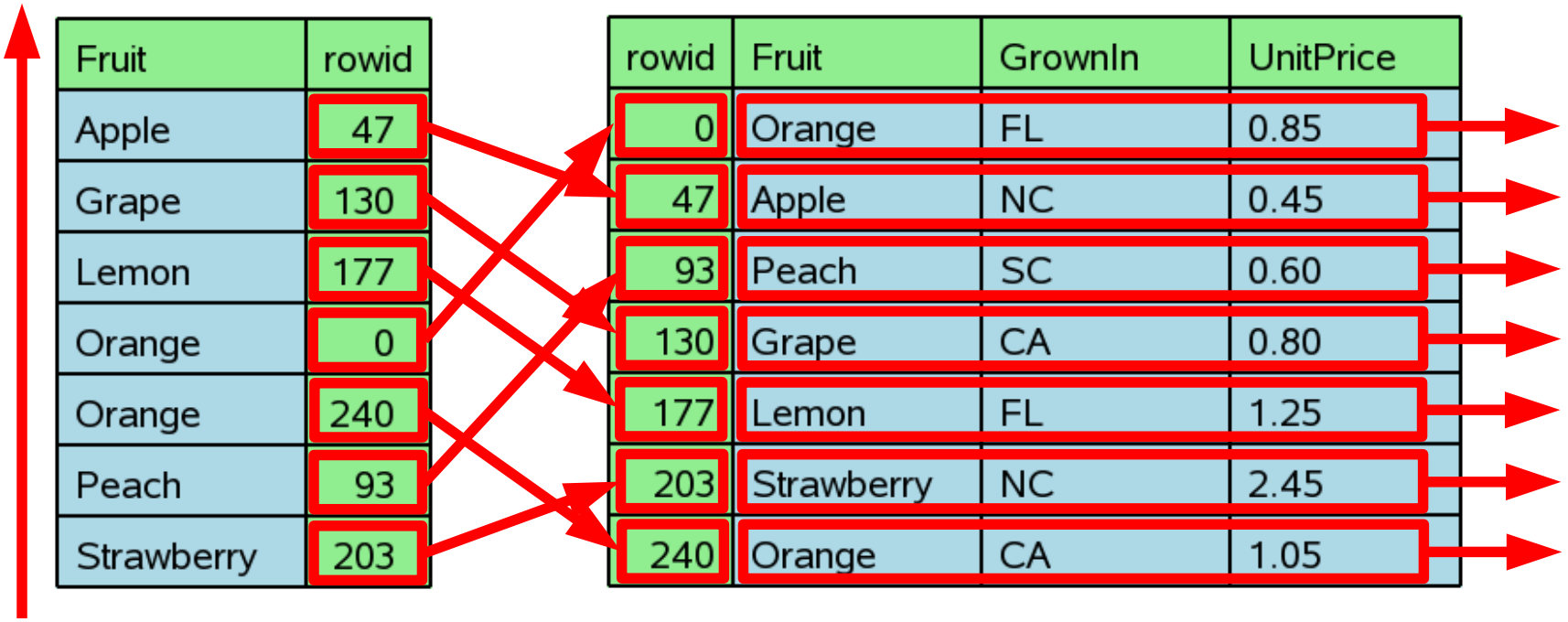


rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

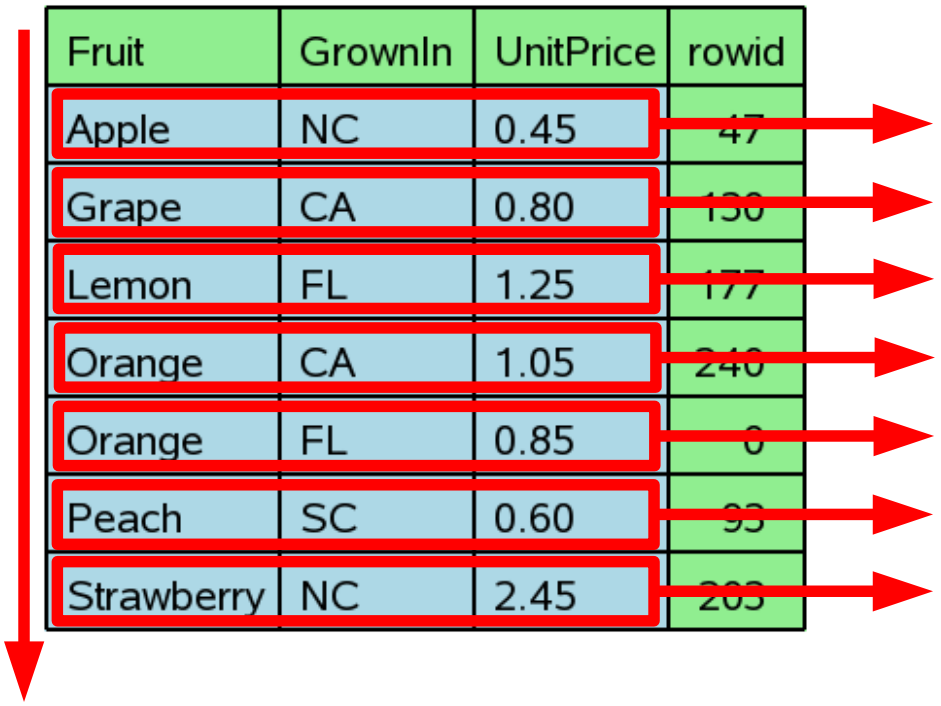
SELECT * FROM tab ORDER BY fruit



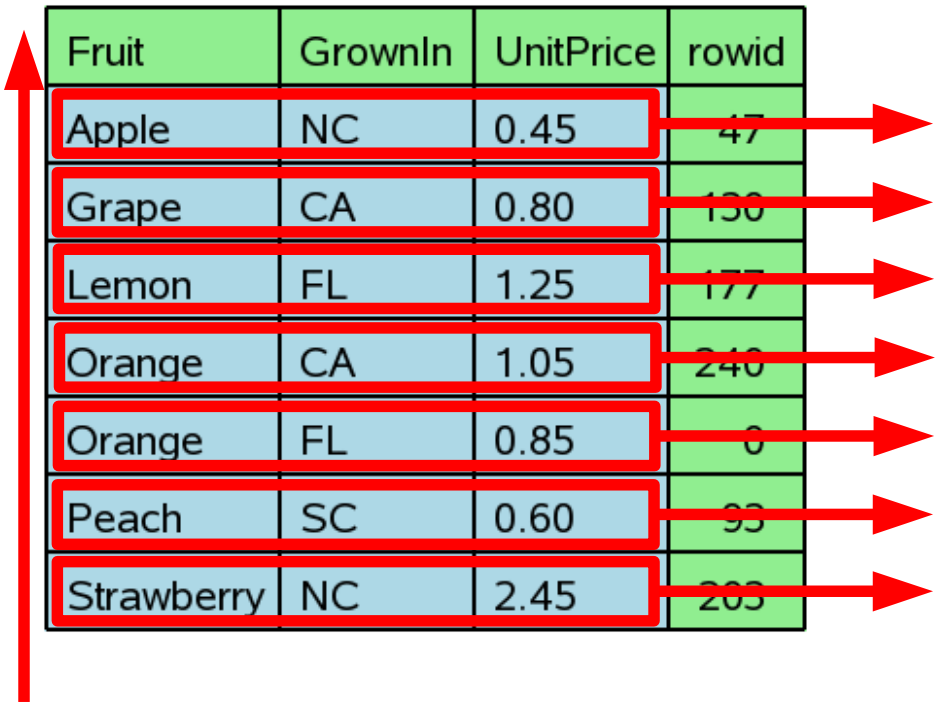
SELECT * FROM tab ORDER BY fruit DESC



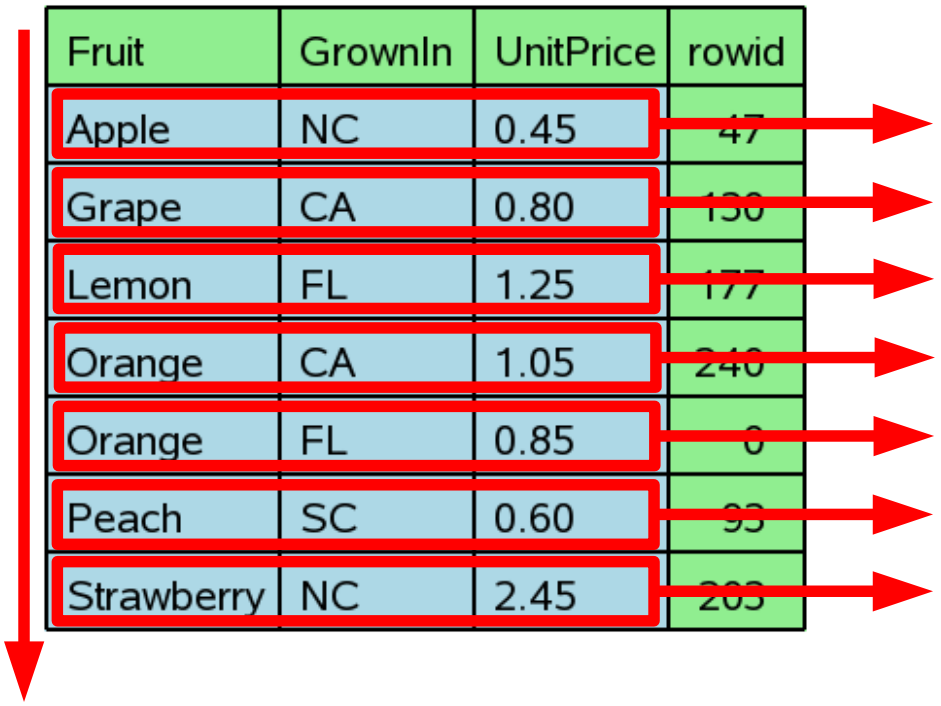
SELECT * FROM tab ORDER BY fruit



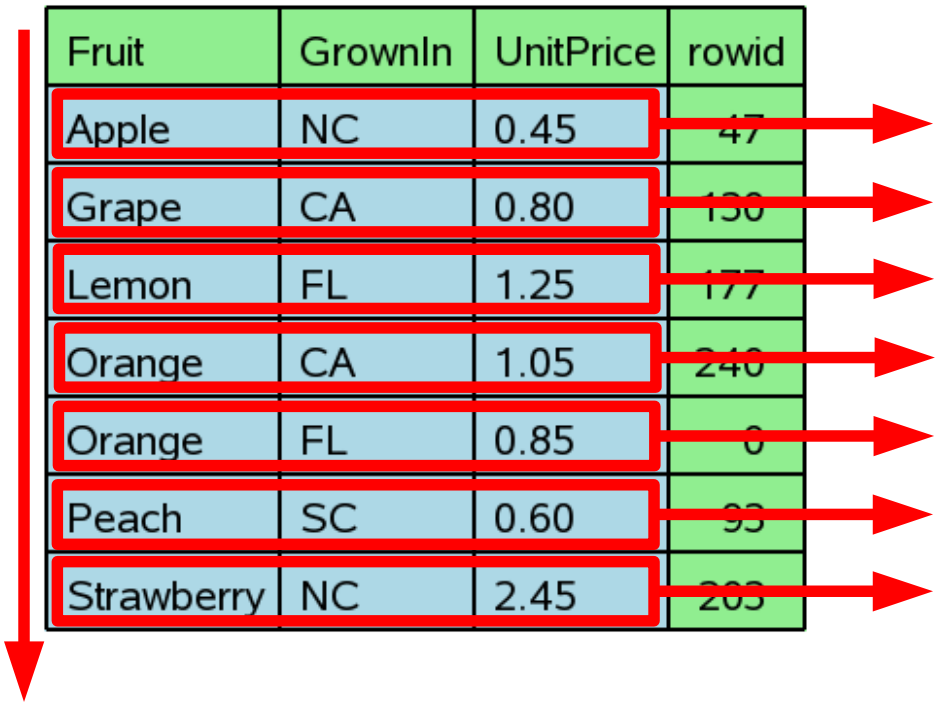
SELECT * FROM tab ORDER BY fruit DESC



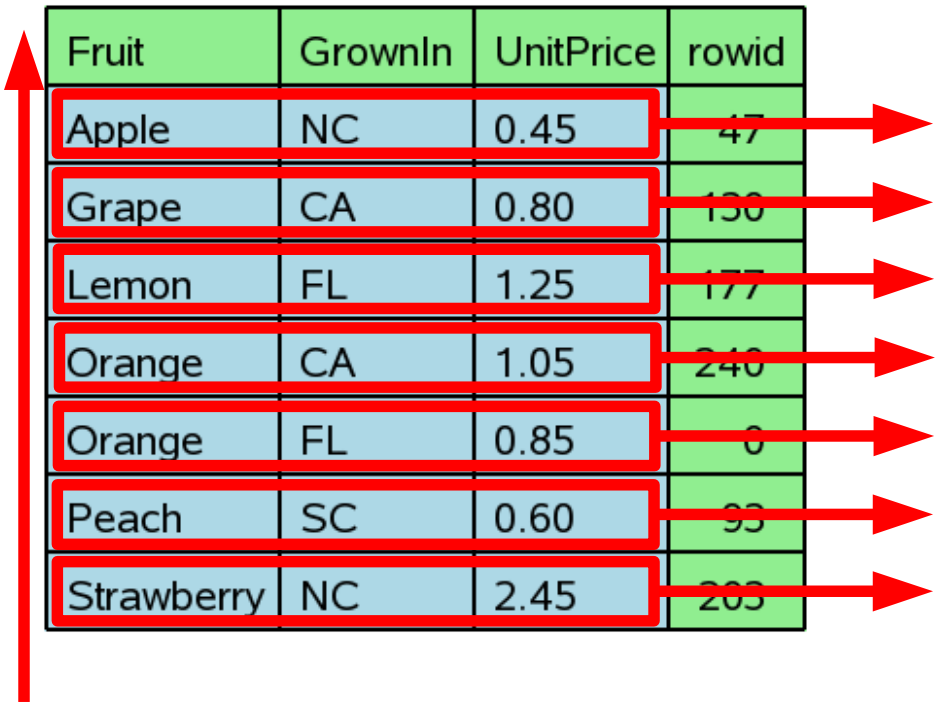
SELECT * FROM tab ORDER BY fruit, grownin



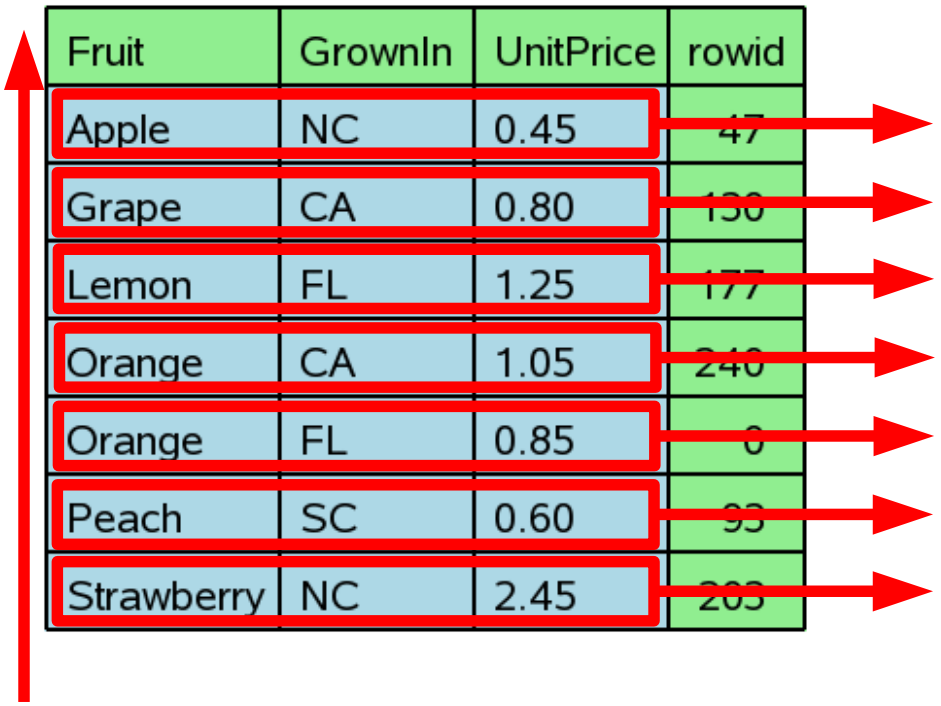
SELECT * FROM tab ORDER BY fruit, grownin, unitprice



SELECT * FROM tab ORDER BY fruit DESC, grownin DESC

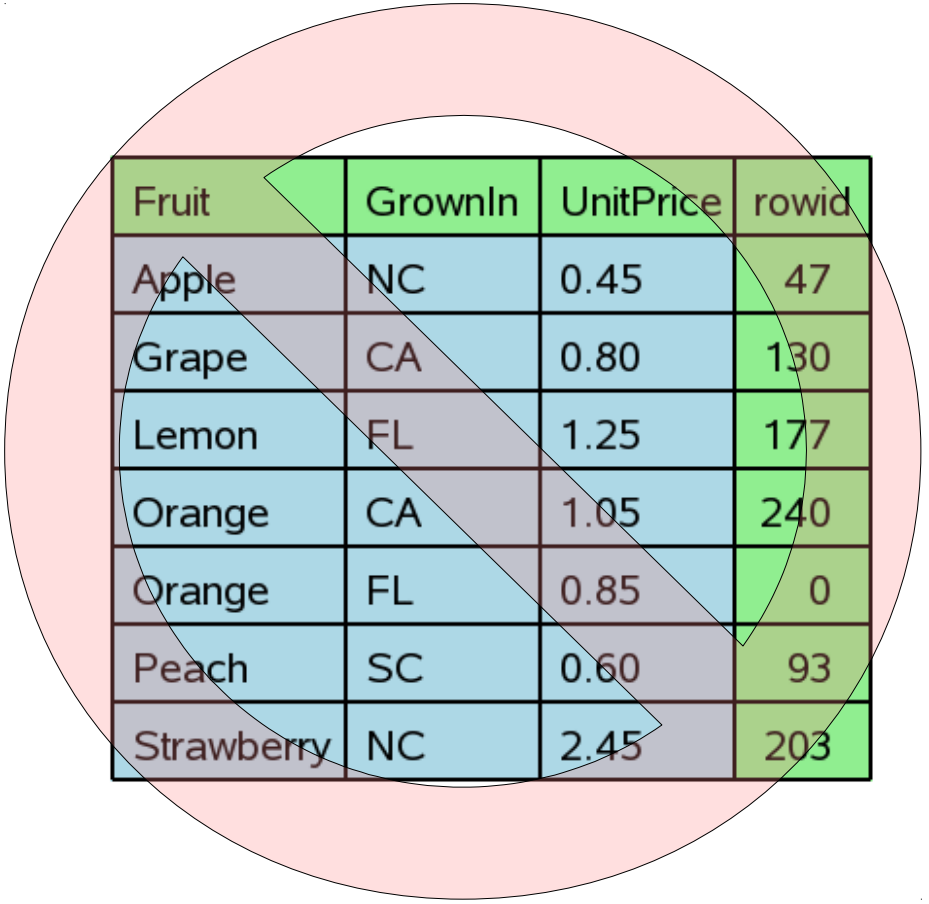


SELECT * FROM tab
ORDER BY fruit DESC, grownin DESC, unitprice DESC



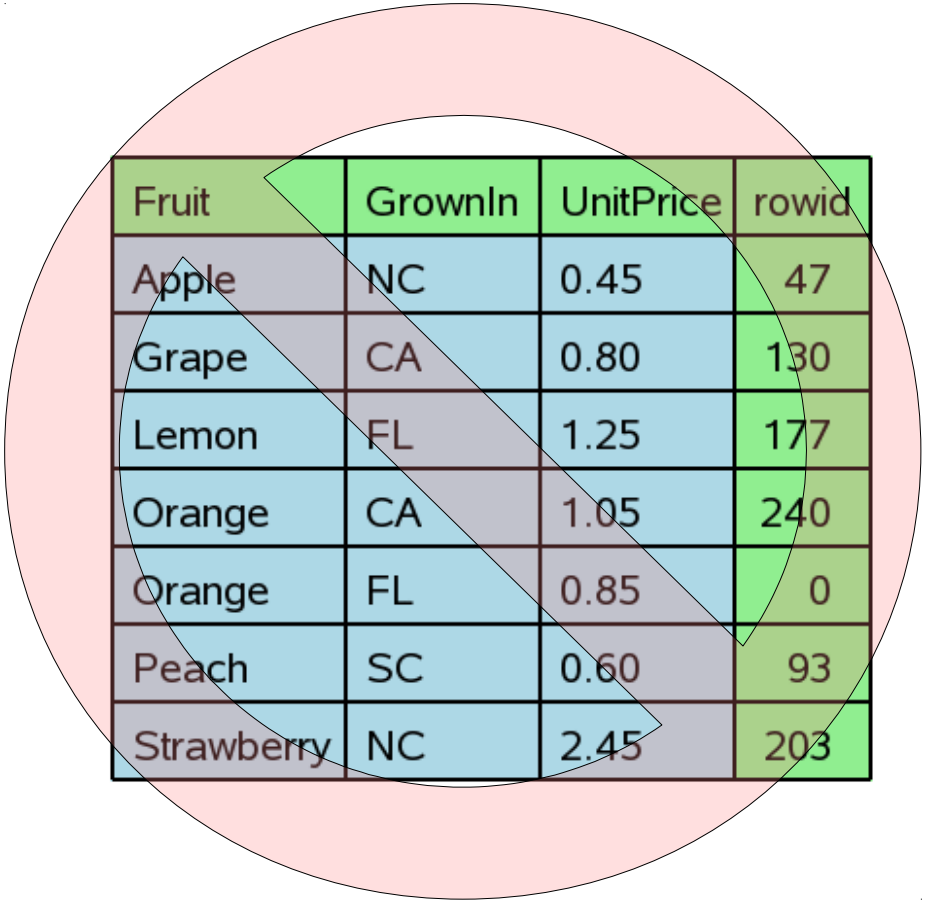
Fruit	GrownIn	UnitPrice	rowid
Apple	NC	0.45	47
Grape	CA	0.80	130
Lemon	FL	1.25	177
Orange	CA	1.05	240
Orange	FL	0.85	0
Peach	SC	0.60	93
Strawberry	NC	2.45	203

SELECT * FROM tab ORDER BY fruit DESC, grownin



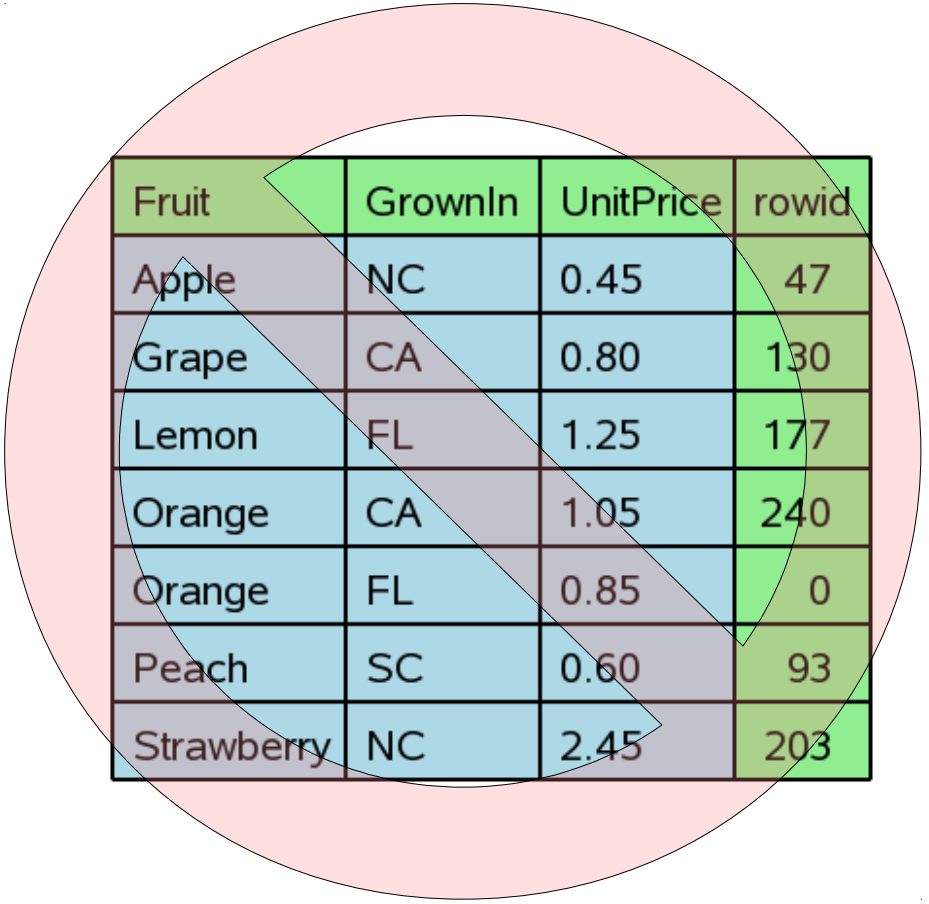
Fruit	GrownIn	UnitPrice	rowid
Apple	NC	0.45	47
Grape	CA	0.80	130
Lemon	FL	1.25	177
Orange	CA	1.05	240
Orange	FL	0.85	0
Peach	SC	0.60	93
Strawberry	NC	2.45	203

SELECT * FROM tab ORDER BY fruit, unitprice



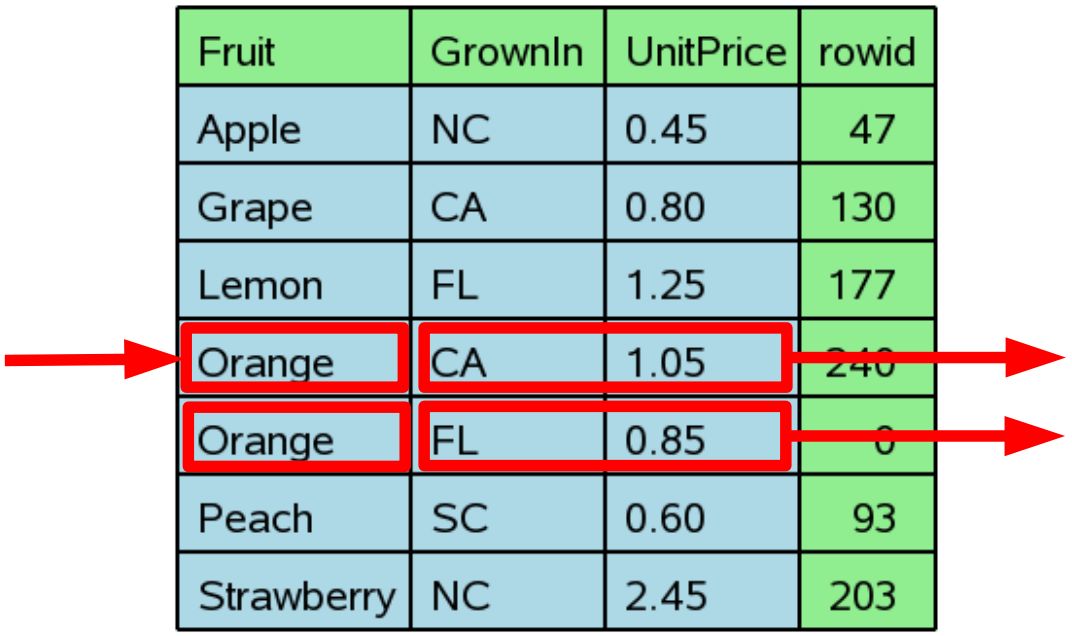
Fruit	GrownIn	UnitPrice	rowid
Apple	NC	0.45	47
Grape	CA	0.80	130
Lemon	FL	1.25	177
Orange	CA	1.05	240
Orange	FL	0.85	0
Peach	SC	0.60	93
Strawberry	NC	2.45	203

SELECT * FROM tab ORDER BY grownin



Fruit	GrownIn	UnitPrice	rowid
Apple	NC	0.45	47
Grape	CA	0.80	130
Lemon	FL	1.25	177
Orange	CA	1.05	240
Orange	FL	0.85	0
Peach	SC	0.60	93
Strawberry	NC	2.45	203

```
SELECT grownin, unitprice FROM tab
 WHERE fruit='Orange'
 ORDER BY grownin
```



Fruit	GrownIn	UnitPrice	rowid
Apple	NC	0.45	47
Grape	CA	0.80	130
Lemon	FL	1.25	177
Orange	CA	1.05	240
Orange	FL	0.85	0
Peach	SC	0.60	93
Strawberry	NC	2.45	203

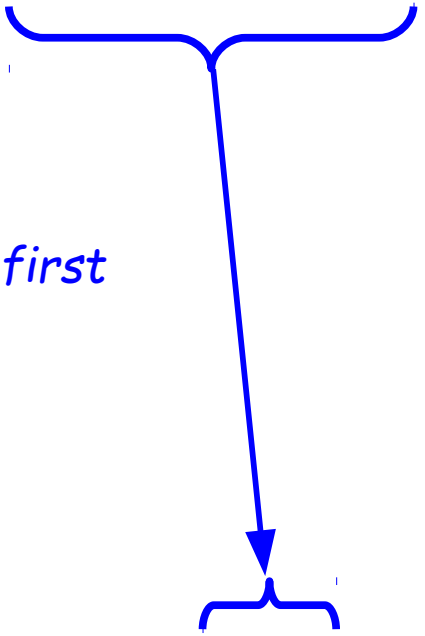
Index Column Order

```
SELECT x, y, z FROM t1 WHERE w=5 AND x=6 ORDER BY x, y;
```

```
CREATE INDEX t1i1 ON t1(w,x,y,z);
```

Index Column Order

```
SELECT x, y, z FROM t1 WHERE w=5 AND x=6 ORDER BY x, y;
```



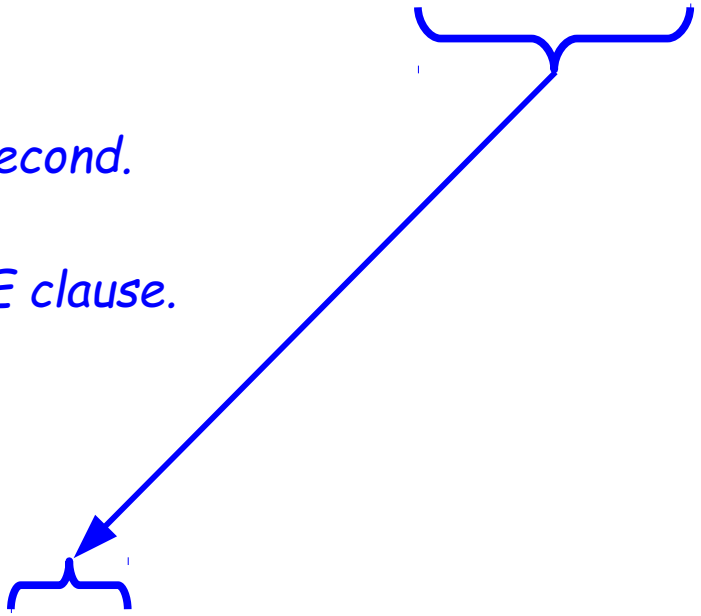
WHERE clause terms first

```
CREATE INDEX t1i1 ON t1(w,x,y,z);
```

Index Column Order

```
SELECT x, y, z FROM t1 WHERE w=5 AND x=6 ORDER BY x, y;
```

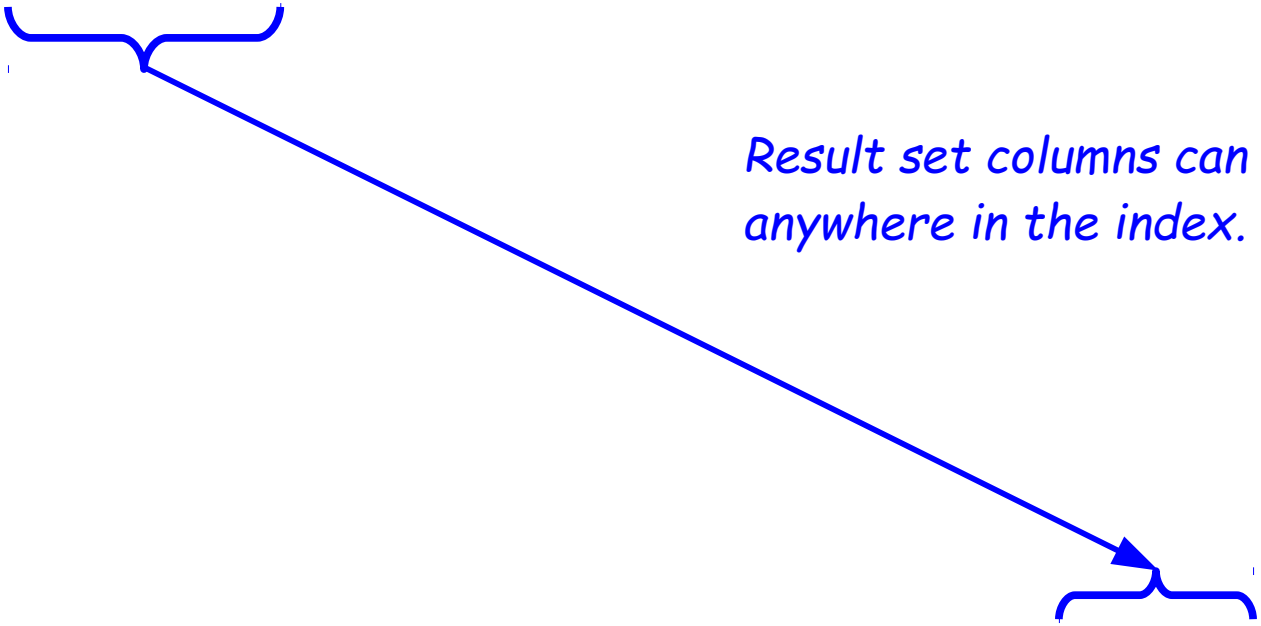
ORDER BY terms second.
No gaps!
Can overlap WHERE clause.



```
CREATE INDEX t1i1 ON t1(w,x,y,z);
```

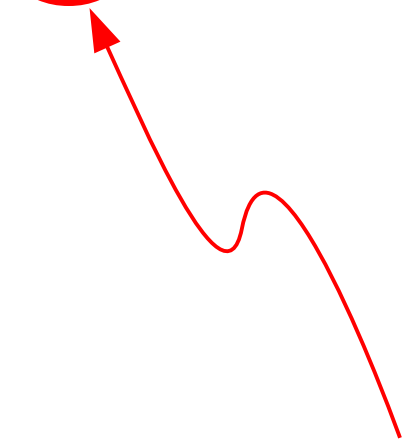
Index Column Order

```
SELECT x, y, z FROM t1 WHERE w=5 AND x=6 ORDER BY x, y;
```



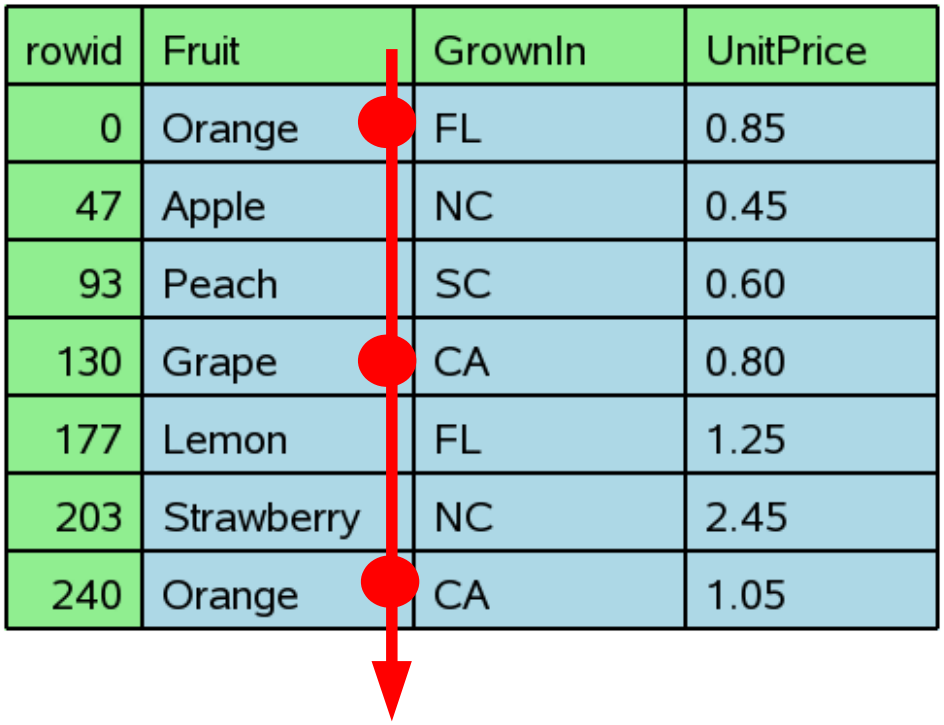

```
CREATE INDEX t1i1 ON t1(w,x,y,z);
```

```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
       OR grownin='CA'
```



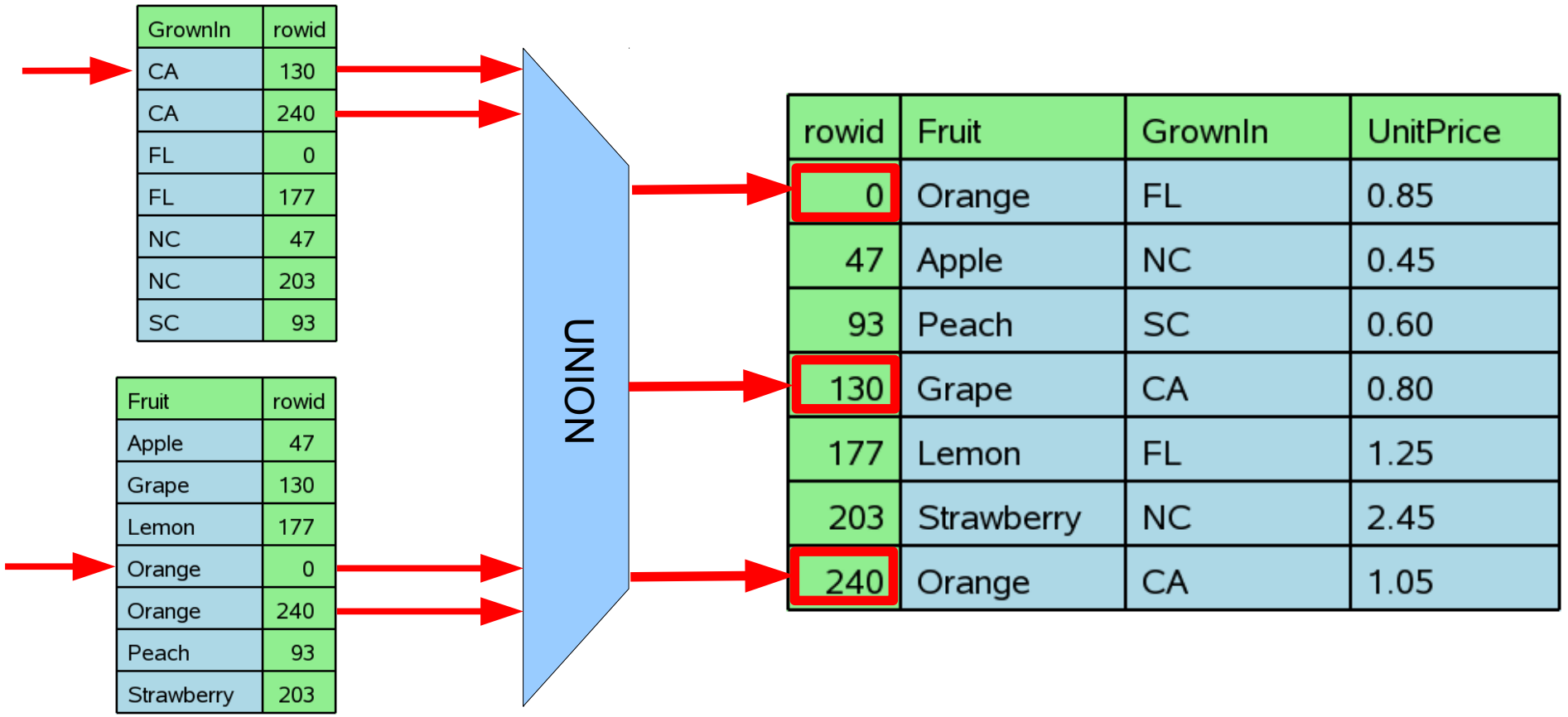
Pesky OR

```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
       OR grownin='CA'
```

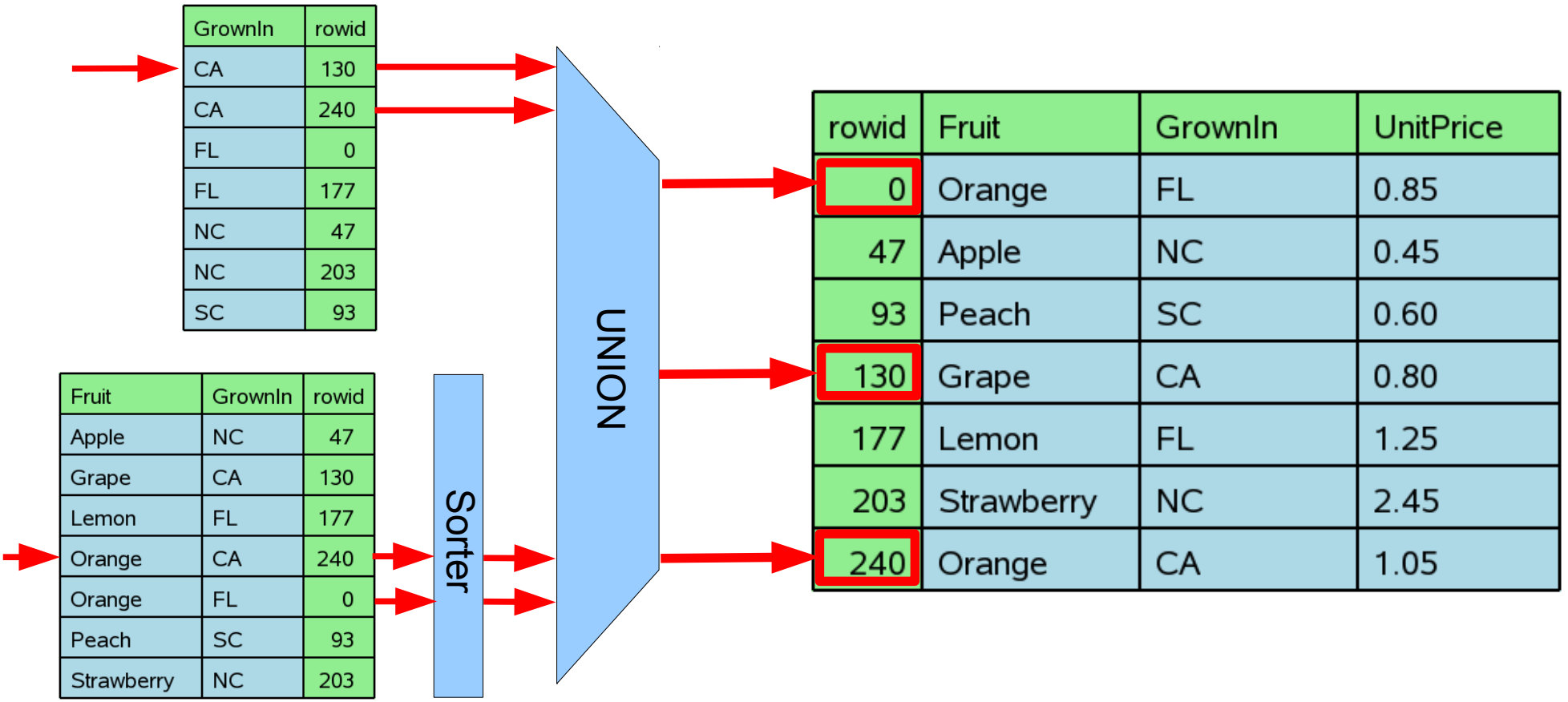


rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

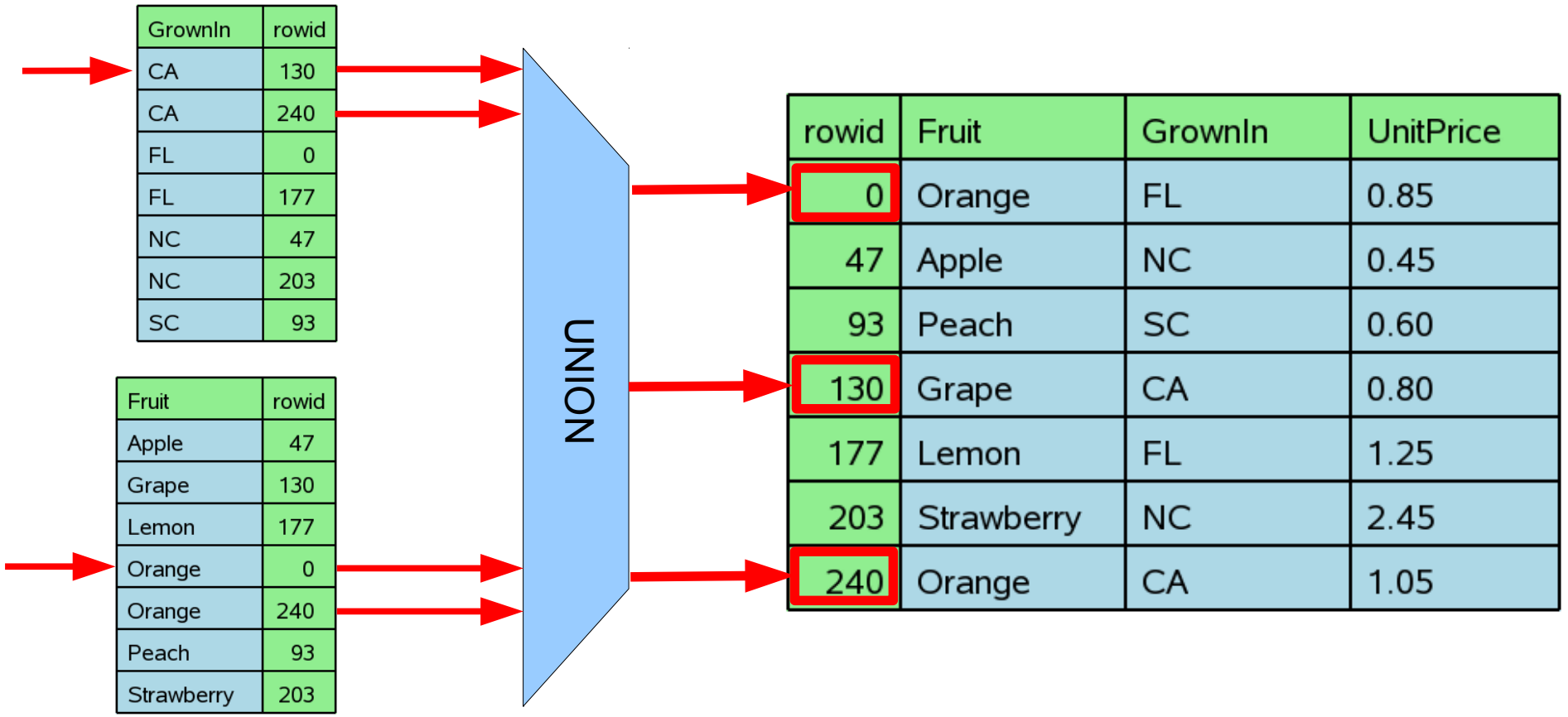
```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
       OR grownin='CA'
```



```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
       OR grownin='CA'
```

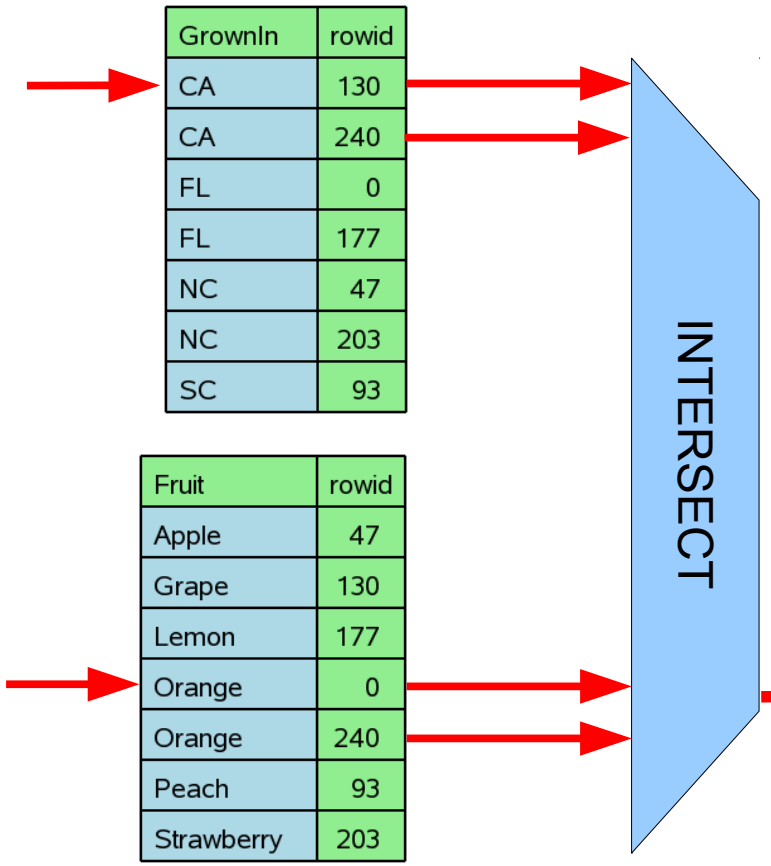


```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
       OR grownin='CA'
```



```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
      AND grownin='CA'
```

6



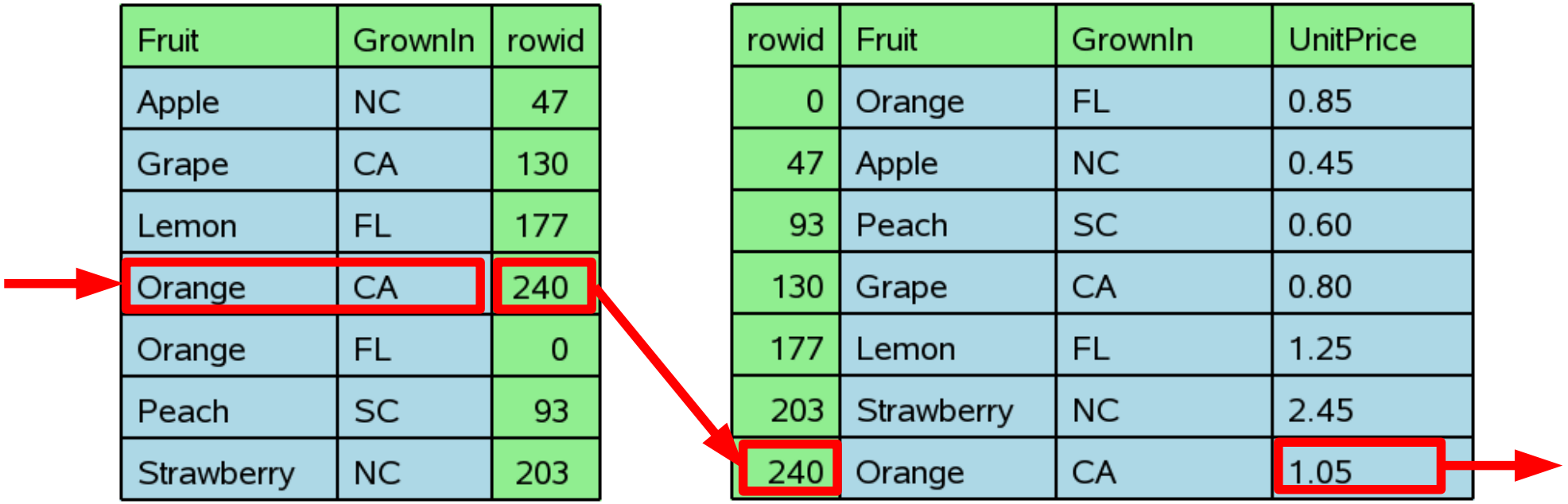
GrownIn	rowid
CA	130
CA	240
FL	0
FL	177
NC	47
NC	203
SC	93

Fruit	rowid
Apple	47
Grape	130
Lemon	177
Orange	0
Orange	240
Peach	93
Strawberry	203

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
     AND  grownin='CA'
```

4

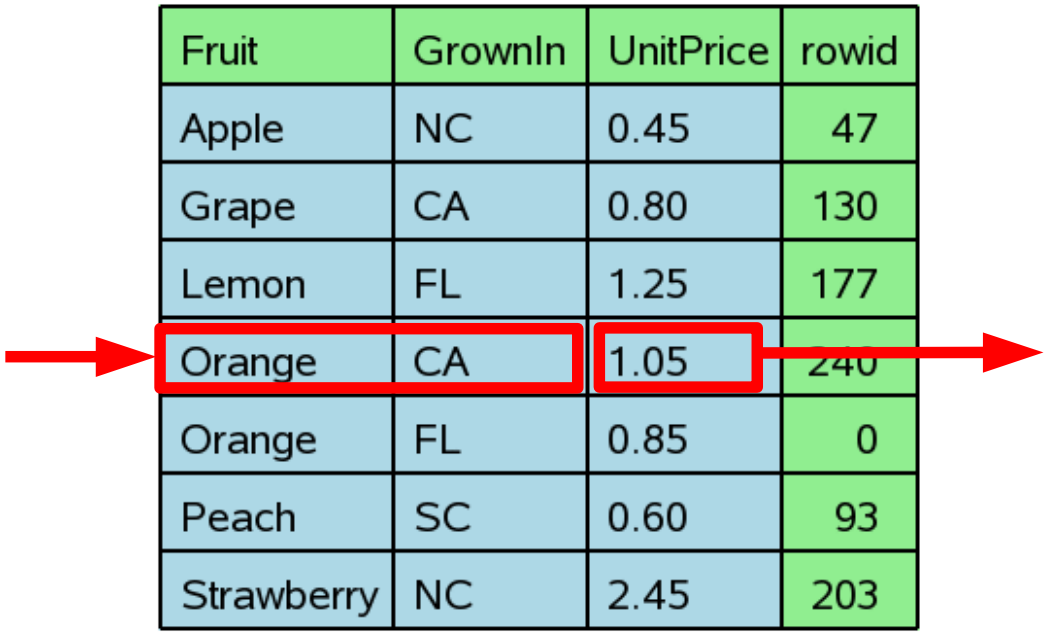


Fruit	GrownIn	rowid
Apple	NC	47
Grape	CA	130
Lemon	FL	177
Orange	CA	240
Orange	FL	0
Peach	SC	93
Strawberry	NC	203

rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

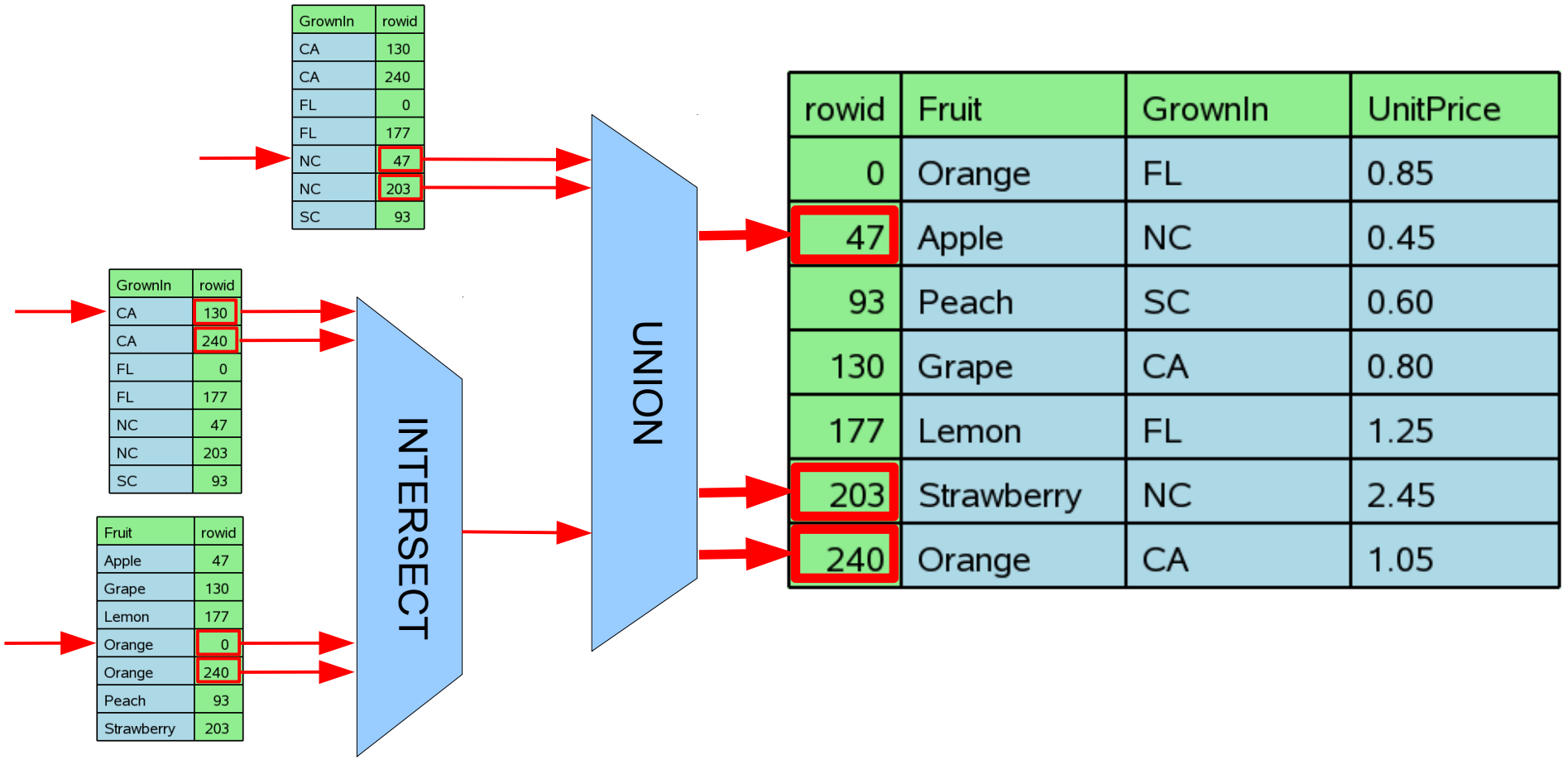
```
SELECT unitprice FROM tab
 WHERE fruit='Orange'
     AND  grownin='CA'
```

5



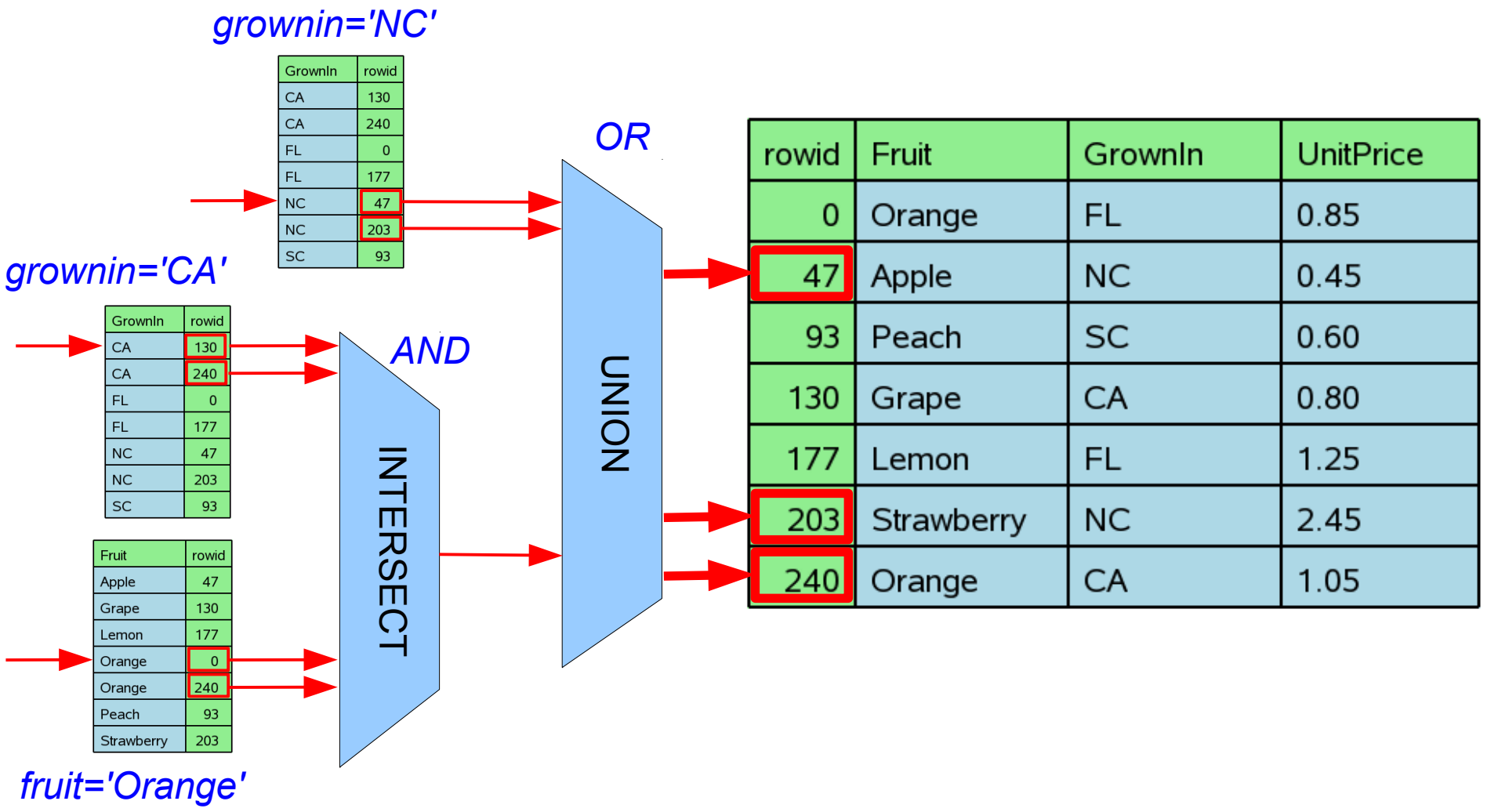
Fruit	GrownIn	UnitPrice	rowid
Apple	NC	0.45	47
Grape	CA	0.80	130
Lemon	FL	1.25	177
Orange	CA	1.05	240
Orange	FL	0.85	0
Peach	SC	0.60	93
Strawberry	NC	2.45	203

```
SELECT unitprice FROM tab
 WHERE (fruit='Orange' AND grownin='CA')
       OR grownin='NC'
```



rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

```
SELECT unitprice FROM tab
 WHERE (fruit='Orange' AND grownin='CA')
       OR grownin='NC'
```



rowid	Fruit	GrownIn	UnitPrice
0	Orange	FL	0.85
47	Apple	NC	0.45
93	Peach	SC	0.60
130	Grape	CA	0.80
177	Lemon	FL	1.25
203	Strawberry	NC	2.45
240	Orange	CA	1.05

Remember This:

SQL is a programming language

There is a compiler

There is a virtual machine

Indices are not magic – they need to make sense

Avoid too many indices

- The more options the optimizer has, the more chances it has to make a poor choice.