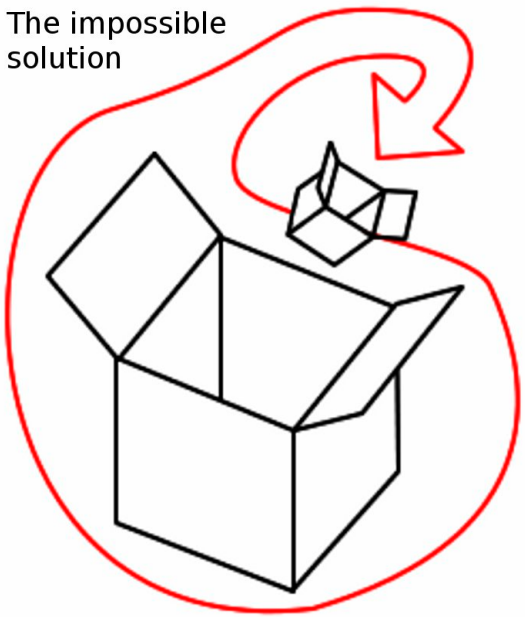
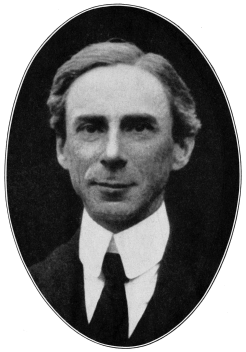
Discussion #2

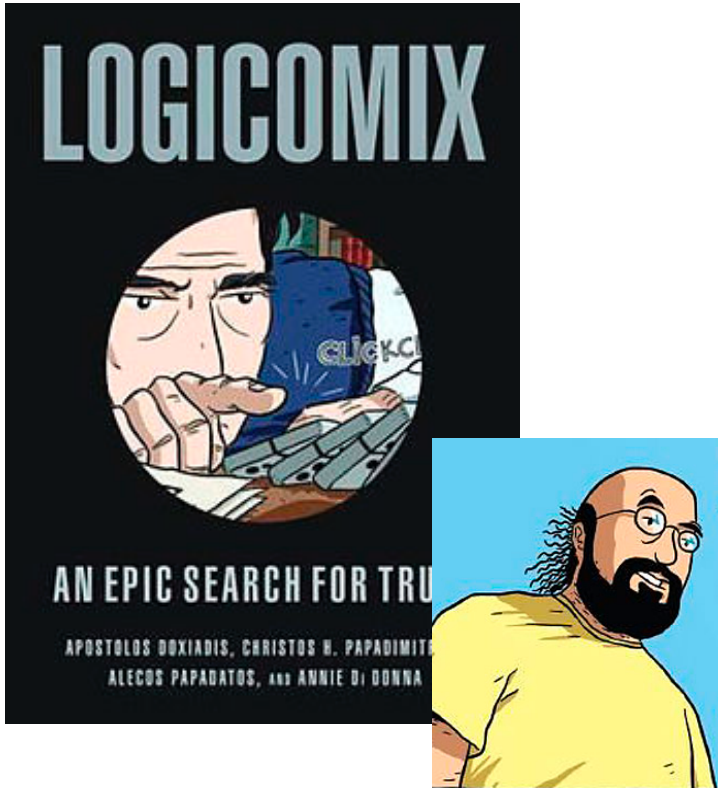
CS186

Fun Aside: Logic

Find the set that contains all set.







http://mindfuckmath.com/post/69927274702/fathom-the-universe-mathematical-paradoxes-get

NULLs

"*It is this writer's opinion that NULLs, at least as currently defined and implemented in SQL, are far more trouble than they are worth and should be avoided; they display very strange and inconsistent behavior and can be a rich source of error and confusion*"

— *A Guide to Sybase And SQL Server, David McGoveran and C. J. Date*

Here are the tables for the three logical operators that come with SQL.

	NOT
TRUE	FALSE
UNKNOWN	UNKNOWN
FALSE	TRUE

AND	TRUE	UNKNOWN	FALSE
TRUE	TRUE	UNKNOWN	FALSE
UNKNOWN	UNKNOWN	UNKNOWN	FALSE
FALSE	FALSE	FALSE	FALSE

OR	TRUE	UNKNOWN	FALSE
TRUE	TRUE	TRUE	TRUE
UNKNOWN	TRUE	UNKNOWN	UNKNOWN
FALSE	TRUE	UNKNOWN	FALSE

Joe Celko's SQL for Smarties: Advanced SQL Programming

NULL

“NULLs are usually implemented with an extra bit somewhere in the row where the column appears, rather than in the column itself. They adversely affect storage requirements, indexing, and searching.”

Joe Celko's SQL for Smarties: Advanced SQL Programming

Everything is a Table

… and most functions are aggregates

Declare the intention

… and leave out the control flow

TR10: Cloud Programming

A new language will improve online applications.

14 comments
ERICA NAONE
May/June 2010

Joseph Hellerstein wants cloud programmers to reach new heights.
Toby Burditt

This article is part of an annual list of what we believe are the 10 most important emerging technologies. See the full list here.

Cloud computing offers the promise of virtually unlimited processing and storage power, courtesy of vast data centers run by companies like Amazon and Google. But programmers don't know how best to exploit this power.

Today, many developers are converting existing programs to run on clouds, rather than creating new types of applications that could work nowhere else. And they are held back by difficulties in keeping track of data and getting reliable information about what's going on across a cloud. If programmers could solve those problems, they could start to really take advantage of what's possible with a cloud. For example, an online music retailer could monitor popular social-media feeds; if a singer suddenly became a hot topic, advertising and special offers across the retailer's site could be instantly reconfigured to make the most of the spike in interest.

General SQL tips

- Don’t think about performance until it’s correct
- When it gets complicated, just create a new table
- Verify your assumptions!

Worksheet!

Using “Window Functions”

Window functions: an SQL idiom to compute with order.

http://www.postgresql.org/docs/9.3/static/tutorial-window.html

```
CREATE VIEW twocounters AS
(SELECT c,
        ROW_NUMBER() OVER (ORDER BY c ASC) AS RowAsc,
        ROW_NUMBER() OVER (ORDER BY c DESC) AS RowDesc
   FROM T
);

SELECT MIN(c)
FROM twocounters
WHERE RowAsc IN (RowDesc, RowDesc - 1, RowDesc + 1);
```

O(n logn!)

Note: handles even # of items.

Binning, Cubing

```
SELECT
Gender,
count(CASE WHEN Age>= 10 AND Age < 20 THEN 1 END) AS [10 - 20],
count(CASE WHEN Age>= 21 AND Age < 30 THEN 1 END) AS [21 - 30],
count(CASE WHEN Age>= 31 AND Age < 35 THEN 1 END) AS [31 - 35],
count(CASE WHEN Age>= 36 AND Age < 40 THEN 1 END) AS [36 - 40]
FROM Attendees AS AgeGroups
GROUP BY Gender
```

http://stackoverflow.com/questions/19103991/create-range-bins-from-sql-server-table-for-histograms

Table 1

whn	description	amount
2006-11-01	Wages	50
2006-11-02	Company Store	-10
2006-11-03	Company Store	-10
2006-11-04	Company Store	-10
2006-11-05	Company Store	-10
2006-11-06	Company Store	-10

Table 2

whn	description	cshIN	cshOUT
2006-11-01	Wages	50	
2006-11-02	Company Store		10
2006-11-03	Company Store		10
2006-11-04	Company Store		10
2006-11-05	Company Store		10
2006-11-06	Company Store		10

Table 1

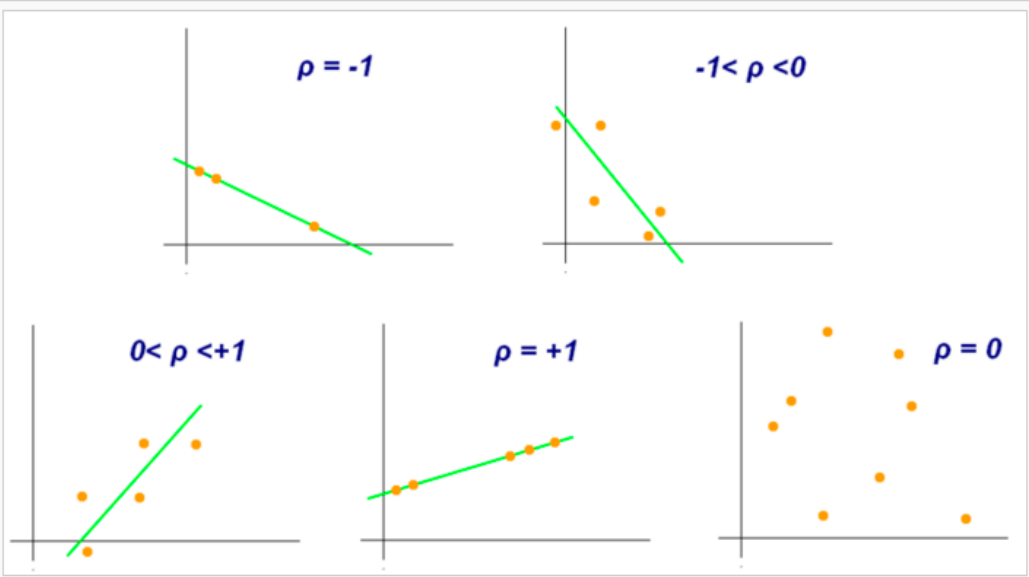
whn	description	amount
2006-11-01	Wages	50
2006-11-02	Company Store	-10
2006-11-03	Company Store	-10
2006-11-04	Company Store	-10
2006-11-05	Company Store	-10
2006-11-06	Company Store	-10

Table 2

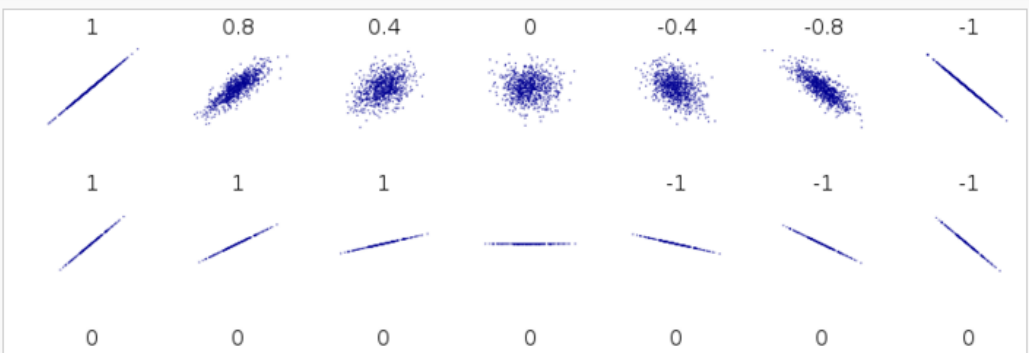
whn	description	cshIN	cshOUT
2006-11-01	Wages	50	
2006-11-02	Company Store		10
2006-11-03	Company Store		10
2006-11-04	Company Store		10
2006-11-05	Company Store		10
2006-11-06	Company Store		10

```
SELECT w AS dte, d AS description,
   CASE WHEN (a>=0) THEN a ELSE NULL END AS cshIN,
   CASE WHEN (a<0) THEN SUBSTR(a,2,10) ELSE NULL END AS cshOUT
   FROM
   (SELECT x.whn AS w, x.description AS d,
           x.amount AS a
      FROM transact x
      JOIN transact y ON (x.whn>=y.whn)
      GROUP BY x.whn, x.description, x.amount) t
```

Pearson Rating



Examples of scatter diagrams with different values of correlation coefficient (ρ)



Pearson Rating

```
-- PEARSON

SELECT user_1, user_2,
       (
         (psum - (sum1 * sum2 / n))/
         ((sum1sq - sum1*sum1) / n)*
         (sum2sq - sum2*sum1) / n)
       AS pearson_r, n
 FROM (SELECT N1.user AS user_1, N2.user AS user_2,
       SUM(N1.rating) AS sum1, SUM(N2.rating) AS sum2,
       SUM(N1.rating * N1.rating) AS sum1sq,
       SUM(N2.rating * N2.rating) AS sum2sq,
       SUM(N1.rating * N2.rating) AS psum,
       COUNT(*) AS N
       FROM Movie_Reviews AS N1
         LEFT OUTER JOIN
         Movie_Reviews AS N2
         ON N1.movie = N2.movie
 WHERE N1.user > N2.user
 GROUP BY N1.user, N2.user)
AS step1;
```