Joins, Skew
and
Histograms

Wolfgang Breitling
www.centrexcc.com

Who am I

Independent consultant since 1996

specializing in Oracle and Peoplesoft setup, administration, and performance tuning

Member of the Oaktable Network

25+ years in database management

DL/1, IMS, ADABAS, SQL/DS, DB2, Oracle

OCP certified DBA - 7, 8, 8*i*, 9*i*

Oracle since 1993 (7.0.12)

Mathematics major from University of Stuttgart

Tuning by swapping Predicate Sides?

el	cpu	pio	lio	rows	sql_text
0.16	0.16	0	10,996	35	select a.c1, b.c1, c.c1 from t1 a, t2 b, t3 c where a.n1 = :n1 and b.n1 = :n2 and a.n2 = c.n2 **and a.n2 = b.n2**
0.07	0.09	0	10,787	35	select a.c1, b.c1, c.c1 from t1 a, t2 b, t3 c where a.n1 = :n1 and b.n1 = :n2 and a.n2 = c.n2 **and b.n2 = a.n2**

Tuning by swapping Predicate Sides?

id	cost	card	operation	rows	elapsed	cr_gets
0	1239		SELECT STATEMENT			
1A	1239	199	HASH JOIN	35	0.160	10,996
2A	1204	8	HASH JOIN	35	0.090	10,706
3F	112	39	TABLE ACCESS FULL T1	39	0.010	979
4F	1092	80	TABLE ACCESS FULL T2	70	0.080	9,727
5	34	4800	TABLE ACCESS FULL T3	4,800	0.020	290

id	cost	card	operation	rows	elapsed	cr_gets
0	1229		SELECT STATEMENT			
1	6	24	TABLE ACCESS BY INDEX ROWID T3	35	0.070	10,787
2	1229	113	NESTED LOOPS	71	0.070	10,783
3A	1204	5	HASH JOIN	35	0.070	10,706
4F	112	39	TABLE ACCESS FULL T1	39	0.010	979
5F	1092	80	TABLE ACCESS FULL T2	70	0.050	9,727
6A	3	24	INDEX RANGE SCAN T3_2	35	0.000	77

Focus

- **Join Cardinality and the Principle of Inclusion**
  - **Cases where the basic formula fails**
- **Histograms and Join Cardinality**
  - **Where Histograms improve the Cardinality Estimate**
  - **Alberto Dell'Era's Join Cardinality Algorithm**
  - **Can Histogram make things worse?**
- **Upgrade Issues**

Principle of Inclusion

"principle of inclusion"
each value of the smaller domain has a match in the larger domain

This is obviously true for joins between foreign keys and primary keys.

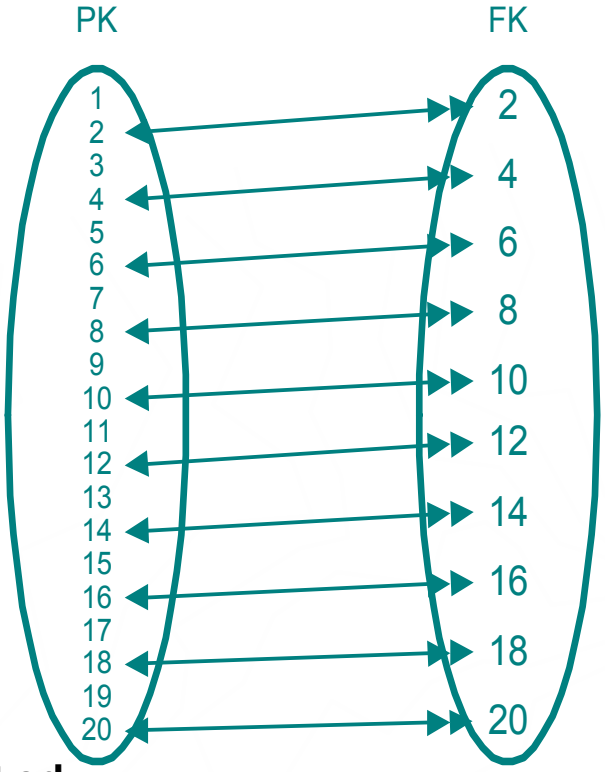
$$\text{join cardinality} = \text{card}_{\mathbf{A}} * \text{card}_{\mathbf{B}} * \text{join selectivity}$$

$$\text{join selectivity} = 1/\max(\text{ndv}_{\mathbf{A}}, \text{ndv}_{\mathbf{B}})$$

Principle of Inclusion

```
SQL> select 'A.'||a.id c1, 'B.'||b.id c2
from j1 A, j2 B
where a.id = b.id;

  20  SELECT STATEMENT
  20   HASH JOIN
  20    TABLE ACCESS  J1
  20    TABLE ACCESS  J2
```

```
C1     C2
-----  -----
A.2    B.2
A.4    B.4
A.6    B.6
A.8    B.8
A.10   B.10
A.12   B.12
A.14   B.14
A.16   B.16
A.18   B.18
A.20   B.20
 …
20 rows selected.
```



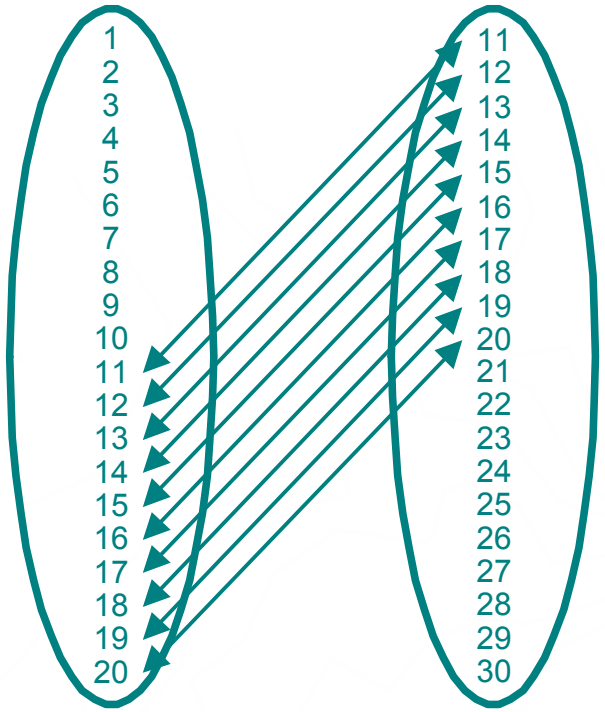
$$\text{Join cardinality} = \text{card}_A * \text{card}_B * \text{join selectivity}$$
$$= \text{card}_A * \text{card}_B * 1/\max(\text{ndv}_a, \text{ndv}_b)$$
$$= 20 * 20 * 1/\max(20, 10) = 20$$

Principle of Inclusion violated

```
SQL> select 'A.'||a.id c1, 'B.'||b.id c2
from j1 A, j2 B
where a.id = b.id;
```

```
20  SELECT STATEMENT
20   HASH JOIN
20    TABLE ACCESS J1
20    TABLE ACCESS J2
```

```
C1     C1
-----  -----
A.11   B.11
A.12   B.12
A.13   B.13
A.14   B.14
A.15   B.15
A.16   B.16
A.17   B.17
A.18   B.18
A.19   B.19
A.20   B.20

10 rows selected.
```



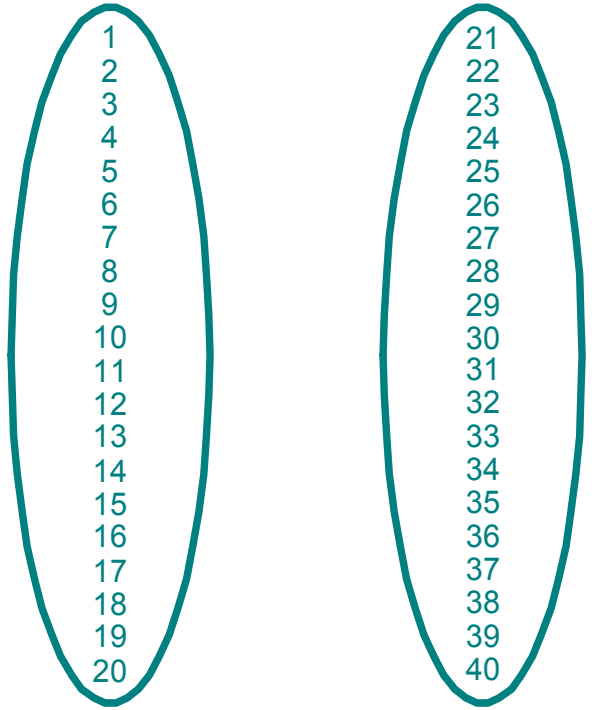
$$\text{Join cardinality} = \text{card}_A * \text{card}_B * \text{join selectivity}$$
$$= \text{card}_A * \text{card}_B * 1/\max(\text{ndv}_a, \text{ndv}_b)$$
$$= 20 * 20 * 1/\max(20, 10) = 20$$

Principle of Inclusion violated

```
SQL> select 'A.'||a.id c1, 'B.'||b.id c2
from j1 A, j2 B
where a.id = b.id;
```

```
  1  SELECT STATEMENT
  1   HASH JOIN
 20    TABLE ACCESS J1
 20    TABLE ACCESS J2
```

```
C1    C2
----- -----

no rows selected.
```

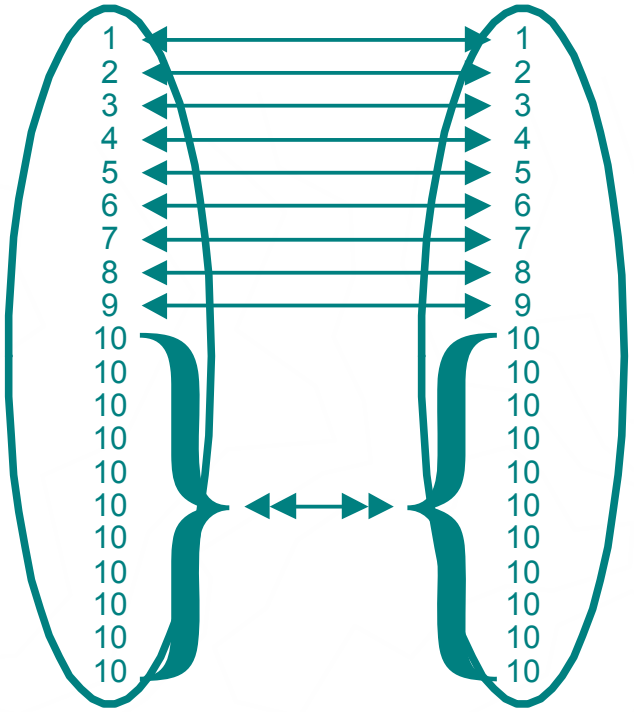


$$\text{Join cardinality} = \text{card}_A * \text{card}_B * \text{join selectivity}$$
$$= \text{card}_A * \text{card}_B * 1/\max(\text{ndv}_a, \text{ndv}_b)$$
$$= 20 * 20 * 1/\max(20, 20) = 20$$

Skew

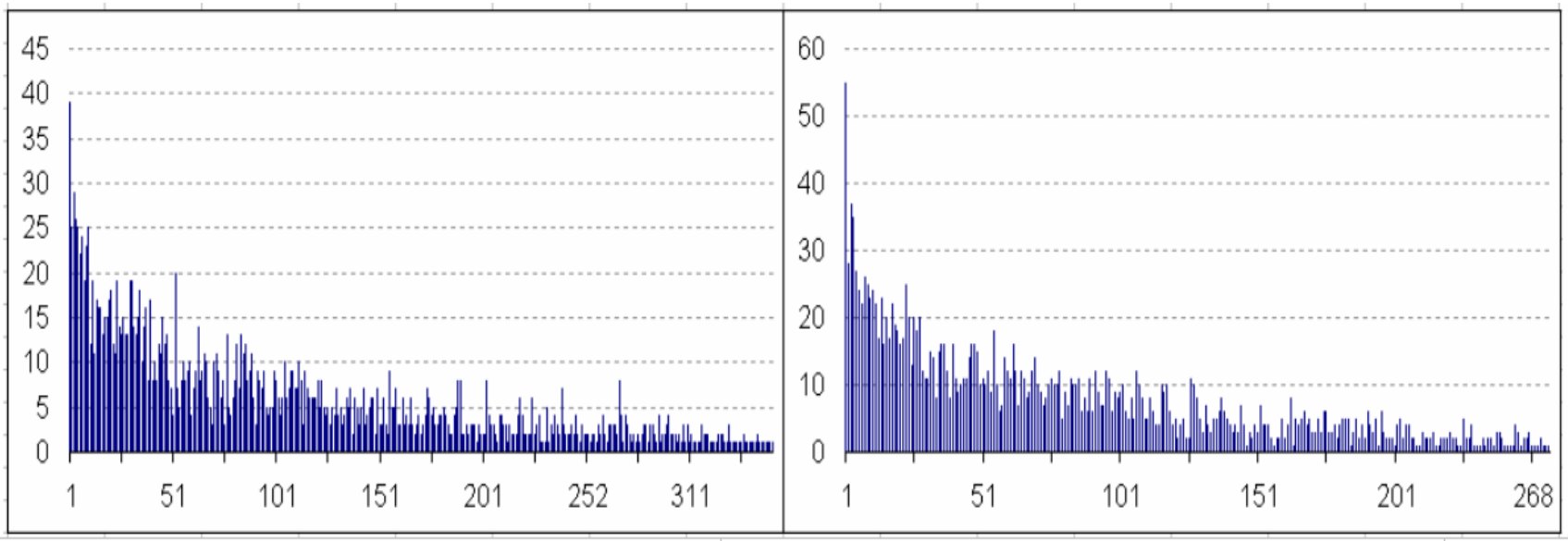
```
SQL> select 'A.'||a.id c1, 'B.'||b.id c2
from j1 A, j2 B
where a.id = b.id;

  40  SELECT STATEMENT
  40   HASH JOIN
  20    TABLE ACCESS J1
  20    TABLE ACCESS J2
```

```
C1    C2
----- -----
A.1   B.1
A.2   B.2
A.3   B.3
A.4   B.4
A.5   B.5
A.6   B.6
A.7   B.7
A.8   B.8
A.9   B.9
A.10  B.10
A.10  B.10
…
130 rows selected.
```



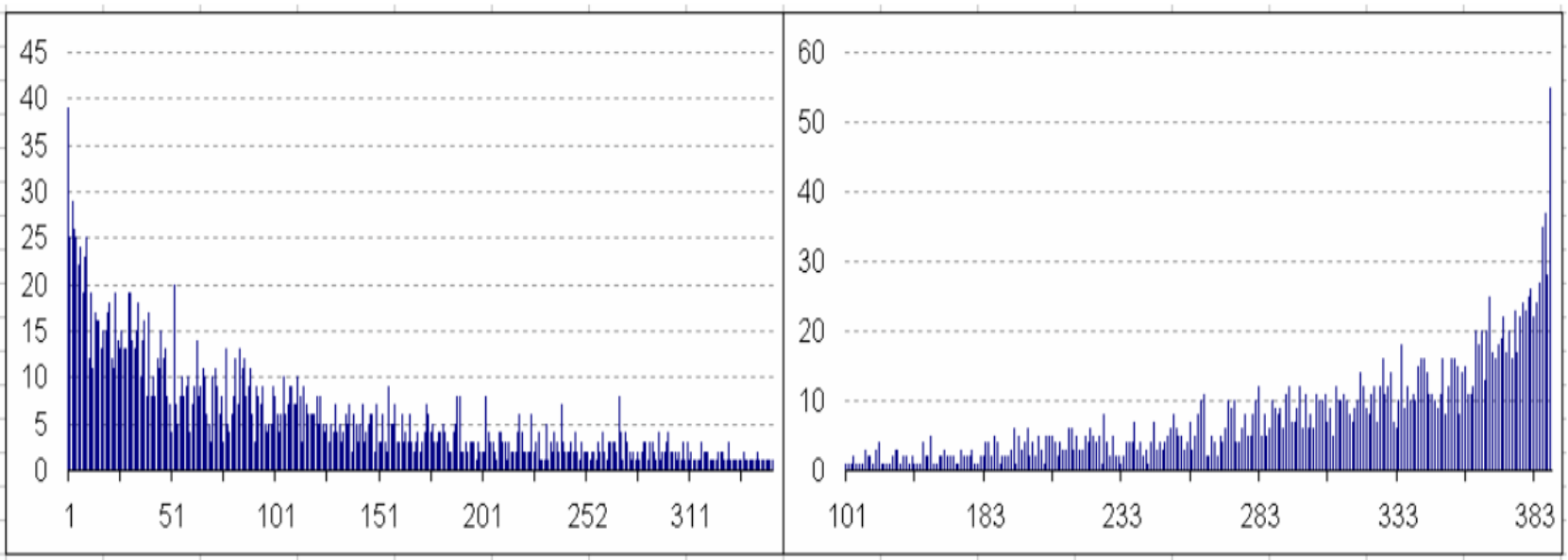
$$\text{Join cardinality} = \text{card}_A * \text{card}_B * \text{join selectivity}$$
$$= \text{card}_A * \text{card}_B * 1/\max(\text{ndv}_a, \text{ndv}_b)$$
$$= 20 * 20 * 1/\max(10, 10) = 40$$

“Symmetrical” Skew



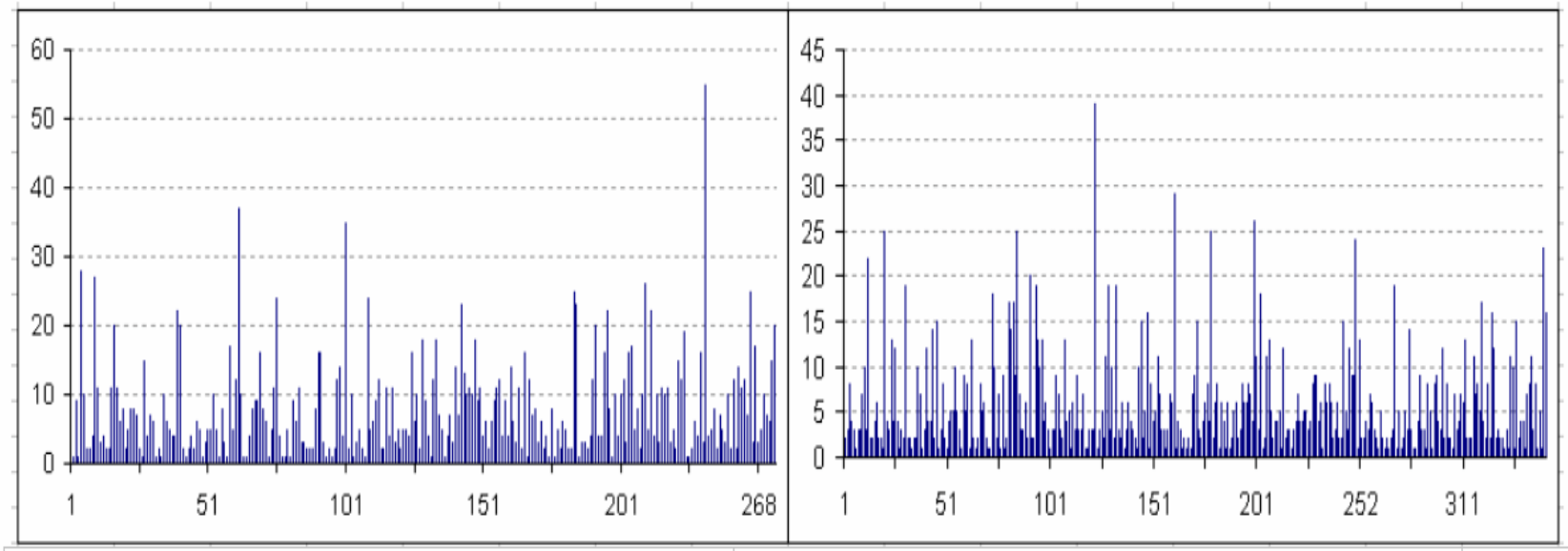
With histograms				Without histograms				actual
Id	Operation	Name	Rows	Id	Operation	Name	Rows	
0	SELECT STATEMENT		21350	0	SELECT STATEMENT		11730	
* 1	HASH JOIN		**21350**	* 1	HASH JOIN		**11730**	**23568**
	error		9.4%		error		50.2%	

“Anti-symmetrical” Skew



With histograms				Without histograms				actual
Id	Operation	Name	Rows	Id	Operation	Name	Rows	
0	SELECT STATEMENT		10381	0	SELECT STATEMENT		11730	
* 1	HASH JOIN		**10381**	* 1	HASH JOIN		**11730**	2712
error			282.8%	error			332.5%	

“Random” Skew



With histograms

```
| Id | Operation            | Name | Rows |
-----------------------------------------
|  0 | SELECT STATEMENT     |      |15270 |
|* 1 |  HASH JOIN           |      |15270 |
-----------------------------------------
```

error 35%

Without histograms

```
| Id | Operation            | Name | Rows |
-----------------------------------------
|  0 | SELECT STATEMENT     |      |11730 |
|* 1 |  HASH JOIN           |      |11730 |
-----------------------------------------
```

error 3.5%

actual

11328

Testcase Rowcounts

value	count
1	54
6	18
8	17
9	19
10	25
15	111
16	31
17	81
18	37
20	64
21	24
22	27
23	29
26	23
27	22
30	193
31	13
32	20
33	47
35	34
38	21
45	17
46	41
47	16
48	16

value	count
4	41
5	18
7	193
8	16
9	33
10	21
15	30
17	101
19	31
21	20
22	19
26	81
30	104
35	21
38	53
40	16
43	30
47	24
55	17
59	25
60	29
61	25
63	13
66	22
79	17

Cardinality Actual vs. Estimate

table	column	NDV	density	nulls	lo	hi	av lg	bkts
TEST1	N1	25	3.0601E-02	0	1	48	3	13
TEST2	N1	25	2.8248E-02	0	4	79	3	14

select count(0) from (select a.c1, b.c1 from test1 a, test2 b where a.n1 = b.n1)

COUNT(0)
38074

1 row selected.

id	cost	card	operation
0	9		SELECT STATEMENT
1		1	SORT AGGREGATE
2A	9	42,665	HASH JOIN
3	4	1,000	TABLE ACCESS FULL TEST1
4	4	1,000	TABLE ACCESS FULL TEST2

id	PREDICATES
2	Access= "N1"="N1"

Histograms

EP	value	bkts	count	pop
1	1	1	50	0
2	9	1	50	0
4	15	2	100	1
5	16	1	50	0
7	17	2	100	1
9	20	2	100	1
10	22	1	50	0
11	26	1	50	0
15	30	4	200	1
16	32	1	50	0
17	33	1	50	0
18	38	1	50	0
19	46	1	50	0
20	48	1	50	0
		20	1000	

EP	value	bkts	count	pop
0	4	0	0	0
1	5	1	50	0
5	7	4	200	1
6	9	1	50	0
7	15	1	50	0
9	17	2	100	1
10	21	1	50	0
12	26	2	100	1
14	30	2	100	1
15	38	1	50	0
16	43	1	50	0
17	47	1	50	0
18	60	1	50	0
19	63	1	50	0
20	79	1	50	0
		20	1000	

Join Histogram

Value	Count-1	Count-2	Pop-1	Pop-2	Pop-J
1	50		0		0
4		0		0	0
5		50		0	0
7		200		1	1
9	50	50	0	0	0
15	100	50	1	0	1
16	50		0		0
17	100	100	1	1	2
20	100		1		1
21		50		0	0
22	50		0		0
26	50	100	0	1	1
30	200	100	1	1	2
32	50		0		0
33	50		0		0
38	50	50	0	0	0
43		50		0	0
46	50		0		0
47		50		0	0
48	50		0		0
60		50		0	0
63		50		0	0
79		50		0	0

CJH

CJH+2

Contribution of Join Popularity 2

		0.030601	0.028248			
Value	Count-1	Count-2	Pop-1	Pop-2	Pop-J	card
9	50	50	0	0	0	
15	100	50	1	0	1	
16	50		0		0	
17	**100**	**100**	**1**	**1**	**2**	**10,000**
20	100		1		1	
21		50		0	0	
22	50		0		0	
26	50	100	0	1	1	
30	**200**	**100**	**1**	**1**	**2**	**20,000**
32	50		0		0	
33	50		0		0	
38	50	50	0	0	0	
43		50		0	0	
46	50		0		0	
47		50		0	0	
48	50		0		0	
60		50		0	0	
63		50		0	0	

30,000

Contribution of Join Popularity 1

			0.030601	0.028248		
Value	Count-1	Count-2	Pop-1	Pop-2	Pop-J	card
9	50	50	0	0	0	
15	**100**	**50**	**1**	**0**	**1**	**2,825**
16	50		0		0	
17	100	100	1	1	2	10,000
20	**100**		**1**		**1**	**2,825**
21		50		0	0	
22	50		0		0	
26	**50**	**100**	**0**	**1**	**1**	**3,060**
30	200	100	1	1	2	20,000
32	50		0		0	
33	50		0		0	
38	50	50	0	0	0	
43		50		0	0	
46	50		0		0	
47		50		0	0	
48	50		0		0	
60		50		0	0	
63		50		0	0	

38,710

Contribution of non-popular Values

	0.030601	0.028248				
Value	Count-1	Count-2	Pop-1	Pop-2	Pop-J	card
9	50	50	0	0	0	
15	100	50	1	0	1	2,825
16	50		0		0	
17	100	100	1	1	2	10,000
20	100		1		1	2,825
21		50		0	0	
22	50		0		0	
26	50	100	0	1	1	3,060
30	200	100	1	1	2	20,000
32	50		0		0	
33	50		0		0	
38	50	50	0	0	0	
43		50		0	0	
46	50		0		0	
47		50		0	0	
48	50		0		0	
60		50		0	0	
63		50		0	0	
	400	350				3,955
						42,665

Oddities - Asymmetry

EP	value	bkts	count	pop
0	-48	0	0	0
1	-46	1	50	0
2	-38	1	50	0
3	-33	1	50	0
4	-32	1	50	0
8	-30	4	200	1
9	-26	1	50	0
10	-22	1	50	0
12	-20	2	100	1
14	-17	2	100	1
15	-16	1	50	0
17	-15	2	100	1
18	-9	1	50	0
20	-1	2	100	1
		20	1000	1

EP	value	bkts	count	pop
0	-79	0	0	0
1	-63	1	50	0
2	-60	1	50	0
3	-47	1	50	0
4	-43	1	50	0
5	-38	1	50	0
7	-30	2	100	1
9	-26	2	100	1
10	-21	1	50	0
12	-17	2	100	1
13	-15	1	50	0
14	-9	1	50	0
18	-7	4	200	1
19	-5	1	50	0
20	-4	1	50	0
		20	1000	

Oddities - Asymmetry

Value	Count-1	Count-2	Pop-1	Pop-2	Pop-J	card
-79	0	0	0	0	0	
-63	0	50	0	0	0	
-60	0	50	0	0	0	
-48	0	0	0	0	0	
-47	0	50	0	0	0	
-46	50	0	0	0	0	
-43	0	50	0	0	0	
-38	50	50	0	0	0	
-33	50	0	0	0	0	
-32	50	0	0	0	0	
-30	200	100	1	1	2	20,000
-26	50	100	0	1	1	2,806
-22	50	0	0	0	0	
-21	0	50	0	0	0	
-20	100	0	1	0	1	2,825
-17	100	100	1	1	2	10,000
-16	50	0	0	0	0	
-15	100	50	1	0	1	2,825
-9	50	50	0	0	0	
-7	0	200	0	1	1	5,612
-5	0	50	0	0	0	
-4	0	50	0	0	0	
-1	100	0	1	0	1	
	300	250				2,104
						46,172

Oddities – another Asymmetry

9.2.0.7>explain plan for select 1 from t1 a, t2 b where a.n1 = b.n1;

```
-------------------------------------------------------------
| Id  | Operation                | Name         | Rows  |
-------------------------------------------------------------
|   0 | SELECT STATEMENT         |              | 46172 |
|*  1 |  HASH JOIN               |              | 46172 |
```

Oddities – another Asymmetry

9.2.0.7>explain plan for select 1 from t1 a, t2 b where **a.n1 = b.n1**;

```
-------------------------------------------------------
| Id  | Operation           | Name        | Rows  |
-------------------------------------------------------
|   0 | SELECT STATEMENT    |             | 46172 |
|*  1 |  HASH JOIN          |             | 46172 |
```

9.2.0.7>explain plan for select 1 from t1 a, t2 b where **b.n1 = a.n1**;

```
-------------------------------------------------------
| Id  | Operation           | Name        | Rows  |
-------------------------------------------------------
|   0 | SELECT STATEMENT    |             | 47855 |
|*  1 |  HASH JOIN          |             | 47855 |
```

Halving "Mystery"

Join column 1 25 unique values Join column 2 25 matching distinct values

Frequency histograms with 25 buckets for both columns

EP	value	counts	pop
1	1	1	0
2	2	1	0
3	3	1	0
4	4	1	0
5	5	1	0
6	6	1	0
7	7	1	0
8	8	1	0
9	9	1	0
10	10	1	0
11	11	1	0
12	12	1	0
13	13	1	0
14	14	1	0
15	15	1	0
16	16	1	0
17	17	1	0
18	18	1	0
19	19	1	0
20	20	1	0
21	21	1	0
22	22	1	0
23	23	1	0
24	24	1	0
25	25	1	0

EP	value	counts	pop
196	1	196	1
392	2	196	1
588	3	196	1
784	4	196	1
980	5	196	1
981	6	1	0
982	7	1	0
983	8	1	0
984	9	1	0
985	10	1	0
986	11	1	0
987	12	1	0
988	13	1	0
989	14	1	0
990	15	1	0
991	16	1	0
992	17	1	0
993	18	1	0
994	19	1	0
995	20	1	0
996	21	1	0
997	22	1	0
998	23	1	0
999	24	1	0
1000	25	1	0

Halving “Mystery”

Value	Count-1	Count-2	Pop-1	Pop-2	Pop-J	card
1	1	196	0	1	1	98
2	1	196	0	1	1	98
3	1	196	0	1	1	98
4	1	196	0	1	1	98
5	1	196	0	1	1	98
6	1	1	0	0	0	0
7	1	1	0	0	0	0
8	1	1	0	0	0	0
9	1	1	0	0	0	0
10	1	1	0	0	0	0
11	1	1	0	0	0	0
12	1	1	0	0	0	0
13	1	1	0	0	0	0
14	1	1	0	0	0	0
15	1	1	0	0	0	0
16	1	1	0	0	0	0
17	1	1	0	0	0	0
18	1	1	0	0	0	0
19	1	1	0	0	0	0
20	1	1	0	0	0	0
21	1	1	0	0	0	0
22	1	1	0	0	0	0
23	1	1	0	0	0	0
24	1	1	0	0	0	0
25	1	1	0	0	0	0
						490

Special Cardinality

Join column 1

50 unique values and height-balanced histogram with 25 buckets

9.2.0.7 and 10.2.0.2

EP	value	bkts	count	pop
0	1	0	0	0
1	2	1	2	0
2	4	1	2	0
3	6	1	2	0
4	8	1	2	0
5	10	1	2	0
6	12	1	2	0
7	14	1	2	0
8	16	1	2	0
9	18	1	2	0
10	20	1	2	0
11	22	1	2	0
12	24	1	2	0
13	26	1	2	0
14	28	1	2	0
15	30	1	2	0
16	32	1	2	0
17	34	1	2	0
18	36	1	2	0
19	38	1	2	0
20	40	1	2	0
21	42	1	2	0
22	44	1	2	0
23	46	1	2	0
24	48	1	2	0
25	50	1	2	0

10.1.0.5 and 10.2.0.1

EP	value	bkts	count	pop
0	1	0	0	0
1	3	1	2	0
2	5	1	2	0
3	7	1	2	0
4	9	1	2	0
5	11	1	2	0
6	13	1	2	0
7	15	1	2	0
8	17	1	2	0
9	19	1	2	0
10	21	1	2	0
11	23	1	2	0
12	25	1	2	0
13	27	1	2	0
14	29	1	2	0
15	31	1	2	0
16	33	1	2	0
17	35	1	2	0
18	37	1	2	0
19	39	1	2	0
20	41	1	2	0
21	43	1	2	0
22	45	1	2	0
23	47	1	2	0
24	49	1	2	0
25	50	1	2	0

Special Cardinality

Join column 2

20 matching distinct value. Histogram gathered with "size 25"

9.2.0.7

EP	value	bkts	count	pop
1	1	1	40	0
2	2	1	40	0
3	3	1	40	0
5	4	2	80	1
6	5	1	40	0
7	6	1	40	0
8	7	1	40	0
10	8	2	80	1
11	9	1	40	0
12	10	1	40	0
13	11	1	40	0
15	12	2	80	1
16	13	1	40	0
17	14	1	40	0
18	15	1	40	0
20	16	2	80	1
21	17	1	40	0
22	18	1	40	0
23	19	1	40	0
25	20	2	80	1

Value	Count-1	Count-2	Pop-1	Pop-2	Pop-J	card
1	0	40	0	0	0	
2	2	40	0	0	0	
3	0	40	0	0	0	
4	2	80	0	1	1	80
5	0	40	0	0	0	
6	2	40	0	0	0	
7	0	40	0	0	0	
8	2	80	0	1	1	80
9	0	40	0	0	0	
10	2	40	0	0	0	
11	0	40	0	0	0	
12	2	80	0	1	1	80
13	0	40	0	0	0	
14	2	40	0	0	0	
15	0	40	0	0	0	
16	2	80	0	1	1	80
17	0	40	0	0	0	
18	2	40	0	0	0	
19	0	40	0	0	0	
20	2	80	0	1	1	80
22	2	0	0	0	0	
24	2	0	0	0	0	
50	2	0	0	0	0	
	24	560				269
						80
						749

Oracle 9.2 creates a height-balanced histogram for the first join column **and** for the second, creating the condition for the special cardinality: the lower of the high values is also a matching value

Special Cardinality

Join column 2

20 matching distinct value. Histogram gathered with "size 25"

10.1.0.5

EP	value	bkts	count	pop
50	1	1	50	1
100	2	1	50	1
150	3	1	50	1
200	4	1	50	1
250	5	1	50	1
300	6	1	50	1
350	7	1	50	1
400	8	1	50	1
450	9	1	50	1
500	10	1	50	1
550	11	1	50	1
600	12	1	50	1
650	13	1	50	1
700	14	1	50	1
750	15	1	50	1
800	16	1	50	1
850	17	1	50	1
900	18	1	50	1
950	19	1	50	1
1000	20	1	50	1

Value	Count-1	Count-2	Pop-1	Pop-2	Pop-J	card
1	0	50	0	1	1	50
2	0	50	0	1	1	50
3	1	50	0	1	1	50
4	0	50	0	1	1	50
5	1	50	0	1	1	50
6	0	50	0	1	1	50
7	1	50	0	1	1	50
8	0	50	0	1	1	50
9	1	50	0	1	1	50
10	0	50	0	1	1	50
11	1	50	0	1	1	50
12	0	50	0	1	1	50
13	1	50	0	1	1	50
14	0	50	0	1	1	50
15	1	50	0	1	1	50
16	0	50	0	1	1	50
17	1	50	0	1	1	50
18	0	50	0	1	1	50
19	1	50	0	1	1	50
20	0	50	0	1	1	50
21	1	0	0	0	0	
23	1	0	0	0	0	
50	1	0	0	0	0	
	2	0				0
						1000

Oracle 10.1.0.5 creates an in**correct** height-balanced histogram for the first join column and a frequency histogram for the second without the condition for the special cardinality.

Special Cardinality

Join column 2

20 matching distinct value. Histogram gathered with "size 25"

10.2.0.2

EP	value	bkts	count	pop
50	1	1	50	1
100	2	1	50	1
150	3	1	50	1
200	4	1	50	1
250	5	1	50	1
300	6	1	50	1
350	7	1	50	1
400	8	1	50	1
450	9	1	50	1
500	10	1	50	1
550	11	1	50	1
600	12	1	50	1
650	13	1	50	1
700	14	1	50	1
750	15	1	50	1
800	16	1	50	1
850	17	1	50	1
900	18	1	50	1
950	19	1	50	1
1000	20	1	50	1

Value	Count-1	Count-2	Pop-1	Pop-2	Pop-J	card
1	0	50	0	1	1	50
2	1	50	0	1	1	50
3	0	50	0	1	1	50
4	1	50	0	1	1	50
5	0	50	0	1	1	50
6	1	50	0	1	1	50
7	0	50	0	1	1	50
8	1	50	0	1	1	50
9	0	50	0	1	1	50
10	1	50	0	1	1	50
11	0	50	0	1	1	50
12	1	50	0	1	1	50
13	0	50	0	1	1	50
14	1	50	0	1	1	50
15	0	50	0	1	1	50
16	1	50	0	1	1	50
17	0	50	0	1	1	50
18	1	50	0	1	1	50
19	0	50	0	1	1	50
20	1	50	0	1	1	50
22	1	0	0	0	0	
24	1	0	0	0	0	
50	1	0	0	0	0	
	2	0				0
						50
						1050

Oracle 10.2.0.2 creates a **correct** height-balanced histogram for the first join column **and** a frequency histogram for the second with the condition for the special cardinality.

Revert to "standard" Formula

The algorithm reverts to the standard formula

$$\text{join cardinality} = \text{card}_{\mathbf{A}} * \text{card}_{\mathbf{B}} * 1/\max(\text{ndv}_{\mathbf{A}}, \text{ndv}_{\mathbf{B}})$$

if

- if LMV is NULL – no matching values (in the histograms)
- if HPV❶ is NULL – no popular values in CJH
- if LMV > HPV – all popular values < any matching value

❶ Highest Popular Value within the CJH

No Matching Values

1	13
2	13
3	13
4	13
5	13
6	688
7	13
8	13
9	13
10	13
11	13
12	13
13	13
14	13
15	13
16	13
17	13
18	13
19	13
20	13
21	13
22	13
23	13
24	13
25	13
26	
	1000

		0
2	13	169
3	688	8944
4	13	169
5	13	169
6	13	8944
7	13	169
8	13	169
9	13	169
10	13	169
11	13	169
12	13	169
13	13	169
14	13	169
15	13	169
16	13	169
17	13	169
18	13	169
19	13	169
20	13	169
21	13	169
22	13	169
23	13	169
24	13	169
25	13	169
26	13	0
	1000	21606

EP	value	bkts	count	pop
0	1	0	0	0
1	4	1	50	0
15	6	14	700	1
16	10	1	50	0
17	14	1	50	0
18	18	1	50	0
19	22	1	50	0
20	25	1	50	0
		20	1000	

EP	value	bkts	count	pop
0	2	0	0	0
14	3	14	700	1
15	7	1	50	0
16	11	1	50	0
17	15	1	50	0
18	19	1	50	0
19	23	1	50	0
20	26	1	50	0
		20	1000	

```
PLAN_TABLE_OUTPUT
-----------------------------------------
| Id  | Operation           | Name | Rows  |
-----------------------------------------
|   0 | SELECT STATEMENT    |      | 40000 |
|*  1 |  HASH JOIN          |      | 40000 |
|   2 |   TABLE ACCESS FULL| J1   |  1000 |
|   3 |   TABLE ACCESS FULL| J2   |  1000 |
-----------------------------------------
```

Gathering histograms of "size 20" for the distribution on the left produces the histograms on the right – with no matching values

No Popular Values in CJH

1	76
2	1
3	72
4	5
5	76
6	1
7	68
8	9
9	76
10	1
11	64
12	13
13	76
14	1
15	60
16	2
17	75
18	1
19	72
20	2
21	75
22	72
23	3
24	50
25	49
	1000

1	76	5776
2	1	1
3	72	5184
4	5	25
5	76	5776
6	1	1
7	68	4624
8	9	81
9	76	5776
10	1	1
11	64	4096
12	13	169
13	76	5776
14	1	1
15	60	3600
16	2	4
17	75	5625
18	1	1
19	72	5184
20	2	4
21	75	5625
22	72	5184
23	3	9
24	50	2500
25	49	2401
	1000	67424

Value	Count-1	Count-2	Pop-1	Pop-2	Pop-J
1	50	50	0	0	0
3	50	50	0	0	0
4	50	50	0	0	0
5	50	50	0	0	0
7	50	50	0	0	0
8	50	50	0	0	0
9	50	50	0	0	0
11	50	50	0	0	0
12	50	50	0	0	0
13	50	50	0	0	0
15	50	50	0	0	0
16	50	50	0	0	0
17	50	50	0	0	0
19	50	50	0	0	0
20	50	50	0	0	0
21	50	50	0	0	0
22	50	50	0	0	0
23	50	50	0	0	0
24	50	50	0	0	0
25	50	50	0	0	0
	1000	1000			

```
PLAN_TABLE_OUTPUT
-----------------------------------------
| Id  | Operation           | Name | Rows  |
-----------------------------------------
|   0 | SELECT STATEMENT    |      | 40000 |
|*  1 |  HASH JOIN          |      | 40000 |
|   2 |   TABLE ACCESS FULL | J1   |  1000 |
|   3 |   TABLE ACCESS FULL | J2   |  1000 |
-----------------------------------------
```

Gathering histograms of "size 20" for the distribution on the left produces the joint histograms on the right – with no popular values

All Popular Values < Any Matching Value

1	7
2	7
3	832
4	7
5	7
6	7
7	7
8	7
9	7
10	7
11	7
12	7
13	7
14	7
15	7
16	7
17	7
18	7
19	7
20	7
21	7
22	7
23	7
24	7
25	7
	1000

2	14
3	7
4	7
5	7
6	7
7	7
8	7
9	832
10	7
11	7
12	7
13	7
14	7
15	7
16	7
17	7
18	7
19	7
20	7
21	7
22	7
23	7
24	7
25	7
	1000

0
98
5824
49
49
49
49
49
49
5824
49
49
49
49
49
49
49
49
49
49
49
49
49
49
49
12775

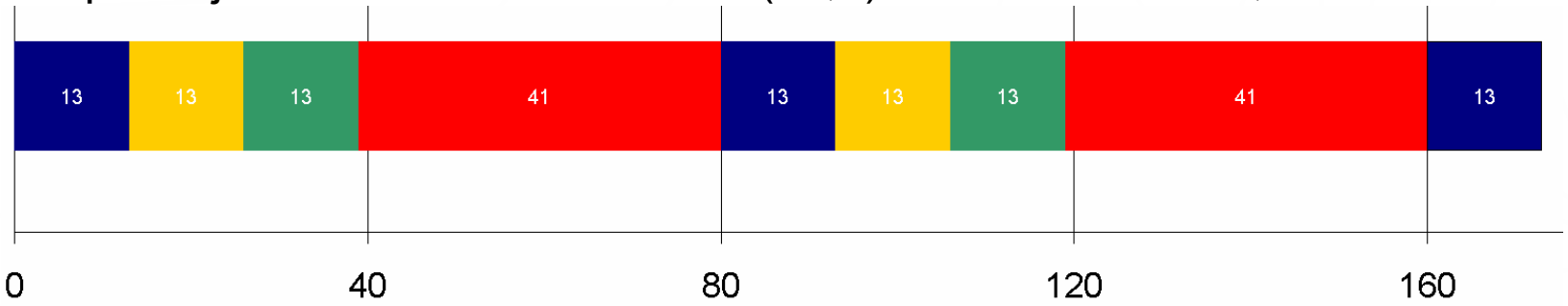
Value	Count-1	Count-2	Pop-1	Pop-2	Pop-J
1	0		0		
2		0		0	
3	800		1		1
4	50		0		0
8		50	0		0
9		800	1		1
11	50	50	0	0	0
18	50	50	0	0	0
25	50	50	0	0	0

```
PLAN_TABLE_OUTPUT
----------------------------------------
| Id  | Operation           | Name | Rows  |
----------------------------------------
|   0 | SELECT STATEMENT    |      | 40000 |
|*  1 |  HASH JOIN          |      | 40000 |
|   2 |   TABLE ACCESS FULL| J1   |  1000 |
|   3 |   TABLE ACCESS FULL| J2   |  1000 |
----------------------------------------
```

Gathering histograms of "size 20" for the distribution on the left produces the joint histogram on the right – with no popular values above the LMV

Upgrade Trap

```
create table j1 as select rownum id, mod(rownum-1,508)+1 n1
      from dual connect by level <= 10160;
update j1 set n1=n1+3 where mod(n1,4)=1 and id <= 3556; -- =7*508
update j1 set n1=n1+2 where mod(n1,4)=2 and id <= 3556;
update j1 set n1=n1+1 where mod(n1,4)=3 and id <= 3556;
```



9.2.0.7

EP	value
0	1
2	4
4	8
6	12
8	16

10.1.0.5

EP	value
0	1
1	4
2	5
3	8
4	9
5	12
6	13
7	16

10.2.0.2

EP	value
0	1
2	4
4	8
6	12
8	16

Summary

- Little Benefit for properly designed FK joins unless significant "symmetrical" skew
- Frequency Histogram provide the best source for improved cardinality estimates – except for "halving"
- Height-balanced Histograms in particular can be volatile.
  Use of skewonly can avoid problem scenarios
- Histograms are "high maintenance"

References

Lewis, Jonathan. *Cost-Based Oracle: Fundamentals*. Apress. ISBN: 1590596366.

http://www.ixora.com.au/newsletter/2001_04.htm

44961.1 Statistics Gathering: Frequency and Strategy Guidelines

72539.1 Interpreting Histogram Information

100229.1 Indexes - Selectivity

175258.1 How to Compute Statistics on Partitioned Tables and Indexes

1031826.6 Histograms: An Overview

Wolfgang Breitling

breitliw@centrexcc.com

Centrex Consulting Corp.

www.centrexcc.com