



## Het juiste executieplan of de juiste statistieken

Bram van der Vos  
[www.axisintoict.nl](http://www.axisintoict.nl)  
[ict@axisinto.nl](mailto:ict@axisinto.nl)



# AXIS INTO ICT Agenda



- De Optimizer
- Statistieken
- Selectiviteit en impact op executieplan
- Demo's

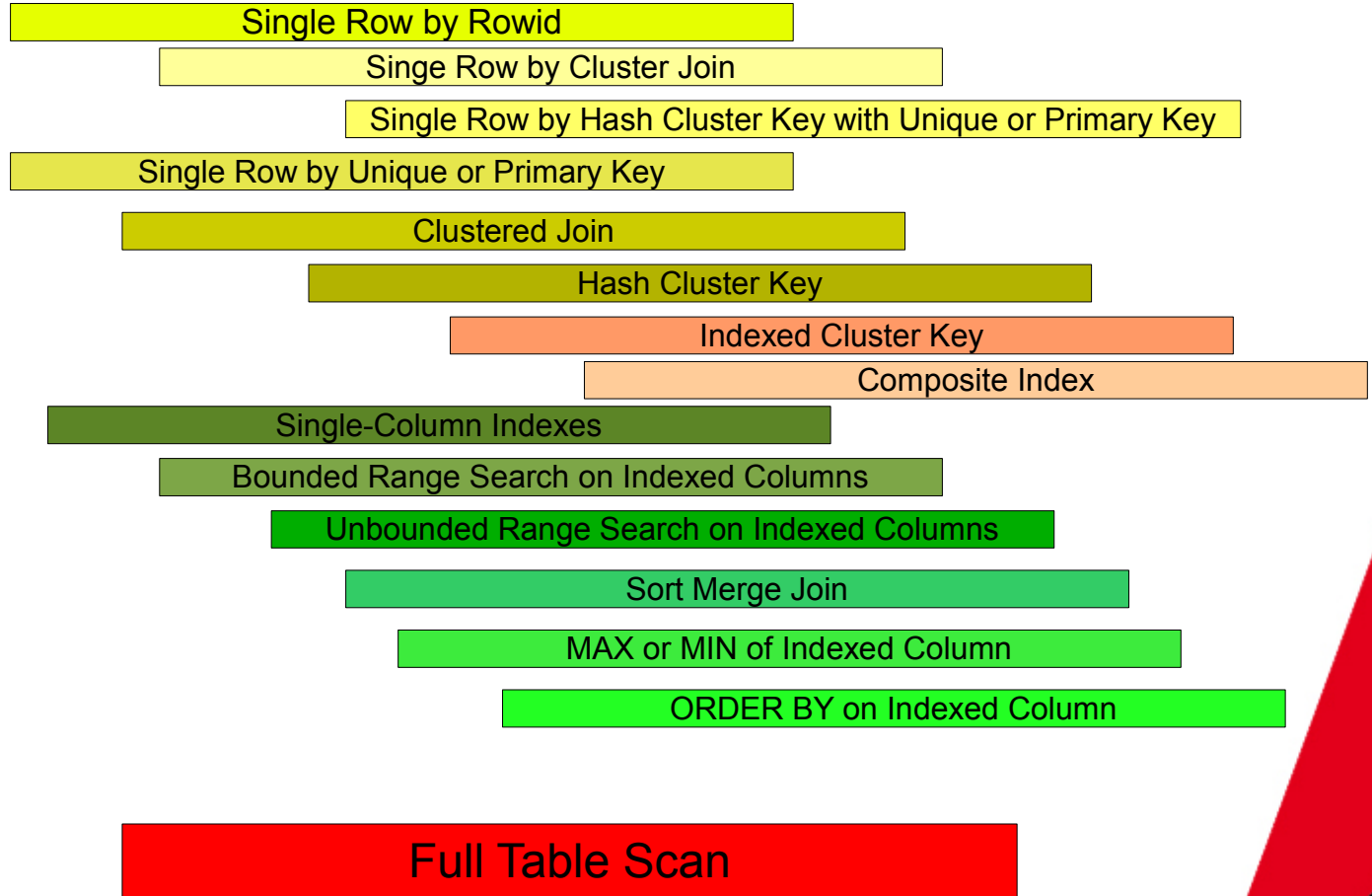
Statistics for measurements of: \_bioassay.tif

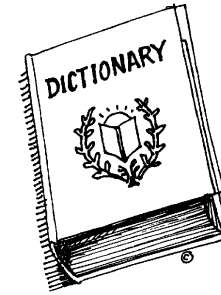
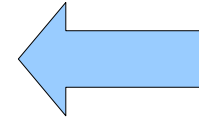
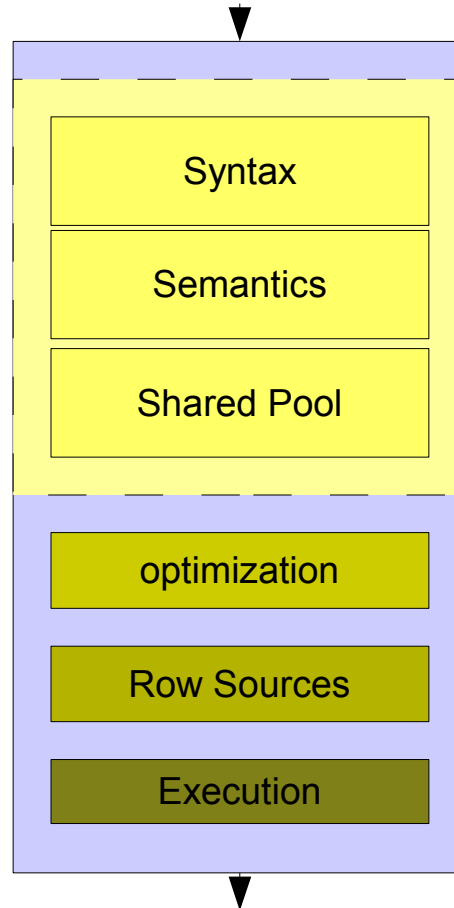
	Length/Per	Area	Angle	Points	Centroid X	Centroid Y	Width	Height
Mean	115.7			2.8	197.4	164.4		
Median	90.0			2.0	200.0	122.0		
Min	67.9			2.0	0.0	0.0		
Max	232.0			5.0	337.0	360.0		
Range	164.1			3.0	337.0	360.0		
PopVariance	3533.9			1.7	10140.8	12425.4		
Variance	3533.9			1.7	10140.8	12425.4		
PopStDev	59.4			1.3	100.7	111.5		
St. Dev.	66.5			1.5	103.5	114.5		
Std.Error	29.7			0.8	23.7	26.3		
Coef.var.	57.5			54.5	52.4	63.7		
Skew	1.3			1.2	-0.2	0.6		
Kurtosis	3.0			2.3	1.8	2.0		
Sum	578.4			11.0	3751.0	3124.0		
# Samples	5	0	0	4	19	19	0	0



- 7.0 Introductie CBO (1992)
  - 7.3 gangbaar
- 10g RBO desupported maar beschikbaar
- 11g RULE niet meer bruikbaar

We werken dus altijd Cost Based

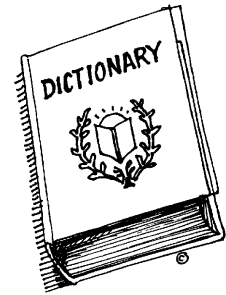
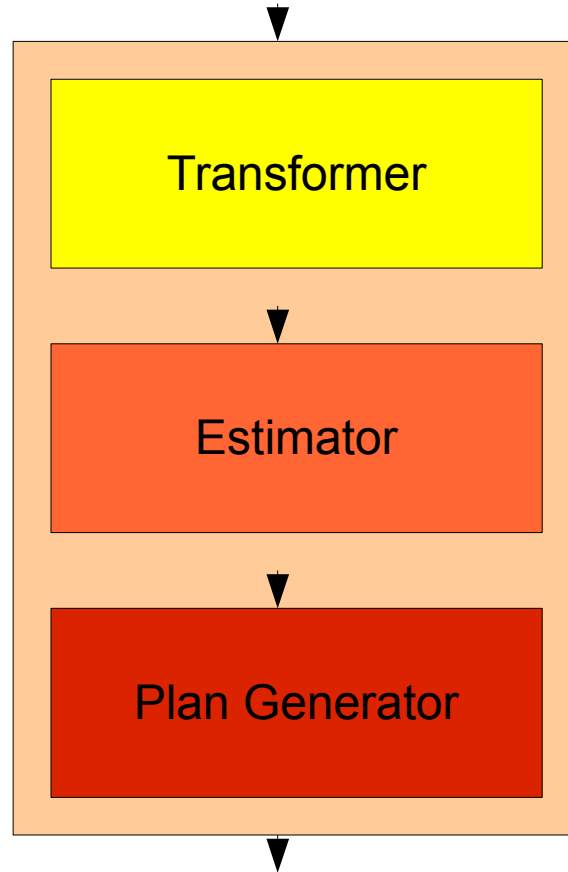






AXIS INTO ICT

# Cost – Based Optimizer





AXIS<sup>INTO</sup>  
ICT

# Transformations



- OR expansion
- SubQuery Unnesting
- View Merging
- Predicate Pushing
- Transitivity



# AXIS<sup>INTO</sup> ICT Transformation Examples



```
select *  
from <TABLE>  
where c1=value1  
or c1 is value2
```

```
select * From <TABLE>  
where C1=value2  
UNION ALL  
select * from <TABLE>  
where c2 value2 and C1 <> value1
```

```
select * from emp, dept  
where emp.deptno = 20  
and emp.deptno=dept.deptno
```

```
select * from emp, dept  
where emp.deptno = 20  
and dept.deptno = 20  
and emp.deptno=dept.deptno
```



- Selectivity
  - % van op te halen rijen t.o.v. het totaal
- Cardinality
  - # rijen
- Cost

$$\frac{\text{Single Block} + \text{Multi Block} + \text{Cpu Cost}}{\text{sreadtime}}$$

Selectivity en Cardinality bepalen Cost en dus het executie plan



- Join Order
- Join methode
- Query Subplans
- Beperk het aantal permutaties
- Acces paths

## Execution Plan

Plan hash value: 2385808155

Id	Operation	Name	Rows	Bytes	Cost (%CPU)	Time
0	SELECT STATEMENT		1	26	2 (<0)	00:00:01
1	NESTED LOOPS		1	26	2 (<0)	00:00:01
2	TABLE ACCESS BY INDEX ROWID	EMP	1	13	1 (<0)	00:00:01
* 3	INDEX UNIQUE SCAN	PK_EMP	1		0 (<0)	00:00:01
4	TABLE ACCESS BY INDEX ROWID	DEPT	4	52	1 (<0)	00:00:01
* 5	INDEX UNIQUE SCAN	PK_DEPT	1		0 (<0)	00:00:01



- DBMS\_XPLAN (sinds 9i, verbeterd in 11g)
  - display()
  - display\_cursor()
  - display\_awr()

```
DBMS_XPLAN.DISPLAY_CURSOR(  
  sql_id          IN  VARCHAR2  DEFAULT  NULL,  
  child_number   IN  NUMBER     DEFAULT  NULL,  
  format         IN  VARCHAR2   DEFAULT  'TYPICAL');
```

- iostat & memstat
- Grafische ondersteuning in Enterprise manager7



- *De* input voor de Cost based optimizer
- Door de DBA te genereren
  - Object statistieken
    - DBA\_[TAB|IND]\_STATISTICS
    - DBA\_[TABLES|INDEXES]
  - Dictionary statistieken
  - Systeem statistieken



- Hardware karakteristieken
  - IO
    - Single block
    - Multi block
  - CPU cycles

```
ioseektim = 10ms  
iotrfspeed = 4096 bytes/ms  
cpuspeednw = gathered value,
```

- Workload vs Noworkload Statistics

```
dbms_stats.gather_system_stats('start'/'stop')  
VS  
dbms_stats.gather_system_stats()
```



- Delete system stats
- Select pname, pval1 from sys.aux\_stats\$
- gather\_system\_stats()
- Select pname, pval1 from sys.aux\_stats\$



- Relevante statistieken
  - # rijen
  - # Blokken
  - #lege blokken
  - Avg\_space
  - Avg\_row\_len
  - Distinct Keys
  - Clustering factor
  - Branch Level



- Analyze table/ `dbms_utility`
- `DBMS_STATS`
  - Automatisch
    - `GATHER_STATS_JOB` (10g)
    - `DBMS_AUTO_TASK_ADMIN` (11g)
  - Handmatig
    - Eigen job
    - Command line



```
BEGIN
  DBMS_STATS.GATHER_TABLE_STATS (
    OWNNAME    => 'OE',
    TABNAME    => 'INVENTORIES');
```

Maar wat betekent dit ???

```
DBMS_STATS.GATHER_TABLE_STATS (
  ownname          VARCHAR2,
  tablename       VARCHAR2,
  partname        VARCHAR2,
  estimate_percent NUMBER,
  block_sample   BOOLEAN,
  method_opt     VARCHAR2,
  degree         NUMBER,
  granularity    VARCHAR2,
  cascade        BOOLEAN,
  stattab        VARCHAR2,
  statid         VARCHAR2,
  statown       VARCHAR2,
  no_invalidate BOOLEAN,
  stattype       VARCHAR2,
  force          BOOLEAN)
```



auto_stats_target	Auto
Cascade	NULL
Degree	32768
Estimate_Percentage	DBMS_STATS.AUTO_SAMPLE_SIZE
Method_opt	FOR ALL COLUMNS SIZE AUTO
No_invalidate	DBMS_STATS.AUTO_INVALIDATE
Granularity	Auto
Publish	True
Incremental	False
Stale_percentage	10%

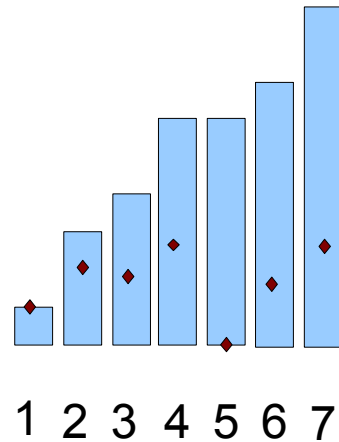


- Aanname van oracle: Data is uniform verdeeld
- $1/\#$  verschillende waarden
- Is dat zo ?
  - Status
  - Geslacht
  - order\_id bij webwinkel
  - Merk/ Kleur van auto
- Wat is het effect?





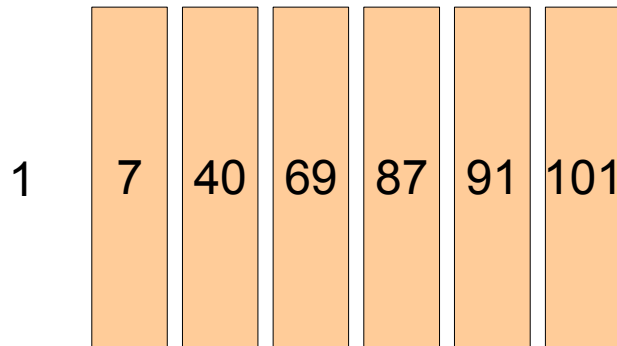
- Frequency based histograms:



- Beperkt aantal waarden
- # voorkomens per waarde



- Height balanced histograms:
  - Veel waarden
  - Beperkt aantal buckets
  - Meerdere waarden per bucket
  - Selectiviteit afhankelijk #buckets





- Alleen histogrammen als het nodig is
  - Join conditie
  - Where clause
  - Non-unique
- Gebruik alleen histogrammen als het iets toevoegt
- Is je aantal waarden  $\leq 254$  gebruik frequency based
- Height based: denk na over aantal buckets
- Histogram waarde max 32 bytes



- Onderdeel van `dbms_stats.gather*_stats`
- `method_opt`
  - FOR ALL [ INDEXED| HIDDEN] columns
    - SIZE <n>
    - REPEAT
    - SKEWONLY
    - AUTO
- Default
  - FOR ALL COLUMNS SIZE AUTO
- Opvallend : Default: All Columns



- Geen statistieken
- Statistieken maar geen histogrammen
- Statistieken en Frequency Based Histogrammen
- Statistieken en Height balanced histogrammen
  - Auto
  - Manual

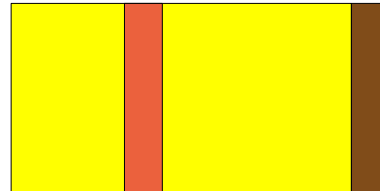


- `Lower(ename)='jansen'` of `ename='jansen'`
  - Selectiviteit ?
- `to_number (value) =12` of `value=12`
  - Selectiviteit ?
- Aanne: `Function (Column) =1%`



```
DBMS_STATS.CREATE_EXTENDED_STATS (  
  ownname      VARCHAR2,  
  tabname      VARCHAR2,  
  extension    VARCHAR2)  
RETURN VARCHAR2;
```

Column1



sys\_<key>

- Virtual column infrastructuur
- dba\_stat\_extensions



## Multi column statistics

- kolommen hebben relaties
  - Merk → type
  - Geslacht → aanhef
  - Woonplaats → verzekeringsopslag
  
- Vraag: hoe gaan we hiermee om??



- Waarom doet Oracle het niet zelf.
  - `Optimizer_dynamic_sampling` 0-10
- Wanneer
  - Als jij het wil
  - Statistieken ontbreken
  - 'Poor estimates' extended statistics
  - Maar wanneer gebeurt het?



## Dynamic Sampling (2)



0	Geen sampling
1	Geen statistieken, geen indexen, grote table
2	Als ten minste 1 table geen statistieken heeft
3	Een of meer expressies
4+	Complexe constructies aanwezig

- Level afhankelijk van de release
- Kost ook performance



selectivity	
=	1%
!=	5%
other	5%
Row length	20
# index leaf blocks	25
#distinct values	100
Table Cardinality	100



# Slim tonen van je executie plan locatie ????



- DBMS\_XPLAN (sinds 9i, verbeterd in 11g)
- Wat krijg je (format parameter)
  - E-rows vs A- Rows ← estimator
  - Display/display\_cursor/display\_awr
  - v\$SQL\_PLAN ?



DBMS\_STATS

- DBMS\_XPLAN
- dbms\_monitor
- dbms\_session
- +gather\_optimizer\_statistics
- statistics\_level
- optimizer\_dynamic\_sampling
- Event 10053



- Alleen focussen op het executieplan is gevaarlijk
- Statistieken zijn de start van een goed executieplan
- Gebruik de maintenance job
- Denk aan de juiste preferences
- Wees verdacht op de effecten van dynamic sampling

