Prague, Czech Republic | November 2014

25 years of missed opportunities? SQL Tuning Revisited

SOFTWARE ENGINEERING GmbH

Platform: z/OS DB2



AGENDA

- 1. Tuning SQL
 - How we have always done it
 - Single SQL, Package, Application...
 - Year 2004 AP comparisons and simulation
- Tuning SQL Revisited A new methodology
- Harvesting the low hanging fruit



Tuning SQL – How we have always done it

- Get an SQL from development
- **EXPLAIN** it
- Tune it if required
- Stage to next Level (Dev -> QA, QA -> Prod)
- Fire fight

IDOG DDZ Tech Come

Prague, Czech Republic | November 2014



Single SQL, Package, Application...

- Get an SQL, Package or list of Packages from development
- Fight for (and against!) Dynamic SQL
- EXPLAIN them all
- See if any have gone belly up
- Tune it if required and if you have the time
- Stage to next Level (Dev -> QA, QA -> Prod)
- Fire fight



Tuning SQL - Year 2004

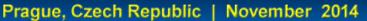
- Get an SQL, Package or list of Packages from development
- Propagate Production statistics down the environment chain (Prod -> QA, Prod -> Dev)
- Simulate Hardware, ZPARMS, and BUFFERPOOLS
- Fight for (and against!) Dynamic SQL
- EXPLAIN them all
- Compare with existing Access Paths Reject any that have got worse
- Tune it if required and if you have the time
- Stage to next Level (Dev -> QA, QA -> Prod)
- Fire fight



Tuning SQL Revisited

- Get *all* Dynamic and Static SQL running in the Plex
- Propagate Production statistics down the environment chain (Prod -> QA, Prod -> Dev)
- Simulate Hardware, ZPARMS, and BUFFERPOOLS
- EXPLAIN them all
- Compare with existing Access Paths Tune any that have got worse
 - Pick the "low hanging fruit"
- Stage to next Level (Dev -> QA, QA -> Prod)







Tuning SQL Revisited

So how to get there?

- 1. Collect as much data as you can
- 2. Store it in a Data Warehouse
- 3. Analyze it
- 4. Take Actions!

Collect as much data as you can

- How many resources do you spend on capturing DB2 SQL workload and its metrics?
- There seems to be out-of-the-box metrics delivered by DB2, but does it give me all the data I need, when I need it?
- How does the smarter database, how does DB2 10, or 11 for z/OS deal with it?...

Brague Czech Benublie I Nevem

Prague, Czech Republic | November 2014



Collect as much data as you can

- DB2 10 Monitoring Enhancements and Changes:
 - Statement Level Statistics
 - Enhanced messages and traces to capture statement level information
 - Statement information in real-time
 - STMT_ID unique statement identifier assigned when statement first inserted into DSC
 - Statement type static or dynamic
 - Bind TS 10 byte TS when stmt was bound, or prepared
 - Statement level execution statistics (per execution)
 - New Monitor class 29 for statement detail level monitoring
 - Monitor Class 29 (overhead is ~1-3%)
 - New for statement level detail



What's exactly new since DB2 10:



- IFCID 316 was enhanced to externalize the data from the Dynamic Statement Cache (DSC) when a flushing situation occurs (LRU, RUNSTATs, ALTER, DROP, REVOKE, ...)
 - NO DATA LOSS



- New IFCIDs 400* and 401 additionally EDM pool data let's call it the Static Statement Cache
 - Memory resident storage of static SQL statements
 - Like with the enhanced 316, data is externalized when the EDM pool is full.
 - NO DATA LOSS



DSC and EDM provide detailed workload insights:

- SQL text
- Statement ID
- Date/time
- **Current status**
- Resource consumption
- Identification/environmental data



DB2 10 also introduced some additional information from the DSC trace we all know today:

- Wait time accumulation for
 - Latch requests
 - Page latches
 - **Drain locks**
 - Drains during waits for claims to be released
 - Log writers

- Date and time in store clock format for Stmt insertion and update (along with internal format)
- Number of times that
 - a RID list overflowed because of
 - storage shortage
 - # of RIDs exceeded internal limit(s)
 - a RID list append for a hybrid join interrupted
 - because of RID pool storage shortage
 - # of RIDs exceeded internal limit(s)
 - a RID list retrieval failed for multiple IX access. The result of IX AND/ORing could not be determined





Prague, Czech Republic | November 2014

Collect as much data as you can

Counters # executions of the statement. For a cursor statement, this is the # of opens. # of synchronous buffer reads performed for statement. # of getpage operations performed for statement. # of rows examined for statement. # of rows processed for statement - for example, the # of rows returned for a select, or the number of rows affected by an insert, update, or delete. # of sorts performed for statement. # of index scans performed for statement. # of tablespace scans performed for statement. * # of parallel groups created for statement. # of synchronous buffer write operations performed for statement. # of times that rid list retrieval for multiple index access was not done because db2 could determine the outcome of index anding or oring*.

O Counters # of times that a rid list was not used because # of rid exceeded one or more internal db2 limits, and the # of rid blocks

EXCEEDED THE VALUE OF SUBSYSTEM PARAMETER MAXTEMPS. FILD # OF TIMES THAT A RID LIST WAS NOT USED BECAUSE NOT ENOUGH STORAGE WAS AVAILABLE TO HOLD

THE RID LIST, OR WORK FILE STORAGE OR RESOURCES WERE NOT. WAILABLE. # OF TIMES THAT IS NOT EVER LOWED TO A WORK FILE BECAUSE NO RID POOL STORAGE

WAS AVAILABLE TO HOLD THE LIST OF RIDS. # OF TIMES ATHAT RID LIST OVERFLOWED TO A WORK FILE BECAUSE THE NUMBER OF RIDS EXCEEDED ONE OR MORE INTERNAL

LIMITS. # OF TIMES THAT APPENDING TO A RID LIST FOR A HYBRID JOIN WAS INTERRUPTED BECAUSE TO RID POOL STORAGE WAS AVAILABLE TO HOLD THE LIST OF RIDS.

OF TIMES THAT APPENDING TO A RID LIST FOR A HYBRID JOIN WAS INTERRUPTED BECAUSE TO RIDS EXCEEDED ONE OR MORE INTERNAL LIMITS.

TIMINGS ACCUMULATED CPU TIME. THIS VALUE INCLUDES CPU TIME THAT IS CONSUMED ON AN IBM SPECIALTY ENGINE. ACCUMULATED ELAPSED TIME USED FOR STATEMENT. ACCUMULATED WAIT TIME FOR LATCH REQUESTS*. ACCUMULATED WAIT TIME FOR DRAINS DURING WAITS FOR CLAIMS TO BE RELEASED. ACCUMULATED WAIT TIME FOR LOCK REQUESTS ACCUMULATED WAIT TIME FOR SYNCHRONOUS I/O. ACCUMULATED WAIT TIME FOR LOCK REQUESTS, ACCUMULATED WAIT TIME FOR SYNCHRONOUS EXECUTION UNIT SWITCH. ACCUMULATED WAIT TIME FOR GLOBAL LOCKS. ACCUMULATED WAIT TIME FOR READ ACTIVITY THAT IS DONE BY ANOTHER THREAD.

IDENTIFICATION DATA SHARING MEMBER THAT CACHED THE SQL STATEMENT*. PROGRAM NAME, PROGRAM NAME IS THE NAME OF THE PACKAGE OR

DBRM THAT PERFORMED THE PREPARE/SQL. PRECOMPILER LINE NUMBER FOR THE PREPARE STATEMENT OR SQL STATEMENT. TRANSACTION NAME. THIS VALUE IS PROVIDED DURING RRS SIGNON OR RE-SIGNON. WORKSTATION NAME*. THIS VALUE IS PROVIDED DURING RRS SIGNON OR RE-SIGNON. USER ID. USER ID. USER ID IS THE PRIMARY AUTH, ID OF THE USER WHO DID THE INITIAL PREPARE. USER GROUP, USER GROUP IS THE CURRENT SQLID OF THE USER WHO DID THE INITIAL PREPARE. USER FROVIDED IDENTIFICATION STRING.

ENVIRONMENTAL REFERENCED TABLE NAME. FOR STATEMENTS THAT REFERENCE HORE THAN ONE TABLE, ONLY THE NAME OF THE FIRST TABLE THAT IS REFERENCED IS REPORTED. (ALL REFERENCED OBJECTS ARE STORED IN THE WLI DATA MODEL) LITERAL REPLACEMENT FLAG* CURRENT SCHEMA. QUALIFIER THAT IS USED FOR UNQUALIFIED TABLE NAMES. BIND OPTIONS: ISOLATION, CURRENTDATA, AND DYNAMICRULES. SPECIAL REGISTER VALUES: CURRENT DEGREE, CURRENT RULES, AND CURRENT PRECISION. WHETHER THE STATEMENT CURSOR IS A HELD CURSOR. TIMESTAMP WHEN STATISTICS COLLECTION BEGAN. DATA COLLECTION BEGINS WHEN A TRACE FOR IFCID 318 IS STARTED. DATE AND TIME WHEN THE STATEMENT WAS INSERTED INTO THE CACHE IN STORE CLOCK FORMAT. DATE AND TIME WHEN THE STATEMENT WAS UPDATED, IN INTERNAL FORMAT.



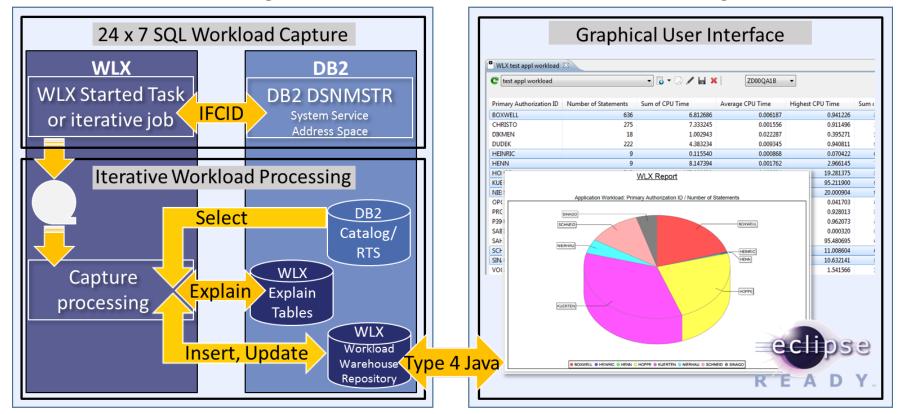
Store it in a SQL Workload Data Warehouse

Captures the hard to get SQLs, even the ones that disappear ...

Mainframe Engine

DUG

Workstation Engine





Store it in a SQL Workload Data Warehouse

Capture the data e.g. using a STC:

Run a started task 24x7 to catch all the IFCIDs that DB2 will be throwing and store the data.

Process the workload:

Externalize and process the data, such as every 60 min:

- customizable (e.g. 30 180 minutes)
- allow Ad hoc data refresh triggered via operator command for the started task (MODIFY)
- capture the SQL Text at trace time
- gather additional catalog and RTS data
- add explain data if needed

Store it in a SQL Workload Data Warehouse

Use a GUI front end, preferably Eclipse:

Exploit and integrate into Eclipse based GUI front ends

- GUIs can come as a Plug-in for
 - IBM Rational
 - IBM Data Studio
 - Eclipse native
- Existing DB2 connections are used to connect to the mainframe
- Interactive dialogs allow complex and powerful analysis
- Export features can create PDF reports and allow MS Excel hand over
- Additional plug-ins interface with other tools, such as **SQL PerformanceExpert** (SPX) and **Bind ImpactExpert** (BIX)

Store it in a SQL Workload Data Warehouse

Make the SQL Workload Warehouse Repository a set of DB2 tables that can also be created in LUW on a x86 server (E.g. DB2) Express-C).

If this is done then you can simply unload from the z/OS DB2 tables and then load the LUW Tables directly from within the GUI which enables you to run all the analytics queries "locally".

This can obviously save a lot of space on the z/OS side!

And remember that all of the Type 4 JAVA SQL is ZiiP eligible!

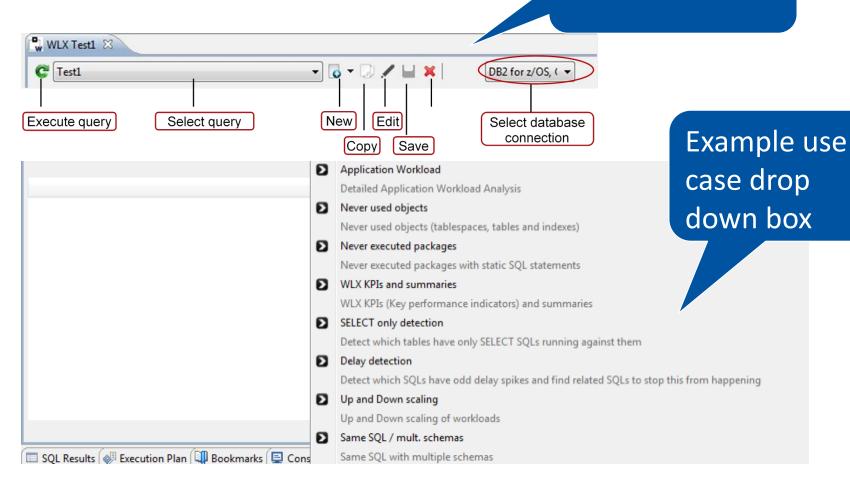


Prague, Czech Republic | November 2014





GUI features – button overview





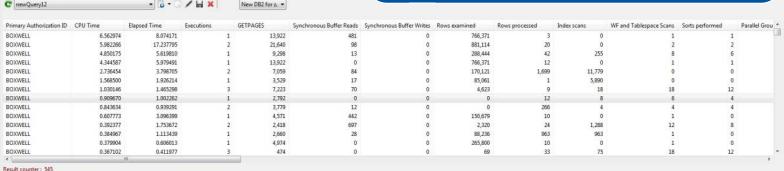


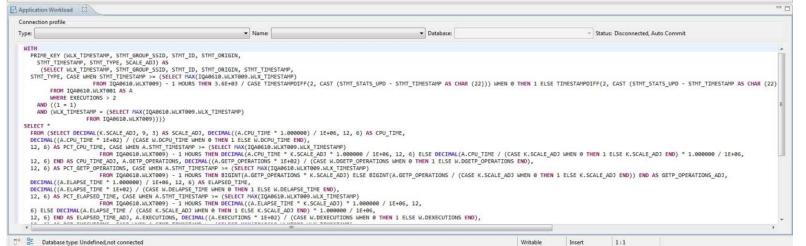


File Edit Navigate Search Project Run Window Help

WLX newQuery12 🖾 📮 WLX newQuery1

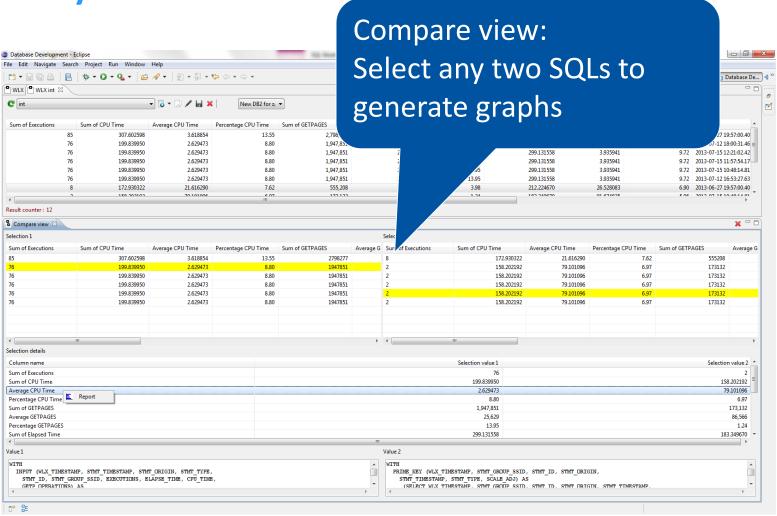




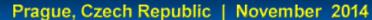


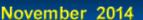




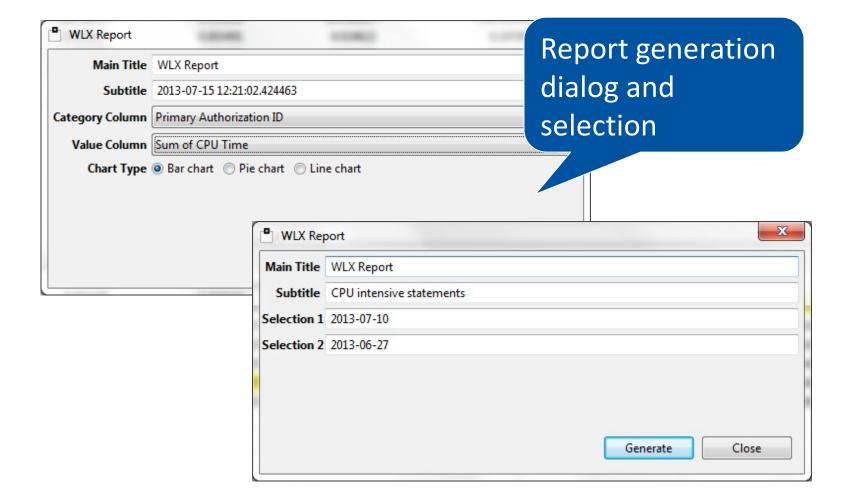


#IDUG









OK, so assuming you have all the data where shall we begin???

How about Intensive CPU?



The definition is simply the percentage of the CPU for a given period of time for an SQL. Here is two days of data:

CPU Intensive		- 0 - 0 / H ×	ZD00QA	ZD00QA1B ▼	
Sum of Executions	Sum of CPU Time	Average CPU Time	Percentage CPU Time	Sum of GETPAGES	
20,234	2,654.196293	0.131175	61.21	329,091,770	
39	412.657181	10.580953	9.52	1,339,000	
1,338	227.825821	0.170273	5.25	7,587,830	
13,636	158.568243	0.011628	3.66	5,698,318	
24,912	155.575122	0.006244	3.59	2,334,284	
1,541	121.625819	0.078926	2.81	7,232,26	
17,204	117.332832	0.006820	2.71	1,669,01	
1,000	110.225396	0.110225	2.54	5,956,295	
3,568	103.097601	0.028895	2.38	5,493,278	
2,074	97.765814	0.047138	2.25	6,457,211	
31,256	89.828375	0.002873	2.07	10,608,667	
(III				

As you can see one SQL executed over 20,000 times and soaked up the lion's share of the machine! Drilling down reveals the SQL:

WHERE KA_BEARB_ZK <> '1' AND KA_BEARB_ZK <> 'L' WITH CS

Harvesting the low hanging fruit

OK, so assuming you have all the data where shall we begin???

- How about Intensive CPU?
- 2. What about by Application?



The Application definition is simply the Primary Authorization Id or the Collection/Package. Here is one snapshot of data:

Number of Statements	Sum of CPU Time	Average CPU Time	Highest CPU Time	Sum of Elapsed Time	Average Elapsed Time
41	1,145.929112	28.648227	673.160930	2,171.410912	54.285272
6	122.393241	20.398873	38.379085	674.223872	112.370645

The average CPU is pretty high and the "highest" is very high! Drilling on down:

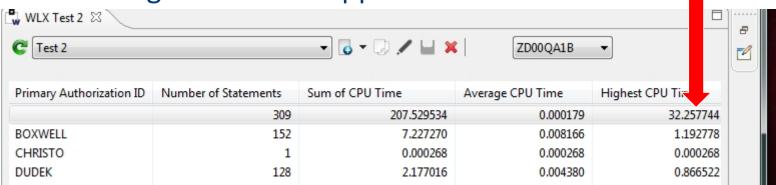
CPU Time	Elapsed Time	Executions	GETPAGES	Synchronous Buffer Reads
673.160930	1,076.620276	1	9,731,686	42,481

Only one execution for this guy and the SQL was a pretty horrible three table join with about 20 predicates.

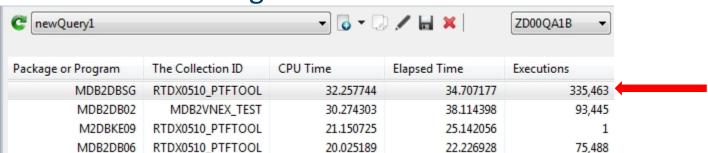


Harvesting the low hanging fruit

Here is a high CPU Static application:



Drill down to Package level:





Drill down to SQL level:

```
SELECT CHAR ( SUBSTR ( DIGITS ( YEAR ( STATSTIME ) ) , 9 , 2 ) CONCAT
SUBSTR ( DIGITS ( DAYOFYEAR ( STATSTIME ) ) , 8 , 3 ) , 5 ) INTO : H
FROM SE_STOGROUP
WHERE NAME = : H
WITH UR
```

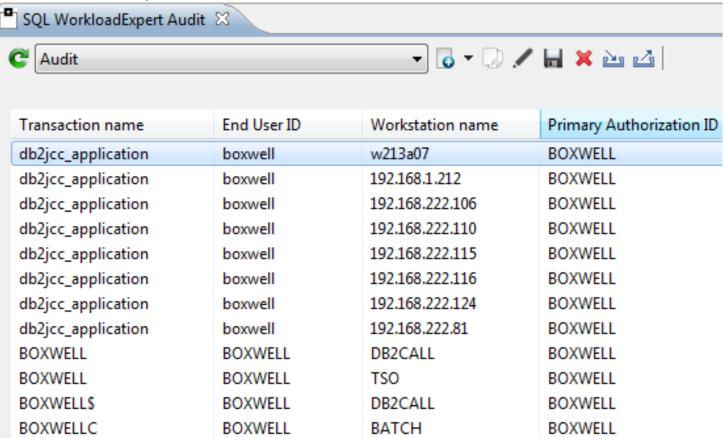
For every physical object a select from SYSSTOGROUP... Rewrite to a LEFT OUTER JOIN and the problem is solved!

OK, so assuming you have all the data where shall we begin???

- 1. How about Intensive CPU?
- 2. What about by Application?
- 3. Auditing???



Choose how you like to find out who did what and when...



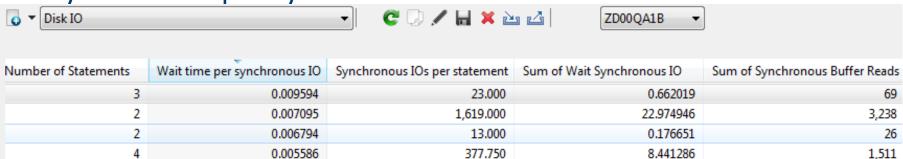


OK, so assuming you have all the data where shall we begin???

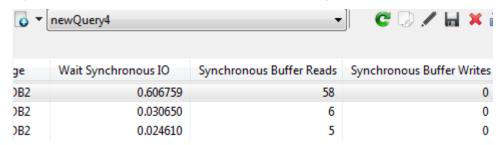
- How about Intensive CPU? 1.
- What about by Application?
- 3. Auditing?
- Disk I/O Performance? 4.



Any Wait time per synchronous IO over 0.002 seconds is bad:



For OLTP transactions any with more than one Synchronous IOs per statement is "sub optimal"! Drill down shows details:



OK, so assuming you have all the data where shall we begin???

- How about Intensive CPU? 1.
- 2. What about by Application?
- 3. Auditing?
- Disk I/O Performance?
- 5. Up and Down Scaling?



Harvesting the low hanging fruit

Up and Down scaling is all about getting a "level playing field" when looking at the cache data. Simply displaying the data for SQLs that have been in the cache a week next to SQLs that have been in the cache for only 10 minutes is a bit biased!

CPU Time	Percentage CPU Time	CPU time adjusted	GETPAGES	Percentage GETPAGES	GETPAGES adjusted
29.231328	1.857795	1.238178	681,568	4.427895	28,869
23.371722	1.485388	0.989977	593,016	3.852606	25,118
16.904098	1.074338	0.604954	446,936	2.903578	15,994
174.386924	11.083150	0.558840	1,622,158	10.538562	5,198

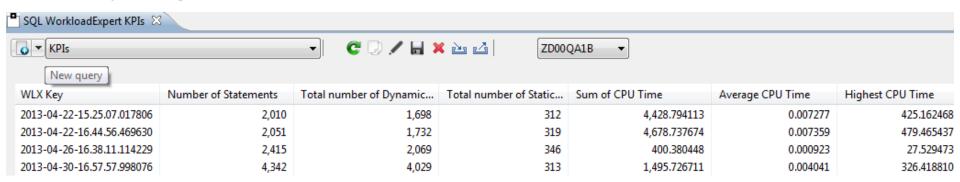
Here you can easily see the "normal Top 10" values and the "adjusted" values. Your "Top 10" suddenly contains completely new candidates that you were *never* even aware of!

OK, so assuming you have all the data where shall we begin???

- 1. How about Intensive CPU?
- 2. What about by Application?
- 3. Auditing?
- 4. Disk I/O Performance?
- 5. Up and Down Scaling?
- 6. KPIs for your Enterprise?

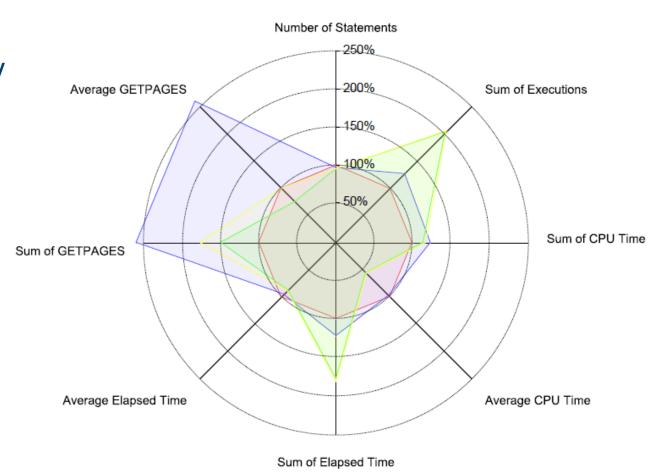


Naturally all this data also lets you build up a great set of KPIs to keep track of how many, what type, and how CPU & I/O hungry everything is:



Not just CPU but GetPages etc. are also available.

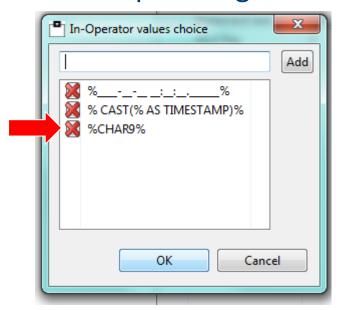
Then you can play with radar charts:





- How about Intensive CPU? 1.
- 2. What about by Application?
- 3. Auditing?
- Disk I/O Performance? 4.
- 5. Up and Down Scaling?
- 6. KPIs for your Enterprise?
- Searching for SQL? 7.

These days it is sometimes pretty neat to see what text is in the SQL. Currently two things spring to mind, first is CHAR9 usage and then dodgy Timestamp casting.





And then...



Drill down to get a better view

```
SELECT COUNT_BETTER_PROG, COUNT_WORSE_PROG, COUNT_BETTER_STMT,

COUNT_WORSE_STMT
FROM IQA0610.BAIM_RUNIDS
WHERE RUN_MODE IN ('DYNA') AND (RUNID = '2014-04-22 14:27:18.84815')
ORDER BY RUNID DESC
```

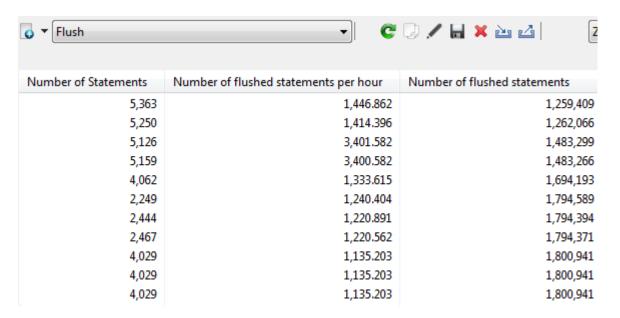


Harvesting the low hanging fruit

- How about Intensive CPU? 1.
- 2. What about by Application?
- 3. Auditing?
- Disk I/O Performance? 4.
- 5. Up and Down Scaling?
- 6. KPIs for your Enterprise?
- 7. Searching for SQL?
- Flushed with success? 8.



If you are catching and storing all the SQL then you can easily see how good the size and performance of your cache is:



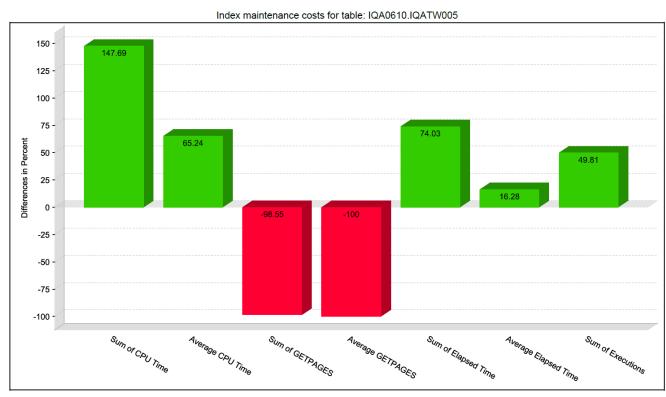
Rule of thumb is to make the EDMSTMTC as big as it can be! 200,000 is a good start!

- How about Intensive CPU? 1.
- 2. What about by Application?
- 3. Auditing?
- Disk I/O Performance? 4.
- 5. Up and Down Scaling?
- 6. KPIs for your Enterprise?
- 7. Searching for SQL?
- Flushed with success? 8.
- **Index Comparison?** 9.



Compare KPIs before and after Index creation. Especially twinned with Virtual Index

usage this is a real winner!
Did that new Index help or hinder my DB2?





Harvesting the low hanging fruit

- How about Intensive CPU? 1.
- 2. What about by Application?
- 3. Auditing?
- Disk I/O Performance? 4.
- 5. Up and Down Scaling?
- KPIs for your Enterprise? 6.
- 7. Searching for SQL?
- Flushed with success? 8.
- 9. **Index Comparison?**
- Miscellaneous other possibilities... 10.

Again, if you are catching and storing all the SQL then you can do:

- Sub-system loading checking
- Delay detection
- Object Quiet Times Alter & Reorg
- Find all non-executed Packages Free
- Never executed SQLs within executed Packages Delete
- Never referenced Tables/Indexes Drop
- Select only usage of objects Locksize tuning

Why stop with just these IFCIDs? If you have a technology for high speed catching and writing why not expand it to handle:

172 – Deadlocks

196 – Timeouts

337 – Lock Escalations

359 – Index page Splits

366/376 – BIF Usage



Harvesting the low hanging fruit

Some real world numbers to amaze and astound:

- On one member of a Data Sharing group the SQLs that normally ran fast were running 45% slower than on other members. After using a WLX it was discovered that this member had orders of magnitude more updates – Increase Log Buffer, Active Log, and Archive Log sizes then redirect some traffic. Et Voila!
- 450,000,000 Get pages per hour saved! -- New index created which gave a knock on performance boost effect to the whole DB2 sub-system
- CPU Reduction from 17,111 seconds per hour to 16 seconds per hour! One "Bad Guy" query tuned
- Elapsed time from 30,000 seconds per hour to 30 seconds per hour! Another single SQL "Bad Guy" query tuned



<u>Application Development:</u>

- Application Workload Analysis: E.g. which machine load is produced by a certain Application?
- Explain Tool link (e.g. **SQL PerformanceExpert**, IBM DataStudio)
- Show same SQL on Multiple Schemas to detect "heavy-hitters"
- SQL Text Analysis for free text search (e.g.: BIF [Built-in Function] and UDF [User-Defined Functions] -usage, Java-formatted timestamps, etc.)
- View to detect "heavy-hitters" resulting from identical statements using different predicates
- Find unused (orphaned) SQL



WLX typical use cases

Workload/Performance management:

- Workload-Change, Problem-Detection and Trending, Comparison of CPU consumption, I/O, execution rates, current KPIs and deltas – Calculated and summarized to the costs of multiple apps
- Disc Problem Detection I/O Rates
- SQL KPIs Background Noise and Exceptions
- SELECT Only Table Detection (READ only activity)
- Delay Detection (All queries which are delayed)
- Up and Down Scaling of SQL Workloads
- DSC Flush Analysis
- **CPU Intensive Statements**
- Index Maintenance Costs



WLX functional packages of use cases

Database Administration:

- Find never used Objects (Tables, Indexes, and Tablespaces)
- Find never executed Packages

Audit and Security:

- AUDIT tables being accessed
- AUDIT DB2 data being accessed
- AUDIT data manipulation (insert/update/delete)
- See where updates came from (inside or outside the local network)
- See where data is being accessed from (IP address, intra-/extranet, etc.)
- SQL Text Analysis for free text search (BIF [Built-in Function] and UDF [User-Defined Functions] -usage, Java-formatted timestamps, etc.)



Appendix

DB2 APARs to check for:

- PM77114 DB2 10 UK91560 Abend S04E
- PM78143 DB2 10 UK93065 SOS HIPER
- PM80371 DB2 10 UK93127 Serviceability for SHTE
- PM83370 DB2 10 UK94511 Fields TB and IX sometimes wrong
- PM85376 DB2 10 UK96310 Abend S04E
- PM89121 DB2 10 UK95683 Storage leak leading to abend HIPER
- PM91159 DB2 10 UK97197 Improve accuracy of IFCID 316 and 401
- PM92610 DB2 11 UK96376 Abend with IFCID 400 or 401
- PM93437 DB2 11 UK97361 IFCID 316 fields length value problem
- PM97922 DB2 10 UI12375 DB2 11 UI12376 Invalid or empty IFCID 316
- PI07461 OPEN Inconsistent QA0401EU, GL and GP
- PI09147 DB2 10 UI15679 DB2 11 UI15680 Abend S04E
- PI09408 DB2 10 UI15740 DB2 11 UI15741 Abend S04E
- PI09788 DB2 11 UI15739 SOS with IFCID400
- PI16183 OPEN MISSING IFCID401

Advertisement ;-)

So now you know...

- Of course it is easier with **SQL WorkLoadExpert for DB2 z/OS**
 - Data Warehouse
 - Extensible and Extendable
 - Low CPU cost
- For Single SQL tuning it links to **SQLPerformanceExpert for DB2** z/OS
- Both work with **BindImpactExpert for DB2 z/OS** for Access Path comparison and release control

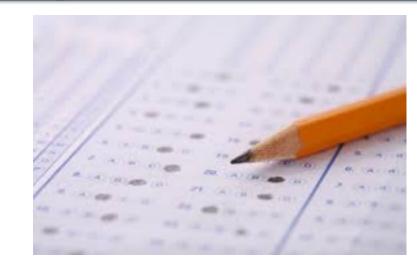
IDUG DB2 Tech Conference

#IDUG

Prague, Czech Republic | November 2014

Ulf Heinric

SOFTWARE ENGINEERING GmbH *u.heinrich@seg.de*



25 years of missed opportunities? SQL Tuning Revisited

