

IMAGE Database Analysis and Optimization

How messy are your databases?

Craig Lalley

Analysis and Optimization



- How messy is your database?
- Blocks vs. Paging
- Master Data Sets
- Hashing Algorithms
- Migrating Secondaries
- HowMessy
- Master Data Sets
- Detail Data Sets
- When to Repack

How Messy Is Your Database?

- 60% of performance improvements come from database optimization
- The goal is to minimize the amount of I/O and improve memory efficiency and CPU utilization
- Databases are dynamic and the law of entropy prevails
- HowMessy from Robelle, Adager or DBLOADNG from the CSL, LPS

Blocks

- IMAGE data is accessed in pages (RISC architecture Howmessy reports in blocks)
- A block may contain multiple records, depending on block and record size
- More entries per block means fewer I/O
- Rule of thumb for block size is 2048 and tweak from there
- Sample MPEX %listf MACORD@,DB

MPEX %LISTF MACORD@,DB

ACCOUNT=		MACS	GROUP=		MACORD		(CONTINUED)					
FILENAME	SET NAME	TYPE	LDEV	ENTRY COUNT	CAPA-CITY	%FULL	BLK FCT	SECTORS	%BLOCK WASTED	SECTRS WASTED		
MACORD36	OPEN-AR	D		8936	2000040	0.4%	35	571456	1.3%	7589		
MACORD37	CLOSED-AR	D		803153	3000025	26.8%	35	857152	1.3%	11384		
MACORD38	MESSAGE-FILE	D		0	1000008	0.0%	9	88912	0.6%	520		
MACORD39	ORDER-ACTIONS	D		425540240179099664		70.0%	28	32	0.0%	-263		
	39.001	C										
	39.002	C										
MACORD40	PROCESSING	D		89147	3600047	2.5%	11	3272784	0.2%	7670		
MACORD41	RETURNS	D		439444	907834	48.4%	17	534032	0.2%	1251		
MACORD42	CUSTOM-CORR	D		51167	435372	11.8%	21	165872	1.4%	2267		
MACORD43	TRANSMITTAL	D		2308030432167230		71.8%	30	8577936	0.2%	16753		
MACORD44	FRAUDS	D		16487	592384	2.8%	26	182288	0.8%	1424		
MACORD45	ORDER-XREF	D		424656245676020215		61.3%	31	32	0.0%	-255		
	45.001	C										
	45.002	C										
MACORD46	AR-COMMENTS	D		870016	1246544	69.8%	26	479456	0.3%	1498		
MACORD47	PICK-LOT-DETAIL	D		411368688823400850		58.5%	10	32	0.0%	-263		
	47.001	C										
	47.002	C										
MACORD48	ORDER-KITS	D	40	0	1001	0.0%	91	112	0.0%	0		
MACORD49	ADDED-DATA-1	D		83670	1600040	52.3%	10	160048	7.7%	12378		
MACORD50	ADDED-DATA-2	D		468815	3102210	15.1%	10	3102224	7.7%	239937		

Master Data Sets

- Search item (key) must be unique
- Record location is determined by a hashing algorithm based on capacity
- Different algorithms are used based on data type
- Imagine a Roulette wheel: after each spin the area where the ball lands turns black.



Hashing Algorithms

- Collisions
- If in the same block, no extra I/O needed
- If block is full, an extra I/O is required
- MaxBlk > 15 could be an issue

Migrating Secondaries

- Deleting the primary value will cause added I/O by adjusting secondaries
- Clustering problems



HOWMESSY

HowMessy/XL (Version 2.7) Data Base: MACORD.MACORD.MACS Run on: SAT, MAR 9, 2002, 11:52 PM
for IMAGE/3000 databases By Robelle Consulting Ltd. Page: 1

Data Set	Type	Capacity	Entries	Load Factor	Secon-		Blks	Blk	Fact	Search Field	Chain	Ave	Std	Expd	Avg	Ineff	Elong-
					daries	Max											
REF-NOS	Man	1009	0	0.0%	0.0%	0	115	0	REF-NO	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ACTION-MAST	Ato	1	0	0.0%	0.0%	1	1	1	ACTION-SEARCH	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CONTROL-NOS	Ato	816251	363135	44.5%	3.0%	529	88	3	CONTROL-1	3	1.03	0.17	1.00	1.85	85.2%	1.85	
CUST-SEARCH	Ato	2724300120970653	77.0%	30.3%	7	53	XREF-NO	9	1.43	0.28	1.00	1.38	29.2%	1.38			
FULL-ORDER	Man	50021	0	0.0%	0.0%	0	115	0	FULL-ORDER-NO	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ORDER-SEOS	Ato	4665373	3192539	68.4%	32.0%	37	75	17	ORDER-SEQ	17	1.47	0.83	1.00	1.40	27.9%	1.40	
FRAUDS-MAST	Ato	1201251	27769	2.3%	1.2%	0	60	3	FRAUD-SEARCH	3	1.01	0.11	1.00	1.00	0.0%	1.00	
ORDER-SEARCH	Ato	7470281943422461	58.1%	26.5%	3	53	XREF-NO	35	1.36	0.00	0.53	1.23	18.1%	2.35			
ORD-KIT-MAST	Ato	1001	0	0.0%	0.0%	0	75	0	ORDER-LINE	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GENERIC-MAST	Ato	3217547	429616	13.4%	6.7%	0	41	6	GENERIC-SEARCH	6	1.07	0.27	1.00	1.02	1.6%	1.02	
ITEM-EDP-NOS	Ato	150000	71	0.0%	0.0%	0	88	1	EDP-NO	1	1.00	0.00	1.00	1.00	0.0%	1.00	
TRACK-KEYS	Ato	3253933	1949631	59.9%	42.7%	81	76	63	INQ-TRACK-NO	63	1.74	1.94	1.00	1.60	27.0%	1.60	
GIFT-CERT-NOS	Ato	331	247	74.6%	30.4%	0	76	4	GIFT-CERT-NO	4	1.44	0.72	1.00	1.20	14.7%	1.20	
ORDER-KEYS	Ato	3957155923153560	58.5%	49.9%	1526	76	ORDER-NO	161	1.99	3.46	1.00	2.04	34.6%	2.04			
DB-SET-NUMS	Man	100	1	1.0%	0.0%	0	6	1	DB-SET-KEY	1	1.00	0.00	1.00	1.00	0.0%	1.00	
CUSTOMERS	Man	3055260021578059	70.6%	0.0%	2396671	5	CUST-EDP	1	1.00	0.00	0.78	1.00	0.0%	1.2			
OH-ADDITIONS	Man	3530488121234662	60.1%	51.5%	1366	17	FULL-ORDER-NO	156	2.06	3.45	1.01	2.03	32.7%	2.00			
PICK-LOT-HEADER	Man	385739	172351	44.7%	20.4%	2	17	6	LOT-NO	6	1.26	0.53	1.00	1.01	0.5%	1.01	
STANDING-HEADER	Man	1001	0	0.0%	0.0%	0	18	0	CUST-EDP	0	0.00	0.00	0.00	0.00	0.0%	0.00	
ORDER-BATCH	Man	192781	128806	66.8%	30.9%	24	8	9	BATCH-NO	9	1.45	0.76	1.00	1.68	49.6%	1.68	

Master Datasets

- High number of secondaries > 30%
- MaxBlock > 15 (Clustering)
- High, inefficient pointers when clustering exists

Detail Datasets

- Highwater mark – delete chains
- Blocking factor
- Primary path
- Sorted chains
- Maximum and Average chain lengths
- Search fields
- Elongation

HowMessy Details

HowMessy/XL (Version 2.8) Data Base: MACITM.MACSDATA.SGAII Run on: FRI, MAR 22, 2002, 3:45 PM
for IMAGE/3000 databases By Robelle Solutions Technology Inc. Page: 2

Data Set	Type	Capacity	Entries	Load Factor	Secon- daries	Max Blks	Blk (Highwater) Fact	Search Field	Max Chain	Ave Chain	Std Dev	Expd Blocks	Avg Blocks	Ineff Ptrs	Elong- ation
SOURCE-XREF	Det	19737	15256	77.3%	(15256)	51	REF-SOURCE		28	1.00	0.22	1.00	1.00	0.2%	1.00
CHAINS	Det	2387	693	29.0%	(693)	11	ISOURCE	CHAIN-ACTION	9	1.01	0.12	1.00	1.00	0.3%	1.00
								SICHAIN-NO	8	2.19	1.17	1.00	1.15	6.8%	1.15
SOURCES	Det	25002	15657	62.6%	(15657)	3	ISOURCE		1	1.00	0.00	1.00	1.00	0.0%	1.00
								S-OFFER-NO	393	11.55	34.68	4.44	7.34	54.3%	1.65
SOURCES-S	Det	20001	0	0.0%	(0)	3	ISOURCE		0	0.00	0.00	0.00	0.00	0.0%	0.00
								S-OFFER-NO	0	0.00	0.00	0.00	0.00	0.0%	0.00
DAILY-SOURCES	Det	713520	577013	80.9%	(577011)	8	ISOURCE		1364	46.08	84.03	6.27	45.99	97.6%	7.34
								S-OFFER-NO	0	0.00	0.00	0.00	0.00	0.0%	0.00
ITEM-WEEKS	Det	449298	333986	75.9%	(333984)	2	ITEM-YR		12	1.93	1.53	1.25	1.93	48.3%	1.55
ITEM-WEEKS-S	Det	100000	0	0.0%	(0)	2	ITEM-YR		0	0.00	0.00	0.00	0.00	0.0%	0.00
CORRESPONDENCE	Det	1296	854	65.9%	(854)	18	SICORR-CODE		48	17.79	12.20	1.48	3.06	11.6%	2.07
CUSTOMIZE-CODE	Det	108	12	11.1%	(12)	36	SICUSTOMIZE-CODE		8	6.00	2.83	1.00	1.00	0.0%	1.00
PO-DETAILS	Det	778830	546419	70.2%	(546419)	15	SIP0-NUMBER		478	5.43	11.39	1.14	1.61	11.3%	1.42
								EDP-NO	204	6.74	8.84	1.15	6.62	83.5%	5.74
								VENDOR-NO	30241	683.02	1946.15	46.04	196.66	28.0%	4.27
PO-COMMENTS	Det	162449	120945	74.5%	(120945)	7	IP0-NO-LINE-NO		2	1.00	0.01	1.00	1.00	0.0%	1.00
PO-ACTION	Det	882840	691353	78.3%	(691351)	21	IP0-NO-LINE-NO		27	2.12	1.73	1.00	2.02	47.9%	2.02
VENDOR-ITEMS	Det	550008	354086	64.4%	(354067)	9	IEDP-NO		19	1.01	0.26	1.00	1.01	1.3%	1.01
								ITEM-NO	1555	2.19	5.10	1.02	1.81	36.8%	1.77
								VENDOR-NO	55657	461.39	2140.29	45.03	175.73	43.5%	3.90
EDP-ITEM-XREF	Det	2100010	1644464	78.3%	(1644452)	85	IEDP-NO		246	3.51	7.64	1.00	3.15	61.1%	3.15
								REF-ITEM-NO	15124	2.67	118.99	1.02	1.07	2.8%	1.05
INVENTORY-TRANS	Det	6999960	5398637	77.1%	(5398611)	24	IEDP-NO		9079	46.71	148.94	2.56	40.68	85.0%	15.86
								TRANSACTION-NO	51998	4.45	69.40	1.07	2.16	26.1%	2.01
								FROM-REF	99999	10.31	621.73	1.32	5.40	42.7%	4.11
ITEM-DESC	Det	625023	445043	71.2%	(445027)	63	IEDP-NO		12	1.77	1.14	1.00	1.14	7.7%	1.14

When to Repack

- Effects of archiving (deleting records)
- How data is loaded
- Effects of delete chains

Timing is everything

Internal DBswitching

- The icing on the cake
- May not have a positive effect
- Only utilize with adequate CPU and Memory resources

IMAGE internal switches

- DSEM Creates sub databases and improves concurrency
- PREFETCH Shortens lock time

Improving I/O

- Omnidex fragmentation
- Spread Automatic and Manual masters across the I/O bandwidth
- Put more thought into detail spreads
- De-Frag/X vs. Reload