Distribution of Parameters for Generated SQL Queries on the TPC-H Database

Scenario: 1-1jp__ng

Scale factor: 2000 GB

The 1000-query pack was generated on: 2023-11-08

First column ('Query parameter or value') displays either:

– a parameter name, or

– a value of the current parameter (only for low-cardinality parameters).

Second column ('Descriptive statistics') displays:

– for low cardinality parameters: frequency and proportion from total (of 1000 queries)

– for higher cardinality parameters: [min, median, max] mean / SD

Query parameter or value	Descriptive statistics
SELECT_n_of_columns	[1, 6, 43] 10 / 10
SELECT_n_of_non_aggr_func___RTRIM	
0	803 (80%)
1	159 (16%)
2	31 (3.1%)
3	7 (0.7%)
SELECT_n_of_non_aggr_func___SUBSTR	
0	765 (77%)
1	182 (18%)
2	41 (4.1%)
3	11 (1.1%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___DOW	
0	878 (88%)
1	89 (8.9%)

Query parameter or value	Descriptive statistics
2	29 (2.9%)
3	4 (0.4%)
SELECT_n_of_non_aggr_func___LOWER	
0	800 (80%)
1	156 (16%)
2	38 (3.8%)
3	4 (0.4%)
4	1 (0.1%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___LTRIM	
0	778 (78%)
1	171 (17%)
2	37 (3.7%)
3	13 (1.3%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___MONTH	
0	842 (84%)
1	105 (11%)
2	38 (3.8%)
3	14 (1.4%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___SQRT	
0	840 (84%)
1	132 (13%)
2	23 (2.3%)
3	5 (0.5%)
SELECT_n_of_non_aggr_func___TRUNC	
0	941 (94%)
1	52 (5.2%)
2	7 (0.7%)
SELECT_n_of_non_aggr_func___UPPER	
0	776 (78%)
1	176 (18%)
2	35 (3.5%)
3	12 (1.2%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___YEAR	
0	858 (86%)
1	99 (9.9%)
2	36 (3.6%)
3	7 (0.7%)
SELECT_n_of_non_aggr_func___DAY	
0	904 (90%)
1	76 (7.6%)
2	16 (1.6%)
3	4 (0.4%)

Query parameter or value	Descriptive statistics
SELECT_n_of_non_aggr_func___ABS	
0	954 (95%)
1	39 (3.9%)
2	6 (0.6%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___LOG	
0	880 (88%)
1	102 (10%)
2	14 (1.4%)
3	4 (0.4%)
SELECT_n_of_non_aggr_func___ROUND	
0	928 (93%)
1	66 (6.6%)
2	6 (0.6%)
SELECT_n_of_non_aggr_func___FLOOR	
0	947 (95%)
1	46 (4.6%)
2	6 (0.6%)
3	1 (0.1%)
SELECT_n_of_all_non_aggr_func	[0.0, 1.0, 22.0] 2.7 / 3.8
SELECT_n_of_aggr_func___MAX	
0	930 (93%)
1	31 (3.1%)
2	19 (1.9%)
3	13 (1.3%)
4	6 (0.6%)
5	1 (0.1%)
SELECT_n_of_aggr_func___COUNT	
0	935 (94%)
1	28 (2.8%)
2	25 (2.5%)
3	4 (0.4%)
4	3 (0.3%)
5	3 (0.3%)
6	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_aggr_func___COUNT_DISTINCT	
0	940 (94%)
1	22 (2.2%)
2	19 (1.9%)
3	10 (1.0%)
4	5 (0.5%)
5	3 (0.3%)
8	1 (0.1%)
SELECT_n_of_aggr_func___MIN	
0	934 (93%)

Query parameter or value	Descriptive statistics
1	24 (2.4%)
2	15 (1.5%)
3	15 (1.5%)
4	7 (0.7%)
5	3 (0.3%)
6	2 (0.2%)
SELECT_n_of_aggr_func___AVG	
0	989 (99%)
1	7 (0.7%)
2	3 (0.3%)
3	1 (0.1%)
SELECT_n_of_aggr_func___SUM	
0	988 (99%)
1	11 (1.1%)
2	1 (0.1%)
SELECT_n_of_all_aggr_func	
0	764 (76%)
1	66 (6.6%)
2	75 (7.5%)
3	47 (4.7%)
4	27 (2.7%)
5	12 (1.2%)
6	6 (0.6%)
7	2 (0.2%)
8	1 (0.1%)
FROM_n_of_join_paths	
1	1,000 (100%)
FROM_n_of_joins___INNER	
0	502 (50%)
1	340 (34%)
2	122 (12%)
3	32 (3.2%)
4	4 (0.4%)
FROM_n_of_joins___RIGHT	
0	505 (51%)
1	320 (32%)
2	138 (14%)
3	30 (3.0%)
4	7 (0.7%)
FROM_n_of_processed_rows	[5, 2,000,000,000, 15,300,027,829] 5,370,568,963 / 6,156,681,424
WHERE_n_of_predicates	
0	228 (23%)
1	253 (25%)
2	224 (22%)
3	153 (15%)

Query parameter or value	Descriptive statistics
4	93 (9.3%)
5	39 (3.9%)
6	10 (1.0%)
WHERE_n_of_attribs_of_type___date	
0	905 (91%)
1	88 (8.8%)
2	7 (0.7%)
WHERE_n_of_attribs_of_type___integer	
0	569 (57%)
1	327 (33%)
2	89 (8.9%)
3	12 (1.2%)
4	3 (0.3%)
WHERE_n_of_attribs_of_type___numeric	
0	735 (74%)
1	222 (22%)
2	37 (3.7%)
3	6 (0.6%)
WHERE_n_of_attribs_of_type___character_varying	
0	533 (53%)
1	330 (33%)
2	118 (12%)
3	17 (1.7%)
4	2 (0.2%)
WHERE_n_of_attribs_of_type___character	
0	832 (83%)
1	143 (14%)
2	25 (2.5%)
WHERE_n_of_pkey_attribs	
0	724 (72%)
1	245 (25%)
2	29 (2.9%)
3	2 (0.2%)
WHERE_n_of_connect_OR	
0	621 (62%)
1	212 (21%)
2	108 (11%)
3	42 (4.2%)
4	15 (1.5%)
5	2 (0.2%)
WHERE_n_of_operators___greater_or_less	
0	546 (55%)
1	306 (31%)
2	104 (10%)
3	31 (3.1%)
4	12 (1.2%)

Query parameter or value	Descriptive statistics
5	1 (0.1%)
WHERE_n_of_operators___in	
0	736 (74%)
1	213 (21%)
2	46 (4.6%)
3	3 (0.3%)
4	2 (0.2%)
WHERE_n_of_operators___between	
0	719 (72%)
1	239 (24%)
2	36 (3.6%)
3	6 (0.6%)
WHERE_n_of_operators___like	
0	905 (91%)
1	90 (9.0%)
2	5 (0.5%)
WHERE_n_of_non_aggr_func___DAY	
0	988 (99%)
1	12 (1.2%)
WHERE_n_of_non_aggr_func___MONTH	
0	985 (99%)
1	14 (1.4%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___FLOOR	
0	976 (98%)
1	23 (2.3%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___SQRT	
0	943 (94%)
1	55 (5.5%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___LOG	
0	957 (96%)
1	42 (4.2%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___TRUNC	
0	983 (98%)
1	17 (1.7%)
WHERE_n_of_non_aggr_func___ROUND	
0	979 (98%)
1	21 (2.1%)
WHERE_n_of_non_aggr_func___YEAR	
0	982 (98%)
1	17 (1.7%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___ABS	

Query parameter or value	Descriptive statistics
0	980 (98%)
1	19 (1.9%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___DOW	
0	984 (98%)
1	15 (1.5%)
2	1 (0.1%)
WHERE_n_of_all_non_aggr_func	
0	781 (78%)
1	191 (19%)
2	24 (2.4%)
3	4 (0.4%)
GROUP_BY_n_of_columns	
0	1,000 (100%)
HAVING_n_of_main_predicates	
0	1,000 (100%)
ORDER_BY_n_of_columns	
0	451 (45%)
1	284 (28%)
2	169 (17%)
3	66 (6.6%)
4	23 (2.3%)
5	7 (0.7%)
limit	[1, 509, 1,000] 504 / 294
offset	[0, 26, 1,000] 252 / 322