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

Scenario: 1-1jp__ng

Scale factor: 500 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___DOW	
0	894 (89%)
1	89 (8.9%)
2	14 (1.4%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func___MONTH	
0	834 (83%)
1	121 (12%)
2	38 (3.8%)
3	7 (0.7%)
SELECT_n_of_non_aggr_func___SUBSTR	
0	801 (80%)
1	144 (14%)
2	47 (4.7%)

Query parameter or value	Descriptive statistics
3	7 (0.7%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___YEAR	
0	860 (86%)
1	99 (9.9%)
2	28 (2.8%)
3	13 (1.3%)
SELECT_n_of_non_aggr_func___RTRIM	
0	799 (80%)
1	163 (16%)
2	27 (2.7%)
3	9 (0.9%)
4	2 (0.2%)
SELECT_n_of_non_aggr_func___UPPER	
0	789 (79%)
1	169 (17%)
2	32 (3.2%)
3	8 (0.8%)
4	2 (0.2%)
SELECT_n_of_non_aggr_func___ABS	
0	918 (92%)
1	73 (7.3%)
2	9 (0.9%)
SELECT_n_of_non_aggr_func___DAY	
0	907 (91%)
1	78 (7.8%)
2	12 (1.2%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func___LOG	
0	870 (87%)
1	114 (11%)
2	12 (1.2%)
3	4 (0.4%)
SELECT_n_of_non_aggr_func___LOWER	
0	801 (80%)
1	151 (15%)
2	42 (4.2%)
3	5 (0.5%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___SQRT	
0	842 (84%)
1	122 (12%)
2	30 (3.0%)
3	5 (0.5%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___LTRIM	

Query parameter or value	Descriptive statistics
0	775 (78%)
1	172 (17%)
2	36 (3.6%)
3	13 (1.3%)
4	4 (0.4%)
SELECT_n_of_non_aggr_func___TRUNC	
0	945 (95%)
1	48 (4.8%)
2	7 (0.7%)
SELECT_n_of_non_aggr_func___FLOOR	
0	943 (94%)
1	56 (5.6%)
2	1 (0.1%)
SELECT_n_of_non_aggr_func___ROUND	
0	943 (94%)
1	53 (5.3%)
2	4 (0.4%)
SELECT_n_of_all_non_aggr_func	[0.0, 1.0, 22.0] 2.6 / 3.7
SELECT_n_of_aggr_func___COUNT_DISTINCT	
0	947 (95%)
1	23 (2.3%)
2	18 (1.8%)
3	7 (0.7%)
4	2 (0.2%)
5	1 (0.1%)
6	2 (0.2%)
SELECT_n_of_aggr_func___MAX	
0	918 (92%)
1	33 (3.3%)
2	24 (2.4%)
3	11 (1.1%)
4	8 (0.8%)
5	6 (0.6%)
SELECT_n_of_aggr_func___SUM	
0	982 (98%)
1	13 (1.3%)
2	4 (0.4%)
3	1 (0.1%)
SELECT_n_of_aggr_func___COUNT	
0	921 (92%)
1	39 (3.9%)
2	22 (2.2%)
3	12 (1.2%)
4	3 (0.3%)
5	2 (0.2%)
6	1 (0.1%)

Query parameter or value	Descriptive statistics
SELECT_n_of_aggr_func___MIN	
0	927 (93%)
1	23 (2.3%)
2	28 (2.8%)
3	13 (1.3%)
4	6 (0.6%)
5	1 (0.1%)
6	2 (0.2%)
SELECT_n_of_aggr_func___AVG	
0	983 (98%)
1	16 (1.6%)
2	1 (0.1%)
SELECT_n_of_all_aggr_func	
0	742 (74%)
1	80 (8.0%)
2	74 (7.4%)
3	50 (5.0%)
4	29 (2.9%)
5	17 (1.7%)
6	6 (0.6%)
7	2 (0.2%)
FROM_n_of_join_paths	
1	1,000 (100%)
FROM_n_of_joins___INNER	
0	492 (49%)
1	324 (32%)
2	140 (14%)
3	40 (4.0%)
4	4 (0.4%)
FROM_n_of_joins___RIGHT	
0	496 (50%)
1	313 (31%)
2	142 (14%)
3	42 (4.2%)
4	7 (0.7%)
FROM_n_of_processed_rows	[5, 750,000,000, 3,825,028,272] 1,472,975,152 / 1,575,685,918
WHERE_n_of_predicates	
0	248 (25%)
1	226 (23%)
2	221 (22%)
3	152 (15%)
4	91 (9.1%)
5	51 (5.1%)
6	11 (1.1%)
WHERE_n_of_attribs_of_type___character_varying	

Query parameter or value	Descriptive statistics
0	541 (54%)
1	316 (32%)
2	128 (13%)
3	15 (1.5%)
WHERE_n_of_attribs_of_type___integer	
0	561 (56%)
1	316 (32%)
2	108 (11%)
3	14 (1.4%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type___numeric	
0	763 (76%)
1	192 (19%)
2	41 (4.1%)
3	4 (0.4%)
WHERE_n_of_attribs_of_type___character	
0	819 (82%)
1	148 (15%)
2	30 (3.0%)
3	3 (0.3%)
WHERE_n_of_attribs_of_type___date	
0	898 (90%)
1	93 (9.3%)
2	9 (0.9%)
WHERE_n_of_pkey_attribs	
0	728 (73%)
1	227 (23%)
2	43 (4.3%)
3	2 (0.2%)
WHERE_n_of_connect_OR	
0	601 (60%)
1	215 (22%)
2	110 (11%)
3	53 (5.3%)
4	18 (1.8%)
5	3 (0.3%)
WHERE_n_of_operators___between	
0	713 (71%)
1	223 (22%)
2	58 (5.8%)
3	5 (0.5%)
4	1 (0.1%)
WHERE_n_of_operators___in	
0	719 (72%)
1	239 (24%)
2	34 (3.4%)

Query parameter or value	Descriptive statistics
3	7 (0.7%)
4	1 (0.1%)
WHERE_n_of_operators___greater_or_less	
0	520 (52%)
1	325 (33%)
2	120 (12%)
3	31 (3.1%)
4	4 (0.4%)
WHERE_n_of_operators___like	
0	881 (88%)
1	113 (11%)
2	5 (0.5%)
3	1 (0.1%)
WHERE_n_of_non_aggr_func___SQRT	
0	949 (95%)
1	49 (4.9%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___DAY	
0	985 (99%)
1	14 (1.4%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___TRUNC	
0	985 (99%)
1	13 (1.3%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___ABS	
0	987 (99%)
1	12 (1.2%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___MONTH	
0	981 (98%)
1	19 (1.9%)
WHERE_n_of_non_aggr_func___ROUND	
0	978 (98%)
1	21 (2.1%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___FLOOR	
0	988 (99%)
1	12 (1.2%)
WHERE_n_of_non_aggr_func___YEAR	
0	981 (98%)
1	19 (1.9%)
WHERE_n_of_non_aggr_func___LOG	
0	973 (97%)
1	27 (2.7%)
WHERE_n_of_non_aggr_func___DOW	

Query parameter or value	Descriptive statistics
0	989 (99%)
1	10 (1.0%)
2	1 (0.1%)
WHERE_n_of_all_non_aggr_func	
0	818 (82%)
1	156 (16%)
2	22 (2.2%)
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	450 (45%)
1	275 (28%)
2	172 (17%)
3	63 (6.3%)
4	28 (2.8%)
5	10 (1.0%)
6	2 (0.2%)
limit	[1, 489, 999] 489 / 283
offset	[0, 0, 1,000] 244 / 325