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

Scenario: 1-1jp__ng

Scale factor: 1000 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, 42] 10 / 10
SELECT_n_of_non_aggr_func___SUBSTR	
0	784 (78%)
1	165 (17%)
2	42 (4.2%)
3	9 (0.9%)
SELECT_n_of_non_aggr_func___DAY	
0	907 (91%)
1	75 (7.5%)
2	16 (1.6%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___MONTH	
0	841 (84%)
1	103 (10%)
2	45 (4.5%)

Query parameter or value	Descriptive statistics
3	11 (1.1%)
SELECT_n_of_non_aggr_func___SQRT	
0	824 (82%)
1	134 (13%)
2	36 (3.6%)
3	5 (0.5%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___LOWER	
0	784 (78%)
1	176 (18%)
2	36 (3.6%)
3	4 (0.4%)
SELECT_n_of_non_aggr_func___RTRIM	
0	799 (80%)
1	152 (15%)
2	36 (3.6%)
3	10 (1.0%)
4	3 (0.3%)
SELECT_n_of_non_aggr_func___YEAR	
0	872 (87%)
1	88 (8.8%)
2	29 (2.9%)
3	11 (1.1%)
SELECT_n_of_non_aggr_func___UPPER	
0	795 (80%)
1	163 (16%)
2	28 (2.8%)
3	11 (1.1%)
4	3 (0.3%)
SELECT_n_of_non_aggr_func___LTRIM	
0	772 (77%)
1	171 (17%)
2	46 (4.6%)
3	10 (1.0%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___LOG	
0	883 (88%)
1	106 (11%)
2	8 (0.8%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func___ABS	
0	929 (93%)
1	64 (6.4%)
2	6 (0.6%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___DOW	

Query parameter or value	Descriptive statistics
0	889 (89%)
1	83 (8.3%)
2	23 (2.3%)
3	5 (0.5%)
SELECT_n_of_non_aggr_func___FLOOR	
0	926 (93%)
1	70 (7.0%)
2	4 (0.4%)
SELECT_n_of_non_aggr_func___ROUND	
0	940 (94%)
1	57 (5.7%)
2	3 (0.3%)
SELECT_n_of_non_aggr_func___TRUNC	
0	960 (96%)
1	37 (3.7%)
2	3 (0.3%)
SELECT_n_of_all_non_aggr_func	[0.0, 1.0, 21.0] 2.6 / 3.8
SELECT_n_of_aggr_func___MAX	
0	929 (93%)
1	32 (3.2%)
2	16 (1.6%)
3	9 (0.9%)
4	6 (0.6%)
5	5 (0.5%)
7	2 (0.2%)
8	1 (0.1%)
SELECT_n_of_aggr_func___AVG	
0	986 (99%)
1	10 (1.0%)
2	4 (0.4%)
SELECT_n_of_aggr_func___MIN	
0	933 (93%)
1	29 (2.9%)
2	22 (2.2%)
3	11 (1.1%)
4	3 (0.3%)
5	2 (0.2%)
SELECT_n_of_aggr_func___SUM	
0	985 (99%)
1	12 (1.2%)
2	3 (0.3%)
SELECT_n_of_aggr_func___COUNT_DISTINCT	
0	940 (94%)
1	23 (2.3%)
2	16 (1.6%)
3	16 (1.6%)

Query parameter or value	Descriptive statistics
4	2 (0.2%)
5	2 (0.2%)
6	1 (0.1%)
SELECT_n_of_aggr_func___COUNT	
0	932 (93%)
1	31 (3.1%)
2	20 (2.0%)
3	13 (1.3%)
4	2 (0.2%)
5	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_all_aggr_func	
0	762 (76%)
1	71 (7.1%)
2	78 (7.8%)
3	41 (4.1%)
4	22 (2.2%)
5	17 (1.7%)
6	5 (0.5%)
7	2 (0.2%)
8	2 (0.2%)
FROM_n_of_join_paths	
1	1,000 (100%)
FROM_n_of_joins___INNER	
0	499 (50%)
1	318 (32%)
2	144 (14%)
3	34 (3.4%)
4	5 (0.5%)
FROM_n_of_joins___RIGHT	
0	463 (46%)
1	350 (35%)
2	152 (15%)
3	32 (3.2%)
4	3 (0.3%)
FROM_n_of_processed_rows	[5, 1,000,000,000, 7,649,989,739] 2,859,466,402 / 3,159,426,448
WHERE_n_of_predicates	
0	235 (24%)
1	240 (24%)
2	228 (23%)
3	160 (16%)
4	78 (7.8%)
5	50 (5.0%)
6	9 (0.9%)
WHERE_n_of_attribs_of_type___character_varying	

Query parameter or value	Descriptive statistics
0	541 (54%)
1	323 (32%)
2	113 (11%)
3	22 (2.2%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type___character	
0	821 (82%)
1	150 (15%)
2	27 (2.7%)
3	2 (0.2%)
WHERE_n_of_attribs_of_type___integer	
0	574 (57%)
1	298 (30%)
2	100 (10%)
3	24 (2.4%)
4	3 (0.3%)
5	1 (0.1%)
WHERE_n_of_attribs_of_type___numeric	
0	782 (78%)
1	182 (18%)
2	32 (3.2%)
3	4 (0.4%)
WHERE_n_of_attribs_of_type___date	
0	890 (89%)
1	102 (10%)
2	8 (0.8%)
WHERE_n_of_pkey_attribs	
0	707 (71%)
1	235 (24%)
2	54 (5.4%)
3	3 (0.3%)
4	1 (0.1%)
WHERE_n_of_connect_OR	
0	612 (61%)
1	220 (22%)
2	102 (10%)
3	51 (5.1%)
4	14 (1.4%)
5	1 (0.1%)
WHERE_n_of_operators___like	
0	874 (87%)
1	122 (12%)
2	4 (0.4%)
WHERE_n_of_operators___between	
0	714 (71%)
1	234 (23%)

Query parameter or value	Descriptive statistics
2	49 (4.9%)
3	3 (0.3%)
WHERE_n_of_operators___greater_or_less	
0	529 (53%)
1	315 (32%)
2	123 (12%)
3	27 (2.7%)
4	6 (0.6%)
WHERE_n_of_operators___in	
0	703 (70%)
1	247 (25%)
2	48 (4.8%)
3	2 (0.2%)
WHERE_n_of_non_aggr_func___LOG	
0	954 (95%)
1	45 (4.5%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___SQRT	
0	926 (93%)
1	72 (7.2%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___ABS	
0	988 (99%)
1	12 (1.2%)
WHERE_n_of_non_aggr_func___DOW	
0	991 (99%)
1	9 (0.9%)
WHERE_n_of_non_aggr_func___YEAR	
0	986 (99%)
1	14 (1.4%)
WHERE_n_of_non_aggr_func___ROUND	
0	987 (99%)
1	13 (1.3%)
WHERE_n_of_non_aggr_func___FLOOR	
0	983 (98%)
1	17 (1.7%)
WHERE_n_of_non_aggr_func___MONTH	
0	985 (99%)
1	15 (1.5%)
WHERE_n_of_non_aggr_func___DAY	
0	980 (98%)
1	20 (2.0%)
WHERE_n_of_non_aggr_func___TRUNC	
0	987 (99%)
1	13 (1.3%)
WHERE_n_of_all_non_aggr_func	

Query parameter or value	Descriptive statistics
0	794 (79%)
1	177 (18%)
2	28 (2.8%)
3	1 (0.1%)
GROUP_BY_n_of_columns	
0	1,000 (100%)
HAVING_n_of_main_predicates	
0	1,000 (100%)
ORDER_BY_n_of_columns	
0	437 (44%)
1	286 (29%)
2	147 (15%)
3	85 (8.5%)
4	36 (3.6%)
5	8 (0.8%)
6	1 (0.1%)
limit	[1, 493, 1,000] 488 / 287
offset	[0, 24, 997] 261 / 329