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

Scenario: 1-1jp__g__h0-3ss_h0-3nss

Scale factor: 500 GB

The 1000-query pack was generated on: 2023-11-05

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	[1.0, 4.0, 37.0] 5.7 / 5.2
SELECT_n_of_non_aggr_func___LTRIM	
0	795 (80%)
1	185 (19%)
2	18 (1.8%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___UPPER	
0	775 (78%)
1	194 (19%)
2	26 (2.6%)
3	4 (0.4%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___RTRIM	
0	804 (80%)
1	164 (16%)

Query parameter or value	Descriptive statistics
2	29 (2.9%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func___DAY	
0	949 (95%)
1	47 (4.7%)
2	4 (0.4%)
SELECT_n_of_non_aggr_func___LOG	
0	936 (94%)
1	62 (6.2%)
2	1 (0.1%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___MONTH	
0	899 (90%)
1	82 (8.2%)
2	16 (1.6%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func___YEAR	
0	908 (91%)
1	83 (8.3%)
2	7 (0.7%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___LOWER	
0	787 (79%)
1	191 (19%)
2	20 (2.0%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___SQRT	
0	910 (91%)
1	84 (8.4%)
2	6 (0.6%)
SELECT_n_of_non_aggr_func___FLOOR	
0	978 (98%)
1	21 (2.1%)
2	1 (0.1%)
SELECT_n_of_non_aggr_func___ABS	
0	960 (96%)
1	36 (3.6%)
2	4 (0.4%)
SELECT_n_of_non_aggr_func___DOW	
0	924 (92%)
1	67 (6.7%)
2	8 (0.8%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___SUBSTR	
0	812 (81%)
1	167 (17%)

Query parameter or value	Descriptive statistics
2	18 (1.8%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___TRUNC	
0	966 (97%)
1	33 (3.3%)
2	1 (0.1%)
SELECT_n_of_non_aggr_func___ROUND	
0	957 (96%)
1	42 (4.2%)
2	1 (0.1%)
SELECT_n_of_all_non_aggr_func	[0.00, 1.00, 18.00] 1.85 / 2.08
SELECT_n_of_aggr_func___MIN	
0	793 (79%)
1	104 (10%)
2	57 (5.7%)
3	23 (2.3%)
4	13 (1.3%)
5	9 (0.9%)
6	1 (0.1%)
SELECT_n_of_aggr_func___COUNT	
0	778 (78%)
1	114 (11%)
2	55 (5.5%)
3	31 (3.1%)
4	14 (1.4%)
5	7 (0.7%)
6	1 (0.1%)
SELECT_n_of_aggr_func___COUNT_DISTINCT	
0	758 (76%)
1	112 (11%)
2	73 (7.3%)
3	33 (3.3%)
4	15 (1.5%)
5	8 (0.8%)
7	1 (0.1%)
SELECT_n_of_aggr_func___MAX	
0	756 (76%)
1	118 (12%)
2	72 (7.2%)
3	28 (2.8%)
4	17 (1.7%)
5	7 (0.7%)
6	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_aggr_func___SUM	

Query parameter or value	Descriptive statistics
0	965 (97%)
1	29 (2.9%)
2	6 (0.6%)
SELECT_n_of_aggr_func___AVG	
0	949 (95%)
1	43 (4.3%)
2	6 (0.6%)
3	2 (0.2%)
SELECT_n_of_all_aggr_func	
0	193 (19%)
1	289 (29%)
2	244 (24%)
3	131 (13%)
4	74 (7.4%)
5	52 (5.2%)
6	13 (1.3%)
7	4 (0.4%)
FROM_n_of_join_paths	
1	1,000 (100%)
FROM_n_of_joins___INNER	
0	499 (50%)
1	316 (32%)
2	137 (14%)
3	45 (4.5%)
4	3 (0.3%)
FROM_n_of_joins___RIGHT	
0	490 (49%)
1	340 (34%)
2	129 (13%)
3	39 (3.9%)
4	2 (0.2%)
FROM_n_of_processed_rows	[5, 500,000,000, 3,825,028,272] 1,485,725,321 / 1,577,489,825
WHERE_n_of_predicates	
0	239 (24%)
1	203 (20%)
2	226 (23%)
3	158 (16%)
4	112 (11%)
5	52 (5.2%)
6	10 (1.0%)
WHERE_n_of_attribs_of_type___date	
0	898 (90%)
1	95 (9.5%)
2	7 (0.7%)
WHERE_n_of_attribs_of_type___integer	

Query parameter or value	Descriptive statistics
0	536 (54%)
1	338 (34%)
2	97 (9.7%)
3	29 (2.9%)
WHERE_n_of_attribs_of_type___character_varying	
0	554 (55%)
1	306 (31%)
2	116 (12%)
3	22 (2.2%)
4	2 (0.2%)
WHERE_n_of_attribs_of_type___character	
0	801 (80%)
1	174 (17%)
2	25 (2.5%)
WHERE_n_of_attribs_of_type___numeric	
0	727 (73%)
1	222 (22%)
2	44 (4.4%)
3	5 (0.5%)
4	2 (0.2%)
WHERE_n_of_pkey_attribs	
0	684 (68%)
1	273 (27%)
2	37 (3.7%)
3	6 (0.6%)
WHERE_n_of_connect_OR	
0	565 (57%)
1	248 (25%)
2	119 (12%)
3	41 (4.1%)
4	24 (2.4%)
5	3 (0.3%)
WHERE_n_of_operators___between	
0	720 (72%)
1	235 (24%)
2	40 (4.0%)
3	5 (0.5%)
WHERE_n_of_operators___greater_or_less	
0	513 (51%)
1	309 (31%)
2	134 (13%)
3	37 (3.7%)
4	6 (0.6%)
5	1 (0.1%)
WHERE_n_of_operators___like	
0	874 (87%)

Query parameter or value	Descriptive statistics
1	119 (12%)
2	6 (0.6%)
3	1 (0.1%)
WHERE_n_of_operators___in	
0	692 (69%)
1	254 (25%)
2	47 (4.7%)
3	7 (0.7%)
WHERE_n_of_non_aggr_func___MONTH	
0	980 (98%)
1	19 (1.9%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___FLOOR	
0	970 (97%)
1	30 (3.0%)
WHERE_n_of_non_aggr_func___SQRT	
0	947 (95%)
1	48 (4.8%)
2	5 (0.5%)
WHERE_n_of_non_aggr_func___LOG	
0	961 (96%)
1	37 (3.7%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___DOW	
0	984 (98%)
1	15 (1.5%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___ROUND	
0	980 (98%)
1	19 (1.9%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___DAY	
0	984 (98%)
1	16 (1.6%)
WHERE_n_of_non_aggr_func___TRUNC	
0	980 (98%)
1	20 (2.0%)
WHERE_n_of_non_aggr_func___YEAR	
0	989 (99%)
1	11 (1.1%)
WHERE_n_of_non_aggr_func___ABS	
0	979 (98%)
1	21 (2.1%)
WHERE_n_of_all_non_aggr_func	
0	791 (79%)
1	167 (17%)

Query parameter or value	Descriptive statistics
2	37 (3.7%)
3	5 (0.5%)
GROUP_BY_n_of_columns	
0	260 (26%)
1	279 (28%)
2	209 (21%)
3	139 (14%)
4	58 (5.8%)
5	48 (4.8%)
6	7 (0.7%)
HAVING_n_of_main_predicates	[0.00, 4.00, 9.00] 3.33 / 2.61
HAVING_n_of_main_predicates___non_scalar_subquery	
0	446 (45%)
1	185 (19%)
2	167 (17%)
3	202 (20%)
HAVING_n_of_main_predicates___scalar_subquery	
0	469 (47%)
1	165 (17%)
2	182 (18%)
3	184 (18%)
HAVING_n_of_subqueries___non_scalar_subquery	
0	446 (45%)
1	185 (19%)
2	167 (17%)
3	202 (20%)
HAVING_n_of_subqueries___scalar_subquery	[0.0, 1.0, 21.0] 2.3 / 3.6
HAVING_n_of_processed_rows_by_subqueries	[0, 375,000,075, 91,800,678,408] 5,087,460,433 / 10,526,280,033
ORDER_BY_n_of_columns	
0	321 (32%)
1	441 (44%)
2	207 (21%)
3	26 (2.6%)
4	5 (0.5%)
limit	[6, 515, 1,000] 508 / 288
offset	[0, 0, 998] 234 / 314