

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

## Scenario: 1-6jp\_rjpp\_\_g\_\_h0-3ss\_h0-3nss

## Scale factor: 500 GB

## The 1000-query pack was generated on: 2023-11-13

## 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, 9, 180] 15 / 24
SELECT_n_of_non_aggr_func__LTRIM	
0	652 (65%)
1	239 (24%)
2	71 (7.1%)
3	19 (1.9%)
4	7 (0.7%)
5	6 (0.6%)
6	4 (0.4%)
7	1 (0.1%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func__RTRIM	
0	633 (63%)
1	236 (24%)
13	1 (0.1%)

Query parameter or value	Descriptive statistics
2	86 (8.6%)
3	19 (1.9%)
4	11 (1.1%)
5	9 (0.9%)
6	3 (0.3%)
8	2 (0.2%)
SELECT__n_of__non__aggr__func__SUBSTR	
0	645 (65%)
1	248 (25%)
10	1 (0.1%)
2	65 (6.5%)
3	24 (2.4%)
4	10 (1.0%)
5	3 (0.3%)
6	4 (0.4%)
SELECT__n_of__non__aggr__func__UPPER	
0	607 (61%)
1	274 (27%)
2	82 (8.2%)
3	21 (2.1%)
4	10 (1.0%)
5	4 (0.4%)
6	2 (0.2%)
SELECT__n_of__non__aggr__func__LOWER	[0.00, 0.00, 11.00] 0.59 / 1.04
SELECT__n_of__non__aggr__func__ROUND	
0	933 (93%)
1	57 (5.7%)
2	6 (0.6%)
3	2 (0.2%)
4	1 (0.1%)
5	1 (0.1%)
SELECT__n_of__non__aggr__func__YEAR	
0	853 (85%)
1	95 (9.5%)
2	29 (2.9%)
3	11 (1.1%)
4	7 (0.7%)
5	3 (0.3%)
7	2 (0.2%)
SELECT__n_of__non__aggr__func__ABS	
0	930 (93%)
1	58 (5.8%)
2	10 (1.0%)
3	1 (0.1%)
4	1 (0.1%)
SELECT__n_of__non__aggr__func__DOW	

Query parameter or value	Descriptive statistics
0	855 (86%)
1	107 (11%)
2	22 (2.2%)
3	12 (1.2%)
4	3 (0.3%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func__MONTH	
0	836 (84%)
1	107 (11%)
2	23 (2.3%)
3	18 (1.8%)
4	8 (0.8%)
5	1 (0.1%)
6	3 (0.3%)
7	2 (0.2%)
8	2 (0.2%)
SELECT_n_of_non_aggr_func__SQRT	
0	826 (83%)
1	128 (13%)
2	24 (2.4%)
3	10 (1.0%)
4	7 (0.7%)
5	2 (0.2%)
7	2 (0.2%)
9	1 (0.1%)
SELECT_n_of_non_aggr_func__DAY	
0	899 (90%)
1	74 (7.4%)
2	19 (1.9%)
3	4 (0.4%)
4	3 (0.3%)
6	1 (0.1%)
SELECT_n_of_non_aggr_func__FLOOR	
0	923 (92%)
1	63 (6.3%)
2	9 (0.9%)
3	3 (0.3%)
4	1 (0.1%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func__LOG	
0	875 (88%)
1	102 (10%)
2	18 (1.8%)
3	5 (0.5%)
SELECT_n_of_non_aggr_func__TRUNC	
0	929 (93%)

Query parameter or value	Descriptive statistics
1	61 (6.1%)
2	7 (0.7%)
3	3 (0.3%)
SELECT_n_of_all_non_aggr_func	[0.0, 3.0, 69.0] 4.5 / 6.9
SELECT_n_of_aggr_func__COUNT	[0.00, 0.00, 14.00] 0.95 / 2.15
SELECT_n_of_aggr_func__MAX	[0.00, 0.00, 11.00] 0.87 / 1.92
SELECT_n_of_aggr_func__COUNT_DISTINCT	[0.00, 0.00, 13.00] 0.92 / 2.10
SELECT_n_of_aggr_func__SUM	
0	949 (95%)
1	35 (3.5%)
2	12 (1.2%)
3	3 (0.3%)
4	1 (0.1%)
SELECT_n_of_aggr_func__AVG	
0	927 (93%)
1	47 (4.7%)
2	18 (1.8%)
3	6 (0.6%)
4	2 (0.2%)
SELECT_n_of_aggr_func__MIN	[0.00, 0.00, 11.00] 0.82 / 1.84
SELECT_n_of_all_aggr_func	[0.0, 3.0, 15.0] 3.7 / 3.1
FROM_n_of_join_paths	
1	154 (15%)
2	155 (16%)
3	183 (18%)
4	166 (17%)
5	176 (18%)
6	166 (17%)
FROM_n_of_super_joins__LEFT	
0	568 (57%)
1	312 (31%)
2	95 (9.5%)
3	24 (2.4%)
4	1 (0.1%)
FROM_n_of_super_joins__RIGHT	
0	535 (54%)
1	326 (33%)
2	102 (10%)
3	30 (3.0%)
4	7 (0.7%)
FROM_n_of_super_joins__FULL	
0	539 (54%)
1	312 (31%)
2	113 (11%)
3	28 (2.8%)
4	7 (0.7%)

Query parameter or value	Descriptive statistics
5	1 (0.1%)
FROM_n_of_joins__INNER	[0.00, 3.00, 13.00] 3.37 / 2.48
FROM_n_of_joins__RIGHT	[0.00, 3.00, 13.00] 3.37 / 2.52
FROM_n_of_processed_rows	[5, 4,725,028,270, 22,150,169,477] 6,032,492,309 / 4,853,940,554
WHERE_n_of_predicates	[0.00, 3.00, 13.00] 3.93 / 3.03
WHERE_n_of_attribs_of_type__character	
0	650 (65%)
1	238 (24%)
2	88 (8.8%)
3	21 (2.1%)
4	3 (0.3%)
WHERE_n_of_attribs_of_type__character_varying	
0	377 (38%)
1	280 (28%)
2	162 (16%)
3	104 (10%)
4	47 (4.7%)
5	21 (2.1%)
6	4 (0.4%)
7	4 (0.4%)
9	1 (0.1%)
WHERE_n_of_attribs_of_type__date	
0	768 (77%)
1	197 (20%)
2	29 (2.9%)
3	5 (0.5%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type__integer	
0	352 (35%)
1	310 (31%)
2	172 (17%)
3	97 (9.7%)
4	55 (5.5%)
5	10 (1.0%)
6	4 (0.4%)
WHERE_n_of_attribs_of_type__numeric	
0	550 (55%)
1	304 (30%)
2	98 (9.8%)
3	35 (3.5%)
4	13 (1.3%)
WHERE_n_of_pkey_attribs	
0	509 (51%)
1	324 (32%)
2	114 (11%)

Query parameter or value	Descriptive statistics
3	45 (4.5%)
4	7 (0.7%)
5	1 (0.1%)
WHERE_n_of_connect_OR	[0.00, 1.00, 11.00] 2.30 / 2.50
WHERE_n_of_operators__between	
0	514 (51%)
1	302 (30%)
2	133 (13%)
3	41 (4.1%)
4	7 (0.7%)
5	3 (0.3%)
WHERE_n_of_operators__greater_or_less	
0	337 (34%)
1	238 (24%)
2	202 (20%)
3	122 (12%)
4	51 (5.1%)
5	32 (3.2%)
6	15 (1.5%)
7	3 (0.3%)
WHERE_n_of_operators__like	
0	788 (79%)
1	176 (18%)
2	29 (2.9%)
3	6 (0.6%)
4	1 (0.1%)
WHERE_n_of_operators__in	
0	521 (52%)
1	291 (29%)
2	135 (14%)
3	34 (3.4%)
4	17 (1.7%)
5	2 (0.2%)
WHERE_n_of_non_aggr_func__LOG	
0	936 (94%)
1	64 (6.4%)
WHERE_n_of_non_aggr_func__MONTH	
0	962 (96%)
1	36 (3.6%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func__TRUNC	
0	961 (96%)
1	38 (3.8%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__ABS	
0	954 (95%)

Query parameter or value	Descriptive statistics
1	45 (4.5%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__FLOOR	
0	960 (96%)
1	40 (4.0%)
WHERE_n_of_non_aggr_func__SQRT	
0	884 (88%)
1	104 (10%)
2	12 (1.2%)
WHERE_n_of_non_aggr_func__ROUND	
0	957 (96%)
1	43 (4.3%)
WHERE_n_of_non_aggr_func__YEAR	
0	963 (96%)
1	36 (3.6%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__DAY	
0	963 (96%)
1	37 (3.7%)
WHERE_n_of_non_aggr_func__DOW	
0	968 (97%)
1	31 (3.1%)
2	1 (0.1%)
WHERE_n_of_all_non_aggr_func	
0	639 (64%)
1	246 (25%)
2	90 (9.0%)
3	16 (1.6%)
4	9 (0.9%)
GROUP_BY_n_of_columns	[0.00, 2.00, 13.00] 3.10 / 2.86
HAVING_n_of_main_predicates	[0.00, 4.00, 9.00] 3.33 / 2.53
HAVING_n_of_main_predicates__non_scalar_subquery	
0	436 (44%)
1	178 (18%)
2	206 (21%)
3	180 (18%)
HAVING_n_of_main_predicates__scalar_subquery	
0	450 (45%)
1	180 (18%)
2	196 (20%)
3	174 (17%)
HAVING_n_of_subqueries__non_scalar_subquery	
0	436 (44%)
1	178 (18%)
2	206 (21%)
3	180 (18%)

Query parameter or value	Descriptive statistics
HAVING_n_of_subqueries__scalar_subquery	[0.0, 1.0, 25.0] 2.5 / 3.8
HAVING_n_of_processed_rows_by_subquery	[0, 610,000,630, 348,702,825,300] 21,871,501,715 / 37,448,566,549
ORDER_BY_n_of_columns	[0.00, 1.00, 9.00] 1.27 / 1.19
limit	[1, 487, 999] 500 / 281
offset	[0, 0, 996] 233 / 315