

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

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

## Scale factor: 200 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, 40.0] 5.8 / 5.5
SELECT_n_of_non_aggr_func__LOWER	
0	790 (79%)
1	187 (19%)
2	20 (2.0%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func__UPPER	
0	808 (81%)
1	168 (17%)
2	22 (2.2%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func__LTRIM	
0	802 (80%)
1	168 (17%)
2	27 (2.7%)

Query parameter or value	Descriptive statistics
3	2 (0.2%)
5	1 (0.1%)
SELECT__n_of__non__aggr__func__RTRIM	
0	788 (79%)
1	187 (19%)
2	21 (2.1%)
3	4 (0.4%)
SELECT__n_of__non__aggr__func__SUBSTR	
0	773 (77%)
1	194 (19%)
2	31 (3.1%)
3	2 (0.2%)
SELECT__n_of__non__aggr__func__SQRT	
0	880 (88%)
1	113 (11%)
2	5 (0.5%)
3	1 (0.1%)
4	1 (0.1%)
SELECT__n_of__non__aggr__func__DAY	
0	936 (94%)
1	58 (5.8%)
2	3 (0.3%)
3	3 (0.3%)
SELECT__n_of__non__aggr__func__DOW	
0	934 (93%)
1	57 (5.7%)
2	8 (0.8%)
3	1 (0.1%)
SELECT__n_of__non__aggr__func__YEAR	
0	911 (91%)
1	81 (8.1%)
2	7 (0.7%)
3	1 (0.1%)
SELECT__n_of__non__aggr__func__MONTH	
0	897 (90%)
1	85 (8.5%)
2	14 (1.4%)
3	4 (0.4%)
SELECT__n_of__non__aggr__func__ABS	
0	970 (97%)
1	30 (3.0%)
SELECT__n_of__non__aggr__func__ROUND	
0	961 (96%)
1	37 (3.7%)
2	2 (0.2%)
SELECT__n_of__non__aggr__func__FLOOR	

Query parameter or value	Descriptive statistics
0	960 (96%)
1	39 (3.9%)
2	1 (0.1%)
SELECT_n_of_non_aggr_func__LOG	
0	946 (95%)
1	51 (5.1%)
2	3 (0.3%)
SELECT_n_of_non_aggr_func__TRUNC	
0	966 (97%)
1	32 (3.2%)
2	2 (0.2%)
SELECT_n_of_all_non_aggr_func	[0.00, 1.00, 20.00] 1.90 / 2.28
SELECT_n_of_aggr_func__MIN	
0	777 (78%)
1	111 (11%)
2	66 (6.6%)
3	26 (2.6%)
4	12 (1.2%)
5	6 (0.6%)
6	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_aggr_func__COUNT	
0	771 (77%)
1	109 (11%)
2	48 (4.8%)
3	41 (4.1%)
4	22 (2.2%)
5	8 (0.8%)
6	1 (0.1%)
SELECT_n_of_aggr_func__MAX	
0	776 (78%)
1	107 (11%)
2	61 (6.1%)
3	36 (3.6%)
4	9 (0.9%)
5	8 (0.8%)
6	1 (0.1%)
7	2 (0.2%)
SELECT_n_of_aggr_func__COUNT_DISTINCT	
0	792 (79%)
1	104 (10%)
2	49 (4.9%)
3	31 (3.1%)
4	14 (1.4%)
5	8 (0.8%)
6	1 (0.1%)

Query parameter or value	Descriptive statistics
7	1 (0.1%)
SELECT_n_of_aggr_func__SUM	
0	955 (96%)
1	40 (4.0%)
2	4 (0.4%)
3	1 (0.1%)
SELECT_n_of_aggr_func__AVG	
0	951 (95%)
1	42 (4.2%)
2	7 (0.7%)
SELECT_n_of_all_aggr_func	
0	207 (21%)
1	286 (29%)
2	217 (22%)
3	156 (16%)
4	66 (6.6%)
5	48 (4.8%)
6	14 (1.4%)
7	6 (0.6%)
FROM_n_of_join_paths	
1	1,000 (100%)
FROM_n_of_joins__INNER	
0	491 (49%)
1	347 (35%)
2	133 (13%)
3	25 (2.5%)
4	4 (0.4%)
FROM_n_of_joins__RIGHT	
0	496 (50%)
1	324 (32%)
2	146 (15%)
3	31 (3.1%)
4	3 (0.3%)
FROM_n_of_processed_rows	[5, 200,000,000, 1,530,018,464] 562,962,300 / 624,216,112
WHERE_n_of_predicates	
0	231 (23%)
1	263 (26%)
2	222 (22%)
3	151 (15%)
4	89 (8.9%)
5	34 (3.4%)
6	10 (1.0%)
WHERE_n_of_attribs_of_type__character_varying	
0	546 (55%)
1	316 (32%)

Query parameter or value	Descriptive statistics
2	113 (11%)
3	23 (2.3%)
4	2 (0.2%)
WHERE_n_of_attribs_of_type__numeric	
0	761 (76%)
1	206 (21%)
2	29 (2.9%)
3	4 (0.4%)
WHERE_n_of_attribs_of_type__character	
0	832 (83%)
1	142 (14%)
2	24 (2.4%)
3	2 (0.2%)
WHERE_n_of_attribs_of_type__integer	
0	583 (58%)
1	325 (33%)
2	82 (8.2%)
3	6 (0.6%)
4	4 (0.4%)
WHERE_n_of_attribs_of_type__date	
0	882 (88%)
1	106 (11%)
2	10 (1.0%)
3	2 (0.2%)
WHERE_n_of_pkey_attribs	
0	737 (74%)
1	234 (23%)
2	28 (2.8%)
3	1 (0.1%)
WHERE_n_of_connect_OR	
0	641 (64%)
1	209 (21%)
2	88 (8.8%)
3	44 (4.4%)
4	14 (1.4%)
5	4 (0.4%)
WHERE_n_of_operators__in	
0	730 (73%)
1	231 (23%)
2	34 (3.4%)
3	4 (0.4%)
5	1 (0.1%)
WHERE_n_of_operators__greater_or_less	
0	553 (55%)
1	302 (30%)
2	111 (11%)

Query parameter or value	Descriptive statistics
3	32 (3.2%)
4	2 (0.2%)
WHERE_n_of_operators_between	
0	714 (71%)
1	240 (24%)
2	37 (3.7%)
3	9 (0.9%)
WHERE_n_of_operators_like	
0	880 (88%)
1	111 (11%)
2	9 (0.9%)
WHERE_n_of_non_aggr_func_ABS	
0	984 (98%)
1	16 (1.6%)
WHERE_n_of_non_aggr_func_LOG	
0	972 (97%)
1	25 (2.5%)
2	3 (0.3%)
WHERE_n_of_non_aggr_func_ROUND	
0	981 (98%)
1	18 (1.8%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func_DOW	
0	975 (98%)
1	24 (2.4%)
3	1 (0.1%)
WHERE_n_of_non_aggr_func_YEAR	
0	985 (99%)
1	15 (1.5%)
WHERE_n_of_non_aggr_func_MONTH	
0	983 (98%)
1	17 (1.7%)
WHERE_n_of_non_aggr_func_SQRT	
0	946 (95%)
1	53 (5.3%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func_DAY	
0	980 (98%)
1	20 (2.0%)
WHERE_n_of_non_aggr_func_TRUNC	
0	983 (98%)
1	17 (1.7%)
WHERE_n_of_non_aggr_func_FLOOR	
0	983 (98%)
1	17 (1.7%)
WHERE_n_of_all_non_aggr_func	

Query parameter or value	Descriptive statistics
0	804 (80%)
1	162 (16%)
2	29 (2.9%)
3	5 (0.5%)
GROUP_BY_n_of_columns	
0	267 (27%)
1	284 (28%)
2	210 (21%)
3	124 (12%)
4	68 (6.8%)
5	40 (4.0%)
6	7 (0.7%)
HAVING_n_of_main_predicates	[0.00, 4.00, 9.00] 3.29 / 2.58
HAVING_n_of_main_predicates__scalar_subquery	
0	470 (47%)
1	163 (16%)
2	184 (18%)
3	183 (18%)
HAVING_n_of_main_predicates__non_scalar_subquery	
0	441 (44%)
1	177 (18%)
2	192 (19%)
3	190 (19%)
HAVING_n_of_subqueries__scalar_subquery	[0.0, 1.0, 20.0] 2.3 / 3.6
HAVING_n_of_subqueries__non_scalar_subquery	
0	441 (44%)
1	177 (18%)
2	192 (19%)
3	190 (19%)
HAVING_n_of_processed_rows_by_subqueries	[0, 120,000,110, 33,000,405,548] 1,878,466,860 / 4,012,922,012
ORDER_BY_n_of_columns	
0	283 (28%)
1	441 (44%)
2	223 (22%)
3	42 (4.2%)
4	10 (1.0%)
5	1 (0.1%)
limit	[3, 500, 1,000] 504 / 286
offset	[0, 0, 1,000] 241 / 318