

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

## Scenario: 1-1jp\_\_g\_\_h0ss\_h0nss

## Scale factor: 5 GB

## The 1000-query pack was generated on: 2023-11-04

## 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] 5.6 / 5.4
SELECT_n_of_non_aggr_func__LOWER	
0	777 (78%)
1	201 (20%)
2	20 (2.0%)
3	1 (0.1%)
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__DOW	
0	929 (93%)
1	65 (6.5%)
2	6 (0.6%)

Query parameter or value	Descriptive statistics
SELECT__n_of__non__aggr__func__RTRIM	
0	792 (79%)
1	185 (19%)
2	20 (2.0%)
3	3 (0.3%)
SELECT__n_of__non__aggr__func__UPPER	
0	786 (79%)
1	183 (18%)
2	30 (3.0%)
4	1 (0.1%)
SELECT__n_of__non__aggr__func__SQRT	
0	881 (88%)
1	106 (11%)
2	12 (1.2%)
4	1 (0.1%)
SELECT__n_of__non__aggr__func__FLOOR	
0	960 (96%)
1	40 (4.0%)
SELECT__n_of__non__aggr__func__LTRIM	
0	811 (81%)
1	165 (17%)
2	22 (2.2%)
3	2 (0.2%)
SELECT__n_of__non__aggr__func__SUBSTR	
0	778 (78%)
1	188 (19%)
2	29 (2.9%)
3	5 (0.5%)
SELECT__n_of__non__aggr__func__DAY	
0	942 (94%)
1	52 (5.2%)
2	5 (0.5%)
3	1 (0.1%)
SELECT__n_of__non__aggr__func__MONTH	
0	905 (91%)
1	79 (7.9%)
2	13 (1.3%)
3	3 (0.3%)
SELECT__n_of__non__aggr__func__LOG	
0	935 (94%)
1	60 (6.0%)
2	4 (0.4%)
3	1 (0.1%)
SELECT__n_of__non__aggr__func__YEAR	
0	932 (93%)
1	60 (6.0%)

Query parameter or value	Descriptive statistics
2	6 (0.6%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___ABS	
0	974 (97%)
1	25 (2.5%)
2	1 (0.1%)
SELECT_n_of_non_aggr_func___ROUND	
0	960 (96%)
1	36 (3.6%)
2	4 (0.4%)
SELECT_n_of_all_non_aggr_func	[0.00, 1.00, 16.00] 1.89 / 2.12
SELECT_n_of_aggr_func___COUNT	
0	794 (79%)
1	92 (9.2%)
2	68 (6.8%)
3	27 (2.7%)
4	10 (1.0%)
5	6 (0.6%)
6	3 (0.3%)
SELECT_n_of_aggr_func___MIN	
0	765 (77%)
1	126 (13%)
2	57 (5.7%)
3	29 (2.9%)
4	14 (1.4%)
5	5 (0.5%)
6	2 (0.2%)
7	2 (0.2%)
SELECT_n_of_aggr_func___MAX	
0	769 (77%)
1	101 (10%)
2	68 (6.8%)
3	37 (3.7%)
4	19 (1.9%)
5	4 (0.4%)
6	2 (0.2%)
SELECT_n_of_aggr_func___AVG	
0	955 (96%)
1	34 (3.4%)
2	11 (1.1%)
SELECT_n_of_aggr_func___COUNT_DISTINCT	
0	776 (78%)
1	107 (11%)
2	68 (6.8%)
3	29 (2.9%)
4	12 (1.2%)

Query parameter or value	Descriptive statistics
5	5 (0.5%)
6	2 (0.2%)
7	1 (0.1%)
SELECT_n_of_aggr_func__SUM	
0	954 (95%)
1	38 (3.8%)
2	7 (0.7%)
3	1 (0.1%)
SELECT_n_of_all_aggr_func	
0	201 (20%)
1	280 (28%)
2	240 (24%)
3	142 (14%)
4	77 (7.7%)
5	38 (3.8%)
6	18 (1.8%)
7	4 (0.4%)
FROM_n_of_join_paths	
1	1,000 (100%)
FROM_n_of_joins__INNER	
0	487 (49%)
1	316 (32%)
2	150 (15%)
3	45 (4.5%)
4	2 (0.2%)
FROM_n_of_joins__RIGHT	
0	496 (50%)
1	338 (34%)
2	130 (13%)
3	35 (3.5%)
4	1 (0.1%)
FROM_n_of_processed_rows	[5, 5,000,323, 38,249,944] 13,808,988 / 15,510,820
WHERE_n_of_predicates	
0	253 (25%)
1	222 (22%)
2	214 (21%)
3	165 (17%)
4	96 (9.6%)
5	41 (4.1%)
6	9 (0.9%)
WHERE_n_of_attribs_of_type__character	
0	816 (82%)
1	166 (17%)
2	16 (1.6%)
3	2 (0.2%)
WHERE_n_of_attribs_of_type__character_varying	

Query parameter or value	Descriptive statistics
0	552 (55%)
1	310 (31%)
2	123 (12%)
3	15 (1.5%)
WHERE_n_of_attribs_of_type__date	
0	902 (90%)
1	88 (8.8%)
2	9 (0.9%)
3	1 (0.1%)
WHERE_n_of_attribs_of_type__integer	
0	555 (56%)
1	319 (32%)
2	104 (10%)
3	21 (2.1%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type__numeric	
0	768 (77%)
1	191 (19%)
2	34 (3.4%)
3	7 (0.7%)
WHERE_n_of_pkey_attribs	
0	706 (71%)
1	245 (25%)
2	43 (4.3%)
3	6 (0.6%)
WHERE_n_of_connect_OR	
0	613 (61%)
1	220 (22%)
2	108 (11%)
3	44 (4.4%)
4	13 (1.3%)
5	2 (0.2%)
WHERE_n_of_operators__in	
0	718 (72%)
1	224 (22%)
2	46 (4.6%)
3	12 (1.2%)
WHERE_n_of_operators__like	
0	900 (90%)
1	93 (9.3%)
2	7 (0.7%)
WHERE_n_of_operators__greater_or_less	
0	552 (55%)
1	289 (29%)
2	121 (12%)
3	28 (2.8%)

Query parameter or value	Descriptive statistics
4	7 (0.7%)
5	3 (0.3%)
WHERE_n_of_operators_between	
0	704 (70%)
1	254 (25%)
2	38 (3.8%)
3	3 (0.3%)
4	1 (0.1%)
WHERE_n_of_non_aggr_func_DOW	
0	986 (99%)
1	14 (1.4%)
WHERE_n_of_non_aggr_func_ABS	
0	983 (98%)
1	17 (1.7%)
WHERE_n_of_non_aggr_func_SQRT	
0	937 (94%)
1	60 (6.0%)
2	3 (0.3%)
WHERE_n_of_non_aggr_func_LOG	
0	962 (96%)
1	37 (3.7%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func_ROUND	
0	980 (98%)
1	20 (2.0%)
WHERE_n_of_non_aggr_func_FLOOR	
0	988 (99%)
1	12 (1.2%)
WHERE_n_of_non_aggr_func_YEAR	
0	987 (99%)
1	12 (1.2%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func_MONTH	
0	982 (98%)
1	18 (1.8%)
WHERE_n_of_non_aggr_func_DAY	
0	982 (98%)
1	18 (1.8%)
WHERE_n_of_non_aggr_func_TRUNC	
0	983 (98%)
1	17 (1.7%)
WHERE_n_of_all_non_aggr_func	
0	795 (80%)
1	178 (18%)
2	24 (2.4%)
3	3 (0.3%)

Query parameter or value	Descriptive statistics
GROUP_BY_n_of_columns	
0	254 (25%)
1	294 (29%)
2	225 (23%)
3	136 (14%)
4	64 (6.4%)
5	26 (2.6%)
6	1 (0.1%)
HAVING_n_of_main_predicates	
0	440 (44%)
1	184 (18%)
2	186 (19%)
3	190 (19%)
ORDER_BY_n_of_columns	
0	301 (30%)
1	485 (49%)
2	187 (19%)
3	19 (1.9%)
4	8 (0.8%)
limit	[1, 500, 998] 500 / 292
offset	[0, 29, 998] 244 / 319