

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

## Scenario: 1-1jp\_\_ng

## Scale factor: 50 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, 43] 10 / 10
SELECT_n_of_non_aggr_func__LOG	
0	889 (89%)
1	100 (10%)
2	11 (1.1%)
SELECT_n_of_non_aggr_func__SUBSTR	
0	790 (79%)
1	162 (16%)
2	38 (3.8%)
3	7 (0.7%)
4	2 (0.2%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func__UPPER	
0	779 (78%)
1	164 (16%)

Query parameter or value	Descriptive statistics
2	44 (4.4%)
3	13 (1.3%)
SELECT_n_of_non_aggr_func___ABS	
0	934 (93%)
1	61 (6.1%)
2	5 (0.5%)
SELECT_n_of_non_aggr_func___SQRT	
0	814 (81%)
1	141 (14%)
2	41 (4.1%)
3	4 (0.4%)
SELECT_n_of_non_aggr_func___TRUNC	
0	947 (95%)
1	46 (4.6%)
2	7 (0.7%)
SELECT_n_of_non_aggr_func___LTRIM	
0	771 (77%)
1	174 (17%)
2	44 (4.4%)
3	9 (0.9%)
4	2 (0.2%)
SELECT_n_of_non_aggr_func___FLOOR	
0	948 (95%)
1	50 (5.0%)
2	2 (0.2%)
SELECT_n_of_non_aggr_func___LOWER	
0	797 (80%)
1	158 (16%)
2	33 (3.3%)
3	12 (1.2%)
SELECT_n_of_non_aggr_func___RTRIM	
0	776 (78%)
1	172 (17%)
2	39 (3.9%)
3	10 (1.0%)
4	3 (0.3%)
SELECT_n_of_non_aggr_func___DAY	
0	918 (92%)
1	67 (6.7%)
2	14 (1.4%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___MONTH	
0	858 (86%)
1	99 (9.9%)
2	32 (3.2%)
3	10 (1.0%)

Query parameter or value	Descriptive statistics
4	1 (0.1%)
SELECT_n_of_non_aggr_func__YEAR	
0	858 (86%)
1	92 (9.2%)
2	45 (4.5%)
3	4 (0.4%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__ROUND	
0	927 (93%)
1	69 (6.9%)
2	4 (0.4%)
SELECT_n_of_non_aggr_func__DOW	
0	893 (89%)
1	84 (8.4%)
2	17 (1.7%)
3	6 (0.6%)
SELECT_n_of_all_non_aggr_func	[0.0, 1.0, 24.0] 2.7 / 3.8
SELECT_n_of_aggr_func__COUNT_DISTINCT	
0	937 (94%)
1	24 (2.4%)
2	20 (2.0%)
3	8 (0.8%)
4	7 (0.7%)
5	2 (0.2%)
6	2 (0.2%)
SELECT_n_of_aggr_func__MAX	
0	936 (94%)
1	23 (2.3%)
2	14 (1.4%)
3	17 (1.7%)
4	5 (0.5%)
5	4 (0.4%)
6	1 (0.1%)
SELECT_n_of_aggr_func__MIN	
0	921 (92%)
1	34 (3.4%)
2	16 (1.6%)
3	15 (1.5%)
4	11 (1.1%)
5	3 (0.3%)
SELECT_n_of_aggr_func__COUNT	
0	934 (93%)
1	28 (2.8%)
2	21 (2.1%)
3	10 (1.0%)
4	5 (0.5%)

Query parameter or value	Descriptive statistics
5	1 (0.1%)
6	1 (0.1%)
SELECT_n_of_aggr_func__SUM	
0	986 (99%)
1	10 (1.0%)
2	4 (0.4%)
SELECT_n_of_aggr_func__AVG	
0	988 (99%)
1	7 (0.7%)
2	4 (0.4%)
3	1 (0.1%)
SELECT_n_of_all_aggr_func	
0	752 (75%)
1	71 (7.1%)
2	68 (6.8%)
3	53 (5.3%)
4	31 (3.1%)
5	19 (1.9%)
6	4 (0.4%)
7	2 (0.2%)
FROM_n_of_join_paths	
1	1,000 (100%)
FROM_n_of_joins__RIGHT	
0	460 (46%)
1	347 (35%)
2	146 (15%)
3	36 (3.6%)
4	11 (1.1%)
FROM_n_of_joins__INNER	
0	482 (48%)
1	329 (33%)
2	143 (14%)
3	43 (4.3%)
4	3 (0.3%)
FROM_n_of_processed_rows	[5, 49,990,245, 382,516,427] 140,504,415 / 157,428,481
WHERE_n_of_predicates	
0	228 (23%)
1	253 (25%)
2	193 (19%)
3	178 (18%)
4	91 (9.1%)
5	42 (4.2%)
6	15 (1.5%)
WHERE_n_of_attribs_of_type__character_varying	
0	523 (52%)

Query parameter or value	Descriptive statistics
1	344 (34%)
2	110 (11%)
3	21 (2.1%)
4	2 (0.2%)
WHERE_n_of_attribs_of_type__integer	
0	563 (56%)
1	334 (33%)
2	84 (8.4%)
3	18 (1.8%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type__character	
0	827 (83%)
1	153 (15%)
2	17 (1.7%)
3	3 (0.3%)
WHERE_n_of_attribs_of_type__numeric	
0	738 (74%)
1	206 (21%)
2	48 (4.8%)
3	7 (0.7%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type__date	
0	890 (89%)
1	101 (10%)
2	9 (0.9%)
WHERE_n_of_pkey_attribs	
0	713 (71%)
1	252 (25%)
2	32 (3.2%)
3	3 (0.3%)
WHERE_n_of_connect_OR	
0	601 (60%)
1	213 (21%)
2	121 (12%)
3	47 (4.7%)
4	15 (1.5%)
5	3 (0.3%)
WHERE_n_of_operators__greater_or_less	
0	539 (54%)
1	302 (30%)
2	125 (13%)
3	26 (2.6%)
4	7 (0.7%)
5	1 (0.1%)
WHERE_n_of_operators__in	
0	711 (71%)

Query parameter or value	Descriptive statistics
1	234 (23%)
2	49 (4.9%)
3	5 (0.5%)
4	1 (0.1%)
WHERE_n_of_operators_like	
0	881 (88%)
1	113 (11%)
2	6 (0.6%)
WHERE_n_of_operators_between	
0	714 (71%)
1	236 (24%)
2	41 (4.1%)
3	8 (0.8%)
4	1 (0.1%)
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_ROUND	
0	979 (98%)
1	21 (2.1%)
WHERE_n_of_non_aggr_func_YEAR	
0	974 (97%)
1	26 (2.6%)
WHERE_n_of_non_aggr_func_DOW	
0	979 (98%)
1	21 (2.1%)
WHERE_n_of_non_aggr_func_TRUNC	
0	971 (97%)
1	29 (2.9%)
WHERE_n_of_non_aggr_func_FLOOR	
0	975 (98%)
1	25 (2.5%)
WHERE_n_of_non_aggr_func_LOG	
0	962 (96%)
1	37 (3.7%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func_ABS	
0	982 (98%)
1	18 (1.8%)
WHERE_n_of_non_aggr_func_MONTH	
0	989 (99%)
1	11 (1.1%)

Query parameter or value	Descriptive statistics
WHERE_n_of_all_non_aggr_func	
0	780 (78%)
1	179 (18%)
2	37 (3.7%)
3	4 (0.4%)
GROUP_BY_n_of_columns	
0	1,000 (100%)
HAVING_n_of_main_predicates	
0	1,000 (100%)
ORDER_BY_n_of_columns	
0	449 (45%)
1	292 (29%)
2	154 (15%)
3	64 (6.4%)
4	29 (2.9%)
5	7 (0.7%)
6	5 (0.5%)
limit	[2, 470, 1,000] 489 / 289
offset	[0, 0, 1,000] 222 / 311