

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

## Scenario: 1-6jp\_rjpp\_\_g\_\_h0ss\_h0nss

## Scale factor: 50 GB

## The 1000-query pack was generated on: 2023-11-12

## 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, 166] 15 / 22
SELECT_n_of_non_aggr_func___DOW	
0	841 (84%)
1	122 (12%)
2	23 (2.3%)
3	7 (0.7%)
4	3 (0.3%)
5	2 (0.2%)
7	1 (0.1%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func___LTRIM	
0	580 (58%)
1	286 (29%)
2	85 (8.5%)
3	21 (2.1%)

Query parameter or value	Descriptive statistics
4	13 (1.3%)
5	9 (0.9%)
6	2 (0.2%)
7	2 (0.2%)
8	2 (0.2%)
SELECT_n_of_non_aggr_func__MONTH	
0	799 (80%)
1	131 (13%)
2	42 (4.2%)
3	14 (1.4%)
4	6 (0.6%)
5	4 (0.4%)
6	2 (0.2%)
7	1 (0.1%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func__SQRT	
0	796 (80%)
1	152 (15%)
2	20 (2.0%)
3	12 (1.2%)
4	14 (1.4%)
5	2 (0.2%)
6	3 (0.3%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func__UPPER	
0	597 (60%)
1	264 (26%)
2	90 (9.0%)
3	28 (2.8%)
4	13 (1.3%)
5	3 (0.3%)
6	3 (0.3%)
7	1 (0.1%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func__YEAR	
0	817 (82%)
1	135 (14%)
2	27 (2.7%)
3	11 (1.1%)
4	4 (0.4%)
5	3 (0.3%)
6	2 (0.2%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func__LOG	
0	873 (87%)
1	98 (9.8%)

Query parameter or value	Descriptive statistics
2	18 (1.8%)
3	6 (0.6%)
4	5 (0.5%)
SELECT_n_of_non_aggr_func__LOWER	
0	597 (60%)
1	283 (28%)
11	1 (0.1%)
2	69 (6.9%)
3	28 (2.8%)
4	9 (0.9%)
5	7 (0.7%)
6	5 (0.5%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func__SUBSTR	
0	619 (62%)
1	265 (27%)
2	82 (8.2%)
3	12 (1.2%)
4	13 (1.3%)
5	3 (0.3%)
6	3 (0.3%)
7	2 (0.2%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func__ROUND	
0	908 (91%)
1	73 (7.3%)
2	15 (1.5%)
3	2 (0.2%)
4	1 (0.1%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func__RTRIM	
0	598 (60%)
1	270 (27%)
2	84 (8.4%)
3	21 (2.1%)
4	13 (1.3%)
5	5 (0.5%)
6	4 (0.4%)
7	3 (0.3%)
8	2 (0.2%)
SELECT_n_of_non_aggr_func__DAY	
0	880 (88%)
1	91 (9.1%)
2	20 (2.0%)
3	4 (0.4%)
4	3 (0.3%)

Query parameter or value	Descriptive statistics
5	2 (0.2%)
SELECT_n_of_non_aggr_func___ABS	
0	927 (93%)
1	62 (6.2%)
2	6 (0.6%)
3	4 (0.4%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___FLOOR	
0	908 (91%)
1	78 (7.8%)
2	11 (1.1%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func___TRUNC	
0	920 (92%)
1	68 (6.8%)
2	10 (1.0%)
3	1 (0.1%)
4	1 (0.1%)
SELECT_n_of_all_non_aggr_func	[0, 3, 66] 5 / 7
SELECT_n_of_aggr_func___MIN	[0.00, 0.00, 12.00] 0.82 / 1.84
SELECT_n_of_aggr_func___COUNT_DISTINCT	[0.00, 0.00, 12.00] 0.84 / 2.02
SELECT_n_of_aggr_func___SUM	
0	935 (94%)
1	41 (4.1%)
2	20 (2.0%)
3	3 (0.3%)
5	1 (0.1%)
SELECT_n_of_aggr_func___COUNT	[0.00, 0.00, 12.00] 0.80 / 1.88
SELECT_n_of_aggr_func___AVG	
0	948 (95%)
1	36 (3.6%)
2	14 (1.4%)
3	2 (0.2%)
SELECT_n_of_aggr_func___MAX	[0.00, 0.00, 13.00] 0.98 / 2.09
SELECT_n_of_all_aggr_func	[0.0, 3.0, 14.0] 3.6 / 3.1
FROM_n_of_join_paths	
1	157 (16%)
2	158 (16%)
3	172 (17%)
4	174 (17%)
5	170 (17%)
6	169 (17%)
FROM_n_of_super_joins___FULL	
0	532 (53%)
1	324 (32%)
2	114 (11%)

Query parameter or value	Descriptive statistics
3	26 (2.6%)
4	2 (0.2%)
5	2 (0.2%)
FROM_n_of_super_joins__RIGHT	
0	560 (56%)
1	308 (31%)
2	99 (9.9%)
3	30 (3.0%)
4	3 (0.3%)
FROM_n_of_super_joins__LEFT	
0	530 (53%)
1	333 (33%)
2	104 (10%)
3	27 (2.7%)
4	6 (0.6%)
FROM_n_of_joins__INNER	[0.00, 3.00, 13.00] 3.32 / 2.37
FROM_n_of_joins__RIGHT	[0.00, 3.00, 12.00] 3.42 / 2.56
FROM_n_of_processed_rows	[5, 465,031,240, 2,178,024,940] 590,479,503 / 486,382,553
WHERE_n_of_predicates	[0.00, 3.00, 12.00] 3.87 / 2.95
WHERE_n_of_attribs_of_type__character_varying	
0	333 (33%)
1	301 (30%)
2	194 (19%)
3	105 (11%)
4	43 (4.3%)
5	20 (2.0%)
6	3 (0.3%)
7	1 (0.1%)
WHERE_n_of_attribs_of_type__date	
0	792 (79%)
1	168 (17%)
2	35 (3.5%)
3	5 (0.5%)
WHERE_n_of_attribs_of_type__integer	
0	362 (36%)
1	316 (32%)
2	165 (17%)
3	93 (9.3%)
4	51 (5.1%)
5	9 (0.9%)
6	4 (0.4%)
WHERE_n_of_attribs_of_type__numeric	
0	551 (55%)
1	295 (30%)
2	117 (12%)

Query parameter or value	Descriptive statistics
3	28 (2.8%)
4	7 (0.7%)
5	2 (0.2%)
WHERE_n_of_attribs_of_type__character	
0	657 (66%)
1	247 (25%)
2	76 (7.6%)
3	15 (1.5%)
4	4 (0.4%)
5	1 (0.1%)
WHERE_n_of_pkey_attribs	
0	545 (55%)
1	279 (28%)
2	120 (12%)
3	47 (4.7%)
4	9 (0.9%)
WHERE_n_of_connect_OR	[0.00, 1.00, 11.00] 2.25 / 2.43
WHERE_n_of_operators__between	
0	532 (53%)
1	279 (28%)
2	128 (13%)
3	44 (4.4%)
4	15 (1.5%)
5	2 (0.2%)
WHERE_n_of_operators__greater_or_less	
0	353 (35%)
1	246 (25%)
2	182 (18%)
3	124 (12%)
4	53 (5.3%)
5	33 (3.3%)
6	7 (0.7%)
7	2 (0.2%)
WHERE_n_of_operators__in	
0	521 (52%)
1	293 (29%)
2	129 (13%)
3	48 (4.8%)
4	9 (0.9%)
WHERE_n_of_operators__like	
0	784 (78%)
1	184 (18%)
2	30 (3.0%)
3	2 (0.2%)
WHERE_n_of_non_aggr_func__DAY	
0	960 (96%)

Query parameter or value	Descriptive statistics
1	37 (3.7%)
2	3 (0.3%)
WHERE_n_of_non_aggr_func__SQRT	
0	890 (89%)
1	101 (10%)
2	7 (0.7%)
3	2 (0.2%)
WHERE_n_of_non_aggr_func__LOG	
0	939 (94%)
1	58 (5.8%)
2	2 (0.2%)
3	1 (0.1%)
WHERE_n_of_non_aggr_func__FLOOR	
0	958 (96%)
1	42 (4.2%)
WHERE_n_of_non_aggr_func__DOW	
0	969 (97%)
1	28 (2.8%)
2	3 (0.3%)
WHERE_n_of_non_aggr_func__ABS	
0	959 (96%)
1	40 (4.0%)
3	1 (0.1%)
WHERE_n_of_non_aggr_func__ROUND	
0	957 (96%)
1	41 (4.1%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func__MONTH	
0	963 (96%)
1	37 (3.7%)
WHERE_n_of_non_aggr_func__YEAR	
0	964 (96%)
1	35 (3.5%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__TRUNC	
0	964 (96%)
1	35 (3.5%)
2	1 (0.1%)
WHERE_n_of_all_non_aggr_func	
0	655 (66%)
1	227 (23%)
2	86 (8.6%)
3	27 (2.7%)
4	2 (0.2%)
5	2 (0.2%)
6	1 (0.1%)

Query parameter or value	Descriptive statistics
GROUP_BY_n_of_columns	[0.00, 3.00, 12.00] 3.27 / 2.92
HAVING_n_of_main_predicates	
0	438 (44%)
1	178 (18%)
2	176 (18%)
3	208 (21%)
ORDER_BY_n_of_columns	[0.00, 1.00, 9.00] 1.30 / 1.20
limit	[2, 499, 1,000] 502 / 282
offset	[0, 102, 1,000] 283 / 335