

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

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

## Scale factor: 2000 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 / 5.1
SELECT_n_of_non_aggr_func__RTRIM	
0	778 (78%)
1	182 (18%)
2	37 (3.7%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func__LOG	
0	932 (93%)
1	61 (6.1%)
2	5 (0.5%)
3	1 (0.1%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__LTRIM	
0	783 (78%)
1	180 (18%)

Query parameter or value	Descriptive statistics
2	34 (3.4%)
3	3 (0.3%)
SELECT__n_of__non__aggr__func__SUBSTR	
0	794 (79%)
1	187 (19%)
2	15 (1.5%)
3	3 (0.3%)
4	1 (0.1%)
SELECT__n_of__non__aggr__func__SQRT	
0	883 (88%)
1	106 (11%)
2	11 (1.1%)
SELECT__n_of__non__aggr__func__FLOOR	
0	973 (97%)
1	27 (2.7%)
SELECT__n_of__non__aggr__func__LOWER	
0	785 (79%)
1	190 (19%)
2	19 (1.9%)
3	4 (0.4%)
4	1 (0.1%)
5	1 (0.1%)
SELECT__n_of__non__aggr__func__DAY	
0	949 (95%)
1	46 (4.6%)
2	5 (0.5%)
SELECT__n_of__non__aggr__func__MONTH	
0	901 (90%)
1	82 (8.2%)
2	17 (1.7%)
SELECT__n_of__non__aggr__func__YEAR	
0	909 (91%)
1	75 (7.5%)
2	15 (1.5%)
3	1 (0.1%)
SELECT__n_of__non__aggr__func__TRUNC	
0	956 (96%)
1	41 (4.1%)
2	2 (0.2%)
3	1 (0.1%)
SELECT__n_of__non__aggr__func__DOW	
0	927 (93%)
1	67 (6.7%)
2	6 (0.6%)
SELECT__n_of__non__aggr__func__UPPER	
0	785 (79%)

Query parameter or value	Descriptive statistics
1	179 (18%)
2	35 (3.5%)
3	1 (0.1%)
SELECT__n_of__non_aggr_func__ROUND	
0	971 (97%)
1	29 (2.9%)
SELECT__n_of__non_aggr_func__ABS	
0	959 (96%)
1	39 (3.9%)
2	2 (0.2%)
SELECT__n_of__all__non_aggr_func	[0.00, 1.00, 20.00] 1.97 / 2.29
SELECT__n_of__aggr_func__MAX	
0	787 (79%)
1	100 (10%)
2	60 (6.0%)
3	30 (3.0%)
4	14 (1.4%)
5	5 (0.5%)
6	3 (0.3%)
7	1 (0.1%)
SELECT__n_of__aggr_func__MIN	
0	764 (76%)
1	121 (12%)
2	75 (7.5%)
3	27 (2.7%)
4	10 (1.0%)
5	3 (0.3%)
SELECT__n_of__aggr_func__COUNT_DISTINCT	
0	785 (79%)
1	98 (9.8%)
2	59 (5.9%)
3	37 (3.7%)
4	14 (1.4%)
5	4 (0.4%)
6	2 (0.2%)
7	1 (0.1%)
SELECT__n_of__aggr_func__AVG	
0	944 (94%)
1	43 (4.3%)
2	10 (1.0%)
3	3 (0.3%)
SELECT__n_of__aggr_func__COUNT	
0	781 (78%)
1	116 (12%)
2	47 (4.7%)
3	34 (3.4%)

Query parameter or value	Descriptive statistics
4	10 (1.0%)
5	10 (1.0%)
6	2 (0.2%)
SELECT__n_of__aggr__func__SUM	
0	952 (95%)
1	37 (3.7%)
2	10 (1.0%)
3	1 (0.1%)
SELECT__n_of__all__aggr__func	
0	189 (19%)
1	302 (30%)
2	241 (24%)
3	145 (15%)
4	67 (6.7%)
5	37 (3.7%)
6	14 (1.4%)
7	5 (0.5%)
FROM__n_of__join__paths	
1	1,000 (100%)
FROM__n_of__joins__RIGHT	
0	524 (52%)
1	317 (32%)
2	126 (13%)
3	29 (2.9%)
4	4 (0.4%)
FROM__n_of__joins__INNER	
0	501 (50%)
1	352 (35%)
2	107 (11%)
3	36 (3.6%)
4	4 (0.4%)
FROM__n_of__processed__rows	[5, 2,000,000,000, 15,300,027,829] 5,580,489,407 / 6,233,847,794
WHERE__n_of__predicates	
0	230 (23%)
1	248 (25%)
2	218 (22%)
3	142 (14%)
4	109 (11%)
5	43 (4.3%)
6	10 (1.0%)
WHERE__n_of__attribs__of__type__character	
0	815 (82%)
1	156 (16%)
2	27 (2.7%)
3	2 (0.2%)

Query parameter or value	Descriptive statistics
WHERE_n_of_attribs_of_type__numeric	
0	745 (75%)
1	210 (21%)
2	42 (4.2%)
3	2 (0.2%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type__character_varying	
0	535 (54%)
1	335 (34%)
2	108 (11%)
3	20 (2.0%)
4	2 (0.2%)
WHERE_n_of_attribs_of_type__integer	
0	562 (56%)
1	324 (32%)
2	99 (9.9%)
3	12 (1.2%)
4	3 (0.3%)
WHERE_n_of_attribs_of_type__date	
0	897 (90%)
1	95 (9.5%)
2	7 (0.7%)
3	1 (0.1%)
WHERE_n_of_pkey_attribs	
0	716 (72%)
1	250 (25%)
2	33 (3.3%)
3	1 (0.1%)
WHERE_n_of_connect_OR	
0	598 (60%)
1	210 (21%)
2	112 (11%)
3	60 (6.0%)
4	15 (1.5%)
5	5 (0.5%)
WHERE_n_of_operators__greater_or_less	
0	527 (53%)
1	322 (32%)
2	103 (10%)
3	37 (3.7%)
4	10 (1.0%)
5	1 (0.1%)
WHERE_n_of_operators__between	
0	703 (70%)
1	243 (24%)
2	48 (4.8%)

Query parameter or value	Descriptive statistics
3	5 (0.5%)
4	1 (0.1%)
WHERE_n_of_operators_in	
0	718 (72%)
1	230 (23%)
2	44 (4.4%)
3	7 (0.7%)
4	1 (0.1%)
WHERE_n_of_operators_like	
0	884 (88%)
1	112 (11%)
2	4 (0.4%)
WHERE_n_of_non_aggr_func_SQRT	
0	941 (94%)
1	58 (5.8%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func_TRUNC	
0	978 (98%)
1	22 (2.2%)
WHERE_n_of_non_aggr_func_DAY	
0	979 (98%)
1	21 (2.1%)
WHERE_n_of_non_aggr_func_LOG	
0	968 (97%)
1	31 (3.1%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func_ROUND	
0	984 (98%)
1	16 (1.6%)
WHERE_n_of_non_aggr_func_MONTH	
0	984 (98%)
1	16 (1.6%)
WHERE_n_of_non_aggr_func_ABS	
0	985 (99%)
1	15 (1.5%)
WHERE_n_of_non_aggr_func_DOW	
0	984 (98%)
1	16 (1.6%)
WHERE_n_of_non_aggr_func_FLOOR	
0	982 (98%)
1	18 (1.8%)
WHERE_n_of_non_aggr_func_YEAR	
0	982 (98%)
1	18 (1.8%)
WHERE_n_of_all_non_aggr_func	
0	793 (79%)

Query parameter or value	Descriptive statistics
<i>1</i>	182 (18%)
<i>2</i>	22 (2.2%)
<i>3</i>	3 (0.3%)
GROUP_BY_n_of_columns	
<i>0</i>	253 (25%)
<i>1</i>	274 (27%)
<i>2</i>	241 (24%)
<i>3</i>	119 (12%)
<i>4</i>	66 (6.6%)
<i>5</i>	41 (4.1%)
<i>6</i>	5 (0.5%)
<i>7</i>	1 (0.1%)
HAVING_n_of_main_predicates	
<i>0</i>	426 (43%)
<i>1</i>	204 (20%)
<i>2</i>	195 (20%)
<i>3</i>	175 (18%)
ORDER_BY_n_of_columns	
<i>0</i>	285 (29%)
<i>1</i>	468 (47%)
<i>2</i>	217 (22%)
<i>3</i>	24 (2.4%)
<i>4</i>	5 (0.5%)
<i>5</i>	1 (0.1%)
limit	[1, 475, 1,000] 493 / 288
offset	[0, 13, 1,000] 254 / 331