

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

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

## Scale factor: 100 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___ABS	
0	945 (95%)
1	50 (5.0%)
2	5 (0.5%)
SELECT_n_of_non_aggr_func___UPPER	
0	801 (80%)
1	152 (15%)
2	35 (3.5%)
3	10 (1.0%)
4	2 (0.2%)
SELECT_n_of_non_aggr_func___SQRT	
0	837 (84%)
1	128 (13%)
2	29 (2.9%)

Query parameter or value	Descriptive statistics
3	5 (0.5%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__SUBSTR	
0	782 (78%)
1	165 (17%)
2	41 (4.1%)
3	10 (1.0%)
4	2 (0.2%)
SELECT_n_of_non_aggr_func__YEAR	
0	879 (88%)
1	80 (8.0%)
2	37 (3.7%)
3	3 (0.3%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__DAY	
0	917 (92%)
1	69 (6.9%)
2	11 (1.1%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func__DOW	
0	891 (89%)
1	87 (8.7%)
2	18 (1.8%)
3	3 (0.3%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__LOG	
0	892 (89%)
1	87 (8.7%)
2	18 (1.8%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func__LOWER	
0	797 (80%)
1	155 (16%)
2	41 (4.1%)
3	6 (0.6%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__LTRIM	
0	790 (79%)
1	156 (16%)
2	43 (4.3%)
3	11 (1.1%)
SELECT_n_of_non_aggr_func__MONTH	
0	850 (85%)
1	109 (11%)
2	34 (3.4%)
3	7 (0.7%)

Query parameter or value	Descriptive statistics
SELECT_n_of_non_aggr_func___RTRIM	
0	799 (80%)
1	150 (15%)
2	42 (4.2%)
3	6 (0.6%)
4	3 (0.3%)
SELECT_n_of_non_aggr_func___FLOOR	
0	929 (93%)
1	65 (6.5%)
2	6 (0.6%)
SELECT_n_of_non_aggr_func___TRUNC	
0	939 (94%)
1	58 (5.8%)
2	2 (0.2%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___ROUND	
0	938 (94%)
1	58 (5.8%)
2	4 (0.4%)
SELECT_n_of_all_non_aggr_func	[0.0, 1.0, 22.0] 2.5 / 3.6
SELECT_n_of_aggr_func___MAX	
0	927 (93%)
1	39 (3.9%)
2	18 (1.8%)
3	7 (0.7%)
4	7 (0.7%)
5	2 (0.2%)
SELECT_n_of_aggr_func___MIN	
0	926 (93%)
1	35 (3.5%)
2	17 (1.7%)
3	10 (1.0%)
4	8 (0.8%)
5	2 (0.2%)
7	1 (0.1%)
8	1 (0.1%)
SELECT_n_of_aggr_func___COUNT_DISTINCT	
0	931 (93%)
1	29 (2.9%)
2	21 (2.1%)
3	12 (1.2%)
4	4 (0.4%)
5	2 (0.2%)
7	1 (0.1%)
SELECT_n_of_aggr_func___AVG	
0	985 (99%)

Query parameter or value	Descriptive statistics
1	13 (1.3%)
2	1 (0.1%)
3	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__COUNT	
0	958 (96%)
1	19 (1.9%)
2	12 (1.2%)
3	4 (0.4%)
4	6 (0.6%)
5	1 (0.1%)
SELECT_n_of_all_aggr_func	
0	766 (77%)
1	80 (8.0%)
2	67 (6.7%)
3	40 (4.0%)
4	28 (2.8%)
5	14 (1.4%)
6	2 (0.2%)
7	2 (0.2%)
8	1 (0.1%)
FROM_n_of_join_paths	
1	1,000 (100%)
FROM_n_of_joins__INNER	
0	487 (49%)
1	339 (34%)
2	139 (14%)
3	30 (3.0%)
4	5 (0.5%)
FROM_n_of_joins__RIGHT	
0	503 (50%)
1	339 (34%)
2	115 (12%)
3	37 (3.7%)
4	6 (0.6%)
FROM_n_of_processed_rows	[5, 100,000,000, 765,037,932] 282,155,976 / 310,637,648
WHERE_n_of_predicates	
0	250 (25%)
1	236 (24%)
2	201 (20%)
3	158 (16%)
4	101 (10%)

Query parameter or value	Descriptive statistics
5	44 (4.4%)
6	10 (1.0%)
WHERE_n_of_attribs_of_type__character_varying	
0	561 (56%)
1	329 (33%)
2	95 (9.5%)
3	15 (1.5%)
WHERE_n_of_attribs_of_type__integer	
0	549 (55%)
1	338 (34%)
2	89 (8.9%)
3	21 (2.1%)
4	3 (0.3%)
WHERE_n_of_attribs_of_type__numeric	
0	746 (75%)
1	211 (21%)
2	35 (3.5%)
3	8 (0.8%)
WHERE_n_of_attribs_of_type__date	
0	879 (88%)
1	108 (11%)
2	11 (1.1%)
3	2 (0.2%)
WHERE_n_of_attribs_of_type__character	
0	822 (82%)
1	157 (16%)
2	20 (2.0%)
3	1 (0.1%)
WHERE_n_of_pkey_attribs	
0	713 (71%)
1	248 (25%)
2	37 (3.7%)
3	2 (0.2%)
WHERE_n_of_connect_OR	
0	610 (61%)
1	208 (21%)
2	104 (10%)
3	64 (6.4%)
4	11 (1.1%)
5	3 (0.3%)
WHERE_n_of_operators__greater_or_less	
0	530 (53%)
1	312 (31%)
2	117 (12%)
3	34 (3.4%)
4	6 (0.6%)

Query parameter or value	Descriptive statistics
5	1 (0.1%)
WHERE_n_of_operators__like	
0	900 (90%)
1	98 (9.8%)
2	2 (0.2%)
WHERE_n_of_operators__in	
0	740 (74%)
1	217 (22%)
2	38 (3.8%)
3	5 (0.5%)
WHERE_n_of_operators__between	
0	710 (71%)
1	241 (24%)
2	41 (4.1%)
3	8 (0.8%)
WHERE_n_of_non_aggr_func__TRUNC	
0	973 (97%)
1	26 (2.6%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__MONTH	
0	985 (99%)
1	15 (1.5%)
WHERE_n_of_non_aggr_func__SQRT	
0	949 (95%)
1	51 (5.1%)
WHERE_n_of_non_aggr_func__FLOOR	
0	980 (98%)
1	20 (2.0%)
WHERE_n_of_non_aggr_func__DAY	
0	980 (98%)
1	20 (2.0%)
WHERE_n_of_non_aggr_func__YEAR	
0	980 (98%)
1	19 (1.9%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__LOG	
0	948 (95%)
1	51 (5.1%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__ABS	
0	982 (98%)
1	16 (1.6%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func__DOW	
0	979 (98%)
1	20 (2.0%)

Query parameter or value	Descriptive statistics
<i>2</i>	1 (0.1%)
WHERE_n_of_non_aggr_func__ROUND	
<i>0</i>	985 (99%)
<i>1</i>	15 (1.5%)
WHERE_n_of_all_non_aggr_func	
<i>0</i>	782 (78%)
<i>1</i>	178 (18%)
<i>2</i>	34 (3.4%)
<i>3</i>	5 (0.5%)
<i>4</i>	1 (0.1%)
GROUP_BY_n_of_columns	
<i>0</i>	1,000 (100%)
HAVING_n_of_main_predicates	
<i>0</i>	1,000 (100%)
ORDER_BY_n_of_columns	
<i>0</i>	440 (44%)
<i>1</i>	276 (28%)
<i>2</i>	169 (17%)
<i>3</i>	78 (7.8%)
<i>4</i>	27 (2.7%)
<i>5</i>	7 (0.7%)
<i>6</i>	3 (0.3%)
limit	[3, 478, 1,000] 484 / 290
offset	[0, 0, 1,000] 254 / 324