

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

## Scenario: 1-1jp\_\_g\_\_h0-3ss\_h0-3nss

## Scale factor: 50 GB

## The 1000-query pack was generated on: 2023-11-05

## 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] 5.8 / 5.9
SELECT_n_of_non_aggr_func__LTRIM	
0	790 (79%)
1	190 (19%)
2	15 (1.5%)
3	4 (0.4%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__RTRIM	
0	768 (77%)
1	201 (20%)
2	25 (2.5%)
3	6 (0.6%)
SELECT_n_of_non_aggr_func__UPPER	
0	781 (78%)
1	183 (18%)

Query parameter or value	Descriptive statistics
2	34 (3.4%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func__LOG	
0	944 (94%)
1	50 (5.0%)
2	3 (0.3%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__DOW	
0	939 (94%)
1	54 (5.4%)
2	7 (0.7%)
SELECT_n_of_non_aggr_func__MONTH	
0	888 (89%)
1	88 (8.8%)
2	22 (2.2%)
3	1 (0.1%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__YEAR	
0	921 (92%)
1	64 (6.4%)
2	14 (1.4%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func__LOWER	
0	782 (78%)
1	197 (20%)
2	18 (1.8%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__SUBSTR	
0	794 (79%)
1	178 (18%)
2	20 (2.0%)
3	8 (0.8%)
SELECT_n_of_non_aggr_func__FLOOR	
0	970 (97%)
1	30 (3.0%)
SELECT_n_of_non_aggr_func__SQRT	
0	892 (89%)
1	96 (9.6%)
2	9 (0.9%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func__DAY	
0	933 (93%)
1	64 (6.4%)
2	3 (0.3%)

Query parameter or value	Descriptive statistics
SELECT_n_of_non_aggr_func___ABS	
0	972 (97%)
1	28 (2.8%)
SELECT_n_of_non_aggr_func___ROUND	
0	971 (97%)
1	27 (2.7%)
2	2 (0.2%)
SELECT_n_of_non_aggr_func___TRUNC	
0	974 (97%)
1	23 (2.3%)
2	3 (0.3%)
SELECT_n_of_all_non_aggr_func	[0.00, 1.00, 19.00] 1.93 / 2.37
SELECT_n_of_aggr_func___COUNT_DISTINCT	
0	792 (79%)
1	97 (9.7%)
2	66 (6.6%)
3	33 (3.3%)
4	7 (0.7%)
5	4 (0.4%)
6	1 (0.1%)
SELECT_n_of_aggr_func___MAX	
0	772 (77%)
1	101 (10%)
2	77 (7.7%)
3	27 (2.7%)
4	10 (1.0%)
5	12 (1.2%)
6	1 (0.1%)
SELECT_n_of_aggr_func___COUNT	
0	787 (79%)
1	105 (11%)
2	59 (5.9%)
3	30 (3.0%)
4	12 (1.2%)
5	6 (0.6%)
6	1 (0.1%)
SELECT_n_of_aggr_func___MIN	
0	789 (79%)
1	104 (10%)
2	56 (5.6%)
3	35 (3.5%)
4	10 (1.0%)
5	3 (0.3%)
6	3 (0.3%)
SELECT_n_of_aggr_func___SUM	
0	954 (95%)

Query parameter or value	Descriptive statistics
1	40 (4.0%)
2	5 (0.5%)
3	1 (0.1%)
SELECT__n_of__aggr__func__AVG	
0	944 (94%)
1	51 (5.1%)
2	4 (0.4%)
3	1 (0.1%)
SELECT__n_of__all__aggr__func	
0	208 (21%)
1	289 (29%)
2	243 (24%)
3	151 (15%)
4	59 (5.9%)
5	37 (3.7%)
6	10 (1.0%)
7	3 (0.3%)
FROM__n_of__join__paths	
1	1,000 (100%)
FROM__n_of__joins__RIGHT	
0	512 (51%)
1	350 (35%)
2	112 (11%)
3	22 (2.2%)
4	4 (0.4%)
FROM__n_of__joins__INNER	
0	487 (49%)
1	319 (32%)
2	155 (16%)
3	35 (3.5%)
4	4 (0.4%)
FROM__n_of__processed__rows	[5, 49,990,245, 382,516,427] 138,474,246 / 156,041,046
WHERE__n_of__predicates	
0	249 (25%)
1	217 (22%)
2	211 (21%)
3	162 (16%)
4	106 (11%)
5	41 (4.1%)
6	14 (1.4%)
WHERE__n_of__attrs_of__type__character__varying	
0	526 (53%)
1	322 (32%)
2	125 (13%)
3	25 (2.5%)

Query parameter or value	Descriptive statistics
4	2 (0.2%)
WHERE_n_of_attribs_of_type__numeric	
0	750 (75%)
1	210 (21%)
2	38 (3.8%)
3	2 (0.2%)
WHERE_n_of_attribs_of_type__character	
0	796 (80%)
1	181 (18%)
2	21 (2.1%)
3	2 (0.2%)
WHERE_n_of_attribs_of_type__integer	
0	581 (58%)
1	308 (31%)
2	95 (9.5%)
3	14 (1.4%)
4	2 (0.2%)
WHERE_n_of_attribs_of_type__date	
0	902 (90%)
1	83 (8.3%)
2	14 (1.4%)
3	1 (0.1%)
WHERE_n_of_pkey_attribs	
0	725 (73%)
1	238 (24%)
2	36 (3.6%)
3	1 (0.1%)
WHERE_n_of_connect_OR	
0	614 (61%)
1	191 (19%)
2	125 (13%)
3	54 (5.4%)
4	13 (1.3%)
5	3 (0.3%)
WHERE_n_of_operators__between	
0	693 (69%)
1	255 (26%)
2	43 (4.3%)
3	9 (0.9%)
WHERE_n_of_operators__in	
0	712 (71%)
1	234 (23%)
2	45 (4.5%)
3	8 (0.8%)
4	1 (0.1%)
WHERE_n_of_operators__greater_or_less	

Query parameter or value	Descriptive statistics
0	556 (56%)
1	284 (28%)
2	112 (11%)
3	44 (4.4%)
4	4 (0.4%)
WHERE_n_of_operators__like	
0	887 (89%)
1	105 (11%)
2	8 (0.8%)
WHERE_n_of_non_aggr_func__ROUND	
0	978 (98%)
1	22 (2.2%)
WHERE_n_of_non_aggr_func__YEAR	
0	985 (99%)
1	14 (1.4%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__ABS	
0	979 (98%)
1	20 (2.0%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__FLOOR	
0	982 (98%)
1	18 (1.8%)
WHERE_n_of_non_aggr_func__MONTH	
0	986 (99%)
1	14 (1.4%)
WHERE_n_of_non_aggr_func__DOW	
0	980 (98%)
1	20 (2.0%)
WHERE_n_of_non_aggr_func__SQRT	
0	940 (94%)
1	58 (5.8%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func__DAY	
0	986 (99%)
1	12 (1.2%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func__LOG	
0	969 (97%)
1	31 (3.1%)
WHERE_n_of_non_aggr_func__TRUNC	
0	985 (99%)
1	15 (1.5%)
WHERE_n_of_all_non_aggr_func	
0	802 (80%)
1	165 (17%)

Query parameter or value	Descriptive statistics
2	29 (2.9%)
3	3 (0.3%)
4	1 (0.1%)
GROUP_BY_n_of_columns	
0	267 (27%)
1	272 (27%)
2	222 (22%)
3	155 (16%)
4	49 (4.9%)
5	31 (3.1%)
6	4 (0.4%)
HAVING_n_of_main_predicates	[0.00, 3.00, 9.00] 3.21 / 2.54
HAVING_n_of_main_predicates__non_scalar_subquery	
0	476 (48%)
1	176 (18%)
2	179 (18%)
3	169 (17%)
HAVING_n_of_main_predicates__scalar_subquery	
0	451 (45%)
1	180 (18%)
2	190 (19%)
3	179 (18%)
HAVING_n_of_subqueries__non_scalar_subquery	
0	476 (48%)
1	176 (18%)
2	179 (18%)
3	169 (17%)
HAVING_n_of_subqueries__scalar_subquery	[0.0, 1.0, 21.0] 2.5 / 3.7
HAVING_n_of_processed_rows_by_subqueries	[0, 15,000,058, 8,797,877,131] 513,301,288 / 1,120,929,355
ORDER_BY_n_of_columns	
0	316 (32%)
1	457 (46%)
2	196 (20%)
3	21 (2.1%)
4	8 (0.8%)
5	2 (0.2%)
limit	[2, 490, 1,000] 495 / 289
offset	[0, 2, 999] 251 / 324