

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

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

## Scale factor: 2 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] 6.0 / 6.0
SELECT_n_of_non_aggr_func__LTRIM	
0	797 (80%)
1	179 (18%)
2	21 (2.1%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func__ROUND	
0	956 (96%)
1	43 (4.3%)
2	1 (0.1%)
SELECT_n_of_non_aggr_func__RTRIM	
0	778 (78%)
1	186 (19%)
2	30 (3.0%)
3	4 (0.4%)

Query parameter or value	Descriptive statistics
4	1 (0.1%)
5	1 (0.1%)
SELECT__n_of__non__aggr__func__SQRT	
0	885 (89%)
1	105 (11%)
2	8 (0.8%)
3	2 (0.2%)
SELECT__n_of__non__aggr__func__SUBSTR	
0	772 (77%)
1	191 (19%)
2	30 (3.0%)
3	5 (0.5%)
4	2 (0.2%)
SELECT__n_of__non__aggr__func__TRUNC	
0	957 (96%)
1	38 (3.8%)
2	4 (0.4%)
3	1 (0.1%)
SELECT__n_of__non__aggr__func__DAY	
0	939 (94%)
1	55 (5.5%)
2	6 (0.6%)
SELECT__n_of__non__aggr__func__LOWER	
0	767 (77%)
1	203 (20%)
2	26 (2.6%)
3	3 (0.3%)
5	1 (0.1%)
SELECT__n_of__non__aggr__func__UPPER	
0	799 (80%)
1	172 (17%)
2	25 (2.5%)
3	4 (0.4%)
SELECT__n_of__non__aggr__func__LOG	
0	928 (93%)
1	66 (6.6%)
2	5 (0.5%)
3	1 (0.1%)
SELECT__n_of__non__aggr__func__FLOOR	
0	959 (96%)
1	39 (3.9%)
2	1 (0.1%)
3	1 (0.1%)
SELECT__n_of__non__aggr__func__ABS	
0	963 (96%)
1	35 (3.5%)

Query parameter or value	Descriptive statistics
2	1 (0.1%)
3	1 (0.1%)
SELECT__n_of__non__aggr__func____DOW	
0	927 (93%)
1	64 (6.4%)
2	9 (0.9%)
SELECT__n_of__non__aggr__func____MONTH	
0	891 (89%)
1	98 (9.8%)
2	8 (0.8%)
3	2 (0.2%)
4	1 (0.1%)
SELECT__n_of__non__aggr__func____YEAR	
0	918 (92%)
1	72 (7.2%)
2	9 (0.9%)
3	1 (0.1%)
SELECT__n_of__all__non__aggr__func	[0.00, 2.00, 18.00] 2.02 / 2.29
SELECT__n_of__aggr__func____COUNT	
0	758 (76%)
1	109 (11%)
2	73 (7.3%)
3	37 (3.7%)
4	14 (1.4%)
5	4 (0.4%)
6	1 (0.1%)
7	2 (0.2%)
8	2 (0.2%)
SELECT__n_of__aggr__func____MIN	
0	772 (77%)
1	102 (10%)
2	73 (7.3%)
3	33 (3.3%)
4	14 (1.4%)
5	5 (0.5%)
6	1 (0.1%)
SELECT__n_of__aggr__func____COUNT_DISTINCT	
0	793 (79%)
1	94 (9.4%)
2	56 (5.6%)
3	35 (3.5%)
4	15 (1.5%)
5	7 (0.7%)
SELECT__n_of__aggr__func____MAX	
0	754 (75%)
1	118 (12%)

Query parameter or value	Descriptive statistics
2	65 (6.5%)
3	36 (3.6%)
4	19 (1.9%)
5	6 (0.6%)
6	1 (0.1%)
8	1 (0.1%)
SELECT_n_of_aggr_func__SUM	
0	956 (96%)
1	38 (3.8%)
2	6 (0.6%)
SELECT_n_of_aggr_func__AVG	
0	948 (95%)
1	51 (5.1%)
2	1 (0.1%)
SELECT_n_of_all_aggr_func	
0	179 (18%)
1	271 (27%)
2	248 (25%)
3	168 (17%)
4	74 (7.4%)
5	39 (3.9%)
6	15 (1.5%)
7	3 (0.3%)
8	3 (0.3%)
FROM_n_of_join_paths	
1	1,000 (100%)
FROM_n_of_joins__INNER	
0	478 (48%)
1	322 (32%)
2	165 (17%)
3	30 (3.0%)
4	5 (0.5%)
FROM_n_of_joins__RIGHT	
0	514 (51%)
1	300 (30%)
2	145 (15%)
3	35 (3.5%)
4	6 (0.6%)
FROM_n_of_processed_rows	[5, 2,999,811, 15,295,747] 5,896,773 / 6,272,610
WHERE_n_of_predicates	
0	228 (23%)
1	222 (22%)
2	229 (23%)
3	165 (17%)
4	101 (10%)
5	47 (4.7%)

Query parameter or value	Descriptive statistics
6	8 (0.8%)
WHERE_n_of_attribs_of_type__character_varying	
0	553 (55%)
1	308 (31%)
2	112 (11%)
3	25 (2.5%)
4	1 (0.1%)
5	1 (0.1%)
WHERE_n_of_attribs_of_type__numeric	
0	723 (72%)
1	237 (24%)
2	39 (3.9%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type__integer	
0	547 (55%)
1	329 (33%)
2	99 (9.9%)
3	24 (2.4%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type__date	
0	887 (89%)
1	103 (10%)
2	9 (0.9%)
3	1 (0.1%)
WHERE_n_of_attribs_of_type__character	
0	826 (83%)
1	148 (15%)
2	26 (2.6%)
WHERE_n_of_pkey_attribs	
0	700 (70%)
1	254 (25%)
2	40 (4.0%)
3	6 (0.6%)
WHERE_n_of_connect_OR	
0	606 (61%)
1	218 (22%)
2	103 (10%)
3	56 (5.6%)
4	13 (1.3%)
5	4 (0.4%)
WHERE_n_of_operators__in	
0	725 (73%)
1	219 (22%)
2	52 (5.2%)
3	3 (0.3%)
4	1 (0.1%)

Query parameter or value	Descriptive statistics
WHERE_n_of_operators__between	
0	690 (69%)
1	259 (26%)
2	47 (4.7%)
3	4 (0.4%)
WHERE_n_of_operators__greater_or_less	
0	511 (51%)
1	330 (33%)
2	121 (12%)
3	33 (3.3%)
4	5 (0.5%)
WHERE_n_of_operators__like	
0	880 (88%)
1	113 (11%)
2	7 (0.7%)
WHERE_n_of_non_aggr_func__TRUNC	
0	980 (98%)
1	20 (2.0%)
WHERE_n_of_non_aggr_func__DOW	
0	984 (98%)
1	16 (1.6%)
WHERE_n_of_non_aggr_func__ROUND	
0	975 (98%)
1	25 (2.5%)
WHERE_n_of_non_aggr_func__FLOOR	
0	980 (98%)
1	20 (2.0%)
WHERE_n_of_non_aggr_func__MONTH	
0	991 (99%)
1	9 (0.9%)
WHERE_n_of_non_aggr_func__DAY	
0	982 (98%)
1	17 (1.7%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__SQRT	
0	936 (94%)
1	60 (6.0%)
2	4 (0.4%)
WHERE_n_of_non_aggr_func__LOG	
0	967 (97%)
1	32 (3.2%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__YEAR	
0	981 (98%)
1	19 (1.9%)
WHERE_n_of_non_aggr_func__ABS	

Query parameter or value	Descriptive statistics
0	978 (98%)
1	22 (2.2%)
WHERE_n_of_all_non_aggr_func	
0	796 (80%)
1	160 (16%)
2	40 (4.0%)
3	4 (0.4%)
GROUP_BY_n_of_columns	
0	244 (24%)
1	275 (28%)
2	233 (23%)
3	134 (13%)
4	65 (6.5%)
5	38 (3.8%)
6	11 (1.1%)
HAVING_n_of_main_predicates	[0.00, 4.00, 9.00] 3.38 / 2.60
HAVING_n_of_main_predicates__non_scalar_subquery	
0	437 (44%)
1	209 (21%)
2	175 (18%)
3	179 (18%)
HAVING_n_of_main_predicates__scalar_subquery	
0	430 (43%)
1	185 (19%)
2	186 (19%)
3	199 (20%)
HAVING_n_of_subqueries__non_scalar_subquery	
0	437 (44%)
1	209 (21%)
2	175 (18%)
3	179 (18%)
HAVING_n_of_subqueries__scalar_subquery	[0.0, 1.0, 19.0] 2.4 / 3.4
HAVING_n_of_processed_rows_by_subqueries	[0, 1,999,865, 245,084,418] 18,931,001 / 37,617,337
ORDER_BY_n_of_columns	
0	290 (29%)
1	443 (44%)
2	227 (23%)
3	29 (2.9%)
4	8 (0.8%)
5	2 (0.2%)
6	1 (0.1%)
limit	[1, 499, 1,000] 493 / 296
offset	[0, 0, 1,000] 239 / 323