

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

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

## Scale factor: 100 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] 5.6 / 5.3
SELECT_n_of_non_aggr_func__SQRT	
0	912 (91%)
1	81 (8.1%)
2	5 (0.5%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func__SUBSTR	
0	792 (79%)
1	178 (18%)
2	25 (2.5%)
3	5 (0.5%)
SELECT_n_of_non_aggr_func__RTRIM	
0	776 (78%)
1	197 (20%)
2	24 (2.4%)

Query parameter or value	Descriptive statistics
3	3 (0.3%)
SELECT_n_of_non_aggr_func___LTRIM	
0	766 (77%)
1	200 (20%)
2	30 (3.0%)
3	3 (0.3%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___UPPER	
0	797 (80%)
1	175 (18%)
2	23 (2.3%)
3	4 (0.4%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___DOW	
0	926 (93%)
1	70 (7.0%)
2	4 (0.4%)
SELECT_n_of_non_aggr_func___MONTH	
0	907 (91%)
1	82 (8.2%)
2	8 (0.8%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func___LOWER	
0	794 (79%)
1	185 (19%)
2	19 (1.9%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___YEAR	
0	919 (92%)
1	69 (6.9%)
2	11 (1.1%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___FLOOR	
0	957 (96%)
1	42 (4.2%)
2	1 (0.1%)
SELECT_n_of_non_aggr_func___TRUNC	
0	959 (96%)
1	39 (3.9%)
2	2 (0.2%)
SELECT_n_of_non_aggr_func___DAY	
0	943 (94%)
1	55 (5.5%)
2	2 (0.2%)
SELECT_n_of_non_aggr_func___LOG	
0	939 (94%)

Query parameter or value	Descriptive statistics
1	56 (5.6%)
2	3 (0.3%)
3	1 (0.1%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__ROUND	
0	967 (97%)
1	30 (3.0%)
2	3 (0.3%)
SELECT_n_of_non_aggr_func__ABS	
0	959 (96%)
1	39 (3.9%)
2	1 (0.1%)
3	1 (0.1%)
SELECT_n_of_all_non_aggr_func	[0.00, 1.00, 21.00] 1.91 / 2.14
SELECT_n_of_aggr_func__COUNT	
0	780 (78%)
1	100 (10%)
2	69 (6.9%)
3	31 (3.1%)
4	10 (1.0%)
5	8 (0.8%)
6	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_aggr_func__MIN	
0	769 (77%)
1	118 (12%)
2	62 (6.2%)
3	29 (2.9%)
4	16 (1.6%)
5	5 (0.5%)
6	1 (0.1%)
SELECT_n_of_aggr_func__MAX	
0	775 (78%)
1	113 (11%)
2	65 (6.5%)
3	35 (3.5%)
4	6 (0.6%)
5	2 (0.2%)
6	3 (0.3%)
7	1 (0.1%)
SELECT_n_of_aggr_func__COUNT_DISTINCT	
0	734 (73%)
1	126 (13%)
2	71 (7.1%)
3	41 (4.1%)
4	15 (1.5%)

Query parameter or value	Descriptive statistics
5	10 (1.0%)
6	1 (0.1%)
7	1 (0.1%)
8	1 (0.1%)
SELECT__n_of__aggr_func__SUM	
0	945 (95%)
1	49 (4.9%)
2	6 (0.6%)
SELECT__n_of__aggr_func__AVG	
0	961 (96%)
1	34 (3.4%)
2	2 (0.2%)
3	3 (0.3%)
SELECT__n_of__all__aggr_func	
0	159 (16%)
1	306 (31%)
2	250 (25%)
3	148 (15%)
4	77 (7.7%)
5	40 (4.0%)
6	14 (1.4%)
7	5 (0.5%)
8	1 (0.1%)
FROM__n_of__join_paths	
1	1,000 (100%)
FROM__n_of__joins__RIGHT	
0	482 (48%)
1	326 (33%)
2	154 (15%)
3	33 (3.3%)
4	5 (0.5%)
FROM__n_of__joins__INNER	
0	501 (50%)
1	327 (33%)
2	130 (13%)
3	39 (3.9%)
4	3 (0.3%)
FROM__n_of__processed_rows	[5, 100,000,000, 765,037,932] 278,143,825 / 310,623,815
WHERE__n_of__predicates	
0	267 (27%)
1	235 (24%)
2	206 (21%)
3	147 (15%)
4	109 (11%)
5	31 (3.1%)

Query parameter or value	Descriptive statistics
6	5 (0.5%)
WHERE_n_of_attribs_of_type__character_varying	
0	575 (58%)
1	304 (30%)
2	104 (10%)
3	17 (1.7%)
WHERE_n_of_attribs_of_type__character	
0	822 (82%)
1	154 (15%)
2	24 (2.4%)
WHERE_n_of_attribs_of_type__integer	
0	589 (59%)
1	306 (31%)
2	90 (9.0%)
3	15 (1.5%)
WHERE_n_of_attribs_of_type__numeric	
0	750 (75%)
1	212 (21%)
2	33 (3.3%)
3	5 (0.5%)
WHERE_n_of_attribs_of_type__date	
0	891 (89%)
1	98 (9.8%)
2	11 (1.1%)
WHERE_n_of_pkey_attribs	
0	746 (75%)
1	215 (22%)
2	38 (3.8%)
3	1 (0.1%)
WHERE_n_of_connect_OR	
0	632 (63%)
1	214 (21%)
2	94 (9.4%)
3	54 (5.4%)
4	6 (0.6%)
WHERE_n_of_operators__greater_or_less	
0	564 (56%)
1	283 (28%)
2	122 (12%)
3	27 (2.7%)
4	4 (0.4%)
WHERE_n_of_operators__like	
0	906 (91%)
1	83 (8.3%)
2	11 (1.1%)
WHERE_n_of_operators__in	

Query parameter or value	Descriptive statistics
0	737 (74%)
1	222 (22%)
2	37 (3.7%)
3	4 (0.4%)
WHERE_n_of_operators__between	
0	725 (73%)
1	220 (22%)
2	51 (5.1%)
3	4 (0.4%)
WHERE_n_of_non_aggr_func__SQRT	
0	950 (95%)
1	49 (4.9%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__ABS	
0	976 (98%)
1	24 (2.4%)
WHERE_n_of_non_aggr_func__MONTH	
0	981 (98%)
1	19 (1.9%)
WHERE_n_of_non_aggr_func__FLOOR	
0	979 (98%)
1	21 (2.1%)
WHERE_n_of_non_aggr_func__DAY	
0	986 (99%)
1	14 (1.4%)
WHERE_n_of_non_aggr_func__LOG	
0	969 (97%)
1	31 (3.1%)
WHERE_n_of_non_aggr_func__DOW	
0	986 (99%)
1	14 (1.4%)
WHERE_n_of_non_aggr_func__ROUND	
0	981 (98%)
1	19 (1.9%)
WHERE_n_of_non_aggr_func__TRUNC	
0	987 (99%)
1	13 (1.3%)
WHERE_n_of_non_aggr_func__YEAR	
0	983 (98%)
1	17 (1.7%)
WHERE_n_of_all_non_aggr_func	
0	807 (81%)
1	165 (17%)
2	27 (2.7%)
4	1 (0.1%)
GROUP_BY_n_of_columns	

Query parameter or value	Descriptive statistics
0	229 (23%)
1	276 (28%)
2	236 (24%)
3	136 (14%)
4	80 (8.0%)
5	34 (3.4%)
6	9 (0.9%)
HAVING_n_of_main_predicates	[0.00, 4.00, 9.00] 3.50 / 2.54
HAVING_n_of_main_predicates__non_scalar_subquery	
0	398 (40%)
1	191 (19%)
2	201 (20%)
3	210 (21%)
HAVING_n_of_main_predicates__scalar_subquery	
0	423 (42%)
1	192 (19%)
2	209 (21%)
3	176 (18%)
HAVING_n_of_subqueries__non_scalar_subquery	
0	398 (40%)
1	191 (19%)
2	201 (20%)
3	210 (21%)
HAVING_n_of_subqueries__scalar_subquery	[0.0, 1.0, 24.0] 2.5 / 3.6
HAVING_n_of_processed_rows_by_subqueries	[0, 81,000,030, 15,000,758,040] 1,068,073,473 / 2,111,872,504
ORDER_BY_n_of_columns	
0	279 (28%)
1	472 (47%)
2	214 (21%)
3	23 (2.3%)
4	9 (0.9%)
5	3 (0.3%)
limit	[2, 498, 999] 498 / 289
offset	[0, 0, 998] 244 / 321