

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

```
## Scenario: 1-3jp_rjpp__g__h0-3ss_h0nss

## Scale factor: 500 GB

## The 1000-query pack was generated on: 2023-11-07

## 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, 120] 10 / 12
SELECT_n_of_non_aggr_func__FLOOR	
0	926 (93%)
1	67 (6.7%)
2	6 (0.6%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func__LTRIM	
0	715 (72%)
1	215 (22%)
2	51 (5.1%)
3	18 (1.8%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__RTRIM	
0	688 (69%)
1	245 (25%)

Query parameter or value	Descriptive statistics
2	55 (5.5%)
3	9 (0.9%)
4	2 (0.2%)
7	1 (0.1%)
SELECT__n_of__non__aggr__func__LOWER	
0	696 (70%)
1	249 (25%)
2	42 (4.2%)
3	10 (1.0%)
5	1 (0.1%)
6	2 (0.2%)
SELECT__n_of__non__aggr__func__DAY	
0	911 (91%)
1	75 (7.5%)
2	13 (1.3%)
3	1 (0.1%)
SELECT__n_of__non__aggr__func__DOW	
0	901 (90%)
1	83 (8.3%)
2	13 (1.3%)
3	1 (0.1%)
4	2 (0.2%)
SELECT__n_of__non__aggr__func__MONTH	
0	844 (84%)
1	124 (12%)
2	21 (2.1%)
3	8 (0.8%)
4	3 (0.3%)
SELECT__n_of__non__aggr__func__SUBSTR	
0	706 (71%)
1	222 (22%)
2	51 (5.1%)
3	12 (1.2%)
4	4 (0.4%)
5	3 (0.3%)
6	1 (0.1%)
9	1 (0.1%)
SELECT__n_of__non__aggr__func__UPPER	
0	705 (71%)
1	229 (23%)
2	51 (5.1%)
3	8 (0.8%)
4	5 (0.5%)
6	1 (0.1%)
7	1 (0.1%)
SELECT__n_of__non__aggr__func__YEAR	

Query parameter or value	Descriptive statistics
0	882 (88%)
1	91 (9.1%)
2	16 (1.6%)
3	7 (0.7%)
4	4 (0.4%)
SELECT_n_of_non_aggr_func___SQRT	
0	866 (87%)
1	108 (11%)
2	17 (1.7%)
3	6 (0.6%)
4	2 (0.2%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___TRUNC	
0	952 (95%)
1	41 (4.1%)
2	6 (0.6%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___ROUND	
0	943 (94%)
1	51 (5.1%)
2	6 (0.6%)
SELECT_n_of_non_aggr_func___ABS	
0	940 (94%)
1	56 (5.6%)
2	4 (0.4%)
SELECT_n_of_non_aggr_func___LOG	
0	907 (91%)
1	77 (7.7%)
2	13 (1.3%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_all_non_aggr_func	[0.0, 2.0, 33.0] 3.1 / 3.9
SELECT_n_of_aggr_func___SUM	
0	938 (94%)
1	41 (4.1%)
2	18 (1.8%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_aggr_func___COUNT	[0.00, 0.00, 9.00] 0.63 / 1.44
SELECT_n_of_aggr_func___MIN	
0	761 (76%)
1	82 (8.2%)
2	50 (5.0%)
3	42 (4.2%)
4	29 (2.9%)
5	10 (1.0%)

Query parameter or value	Descriptive statistics
6	16 (1.6%)
7	5 (0.5%)
8	5 (0.5%)
SELECT_n_of_aggr_func__COUNT_DISTINCT	[0.00, 0.00, 11.00] 0.65 / 1.47
SELECT_n_of_aggr_func__MAX	[0.00, 0.00, 13.00] 0.64 / 1.47
SELECT_n_of_aggr_func__AVG	
0	947 (95%)
1	42 (4.2%)
2	8 (0.8%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_all_aggr_func	[0.00, 2.00, 13.00] 2.71 / 2.30
FROM_n_of_join_paths	
1	350 (35%)
2	334 (33%)
3	316 (32%)
FROM_n_of_super_joins__RIGHT	
0	758 (76%)
1	224 (22%)
2	18 (1.8%)
FROM_n_of_super_joins__LEFT	
0	805 (81%)
1	181 (18%)
2	14 (1.4%)
FROM_n_of_super_joins__FULL	
0	770 (77%)
1	205 (21%)
2	25 (2.5%)
FROM_n_of_joins__INNER	
0	268 (27%)
1	254 (25%)
2	209 (21%)
3	148 (15%)
4	78 (7.8%)
5	31 (3.1%)
6	8 (0.8%)
7	3 (0.3%)
8	1 (0.1%)
FROM_n_of_joins__RIGHT	[0.00, 1.00, 10.00] 1.73 / 1.59
FROM_n_of_processed_rows	[5, 3,400,028,242, 11,475,084,816] 3,138,451,692 / 2,955,957,578
WHERE_n_of_predicates	[0.00, 2.00, 10.00] 2.75 / 2.15
WHERE_n_of_attribs_of_type__character_varying	
0	442 (44%)
1	302 (30%)
2	161 (16%)

Query parameter or value	Descriptive statistics
3	74 (7.4%)
4	18 (1.8%)
5	3 (0.3%)
WHERE_n_of_attribs_of_type__date	
0	831 (83%)
1	142 (14%)
2	25 (2.5%)
3	2 (0.2%)
WHERE_n_of_attribs_of_type__integer	
0	466 (47%)
1	311 (31%)
2	158 (16%)
3	52 (5.2%)
4	8 (0.8%)
5	5 (0.5%)
WHERE_n_of_attribs_of_type__numeric	
0	639 (64%)
1	273 (27%)
2	71 (7.1%)
3	16 (1.6%)
5	1 (0.1%)
WHERE_n_of_attribs_of_type__character	
0	743 (74%)
1	207 (21%)
2	44 (4.4%)
3	4 (0.4%)
4	2 (0.2%)
WHERE_n_of_pkey_attribs	
0	616 (62%)
1	297 (30%)
2	73 (7.3%)
3	12 (1.2%)
4	1 (0.1%)
5	1 (0.1%)
WHERE_n_of_connect_OR	
0	459 (46%)
1	185 (19%)
2	130 (13%)
3	120 (12%)
4	54 (5.4%)
5	26 (2.6%)
6	18 (1.8%)
7	6 (0.6%)
8	2 (0.2%)
WHERE_n_of_operators__greater_or_less	
0	402 (40%)

Query parameter or value	Descriptive statistics
1	301 (30%)
2	171 (17%)
3	89 (8.9%)
4	25 (2.5%)
5	8 (0.8%)
6	4 (0.4%)
WHERE_n_of_operators_in	
0	627 (63%)
1	283 (28%)
2	59 (5.9%)
3	24 (2.4%)
4	7 (0.7%)
WHERE_n_of_operators_like	
0	841 (84%)
1	142 (14%)
2	16 (1.6%)
3	1 (0.1%)
WHERE_n_of_operators_between	
0	639 (64%)
1	253 (25%)
2	86 (8.6%)
3	20 (2.0%)
4	2 (0.2%)
WHERE_n_of_non_aggr_func_DOW	
0	970 (97%)
1	30 (3.0%)
WHERE_n_of_non_aggr_func_SQRT	
0	906 (91%)
1	88 (8.8%)
2	6 (0.6%)
WHERE_n_of_non_aggr_func_DAY	
0	968 (97%)
1	32 (3.2%)
WHERE_n_of_non_aggr_func_ROUND	
0	966 (97%)
1	34 (3.4%)
WHERE_n_of_non_aggr_func_YEAR	
0	969 (97%)
1	30 (3.0%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func_TRUNC	
0	973 (97%)
1	27 (2.7%)
WHERE_n_of_non_aggr_func_FLOOR	
0	967 (97%)
1	32 (3.2%)

Query parameter or value	Descriptive statistics
2	1 (0.1%)
WHERE_n_of_non_aggr_func__MONTH	
0	974 (97%)
1	25 (2.5%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__ABS	
0	982 (98%)
1	18 (1.8%)
WHERE_n_of_non_aggr_func__LOG	
0	948 (95%)
1	52 (5.2%)
WHERE_n_of_all_non_aggr_func	
0	699 (70%)
1	230 (23%)
2	59 (5.9%)
3	10 (1.0%)
4	2 (0.2%)
GROUP_BY_n_of_columns	[0.00, 2.00, 11.00] 2.36 / 2.12
HAVING_n_of_main_predicates	
0	289 (29%)
1	87 (8.7%)
2	133 (13%)
3	200 (20%)
4	148 (15%)
5	94 (9.4%)
6	49 (4.9%)
HAVING_n_of_main_predicates__scalar_subquery	
0	435 (44%)
1	195 (20%)
2	194 (19%)
3	176 (18%)
HAVING_n_of_subqueries__scalar_subquery	[0.0, 1.0, 20.0] 2.3 / 3.4
HAVING_n_of_processed_rows_by_subqueries	[6,105,000,605, 126,225,932,866] 7,198,189,974 / 15,805,043,075
ORDER_BY_n_of_columns	
0	309 (31%)
1	408 (41%)
2	197 (20%)
3	62 (6.2%)
4	13 (1.3%)
5	7 (0.7%)
6	2 (0.2%)
7	1 (0.1%)
8	1 (0.1%)
limit	[1, 494, 1,000] 499 / 289

offset

[0, 66, 998] 265 / 324

---