

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

Scenario: 1-6jp_ujpp__g__h0-3ss_h0-3nss

Scale factor: 100 GB

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

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, 9, 170] 16 / 24
SELECT_n_of_non_aggr_func__LOWER	
0	643 (64%)
1	232 (23%)
2	77 (7.7%)
3	23 (2.3%)
4	16 (1.6%)
5	3 (0.3%)
6	2 (0.2%)
7	3 (0.3%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func__LTRIM	
0	648 (65%)
1	240 (24%)
2	72 (7.2%)

Query parameter or value	Descriptive statistics
3	22 (2.2%)
4	6 (0.6%)
5	9 (0.9%)
6	1 (0.1%)
9	2 (0.2%)
SELECT_n_of_non_aggr_func___SQRT	
0	809 (81%)
1	139 (14%)
2	27 (2.7%)
3	8 (0.8%)
4	13 (1.3%)
5	2 (0.2%)
6	1 (0.1%)
9	1 (0.1%)
SELECT_n_of_non_aggr_func___UPPER	
0	631 (63%)
1	269 (27%)
2	73 (7.3%)
3	13 (1.3%)
4	5 (0.5%)
5	5 (0.5%)
6	2 (0.2%)
7	1 (0.1%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func___SUBSTR	
0	622 (62%)
1	266 (27%)
10	1 (0.1%)
2	74 (7.4%)
3	12 (1.2%)
4	14 (1.4%)
5	6 (0.6%)
6	2 (0.2%)
7	3 (0.3%)
SELECT_n_of_non_aggr_func___DOW	
0	864 (86%)
1	91 (9.1%)
2	27 (2.7%)
3	9 (0.9%)
4	6 (0.6%)
5	2 (0.2%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func___MONTH	
0	799 (80%)
1	137 (14%)
10	1 (0.1%)

Query parameter or value	Descriptive statistics
2	37 (3.7%)
3	12 (1.2%)
4	9 (0.9%)
5	4 (0.4%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func___ABS	
0	934 (93%)
1	54 (5.4%)
2	10 (1.0%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___DAY	
0	884 (88%)
1	90 (9.0%)
2	18 (1.8%)
3	5 (0.5%)
4	3 (0.3%)
SELECT_n_of_non_aggr_func___RTRIM	
0	630 (63%)
1	235 (24%)
11	1 (0.1%)
2	94 (9.4%)
3	29 (2.9%)
4	6 (0.6%)
5	4 (0.4%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func___YEAR	
0	826 (83%)
1	127 (13%)
2	26 (2.6%)
3	7 (0.7%)
4	5 (0.5%)
5	7 (0.7%)
6	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func___FLOOR	
0	927 (93%)
1	61 (6.1%)
2	10 (1.0%)
3	1 (0.1%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___LOG	
0	881 (88%)
1	87 (8.7%)
2	26 (2.6%)
3	3 (0.3%)
4	3 (0.3%)

Query parameter or value	Descriptive statistics
SELECT__n_of__non_aggr_func__ROUND	
0	926 (93%)
1	58 (5.8%)
2	11 (1.1%)
3	3 (0.3%)
4	1 (0.1%)
5	1 (0.1%)
SELECT__n_of__non_aggr_func__TRUNC	
0	925 (93%)
1	67 (6.7%)
2	7 (0.7%)
4	1 (0.1%)
SELECT__n_of__all__non_aggr_func	[0.0, 3.0, 84.0] 4.6 / 7.0
SELECT__n_of__aggr_func__MIN	[0.00, 0.00, 11.00] 0.83 / 1.93
SELECT__n_of__aggr_func__COUNT	[0.00, 0.00, 12.00] 0.87 / 1.90
SELECT__n_of__aggr_func__COUNT_DISTINCT	[0.00, 0.00, 12.00] 0.84 / 2.03
SELECT__n_of__aggr_func__MAX	[0.00, 0.00, 14.00] 0.82 / 1.96
SELECT__n_of__aggr_func__SUM	
0	943 (94%)
1	39 (3.9%)
2	13 (1.3%)
3	4 (0.4%)
7	1 (0.1%)
SELECT__n_of__aggr_func__AVG	
0	918 (92%)
1	53 (5.3%)
2	23 (2.3%)
3	4 (0.4%)
4	1 (0.1%)
5	1 (0.1%)
SELECT__n_of__all__aggr_func	[0.0, 3.0, 14.0] 3.6 / 3.2
FROM__n_of__join_paths	
1	162 (16%)
2	165 (17%)
3	176 (18%)
4	154 (15%)
5	156 (16%)
6	187 (19%)
FROM__n_of__super_joins__FULL	
0	550 (55%)
1	312 (31%)
2	112 (11%)
3	23 (2.3%)
4	3 (0.3%)
FROM__n_of__super_joins__RIGHT	
0	545 (55%)

Query parameter or value	Descriptive statistics
1	300 (30%)
2	123 (12%)
3	30 (3.0%)
4	2 (0.2%)
FROM_n_of_super_joins__LEFT	
0	550 (55%)
1	297 (30%)
2	134 (13%)
3	16 (1.6%)
4	3 (0.3%)
FROM_n_of_joins__INNER	[0.00, 3.00, 12.00] 3.23 / 2.45
FROM_n_of_joins__RIGHT	[0.00, 3.00, 14.00] 3.38 / 2.60
FROM_n_of_processed_rows	[5, 900,037,902, 4,256,227,497] 1,144,041,691 / 937,095,391
WHERE_n_of_predicates	[0.00, 3.00, 13.00] 3.84 / 2.93
WHERE_n_of_attribs_of_type__integer	
0	374 (37%)
1	315 (32%)
2	167 (17%)
3	87 (8.7%)
4	41 (4.1%)
5	10 (1.0%)
6	4 (0.4%)
7	1 (0.1%)
8	1 (0.1%)
WHERE_n_of_attribs_of_type__numeric	
0	546 (55%)
1	283 (28%)
2	130 (13%)
3	23 (2.3%)
4	18 (1.8%)
WHERE_n_of_attribs_of_type__character_varying	
0	361 (36%)
1	302 (30%)
2	178 (18%)
3	98 (9.8%)
4	38 (3.8%)
5	15 (1.5%)
6	7 (0.7%)
7	1 (0.1%)
WHERE_n_of_attribs_of_type__date	
0	777 (78%)
1	182 (18%)
2	34 (3.4%)
3	6 (0.6%)
4	1 (0.1%)

Query parameter or value	Descriptive statistics
WHERE_n_of_attribs_of_type__character	
0	638 (64%)
1	253 (25%)
2	91 (9.1%)
3	15 (1.5%)
4	3 (0.3%)
WHERE_n_of_pkey_attribs	
0	548 (55%)
1	295 (30%)
2	113 (11%)
3	36 (3.6%)
4	5 (0.5%)
5	1 (0.1%)
6	2 (0.2%)
WHERE_n_of_connect_OR	[0.00, 1.00, 11.00] 2.22 / 2.43
WHERE_n_of_operators__greater_or_less	
0	327 (33%)
1	271 (27%)
2	172 (17%)
3	129 (13%)
4	59 (5.9%)
5	35 (3.5%)
6	5 (0.5%)
8	1 (0.1%)
9	1 (0.1%)
WHERE_n_of_operators__between	
0	514 (51%)
1	315 (32%)
2	110 (11%)
3	46 (4.6%)
4	13 (1.3%)
5	2 (0.2%)
WHERE_n_of_operators__in	
0	548 (55%)
1	289 (29%)
2	122 (12%)
3	29 (2.9%)
4	8 (0.8%)
5	1 (0.1%)
6	2 (0.2%)
7	1 (0.1%)
WHERE_n_of_operators__like	
0	780 (78%)
1	189 (19%)
2	29 (2.9%)
3	1 (0.1%)

Query parameter or value	Descriptive statistics
4	1 (0.1%)
WHERE_n_of_non_aggr_func__DOW	
0	962 (96%)
1	38 (3.8%)
WHERE_n_of_non_aggr_func__SQRT	
0	867 (87%)
1	115 (12%)
2	18 (1.8%)
WHERE_n_of_non_aggr_func__ABS	
0	958 (96%)
1	42 (4.2%)
WHERE_n_of_non_aggr_func__ROUND	
0	943 (94%)
1	55 (5.5%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func__LOG	
0	919 (92%)
1	75 (7.5%)
2	5 (0.5%)
3	1 (0.1%)
WHERE_n_of_non_aggr_func__MONTH	
0	959 (96%)
1	40 (4.0%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__DAY	
0	969 (97%)
1	31 (3.1%)
WHERE_n_of_non_aggr_func__FLOOR	
0	956 (96%)
1	42 (4.2%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func__YEAR	
0	946 (95%)
1	53 (5.3%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__TRUNC	
0	969 (97%)
1	29 (2.9%)
2	2 (0.2%)
WHERE_n_of_all_non_aggr_func	
0	593 (59%)
1	280 (28%)
2	90 (9.0%)
3	26 (2.6%)
4	9 (0.9%)
5	1 (0.1%)

Query parameter or value	Descriptive statistics
6	1 (0.1%)
GROUP_BY_n_of_columns	[0.00, 2.00, 12.00] 3.03 / 2.88
HAVING_n_of_main_predicates	[0.00, 3.00, 9.00] 3.16 / 2.53
HAVING_n_of_main_predicates__non_scalar_subquery	
0	491 (49%)
1	169 (17%)
2	189 (19%)
3	151 (15%)
HAVING_n_of_main_predicates__scalar_subquery	
0	446 (45%)
1	210 (21%)
2	173 (17%)
3	171 (17%)
HAVING_n_of_subqueries__non_scalar_subquery	
0	491 (49%)
1	169 (17%)
2	189 (19%)
3	151 (15%)
HAVING_n_of_subqueries__scalar_subquery	[0.0, 1.0, 19.0] 2.3 / 3.5
HAVING_n_of_processed_rows_by_subquery_offset	[0, 1,018,519,089, 56,622,881,027] 3,889,287,754 / 6,701,192,660
ORDER_BY_n_of_columns	
0	290 (29%)
1	345 (35%)
2	227 (23%)
3	90 (9.0%)
4	34 (3.4%)
5	9 (0.9%)
6	4 (0.4%)
7	1 (0.1%)
limit	[1, 510, 999] 506 / 285
offset	[0, 0, 995] 243 / 319