

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

Scenario: 1-1jp__g__h0-3ss_h0-3nss

Scale factor: 1000 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 / 5.4
SELECT_n_of_non_aggr_func__LOWER	
0	779 (78%)
1	192 (19%)
2	25 (2.5%)
3	3 (0.3%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__UPPER	
0	776 (78%)
1	201 (20%)
2	19 (1.9%)
3	2 (0.2%)
4	1 (0.1%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func__TRUNC	

Query parameter or value	Descriptive statistics
0	960 (96%)
1	39 (3.9%)
2	1 (0.1%)
SELECT_n_of_non_aggr_func__DAY	
0	954 (95%)
1	38 (3.8%)
2	8 (0.8%)
SELECT_n_of_non_aggr_func__RTRIM	
0	781 (78%)
1	194 (19%)
2	23 (2.3%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func__FLOOR	
0	958 (96%)
1	42 (4.2%)
SELECT_n_of_non_aggr_func__LTRIM	
0	771 (77%)
1	193 (19%)
2	34 (3.4%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func__SQRT	
0	895 (90%)
1	95 (9.5%)
2	9 (0.9%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func__DOW	
0	927 (93%)
1	66 (6.6%)
2	6 (0.6%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func__LOG	
0	935 (94%)
1	62 (6.2%)
2	3 (0.3%)
SELECT_n_of_non_aggr_func__MONTH	
0	915 (92%)
1	66 (6.6%)
2	14 (1.4%)
3	5 (0.5%)
SELECT_n_of_non_aggr_func__SUBSTR	
0	794 (79%)
1	187 (19%)
2	16 (1.6%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__YEAR	

Query parameter or value	Descriptive statistics
0	929 (93%)
1	64 (6.4%)
2	7 (0.7%)
SELECT_n_of_non_aggr_func__ROUND	
0	969 (97%)
1	29 (2.9%)
2	2 (0.2%)
SELECT_n_of_non_aggr_func__ABS	
0	959 (96%)
1	40 (4.0%)
2	1 (0.1%)
SELECT_n_of_all_non_aggr_func	[0.00, 1.00, 15.00] 1.92 / 2.09
SELECT_n_of_aggr_func__COUNT_DISTINCT	
0	778 (78%)
1	108 (11%)
2	59 (5.9%)
3	36 (3.6%)
4	10 (1.0%)
5	4 (0.4%)
6	3 (0.3%)
7	1 (0.1%)
8	1 (0.1%)
SELECT_n_of_aggr_func__MIN	
0	753 (75%)
1	107 (11%)
2	78 (7.8%)
3	39 (3.9%)
4	13 (1.3%)
5	6 (0.6%)
6	4 (0.4%)
SELECT_n_of_aggr_func__COUNT	
0	763 (76%)
1	112 (11%)
2	67 (6.7%)
3	29 (2.9%)
4	19 (1.9%)
5	8 (0.8%)
6	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_aggr_func__MAX	
0	768 (77%)
1	114 (11%)
2	63 (6.3%)
3	30 (3.0%)
4	13 (1.3%)
5	10 (1.0%)

Query parameter or value	Descriptive statistics
6	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_aggr_func__AVG	
0	956 (96%)
1	36 (3.6%)
2	4 (0.4%)
3	4 (0.4%)
SELECT_n_of_aggr_func__SUM	
0	954 (95%)
1	38 (3.8%)
2	7 (0.7%)
3	1 (0.1%)
SELECT_n_of_all_aggr_func	
0	181 (18%)
1	277 (28%)
2	239 (24%)
3	145 (15%)
4	80 (8.0%)
5	55 (5.5%)
6	16 (1.6%)
7	6 (0.6%)
8	1 (0.1%)
FROM_n_of_join_paths	
1	1,000 (100%)
FROM_n_of_joins__RIGHT	
0	485 (49%)
1	341 (34%)
2	135 (14%)
3	35 (3.5%)
4	4 (0.4%)
FROM_n_of_joins__INNER	
0	485 (49%)
1	337 (34%)
2	136 (14%)
3	39 (3.9%)
4	3 (0.3%)
FROM_n_of_processed_rows	[5, 1,000,000,000, 7,649,989,739] 2,925,686,268 / 3,175,959,180
WHERE_n_of_predicates	
0	220 (22%)
1	237 (24%)
2	231 (23%)
3	146 (15%)
4	105 (11%)
5	50 (5.0%)
6	11 (1.1%)

Query parameter or value	Descriptive statistics
WHERE_n_of_attribs_of_type__character	
0	813 (81%)
1	164 (16%)
2	20 (2.0%)
3	3 (0.3%)
WHERE_n_of_attribs_of_type__integer	
0	522 (52%)
1	345 (35%)
2	112 (11%)
3	20 (2.0%)
5	1 (0.1%)
WHERE_n_of_attribs_of_type__numeric	
0	745 (75%)
1	211 (21%)
2	40 (4.0%)
3	4 (0.4%)
WHERE_n_of_attribs_of_type__character_varying	
0	532 (53%)
1	338 (34%)
2	108 (11%)
3	22 (2.2%)
WHERE_n_of_attribs_of_type__date	
0	909 (91%)
1	80 (8.0%)
2	10 (1.0%)
3	1 (0.1%)
WHERE_n_of_pkey_attribs	
0	690 (69%)
1	270 (27%)
2	36 (3.6%)
3	3 (0.3%)
4	1 (0.1%)
WHERE_n_of_connect_OR	
0	588 (59%)
1	233 (23%)
2	104 (10%)
3	57 (5.7%)
4	18 (1.8%)
WHERE_n_of_operators__in	
0	720 (72%)
1	224 (22%)
2	50 (5.0%)
3	5 (0.5%)
4	1 (0.1%)
WHERE_n_of_operators__greater_or_less	
0	542 (54%)

Query parameter or value	Descriptive statistics
1	287 (29%)
2	130 (13%)
3	34 (3.4%)
4	6 (0.6%)
5	1 (0.1%)
WHERE_n_of_operators__between	
0	695 (70%)
1	243 (24%)
2	56 (5.6%)
3	5 (0.5%)
4	1 (0.1%)
WHERE_n_of_operators__like	
0	864 (86%)
1	126 (13%)
2	10 (1.0%)
WHERE_n_of_non_aggr_func__FLOOR	
0	975 (98%)
1	24 (2.4%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__LOG	
0	962 (96%)
1	36 (3.6%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func__SQRT	
0	935 (94%)
1	63 (6.3%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func__MONTH	
0	989 (99%)
1	11 (1.1%)
WHERE_n_of_non_aggr_func__ABS	
0	985 (99%)
1	15 (1.5%)
WHERE_n_of_non_aggr_func__DAY	
0	986 (99%)
1	13 (1.3%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__DOW	
0	987 (99%)
1	12 (1.2%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__ROUND	
0	969 (97%)
1	30 (3.0%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__TRUNC	

Query parameter or value	Descriptive statistics
0	986 (99%)
1	14 (1.4%)
WHERE_n_of_non_aggr_func__YEAR	
0	988 (99%)
1	12 (1.2%)
WHERE_n_of_all_non_aggr_func	
0	803 (80%)
1	154 (15%)
2	37 (3.7%)
3	6 (0.6%)
GROUP_BY_n_of_columns	
0	241 (24%)
1	295 (30%)
2	211 (21%)
3	153 (15%)
4	60 (6.0%)
5	35 (3.5%)
6	5 (0.5%)
HAVING_n_of_main_predicates	[0.00, 4.00, 9.00] 3.40 / 2.57
HAVING_n_of_main_predicates__non_scalar_subquery	
0	425 (43%)
1	186 (19%)
2	192 (19%)
3	197 (20%)
HAVING_n_of_main_predicates__scalar_subquery	
0	430 (43%)
1	204 (20%)
2	174 (17%)
3	192 (19%)
HAVING_n_of_subqueries__non_scalar_subquery	
0	425 (43%)
1	186 (19%)
2	192 (19%)
3	197 (20%)
HAVING_n_of_subqueries__scalar_subquery	[0.0, 1.0, 21.0] 2.5 / 3.7
HAVING_n_of_processed_rows_by_subqueries	[0, 800,000,000, 172,499,763,307] 10,391,486,999 / 19,979,026,334
ORDER_BY_n_of_columns	
0	301 (30%)
1	461 (46%)
2	202 (20%)
3	24 (2.4%)
4	7 (0.7%)
5	4 (0.4%)
6	1 (0.1%)
limit	[3, 480, 999] 493 / 288

Query parameter or value	Descriptive statistics
offset	[0, 0, 998] 240 / 320