

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

Scenario: 1-6jp_ujpp__g__h0ss_h0nss

Scale factor: 100 GB

The 1000-query pack was generated on: 2023-11-10

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, 174] 16 / 24
SELECT_n_of_non_aggr_func__DAY	
0	883 (88%)
1	97 (9.7%)
2	9 (0.9%)
3	6 (0.6%)
4	1 (0.1%)
5	2 (0.2%)
6	1 (0.1%)
9	1 (0.1%)
SELECT_n_of_non_aggr_func__LOWER	[0.00, 0.00, 9.00] 0.62 / 1.02
SELECT_n_of_non_aggr_func__LTRIM	[0.00, 0.00, 9.00] 0.58 / 1.07
SELECT_n_of_non_aggr_func__SUBSTR	[0.00, 0.00, 10.00] 0.60 / 1.01
SELECT_n_of_non_aggr_func__RTRIM	
0	592 (59%)

Query parameter or value	Descriptive statistics
1	287 (29%)
2	83 (8.3%)
3	16 (1.6%)
4	8 (0.8%)
5	10 (1.0%)
6	2 (0.2%)
7	1 (0.1%)
9	1 (0.1%)
SELECT__n_of__non_aggr_func__UPPER	
0	626 (63%)
1	267 (27%)
2	68 (6.8%)
3	17 (1.7%)
4	10 (1.0%)
5	4 (0.4%)
6	3 (0.3%)
7	4 (0.4%)
8	1 (0.1%)
SELECT__n_of__non_aggr_func__YEAR	[0.00, 0.00, 10.00] 0.27 / 0.87
SELECT__n_of__non_aggr_func__DOW	
0	856 (86%)
1	108 (11%)
2	23 (2.3%)
3	6 (0.6%)
4	2 (0.2%)
5	1 (0.1%)
6	3 (0.3%)
7	1 (0.1%)
SELECT__n_of__non_aggr_func__MONTH	[0.00, 0.00, 11.00] 0.32 / 0.99
SELECT__n_of__non_aggr_func__SQRT	
0	791 (79%)
1	158 (16%)
2	27 (2.7%)
3	11 (1.1%)
4	5 (0.5%)
5	5 (0.5%)
6	1 (0.1%)
7	1 (0.1%)
8	1 (0.1%)
SELECT__n_of__non_aggr_func__FLOOR	
0	929 (93%)
1	64 (6.4%)
2	5 (0.5%)
3	1 (0.1%)
4	1 (0.1%)
SELECT__n_of__non_aggr_func__ROUND	

Query parameter or value	Descriptive statistics
0	919 (92%)
1	69 (6.9%)
2	5 (0.5%)
3	3 (0.3%)
4	3 (0.3%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func__LOG	
0	875 (88%)
1	97 (9.7%)
2	19 (1.9%)
3	6 (0.6%)
4	2 (0.2%)
6	1 (0.1%)
SELECT_n_of_non_aggr_func__ABS	
0	919 (92%)
1	69 (6.9%)
2	9 (0.9%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func__TRUNC	
0	919 (92%)
1	72 (7.2%)
2	7 (0.7%)
3	2 (0.2%)
SELECT_n_of_all_non_aggr_func	[0.0, 3.0, 81.0] 4.8 / 7.8
SELECT_n_of_aggr_func__MAX	[0.00, 0.00, 14.00] 0.93 / 2.11
SELECT_n_of_aggr_func__MIN	[0.00, 0.00, 11.00] 0.75 / 1.75
SELECT_n_of_aggr_func__COUNT_DISTINCT	[0.00, 0.00, 12.00] 0.93 / 2.04
SELECT_n_of_aggr_func__AVG	
0	936 (94%)
1	46 (4.6%)
2	12 (1.2%)
3	6 (0.6%)
SELECT_n_of_aggr_func__COUNT	[0.00, 0.00, 15.00] 0.82 / 1.96
SELECT_n_of_aggr_func__SUM	
0	936 (94%)
1	42 (4.2%)
2	19 (1.9%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_all_aggr_func	[0.0, 3.0, 15.0] 3.6 / 3.1
FROM_n_of_join_paths	
1	143 (14%)
2	182 (18%)
3	177 (18%)
4	152 (15%)
5	179 (18%)

Query parameter or value	Descriptive statistics
6	167 (17%)
FROM_n_of_super_joins___FULL	
0	546 (55%)
1	310 (31%)
2	117 (12%)
3	26 (2.6%)
4	1 (0.1%)
FROM_n_of_super_joins___LEFT	
0	521 (52%)
1	333 (33%)
2	121 (12%)
3	23 (2.3%)
4	2 (0.2%)
FROM_n_of_super_joins___RIGHT	
0	522 (52%)
1	332 (33%)
2	123 (12%)
3	21 (2.1%)
4	2 (0.2%)
FROM_n_of_joins___INNER	[0.00, 3.00, 12.00] 3.39 / 2.51
FROM_n_of_joins___RIGHT	[0.00, 3.00, 11.00] 3.32 / 2.47
FROM_n_of_processed_rows	[5, 928,038,020, 4,426,227,472] 1,191,231,738 / 990,940,512
WHERE_n_of_predicates	[0.00, 3.00, 13.00] 3.84 / 2.91
WHERE_n_of_attribs_of_type___date	
0	789 (79%)
1	169 (17%)
2	32 (3.2%)
3	8 (0.8%)
4	2 (0.2%)
WHERE_n_of_attribs_of_type___integer	
0	375 (38%)
1	269 (27%)
2	189 (19%)
3	107 (11%)
4	47 (4.7%)
5	10 (1.0%)
6	3 (0.3%)
WHERE_n_of_attribs_of_type___numeric	
0	555 (56%)
1	307 (31%)
2	104 (10%)
3	27 (2.7%)
4	6 (0.6%)
5	1 (0.1%)
WHERE_n_of_attribs_of_type___character_varying	

Query parameter or value	Descriptive statistics
0	368 (37%)
1	275 (28%)
2	181 (18%)
3	101 (10%)
4	57 (5.7%)
5	15 (1.5%)
6	2 (0.2%)
7	1 (0.1%)
WHERE_n_of_attribs_of_type__character	
0	667 (67%)
1	232 (23%)
2	79 (7.9%)
3	19 (1.9%)
4	3 (0.3%)
WHERE_n_of_pkey_attribs	
0	518 (52%)
1	314 (31%)
2	125 (13%)
3	36 (3.6%)
4	7 (0.7%)
WHERE_n_of_connect_OR	[0.00, 1.00, 10.00] 2.19 / 2.39
WHERE_n_of_operators__between	
0	503 (50%)
1	310 (31%)
2	130 (13%)
3	45 (4.5%)
4	8 (0.8%)
5	3 (0.3%)
7	1 (0.1%)
WHERE_n_of_operators__greater_or_less	
0	334 (33%)
1	279 (28%)
2	188 (19%)
3	98 (9.8%)
4	70 (7.0%)
5	19 (1.9%)
6	8 (0.8%)
7	3 (0.3%)
8	1 (0.1%)
WHERE_n_of_operators__in	
0	527 (53%)
1	294 (29%)
2	137 (14%)
3	28 (2.8%)
4	9 (0.9%)
5	5 (0.5%)

Query parameter or value	Descriptive statistics
WHERE_n_of_operators_like	
0	788 (79%)
1	174 (17%)
2	32 (3.2%)
3	6 (0.6%)
WHERE_n_of_non_aggr_func_DOW	
0	961 (96%)
1	36 (3.6%)
2	3 (0.3%)
WHERE_n_of_non_aggr_func_TRUNC	
0	963 (96%)
1	37 (3.7%)
WHERE_n_of_non_aggr_func_ABS	
0	961 (96%)
1	39 (3.9%)
WHERE_n_of_non_aggr_func_DAY	
0	973 (97%)
1	26 (2.6%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func_SQRT	
0	870 (87%)
1	116 (12%)
2	14 (1.4%)
WHERE_n_of_non_aggr_func_MONTH	
0	963 (96%)
1	34 (3.4%)
2	3 (0.3%)
WHERE_n_of_non_aggr_func_YEAR	
0	963 (96%)
1	34 (3.4%)
2	3 (0.3%)
WHERE_n_of_non_aggr_func_LOG	
0	920 (92%)
1	78 (7.8%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func_FLOOR	
0	967 (97%)
1	32 (3.2%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func_ROUND	
0	944 (94%)
1	54 (5.4%)
2	2 (0.2%)
WHERE_n_of_all_non_aggr_func	
0	601 (60%)
1	287 (29%)

Query parameter or value	Descriptive statistics
2	86 (8.6%)
3	20 (2.0%)
4	5 (0.5%)
5	1 (0.1%)
GROUP_BY_n_of_columns	[0.00, 3.00, 12.00] 3.15 / 2.86
HAVING_n_of_main_predicates	
0	420 (42%)
1	183 (18%)
2	211 (21%)
3	186 (19%)
ORDER_BY_n_of_columns	
0	293 (29%)
1	329 (33%)
2	232 (23%)
3	105 (11%)
4	20 (2.0%)
5	14 (1.4%)
6	4 (0.4%)
7	2 (0.2%)
9	1 (0.1%)
limit	[1, 519, 999] 513 / 288
offset	[0, 60, 1,000] 266 / 326