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

Scenario: 1-3jp_rjjp__g__h0ss_h0nss

Scale factor: 2000 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, 6, 113] 11 / 14
SELECT_n_of_non_aggr_func___UPPER	
0	696 (70%)
1	231 (23%)
2	53 (5.3%)
3	14 (1.4%)
4	3 (0.3%)
5	2 (0.2%)
6	1 (0.1%)
SELECT_n_of_non_aggr_func___ABS	
0	938 (94%)
1	59 (5.9%)
2	3 (0.3%)
SELECT_n_of_non_aggr_func___DOW	
0	891 (89%)

Query parameter or value	Descriptive statistics
1	81 (8.1%)
2	18 (1.8%)
3	6 (0.6%)
4	3 (0.3%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___LTRIM	
0	679 (68%)
1	232 (23%)
2	70 (7.0%)
3	9 (0.9%)
4	7 (0.7%)
5	1 (0.1%)
7	1 (0.1%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func___MONTH	
0	840 (84%)
1	113 (11%)
2	29 (2.9%)
3	12 (1.2%)
4	2 (0.2%)
5	3 (0.3%)
6	1 (0.1%)
SELECT_n_of_non_aggr_func___RTRIM	
0	710 (71%)
1	224 (22%)
2	43 (4.3%)
3	14 (1.4%)
4	2 (0.2%)
5	5 (0.5%)
6	2 (0.2%)
SELECT_n_of_non_aggr_func___SQRT	
0	855 (86%)
1	111 (11%)
2	23 (2.3%)
3	8 (0.8%)
4	2 (0.2%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___YEAR	
0	853 (85%)
1	113 (11%)
2	24 (2.4%)
3	6 (0.6%)
4	2 (0.2%)
5	2 (0.2%)
SELECT_n_of_non_aggr_func___SUBSTR	
0	696 (70%)

Query parameter or value	Descriptive statistics
1	225 (23%)
2	60 (6.0%)
3	12 (1.2%)
4	3 (0.3%)
5	2 (0.2%)
6	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func___FLOOR	
0	942 (94%)
1	53 (5.3%)
2	3 (0.3%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___LOWER	
0	709 (71%)
1	219 (22%)
2	52 (5.2%)
3	14 (1.4%)
4	3 (0.3%)
5	3 (0.3%)
SELECT_n_of_non_aggr_func___DAY	
0	907 (91%)
1	73 (7.3%)
2	17 (1.7%)
3	2 (0.2%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___LOG	
0	903 (90%)
1	79 (7.9%)
2	15 (1.5%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___ROUND	
0	944 (94%)
1	48 (4.8%)
2	6 (0.6%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___TRUNC	
0	940 (94%)
1	55 (5.5%)
2	4 (0.4%)
3	1 (0.1%)
SELECT_n_of_all_non_aggr_func	[0.0, 2.0, 54.0] 3.3 / 5.0
SELECT_n_of_aggr_func___COUNT	[0.00, 0.00, 11.00] 0.63 / 1.46
SELECT_n_of_aggr_func___MAX	[0.00, 0.00, 10.00] 0.64 / 1.51
SELECT_n_of_aggr_func___COUNT_DISTINCT	
0	762 (76%)

Query parameter or value	Descriptive statistics
1	93 (9.3%)
2	46 (4.6%)
3	48 (4.8%)
4	18 (1.8%)
5	16 (1.6%)
6	9 (0.9%)
7	4 (0.4%)
8	4 (0.4%)
SELECT_n_of_aggr_func___MIN	[0.00, 0.00, 9.00] 0.56 / 1.34
SELECT_n_of_aggr_func___AVG	
0	956 (96%)
1	30 (3.0%)
2	11 (1.1%)
3	3 (0.3%)
SELECT_n_of_aggr_func___SUM	
0	954 (95%)
1	34 (3.4%)
2	8 (0.8%)
3	3 (0.3%)
4	1 (0.1%)
SELECT_n_of_all_aggr_func	[0.00, 2.00, 11.00] 2.55 / 2.32
FROM_n_of_join_paths	
1	336 (34%)
2	331 (33%)
3	333 (33%)
FROM_n_of_super_joins___FULL	
0	762 (76%)
1	216 (22%)
2	22 (2.2%)
FROM_n_of_super_joins___LEFT	
0	757 (76%)
1	223 (22%)
2	20 (2.0%)
FROM_n_of_super_joins___RIGHT	
0	780 (78%)
1	188 (19%)
2	32 (3.2%)
FROM_n_of_joins___INNER	
0	232 (23%)
1	262 (26%)
2	220 (22%)
3	134 (13%)
4	81 (8.1%)
5	46 (4.6%)
6	21 (2.1%)
7	3 (0.3%)

Query parameter or value	Descriptive statistics
8	1 (0.1%)
FROM_n_of_joins___RIGHT	[0.00, 1.00, 9.00] 1.78 / 1.60
FROM_n_of_processed_rows	[5, 13,620,027,823, 45,900,083,487] 13,307,923,128 / 12,272,607,089
WHERE_n_of_predicates	[0.00, 2.00, 9.00] 2.82 / 2.25
WHERE_n_of_attribs_of_type___integer	
0	452 (45%)
1	321 (32%)
2	143 (14%)
3	64 (6.4%)
4	16 (1.6%)
5	4 (0.4%)
WHERE_n_of_attribs_of_type___numeric	
0	648 (65%)
1	254 (25%)
2	76 (7.6%)
3	19 (1.9%)
4	3 (0.3%)
WHERE_n_of_attribs_of_type___character_varying	
0	429 (43%)
1	325 (33%)
2	155 (16%)
3	73 (7.3%)
4	15 (1.5%)
5	2 (0.2%)
6	1 (0.1%)
WHERE_n_of_attribs_of_type___character	
0	725 (73%)
1	213 (21%)
2	57 (5.7%)
3	5 (0.5%)
WHERE_n_of_attribs_of_type___date	
0	834 (83%)
1	142 (14%)
2	21 (2.1%)
3	3 (0.3%)
WHERE_n_of_pkey_attribs	
0	624 (62%)
1	274 (27%)
2	83 (8.3%)
3	18 (1.8%)
4	1 (0.1%)
WHERE_n_of_connect_OR	
0	452 (45%)
1	170 (17%)
2	138 (14%)

Query parameter or value	Descriptive statistics
3	115 (12%)
4	61 (6.1%)
5	38 (3.8%)
6	19 (1.9%)
7	4 (0.4%)
8	3 (0.3%)
WHERE_n_of_operators___greater_or_less	
0	429 (43%)
1	289 (29%)
2	163 (16%)
3	74 (7.4%)
4	39 (3.9%)
5	5 (0.5%)
6	1 (0.1%)
WHERE_n_of_operators___in	
0	607 (61%)
1	270 (27%)
2	90 (9.0%)
3	26 (2.6%)
4	5 (0.5%)
5	1 (0.1%)
6	1 (0.1%)
WHERE_n_of_operators___between	
0	625 (63%)
1	263 (26%)
2	84 (8.4%)
3	27 (2.7%)
4	1 (0.1%)
WHERE_n_of_operators___like	
0	830 (83%)
1	153 (15%)
2	15 (1.5%)
3	2 (0.2%)
WHERE_n_of_non_aggr_func___DAY	
0	972 (97%)
1	27 (2.7%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___YEAR	
0	983 (98%)
1	17 (1.7%)
WHERE_n_of_non_aggr_func___SQRT	
0	894 (89%)
1	101 (10%)
2	5 (0.5%)
WHERE_n_of_non_aggr_func___FLOOR	
0	965 (97%)

Query parameter or value	Descriptive statistics
1	34 (3.4%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___DOW	
0	965 (97%)
1	35 (3.5%)
WHERE_n_of_non_aggr_func___ROUND	
0	973 (97%)
1	27 (2.7%)
WHERE_n_of_non_aggr_func___MONTH	
0	974 (97%)
1	25 (2.5%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___LOG	
0	943 (94%)
1	52 (5.2%)
2	5 (0.5%)
WHERE_n_of_non_aggr_func___TRUNC	
0	959 (96%)
1	41 (4.1%)
WHERE_n_of_non_aggr_func___ABS	
0	973 (97%)
1	27 (2.7%)
WHERE_n_of_all_non_aggr_func	
0	682 (68%)
1	238 (24%)
2	67 (6.7%)
3	12 (1.2%)
4	1 (0.1%)
GROUP_BY_n_of_columns	[0.00, 2.00, 10.00] 2.29 / 2.15
HAVING_n_of_main_predicates	
0	432 (43%)
1	193 (19%)
2	181 (18%)
3	194 (19%)
ORDER_BY_n_of_columns	
0	268 (27%)
1	398 (40%)
2	244 (24%)
3	62 (6.2%)
4	21 (2.1%)
5	5 (0.5%)
6	2 (0.2%)
limit	[2, 481, 1,000] 492 / 288
offset	[0, 0, 996] 242 / 321