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

Scenario: 1-3jp_rjjp__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, 6, 90] 10 / 13
SELECT_n_of_non_aggr_func___SUBSTR	
0	712 (71%)
1	226 (23%)
2	48 (4.8%)
3	6 (0.6%)
4	5 (0.5%)
5	1 (0.1%)
6	2 (0.2%)
SELECT_n_of_non_aggr_func___UPPER	
0	669 (67%)
1	255 (26%)
2	52 (5.2%)
3	14 (1.4%)
4	8 (0.8%)

Query parameter or value	Descriptive statistics
5	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func___LOG	
0	906 (91%)
1	80 (8.0%)
2	10 (1.0%)
3	4 (0.4%)
SELECT_n_of_non_aggr_func___RTRIM	
0	661 (66%)
1	264 (26%)
2	53 (5.3%)
3	16 (1.6%)
4	4 (0.4%)
6	2 (0.2%)
SELECT_n_of_non_aggr_func___SQRT	
0	874 (87%)
1	95 (9.5%)
2	18 (1.8%)
3	6 (0.6%)
4	6 (0.6%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___YEAR	
0	856 (86%)
1	116 (12%)
2	18 (1.8%)
3	4 (0.4%)
4	4 (0.4%)
5	1 (0.1%)
6	1 (0.1%)
SELECT_n_of_non_aggr_func___DOW	
0	882 (88%)
1	102 (10%)
2	13 (1.3%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___FLOOR	
0	953 (95%)
1	44 (4.4%)
2	3 (0.3%)
SELECT_n_of_non_aggr_func___MONTH	
0	838 (84%)
1	123 (12%)
2	27 (2.7%)
3	6 (0.6%)
4	3 (0.3%)
5	2 (0.2%)

Query parameter or value	Descriptive statistics
6	1 (0.1%)
SELECT_n_of_non_aggr_func___LTRIM	
0	719 (72%)
1	205 (21%)
2	55 (5.5%)
3	13 (1.3%)
4	4 (0.4%)
5	2 (0.2%)
6	2 (0.2%)
SELECT_n_of_non_aggr_func___TRUNC	
0	934 (93%)
1	59 (5.9%)
2	6 (0.6%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___LOWER	
0	701 (70%)
1	234 (23%)
2	45 (4.5%)
3	11 (1.1%)
4	3 (0.3%)
5	3 (0.3%)
6	2 (0.2%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func___ROUND	
0	957 (96%)
1	41 (4.1%)
2	2 (0.2%)
SELECT_n_of_non_aggr_func___DAY	
0	912 (91%)
1	78 (7.8%)
2	9 (0.9%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___ABS	
0	946 (95%)
1	53 (5.3%)
2	1 (0.1%)
SELECT_n_of_all_non_aggr_func	[0.0, 2.0, 41.0] 3.2 / 4.5
SELECT_n_of_aggr_func___COUNT	[0.00, 0.00, 9.00] 0.65 / 1.43
SELECT_n_of_aggr_func___COUNT_DISTINCT	[0.00, 0.00, 9.00] 0.65 / 1.46
SELECT_n_of_aggr_func___MIN	[0.00, 0.00, 9.00] 0.65 / 1.45
SELECT_n_of_aggr_func___AVG	
0	934 (93%)
1	39 (3.9%)
2	22 (2.2%)
3	2 (0.2%)
4	2 (0.2%)

Query parameter or value	Descriptive statistics
5	1 (0.1%)
SELECT_n_of_aggr_func___MAX	[0.00, 0.00, 9.00] 0.60 / 1.38
SELECT_n_of_aggr_func___SUM	
0	945 (95%)
1	35 (3.5%)
2	16 (1.6%)
3	3 (0.3%)
4	1 (0.1%)
SELECT_n_of_all_aggr_func	[0.00, 2.00, 10.00] 2.73 / 2.29
FROM_n_of_join_paths	
1	301 (30%)
2	349 (35%)
3	350 (35%)
FROM_n_of_super_joins___RIGHT	
0	772 (77%)
1	206 (21%)
2	22 (2.2%)
FROM_n_of_super_joins___LEFT	
0	738 (74%)
1	233 (23%)
2	29 (2.9%)
FROM_n_of_super_joins___FULL	
0	802 (80%)
1	181 (18%)
2	17 (1.7%)
FROM_n_of_joins___INNER	
0	224 (22%)
1	266 (27%)
2	220 (22%)
3	148 (15%)
4	87 (8.7%)
5	37 (3.7%)
6	14 (1.4%)
7	4 (0.4%)
FROM_n_of_joins___RIGHT	
0	235 (24%)
1	267 (27%)
2	196 (20%)
3	147 (15%)
4	96 (9.6%)
5	42 (4.2%)
6	12 (1.2%)
7	5 (0.5%)
FROM_n_of_processed_rows	[5, 681,037,932, 2,295,113,706] 647,143,316 / 593,811,107
WHERE_n_of_predicates	[0.00, 3.00, 9.00] 2.81 / 2.11

Query parameter or value	Descriptive statistics
WHERE_n_of_attribs_of_type___character_varying	
0	430 (43%)
1	312 (31%)
2	167 (17%)
3	72 (7.2%)
4	17 (1.7%)
5	2 (0.2%)
WHERE_n_of_attribs_of_type___integer	
0	461 (46%)
1	320 (32%)
2	141 (14%)
3	58 (5.8%)
4	16 (1.6%)
5	4 (0.4%)
WHERE_n_of_attribs_of_type___numeric	
0	620 (62%)
1	299 (30%)
2	64 (6.4%)
3	15 (1.5%)
4	2 (0.2%)
WHERE_n_of_attribs_of_type___character	
0	724 (72%)
1	228 (23%)
2	39 (3.9%)
3	9 (0.9%)
WHERE_n_of_attribs_of_type___date	
0	822 (82%)
1	157 (16%)
2	20 (2.0%)
3	1 (0.1%)
WHERE_n_of_pkey_attribs	
0	634 (63%)
1	271 (27%)
2	76 (7.6%)
3	16 (1.6%)
4	2 (0.2%)
5	1 (0.1%)
WHERE_n_of_connect_OR	
0	434 (43%)
1	194 (19%)
2	148 (15%)
3	112 (11%)
4	58 (5.8%)
5	33 (3.3%)
6	16 (1.6%)
7	5 (0.5%)

Query parameter or value	Descriptive statistics
WHERE_n_of_operators___between	
0	601 (60%)
1	301 (30%)
2	75 (7.5%)
3	21 (2.1%)
4	1 (0.1%)
5	1 (0.1%)
WHERE_n_of_operators___greater_or_less	
0	389 (39%)
1	330 (33%)
2	162 (16%)
3	81 (8.1%)
4	27 (2.7%)
5	11 (1.1%)
WHERE_n_of_operators___in	
0	593 (59%)
1	305 (31%)
2	76 (7.6%)
3	21 (2.1%)
4	5 (0.5%)
WHERE_n_of_operators___like	
0	845 (85%)
1	140 (14%)
2	15 (1.5%)
WHERE_n_of_non_aggr_func___TRUNC	
0	971 (97%)
1	29 (2.9%)
WHERE_n_of_non_aggr_func___LOG	
0	941 (94%)
1	55 (5.5%)
2	4 (0.4%)
WHERE_n_of_non_aggr_func___ROUND	
0	967 (97%)
1	33 (3.3%)
WHERE_n_of_non_aggr_func___SQRT	
0	907 (91%)
1	91 (9.1%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___DOW	
0	971 (97%)
1	29 (2.9%)
WHERE_n_of_non_aggr_func___ABS	
0	966 (97%)
1	34 (3.4%)
WHERE_n_of_non_aggr_func___MONTH	
0	975 (98%)

Query parameter or value	Descriptive statistics
1	25 (2.5%)
WHERE_n_of_non_aggr_func___FLOOR	
0	980 (98%)
1	19 (1.9%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___YEAR	
0	969 (97%)
1	30 (3.0%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___DAY	
0	967 (97%)
1	32 (3.2%)
2	1 (0.1%)
WHERE_n_of_all_non_aggr_func	
0	685 (69%)
1	244 (24%)
2	62 (6.2%)
3	9 (0.9%)
GROUP_BY_n_of_columns	[0.00, 2.00, 9.00] 2.31 / 2.11
HAVING_n_of_main_predicates	
0	462 (46%)
1	175 (18%)
2	179 (18%)
3	184 (18%)
ORDER_BY_n_of_columns	
0	295 (30%)
1	389 (39%)
2	220 (22%)
3	70 (7.0%)
4	14 (1.4%)
5	5 (0.5%)
6	5 (0.5%)
7	2 (0.2%)
limit	[1, 497, 1,000] 504 / 293
offset	[0, 72, 998] 275 / 325