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

Scenario: 1-3jp_ujjp__g__h0ss_h0nss

Scale factor: 500 GB

The 1000-query pack was generated on: 2023-11-09

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, 97] 10 / 13
SELECT_n_of_non_aggr_func___LTRIM	
0	681 (68%)
1	239 (24%)
2	51 (5.1%)
3	20 (2.0%)
4	6 (0.6%)
5	2 (0.2%)
6	1 (0.1%)
SELECT_n_of_non_aggr_func___ROUND	
0	953 (95%)
1	44 (4.4%)
2	3 (0.3%)
SELECT_n_of_non_aggr_func___RTRIM	
0	691 (69%)

Query parameter or value	Descriptive statistics
1	240 (24%)
2	52 (5.2%)
3	9 (0.9%)
4	3 (0.3%)
5	4 (0.4%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func___SUBSTR	
0	696 (70%)
1	234 (23%)
2	46 (4.6%)
3	15 (1.5%)
4	7 (0.7%)
5	2 (0.2%)
SELECT_n_of_non_aggr_func___UPPER	
0	681 (68%)
1	244 (24%)
2	55 (5.5%)
3	13 (1.3%)
4	6 (0.6%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___DOW	
0	898 (90%)
1	76 (7.6%)
2	15 (1.5%)
3	7 (0.7%)
4	4 (0.4%)
SELECT_n_of_non_aggr_func___MONTH	
0	854 (85%)
1	105 (11%)
2	21 (2.1%)
3	11 (1.1%)
4	3 (0.3%)
5	5 (0.5%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func___DAY	
0	920 (92%)
1	69 (6.9%)
2	8 (0.8%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___ABS	
0	939 (94%)
1	55 (5.5%)
2	5 (0.5%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___SQRT	

Query parameter or value	Descriptive statistics
0	846 (85%)
1	123 (12%)
2	18 (1.8%)
3	9 (0.9%)
4	4 (0.4%)
SELECT_n_of_non_aggr_func___LOWER	
0	702 (70%)
1	215 (22%)
2	58 (5.8%)
3	18 (1.8%)
4	6 (0.6%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func___FLOOR	
0	939 (94%)
1	57 (5.7%)
2	4 (0.4%)
SELECT_n_of_non_aggr_func___YEAR	
0	867 (87%)
1	98 (9.8%)
2	22 (2.2%)
3	5 (0.5%)
4	5 (0.5%)
5	3 (0.3%)
SELECT_n_of_non_aggr_func___TRUNC	
0	953 (95%)
1	42 (4.2%)
2	4 (0.4%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___LOG	
0	904 (90%)
1	83 (8.3%)
2	8 (0.8%)
3	5 (0.5%)
SELECT_n_of_all_non_aggr_func	[0.0, 2.0, 45.0] 3.3 / 4.7
SELECT_n_of_aggr_func___AVG	
0	941 (94%)
1	46 (4.6%)
2	7 (0.7%)
3	5 (0.5%)
5	1 (0.1%)
SELECT_n_of_aggr_func___COUNT_DISTINCT	[0.00, 0.00, 11.00] 0.60 / 1.37
SELECT_n_of_aggr_func___COUNT	[0.00, 0.00, 10.00] 0.65 / 1.43
SELECT_n_of_aggr_func___MIN	[0.00, 0.00, 9.00] 0.69 / 1.50
SELECT_n_of_aggr_func___MAX	[0.00, 0.00, 9.00] 0.58 / 1.36
SELECT_n_of_aggr_func___SUM	
0	940 (94%)

Query parameter or value	Descriptive statistics
1	41 (4.1%)
2	17 (1.7%)
4	1 (0.1%)
5	1 (0.1%)
SELECT_n_of_all_aggr_func	[0.00, 2.00, 11.00] 2.68 / 2.26
FROM_n_of_join_paths	
1	322 (32%)
2	351 (35%)
3	327 (33%)
FROM_n_of_super_joins___FULL	
0	769 (77%)
1	213 (21%)
2	18 (1.8%)
FROM_n_of_super_joins___RIGHT	
0	774 (77%)
1	210 (21%)
2	16 (1.6%)
FROM_n_of_super_joins___LEFT	
0	759 (76%)
1	218 (22%)
2	23 (2.3%)
FROM_n_of_joins___RIGHT	[0.00, 2.00, 9.00] 1.77 / 1.50
FROM_n_of_joins___INNER	
0	253 (25%)
1	274 (27%)
2	200 (20%)
3	130 (13%)
4	89 (8.9%)
5	30 (3.0%)
6	16 (1.6%)
7	6 (0.6%)
8	2 (0.2%)
FROM_n_of_processed_rows	[5, 3,405,028,270, 11,475,084,806] 3,286,013,048 / 3,127,374,298
WHERE_n_of_predicates	[0.00, 2.00, 10.00] 2.79 / 2.19
WHERE_n_of_attribs_of_type___character_varying	
0	412 (41%)
1	346 (35%)
2	159 (16%)
3	63 (6.3%)
4	17 (1.7%)
5	3 (0.3%)
WHERE_n_of_attribs_of_type___integer	
0	457 (46%)
1	316 (32%)
2	150 (15%)

Query parameter or value	Descriptive statistics
3	57 (5.7%)
4	14 (1.4%)
5	6 (0.6%)
WHERE_n_of_attribs_of_type___numeric	
0	668 (67%)
1	228 (23%)
2	80 (8.0%)
3	21 (2.1%)
4	3 (0.3%)
WHERE_n_of_attribs_of_type___character	
0	727 (73%)
1	219 (22%)
2	47 (4.7%)
3	6 (0.6%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type___date	
0	845 (85%)
1	133 (13%)
2	19 (1.9%)
3	3 (0.3%)
WHERE_n_of_pkey_attribs	
0	635 (64%)
1	280 (28%)
2	68 (6.8%)
3	16 (1.6%)
4	1 (0.1%)
WHERE_n_of_connect_OR	[0.00, 1.00, 9.00] 1.31 / 1.64
WHERE_n_of_operators___greater_or_less	
0	399 (40%)
1	319 (32%)
2	174 (17%)
3	76 (7.6%)
4	21 (2.1%)
5	8 (0.8%)
6	2 (0.2%)
7	1 (0.1%)
WHERE_n_of_operators___in	
0	622 (62%)
1	271 (27%)
2	78 (7.8%)
3	24 (2.4%)
4	3 (0.3%)
5	2 (0.2%)
WHERE_n_of_operators___between	
0	654 (65%)
1	243 (24%)

Query parameter or value	Descriptive statistics
2	84 (8.4%)
3	16 (1.6%)
4	3 (0.3%)
WHERE_n_of_operators___like	
0	827 (83%)
1	143 (14%)
2	29 (2.9%)
3	1 (0.1%)
WHERE_n_of_non_aggr_func___ROUND	
0	971 (97%)
1	29 (2.9%)
WHERE_n_of_non_aggr_func___DAY	
0	971 (97%)
1	29 (2.9%)
WHERE_n_of_non_aggr_func___FLOOR	
0	966 (97%)
1	34 (3.4%)
WHERE_n_of_non_aggr_func___DOW	
0	978 (98%)
1	21 (2.1%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___SQRT	
0	903 (90%)
1	91 (9.1%)
2	6 (0.6%)
WHERE_n_of_non_aggr_func___LOG	
0	947 (95%)
1	51 (5.1%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___ABS	
0	975 (98%)
1	25 (2.5%)
WHERE_n_of_non_aggr_func___TRUNC	
0	974 (97%)
1	24 (2.4%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___YEAR	
0	983 (98%)
1	17 (1.7%)
WHERE_n_of_non_aggr_func___MONTH	
0	978 (98%)
1	22 (2.2%)
WHERE_n_of_all_non_aggr_func	
0	728 (73%)
1	194 (19%)
2	63 (6.3%)

Query parameter or value	Descriptive statistics
3	15 (1.5%)
GROUP_BY_n_of_columns	[0.00, 2.00, 9.00] 2.40 / 2.17
HAVING_n_of_main_predicates	
0	426 (43%)
1	203 (20%)
2	180 (18%)
3	191 (19%)
ORDER_BY_n_of_columns	
0	296 (30%)
1	372 (37%)
2	231 (23%)
3	65 (6.5%)
4	23 (2.3%)
5	12 (1.2%)
6	1 (0.1%)
limit	[2, 500, 1,000] 494 / 286
offset	[0, 0, 996] 246 / 319