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

Scenario: 1-1jp__g__h0ss_h0nss

Scale factor: 1000 GB

The 1000-query pack was generated on: 2023-11-04

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, 42.0] 5.8 / 5.5
SELECT_n_of_non_aggr_func___ABS	
0	960 (96%)
1	38 (3.8%)
2	2 (0.2%)
SELECT_n_of_non_aggr_func___LOWER	
0	776 (78%)
1	195 (20%)
2	26 (2.6%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___LTRIM	
0	790 (79%)
1	183 (18%)
2	25 (2.5%)

Query parameter or value	Descriptive statistics
3	2 (0.2%)
SELECT_n_of_non_aggr_func___RTRIM	
0	787 (79%)
1	188 (19%)
2	22 (2.2%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___SUBSTR	
0	771 (77%)
1	194 (19%)
2	30 (3.0%)
3	5 (0.5%)
SELECT_n_of_non_aggr_func___SQRT	
0	901 (90%)
1	92 (9.2%)
2	7 (0.7%)
SELECT_n_of_non_aggr_func___LOG	
0	933 (93%)
1	60 (6.0%)
2	6 (0.6%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___DOW	
0	948 (95%)
1	46 (4.6%)
2	4 (0.4%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___MONTH	
0	913 (91%)
1	69 (6.9%)
2	16 (1.6%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___UPPER	
0	788 (79%)
1	186 (19%)
2	23 (2.3%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func___DAY	
0	949 (95%)
1	45 (4.5%)
2	5 (0.5%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___FLOOR	
0	967 (97%)
1	31 (3.1%)
2	2 (0.2%)
SELECT_n_of_non_aggr_func___YEAR	

Query parameter or value	Descriptive statistics
0	928 (93%)
1	61 (6.1%)
2	9 (0.9%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___ROUND	
0	975 (98%)
1	25 (2.5%)
SELECT_n_of_non_aggr_func___TRUNC	
0	969 (97%)
1	30 (3.0%)
2	1 (0.1%)
SELECT_n_of_all_non_aggr_func	[0.00, 1.00, 18.00] 1.87 / 2.15
SELECT_n_of_aggr_func___MIN	
0	780 (78%)
1	116 (12%)
2	62 (6.2%)
3	26 (2.6%)
4	10 (1.0%)
5	3 (0.3%)
6	2 (0.2%)
7	1 (0.1%)
SELECT_n_of_aggr_func___AVG	
0	955 (96%)
1	34 (3.4%)
2	11 (1.1%)
SELECT_n_of_aggr_func___MAX	
0	768 (77%)
1	94 (9.4%)
2	70 (7.0%)
3	40 (4.0%)
4	15 (1.5%)
5	7 (0.7%)
6	4 (0.4%)
7	2 (0.2%)
SELECT_n_of_aggr_func___COUNT_DISTINCT	
0	804 (80%)
1	93 (9.3%)
2	53 (5.3%)
3	28 (2.8%)
4	12 (1.2%)
5	8 (0.8%)
6	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_aggr_func___SUM	
0	958 (96%)
1	34 (3.4%)

Query parameter or value	Descriptive statistics
2	8 (0.8%)
SELECT_n_of_aggr_func___COUNT	
0	759 (76%)
1	115 (12%)
2	70 (7.0%)
3	32 (3.2%)
4	15 (1.5%)
5	7 (0.7%)
6	2 (0.2%)
SELECT_n_of_all_aggr_func	
0	208 (21%)
1	276 (28%)
2	227 (23%)
3	151 (15%)
4	74 (7.4%)
5	43 (4.3%)
6	16 (1.6%)
7	3 (0.3%)
8	2 (0.2%)
FROM_n_of_join_paths	
1	1,000 (100%)
FROM_n_of_joins___INNER	
0	504 (50%)
1	320 (32%)
2	149 (15%)
3	23 (2.3%)
4	4 (0.4%)
FROM_n_of_joins___RIGHT	
0	481 (48%)
1	341 (34%)
2	138 (14%)
3	33 (3.3%)
4	7 (0.7%)
FROM_n_of_processed_rows	[5, 1,000,000,000, 7,649,989,739] 2,914,486,288 / 3,165,754,122
WHERE_n_of_predicates	
0	247 (25%)
1	234 (23%)
2	212 (21%)
3	156 (16%)
4	90 (9.0%)
5	50 (5.0%)
6	11 (1.1%)
WHERE_n_of_attribs_of_type___character	
0	834 (83%)
1	146 (15%)

Query parameter or value	Descriptive statistics
2	19 (1.9%)
3	1 (0.1%)
WHERE_n_of_attribs_of_type___date	
0	893 (89%)
1	91 (9.1%)
2	16 (1.6%)
WHERE_n_of_attribs_of_type___integer	
0	596 (60%)
1	305 (31%)
2	85 (8.5%)
3	13 (1.3%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type___numeric	
0	723 (72%)
1	227 (23%)
2	38 (3.8%)
3	11 (1.1%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type___character_varying	
0	527 (53%)
1	340 (34%)
2	108 (11%)
3	23 (2.3%)
4	1 (0.1%)
5	1 (0.1%)
WHERE_n_of_pkey_attribs	
0	732 (73%)
1	226 (23%)
2	40 (4.0%)
3	2 (0.2%)
WHERE_n_of_connect_OR	
0	612 (61%)
1	213 (21%)
2	94 (9.4%)
3	63 (6.3%)
4	15 (1.5%)
5	3 (0.3%)
WHERE_n_of_operators___between	
0	737 (74%)
1	222 (22%)
2	37 (3.7%)
3	4 (0.4%)
WHERE_n_of_operators___greater_or_less	
0	538 (54%)
1	305 (31%)
2	117 (12%)

Query parameter or value	Descriptive statistics
3	32 (3.2%)
4	7 (0.7%)
5	1 (0.1%)
WHERE_n_of_operators___in	
0	731 (73%)
1	220 (22%)
2	44 (4.4%)
3	5 (0.5%)
WHERE_n_of_operators___like	
0	870 (87%)
1	123 (12%)
2	7 (0.7%)
WHERE_n_of_non_aggr_func___FLOOR	
0	970 (97%)
1	29 (2.9%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___MONTH	
0	982 (98%)
1	18 (1.8%)
WHERE_n_of_non_aggr_func___SQRT	
0	950 (95%)
1	50 (5.0%)
WHERE_n_of_non_aggr_func___DAY	
0	976 (98%)
1	24 (2.4%)
WHERE_n_of_non_aggr_func___YEAR	
0	982 (98%)
1	17 (1.7%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___DOW	
0	987 (99%)
1	13 (1.3%)
WHERE_n_of_non_aggr_func___ABS	
0	977 (98%)
1	23 (2.3%)
WHERE_n_of_non_aggr_func___ROUND	
0	984 (98%)
1	15 (1.5%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___TRUNC	
0	977 (98%)
1	23 (2.3%)
WHERE_n_of_non_aggr_func___LOG	
0	973 (97%)
1	27 (2.7%)
WHERE_n_of_all_non_aggr_func	

Query parameter or value	Descriptive statistics
0	798 (80%)
1	165 (17%)
2	32 (3.2%)
3	4 (0.4%)
4	1 (0.1%)
GROUP_BY_n_of_columns	
0	277 (28%)
1	272 (27%)
2	215 (22%)
3	121 (12%)
4	75 (7.5%)
5	31 (3.1%)
6	9 (0.9%)
HAVING_n_of_main_predicates	
0	451 (45%)
1	193 (19%)
2	167 (17%)
3	189 (19%)
ORDER_BY_n_of_columns	
0	301 (30%)
1	468 (47%)
2	197 (20%)
3	27 (2.7%)
4	4 (0.4%)
5	3 (0.3%)
limit	[1, 498, 999] 499 / 297
offset	[0, 5, 997] 256 / 328