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

Scenario: 1-3jp_ujjp__g__h0-3ss_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, 94] 10 / 13
SELECT_n_of_non_aggr_func___LOWER	
0	695 (70%)
1	235 (24%)
2	52 (5.2%)
3	12 (1.2%)
4	1 (0.1%)
5	2 (0.2%)
6	2 (0.2%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func___LOG	
0	893 (89%)
1	88 (8.8%)
2	11 (1.1%)
3	7 (0.7%)

Query parameter or value	Descriptive statistics
4	1 (0.1%)
SELECT_n_of_non_aggr_func___ROUND	
0	941 (94%)
1	53 (5.3%)
2	6 (0.6%)
SELECT_n_of_non_aggr_func___RTRIM	
0	691 (69%)
1	237 (24%)
2	46 (4.6%)
3	18 (1.8%)
4	6 (0.6%)
5	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func___LTRIM	
0	700 (70%)
1	224 (22%)
2	55 (5.5%)
3	18 (1.8%)
4	2 (0.2%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func___SQRT	
0	845 (85%)
1	120 (12%)
2	22 (2.2%)
3	6 (0.6%)
4	4 (0.4%)
5	2 (0.2%)
6	1 (0.1%)
SELECT_n_of_non_aggr_func___DOW	
0	891 (89%)
1	85 (8.5%)
2	13 (1.3%)
3	8 (0.8%)
4	3 (0.3%)
SELECT_n_of_non_aggr_func___SUBSTR	
0	702 (70%)
1	226 (23%)
2	48 (4.8%)
3	17 (1.7%)
4	5 (0.5%)
5	2 (0.2%)
SELECT_n_of_non_aggr_func___YEAR	
0	877 (88%)
1	89 (8.9%)
2	20 (2.0%)
3	8 (0.8%)

Query parameter or value	Descriptive statistics
4	4 (0.4%)
5	2 (0.2%)
SELECT_n_of_non_aggr_func___ABS	
0	950 (95%)
1	47 (4.7%)
2	3 (0.3%)
SELECT_n_of_non_aggr_func___MONTH	
0	851 (85%)
1	111 (11%)
2	23 (2.3%)
3	9 (0.9%)
4	4 (0.4%)
5	2 (0.2%)
SELECT_n_of_non_aggr_func___TRUNC	
0	933 (93%)
1	57 (5.7%)
2	9 (0.9%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___UPPER	
0	704 (70%)
1	229 (23%)
2	51 (5.1%)
3	10 (1.0%)
4	4 (0.4%)
5	1 (0.1%)
6	1 (0.1%)
SELECT_n_of_non_aggr_func___DAY	
0	922 (92%)
1	70 (7.0%)
2	5 (0.5%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___FLOOR	
0	932 (93%)
1	63 (6.3%)
2	5 (0.5%)
SELECT_n_of_all_non_aggr_func	[0.0, 2.0, 40.0] 3.3 / 4.6
SELECT_n_of_aggr_func___COUNT_DISTINCT	[0.00, 0.00, 9.00] 0.65 / 1.45
SELECT_n_of_aggr_func___MAX	[0.00, 0.00, 9.00] 0.66 / 1.48
SELECT_n_of_aggr_func___MIN	[0.00, 0.00, 9.00] 0.65 / 1.52
SELECT_n_of_aggr_func___COUNT	
0	763 (76%)
1	79 (7.9%)
2	63 (6.3%)
3	34 (3.4%)
4	27 (2.7%)

Query parameter or value	Descriptive statistics
5	15 (1.5%)
6	15 (1.5%)
7	3 (0.3%)
8	1 (0.1%)
SELECT_n_of_aggr_func___AVG	
0	947 (95%)
1	42 (4.2%)
2	9 (0.9%)
3	2 (0.2%)
SELECT_n_of_aggr_func___SUM	
0	952 (95%)
1	33 (3.3%)
2	11 (1.1%)
3	4 (0.4%)
SELECT_n_of_all_aggr_func	[0.00, 2.00, 10.00] 2.70 / 2.23
FROM_n_of_join_paths	
1	318 (32%)
2	334 (33%)
3	348 (35%)
FROM_n_of_super_joins___LEFT	
0	772 (77%)
1	210 (21%)
2	18 (1.8%)
FROM_n_of_super_joins___FULL	
0	765 (77%)
1	219 (22%)
2	16 (1.6%)
FROM_n_of_super_joins___RIGHT	
0	755 (76%)
1	223 (22%)
2	22 (2.2%)
FROM_n_of_joins___INNER	
0	231 (23%)
1	268 (27%)
2	216 (22%)
3	125 (13%)
4	90 (9.0%)
5	44 (4.4%)
6	23 (2.3%)
7	3 (0.3%)
FROM_n_of_joins___RIGHT	
0	227 (23%)
1	283 (28%)
2	212 (21%)
3	148 (15%)
4	79 (7.9%)

Query parameter or value	Descriptive statistics
5	33 (3.3%)
6	9 (0.9%)
7	8 (0.8%)
8	1 (0.1%)
FROM_n_of_processed_rows	[5, 13,620,027,828, 45,900,083,487] 12,883,382,016 / 11,878,957,778
WHERE_n_of_predicates	[0.00, 2.00, 10.00] 2.77 / 2.22
WHERE_n_of_attribs_of_type___character_varying	
0	441 (44%)
1	310 (31%)
2	169 (17%)
3	58 (5.8%)
4	18 (1.8%)
5	4 (0.4%)
WHERE_n_of_attribs_of_type___date	
0	837 (84%)
1	139 (14%)
2	23 (2.3%)
3	1 (0.1%)
WHERE_n_of_attribs_of_type___integer	
0	439 (44%)
1	317 (32%)
2	176 (18%)
3	50 (5.0%)
4	12 (1.2%)
5	6 (0.6%)
WHERE_n_of_attribs_of_type___character	
0	746 (75%)
1	196 (20%)
2	53 (5.3%)
3	5 (0.5%)
WHERE_n_of_attribs_of_type___numeric	
0	648 (65%)
1	268 (27%)
2	66 (6.6%)
3	16 (1.6%)
4	1 (0.1%)
5	1 (0.1%)
WHERE_n_of_pkey_attribs	
0	599 (60%)
1	303 (30%)
2	80 (8.0%)
3	13 (1.3%)
4	5 (0.5%)
WHERE_n_of_connect_OR	
0	456 (46%)

Query parameter or value	Descriptive statistics
1	171 (17%)
2	149 (15%)
3	110 (11%)
4	44 (4.4%)
5	47 (4.7%)
6	14 (1.4%)
7	7 (0.7%)
8	2 (0.2%)
WHERE_n_of_operators___between	
0	636 (64%)
1	255 (26%)
2	87 (8.7%)
3	19 (1.9%)
4	3 (0.3%)
WHERE_n_of_operators___greater_or_less	
0	423 (42%)
1	285 (29%)
2	183 (18%)
3	80 (8.0%)
4	22 (2.2%)
5	7 (0.7%)
WHERE_n_of_operators___in	
0	627 (63%)
1	259 (26%)
2	93 (9.3%)
3	17 (1.7%)
4	4 (0.4%)
WHERE_n_of_operators___like	
0	830 (83%)
1	151 (15%)
2	18 (1.8%)
3	1 (0.1%)
WHERE_n_of_non_aggr_func___DAY	
0	976 (98%)
1	23 (2.3%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___LOG	
0	942 (94%)
1	54 (5.4%)
2	4 (0.4%)
WHERE_n_of_non_aggr_func___DOW	
0	976 (98%)
1	22 (2.2%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___ROUND	
0	977 (98%)

Query parameter or value	Descriptive statistics
1	22 (2.2%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___SQRT	
0	900 (90%)
1	96 (9.6%)
2	4 (0.4%)
WHERE_n_of_non_aggr_func___YEAR	
0	979 (98%)
1	21 (2.1%)
WHERE_n_of_non_aggr_func___FLOOR	
0	968 (97%)
1	32 (3.2%)
WHERE_n_of_non_aggr_func___ABS	
0	962 (96%)
1	38 (3.8%)
WHERE_n_of_non_aggr_func___MONTH	
0	972 (97%)
1	27 (2.7%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___TRUNC	
0	970 (97%)
1	30 (3.0%)
WHERE_n_of_all_non_aggr_func	
0	704 (70%)
1	220 (22%)
2	58 (5.8%)
3	17 (1.7%)
4	1 (0.1%)
GROUP_BY_n_of_columns	[0.00, 2.00, 9.00] 2.24 / 2.11
HAVING_n_of_main_predicates	
0	328 (33%)
1	92 (9.2%)
2	123 (12%)
3	179 (18%)
4	138 (14%)
5	86 (8.6%)
6	54 (5.4%)
HAVING_n_of_main_predicates___scalar_subquery	
0	467 (47%)
1	178 (18%)
2	170 (17%)
3	185 (19%)
HAVING_n_of_subqueries___scalar_subquery	[0.0, 1.0, 20.0] 2.3 / 3.6
HAVING_n_of_processed_rows_by_subqueries	[0, 40,000,074, 688,501,251,330] 29,762,411,438 / 68,400,803,059
ORDER_BY_n_of_columns	

Query parameter or value	Descriptive statistics
0	311 (31%)
1	382 (38%)
2	212 (21%)
3	64 (6.4%)
4	22 (2.2%)
5	7 (0.7%)
6	2 (0.2%)
limit	[2, 512, 1,000] 505 / 291
offset	[0, 32, 1,000] 257 / 328