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

Scenario: 1-6jp_ujjp__g__h0ss_h0nss

Scale factor: 200 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, 8, 157] 16 / 24
SELECT_n_of_non_aggr_func___LOWER	
0	641 (64%)
1	255 (26%)
2	57 (5.7%)
3	25 (2.5%)
4	9 (0.9%)
5	6 (0.6%)
6	3 (0.3%)
8	2 (0.2%)
9	2 (0.2%)
SELECT_n_of_non_aggr_func___RTRIM	
0	632 (63%)
1	237 (24%)
2	83 (8.3%)

1

Query parameter or value	Descriptive statistics
3	21 (2.1%)
4	9 (0.9%)
5	10 (1.0%)
6	6 (0.6%)
7	2 (0.2%)
SELECT_n_of_non_aggr_func___SUBSTR	
0	625 (63%)
1	248 (25%)
2	71 (7.1%)
3	26 (2.6%)
4	16 (1.6%)
5	6 (0.6%)
6	3 (0.3%)
7	2 (0.2%)
8	3 (0.3%)
SELECT_n_of_non_aggr_func___UPPER	
0	607 (61%)
1	268 (27%)
2	75 (7.5%)
3	23 (2.3%)
4	11 (1.1%)
5	8 (0.8%)
6	5 (0.5%)
7	3 (0.3%)
SELECT_n_of_non_aggr_func___MONTH	
0	819 (82%)
1	124 (12%)
2	24 (2.4%)
3	19 (1.9%)
4	6 (0.6%)
5	1 (0.1%)
6	3 (0.3%)
7	1 (0.1%)
8	3 (0.3%)
SELECT_n_of_non_aggr_func___YEAR	
0	831 (83%)
1	109 (11%)
2	32 (3.2%)
3	13 (1.3%)
4	6 (0.6%)
5	3 (0.3%)
7	3 (0.3%)
8	2 (0.2%)
9	1 (0.1%)
SELECT_n_of_non_aggr_func___LOG	
0	852 (85%)

Query parameter or value	Descriptive statistics
1	105 (11%)
2	25 (2.5%)
3	9 (0.9%)
4	4 (0.4%)
5	2 (0.2%)
6	2 (0.2%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func___LTRIM	
0	611 (61%)
1	257 (26%)
2	79 (7.9%)
3	28 (2.8%)
4	11 (1.1%)
5	9 (0.9%)
6	2 (0.2%)
7	1 (0.1%)
8	2 (0.2%)
SELECT_n_of_non_aggr_func___ROUND	
0	920 (92%)
1	66 (6.6%)
2	12 (1.2%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___SQRT	
0	781 (78%)
1	153 (15%)
2	37 (3.7%)
3	14 (1.4%)
4	7 (0.7%)
5	5 (0.5%)
6	2 (0.2%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func___DOW	
0	835 (84%)
1	120 (12%)
2	22 (2.2%)
3	13 (1.3%)
4	6 (0.6%)
5	1 (0.1%)
6	3 (0.3%)
SELECT_n_of_non_aggr_func___DAY	
0	891 (89%)
1	79 (7.9%)
2	21 (2.1%)
3	6 (0.6%)
4	2 (0.2%)
5	1 (0.1%)

Query parameter or value	Descriptive statistics
SELECT_n_of_non_aggr_func___FLOOR	
0	921 (92%)
1	64 (6.4%)
2	11 (1.1%)
3	4 (0.4%)
SELECT_n_of_non_aggr_func___ABS	
0	919 (92%)
1	66 (6.6%)
2	11 (1.1%)
3	1 (0.1%)
4	3 (0.3%)
SELECT_n_of_non_aggr_func___TRUNC	
0	917 (92%)
1	65 (6.5%)
2	14 (1.4%)
3	4 (0.4%)
SELECT_n_of_all_non_aggr_func	[0, 3, 68] 5 / 8
SELECT_n_of_aggr_func___COUNT_DISTINCT	[0.00, 0.00, 14.00] 0.86 / 2.04
SELECT_n_of_aggr_func___MAX	[0.00, 0.00, 11.00] 0.84 / 1.85
SELECT_n_of_aggr_func___COUNT	[0.00, 0.00, 11.00] 0.73 / 1.73
SELECT_n_of_aggr_func___MIN	[0.00, 0.00, 11.00] 0.73 / 1.76
SELECT_n_of_aggr_func___AVG	
0	933 (93%)
1	43 (4.3%)
2	19 (1.9%)
3	5 (0.5%)
SELECT_n_of_aggr_func___SUM	
0	941 (94%)
1	45 (4.5%)
2	13 (1.3%)
3	1 (0.1%)
SELECT_n_of_all_aggr_func	[0.00, 3.00, 14.00] 3.32 / 2.97
FROM_n_of_join_paths	
1	180 (18%)
2	173 (17%)
3	176 (18%)
4	185 (19%)
5	138 (14%)
6	148 (15%)
FROM_n_of_super_joins___LEFT	
0	567 (57%)
1	310 (31%)
2	100 (10%)
3	19 (1.9%)
4	4 (0.4%)
FROM_n_of_super_joins___RIGHT	

Query parameter or value	Descriptive statistics
0	560 (56%)
1	305 (31%)
2	110 (11%)
3	21 (2.1%)
4	4 (0.4%)
FROM_n_of_super_joins___FULL	
0	572 (57%)
1	314 (31%)
2	89 (8.9%)
3	22 (2.2%)
4	3 (0.3%)
FROM_n_of_joins___INNER	[0.00, 3.00, 13.00] 3.15 / 2.43
FROM_n_of_joins___RIGHT	[0.00, 3.00, 14.00] 3.11 / 2.42
FROM_n_of_processed_rows	[5, 1,693,018,547, 8,390,110,684] 2,262,257,826 / 1,914,541,451
WHERE_n_of_predicates	[0.00, 3.00, 12.00] 3.67 / 2.89
WHERE_n_of_attribs_of_type___character_varying	
0	372 (37%)
1	299 (30%)
2	187 (19%)
3	93 (9.3%)
4	35 (3.5%)
5	9 (0.9%)
6	4 (0.4%)
7	1 (0.1%)
WHERE_n_of_attribs_of_type___integer	
0	374 (37%)
1	288 (29%)
2	171 (17%)
3	99 (9.9%)
4	49 (4.9%)
5	12 (1.2%)
6	5 (0.5%)
7	2 (0.2%)
WHERE_n_of_attribs_of_type___numeric	
0	601 (60%)
1	255 (26%)
2	107 (11%)
3	32 (3.2%)
4	3 (0.3%)
5	2 (0.2%)
WHERE_n_of_attribs_of_type___character	
0	669 (67%)
1	241 (24%)
2	75 (7.5%)
3	15 (1.5%)

Query parameter or value	Descriptive statistics
WHERE_n_of_attribs_of_type___date	
0	796 (80%)
1	161 (16%)
2	37 (3.7%)
3	6 (0.6%)
WHERE_n_of_pkey_attribs	
0	527 (53%)
1	302 (30%)
2	120 (12%)
3	40 (4.0%)
4	10 (1.0%)
5	1 (0.1%)
WHERE_n_of_connect__OR	[0.00, 1.00, 11.00] 2.05 / 2.35
WHERE_n_of_operators___between	
0	531 (53%)
1	267 (27%)
2	151 (15%)
3	39 (3.9%)
4	10 (1.0%)
5	1 (0.1%)
6	1 (0.1%)
WHERE_n_of_operators___greater_or_less	
0	345 (35%)
1	273 (27%)
2	191 (19%)
3	119 (12%)
4	37 (3.7%)
5	27 (2.7%)
6	6 (0.6%)
7	2 (0.2%)
WHERE_n_of_operators___in	
0	555 (56%)
1	280 (28%)
2	117 (12%)
3	31 (3.1%)
4	12 (1.2%)
5	5 (0.5%)
WHERE_n_of_operators___like	
0	795 (80%)
1	177 (18%)
2	25 (2.5%)
3	3 (0.3%)
WHERE_n_of_non_aggr_func___DAY	
0	962 (96%)
1	36 (3.6%)
2	2 (0.2%)

Query parameter or value	Descriptive statistics
WHERE_n_of_non_aggr_func___LOG	
0	905 (91%)
1	91 (9.1%)
2	4 (0.4%)
WHERE_n_of_non_aggr_func___MONTH	
0	961 (96%)
1	37 (3.7%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___YEAR	
0	957 (96%)
1	41 (4.1%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___FLOOR	
0	956 (96%)
1	44 (4.4%)
WHERE_n_of_non_aggr_func___ROUND	
0	955 (96%)
1	43 (4.3%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___SQRT	
0	869 (87%)
1	124 (12%)
2	7 (0.7%)
WHERE_n_of_non_aggr_func___ABS	
0	964 (96%)
1	36 (3.6%)
WHERE_n_of_non_aggr_func___DOW	
0	967 (97%)
1	30 (3.0%)
2	3 (0.3%)
WHERE_n_of_non_aggr_func___TRUNC	
0	952 (95%)
1	46 (4.6%)
2	2 (0.2%)
WHERE_n_of_all_non_aggr_func	
0	597 (60%)
1	271 (27%)
2	95 (9.5%)
3	33 (3.3%)
4	4 (0.4%)
GROUP_BY_n_of_columns	[0.00, 2.00, 12.00] 2.97 / 2.79
HAVING_n_of_main_predicates	
0	439 (44%)
1	211 (21%)
2	178 (18%)
3	172 (17%)

Query parameter or value	Descriptive statistics
ORDER_BY_n_of_columns	
0	289 (29%)
1	328 (33%)
2	235 (24%)
3	112 (11%)
4	21 (2.1%)
5	5 (0.5%)
6	7 (0.7%)
7	1 (0.1%)
8	2 (0.2%)
limit	[2, 494, 1,000] 504 / 290
offset	[0, 0, 997] 238 / 317