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

Scenario: 1-3jp_ujjp__g__h0-3ss_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, 117] 10 / 12
SELECT_n_of_non_aggr_func___LOWER	
0	709 (71%)
1	236 (24%)
2	47 (4.7%)
3	4 (0.4%)
4	2 (0.2%)
5	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func___RTRIM	
0	708 (71%)
1	243 (24%)
2	31 (3.1%)
3	13 (1.3%)
4	3 (0.3%)

Query parameter or value	Descriptive statistics
5	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func___SUBSTR	
0	719 (72%)
1	216 (22%)
2	48 (4.8%)
3	12 (1.2%)
4	4 (0.4%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___LTRIM	
0	686 (69%)
1	249 (25%)
2	52 (5.2%)
3	9 (0.9%)
4	3 (0.3%)
6	1 (0.1%)
SELECT_n_of_non_aggr_func___UPPER	
0	678 (68%)
1	254 (25%)
2	47 (4.7%)
3	11 (1.1%)
4	9 (0.9%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___MONTH	
0	853 (85%)
1	119 (12%)
2	19 (1.9%)
3	4 (0.4%)
4	3 (0.3%)
6	2 (0.2%)
SELECT_n_of_non_aggr_func___YEAR	
0	865 (87%)
1	110 (11%)
2	17 (1.7%)
3	5 (0.5%)
4	1 (0.1%)
5	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func___FLOOR	
0	951 (95%)
1	46 (4.6%)
2	2 (0.2%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___SQRT	
0	865 (87%)
1	111 (11%)

Query parameter or value	Descriptive statistics
2	19 (1.9%)
3	3 (0.3%)
4	1 (0.1%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___ABS	
0	946 (95%)
1	50 (5.0%)
2	4 (0.4%)
SELECT_n_of_non_aggr_func___DAY	
0	915 (92%)
1	72 (7.2%)
2	11 (1.1%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___DOW	
0	885 (89%)
1	103 (10%)
2	8 (0.8%)
3	1 (0.1%)
4	1 (0.1%)
5	2 (0.2%)
SELECT_n_of_non_aggr_func___ROUND	
0	951 (95%)
1	47 (4.7%)
2	2 (0.2%)
SELECT_n_of_non_aggr_func___LOG	
0	914 (91%)
1	78 (7.8%)
2	6 (0.6%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___TRUNC	
0	953 (95%)
1	41 (4.1%)
2	6 (0.6%)
SELECT_n_of_all_non_aggr_func	[0.00, 2.00, 49.00] 3.00 / 3.97
SELECT_n_of_aggr_func___MIN	[0.00, 0.00, 10.00] 0.73 / 1.57
SELECT_n_of_aggr_func___COUNT	[0.00, 0.00, 9.00] 0.64 / 1.47
SELECT_n_of_aggr_func___MAX	[0.00, 0.00, 10.00] 0.65 / 1.43
SELECT_n_of_aggr_func___COUNT_DISTINCT	
0	756 (76%)
1	84 (8.4%)
2	47 (4.7%)
3	33 (3.3%)
4	38 (3.8%)
5	18 (1.8%)
6	10 (1.0%)
7	10 (1.0%)

Query parameter or value	Descriptive statistics
8	4 (0.4%)
SELECT_n_of_aggr_func___SUM	
0	946 (95%)
1	45 (4.5%)
2	8 (0.8%)
3	1 (0.1%)
SELECT_n_of_aggr_func___AVG	
0	935 (94%)
1	45 (4.5%)
2	18 (1.8%)
3	2 (0.2%)
SELECT_n_of_all_aggr_func	[0.00, 2.00, 11.00] 2.86 / 2.30
FROM_n_of_join_paths	
1	332 (33%)
2	322 (32%)
3	346 (35%)
FROM_n_of_super_joins___LEFT	
0	778 (78%)
1	203 (20%)
2	19 (1.9%)
FROM_n_of_super_joins___FULL	
0	734 (73%)
1	232 (23%)
2	34 (3.4%)
FROM_n_of_super_joins___RIGHT	
0	795 (80%)
1	186 (19%)
2	19 (1.9%)
FROM_n_of_joins___RIGHT	[0.00, 1.00, 9.00] 1.75 / 1.51
FROM_n_of_joins___INNER	[0.00, 1.00, 9.00] 1.77 / 1.55
FROM_n_of_processed_rows	[5, 3,405,028,272, 11,475,084,811] 3,229,852,653 / 2,917,106,307
WHERE_n_of_predicates	[0.00, 2.00, 10.00] 2.83 / 2.22
WHERE_n_of_attribs_of_type___character_varying	
0	432 (43%)
1	289 (29%)
2	182 (18%)
3	66 (6.6%)
4	26 (2.6%)
5	5 (0.5%)
WHERE_n_of_attribs_of_type___integer	
0	450 (45%)
1	316 (32%)
2	153 (15%)
3	59 (5.9%)
4	19 (1.9%)

Query parameter or value	Descriptive statistics
5	3 (0.3%)
WHERE_n_of_attribs_of_type___date	
0	836 (84%)
1	146 (15%)
2	16 (1.6%)
3	2 (0.2%)
WHERE_n_of_attribs_of_type___numeric	
0	644 (64%)
1	256 (26%)
2	77 (7.7%)
3	20 (2.0%)
4	2 (0.2%)
5	1 (0.1%)
WHERE_n_of_attribs_of_type___character	
0	763 (76%)
1	188 (19%)
2	45 (4.5%)
3	4 (0.4%)
WHERE_n_of_pkey_attribs	
0	609 (61%)
1	289 (29%)
2	88 (8.8%)
3	12 (1.2%)
4	2 (0.2%)
WHERE_n_of_connect_OR	
0	455 (46%)
1	182 (18%)
2	138 (14%)
3	85 (8.5%)
4	73 (7.3%)
5	36 (3.6%)
6	21 (2.1%)
7	5 (0.5%)
8	5 (0.5%)
WHERE_n_of_operators___greater_or_less	
0	424 (42%)
1	265 (27%)
2	185 (19%)
3	83 (8.3%)
4	27 (2.7%)
5	14 (1.4%)
6	2 (0.2%)
WHERE_n_of_operators___in	
0	614 (61%)
1	277 (28%)
2	88 (8.8%)

Query parameter or value	Descriptive statistics
3	17 (1.7%)
4	3 (0.3%)
5	1 (0.1%)
WHERE_n_of_operators___between	
0	608 (61%)
1	274 (27%)
2	102 (10%)
3	13 (1.3%)
4	3 (0.3%)
WHERE_n_of_operators___like	
0	841 (84%)
1	147 (15%)
2	11 (1.1%)
3	1 (0.1%)
WHERE_n_of_non_aggr_func___ABS	
0	964 (96%)
1	36 (3.6%)
WHERE_n_of_non_aggr_func___FLOOR	
0	964 (96%)
1	36 (3.6%)
WHERE_n_of_non_aggr_func___DAY	
0	972 (97%)
1	27 (2.7%)
3	1 (0.1%)
WHERE_n_of_non_aggr_func___SQRT	
0	912 (91%)
1	85 (8.5%)
2	3 (0.3%)
WHERE_n_of_non_aggr_func___YEAR	
0	979 (98%)
1	20 (2.0%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___DOW	
0	971 (97%)
1	28 (2.8%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___LOG	
0	949 (95%)
1	50 (5.0%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___TRUNC	
0	975 (98%)
1	23 (2.3%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___ROUND	
0	975 (98%)

Query parameter or value	Descriptive statistics
1	23 (2.3%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___MONTH	
0	977 (98%)
1	23 (2.3%)
WHERE_n_of_all_non_aggr_func	
0	714 (71%)
1	219 (22%)
2	47 (4.7%)
3	19 (1.9%)
4	1 (0.1%)
GROUP_BY_n_of_columns	[0.00, 2.00, 9.00] 2.46 / 2.12
HAVING_n_of_main_predicates	
0	285 (29%)
1	97 (9.7%)
2	138 (14%)
3	200 (20%)
4	135 (14%)
5	108 (11%)
6	37 (3.7%)
HAVING_n_of_main_predicates___scalar_subquery	
0	431 (43%)
1	184 (18%)
2	204 (20%)
3	181 (18%)
HAVING_n_of_subqueries___scalar_subquery	[0.0, 1.0, 21.0] 2.4 / 3.7
HAVING_n_of_processed_rows_by_subqueries	[0, 150,000,065, 173,571,440,342] 7,646,899,130 / 16,511,796,288
ORDER_BY_n_of_columns	
0	326 (33%)
1	350 (35%)
2	244 (24%)
3	55 (5.5%)
4	13 (1.3%)
5	5 (0.5%)
6	3 (0.3%)
7	2 (0.2%)
8	2 (0.2%)
limit	[2, 496, 1,000] 494 / 295
offset	[0, 60, 997] 268 / 331