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

Scenario: 1-3jp_ujjp__g__h0ss_h0nss

Scale factor: 100 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, 107] 11 / 14
SELECT_n_of_non_aggr_func___LOG	
0	895 (90%)
1	88 (8.8%)
2	15 (1.5%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___LTRIM	
0	694 (69%)
1	233 (23%)
2	57 (5.7%)
3	9 (0.9%)
4	3 (0.3%)
5	2 (0.2%)
6	2 (0.2%)
SELECT_n_of_non_aggr_func___RTRIM	

Query parameter or value	Descriptive statistics
0	697 (70%)
1	232 (23%)
2	47 (4.7%)
3	17 (1.7%)
4	5 (0.5%)
5	1 (0.1%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func___LOWER	
0	684 (68%)
1	229 (23%)
2	65 (6.5%)
3	14 (1.4%)
4	3 (0.3%)
5	4 (0.4%)
6	1 (0.1%)
SELECT_n_of_non_aggr_func___UPPER	
0	704 (70%)
1	219 (22%)
2	64 (6.4%)
3	7 (0.7%)
4	4 (0.4%)
5	2 (0.2%)
SELECT_n_of_non_aggr_func___ABS	
0	946 (95%)
1	44 (4.4%)
2	6 (0.6%)
3	3 (0.3%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___SUBSTR	
0	680 (68%)
1	240 (24%)
2	65 (6.5%)
3	10 (1.0%)
4	3 (0.3%)
5	1 (0.1%)
6	1 (0.1%)
SELECT_n_of_non_aggr_func___SQRT	
0	829 (83%)
1	133 (13%)
2	24 (2.4%)
3	8 (0.8%)
4	4 (0.4%)
5	1 (0.1%)
8	1 (0.1%)
SELECT_n_of_non_aggr_func___FLOOR	
0	958 (96%)

Query parameter or value	Descriptive statistics
1	37 (3.7%)
2	4 (0.4%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___DOW	
0	891 (89%)
1	85 (8.5%)
2	16 (1.6%)
3	6 (0.6%)
5	2 (0.2%)
SELECT_n_of_non_aggr_func___MONTH	
0	852 (85%)
1	106 (11%)
2	31 (3.1%)
3	4 (0.4%)
4	3 (0.3%)
5	2 (0.2%)
6	2 (0.2%)
SELECT_n_of_non_aggr_func___TRUNC	
0	947 (95%)
1	52 (5.2%)
2	1 (0.1%)
SELECT_n_of_non_aggr_func___YEAR	
0	880 (88%)
1	86 (8.6%)
2	22 (2.2%)
3	9 (0.9%)
4	2 (0.2%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___DAY	
0	901 (90%)
1	80 (8.0%)
2	14 (1.4%)
3	3 (0.3%)
4	1 (0.1%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___ROUND	
0	947 (95%)
1	46 (4.6%)
2	7 (0.7%)
SELECT_n_of_all_non_aggr_func	[0.0, 2.0, 46.0] 3.3 / 4.8
SELECT_n_of_aggr_func___MIN	[0.00, 0.00, 10.00] 0.57 / 1.33
SELECT_n_of_aggr_func___SUM	
0	942 (94%)
1	44 (4.4%)
2	9 (0.9%)
3	4 (0.4%)

Query parameter or value	Descriptive statistics
4	1 (0.1%)
SELECT_n_of_aggr_func___MAX	[0.00, 0.00, 11.00] 0.70 / 1.57
SELECT_n_of_aggr_func___COUNT_DISTINCT	
0	773 (77%)
1	75 (7.5%)
2	57 (5.7%)
3	39 (3.9%)
4	28 (2.8%)
5	18 (1.8%)
6	4 (0.4%)
7	4 (0.4%)
8	2 (0.2%)
SELECT_n_of_aggr_func___AVG	
0	932 (93%)
1	50 (5.0%)
2	13 (1.3%)
3	5 (0.5%)
SELECT_n_of_aggr_func___COUNT	[0.00, 0.00, 9.00] 0.61 / 1.48
SELECT_n_of_all_aggr_func	[0.00, 2.00, 11.00] 2.63 / 2.30
FROM_n_of_join_paths	
1	359 (36%)
2	324 (32%)
3	317 (32%)
FROM_n_of_super_joins___FULL	
0	778 (78%)
1	203 (20%)
2	19 (1.9%)
FROM_n_of_super_joins___RIGHT	
0	795 (80%)
1	184 (18%)
2	21 (2.1%)
FROM_n_of_super_joins___LEFT	
0	784 (78%)
1	192 (19%)
2	24 (2.4%)
FROM_n_of_joins___INNER	
0	263 (26%)
1	263 (26%)
2	224 (22%)
3	118 (12%)
4	72 (7.2%)
5	33 (3.3%)
6	16 (1.6%)
7	6 (0.6%)
8	5 (0.5%)
FROM_n_of_joins___RIGHT	

Query parameter or value	Descriptive statistics
0	245 (25%)
1	250 (25%)
2	230 (23%)
3	136 (14%)
4	79 (7.9%)
5	28 (2.8%)
6	26 (2.6%)
7	6 (0.6%)
FROM_n_of_processed_rows	[5, 681,037,927, 2,295,113,731] 660,041,223 / 622,698,872
WHERE_n_of_predicates	[0.00, 2.00, 10.00] 2.80 / 2.23
WHERE_n_of_attribs_of_type___character_varying	
0	447 (45%)
1	323 (32%)
2	174 (17%)
3	42 (4.2%)
4	9 (0.9%)
5	5 (0.5%)
WHERE_n_of_attribs_of_type___character	
0	700 (70%)
1	243 (24%)
2	47 (4.7%)
3	7 (0.7%)
4	3 (0.3%)
WHERE_n_of_attribs_of_type___numeric	
0	632 (63%)
1	265 (27%)
2	82 (8.2%)
3	18 (1.8%)
4	3 (0.3%)
WHERE_n_of_attribs_of_type___integer	
0	448 (45%)
1	326 (33%)
2	150 (15%)
3	61 (6.1%)
4	14 (1.4%)
6	1 (0.1%)
WHERE_n_of_attribs_of_type___date	
0	832 (83%)
1	133 (13%)
2	31 (3.1%)
3	4 (0.4%)
WHERE_n_of_pkey_attribs	
0	624 (62%)
1	279 (28%)
2	80 (8.0%)

Query parameter or value	Descriptive statistics
3	15 (1.5%)
4	1 (0.1%)
5	1 (0.1%)
WHERE_n_of_connect_OR	
0	455 (46%)
1	185 (19%)
2	124 (12%)
3	109 (11%)
4	62 (6.2%)
5	39 (3.9%)
6	21 (2.1%)
7	3 (0.3%)
8	2 (0.2%)
WHERE_n_of_operators___greater_or_less	
0	417 (42%)
1	309 (31%)
2	158 (16%)
3	78 (7.8%)
4	29 (2.9%)
5	6 (0.6%)
6	2 (0.2%)
7	1 (0.1%)
WHERE_n_of_operators___between	
0	618 (62%)
1	275 (28%)
2	78 (7.8%)
3	23 (2.3%)
4	6 (0.6%)
WHERE_n_of_operators___in	
0	612 (61%)
1	278 (28%)
2	87 (8.7%)
3	19 (1.9%)
4	4 (0.4%)
WHERE_n_of_operators___like	
0	823 (82%)
1	165 (17%)
2	12 (1.2%)
WHERE_n_of_non_aggr_func___TRUNC	
0	971 (97%)
1	29 (2.9%)
WHERE_n_of_non_aggr_func___FLOOR	
0	964 (96%)
1	36 (3.6%)
WHERE_n_of_non_aggr_func___LOG	
0	944 (94%)

Query parameter or value	Descriptive statistics
1	55 (5.5%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___DOW	
0	968 (97%)
1	30 (3.0%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___SQRT	
0	890 (89%)
1	106 (11%)
2	4 (0.4%)
WHERE_n_of_non_aggr_func___YEAR	
0	973 (97%)
1	27 (2.7%)
WHERE_n_of_non_aggr_func___MONTH	
0	968 (97%)
1	32 (3.2%)
WHERE_n_of_non_aggr_func___DAY	
0	975 (98%)
1	25 (2.5%)
WHERE_n_of_non_aggr_func___ROUND	
0	978 (98%)
1	22 (2.2%)
WHERE_n_of_non_aggr_func___ABS	
0	969 (97%)
1	31 (3.1%)
WHERE_n_of_all_non_aggr_func	
0	683 (68%)
1	240 (24%)
2	65 (6.5%)
3	11 (1.1%)
4	1 (0.1%)
GROUP_BY_n_of_columns	[0.00, 2.00, 10.00] 2.37 / 2.14
HAVING_n_of_main_predicates	
0	438 (44%)
1	187 (19%)
2	187 (19%)
3	188 (19%)
ORDER_BY_n_of_columns	
0	299 (30%)
1	366 (37%)
2	245 (25%)
3	66 (6.6%)
4	16 (1.6%)
5	2 (0.2%)
6	5 (0.5%)
7	1 (0.1%)

Query parameter or value	Descriptive statistics
limit	[4, 494, 1,000] 492 / 290
offset	[0, 0, 999] 239 / 321