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

Scenario: 1-1jp__g__h0-3ss_h0-3nss

Scale factor: 10 GB

The 1000-query pack was generated on: 2023-11-05

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, 43.0] 5.7 / 5.3
SELECT_n_of_non_aggr_func___LTRIM	
0	774 (77%)
1	204 (20%)
2	19 (1.9%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func___SUBSTR	
0	764 (76%)
1	204 (20%)
2	28 (2.8%)
3	4 (0.4%)
SELECT_n_of_non_aggr_func___LOWER	
0	779 (78%)
1	186 (19%)
2	29 (2.9%)

Query parameter or value	Descriptive statistics
3	6 (0.6%)
SELECT_n_of_non_aggr_func___SQRT	
0	879 (88%)
1	113 (11%)
2	7 (0.7%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___UPPER	
0	759 (76%)
1	202 (20%)
2	35 (3.5%)
3	2 (0.2%)
4	2 (0.2%)
SELECT_n_of_non_aggr_func___MONTH	
0	903 (90%)
1	82 (8.2%)
2	12 (1.2%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func___YEAR	
0	911 (91%)
1	78 (7.8%)
2	9 (0.9%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___ROUND	
0	954 (95%)
1	41 (4.1%)
2	5 (0.5%)
SELECT_n_of_non_aggr_func___RTRIM	
0	780 (78%)
1	189 (19%)
2	27 (2.7%)
3	4 (0.4%)
SELECT_n_of_non_aggr_func___FLOOR	
0	967 (97%)
1	32 (3.2%)
2	1 (0.1%)
SELECT_n_of_non_aggr_func___LOG	
0	935 (94%)
1	63 (6.3%)
2	2 (0.2%)
SELECT_n_of_non_aggr_func___DAY	
0	945 (95%)
1	48 (4.8%)
2	7 (0.7%)
SELECT_n_of_non_aggr_func___DOW	
0	923 (92%)
1	72 (7.2%)

Query parameter or value	Descriptive statistics
2	5 (0.5%)
SELECT_n_of_non_aggr_func___TRUNC	
0	962 (96%)
1	38 (3.8%)
SELECT_n_of_non_aggr_func___ABS	
0	964 (96%)
1	35 (3.5%)
2	1 (0.1%)
SELECT_n_of_all_non_aggr_func	[0.00, 2.00, 21.00] 2.04 / 2.22
SELECT_n_of_aggr_func___COUNT	
0	786 (79%)
1	119 (12%)
2	54 (5.4%)
3	27 (2.7%)
4	8 (0.8%)
5	2 (0.2%)
6	2 (0.2%)
7	1 (0.1%)
8	1 (0.1%)
SELECT_n_of_aggr_func___AVG	
0	941 (94%)
1	54 (5.4%)
2	5 (0.5%)
SELECT_n_of_aggr_func___MIN	
0	763 (76%)
1	109 (11%)
2	64 (6.4%)
3	40 (4.0%)
4	14 (1.4%)
5	4 (0.4%)
6	4 (0.4%)
7	1 (0.1%)
8	1 (0.1%)
SELECT_n_of_aggr_func___COUNT_DISTINCT	
0	746 (75%)
1	130 (13%)
2	60 (6.0%)
3	33 (3.3%)
4	23 (2.3%)
5	5 (0.5%)
6	3 (0.3%)
SELECT_n_of_aggr_func___MAX	
0	789 (79%)
1	91 (9.1%)
2	61 (6.1%)
3	33 (3.3%)

Query parameter or value	Descriptive statistics
4	16 (1.6%)
5	8 (0.8%)
6	1 (0.1%)
9	1 (0.1%)
SELECT_n_of_aggr_func___SUM	
0	959 (96%)
1	34 (3.4%)
2	5 (0.5%)
3	2 (0.2%)
SELECT_n_of_all_aggr_func	[0.00, 2.00, 9.00] 1.88 / 1.53
FROM_n_of_join_paths	
1	1,000 (100%)
FROM_n_of_joins___RIGHT	
0	474 (47%)
1	342 (34%)
2	141 (14%)
3	37 (3.7%)
4	6 (0.6%)
FROM_n_of_joins___INNER	
0	473 (47%)
1	340 (34%)
2	148 (15%)
3	36 (3.6%)
4	3 (0.3%)
FROM_n_of_processed_rows	[5, 8,101,071, 76,480,513] 28,136,838 / 31,467,415
WHERE_n_of_predicates	
0	228 (23%)
1	244 (24%)
2	217 (22%)
3	171 (17%)
4	86 (8.6%)
5	40 (4.0%)
6	14 (1.4%)
WHERE_n_of_attribs_of_type___character_varying	
0	538 (54%)
1	325 (33%)
2	116 (12%)
3	21 (2.1%)
WHERE_n_of_attribs_of_type___integer	
0	561 (56%)
1	319 (32%)
2	102 (10%)
3	17 (1.7%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type___numeric	
0	748 (75%)

Query parameter or value	Descriptive statistics
1	216 (22%)
2	31 (3.1%)
3	5 (0.5%)
WHERE_n_of_attribs_of_type___character	
0	807 (81%)
1	164 (16%)
2	26 (2.6%)
3	3 (0.3%)
WHERE_n_of_attribs_of_type___date	
0	904 (90%)
1	89 (8.9%)
2	7 (0.7%)
WHERE_n_of_pkey_attribs	
0	719 (72%)
1	238 (24%)
2	42 (4.2%)
3	1 (0.1%)
WHERE_n_of_connect_OR	
0	602 (60%)
1	209 (21%)
2	124 (12%)
3	51 (5.1%)
4	13 (1.3%)
5	1 (0.1%)
WHERE_n_of_operators___between	
0	729 (73%)
1	224 (22%)
2	43 (4.3%)
3	4 (0.4%)
WHERE_n_of_operators___in	
0	713 (71%)
1	229 (23%)
2	53 (5.3%)
3	3 (0.3%)
4	2 (0.2%)
WHERE_n_of_operators___greater_or_less	
0	533 (53%)
1	295 (30%)
2	127 (13%)
3	37 (3.7%)
4	6 (0.6%)
5	2 (0.2%)
WHERE_n_of_operators___like	
0	889 (89%)
1	104 (10%)
2	7 (0.7%)

Query parameter or value	Descriptive statistics
WHERE_n_of_non_aggr_func___SQRT	
0	939 (94%)
1	59 (5.9%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___LOG	
0	951 (95%)
1	49 (4.9%)
WHERE_n_of_non_aggr_func___ROUND	
0	981 (98%)
1	18 (1.8%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___DOW	
0	985 (99%)
1	15 (1.5%)
WHERE_n_of_non_aggr_func___MONTH	
0	986 (99%)
1	14 (1.4%)
WHERE_n_of_non_aggr_func___DAY	
0	987 (99%)
1	13 (1.3%)
WHERE_n_of_non_aggr_func___YEAR	
0	991 (99%)
1	8 (0.8%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___ABS	
0	980 (98%)
1	20 (2.0%)
WHERE_n_of_non_aggr_func___TRUNC	
0	988 (99%)
1	12 (1.2%)
WHERE_n_of_non_aggr_func___FLOOR	
0	980 (98%)
1	19 (1.9%)
2	1 (0.1%)
WHERE_n_of_all_non_aggr_func	
0	805 (81%)
1	157 (16%)
2	34 (3.4%)
3	4 (0.4%)
GROUP_BY_n_of_columns	
0	238 (24%)
1	269 (27%)
2	215 (22%)
3	153 (15%)
4	72 (7.2%)
5	47 (4.7%)

Query parameter or value	Descriptive statistics
6	5 (0.5%)
7	1 (0.1%)
HAVING_n_of_main_predicates	[0.00, 4.00, 9.00] 3.38 / 2.50
HAVING_n_of_main_predicates___non_scalar_subquery	
0	441 (44%)
1	204 (20%)
2	189 (19%)
3	166 (17%)
HAVING_n_of_main_predicates___scalar_subquery	
0	431 (43%)
1	189 (19%)
2	201 (20%)
3	179 (18%)
HAVING_n_of_subqueries___non_scalar_subquery	
0	441 (44%)
1	204 (20%)
2	189 (19%)
3	166 (17%)
HAVING_n_of_subqueries___scalar_subquery	[0.0, 1.0, 18.0] 2.4 / 3.5
HAVING_n_of_processed_rows_by_subqueries	[0, 11,248,740, 1,300,168,636] 94,428,449 / 185,342,903
ORDER_BY_n_of_columns	
0	300 (30%)
1	480 (48%)
2	189 (19%)
3	27 (2.7%)
4	3 (0.3%)
5	1 (0.1%)
limit	[1, 507, 1,000] 506 / 290
offset	[0, 0, 999] 243 / 325