

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

Scenario: 1-6jp_rjpp_g_h0-3ss_h0nss

Scale factor: 200 GB

The 1000-query pack was generated on: 2023-11-07

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, 178] 17 / 26
SELECT_n_of_non_aggr_func__LTRIM	[0.00, 0.00, 9.00] 0.61 / 1.10
SELECT_n_of_non_aggr_func__RTRIM	[0.00, 0.00, 10.00] 0.66 / 1.18
SELECT_n_of_non_aggr_func__SQRT	[0.00, 0.00, 9.00] 0.33 / 0.94
SELECT_n_of_non_aggr_func__SUBSTR	[0.00, 0.00, 9.00] 0.62 / 1.10
SELECT_n_of_non_aggr_func__DOW	
0	847 (85%)
1	114 (11%)
2	22 (2.2%)
3	7 (0.7%)
4	4 (0.4%)
5	5 (0.5%)
6	1 (0.1%)
SELECT_n_of_non_aggr_func__LOWER	[0.00, 0.00, 9.00] 0.64 / 1.17
SELECT_n_of_non_aggr_func__MONTH	[0.00, 0.00, 10.00] 0.36 / 1.01

Query parameter or value	Descriptive statistics
SELECT_n_of_non_aggr_func__YEAR	[0.00, 0.00, 9.00] 0.30 / 0.92
SELECT_n_of_non_aggr_func__UPPER	[0.00, 0.00, 9.00] 0.60 / 1.11
SELECT_n_of_non_aggr_func__ABS	
0	933 (93%)
1	51 (5.1%)
2	13 (1.3%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__FLOOR	
0	918 (92%)
1	72 (7.2%)
2	9 (0.9%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func__LOG	
0	849 (85%)
1	110 (11%)
2	20 (2.0%)
3	11 (1.1%)
4	4 (0.4%)
5	2 (0.2%)
6	3 (0.3%)
7	1 (0.1%)
SELECT_n_of_non_aggr_func__ROUND	
0	912 (91%)
1	71 (7.1%)
2	11 (1.1%)
3	2 (0.2%)
4	2 (0.2%)
5	2 (0.2%)
SELECT_n_of_non_aggr_func__TRUNC	
0	935 (94%)
1	53 (5.3%)
2	7 (0.7%)
3	4 (0.4%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func__DAY	
0	877 (88%)
1	84 (8.4%)
2	24 (2.4%)
3	8 (0.8%)
4	3 (0.3%)
5	2 (0.2%)
6	2 (0.2%)
SELECT_n_of_all_non_aggr_func	[0.0, 3.0, 81.0] 5.1 / 9.1
SELECT_n_of_aggr_func__COUNT	[0.00, 0.00, 11.00] 0.72 / 1.79
SELECT_n_of_aggr_func__MAX	[0.00, 0.00, 12.00] 0.92 / 2.08

Query parameter or value	Descriptive statistics
SELECT_n_of_aggr_func__MIN	[0.00, 0.00, 11.00] 0.93 / 2.01
SELECT_n_of_aggr_func__AVG	
0	933 (93%)
1	42 (4.2%)
2	17 (1.7%)
3	7 (0.7%)
4	1 (0.1%)
SELECT_n_of_aggr_func__SUM	
0	947 (95%)
1	36 (3.6%)
2	13 (1.3%)
3	4 (0.4%)
SELECT_n_of_aggr_func__COUNT_DISTINCT	[0.00, 0.00, 12.00] 0.82 / 1.89
SELECT_n_of_all_aggr_func	[0.00, 3.00, 14.00] 3.56 / 3.13
FROM_n_of_join_paths	
1	177 (18%)
2	148 (15%)
3	142 (14%)
4	173 (17%)
5	167 (17%)
6	193 (19%)
FROM_n_of_super_joins__FULL	
0	525 (53%)
1	330 (33%)
2	120 (12%)
3	25 (2.5%)
FROM_n_of_super_joins__LEFT	
0	554 (55%)
1	304 (30%)
2	119 (12%)
3	17 (1.7%)
4	6 (0.6%)
FROM_n_of_super_joins__RIGHT	
0	534 (53%)
1	300 (30%)
2	134 (13%)
3	30 (3.0%)
4	2 (0.2%)
FROM_n_of_joins__INNER	[0.00, 3.00, 13.00] 3.41 / 2.61
FROM_n_of_joins__RIGHT	[0.00, 3.00, 12.00] 3.36 / 2.55
FROM_n_of_processed_rows	[5, 1,890,018,514, 8,714,110,709] 2,436,925,987 / 2,042,211,810
WHERE_n_of_predicates	[0.0, 3.0, 13.0] 3.9 / 3.0
WHERE_n_of_attribs_of_type__character	
0	653 (65%)
1	238 (24%)

Query parameter or value	Descriptive statistics
2	78 (7.8%)
3	20 (2.0%)
4	11 (1.1%)
WHERE_n_of_attribs_of_type__character_varying	
0	345 (35%)
1	294 (29%)
2	189 (19%)
3	103 (10%)
4	50 (5.0%)
5	15 (1.5%)
7	4 (0.4%)
WHERE_n_of_attribs_of_type__integer	
0	396 (40%)
1	285 (29%)
2	169 (17%)
3	93 (9.3%)
4	38 (3.8%)
5	17 (1.7%)
6	2 (0.2%)
WHERE_n_of_attribs_of_type__numeric	
0	567 (57%)
1	263 (26%)
2	106 (11%)
3	47 (4.7%)
4	12 (1.2%)
5	3 (0.3%)
6	1 (0.1%)
7	1 (0.1%)
WHERE_n_of_attribs_of_type__date	
0	790 (79%)
1	171 (17%)
2	33 (3.3%)
3	4 (0.4%)
4	2 (0.2%)
WHERE_n_of_pkey_attribs	
0	550 (55%)
1	299 (30%)
2	113 (11%)
3	31 (3.1%)
4	3 (0.3%)
5	4 (0.4%)
WHERE_n_of_connect_OR	[0.00, 1.00, 12.00] 2.25 / 2.54
WHERE_n_of_operators__between	
0	537 (54%)
1	279 (28%)
2	124 (12%)

Query parameter or value	Descriptive statistics
3	50 (5.0%)
4	8 (0.8%)
5	2 (0.2%)
WHERE_n_of_operators___greater_or_less	
0	347 (35%)
1	243 (24%)
2	204 (20%)
3	99 (9.9%)
4	58 (5.8%)
5	33 (3.3%)
6	10 (1.0%)
7	6 (0.6%)
WHERE_n_of_operators___like	
0	793 (79%)
1	178 (18%)
2	28 (2.8%)
3	1 (0.1%)
WHERE_n_of_operators___in	
0	515 (52%)
1	304 (30%)
2	120 (12%)
3	50 (5.0%)
4	8 (0.8%)
5	3 (0.3%)
WHERE_n_of_non_aggr_func___LOG	
0	916 (92%)
1	81 (8.1%)
2	3 (0.3%)
WHERE_n_of_non_aggr_func___YEAR	
0	970 (97%)
1	30 (3.0%)
WHERE_n_of_non_aggr_func___ABS	
0	956 (96%)
1	42 (4.2%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___SQRT	
0	888 (89%)
1	104 (10%)
2	7 (0.7%)
3	1 (0.1%)
WHERE_n_of_non_aggr_func___FLOOR	
0	956 (96%)
1	42 (4.2%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___TRUNC	
0	957 (96%)

Query parameter or value	Descriptive statistics
1	43 (4.3%)
WHERE_n_of_non_aggr_func__ROUND	
0	949 (95%)
1	50 (5.0%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__DAY	
0	971 (97%)
1	29 (2.9%)
WHERE_n_of_non_aggr_func__MONTH	
0	960 (96%)
1	38 (3.8%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func__DOW	
0	967 (97%)
1	31 (3.1%)
2	2 (0.2%)
WHERE_n_of_all_non_aggr_func	
0	615 (62%)
1	279 (28%)
2	78 (7.8%)
3	17 (1.7%)
4	10 (1.0%)
5	1 (0.1%)
GROUP_BY_n_of_columns	[0.00, 2.00, 12.00] 3.09 / 2.91
HAVING_n_of_main_predicates	
0	289 (29%)
1	95 (9.5%)
2	125 (13%)
3	217 (22%)
4	125 (13%)
5	101 (10%)
6	48 (4.8%)
HAVING_n_of_main_predicates__scalar_subquery	
0	417 (42%)
1	191 (19%)
2	198 (20%)
3	194 (19%)
HAVING_n_of_subqueries__scalar_subquery	[0.0, 1.0, 21.0] 2.6 / 3.7
HAVING_n_of_processed_rows_by_subqueries	[148,000,135, 138,929,752,845] 6,319,985,795 / 13,403,266,385
ORDER_BY_n_of_columns	
0	322 (32%)
1	338 (34%)
2	225 (23%)
3	76 (7.6%)
4	28 (2.8%)

Query parameter or value	Descriptive statistics
5	4 (0.4%)
6	4 (0.4%)
7	3 (0.3%)
limit	[1, 505, 1,000] 506 / 292
offset	[0, 0, 1,000] 246 / 321