

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

Scenario: 1-1jp__g__h0ss_h0nss

Scale factor: 200 GB

The 1000-query pack was generated on: 2023-11-04

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, 41.0] 5.7 / 5.5
SELECT_n_of_non_aggr_func__LOWER	
0	806 (81%)
1	169 (17%)
2	24 (2.4%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func__SUBSTR	
0	797 (80%)
1	177 (18%)
2	23 (2.3%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__UPPER	
0	801 (80%)
1	170 (17%)

Query parameter or value	Descriptive statistics
2	24 (2.4%)
3	4 (0.4%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__RTRIM	
0	783 (78%)
1	187 (19%)
2	28 (2.8%)
3	1 (0.1%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func__MONTH	
0	904 (90%)
1	83 (8.3%)
2	10 (1.0%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func__LTRIM	
0	768 (77%)
1	201 (20%)
2	31 (3.1%)
SELECT_n_of_non_aggr_func__ROUND	
0	961 (96%)
1	38 (3.8%)
2	1 (0.1%)
SELECT_n_of_non_aggr_func__SQRT	
0	915 (92%)
1	73 (7.3%)
2	10 (1.0%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func__ABS	
0	972 (97%)
1	28 (2.8%)
SELECT_n_of_non_aggr_func__YEAR	
0	922 (92%)
1	67 (6.7%)
2	11 (1.1%)
SELECT_n_of_non_aggr_func__DAY	
0	952 (95%)
1	47 (4.7%)
2	1 (0.1%)
SELECT_n_of_non_aggr_func__DOW	
0	937 (94%)
1	54 (5.4%)
2	8 (0.8%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func__TRUNC	
0	957 (96%)
1	42 (4.2%)

Query parameter or value	Descriptive statistics
3	1 (0.1%)
SELECT_n_of_non_aggr_func__LOG	
0	926 (93%)
1	70 (7.0%)
2	4 (0.4%)
SELECT_n_of_non_aggr_func__FLOOR	
0	961 (96%)
1	38 (3.8%)
2	1 (0.1%)
SELECT_n_of_all_non_aggr_func	[0.00, 1.00, 17.00] 1.85 / 2.09
SELECT_n_of_aggr_func__MAX	
0	786 (79%)
1	111 (11%)
2	53 (5.3%)
3	28 (2.8%)
4	11 (1.1%)
5	7 (0.7%)
6	4 (0.4%)
SELECT_n_of_aggr_func__COUNT	
0	763 (76%)
1	129 (13%)
2	56 (5.6%)
3	37 (3.7%)
4	9 (0.9%)
5	3 (0.3%)
6	2 (0.2%)
7	1 (0.1%)
SELECT_n_of_aggr_func__COUNT_DISTINCT	
0	769 (77%)
1	108 (11%)
2	65 (6.5%)
3	35 (3.5%)
4	12 (1.2%)
5	9 (0.9%)
6	1 (0.1%)
7	1 (0.1%)
SELECT_n_of_aggr_func__AVG	
0	962 (96%)
1	34 (3.4%)
2	4 (0.4%)
SELECT_n_of_aggr_func__MIN	
0	777 (78%)
1	112 (11%)
2	69 (6.9%)
3	19 (1.9%)
4	14 (1.4%)

Query parameter or value	Descriptive statistics
5	8 (0.8%)
6	1 (0.1%)
SELECT_n_of_aggr_func__SUM	
0	948 (95%)
1	42 (4.2%)
2	9 (0.9%)
3	1 (0.1%)
SELECT_n_of_all_aggr_func	
0	192 (19%)
1	302 (30%)
2	239 (24%)
3	144 (14%)
4	62 (6.2%)
5	39 (3.9%)
6	20 (2.0%)
7	2 (0.2%)
FROM_n_of_join_paths	
1	1,000 (100%)
FROM_n_of_joins__RIGHT	
0	515 (52%)
1	297 (30%)
2	147 (15%)
3	35 (3.5%)
4	6 (0.6%)
FROM_n_of_joins__INNER	
0	492 (49%)
1	343 (34%)
2	118 (12%)
3	44 (4.4%)
4	3 (0.3%)
FROM_n_of_processed_rows	[5, 162,000,030, 1,530,018,464] 546,598,189 / 622,473,229
WHERE_n_of_predicates	
0	229 (23%)
1	236 (24%)
2	220 (22%)
3	165 (17%)
4	94 (9.4%)
5	47 (4.7%)
6	9 (0.9%)
WHERE_n_of_attribs_of_type__character	
0	818 (82%)
1	150 (15%)
2	30 (3.0%)
3	2 (0.2%)
WHERE_n_of_attribs_of_type__integer	

Query parameter or value	Descriptive statistics
<i>0</i>	559 (56%)
<i>1</i>	326 (33%)
<i>2</i>	93 (9.3%)
<i>3</i>	19 (1.9%)
<i>4</i>	3 (0.3%)
WHERE_n_of_attribs_of_type__numeric	
<i>0</i>	727 (73%)
<i>1</i>	221 (22%)
<i>2</i>	50 (5.0%)
<i>3</i>	2 (0.2%)
WHERE_n_of_attribs_of_type__character_varying	
<i>0</i>	554 (55%)
<i>1</i>	318 (32%)
<i>2</i>	103 (10%)
<i>3</i>	24 (2.4%)
<i>4</i>	1 (0.1%)
WHERE_n_of_attribs_of_type__date	
<i>0</i>	898 (90%)
<i>1</i>	93 (9.3%)
<i>2</i>	8 (0.8%)
<i>3</i>	1 (0.1%)
WHERE_n_of_pkey_attribs	
<i>0</i>	709 (71%)
<i>1</i>	249 (25%)
<i>2</i>	40 (4.0%)
<i>3</i>	1 (0.1%)
<i>4</i>	1 (0.1%)
WHERE_n_of_connect_OR	
<i>0</i>	608 (61%)
<i>1</i>	212 (21%)
<i>2</i>	106 (11%)
<i>3</i>	54 (5.4%)
<i>4</i>	18 (1.8%)
<i>5</i>	2 (0.2%)
WHERE_n_of_operators__greater_or_less	
<i>0</i>	540 (54%)
<i>1</i>	306 (31%)
<i>2</i>	118 (12%)
<i>3</i>	32 (3.2%)
<i>4</i>	3 (0.3%)
<i>5</i>	1 (0.1%)
WHERE_n_of_operators__between	
<i>0</i>	696 (70%)
<i>1</i>	261 (26%)
<i>2</i>	40 (4.0%)
<i>3</i>	3 (0.3%)

Query parameter or value	Descriptive statistics
WHERE_n_of_operators_in	
0	707 (71%)
1	246 (25%)
2	42 (4.2%)
3	4 (0.4%)
4	1 (0.1%)
WHERE_n_of_operators_like	
0	891 (89%)
1	102 (10%)
2	7 (0.7%)
WHERE_n_of_non_aggr_func__SQRT	
0	947 (95%)
1	47 (4.7%)
2	6 (0.6%)
WHERE_n_of_non_aggr_func__TRUNC	
0	979 (98%)
1	21 (2.1%)
WHERE_n_of_non_aggr_func__DAY	
0	975 (98%)
1	25 (2.5%)
WHERE_n_of_non_aggr_func__FLOOR	
0	982 (98%)
1	18 (1.8%)
WHERE_n_of_non_aggr_func__LOG	
0	964 (96%)
1	35 (3.5%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__MONTH	
0	987 (99%)
1	13 (1.3%)
WHERE_n_of_non_aggr_func__ABS	
0	985 (99%)
1	15 (1.5%)
WHERE_n_of_non_aggr_func__ROUND	
0	977 (98%)
1	22 (2.2%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__DOW	
0	984 (98%)
1	16 (1.6%)
WHERE_n_of_non_aggr_func__YEAR	
0	980 (98%)
1	20 (2.0%)
WHERE_n_of_all_non_aggr_func	
0	789 (79%)
1	176 (18%)

Query parameter or value	Descriptive statistics
<i>2</i>	33 (3.3%)
<i>3</i>	2 (0.2%)
GROUP_BY_n_of_columns	
<i>0</i>	248 (25%)
<i>1</i>	272 (27%)
<i>2</i>	233 (23%)
<i>3</i>	141 (14%)
<i>4</i>	68 (6.8%)
<i>5</i>	30 (3.0%)
<i>6</i>	7 (0.7%)
<i>7</i>	1 (0.1%)
HAVING_n_of_main_predicates	
<i>0</i>	430 (43%)
<i>1</i>	209 (21%)
<i>2</i>	170 (17%)
<i>3</i>	191 (19%)
ORDER_BY_n_of_columns	
<i>0</i>	300 (30%)
<i>1</i>	489 (49%)
<i>2</i>	177 (18%)
<i>3</i>	23 (2.3%)
<i>4</i>	9 (0.9%)
<i>5</i>	2 (0.2%)
limit	[1, 492, 1,000] 494 / 286
offset	[0, 41, 1,000] 258 / 322