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

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

Scale factor: 2000 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, 40.0] 5.4 / 4.9
SELECT_n_of_non_aggr_func___SUBSTR	
0	770 (77%)
1	205 (21%)
2	24 (2.4%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___UPPER	
0	776 (78%)
1	200 (20%)
2	21 (2.1%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func___LOWER	
0	777 (78%)
1	187 (19%)
2	31 (3.1%)

Query parameter or value	Descriptive statistics
3	4 (0.4%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___LTRIM	
0	771 (77%)
1	204 (20%)
2	23 (2.3%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___LOG	
0	929 (93%)
1	59 (5.9%)
2	12 (1.2%)
SELECT_n_of_non_aggr_func___RTRIM	
0	793 (79%)
1	182 (18%)
2	22 (2.2%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func___ROUND	
0	965 (97%)
1	34 (3.4%)
2	1 (0.1%)
SELECT_n_of_non_aggr_func___MONTH	
0	905 (91%)
1	84 (8.4%)
2	10 (1.0%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___SQRT	
0	906 (91%)
1	89 (8.9%)
2	5 (0.5%)
SELECT_n_of_non_aggr_func___DOW	
0	933 (93%)
1	60 (6.0%)
2	5 (0.5%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___YEAR	
0	929 (93%)
1	64 (6.4%)
2	5 (0.5%)
3	2 (0.2%)
SELECT_n_of_non_aggr_func___DAY	
0	950 (95%)
1	46 (4.6%)
2	4 (0.4%)
SELECT_n_of_non_aggr_func___ABS	
0	959 (96%)
1	39 (3.9%)

Query parameter or value	Descriptive statistics
2	2 (0.2%)
SELECT_n_of_non_aggr_func___TRUNC	
0	973 (97%)
1	27 (2.7%)
SELECT_n_of_non_aggr_func___FLOOR	
0	961 (96%)
1	39 (3.9%)
SELECT_n_of_all_non_aggr_func	[0.00, 1.00, 17.00] 1.91 / 1.99
SELECT_n_of_aggr_func___MIN	
0	757 (76%)
1	121 (12%)
2	60 (6.0%)
3	42 (4.2%)
4	11 (1.1%)
5	6 (0.6%)
6	3 (0.3%)
SELECT_n_of_aggr_func___MAX	
0	776 (78%)
1	124 (12%)
2	53 (5.3%)
3	24 (2.4%)
4	12 (1.2%)
5	7 (0.7%)
6	2 (0.2%)
7	1 (0.1%)
8	1 (0.1%)
SELECT_n_of_aggr_func___COUNT_DISTINCT	
0	754 (75%)
1	126 (13%)
2	74 (7.4%)
3	28 (2.8%)
4	8 (0.8%)
5	6 (0.6%)
6	2 (0.2%)
7	2 (0.2%)
SELECT_n_of_aggr_func___SUM	
0	960 (96%)
1	33 (3.3%)
2	6 (0.6%)
3	1 (0.1%)
SELECT_n_of_aggr_func___COUNT	
0	780 (78%)
1	113 (11%)
2	70 (7.0%)
3	23 (2.3%)
4	8 (0.8%)

Query parameter or value	Descriptive statistics
5	5 (0.5%)
6	1 (0.1%)
SELECT_n_of_aggr_func___AVG	
0	957 (96%)
1	34 (3.4%)
2	7 (0.7%)
3	2 (0.2%)
SELECT_n_of_all_aggr_func	[0.00, 2.00, 9.00] 1.80 / 1.47
FROM_n_of_join_paths	
1	1,000 (100%)
FROM_n_of_joins___INNER	
0	503 (50%)
1	341 (34%)
2	121 (12%)
3	34 (3.4%)
4	1 (0.1%)
FROM_n_of_joins___RIGHT	
0	502 (50%)
1	313 (31%)
2	138 (14%)
3	45 (4.5%)
4	2 (0.2%)
FROM_n_of_processed_rows	[5, 2,000,000,000, 15,300,027,829] 5,421,049,045 / 6,199,103,928
WHERE_n_of_predicates	
0	244 (24%)
1	218 (22%)
2	222 (22%)
3	166 (17%)
4	98 (9.8%)
5	43 (4.3%)
6	9 (0.9%)
WHERE_n_of_attribs_of_type___character_varying	
0	535 (54%)
1	329 (33%)
2	116 (12%)
3	20 (2.0%)
WHERE_n_of_attribs_of_type___integer	
0	583 (58%)
1	304 (30%)
2	98 (9.8%)
3	14 (1.4%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type___numeric	
0	734 (73%)
1	224 (22%)

Query parameter or value	Descriptive statistics
2	31 (3.1%)
3	10 (1.0%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type___date	
0	908 (91%)
1	80 (8.0%)
2	12 (1.2%)
WHERE_n_of_attribs_of_type___character	
0	798 (80%)
1	177 (18%)
2	22 (2.2%)
3	3 (0.3%)
WHERE_n_of_pkey_attribs	
0	747 (75%)
1	213 (21%)
2	37 (3.7%)
3	3 (0.3%)
WHERE_n_of_connect_OR	
0	608 (61%)
1	214 (21%)
2	121 (12%)
3	33 (3.3%)
4	21 (2.1%)
5	3 (0.3%)
WHERE_n_of_operators___greater_or_less	
0	521 (52%)
1	322 (32%)
2	118 (12%)
3	32 (3.2%)
4	7 (0.7%)
WHERE_n_of_operators___in	
0	714 (71%)
1	238 (24%)
2	47 (4.7%)
3	1 (0.1%)
WHERE_n_of_operators___like	
0	888 (89%)
1	107 (11%)
2	5 (0.5%)
WHERE_n_of_operators___between	
0	746 (75%)
1	211 (21%)
2	39 (3.9%)
3	2 (0.2%)
4	2 (0.2%)
WHERE_n_of_non_aggr_func___FLOOR	

Query parameter or value	Descriptive statistics
0	983 (98%)
1	16 (1.6%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___DAY	
0	980 (98%)
1	18 (1.8%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___LOG	
0	955 (96%)
1	44 (4.4%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___ABS	
0	976 (98%)
1	23 (2.3%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___SQRT	
0	942 (94%)
1	56 (5.6%)
2	2 (0.2%)
WHERE_n_of_non_aggr_func___YEAR	
0	991 (99%)
1	9 (0.9%)
WHERE_n_of_non_aggr_func___TRUNC	
0	981 (98%)
1	19 (1.9%)
WHERE_n_of_non_aggr_func___ROUND	
0	984 (98%)
1	15 (1.5%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func___MONTH	
0	986 (99%)
1	14 (1.4%)
WHERE_n_of_non_aggr_func___DOW	
0	988 (99%)
1	12 (1.2%)
WHERE_n_of_all_non_aggr_func	
0	794 (79%)
1	177 (18%)
2	24 (2.4%)
3	3 (0.3%)
4	2 (0.2%)
GROUP_BY_n_of_columns	
0	243 (24%)
1	285 (29%)
2	229 (23%)
3	135 (14%)

Query parameter or value	Descriptive statistics
4	70 (7.0%)
5	29 (2.9%)
6	9 (0.9%)
HAVING_n_of_main_predicates	[0.00, 4.00, 9.00] 3.51 / 2.61
HAVING_n_of_main_predicates___non_scalar_subquery	
0	435 (44%)
1	167 (17%)
2	209 (21%)
3	189 (19%)
HAVING_n_of_main_predicates___scalar_subquery	
0	390 (39%)
1	217 (22%)
2	190 (19%)
3	203 (20%)
HAVING_n_of_subqueries___non_scalar_subquery	
0	435 (44%)
1	167 (17%)
2	209 (21%)
3	189 (19%)
HAVING_n_of_subqueries___scalar_subquery	[0.0, 1.0, 20.0] 2.5 / 3.6
HAVING_n_of_processed_rows_by_subqueries	[0, 1,609,999,996, 345,000,639,538] 20,167,233,926 / 41,103,778,071
ORDER_BY_n_of_columns	
0	308 (31%)
1	442 (44%)
2	218 (22%)
3	27 (2.7%)
4	4 (0.4%)
5	1 (0.1%)
limit	[1, 536, 1,000] 519 / 287
offset	[0, 0, 997] 255 / 328