

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

Scenario: 1-1jp__g__h0ss_h0nss

Scale factor: 100 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, 42.0] 5.9 / 5.8
SELECT_n_of_non_aggr_func__RTRIM	
0	790 (79%)
1	169 (17%)
2	34 (3.4%)
3	7 (0.7%)
SELECT_n_of_non_aggr_func__UPPER	
0	778 (78%)
1	186 (19%)
2	30 (3.0%)
3	6 (0.6%)
SELECT_n_of_non_aggr_func__FLOOR	
0	963 (96%)
1	36 (3.6%)
2	1 (0.1%)

Query parameter or value	Descriptive statistics
SELECT__n_of__non__aggr__func__LOWER	
0	767 (77%)
1	196 (20%)
2	33 (3.3%)
3	3 (0.3%)
4	1 (0.1%)
SELECT__n_of__non__aggr__func__MONTH	
0	897 (90%)
1	88 (8.8%)
2	9 (0.9%)
3	6 (0.6%)
SELECT__n_of__non__aggr__func__TRUNC	
0	967 (97%)
1	32 (3.2%)
2	1 (0.1%)
SELECT__n_of__non__aggr__func__YEAR	
0	921 (92%)
1	68 (6.8%)
2	10 (1.0%)
3	1 (0.1%)
SELECT__n_of__non__aggr__func__LTRIM	
0	758 (76%)
1	212 (21%)
2	23 (2.3%)
3	7 (0.7%)
SELECT__n_of__non__aggr__func__ABS	
0	956 (96%)
1	43 (4.3%)
2	1 (0.1%)
SELECT__n_of__non__aggr__func__SUBSTR	
0	812 (81%)
1	158 (16%)
2	26 (2.6%)
3	3 (0.3%)
4	1 (0.1%)
SELECT__n_of__non__aggr__func__SQRT	
0	901 (90%)
1	89 (8.9%)
2	9 (0.9%)
3	1 (0.1%)
SELECT__n_of__non__aggr__func__DAY	
0	949 (95%)
1	46 (4.6%)
2	4 (0.4%)
3	1 (0.1%)
SELECT__n_of__non__aggr__func__DOW	

Query parameter or value	Descriptive statistics
0	928 (93%)
1	60 (6.0%)
2	9 (0.9%)
3	3 (0.3%)
SELECT_n_of_non_aggr_func__LOG	
0	941 (94%)
1	57 (5.7%)
2	2 (0.2%)
SELECT_n_of_non_aggr_func__ROUND	
0	969 (97%)
1	29 (2.9%)
2	2 (0.2%)
SELECT_n_of_all_non_aggr_func	[0.00, 1.00, 17.00] 1.98 / 2.24
SELECT_n_of_aggr_func__COUNT	
0	777 (78%)
1	99 (9.9%)
2	67 (6.7%)
3	33 (3.3%)
4	14 (1.4%)
5	7 (0.7%)
6	2 (0.2%)
7	1 (0.1%)
SELECT_n_of_aggr_func__MIN	
0	788 (79%)
1	107 (11%)
2	50 (5.0%)
3	27 (2.7%)
4	18 (1.8%)
5	8 (0.8%)
7	2 (0.2%)
SELECT_n_of_aggr_func__AVG	
0	959 (96%)
1	37 (3.7%)
2	3 (0.3%)
3	1 (0.1%)
SELECT_n_of_aggr_func__MAX	
0	767 (77%)
1	106 (11%)
2	71 (7.1%)
3	28 (2.8%)
4	17 (1.7%)
5	5 (0.5%)
6	5 (0.5%)
7	1 (0.1%)
SELECT_n_of_aggr_func__COUNT_DISTINCT	
0	778 (78%)

Query parameter or value	Descriptive statistics
1	106 (11%)
2	59 (5.9%)
3	35 (3.5%)
4	14 (1.4%)
5	8 (0.8%)
SELECT_n_of_aggr_func__SUM	
0	943 (94%)
1	45 (4.5%)
2	10 (1.0%)
3	1 (0.1%)
4	1 (0.1%)
SELECT_n_of_all_aggr_func	
0	187 (19%)
1	293 (29%)
2	217 (22%)
3	167 (17%)
4	76 (7.6%)
5	38 (3.8%)
6	15 (1.5%)
7	7 (0.7%)
FROM_n_of_join_paths	
1	1,000 (100%)
FROM_n_of_joins__RIGHT	
0	487 (49%)
1	337 (34%)
2	141 (14%)
3	30 (3.0%)
4	5 (0.5%)
FROM_n_of_joins__INNER	
0	502 (50%)
1	312 (31%)
2	140 (14%)
3	40 (4.0%)
4	6 (0.6%)
FROM_n_of_processed_rows	[5, 100,000,000, 765,037,932] 286,025,165 / 314,561,240
WHERE_n_of_predicates	
0	227 (23%)
1	238 (24%)
2	206 (21%)
3	178 (18%)
4	93 (9.3%)
5	45 (4.5%)
6	13 (1.3%)
WHERE_n_of_attribs_of_type__integer	
0	568 (57%)

Query parameter or value	Descriptive statistics
1	314 (31%)
2	89 (8.9%)
3	25 (2.5%)
4	4 (0.4%)
WHERE_n_of_attribs_of_type__character	
0	824 (82%)
1	144 (14%)
2	30 (3.0%)
3	2 (0.2%)
WHERE_n_of_attribs_of_type__numeric	
0	738 (74%)
1	226 (23%)
2	29 (2.9%)
3	7 (0.7%)
WHERE_n_of_attribs_of_type__character_varying	
0	527 (53%)
1	322 (32%)
2	131 (13%)
3	20 (2.0%)
WHERE_n_of_attribs_of_type__date	
0	897 (90%)
1	89 (8.9%)
2	14 (1.4%)
WHERE_n_of_pkey_attribs	
0	712 (71%)
1	249 (25%)
2	36 (3.6%)
3	3 (0.3%)
WHERE_n_of_connect_OR	
0	584 (58%)
1	217 (22%)
2	125 (13%)
3	55 (5.5%)
4	16 (1.6%)
5	3 (0.3%)
WHERE_n_of_operators__between	
0	706 (71%)
1	248 (25%)
2	43 (4.3%)
3	3 (0.3%)
WHERE_n_of_operators__in	
0	714 (71%)
1	228 (23%)
2	51 (5.1%)
3	7 (0.7%)
WHERE_n_of_operators__greater_or_less	

Query parameter or value	Descriptive statistics
0	516 (52%)
1	319 (32%)
2	128 (13%)
3	31 (3.1%)
4	6 (0.6%)
WHERE_n_of_operators__like	
0	884 (88%)
1	113 (11%)
2	3 (0.3%)
WHERE_n_of_non_aggr_func__LOG	
0	963 (96%)
1	37 (3.7%)
WHERE_n_of_non_aggr_func__TRUNC	
0	978 (98%)
1	22 (2.2%)
WHERE_n_of_non_aggr_func__MONTH	
0	982 (98%)
1	18 (1.8%)
WHERE_n_of_non_aggr_func__SQRT	
0	924 (92%)
1	75 (7.5%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__DOW	
0	986 (99%)
1	14 (1.4%)
WHERE_n_of_non_aggr_func__FLOOR	
0	978 (98%)
1	22 (2.2%)
WHERE_n_of_non_aggr_func__DAY	
0	980 (98%)
1	20 (2.0%)
WHERE_n_of_non_aggr_func__ABS	
0	978 (98%)
1	22 (2.2%)
WHERE_n_of_non_aggr_func__YEAR	
0	988 (99%)
1	12 (1.2%)
WHERE_n_of_non_aggr_func__ROUND	
0	978 (98%)
1	22 (2.2%)
WHERE_n_of_all_non_aggr_func	
0	771 (77%)
1	197 (20%)
2	28 (2.8%)
3	3 (0.3%)
4	1 (0.1%)

Query parameter or value	Descriptive statistics
GROUP_BY_n_of_columns	
0	243 (24%)
1	246 (25%)
2	237 (24%)
3	165 (17%)
4	66 (6.6%)
5	36 (3.6%)
6	7 (0.7%)
HAVING_n_of_main_predicates	
0	437 (44%)
1	181 (18%)
2	200 (20%)
3	182 (18%)
ORDER_BY_n_of_columns	
0	295 (30%)
1	444 (44%)
2	228 (23%)
3	24 (2.4%)
4	7 (0.7%)
5	1 (0.1%)
6	1 (0.1%)
limit	[2, 513, 1,000] 504 / 293
offset	[0, 0, 1,000] 220 / 312