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

Scenario: 1-3jp_ujjp__g__h0-3ss_h0-3nss

Scale factor: 500 GB

The 1000-query pack was generated on: 2023-11-10

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, 6, 103] 10 / 13
SELECT_n_of_non_aggr_func___LOWER	
0	698 (70%)
1	229 (23%)
2	53 (5.3%)
3	15 (1.5%)
4	1 (0.1%)
5	3 (0.3%)
6	1 (0.1%)
SELECT_n_of_non_aggr_func___LTRIM	
0	710 (71%)
1	220 (22%)
2	49 (4.9%)
3	14 (1.4%)
4	6 (0.6%)

Query parameter or value	Descriptive statistics
5	1 (0.1%)
SELECT_n_of_non_aggr_func___RTRIM	
0	685 (69%)
1	245 (25%)
2	51 (5.1%)
3	13 (1.3%)
4	5 (0.5%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___DOW	
0	889 (89%)
1	92 (9.2%)
2	14 (1.4%)
3	5 (0.5%)
SELECT_n_of_non_aggr_func___MONTH	
0	845 (85%)
1	119 (12%)
2	27 (2.7%)
3	4 (0.4%)
4	1 (0.1%)
5	3 (0.3%)
6	1 (0.1%)
SELECT_n_of_non_aggr_func___ROUND	
0	942 (94%)
1	56 (5.6%)
2	1 (0.1%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___SUBSTR	
0	687 (69%)
1	234 (23%)
2	50 (5.0%)
3	17 (1.7%)
4	8 (0.8%)
5	4 (0.4%)
SELECT_n_of_non_aggr_func___TRUNC	
0	943 (94%)
1	51 (5.1%)
2	5 (0.5%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___YEAR	
0	879 (88%)
1	98 (9.8%)
2	13 (1.3%)
3	6 (0.6%)
4	3 (0.3%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___UPPER	

Query parameter or value	Descriptive statistics
0	685 (69%)
1	229 (23%)
2	58 (5.8%)
3	18 (1.8%)
4	7 (0.7%)
5	3 (0.3%)
SELECT_n_of_non_aggr_func___ABS	
0	947 (95%)
1	50 (5.0%)
2	2 (0.2%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___SQRT	
0	847 (85%)
1	121 (12%)
2	19 (1.9%)
3	10 (1.0%)
4	3 (0.3%)
SELECT_n_of_non_aggr_func___DAY	
0	915 (92%)
1	71 (7.1%)
2	12 (1.2%)
3	1 (0.1%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___FLOOR	
0	954 (95%)
1	44 (4.4%)
2	2 (0.2%)
SELECT_n_of_non_aggr_func___LOG	
0	920 (92%)
1	68 (6.8%)
2	10 (1.0%)
3	2 (0.2%)
SELECT_n_of_all_non_aggr_func	[0.0, 2.0, 42.0] 3.2 / 4.3
SELECT_n_of_aggr_func___COUNT	[0.00, 0.00, 9.00] 0.68 / 1.49
SELECT_n_of_aggr_func___AVG	
0	935 (94%)
1	45 (4.5%)
2	16 (1.6%)
3	3 (0.3%)
4	1 (0.1%)
SELECT_n_of_aggr_func___MAX	[0.00, 0.00, 9.00] 0.68 / 1.46
SELECT_n_of_aggr_func___SUM	
0	950 (95%)
1	38 (3.8%)
2	10 (1.0%)
3	2 (0.2%)

Query parameter or value	Descriptive statistics
SELECT_n_of_aggr_func___COUNT_DISTINCT	
0	774 (77%)
1	79 (7.9%)
2	50 (5.0%)
3	46 (4.6%)
4	17 (1.7%)
5	14 (1.4%)
6	11 (1.1%)
7	5 (0.5%)
8	4 (0.4%)
SELECT_n_of_aggr_func___MIN	
0	765 (77%)
1	93 (9.3%)
2	43 (4.3%)
3	33 (3.3%)
4	24 (2.4%)
5	19 (1.9%)
6	13 (1.3%)
7	8 (0.8%)
8	2 (0.2%)
SELECT_n_of_all_aggr_func	[0.00, 2.00, 10.00] 2.72 / 2.27
FROM_n_of_join_paths	
1	331 (33%)
2	323 (32%)
3	346 (35%)
FROM_n_of_super_joins___LEFT	
0	758 (76%)
1	222 (22%)
2	20 (2.0%)
FROM_n_of_super_joins___RIGHT	
0	752 (75%)
1	229 (23%)
2	19 (1.9%)
FROM_n_of_super_joins___FULL	
0	790 (79%)
1	189 (19%)
2	21 (2.1%)
FROM_n_of_joins___INNER	
0	222 (22%)
1	241 (24%)
2	230 (23%)
3	138 (14%)
4	87 (8.7%)
5	63 (6.3%)
6	15 (1.5%)
7	2 (0.2%)

Query parameter or value	Descriptive statistics
8	2 (0.2%)
FROM_n_of_joins___RIGHT	
0	234 (23%)
1	266 (27%)
2	225 (23%)
3	141 (14%)
4	72 (7.2%)
5	43 (4.3%)
6	15 (1.5%)
7	3 (0.3%)
8	1 (0.1%)
FROM_n_of_processed_rows	[5, 3,405,028,267, 11,475,084,811] 3,163,031,808 / 2,872,469,769
WHERE_n_of_predicates	[0.00, 2.00, 10.00] 2.83 / 2.31
WHERE_n_of_attribs_of_type___character_varying	
0	445 (45%)
1	300 (30%)
2	165 (17%)
3	66 (6.6%)
4	23 (2.3%)
5	1 (0.1%)
WHERE_n_of_attribs_of_type___integer	
0	461 (46%)
1	290 (29%)
2	160 (16%)
3	67 (6.7%)
4	19 (1.9%)
5	3 (0.3%)
WHERE_n_of_attribs_of_type___numeric	
0	654 (65%)
1	254 (25%)
2	75 (7.5%)
3	13 (1.3%)
4	4 (0.4%)
WHERE_n_of_attribs_of_type___date	
0	822 (82%)
1	155 (16%)
2	20 (2.0%)
3	2 (0.2%)
4	1 (0.1%)
WHERE_n_of_attribs_of_type___character	
0	736 (74%)
1	206 (21%)
2	45 (4.5%)
3	11 (1.1%)
4	1 (0.1%)

Query parameter or value	Descriptive statistics
5	1 (0.1%)
WHERE_n_of_pkey_attribs	
0	621 (62%)
1	274 (27%)
2	86 (8.6%)
3	16 (1.6%)
4	3 (0.3%)
WHERE_n_of_connect_OR	
0	451 (45%)
1	174 (17%)
2	143 (14%)
3	96 (9.6%)
4	62 (6.2%)
5	38 (3.8%)
6	25 (2.5%)
7	7 (0.7%)
8	4 (0.4%)
WHERE_n_of_operators___between	
0	639 (64%)
1	246 (25%)
2	80 (8.0%)
3	31 (3.1%)
4	3 (0.3%)
5	1 (0.1%)
WHERE_n_of_operators___greater_or_less	
0	400 (40%)
1	289 (29%)
2	180 (18%)
3	86 (8.6%)
4	26 (2.6%)
5	13 (1.3%)
6	6 (0.6%)
WHERE_n_of_operators___like	
0	830 (83%)
1	148 (15%)
2	20 (2.0%)
3	2 (0.2%)
WHERE_n_of_operators___in	
0	607 (61%)
1	285 (29%)
2	89 (8.9%)
3	14 (1.4%)
4	4 (0.4%)
5	1 (0.1%)
WHERE_n_of_non_aggr_func___SQRT	
0	929 (93%)

Query parameter or value	Descriptive statistics
1	68 (6.8%)
2	3 (0.3%)
WHERE_n_of_non_aggr_func___FLOOR	
0	972 (97%)
1	28 (2.8%)
WHERE_n_of_non_aggr_func___DAY	
0	973 (97%)
1	27 (2.7%)
WHERE_n_of_non_aggr_func___LOG	
0	961 (96%)
1	39 (3.9%)
WHERE_n_of_non_aggr_func___YEAR	
0	973 (97%)
1	27 (2.7%)
WHERE_n_of_non_aggr_func___ABS	
0	974 (97%)
1	26 (2.6%)
WHERE_n_of_non_aggr_func___TRUNC	
0	972 (97%)
1	28 (2.8%)
WHERE_n_of_non_aggr_func___DOW	
0	974 (97%)
1	26 (2.6%)
WHERE_n_of_non_aggr_func___ROUND	
0	972 (97%)
1	28 (2.8%)
WHERE_n_of_non_aggr_func___MONTH	
0	976 (98%)
1	22 (2.2%)
2	2 (0.2%)
WHERE_n_of_all_non_aggr_func	
0	738 (74%)
1	211 (21%)
2	39 (3.9%)
3	8 (0.8%)
4	4 (0.4%)
GROUP_BY_n_of_columns	[0.00, 2.00, 9.00] 2.46 / 2.17
HAVING_n_of_main_predicates	[0.00, 4.00, 9.00] 3.39 / 2.55
HAVING_n_of_main_predicates___non_scalar_subquery	
0	431 (43%)
1	172 (17%)
2	197 (20%)
3	200 (20%)
HAVING_n_of_main_predicates___scalar_subquery	
0	440 (44%)
1	192 (19%)

Query parameter or value	Descriptive statistics
2	182 (18%)
3	186 (19%)
HAVING_n_of_subqueries___non_scalar_subquery	
0	431 (43%)
1	172 (17%)
2	197 (20%)
3	200 (20%)
HAVING_n_of_subqueries___scalar_subquery	[0.0, 1.0, 21.0] 2.3 / 3.3
HAVING_n_of_processed_rows_by_subqueries	[0, 2,475,000,090, 150,256,186,794] 10,919,200,491 / 18,822,232,428
ORDER_BY_n_of_columns	
0	289 (29%)
1	372 (37%)
2	258 (26%)
3	59 (5.9%)
4	14 (1.4%)
5	7 (0.7%)
6	1 (0.1%)
limit	[1, 505, 1,000] 506 / 284
offset	[0, 0, 1,000] 246 / 321