

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

## Scenario: 1-3jp\_rjpp\_\_g\_\_h0-3ss\_h0-3nss

## Scale factor: 200 GB

## The 1000-query pack was generated on: 2023-11-06

## 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, 101] 10 / 13
SELECT_n_of_non_aggr_func__MONTH	
0	837 (84%)
1	126 (13%)
2	25 (2.5%)
3	6 (0.6%)
4	2 (0.2%)
5	2 (0.2%)
6	2 (0.2%)
SELECT_n_of_non_aggr_func__YEAR	
0	879 (88%)
1	88 (8.8%)
2	20 (2.0%)
3	9 (0.9%)
4	4 (0.4%)

Query parameter or value	Descriptive statistics
SELECT__n_of__non_aggr_func__LOG	
0	901 (90%)
1	83 (8.3%)
2	11 (1.1%)
3	3 (0.3%)
4	1 (0.1%)
5	1 (0.1%)
SELECT__n_of__non_aggr_func__LOWER	
0	708 (71%)
1	234 (23%)
2	42 (4.2%)
3	11 (1.1%)
4	2 (0.2%)
5	2 (0.2%)
8	1 (0.1%)
SELECT__n_of__non_aggr_func__LTRIM	
0	706 (71%)
1	227 (23%)
2	47 (4.7%)
3	14 (1.4%)
4	2 (0.2%)
6	1 (0.1%)
7	2 (0.2%)
8	1 (0.1%)
SELECT__n_of__non_aggr_func__RTRIM	
0	711 (71%)
1	223 (22%)
2	53 (5.3%)
3	7 (0.7%)
4	4 (0.4%)
5	2 (0.2%)
SELECT__n_of__non_aggr_func__SUBSTR	
0	719 (72%)
1	217 (22%)
2	45 (4.5%)
3	15 (1.5%)
4	3 (0.3%)
5	1 (0.1%)
SELECT__n_of__non_aggr_func__DOW	
0	884 (88%)
1	93 (9.3%)
2	17 (1.7%)
3	5 (0.5%)
4	1 (0.1%)
SELECT__n_of__non_aggr_func__SQRT	
0	855 (86%)

Query parameter or value	Descriptive statistics
1	125 (13%)
2	12 (1.2%)
3	6 (0.6%)
4	1 (0.1%)
5	1 (0.1%)
SELECT_n_of_non_aggr_func___UPPER	
0	699 (70%)
1	230 (23%)
2	48 (4.8%)
3	15 (1.5%)
4	4 (0.4%)
5	4 (0.4%)
SELECT_n_of_non_aggr_func___DAY	
0	906 (91%)
1	83 (8.3%)
2	8 (0.8%)
3	2 (0.2%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___ROUND	
0	949 (95%)
1	46 (4.6%)
2	4 (0.4%)
4	1 (0.1%)
SELECT_n_of_non_aggr_func___TRUNC	
0	945 (95%)
1	50 (5.0%)
2	5 (0.5%)
SELECT_n_of_non_aggr_func___FLOOR	
0	954 (95%)
1	44 (4.4%)
2	1 (0.1%)
3	1 (0.1%)
SELECT_n_of_non_aggr_func___ABS	
0	942 (94%)
1	52 (5.2%)
2	4 (0.4%)
3	2 (0.2%)
SELECT_n_of_all_non_aggr_func	[0.00, 2.00, 40.00] 3.11 / 4.40
SELECT_n_of_aggr_func___COUNT_DISTINCT	[0.00, 0.00, 9.00] 0.59 / 1.41
SELECT_n_of_aggr_func___COUNT	[0.00, 0.00, 10.00] 0.64 / 1.50
SELECT_n_of_aggr_func___MIN	[0.00, 0.00, 14.00] 0.67 / 1.60
SELECT_n_of_aggr_func___MAX	[0.00, 0.00, 9.00] 0.56 / 1.34
SELECT_n_of_aggr_func___AVG	
0	948 (95%)
1	36 (3.6%)
2	12 (1.2%)

Query parameter or value	Descriptive statistics
3	2 (0.2%)
4	2 (0.2%)
SELECT_n_of_aggr_func__SUM	
0	938 (94%)
1	42 (4.2%)
2	16 (1.6%)
3	3 (0.3%)
4	1 (0.1%)
SELECT_n_of_all_aggr_func	[0.00, 2.00, 14.00] 2.63 / 2.37
FROM_n_of_join_paths	
1	349 (35%)
2	307 (31%)
3	344 (34%)
FROM_n_of_super_joins__FULL	
0	771 (77%)
1	203 (20%)
2	26 (2.6%)
FROM_n_of_super_joins__RIGHT	
0	763 (76%)
1	222 (22%)
2	15 (1.5%)
FROM_n_of_super_joins__LEFT	
0	763 (76%)
1	224 (22%)
2	13 (1.3%)
FROM_n_of_joins__RIGHT	[0.00, 1.00, 9.00] 1.79 / 1.61
FROM_n_of_joins__INNER	
0	235 (24%)
1	257 (26%)
2	240 (24%)
3	134 (13%)
4	75 (7.5%)
5	38 (3.8%)
6	14 (1.4%)
7	5 (0.5%)
8	2 (0.2%)
FROM_n_of_processed_rows	[5, 1,362,018,459, 4,590,055,387] 1,300,930,926 / 1,235,116,208
WHERE_n_of_predicates	[0.00, 2.00, 9.00] 2.75 / 2.21
WHERE_n_of_attris_of_type__numeric	
0	674 (67%)
1	238 (24%)
2	71 (7.1%)
3	14 (1.4%)
4	3 (0.3%)
WHERE_n_of_attris_of_type__integer	

Query parameter or value	Descriptive statistics
<i>0</i>	444 (44%)
<i>1</i>	336 (34%)
<i>2</i>	147 (15%)
<i>3</i>	55 (5.5%)
<i>4</i>	16 (1.6%)
<i>5</i>	1 (0.1%)
<i>6</i>	1 (0.1%)
WHERE_n_of_attribs_of_type___character_varying	
<i>0</i>	465 (47%)
<i>1</i>	276 (28%)
<i>2</i>	173 (17%)
<i>3</i>	57 (5.7%)
<i>4</i>	26 (2.6%)
<i>5</i>	2 (0.2%)
<i>6</i>	1 (0.1%)
WHERE_n_of_attribs_of_type___character	
<i>0</i>	726 (73%)
<i>1</i>	215 (22%)
<i>2</i>	48 (4.8%)
<i>3</i>	7 (0.7%)
<i>4</i>	4 (0.4%)
WHERE_n_of_attribs_of_type___date	
<i>0</i>	840 (84%)
<i>1</i>	139 (14%)
<i>2</i>	19 (1.9%)
<i>3</i>	2 (0.2%)
WHERE_n_of_pkey_attribs	
<i>0</i>	593 (59%)
<i>1</i>	302 (30%)
<i>2</i>	88 (8.8%)
<i>3</i>	15 (1.5%)
<i>4</i>	2 (0.2%)
WHERE_n_of_connect_OR	
<i>0</i>	441 (44%)
<i>1</i>	199 (20%)
<i>2</i>	142 (14%)
<i>3</i>	96 (9.6%)
<i>4</i>	62 (6.2%)
<i>5</i>	31 (3.1%)
<i>6</i>	20 (2.0%)
<i>7</i>	6 (0.6%)
<i>8</i>	3 (0.3%)
WHERE_n_of_operators___greater_or_less	
<i>0</i>	425 (43%)
<i>1</i>	294 (29%)
<i>2</i>	144 (14%)

Query parameter or value	Descriptive statistics
3	95 (9.5%)
4	33 (3.3%)
5	9 (0.9%)
WHERE_n_of_operators__between	
0	626 (63%)
1	267 (27%)
2	86 (8.6%)
3	19 (1.9%)
4	2 (0.2%)
WHERE_n_of_operators__in	
0	619 (62%)
1	272 (27%)
2	82 (8.2%)
3	24 (2.4%)
4	2 (0.2%)
5	1 (0.1%)
WHERE_n_of_operators__like	
0	840 (84%)
1	141 (14%)
2	17 (1.7%)
3	2 (0.2%)
WHERE_n_of_non_aggr_func__ROUND	
0	974 (97%)
1	26 (2.6%)
WHERE_n_of_non_aggr_func__LOG	
0	946 (95%)
1	51 (5.1%)
2	3 (0.3%)
WHERE_n_of_non_aggr_func__ABS	
0	967 (97%)
1	33 (3.3%)
WHERE_n_of_non_aggr_func__DOW	
0	967 (97%)
1	33 (3.3%)
WHERE_n_of_non_aggr_func__SQRT	
0	905 (91%)
1	86 (8.6%)
2	9 (0.9%)
WHERE_n_of_non_aggr_func__TRUNC	
0	972 (97%)
1	28 (2.8%)
WHERE_n_of_non_aggr_func__YEAR	
0	971 (97%)
1	28 (2.8%)
2	1 (0.1%)
WHERE_n_of_non_aggr_func__MONTH	

Query parameter or value	Descriptive statistics
0	976 (98%)
1	24 (2.4%)
WHERE_n_of_non_aggr_func__DAY	
0	979 (98%)
1	21 (2.1%)
WHERE_n_of_non_aggr_func__FLOOR	
0	970 (97%)
1	29 (2.9%)
2	1 (0.1%)
WHERE_n_of_all_non_aggr_func	
0	701 (70%)
1	232 (23%)
2	49 (4.9%)
3	16 (1.6%)
4	1 (0.1%)
5	1 (0.1%)
GROUP_BY_n_of_columns	[0.00, 2.00, 9.00] 2.35 / 2.20
HAVING_n_of_main_predicates	[0.00, 3.00, 9.00] 3.19 / 2.56
HAVING_n_of_main_predicates__scalar_subquery	
0	478 (48%)
1	187 (19%)
2	176 (18%)
3	159 (16%)
HAVING_n_of_main_predicates__non_scalar_subquery	
0	449 (45%)
1	187 (19%)
2	173 (17%)
3	191 (19%)
HAVING_n_of_subqueries__scalar_subquery	[0.0, 1.0, 21.0] 2.2 / 3.4
HAVING_n_of_subqueries__non_scalar_subquery	
0	449 (45%)
1	187 (19%)
2	173 (17%)
3	191 (19%)
HAVING_n_of_processed_rows_by_subqueries	[0, 600,000,000, 70,074,940,134] 4,208,495,922 / 7,937,089,033
ORDER_BY_n_of_columns	
0	299 (30%)
1	401 (40%)
2	205 (21%)
3	65 (6.5%)
4	16 (1.6%)
5	9 (0.9%)
6	4 (0.4%)
7	1 (0.1%)
limit	[4, 508, 999] 501 / 284

Query parameter or value	Descriptive statistics
offset	[0, 5, 1,000] 255 / 329