

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

Scenario: 1-6jp_rjpp__g__h0ss_h0nss

Scale factor: 1000 GB

The 1000-query pack was generated on: 2023-11-12

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, 8, 172] 16 / 25 |
| SELECT_n_of_non_aggr_func__LOWER | |
| 0 | 612 (61%) |
| 1 | 254 (25%) |
| 2 | 80 (8.0%) |
| 3 | 29 (2.9%) |
| 4 | 14 (1.4%) |
| 5 | 8 (0.8%) |
| 6 | 2 (0.2%) |
| 7 | 1 (0.1%) |
| SELECT_n_of_non_aggr_func__MONTH | |
| 0 | 801 (80%) |
| 1 | 135 (14%) |
| 2 | 35 (3.5%) |
| 3 | 14 (1.4%) |

| Query parameter or value | Descriptive statistics |
|-----------------------------------|---------------------------------|
| 4 | 5 (0.5%) |
| 5 | 3 (0.3%) |
| 6 | 4 (0.4%) |
| 7 | 1 (0.1%) |
| 8 | 2 (0.2%) |
| SELECT_n_of_non_aggr_func__UPPER | [0.00, 0.00, 11.00] 0.60 / 1.11 |
| SELECT_n_of_non_aggr_func__SUBSTR | [0.00, 0.00, 9.00] 0.58 / 1.01 |
| SELECT_n_of_non_aggr_func__DAY | |
| 0 | 891 (89%) |
| 1 | 79 (7.9%) |
| 2 | 18 (1.8%) |
| 3 | 7 (0.7%) |
| 4 | 4 (0.4%) |
| 7 | 1 (0.1%) |
| SELECT_n_of_non_aggr_func__LTRIM | |
| 0 | 632 (63%) |
| 1 | 238 (24%) |
| 10 | 1 (0.1%) |
| 2 | 80 (8.0%) |
| 3 | 24 (2.4%) |
| 4 | 12 (1.2%) |
| 5 | 8 (0.8%) |
| 6 | 3 (0.3%) |
| 8 | 2 (0.2%) |
| SELECT_n_of_non_aggr_func__SQRT | |
| 0 | 806 (81%) |
| 1 | 140 (14%) |
| 2 | 31 (3.1%) |
| 3 | 12 (1.2%) |
| 4 | 5 (0.5%) |
| 5 | 2 (0.2%) |
| 6 | 2 (0.2%) |
| 7 | 2 (0.2%) |
| SELECT_n_of_non_aggr_func__TRUNC | |
| 0 | 915 (92%) |
| 1 | 69 (6.9%) |
| 2 | 10 (1.0%) |
| 3 | 6 (0.6%) |
| SELECT_n_of_non_aggr_func__LOG | |
| 0 | 866 (87%) |
| 1 | 95 (9.5%) |
| 2 | 24 (2.4%) |
| 3 | 6 (0.6%) |
| 4 | 5 (0.5%) |
| 5 | 4 (0.4%) |
| SELECT_n_of_non_aggr_func__ABS | |

| Query parameter or value | Descriptive statistics |
|----------------------------------|---------------------------------|
| 0 | 924 (92%) |
| 1 | 65 (6.5%) |
| 2 | 6 (0.6%) |
| 3 | 5 (0.5%) |
| SELECT_n_of_non_aggr_func__RTRIM | [0.00, 0.00, 9.00] 0.61 / 1.10 |
| SELECT_n_of_non_aggr_func__DOW | |
| 0 | 847 (85%) |
| 1 | 114 (11%) |
| 2 | 21 (2.1%) |
| 3 | 9 (0.9%) |
| 4 | 4 (0.4%) |
| 5 | 4 (0.4%) |
| 7 | 1 (0.1%) |
| SELECT_n_of_non_aggr_func__YEAR | |
| 0 | 840 (84%) |
| 1 | 104 (10%) |
| 2 | 32 (3.2%) |
| 3 | 11 (1.1%) |
| 4 | 5 (0.5%) |
| 5 | 3 (0.3%) |
| 6 | 3 (0.3%) |
| 8 | 1 (0.1%) |
| 9 | 1 (0.1%) |
| SELECT_n_of_non_aggr_func__FLOOR | |
| 0 | 918 (92%) |
| 1 | 74 (7.4%) |
| 2 | 6 (0.6%) |
| 3 | 2 (0.2%) |
| SELECT_n_of_non_aggr_func__ROUND | |
| 0 | 914 (91%) |
| 1 | 75 (7.5%) |
| 2 | 7 (0.7%) |
| 3 | 3 (0.3%) |
| 4 | 1 (0.1%) |
| SELECT_n_of_all_non_aggr_func | [0.0, 3.0, 60.0] 4.9 / 7.8 |
| SELECT_n_of_aggr_func__MIN | [0.00, 0.00, 11.00] 0.94 / 2.04 |
| SELECT_n_of_aggr_func__AVG | |
| 0 | 927 (93%) |
| 1 | 46 (4.6%) |
| 2 | 17 (1.7%) |
| 3 | 5 (0.5%) |
| 4 | 4 (0.4%) |
| 5 | 1 (0.1%) |
| SELECT_n_of_aggr_func__MAX | [0.00, 0.00, 12.00] 0.87 / 1.86 |
| SELECT_n_of_aggr_func__COUNT | [0.00, 0.00, 13.00] 0.76 / 1.87 |
| SELECT_n_of_aggr_func__SUM | |

| Query parameter or value | Descriptive statistics |
|---------------------------------------|---|
| 0 | 934 (93%) |
| 1 | 47 (4.7%) |
| 2 | 12 (1.2%) |
| 3 | 4 (0.4%) |
| 4 | 2 (0.2%) |
| 5 | 1 (0.1%) |
| SELECT_n_of_aggr_func__COUNT_DISTINCT | [0.00, 0.00, 14.00] 0.87 / 2.04 |
| SELECT_n_of_all_aggr_func | [0.00, 3.00, 15.00] 3.65 / 3.14 |
| FROM_n_of_join_paths | |
| 1 | 160 (16%) |
| 2 | 167 (17%) |
| 3 | 151 (15%) |
| 4 | 174 (17%) |
| 5 | 188 (19%) |
| 6 | 160 (16%) |
| FROM_n_of_super_joins__FULL | |
| 0 | 525 (53%) |
| 1 | 327 (33%) |
| 2 | 116 (12%) |
| 3 | 29 (2.9%) |
| 4 | 3 (0.3%) |
| FROM_n_of_super_joins__RIGHT | |
| 0 | 553 (55%) |
| 1 | 314 (31%) |
| 2 | 108 (11%) |
| 3 | 23 (2.3%) |
| 4 | 1 (0.1%) |
| 5 | 1 (0.1%) |
| FROM_n_of_super_joins__LEFT | |
| 0 | 544 (54%) |
| 1 | 308 (31%) |
| 2 | 117 (12%) |
| 3 | 27 (2.7%) |
| 4 | 4 (0.4%) |
| FROM_n_of_joins__INNER | [0.00, 3.00, 13.00] 3.30 / 2.48 |
| FROM_n_of_joins__RIGHT | [0.00, 3.00, 12.00] 3.35 / 2.50 |
| FROM_n_of_processed_rows | [5, 9,999,989,777, 44,909,938,334] 12,100,984,622 / 9,505,341,456 |
| WHERE_n_of_predicates | [0.00, 3.00, 13.00] 3.94 / 2.98 |
| WHERE_n_of_attribs_of_type__character | |
| 0 | 639 (64%) |
| 1 | 273 (27%) |
| 2 | 74 (7.4%) |
| 3 | 12 (1.2%) |
| 4 | 2 (0.2%) |
| WHERE_n_of_attribs_of_type__integer | |

| Query parameter or value | Descriptive statistics |
|---|---------------------------------|
| 0 | 354 (35%) |
| 1 | 305 (31%) |
| 2 | 179 (18%) |
| 3 | 101 (10%) |
| 4 | 37 (3.7%) |
| 5 | 19 (1.9%) |
| 6 | 4 (0.4%) |
| 7 | 1 (0.1%) |
| WHERE_n_of_attribs_of_type__numeric | |
| 0 | 558 (56%) |
| 1 | 283 (28%) |
| 2 | 117 (12%) |
| 3 | 30 (3.0%) |
| 4 | 11 (1.1%) |
| 5 | 1 (0.1%) |
| WHERE_n_of_attribs_of_type__character_varying | |
| 0 | 341 (34%) |
| 1 | 300 (30%) |
| 2 | 183 (18%) |
| 3 | 100 (10%) |
| 4 | 49 (4.9%) |
| 5 | 12 (1.2%) |
| 6 | 13 (1.3%) |
| 7 | 2 (0.2%) |
| WHERE_n_of_attribs_of_type__date | |
| 0 | 786 (79%) |
| 1 | 168 (17%) |
| 2 | 39 (3.9%) |
| 3 | 7 (0.7%) |
| WHERE_n_of_pkey_attribs | |
| 0 | 526 (53%) |
| 1 | 307 (31%) |
| 2 | 119 (12%) |
| 3 | 41 (4.1%) |
| 4 | 6 (0.6%) |
| 5 | 1 (0.1%) |
| WHERE_n_of_connect_OR | [0.00, 2.00, 12.00] 2.30 / 2.49 |
| WHERE_n_of_operators__greater_or_less | |
| 0 | 334 (33%) |
| 1 | 270 (27%) |
| 2 | 176 (18%) |
| 3 | 116 (12%) |
| 4 | 61 (6.1%) |
| 5 | 24 (2.4%) |
| 6 | 14 (1.4%) |
| 7 | 4 (0.4%) |

| Query parameter or value | Descriptive statistics |
|--------------------------------|------------------------|
| 8 | 1 (0.1%) |
| WHERE_n_of_operators_in | |
| 0 | 506 (51%) |
| 1 | 311 (31%) |
| 2 | 136 (14%) |
| 3 | 39 (3.9%) |
| 4 | 7 (0.7%) |
| 5 | 1 (0.1%) |
| WHERE_n_of_operators_between | |
| 0 | 522 (52%) |
| 1 | 295 (30%) |
| 2 | 126 (13%) |
| 3 | 38 (3.8%) |
| 4 | 12 (1.2%) |
| 5 | 6 (0.6%) |
| 6 | 1 (0.1%) |
| WHERE_n_of_operators_like | |
| 0 | 788 (79%) |
| 1 | 178 (18%) |
| 2 | 29 (2.9%) |
| 3 | 5 (0.5%) |
| WHERE_n_of_non_aggr_func_LOG | |
| 0 | 906 (91%) |
| 1 | 89 (8.9%) |
| 2 | 4 (0.4%) |
| 3 | 1 (0.1%) |
| WHERE_n_of_non_aggr_func_SQRT | |
| 0 | 876 (88%) |
| 1 | 118 (12%) |
| 2 | 6 (0.6%) |
| WHERE_n_of_non_aggr_func_ABS | |
| 0 | 959 (96%) |
| 1 | 38 (3.8%) |
| 2 | 3 (0.3%) |
| WHERE_n_of_non_aggr_func_FLOOR | |
| 0 | 954 (95%) |
| 1 | 44 (4.4%) |
| 2 | 2 (0.2%) |
| WHERE_n_of_non_aggr_func_DAY | |
| 0 | 966 (97%) |
| 1 | 34 (3.4%) |
| WHERE_n_of_non_aggr_func_YEAR | |
| 0 | 965 (97%) |
| 1 | 34 (3.4%) |
| 2 | 1 (0.1%) |
| WHERE_n_of_non_aggr_func_MONTH | |

| Query parameter or value | Descriptive statistics |
|---------------------------------|---------------------------------|
| 0 | 966 (97%) |
| 1 | 34 (3.4%) |
| WHERE_n_of_non_aggr_func__TRUNC | |
| 0 | 966 (97%) |
| 1 | 34 (3.4%) |
| WHERE_n_of_non_aggr_func__ROUND | |
| 0 | 954 (95%) |
| 1 | 43 (4.3%) |
| 2 | 3 (0.3%) |
| WHERE_n_of_non_aggr_func__DOW | |
| 0 | 966 (97%) |
| 1 | 33 (3.3%) |
| 2 | 1 (0.1%) |
| WHERE_n_of_all_non_aggr_func | |
| 0 | 619 (62%) |
| 1 | 262 (26%) |
| 2 | 85 (8.5%) |
| 3 | 26 (2.6%) |
| 4 | 6 (0.6%) |
| 5 | 2 (0.2%) |
| GROUP_BY_n_of_columns | [0.00, 2.00, 12.00] 3.12 / 2.84 |
| HAVING_n_of_main_predicates | |
| 0 | 448 (45%) |
| 1 | 192 (19%) |
| 2 | 184 (18%) |
| 3 | 176 (18%) |
| ORDER_BY_n_of_columns | |
| 0 | 276 (28%) |
| 1 | 338 (34%) |
| 2 | 244 (24%) |
| 3 | 93 (9.3%) |
| 4 | 34 (3.4%) |
| 5 | 10 (1.0%) |
| 6 | 4 (0.4%) |
| 7 | 1 (0.1%) |
| limit | [1, 516, 1,000] 509 / 287 |
| offset | [0, 0, 997] 252 / 327 |