

Сравнение аналитических хранилищ Vertica и ClickHouse

Вставка данных

100, 250, 500, 850, 1_000, ... - количество элементов в списке для вставки данных

Vertica

100	250	500	850	1_000	10_000	50_000	100_000	1_000_000	5_000_000
Median - 0.1892249584197998	Median - 0.09809112548828125	Median - 0.10187697410583496	Median - 0.12445902824401855	Median - 0.14134716987609863	Median - 1.0215561389923096	Median - 0.35663485527038574	Median - 0.6463849544525146	Median - 6.64872193336488	Median - 43.139312982559204
Average - 0.1892249584197998	Average - 0.09809112548828125	Average - 0.10187697410583496	Average - 0.12445902824401855	Average - 0.14134716987609863	Average - 1.0215561389923096	Average - 0.35663485527038574	Average - 0.6463849544525146	Average - 6.64872193336488	Average - 43.139312982559204
Summary - 0.1892249584197998	Summary - 0.09809112548828125	Summary - 0.10187697410583496	Summary - 0.12445902824401855	Summary - 0.14134716987609863	Summary - 1.0215561389923096	Summary - 0.35663485527038574	Summary - 0.6463849544525146	Summary - 6.64872193336488	Summary - 43.139312982559204

ClickHouse

100	250	500	850	1_00 0	10_0 00	50_0 00	100_ 000	1_00 0_00 0	5_00 0_00 0
Median - 0.05 2938 9381 4086 914	Median - 0.016 0298 3474 7314 453	Median - 0.09 3031 8832 3974 61	Median - 0.106 8029 4036 8652 34	Median - 0.05 2123 7850 1892 09	Median - 0.27 3169 9943 5424 805	Median - 0.34 4658 8516 2353 516	Median - 1.100 3768 4440 6128	Median - 9.90 4001 7127 9907 2	Median - 42.0 3411 4122 3907 5
Average - 0.05 293 893 8140 8691 4	Average - 0.01 602 983 4747 3144 53	Average - 0.09 3031 883 2397 461	Average - 0.10 680 294 036 865 234	Average - 0.05 2123 7850 1892 09	Average - 0.27 3169 994 354 248 05	Average - 0.34 465 8851 623 5351 6	Average - 1.10 0376 844 4061 28	Average - 9.90 4001 7127 9907 2	Average - 42.0 3411 4122 3907 5
Summary - 0.05 2938 9381 4086 914	Summary - 0.016 0298 3474 7314 453	Summary - 0.09 3031 8832 3974 61	Summary - 0.106 8029 4036 8652 34	Summary - 0.05 2123 7850 1892 09	Summary - 0.27 3169 9943 5424 805	Summary - 0.34 4658 8516 2353 516	Summary - 1.100 3768 4440 6128	Summary - 9.90 4001 7127 9907 2	Summary - 42.0 3411 4122 3907 5

Вывод: По результату исследования видно, что ClickHouse лидирует **по среднему значению** скорости вставки данных до отметки в 100_000 единиц элементов в списке, после этого **средние показатели** значения Vertica справляется с задачей вставки данных лучше.

Чтение данных

1_000, 10_000, 100_000, ... - количество считываемых данных

Vertica

1_000	10_000	100_000	1_000_0 00	2_000_ 000	5_000_ 000	10_000_ 000
-------	--------	---------	---------------	---------------	---------------	----------------

Median - 0.06271 409988 40332	Median - 0.26510 906219 48242	Median - 0.12449 908256 530762	Median - 0.08140 301704 406738	Median - 0.09993 600845 336914	Median - 0.09772 968292 236328	Median - 0.16525 4116058 3496
Average - 0.06271 40998 840332	Average - 0.2651 090621 948242	Average - 0.1244 990825 653076 2	Average - 0.0814 030170 440673 8	Average - 0.0999 360084 533691 4	Average - 0.0977 296829 223632 8	Average - 0.16525 411605 83496
Summary - 0.06271 409988 40332	Summary - 0.26510 906219 48242	Summary - 0.12449 908256 530762	Summary - 0.08140 301704 406738	Summary - 0.09993 600845 336914	Summary - 0.09772 968292 236328	Summary - 0.16525 4116058 3496

ClickHouse

1_000	10_000	100_000	1_000_000	2_000_000	5_000_000	10_000_000
Median - 0.00272 262096 405029 3	Median - 0.00324 010848 999023 44	Median - 0.00414 085388 183593 75	Median - 0.00307 404994 964599 6	Median - 0.01066 076755 523681 6	Median - 0.00317 895412 445068 36	Median - 0.00425 386428 833007 8
Average - 0.0044 783353 805541 99	Average - 0.0047 469615 936279 3	Average - 0.01140 375137 329101 6	Average - 0.0032 176256 179809 57	Average - 0.05671 994686 126709	Average - 0.0070 851087 570190 43	Average - 0.0069 395303 726196 29
Summary - 0.04478 335380 554199	Summary - 0.04746 961593 62793	Summary - 0.11403 7513732 91016	Summary - 0.03217 625617 980957	Summary - 0.56719 946861 26709	Summary - 0.07085 1087570 19043	Summary - 0.06939 530372 619629

Вывод: По результатам исследования видно, что ClickHouse существенно быстрее справляется с задачей чтения данных по среднему значению.

Аналитические вычисления (эксперимент проводился при равных объемах заполнения данных - 10_000_000 записей)

Vertica

Поиск максимального элемента колонки	Поиск медианного значения	Вычисление суммы
0.13187384605407715	1.0892741680145264	0.15050792694091797

ClickHouse

Поиск максимального элемента колонки	Вычисление энтропии столбца	Вычисление квантиля
0.1840221881866455	5.341540098190308	0.16817092895507812

И та и другая база показали хорошие результаты для нашего проекта и задач. Учитывая эксперименты по вставке и чтению данных выбор был сделан в пользу хранилища ClickHouse. Также ClickHouse полностью бесплатное решение в отличие от Vertica. С кодом можно ознакомиться в тетрадках jupyter-notebook.