

性能测试报告

1. 测试目的:

- 1.1 验证 PrestoSQL 搭载 MySQL 是否可以提高 MySQL 的优化效率?
- 1.2 验证是否可以通过水平扩展 PrestoSQL 的 Worker 节点数量或内存来提升优化效率?

2. 测试方案

2.1. 数据准备

采用 TPC-DS (采用 7 张事实表, 17 张维度表平均每表含有 18 列, 其工作负载包含 99 个 SQL 查询, 覆盖 SQL99 和 2003 的核心部分以及 OLAP) 作为数据来源支撑。利用 TPC-DS 基准测试工具生成 1G、5G、20G 的数据分别进行单大表测试和多表关联测试

对于多表关联测试, 选择了 15 条具有代表性的 SQL 语句, 对于单大表测试, 选择 TPC-DS 生成测试数据集中数据量最大的表 store_sales 并选用了 9 条使用频率高的常规性聚合 SQL 语句进行测试, 几乎所有的测试案例都有很高的 IO 负载和 CPU 计算需求。

	单表数据量(行), 以`store_sales`表为例
1G 数据	3022922
5G 数据	16973841
20G 数据	56561113

2.2. 服务器配置

2.2.1. 数据量 1G 服务器配置 (Presto 单节点部署)

服务器	配置
MySQL-5.7	2 核 2 G
PrestoSQL-346	2 核 16 G

2.2.2. 数据量 5G 的服务器配置 (PrestoSQL 分布式部署)

服务器	配置
MySQL-5.7	2 核 2 G
PrestoSQL-346 (Coordinator)	4 核 6 G
PrestoSQL-346 (Worker)	4 核 6 G
PrestoSQL-346 (Worker)	4 核 6 G

2.2.3. 数据量 20G 的服务器配置 (PrestoSQL 分布式部署)

服务器	配置
MySQL-5.7	2 核 2 G
PrestoSQL-346 (Coordinator)	4 核 8 G
PrestoSQL-346 (Worker)	4 核 8 G

PrestoSQL-346 (Worker)	4 核 8 G
------------------------	---------

3. 测试结果

3.1. 多表测试结果：

数据量为 1G 的多表测试结果 (S)		
SQL (TPC-DS)	Mysql	PrestoSQL
SQL_01	4.45	5.45
SQL_02	0.08	4.63
SQL_03	1.86	1.67
SQL_04	1.84	14.17
SQL_05	4.53	5.21
SQL_06	5.48	7.9
SQL_07	2.55	9.12
SQL_08	4.17	4.34
SQL_09	1.63	0.63
SQL_10	6.41	6.43
SQL_11	3.09	8.93
SQL_12	6.13	6.79
SQL_13	10.98	21.22
SQL_14	12.12	20.22
SQL_15	1.37	3.26

数据量为 5G 的多表测试结果 (S)		
SQL (TPC-DS)	Mysql	PrestoSQL
SQL_01	43.91	22.85
SQL_02	1.12	37
SQL_03	53.01	33.05
SQL_04	33.57	53.64
SQL_05	64.94	36.21
SQL_06	95.99	44.06
SQL_07	41.47	44.84
SQL_08	43.54	20.26
SQL_09	16.03	0.73
SQL_10	84.87	39.86
SQL_11	41.42	44.43
SQL_12	81.41	39.74
SQL_13	261.39	68
SQL_14	272.02	60.06

SQL_15	32.62	28.87
--------	-------	-------

数据量为 20G 的多表测试结果 (S)		
SQL (TPC-DS)	Mysql	PrestoSQL
SQL_01	170.85	113
SQL_02	3.01	139
SQL_03	209.82	109
SQL_04	130.25	213
SQL_05	240.83	140
SQL_06	342.83	175
SQL_07	152.51	175
SQL_08	167.66	105
SQL_09	4.31	0.62
SQL_10	310.28	160
SQL_11	149.03	175
SQL_12	313.91	160
SQL_13	1067	286
SQL_14	1048.43	290
SQL_15	132.08	125

3.2 单表测试数据

数据量为 1G 的单表测试结果 (S)		
SQL (TPC-DS)	Mysql	PrestoSQL
SQL_01	0.92	1.43
SQL_02	1.27	1.33
SQL_03	1.54	2.76
SQL_04	0	0.97
SQL_05	0.75	1.33
SQL_06	12.19	6.89
SQL_07	12.25	5.83
SQL_08	3.8	12.91
SQL_09	4.72	13.31

数据量为 5G 的单表测试结果 (S)		
SQL (TPC-DS)	Mysql	PrestoSQL
SQL_01	22.9	25.19
SQL_02	25.47	36.46
SQL_03	28.22	28.64
SQL_04	0.01	23.52
SQL_05	22.45	25.03

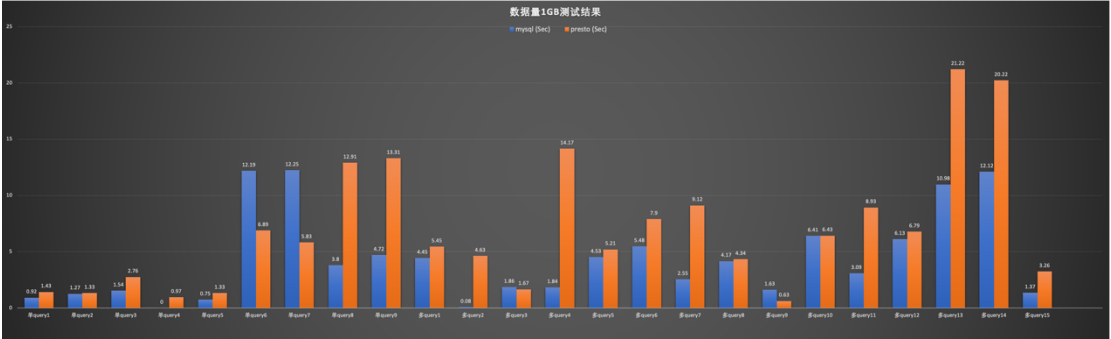
SQL_06	210	38.13
SQL_07	208	37.36
SQL_08	41.05	46.39
SQL_09	28.49	28.89

数据量为 20G 的单表测试结果（S）		
SQL（TPC-DS）	Mysql	PrestoSQL
SQL_01	100.14	110
SQL_02	110.25	111
SQL_03	115.95	118
SQL_04	0.04	98
SQL_05	97.67	108
SQL_06	1087.26	147
SQL_07	1081.66	138
SQL_08	159.97	125
SQL_09	114.98	116

下面是通过直方图形式直观比较 presto 和 mysql 的 SQL 执行时间。

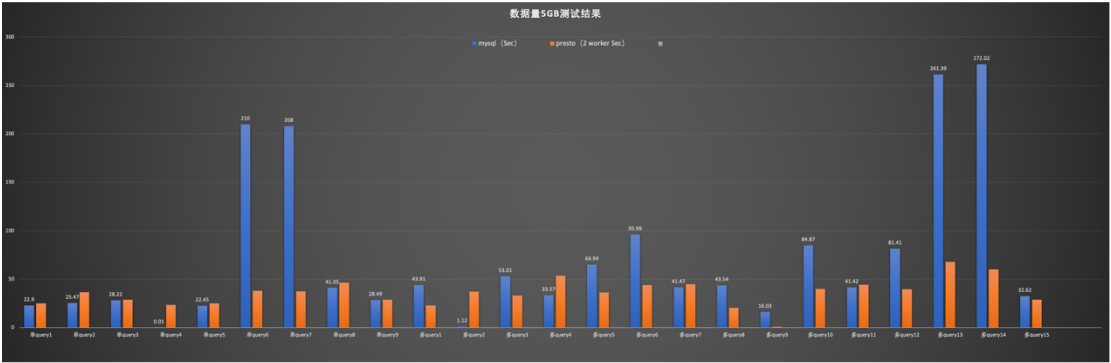
数据量 1G 服务器配置

PrestoSQL（单节点，2 核 16G）



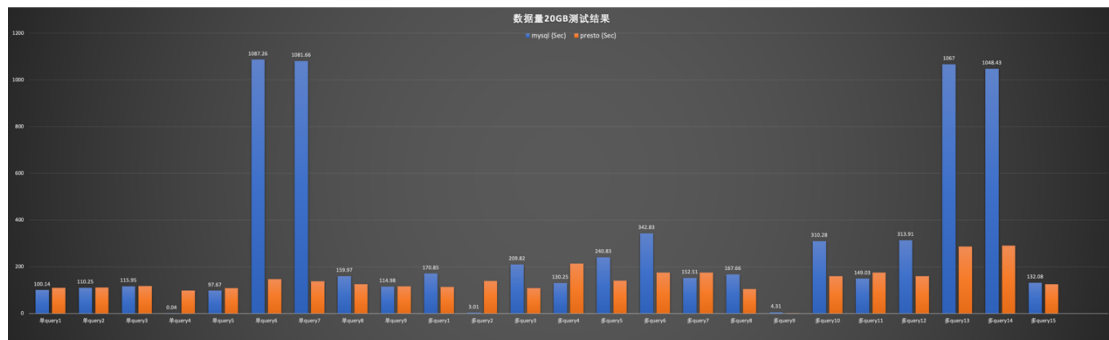
数据量 5G 服务器配置：

PrestoSQL（分布式部署）Coordinator（4 核 6G）、Worker * 2（4 核 6G）



数据量 20G 服务器配置:

PrestoSQL (分布式部署) Coordinator (4 核 8G)、Worker * 2 (4 核 8G)



4. 测试总结

从上面的测试结果可以看出, PrestoSQL 综合性能要比 MySQL 好一些, 在其查询性能和支持的数据源方面要突出一些, 同时在大数据量、多表查询、多节点、单大表聚合操作方面体现出了优势。

由于 PrestoSQL 是完全基于内存计算的, 所以 PrestoSQL 在查询时占用的内存不少。

5. 测试结论

PrestoSQL 搭载 MySQL 可以提高 MySQL 的优化效率 (本次测试)

可以通过水平扩展 PrestoSQL 的 Worker 节点数量或内存来提升优化效率 (本次测试)

6. 参考文档

<https://prestosql.io/docs/current/installation/deployment.html#jvm-config>

(PrestoSQL 官方部署文档)

<http://tpc.org/tpcds/default5.asp> (TPC 测试相关官方文档)

<https://bbs.huaweicloud.com/blogs/173486> (PrestoSQL 调优相关文档)

<https://blog.csdn.net/hancky/article/details/88954567> (TPC-DS 使用文档)