# 分布键推荐实验结果

环境：2cn+6dn 3节点分布式数据库

## 数据集

### 某行

23个表

最大250w

最小29行

表的列数5列到66列不等：

### Tpch 10x

考虑sql执行耗时具有浮动性，对默认分布键场景与推荐分布键场景下分别对22条复杂sql各执行10次后取10次的平均值作为最终结果对比。过程数据见附件《分布键推荐测试数据.xlxs》。



* 默认分布键

|  |  |
| --- | --- |
| 表名 | 列名 |
| customer | c\_custkey |
| lineitem | l\_orderkey |
| orders | o\_orderkey |
| part | p\_partkey |
| partsupp | ps\_partkey |
| region | r\_regionkey |
| supplier | s\_suppkey |
| nation | n\_nationkey |

* 第一版本分布键

|  |  |
| --- | --- |
| 表名 | 列名 |
| customer | c\_custkey |
| lineitem | l\_orderkey |
| orders | o\_orderkey |
| part | p\_partkey |
| partsupp | ps\_partkey |
| region | 复制表 |
| supplier | s\_suppkey |
| nation | 复制表 |

* 第二版本推荐分布键

|  |  |
| --- | --- |
| 表名 | 列名 |
| customer | c\_custkey |
| lineitem | l\_orderkey |
| orders | o\_orderkey |
| part | p\_partkey |
| partsupp | ps\_partkey |
| region | r\_regionkey |
| supplier | s\_suppkey |
| nation | 复制表 |

## 分布键推荐脚本执行时间

* 分布键推荐脚本执行时间

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 单位\s | TPCH | | 某行 | |
| 驱动 | 非驱动 | 驱动 | 非驱动 |
| C00 | 不支持 | 2.102s | 不支持 | 20.056s |
| C10 | 0.993s | 1.518s | 3.323s | 6.118s |
| 提升比 | 52% | 27% | 83% | 68% |

* 分布键推荐脚本执行时

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| /s | sql1 | sql2 | sql3 | sql4 | sql5 |
| default | 0.144121 | 0.178257 | 5.576713 | 0.155217 | 0.211475 |
| C00 | 0.138968658 | 0.137121 | 1.839193 | 0.132667 | 0.141272 |
| C10 | 0.130529255 | 0.131626 | 0.642306 | 0.124246 | 0.131203 |

当前版本推荐的分布键整体优于默认分布键与C00版本推荐结果。其中C00版本推荐结果Sql3性能显著提升2倍。C10版本推荐结果sql3性能提升高达7倍。