OLAP Parquet Proto 对象写入性能

### OLAP数据Parquet写入性能压测

TS5 2\*8CPU 64GB 12\*1TB SATA

写入原数据： 94GB PB数据大小为：5KB/条

Cpu率先达到瓶颈 100%此时写入性能为：

|  |  |  |
| --- | --- | --- |
| 存储介质|压缩方式 | SNAPPY | 无压缩 |
| HDFS | 16000/s | 24GB | 17000/s | 32GB |
| LOCAL DISK | 18400/s | 24GB | 20000/s | 32GB |

TS80 2\*8CPU(E5-2620v4)，64GB，4\*1.8T SSD，1\*240G SSD

Cpu率先达到瓶颈 100%此时写入性能为：

|  |  |  |
| --- | --- | --- |
| 存储介质|压缩方式 | SNAPPY | 无压缩 |
| HDFS | 30000/s | 24GB | 32000/s | 32GB |
| LOCAL DISK | 38000/s | 24GB | 40000/s | 32GB |

Hermes存储设计为HDFS故写入性能评估为：

|  |  |  |
| --- | --- | --- |
|  | 写入极限性能 | 综合读写 |
| TS5 | 16000/s | 10000/s |
| TS80 | 30000/s | 18000/s |

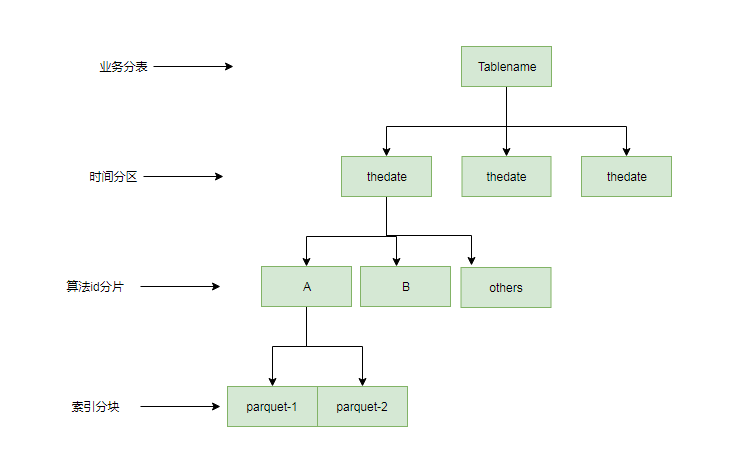
### OLAP业务数据及存储设计

#### 业务数据

1. OLAP数据包含62个业务模块，每个模块单独一份数据；
2. 总数居流量为 154亿/天，单条数据大小为 5~10kb/条，总存储 77TB，写入峰值35w/s；
3. 每个业务表包含数百个算法id数据；
4. 每个业务表包含两个主要算法id A B,占有90%的数据；
5. 每次分析都会指定相关的算法id

#### 存储设计

业务分表🡪时间分区🡪算法分片🡪索引分块



写入压缩比：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 10038 | 100000037 | 100000078 |
| 源数据 | 841M | 905M | 985M |
| 无裁剪索引 | 177M | 242M | 350M |
| 裁剪索引 | 117M | 197M | 107M |
| 裁剪压缩比 | 66% | 81% | 31% |