

高性能消息数据存储引擎的设计解析

李淼

融云-首席架构师

SHANGHAI

极客时间VIP年卡

每天6元, 365天畅看全部技术实战课程

- 20余类硬技能, 培养多岗多能的混合型人才
- 全方位拆解业务实战案例, 快速提升开发效率
- 碎片化时间学习, 不占用大量工作、培训时间



融云，通信云行业领导者



安全·可靠的全球互联网通信云

通讯行业领导者
为开发者、企业提供安全、稳定、可靠
覆盖全球的通讯云服务
连续多年市场占有率稳居第一



李淼-融云首席架构师，联合创始人
10年IM领域设计和研究经验

- 实时通讯云计算平台
- 日均活跃数 6500 万
- 日消息峰值 2200 亿

TABLE OF
CONTENTS 大纲

- 即时通讯消息存储特点
- 消息存储引擎设计
- 消息存储引擎优化
- 消息存储服务架构设计

即时通讯消息存储特点

- 时间顺序进行排序
- 存储具有时效性，定期淘汰
- 写入并发量高
- 写入读取比一般在5 : 1

融云消息存储历程

- 原型验证-MySQL
- 正式阶段一-Redis
- 正式阶段二-LevelDB
- 正式阶段三-Redis

为什么要自研存储

- 满足复杂的数据业务逻辑
- 降低设备成本
- 简化部署模型
- 源码基本可控

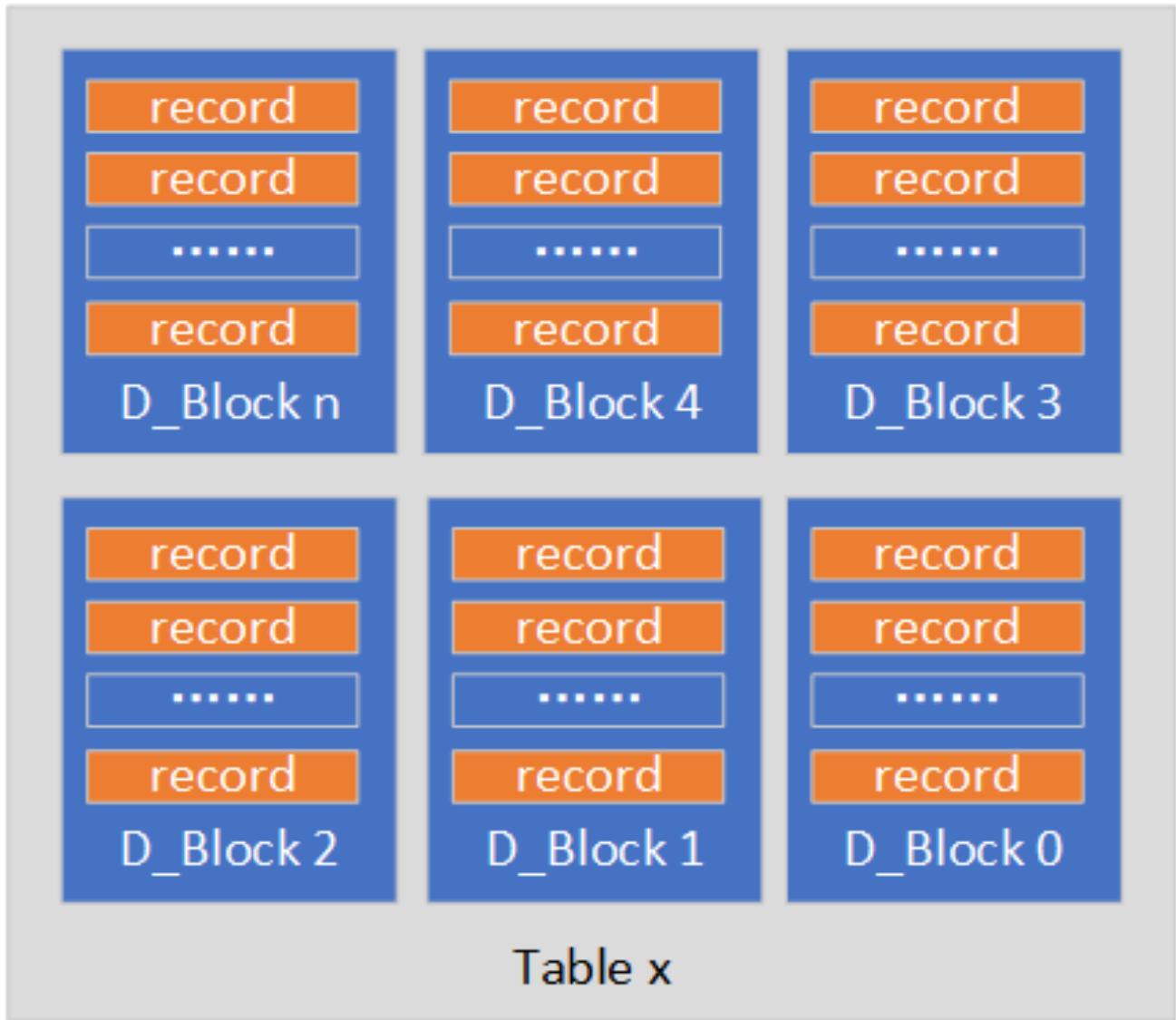
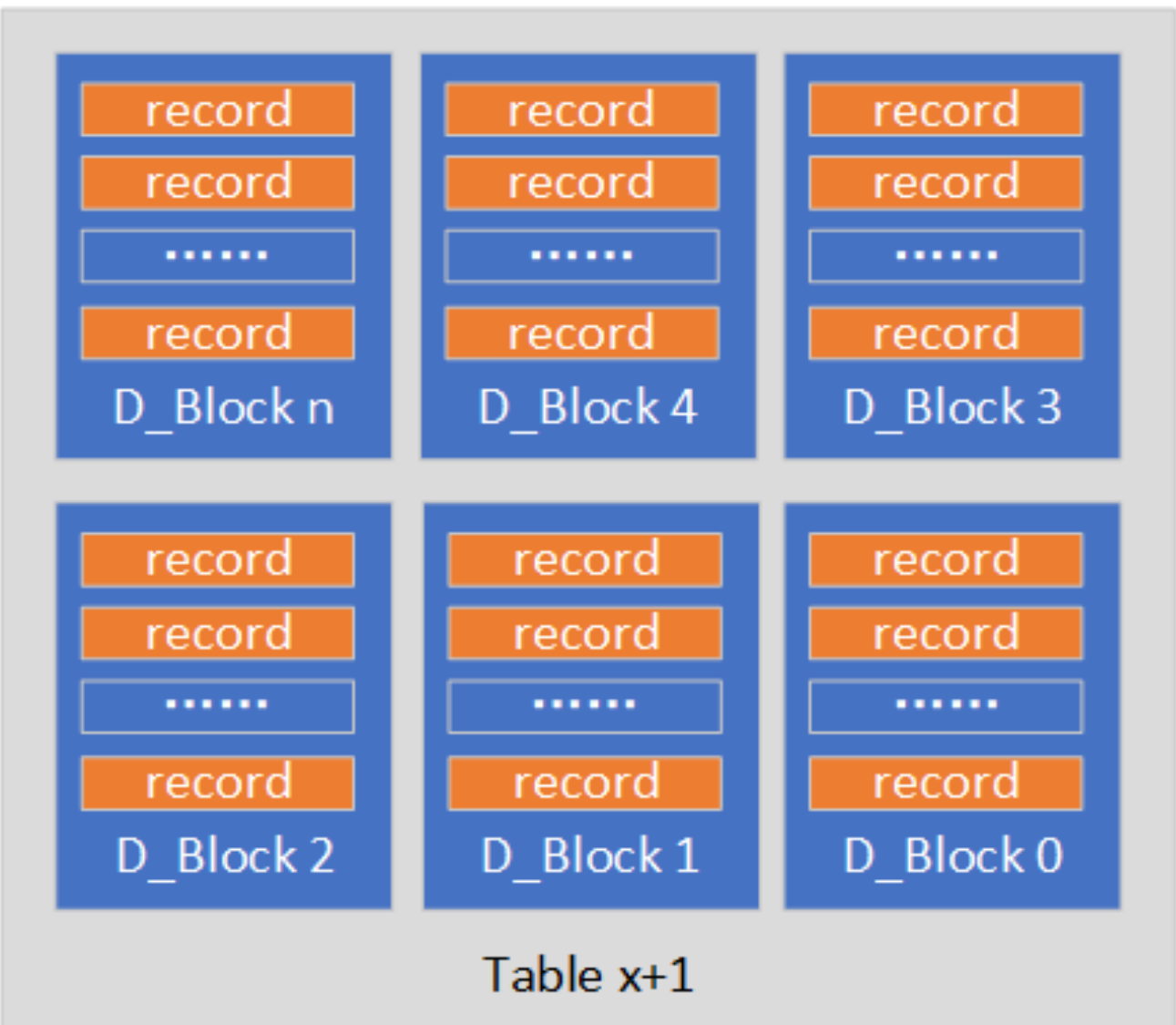
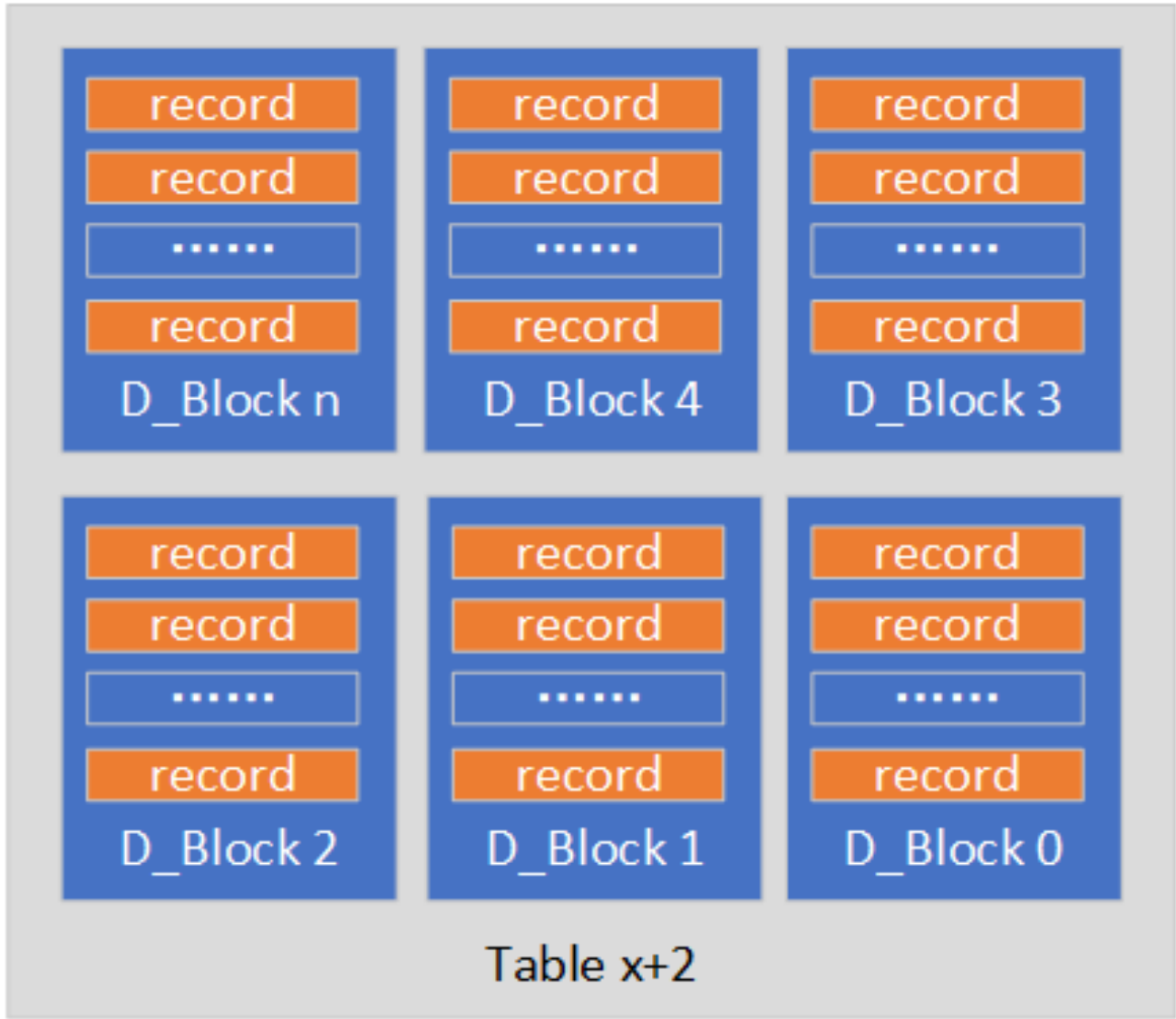
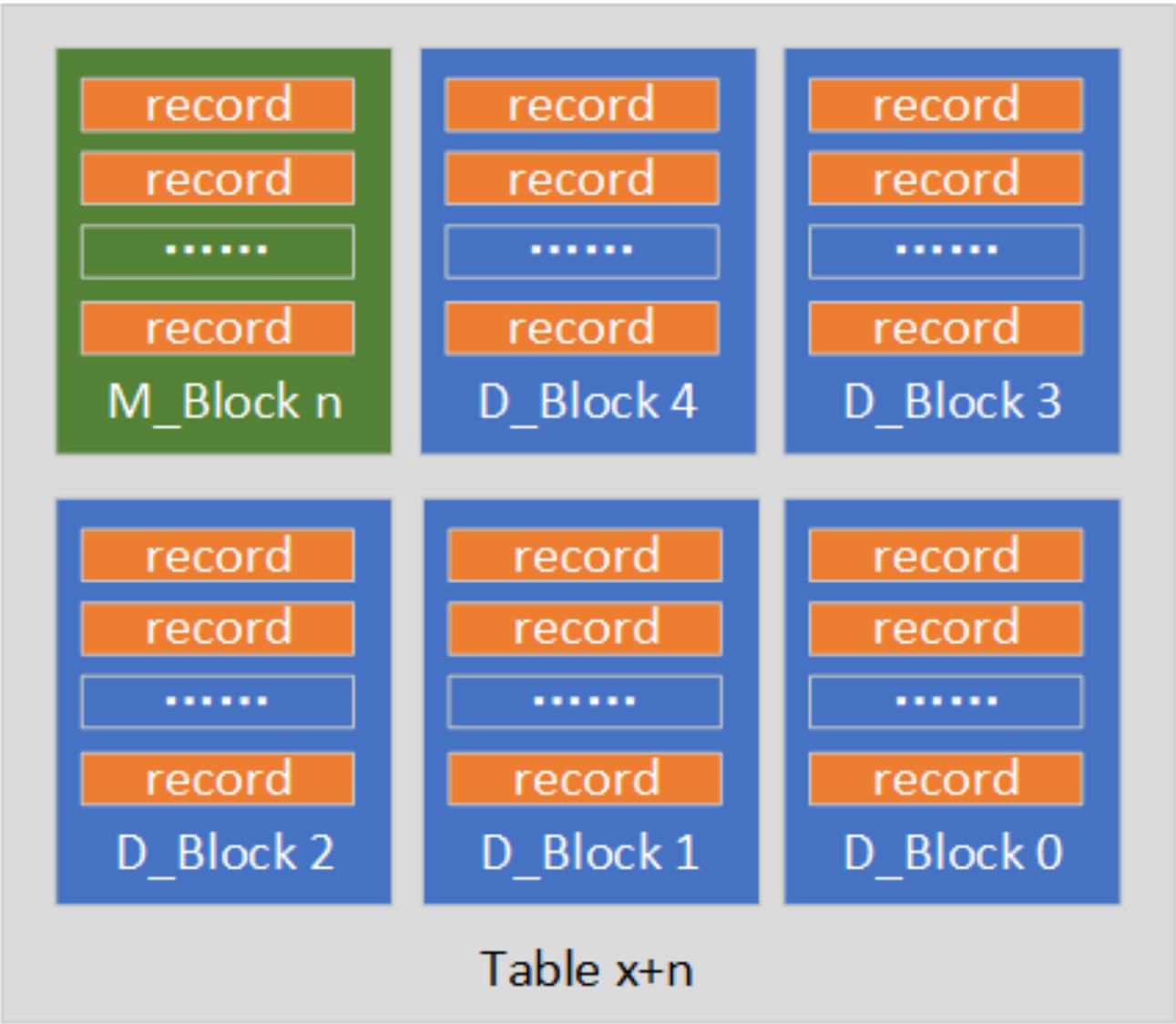
存储设计要求

- 快速数据淘汰
- 避免数据合并
- 读写性能要求高
- 开发使用灵活

站在前人的肩膀

- 数据采用WAL写入
- 借鉴 InfluxDB 的 LSM 树
- 借鉴 whiskey 的 K / V 分离存储
- 借鉴 MyISAM 的文件定义

存储逻辑划分



存储文件规划

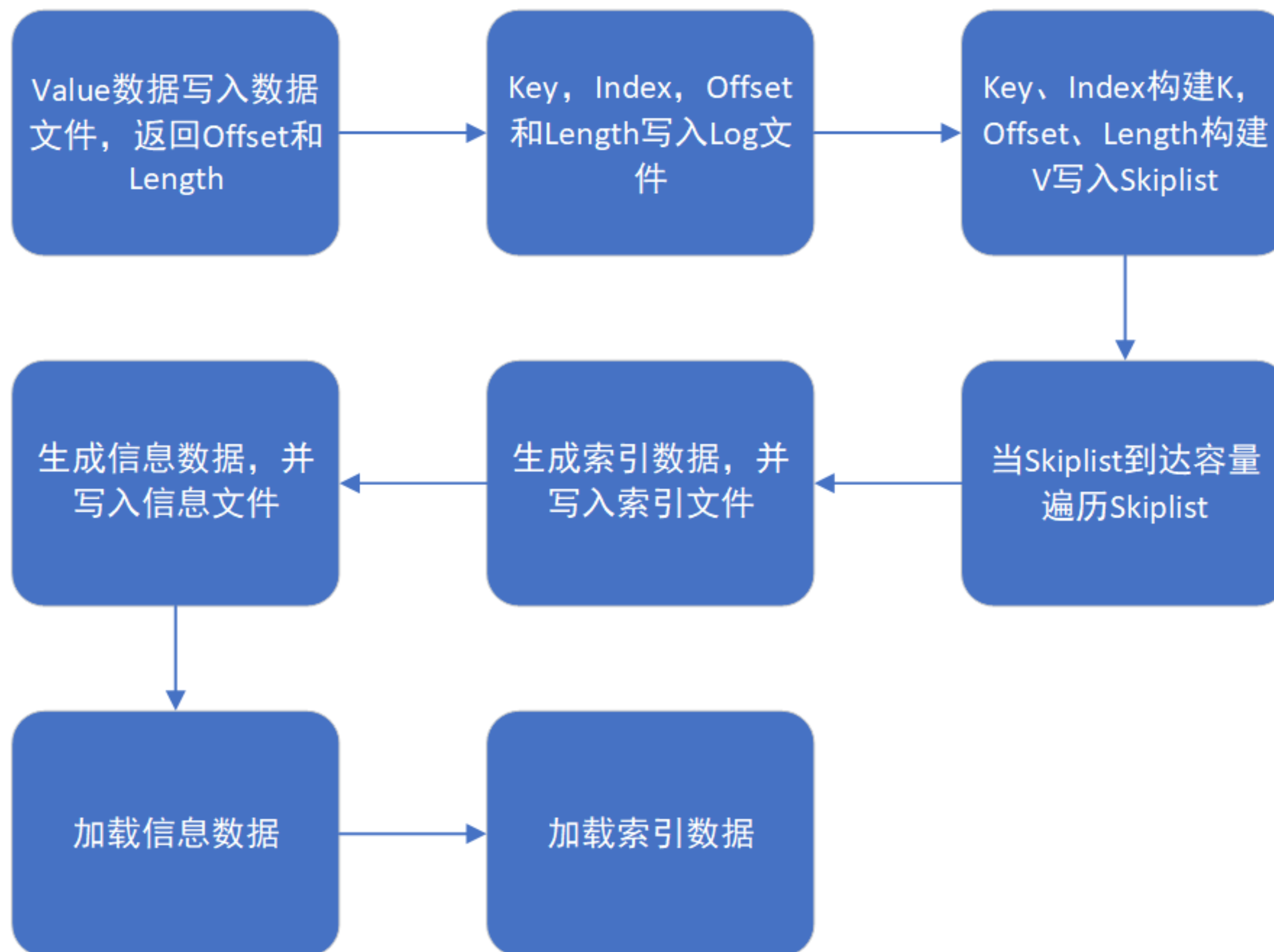
➤ Table

- ✓ xxx.data 数据存储文件
- ✓ xxx.index 数据索引文件
- ✓ xxx.info table信息文件

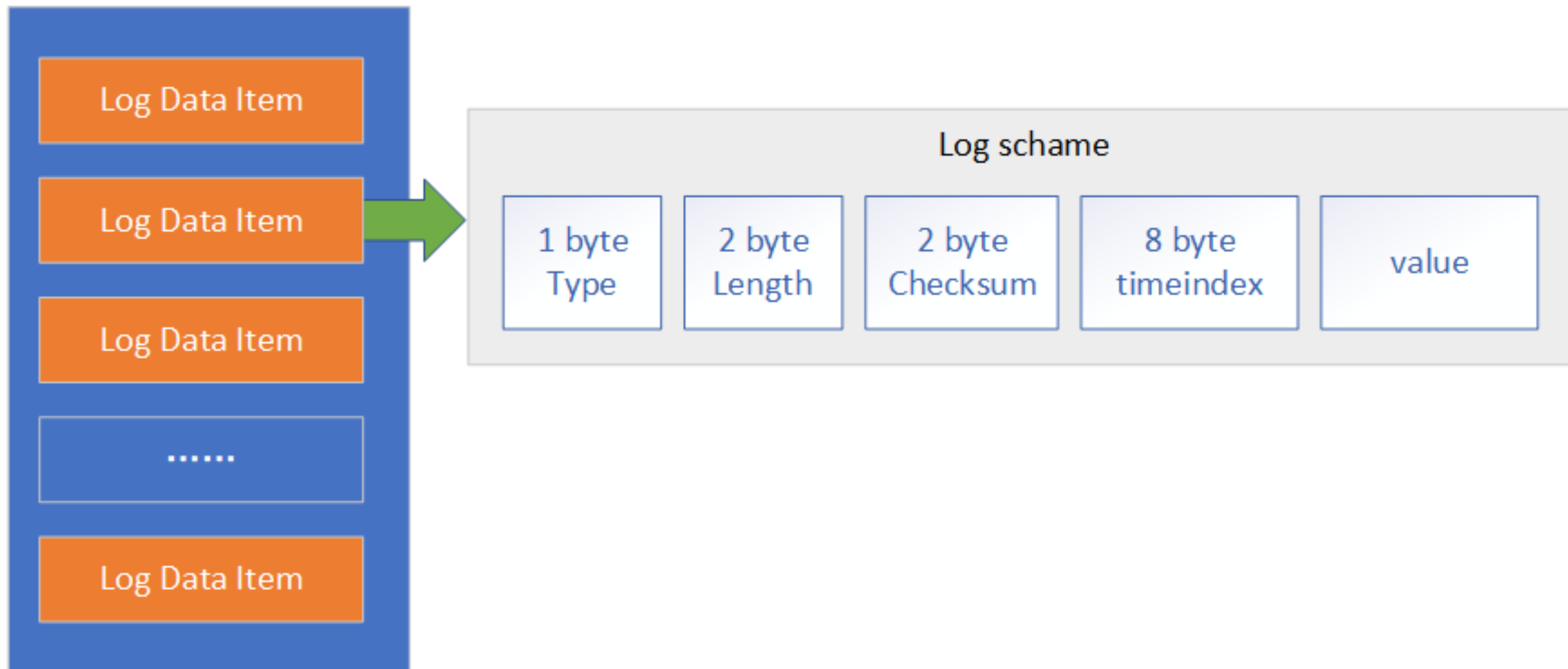
➤ Log

- ✓ xx.log 日志信息文件

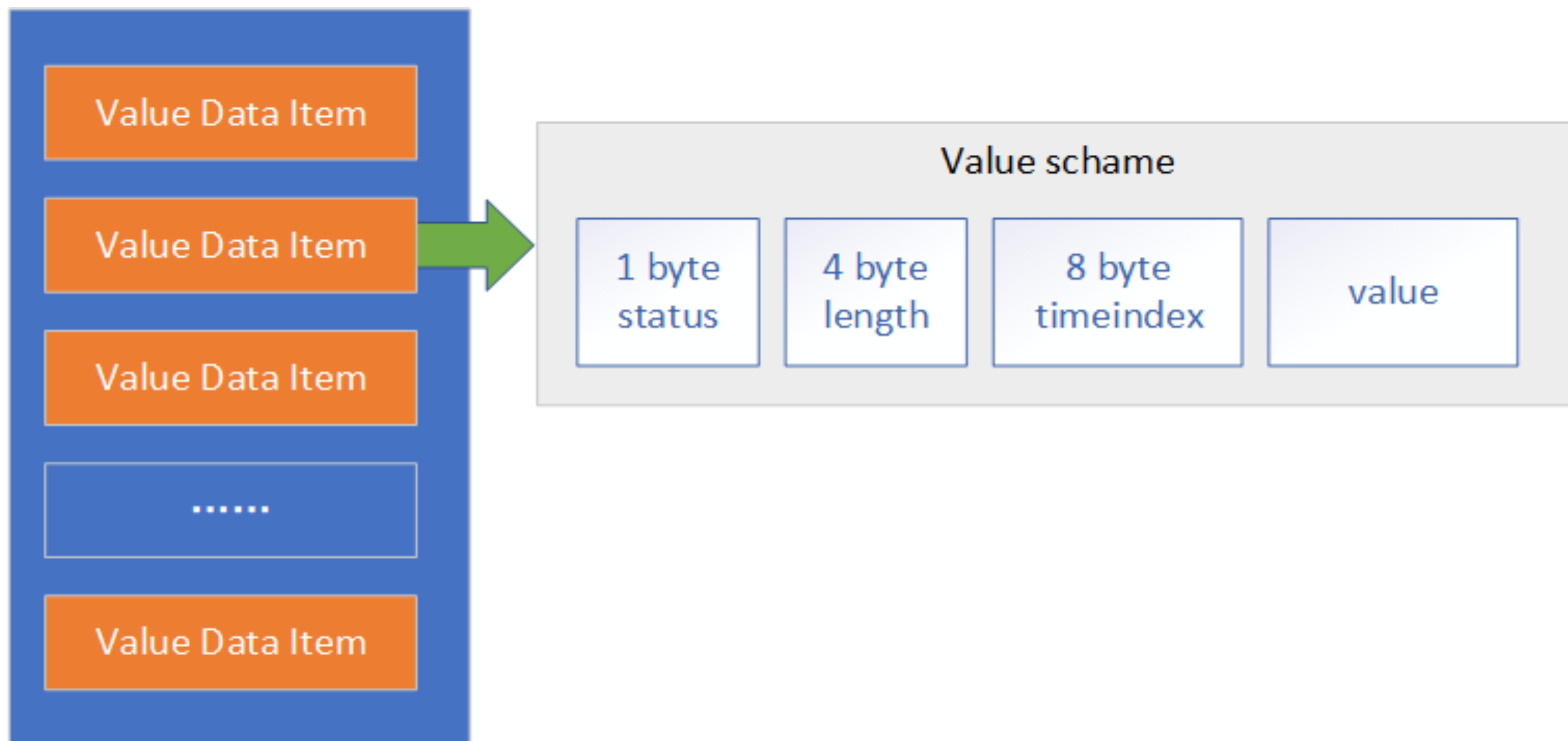
数据写入逻辑



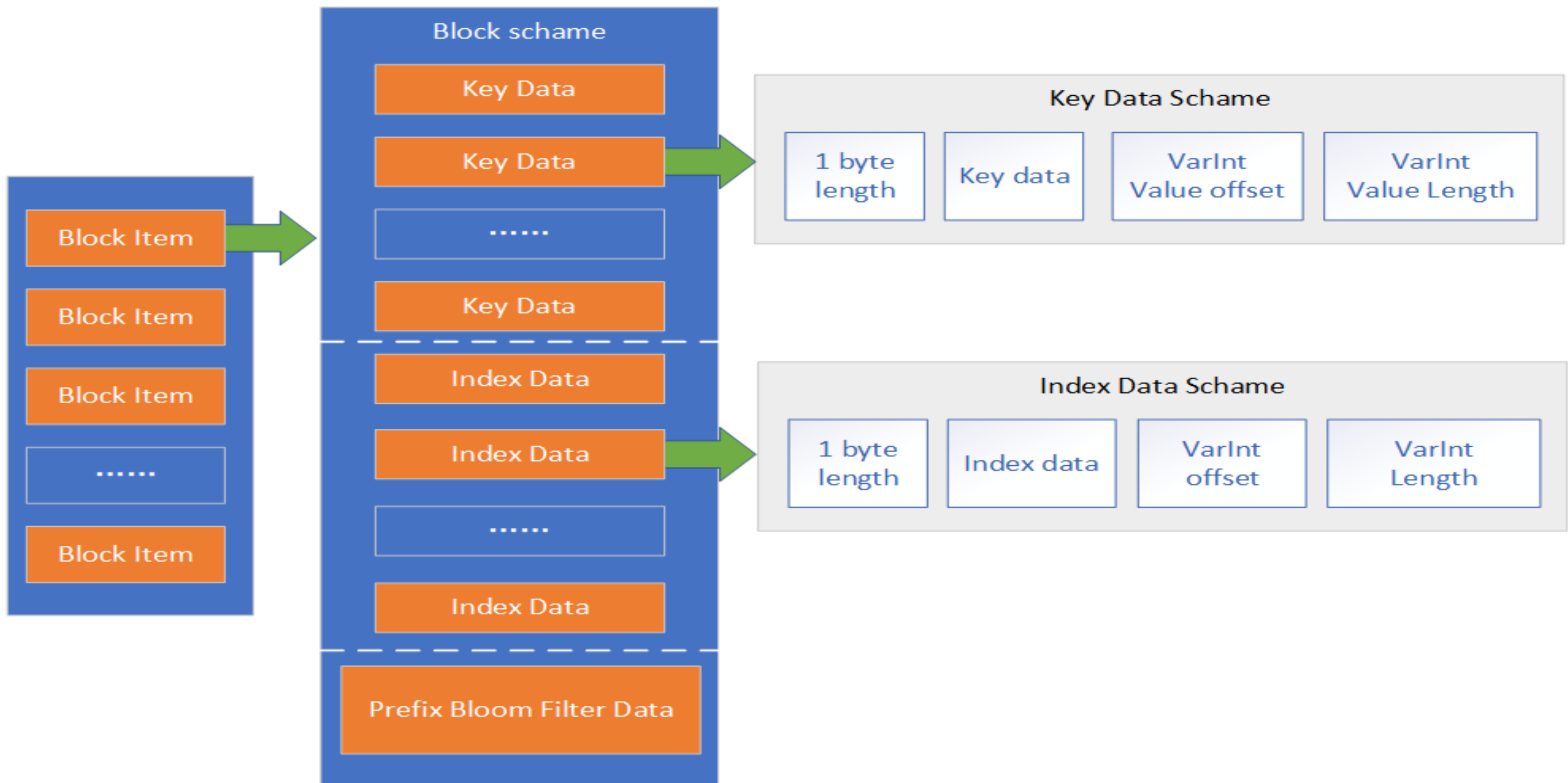
日志文件设计



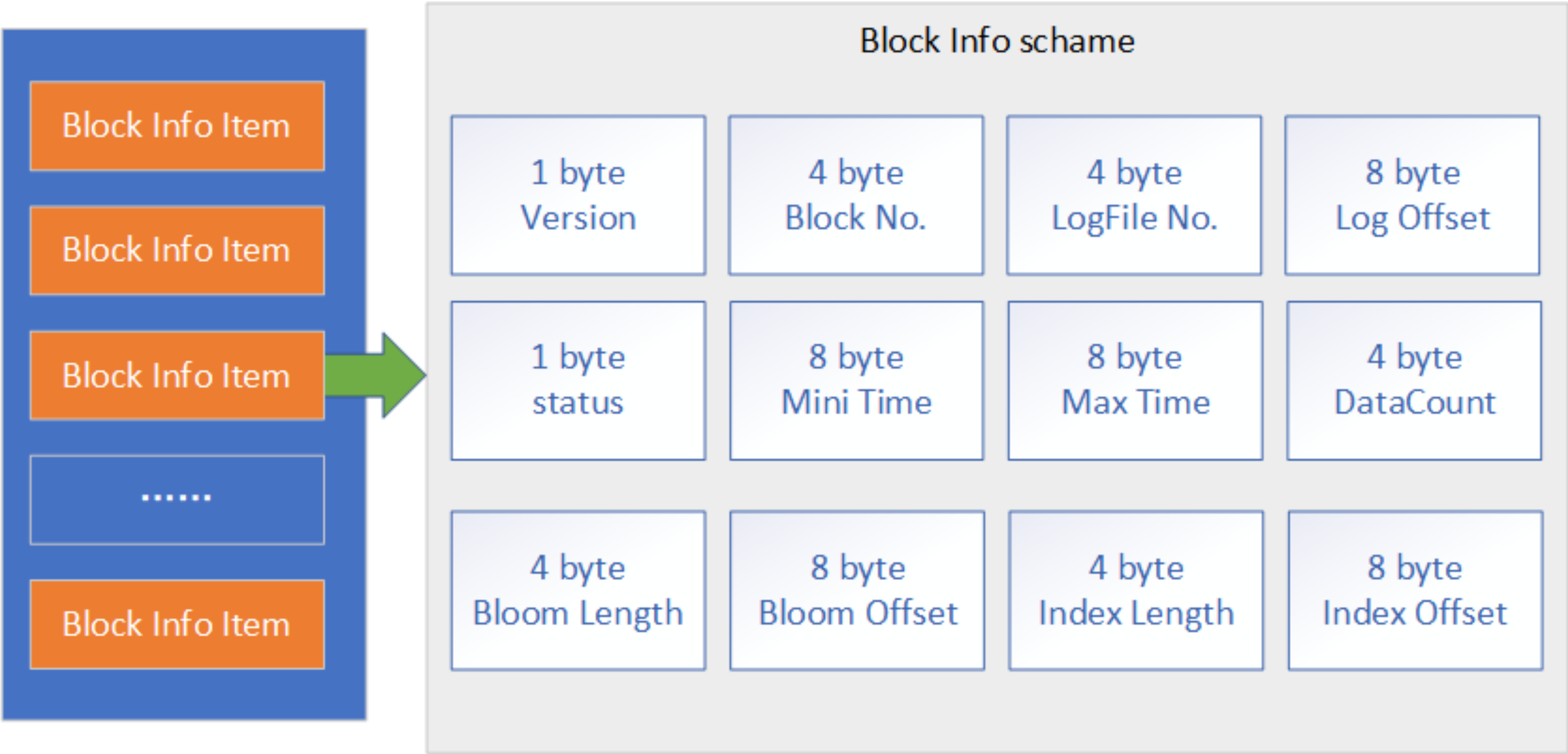
数据存储文件设计



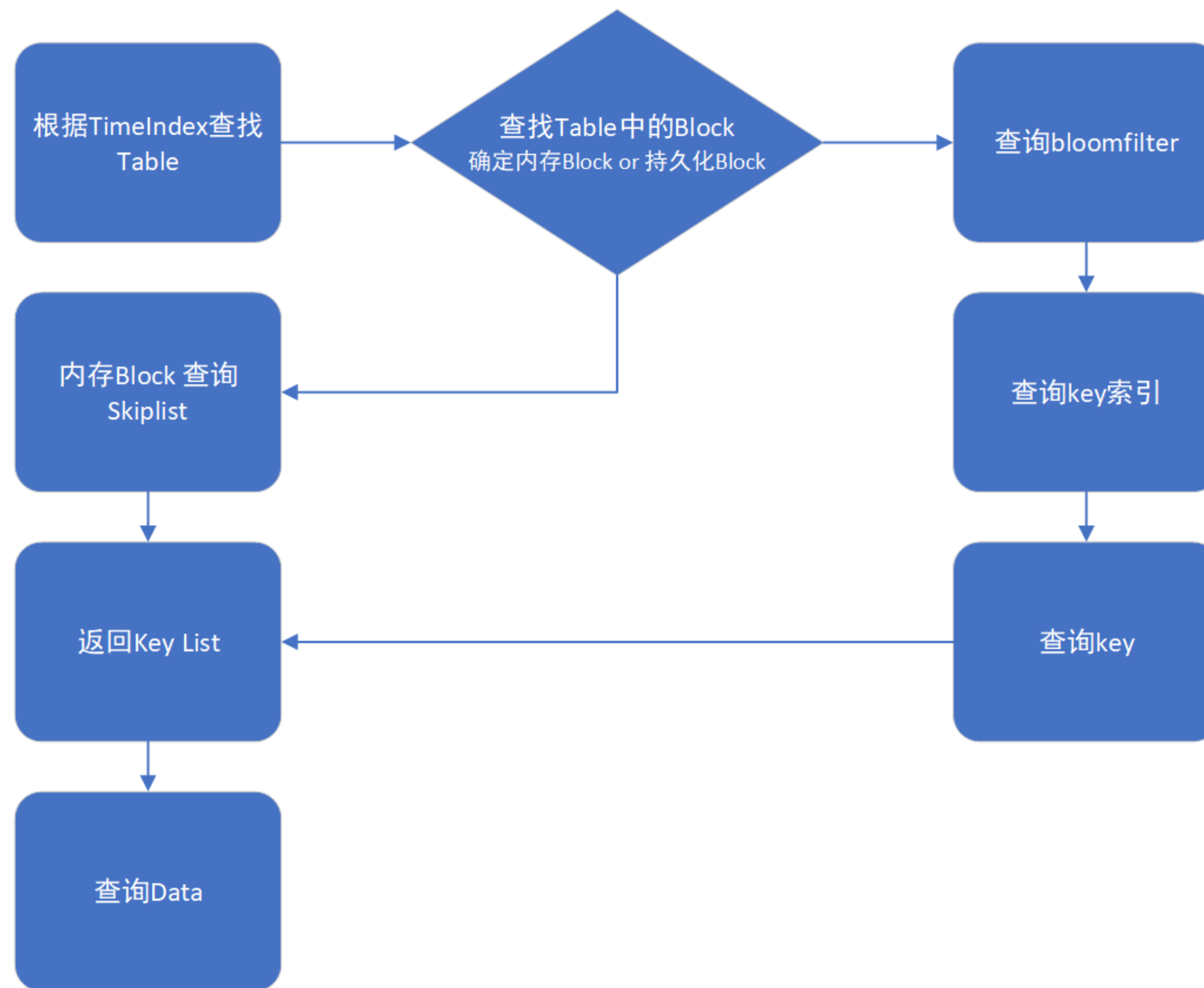
数据索引文件设计



table信息文件设计



数据查询逻辑



内存优化篇

- 重写的SkipList，内存尽有Java中的 SkipList 1/4
- 40亿数据索引，尽消耗400MB内存
- 实现内存对象分配器
- 实现 LRU 进阶的 LIRS 缓存

存储优化篇

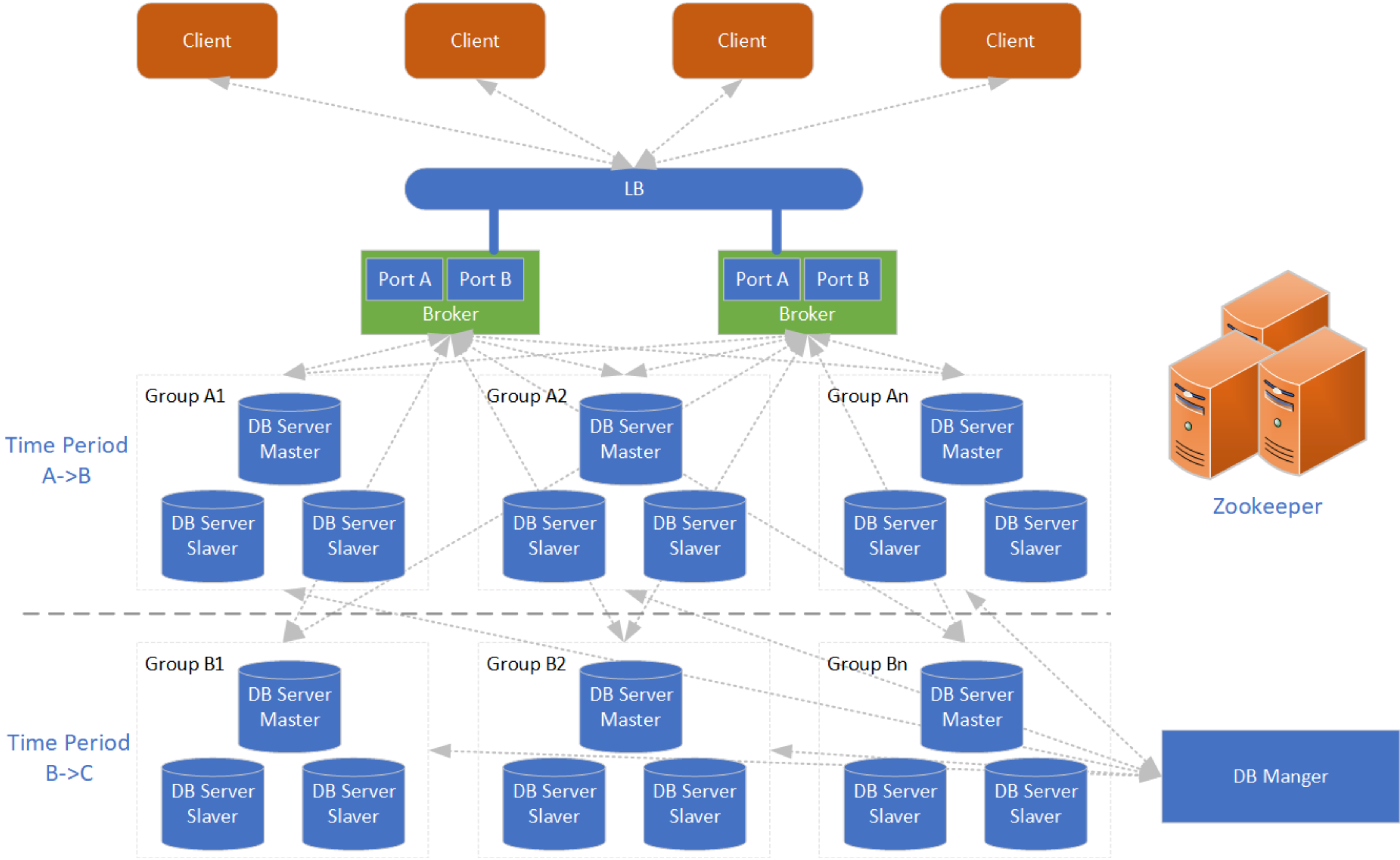
- 索引数据前缀压缩
- 数值数据 VarInt 编码
- 业务数据 quicklz 压缩
- 数据写入采用双循环Buffer
- 重复数据引用写入

性能数据指标

写入速率测试		
数据量	耗时（毫秒）	速率（条/秒）
1,000,000	1,562	640,205
5,000,000	7,584	659,283
10,000,000	15,393	649,646
50,000,000	76,740	651,551
100,000,000	152,027	657,778
读取速率测试		
数据量	耗时（毫秒）	速率（条/秒）
1,000,000	3,483	287,109
2,000,000	6,970	286,944
5,000,000	16,176	309,100
10,000,000	32,135	311,187

CPU : Intel i7 8550U 内存 : 16GB JVM 4GB 硬盘 : PCIe SSD

服务端架构



服务端特点

- ✓ 无数据迁移扩容
- ✓ 自动主从切换
- ✓ 异步长连接客户端
- ✓ 多协议适配 (MQTT、Websocket、HTTP2)

特别提示

- 数据存储引擎未来两个月会进行开源

安全、可控、高效、稳定

的企业即时通讯统一平台就是融云

感谢聆听



- ▶ www.rongcloud.cn
- ▶ support@rongcloud.cn | 400 - 919 - 9066
- ▶ 北京市朝阳区北苑路北甲13号院1号楼 北辰泰岳大厦 14层

AiCon

2018.12.20-23 / 北京·国际会议中心

AI商业化下的技术演进实战干货分享

京东：智能金融

景驰科技：自动驾驶

阿里巴巴：NLP

清华人工智能研究院：机器学习

今日头条：机器学习

Twitter：搜索推荐

AWS：计算机视觉

Netflix：机器学习



扫码了解详情

技术创新的浪潮接踵而来， 继续搬砖还是奋起直追？

云数据

AI

区块链

架构优化

高效运维

CTO技术选型

微服务

新开源框架

会议：2018年12月07-08日 培训：2018年12月09-10日

地址：北京·国际会议中心



THANKS!

SHANGHAI