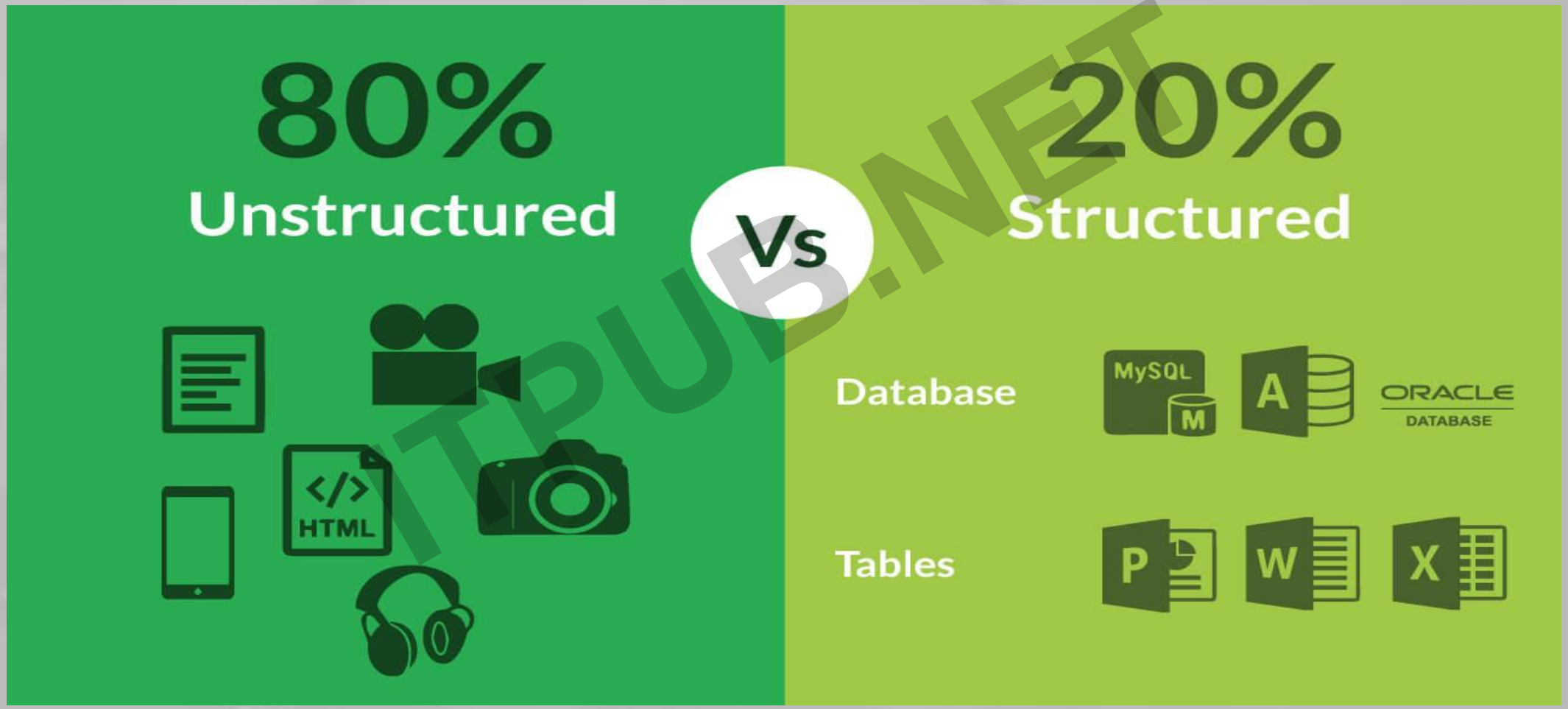
Key/Value SSD的原理和应用

苗宁忠

2019-05-09

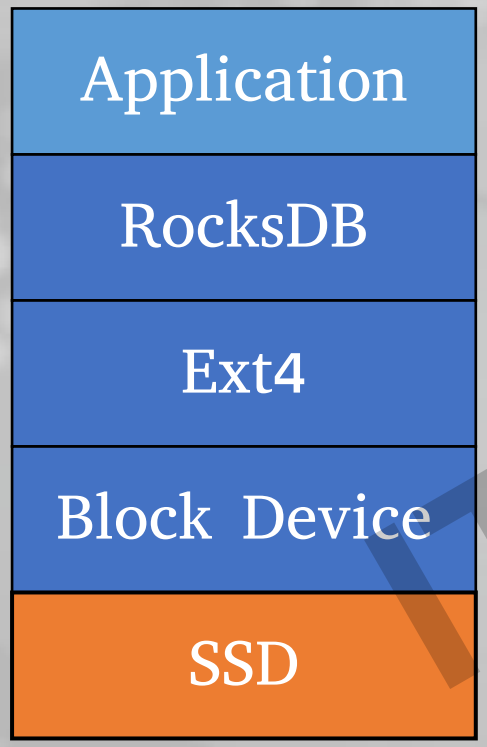
Shannon Systems
宝 存 科 技

Unstructed data dominates

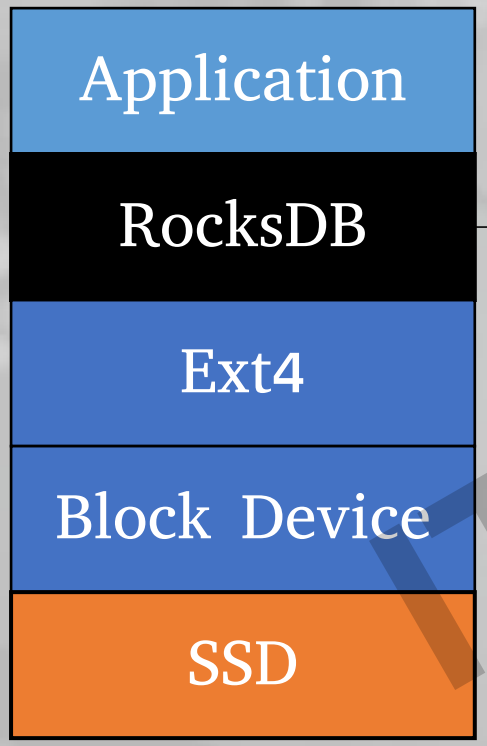


Traditional KV Store

Traditional KV Store

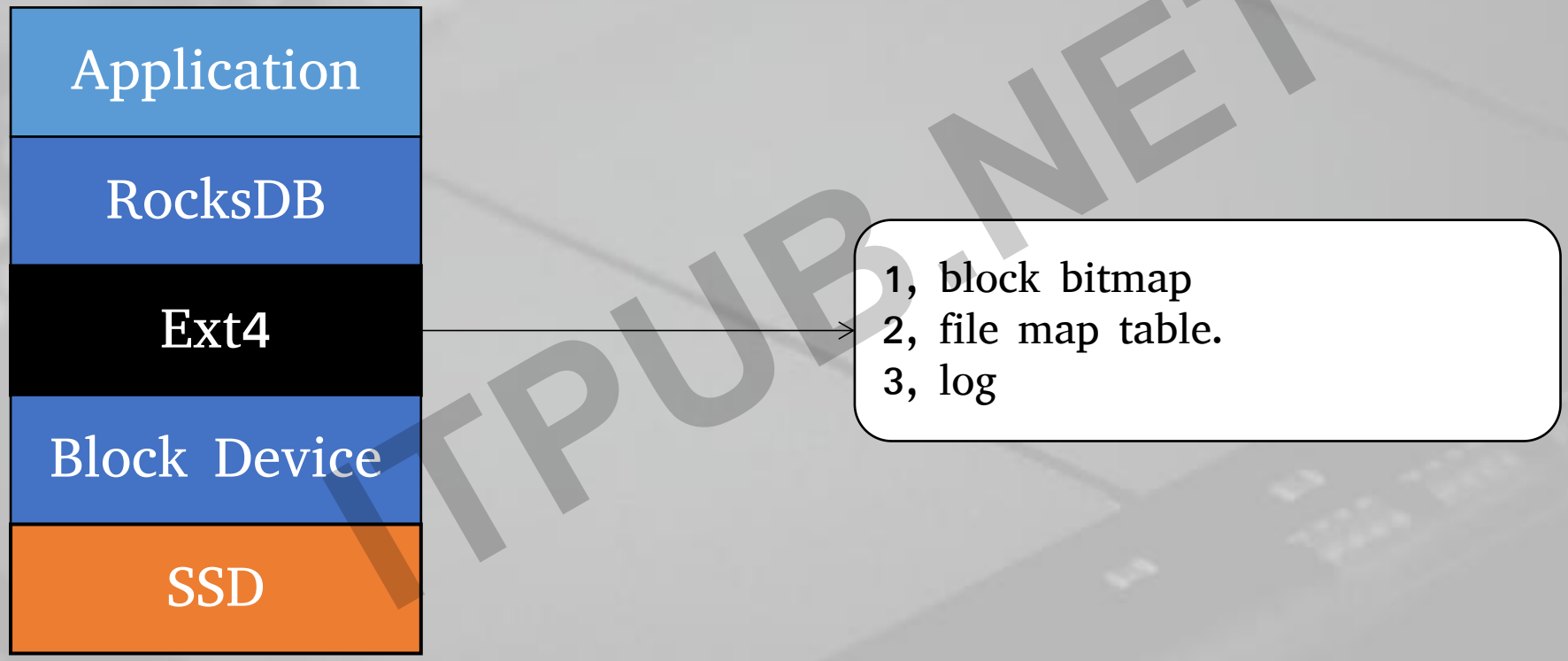


Traditional KV Store

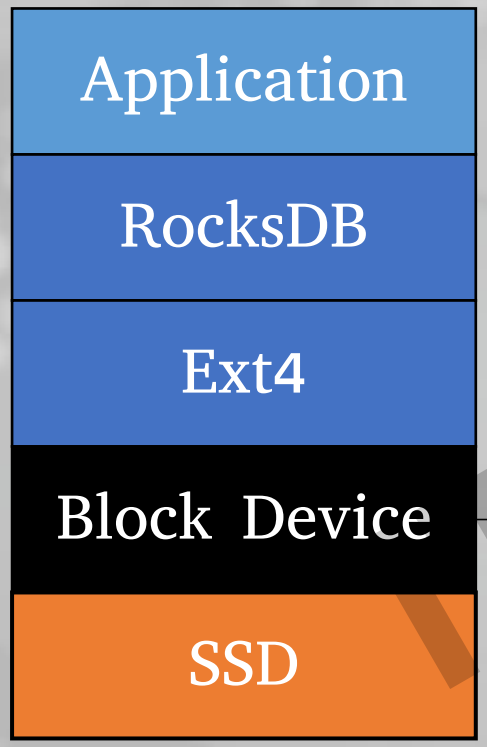


1，Compaction操作导致几倍的写放大.
2，Write Log

Traditional KV Store

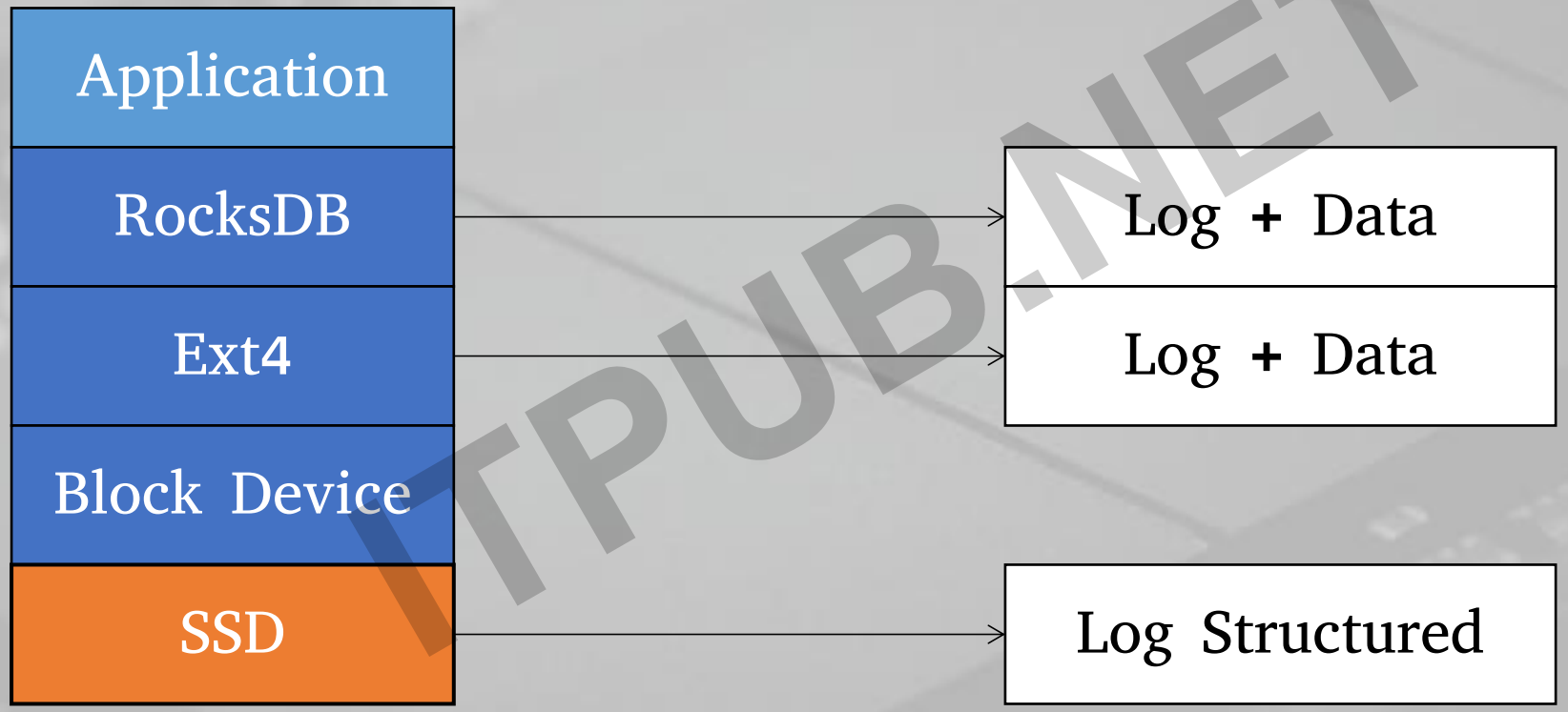


Traditional KV Store

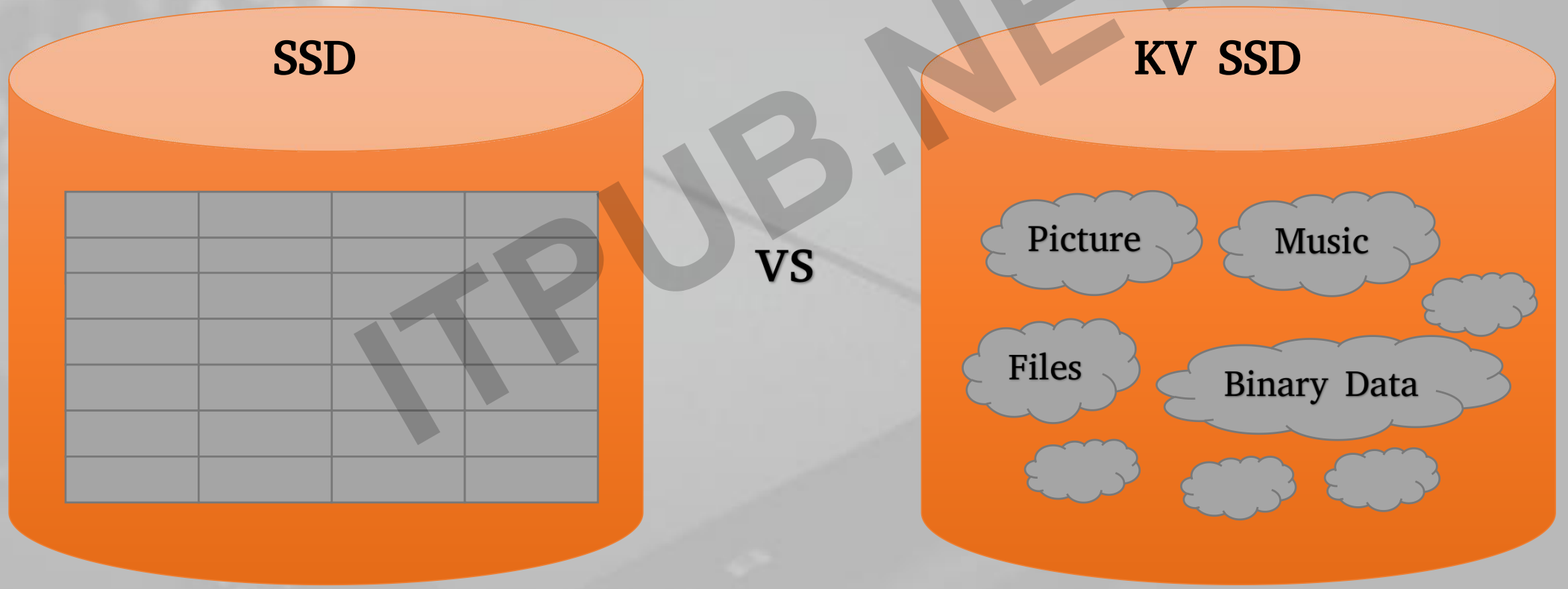


1，Linux Block Device接口是为机械硬盘设计的，并不能发挥出SSD的优势和特性。

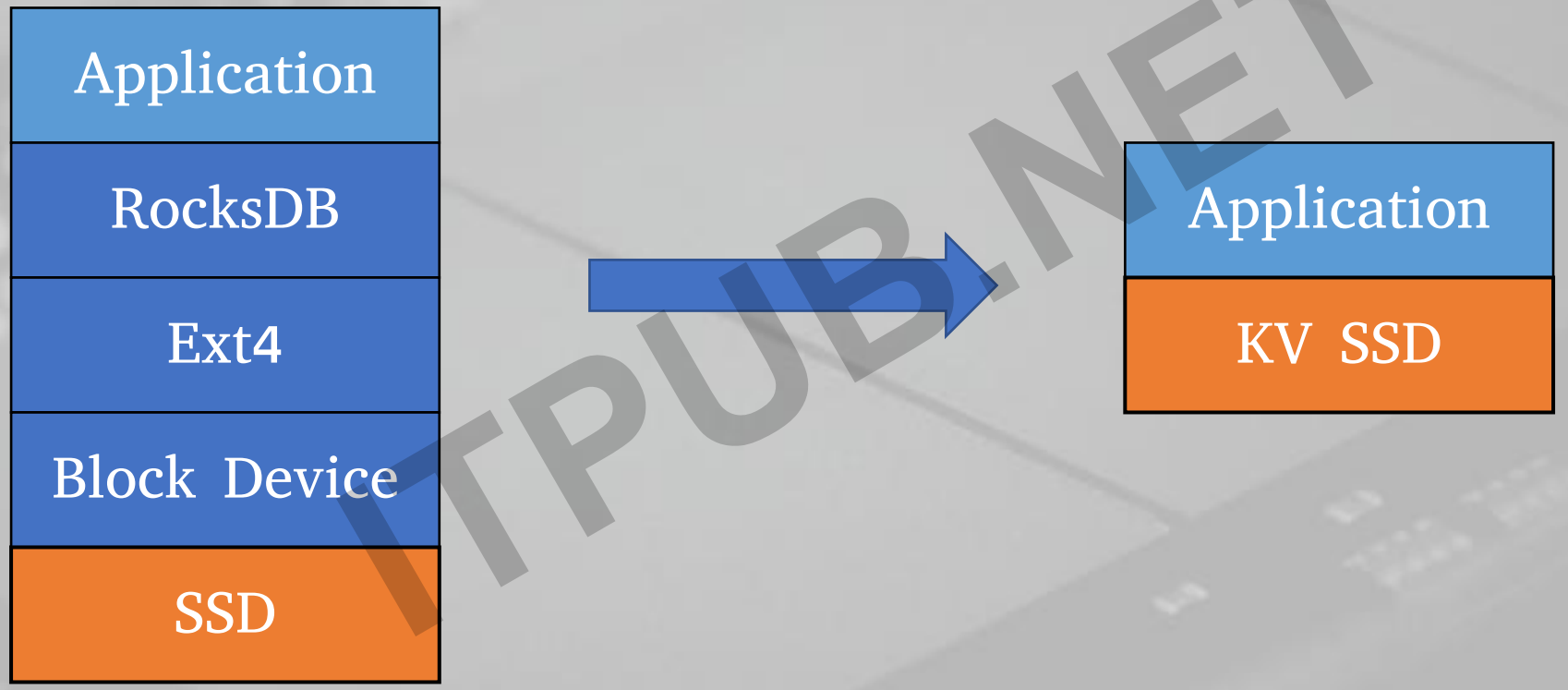
Log on Log



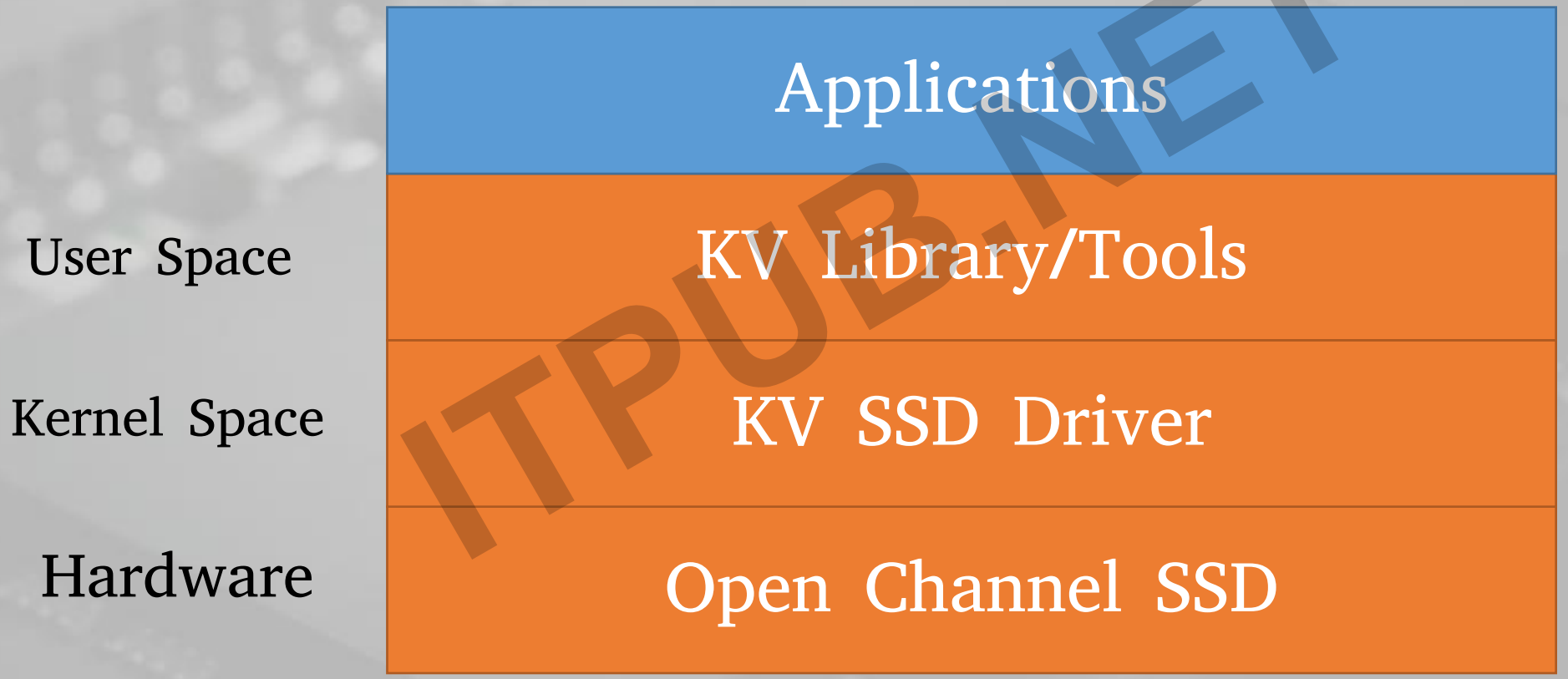
Traditional SSD vs KV SSD



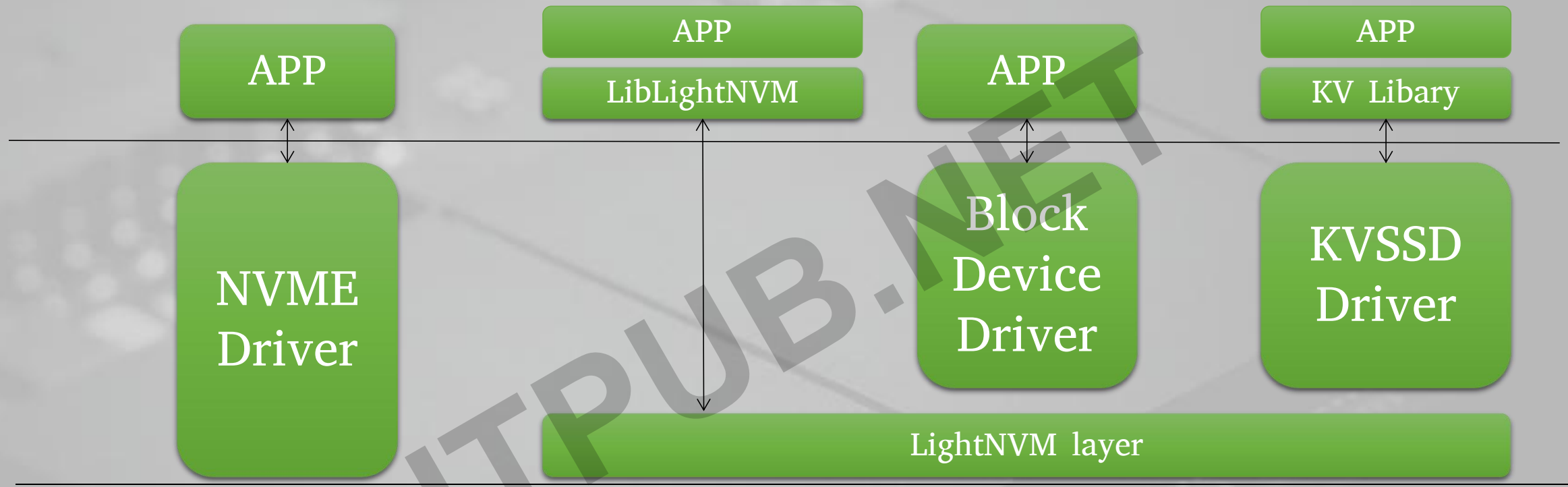
New Software Architecture



KV SSD Software Stack



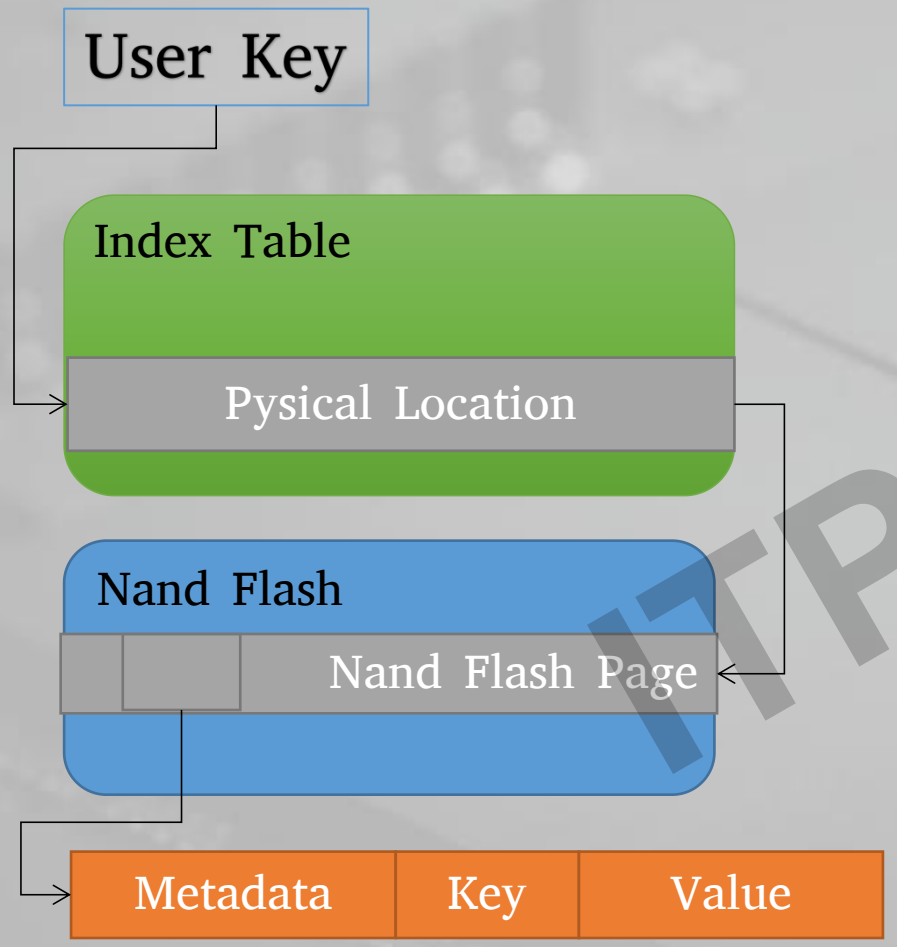
Open Channel SSD

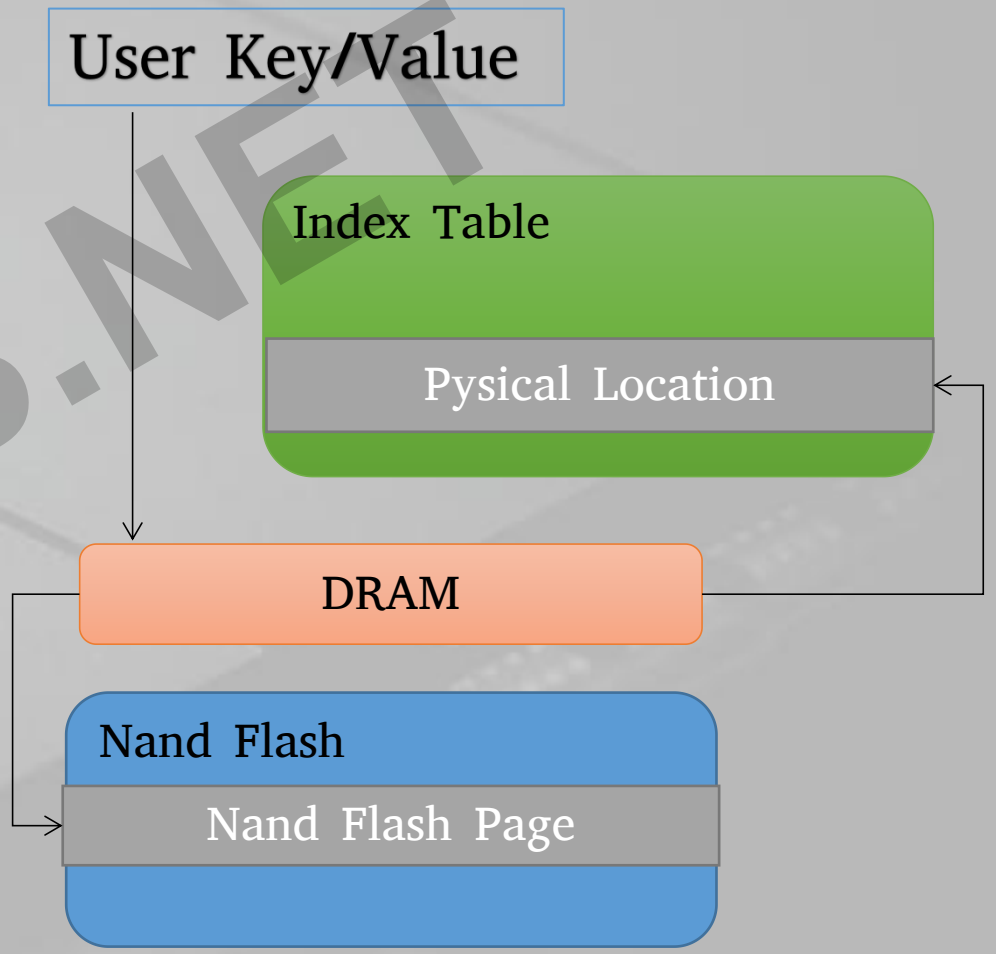


KVSSD Driver

- key2addr index table
- Log structured
- GC
- WL
- Compression

KVSSD Driver Read/Write Flow





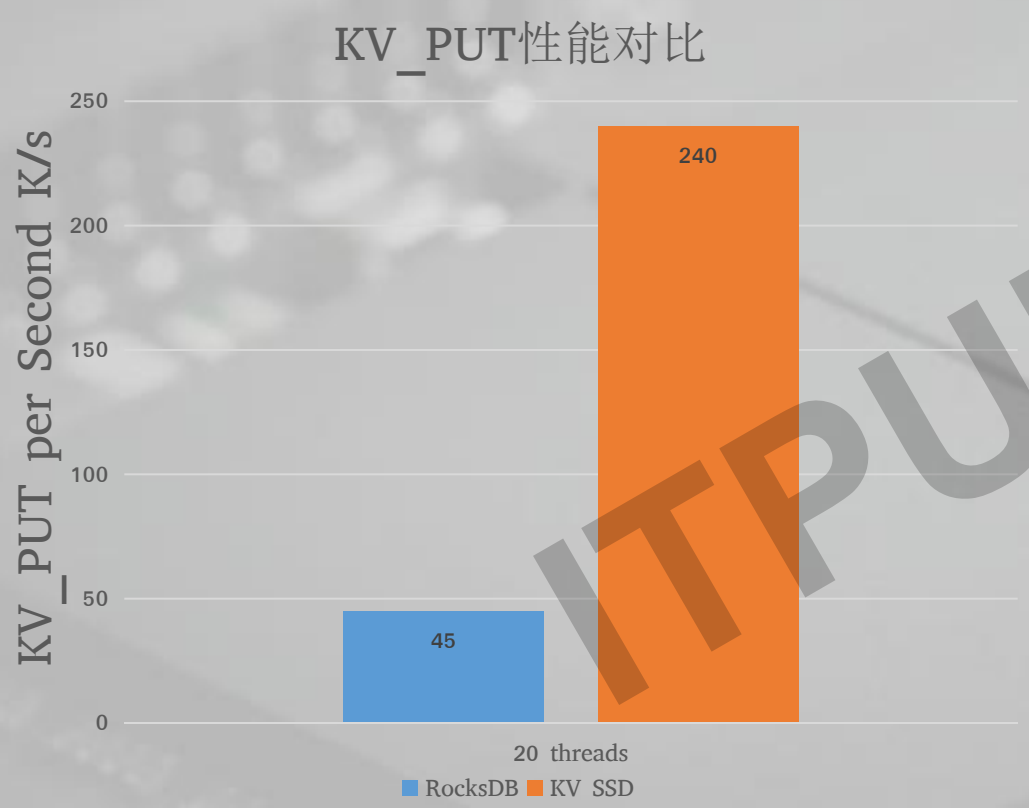
KV Library API

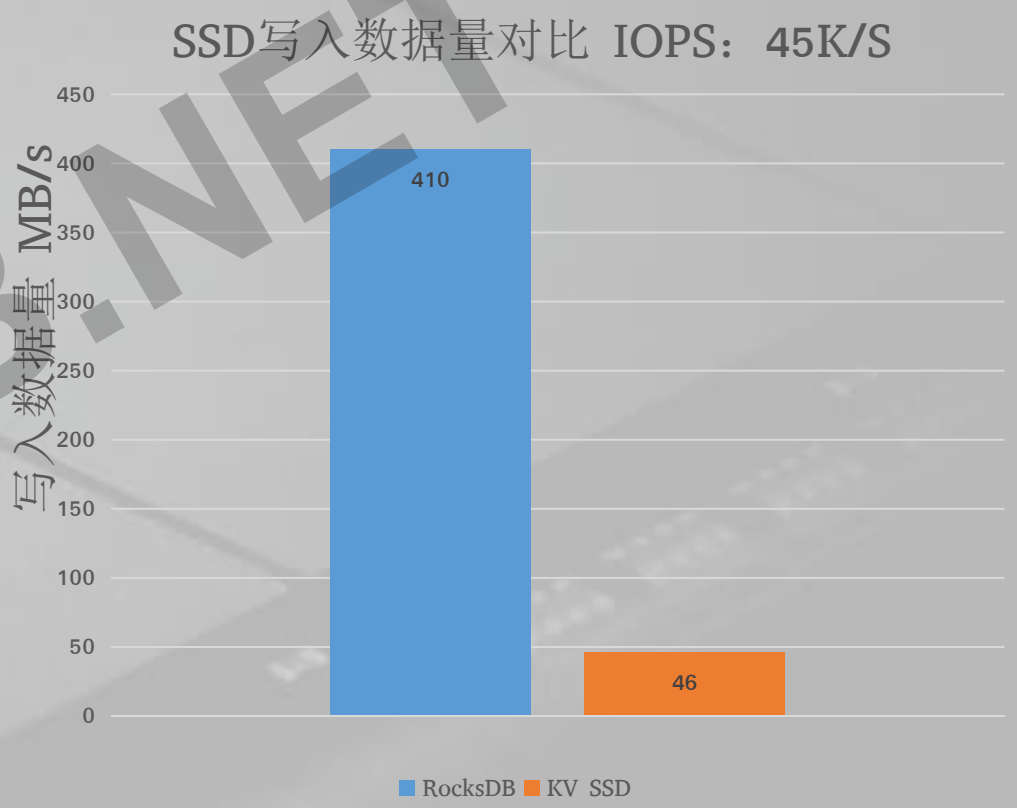
- open_db/close_db
- kv_put/kv_get/kv_delete
- iterator
- write_batch/read_batch
- snapshot
- Column Family
- Read Cache

KVSSD Overview

- KEY_SIZE: 1B ~127B
- VALUE_SIZE: 1B ~ 40MB (1KB+ better)
- Persistency
- Atomic Write

KVSSD和RocksDB的随机写性能对比





测试条件： key_size:8，value_size:1KB，WriteOption.sync=true

Use Cases

mongoDB. ceph Pika redis MyRocks

RocksDB

Q&A

- 我的微信号: stanleymiao

上海宝存信息科技有限公司
Shannon Systems

上海市杨浦区荆州路168号安联大厦9楼
9F Anlian Building, 168 Jingzhou Road, Yangpu, Shanghai
021-5558-0181
contact@shannon-sys.com
www.shannon-sys.com