

重构数智缓存底座 助力金融强国建设

Sudis——金融级缓存数据库，引领数据智能

北京内核科技有限公司 CMO 唐欣

2026.07.17

2026



目录

CONTENTS

01
公司介绍

02
趋势分析

03
Sudis解决方案

04
行业实践

05
总结与愿景

01

公司介绍

专注分布式缓存数据库

公司介绍



内核科技

专注成就专业

成立背景: 2020年创立, 源于对高性能、高可靠缓存技术的深刻洞察与市场需求。

核心团队: 汇聚来自北大、北航、北邮等顶尖学府的资深专家, 在分布式系统、操作系统内核、数据库领域拥有深厚积累。

行业定位: 专注于打造新一代的金融级缓存数据库, 解决行业痛点。

权威认可

参与标准制定

行业地位: 北京信息技术应用创新工作委员会理事单位, 深度参与行业生态建设。

标准贡献: 作为核心成员, 参与起草分布式数据缓存中间件团体标准。

愿景使命: 致力于成为领先的智能数据基础设施提供商。

产品Sudis

自主可控基石

知识产权: 拥有完全自主知识产权, 代码自主率高达99.7%。

权威认证: 已通过工信部电子第五研究所、中国人民银行金融信创生态实验室、信通院可信中间件的测试认证。

安全承诺: 从源头保障金融信息安全, 无任何后门或协议风险。

02

趋势分析

分布式缓存在AI时代的新需求

金融AI时代缓存已成为AI的“最近数据层”



金融行业正从“系统自动化”走向“智能实时决策”

典型场景

实时风控 / 反欺诈
智能推荐 / 定价
大模型客服 / 投研

过去

缓存 = 提速组件

现在

缓存 = AI 实时决策的数据入口



AI 算力越来越强，但数据是否足够快、足够稳？

缓存决定能不能“实时”

金融缓存正在发生变化



量级+架构的变化

从“临时数据存放” →
AI 实时特征与状态的核心承载层

1

规模变化

分片几百节点 → 几千节点
集群规模指数级增长，管理复杂度飙升

2

数据变化

单个对象越来越大
单节点需稳定支撑 20 GB/s 级热点带宽

3

角色变化

从“用完即丢”
到“推理 + 训练的数据闭环”。

现实冲突：传统缓存架构全面承压

Gossip 在大集群下的“信息风暴”



当集群规模扩展到数百甚至上千节点时，基于Gossip的节点间通信导致网络带宽被元数据同步耗尽，集群规模上不去，脑裂风险增加

单线程模型遭遇热点 Key 与大 Key



大Key（用户画像、大JSON）单次读写耗时秒级，阻塞整个请求队列；
热点Key（秒杀、爆款理财）单核CPU瞬时被打满，节点失稳

多版本混用，高危漏洞频出，运维复杂度失控



业务线众多，缓存版本不一，配置千差万别，故障定位难，版本升级与漏洞补丁风险极高

XC环境性能难以释放



在海光、鲲鹏、飞腾芯片及麒麟、UOS操作系统上，NUMA感知缺失，内存访问跨节点，延迟飙升，算力无法充分发挥



03

Sudis解决方案

为金融级而生的重构

性能革命：多核并行与硬件协同

打破单线程瓶颈，释放XC算力



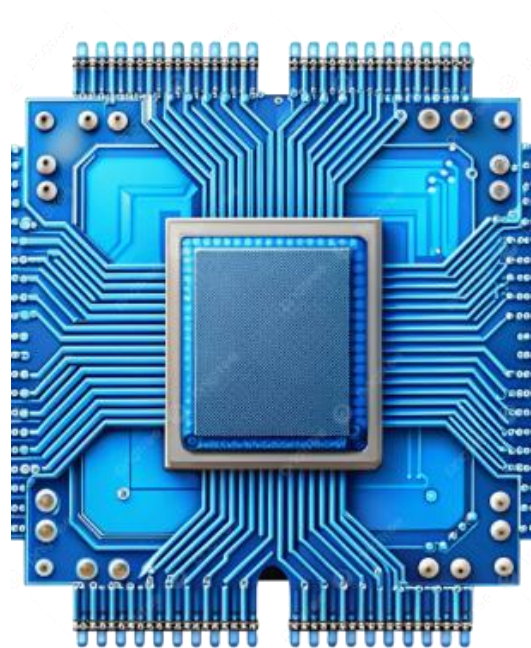
多核多线程架构设计：彻底解决 Redis 单线程导致的单核阻塞问题，实现网络 I/O、命令处理、数据存取全流程并行。



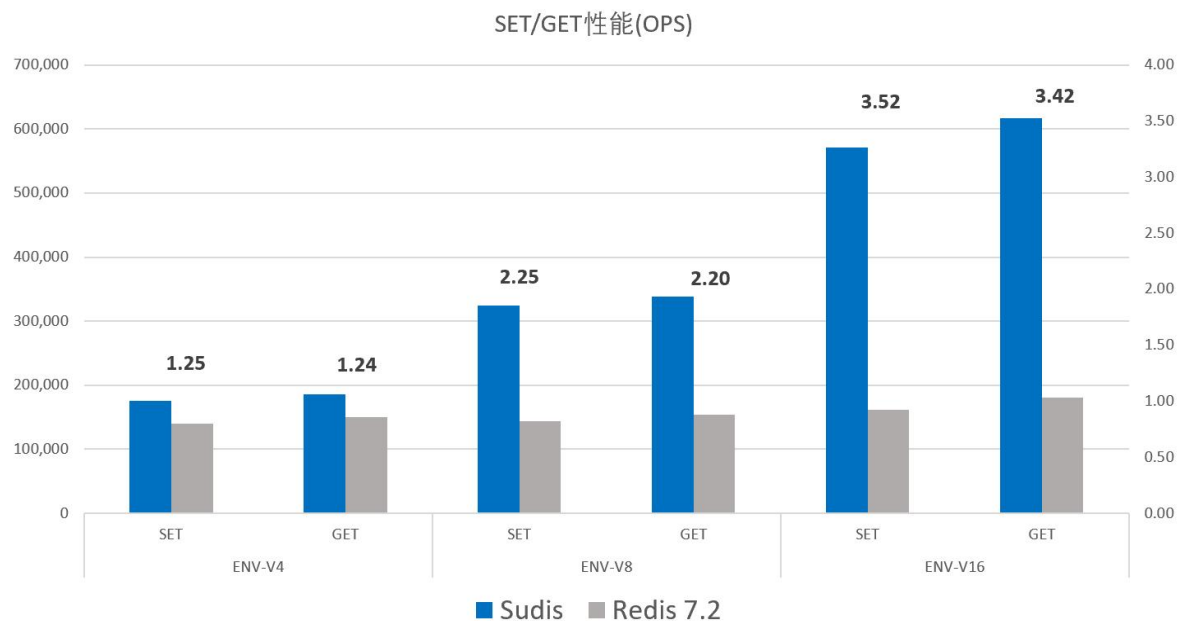
NUMA 感知优化：数据在海光、鲲鹏芯片内部“就近访问”，避开跨节点损耗，XC性能爆发。



底层数据结构和算法重构：单节点吞吐量较开源版本倍数提升，轻松应对金融级 QPS 洪峰。

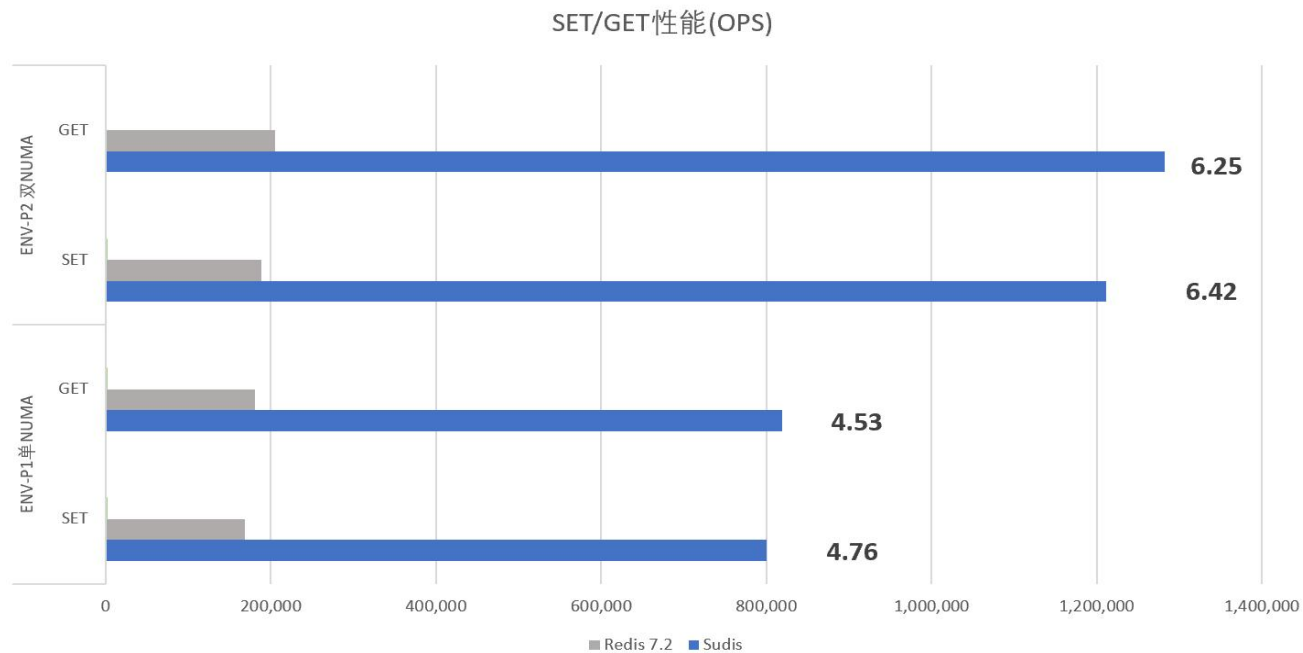


实测海光服务器虚拟机性能对比：随核心增长的线性战力



在4C/8C/16C虚拟机实测中，
随核数增加表现出显著的性能线性扩展优势。

实测吞吐：海光服务器物理机上性能对比



数据说明：在海光7490服务器实测中，在双NUMA64核算力下，吞吐量提升到6.x倍，彻底释放算力潜能。

金融级 SLA 对标：极致稳态响应

场景指标	Sudis V4.3	Redis 7.2	对比结论
单 NUMA P99 延迟	3.62 ms	15.87 ms	延迟降低 77%
Set Key 处理能力	80万 OPS	16.8万 OPS	4.76 倍 性能优势
同城双活容灾切换	5 秒内	> 30 秒	完美满足 RTO 要求

开发语言	Redis总耗时	Sudis总耗时	Sudis/Redis 性能倍率
C++	1884 ms	295 ms	6.38
Go	2300 ms	440 ms	5.23
Python	4030 ms	670 ms	6.01

真实业务场景：总计写入 12800 个 Key，每个 Key 包含 300 个 Field，单 Field 数据长度 761 字节。

稳若磐石：根治大 Key 与热 Key 顽疾



单 Key 并行计算

将单次重量级请求分拆多线程并行处理，结果集自动合并，将单点压力分散至全核心。



超大 Key 自动分片

针对 GB 级大 Key 自动执行底层分片存储，消除单 Key 读写导致的整个请求队列阻塞。



热点 Key 吞吐翻倍

利用多线程并发技术，热点 Key 吞吐量实测提升 3 倍以上，确保秒杀场景下系统不抖动。

3x+

16C 热点 Key 吞吐提升

金融业务实测：

在某行理财秒杀场景中，Sudis 成功承载3倍峰值流量，单 Key 延迟始终保持在毫秒级，完美解决生产中断风险。

极致运维：金融级业务连续性保障



零感知升级，90% 效率提升

🔄 在线滚动升级：支持版本热切换，业务全程零感知，补丁修复与内核升级无需停机。

🖥️ 图形化控制台：统一管理、实时观测、一键扩容，替代繁琐命令行，运维效率提升 90%。

🛡️ 高可用机制：增强的故障感知与自动切换，综合心跳、延迟等多维指标，全链路国密 SM2/3/4 支持，满足最严苛XC合规。

降本增效：成本优化与安全合规



TCO 降低 33-66%

1主1从极简架构设计，大幅降低硬件采购与资源冗余成本，实现商务投资收益最大化。



冷热数据分层

智能冷热数据调度，利用高性能 NVMe 支撑海量数据，大幅精简昂贵的 DRAM 内存占用。



全栈XC合规

全面适配国密标准，具备完全自主知识产权，代码自主率 99.7%，消除供应链安全风险。

66%

硬件总体拥有成本下降

基于架构重构实现极致资源利用率与安全能力的双向闭环

04

行业实践

金融行业成功案例

银行业客户案例：核心业务高可用

某股份制银行 (统一消息平台)

▲ 挑战

原缓存服务因1.8GB超大List Key导致生产系统频繁中断。

🔗 方案

迁移至Sudis，利用其大Key无感限制与多线程能力。

🏆 成果

交易峰值 OPS提升300%，半年实现 0故障运行。

某农商行 (支付清算系统)

▲ 挑战

双集群运行下，业务无感切换，数据实时同步，架构可运维。

🔗 方案

采用Sudis-Proxy+ redis / Sudis双集群主备模式。

🏆 成果

故障感知自动切换，同时支持一键切换主备角色，无需修改任何业务代码，无需重建同步链路。

某全国性银行 (同城双活容灾)

▲ 挑战

核心系统要求RTO < 30s，传统异步方案无法保证零丢失。

🔗 方案

部署Sudis双活架构，配合增强数据一致性同步机制。

🏆 成果

实战演练中在5秒内完成切换，完美满足SLA要求。

证券及保险案例：集群管理与降本

某头部券商 (行情与智能客服)

▲ 挑战

数百个小集群升级耗时长（达8个月），维护压力巨大。

🔧 方案

利用Sudis滚动升级能力与统一图形化管理平台。

📈 成果

运维人力 **节约90%**，升级周期缩短至 **4-6周**。

大型保险集团 (统一缓存平台)

▲ 挑战

原缓存集群超2000节点，硬件成本与管理复杂度极高。

🔧 方案

构建基于Sudis的高性能统一缓存，提升单机资源利用率。

📈 成果

节点精简至 **1000个**，年省云资源成本 **500万元**。

金融控股集团 (多租户统一平台)

▲ 挑战

需统一服务多子公司，对多租户隔离、配额管理要求严。

🔧 方案

通过Sudis命名空间实现多租户逻辑隔离与ACL安全访问。

📈 成果

资源利用率 **提升40%**，实现了多租户TCO精细化管理。

05

总结与愿景

共建数字金融强国缓存底座

总结与愿景 - 共建数字金融强国缓存底座

安全可控

- 99.7%自主代码
- 无后门
- 国密支持
- 主动安全。

性能领先

- 1.x-6.x于Redis的性能
- 根治大Key/热Key问题。

成本优化

- 节点数减少33%-67%
- 硬件TCO下降40%+。

零风险切换: 完全兼容Redis协议, 业务代码零修改。

01

平滑迁移
路径

02

专业护航: 提供SudisSync双向同步、灰度切流、秒级回滚及全流程专家服务。

清晰路线: POC (2周) → 非核心双跑 (1个月) → 核心灰度 (1个月) → 全量上线。

2026

感谢！

携手共建金融级XC缓存底座！

