

Gdevops

Global DevOps Summit

全球敏捷运维峰会





ABC场景驱动下，如何制定数据库与存储技术规划

vivo 肖博



肖 博

vivo 通用存储研发组总监

职业经历

百度数据库团队

同程艺龙数据库团队

技术分享

vivo大规模Redis集群实践

基于 Redis Cluster 的灾备与同城双活实践

大规模多存储场景的数据库选型与服务平台建设之路



目录

- 数据存储技术的本质
- 数据存储技术的现状
- 数据存储技术发展规划

追本溯源



甲骨文

存储介质：龟壳

特点：容量极小、读写效率低



竹简

存储介质：竹子

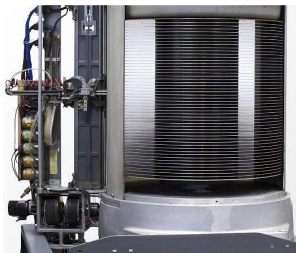
特点：容量较小、读写效率较低



书本

存储介质：纸张

特点：容量较大，读写效率较高



磁盘

存储介质：磁介质

特点：容量大，读写效率较高

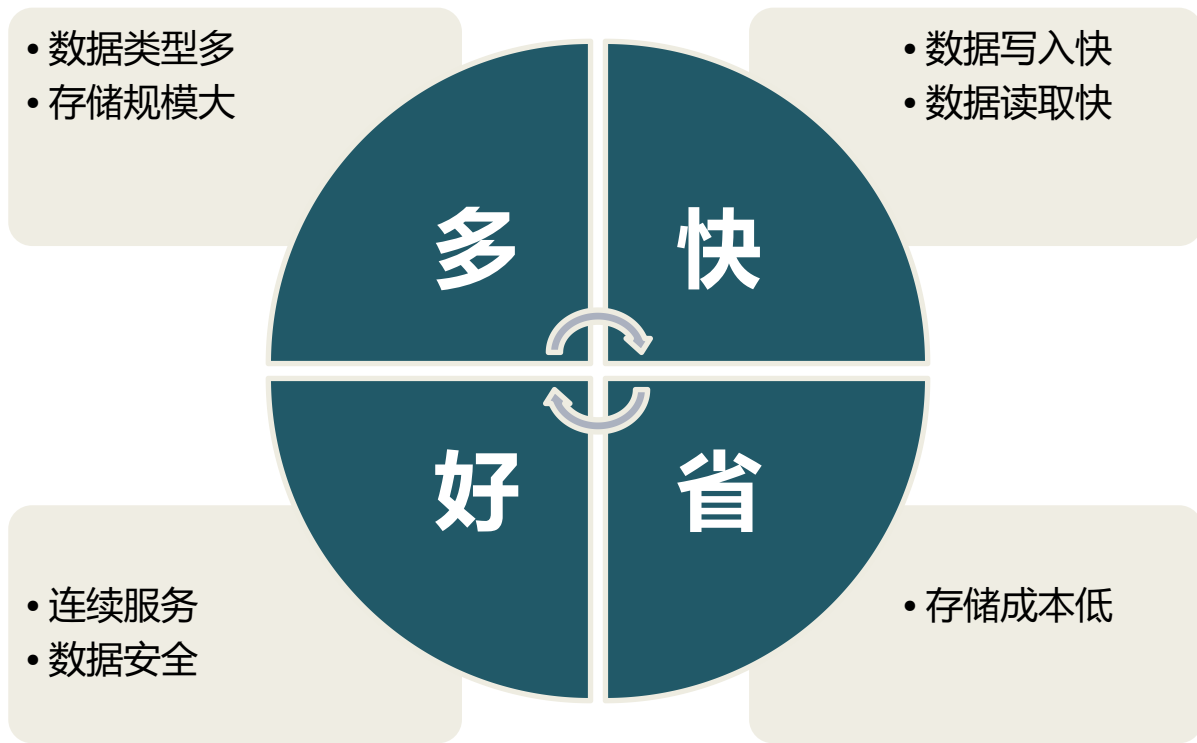


存储介质：NAND

特点：容量大，读写效率高



数据存储的诉求





数据存储的本质

数据存储

数据读取

存储规模



目录

- 数据存储技术的本质
- 数据存储技术的现状
- 数据存储技术发展规划

行业头部在做什么

- ◎ 自研数据库/存储产品
- ◎ 挑战金融场景
- ◎ 选择部分或者全部开源
- ◎ 弹性伸缩、存算分离等分布式特性
- ◎ 智能优化、智能运维
- ◎ 统一管控，HTPA多模等特性
- ◎ 社区运营、生态建设、认证培训

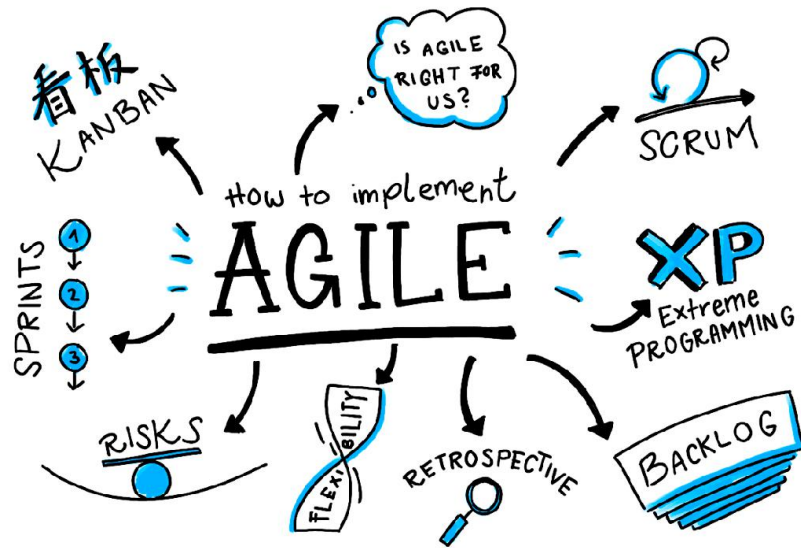


敏捷交付

对数据库以及存储服务而言，在软件研发过程中能够快速
的交付满足业务需要的存储服务和数据库

敏捷变更

对数据库以及存储服务而言，在服务运行和运维过程中的
各种变更能够足够的快捷，足够的轻量。



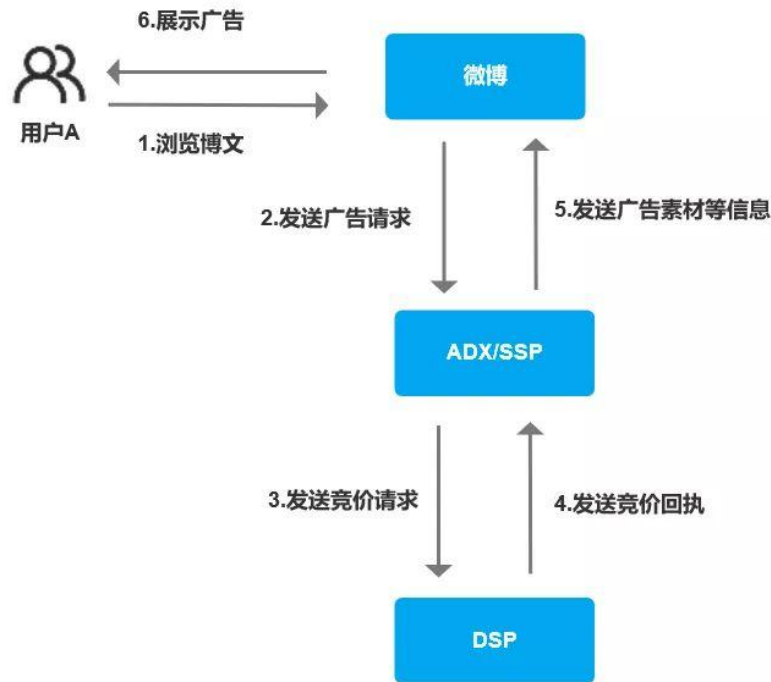
更高的性能

用户体验

内容与场景同质化的情况下，各软件比拼的更多是用户体验，而大多数软件或者应用都是围绕着数据展开的，存储服务与数据库产品的性能显的尤为重要

公司收入

在广告服务、推荐场景、自然语言处理等场景中整体服务的性能都会受到存储和数据库服务的性能影响



更大规模

数据规模

移动互联网、工业互联网等产生大量的数据，各种IOT设备源源不断的产生数据，参与数据生产加工的人群越来越多，数据规模的不断增加，分布式架构已经成为常态

数据类型

结构化、半结构化、非结构化的数据类型以及AI、大数据等场景对数据产生的需求，支撑各类数据存储的产品百花齐放

数据类型

数据分层表现明显，需要实时处理的数据占小部分，有大量的冷数据需要处理，归档存储和冷存储的成本越来越高



更低的存储成本

资源利用率

购买的服务器资源有没有被充分的利用，服务器的CPU、内存、磁盘利用率如何，资源利用率的考核成为企业的常态，数据库和存储服务的虚拟化、容器化、资源混合部署等技术逐渐成熟

成本量化

单个业务或者产品的ROI需要把用到的所有资源进行量化，对存储和数据库服务同样需要，账单服务的诉求也越来越强烈



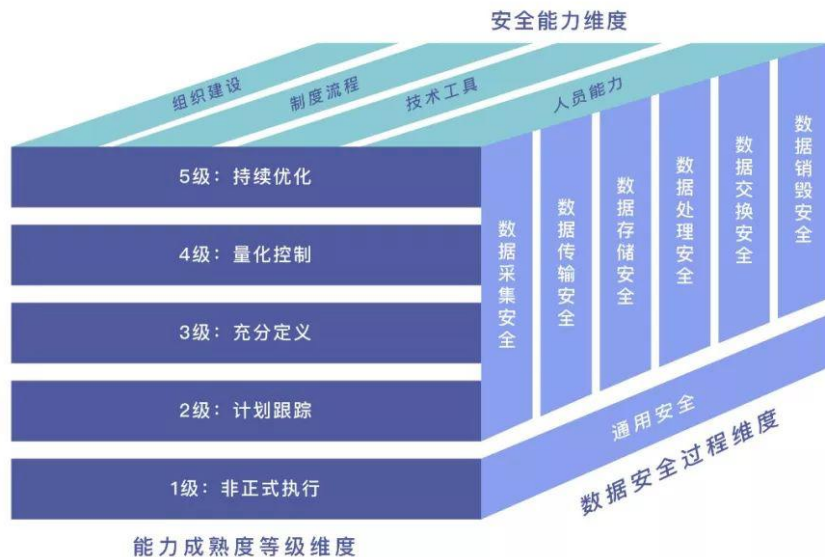
更多的安全诉求

安全意识

全球数据泄露事件频发，敏感数据通过网络进行扩散进而产生不良影响，用户目前对个人隐私数据保护的意识越来越强烈

合规要求

各个国家和地区密集出台数据安全的条例和法规，安全合规已经影响到企业能否正常开展业务



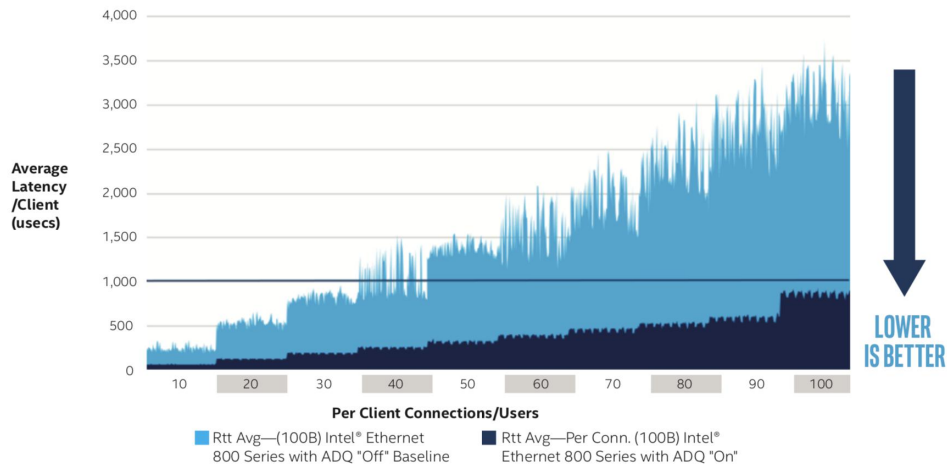
内核旁路

内核旁路在解决软件性能问题中越来越多的被使用, dpdk、spdk等将技术日益成熟并被应用

专用硬件

专有的硬件设备从各个方面来提升软件性能, 和硬件设备结合能更大性能的提升软件性能, 如AEP, DPU, 智能网卡等等

Predictability Results: More Than 50 Percent Increased³





交叉融合

一体化

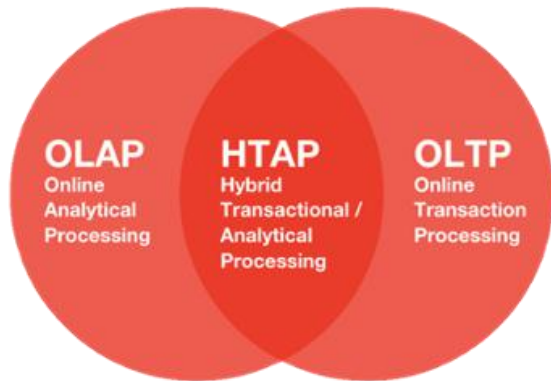
流批一体、湖仓一体、HTAP等可以减少数据流动、降低成本、

提升效率的方案和概念不断被挖掘和实现

多模存储

将多个系统组合使用的解决方案下沉为数据库内置能力，HTAP、

多模数据库等产品不断涌现

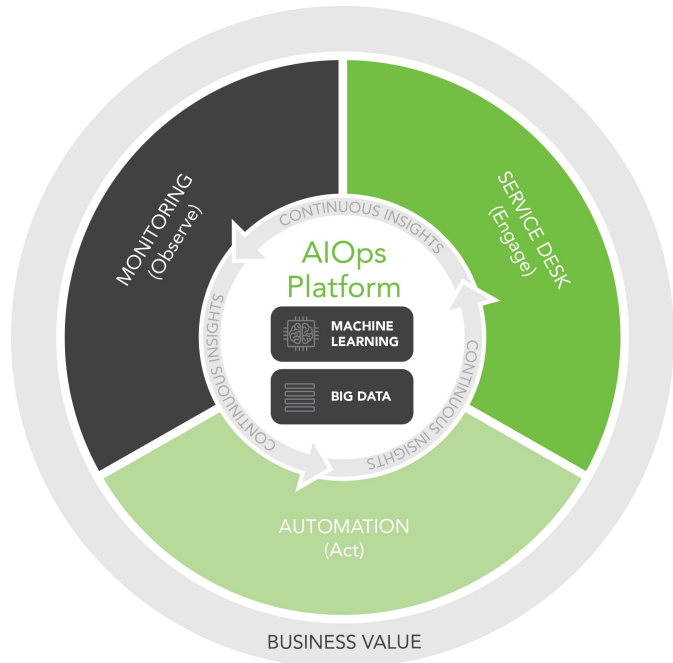


智能运维

基于已有的运维数据，通过机器学习的方式将运维动作智能化，进一步解决因业务扩张，高人力成本难以维系的问题

智能优化

将AI技术应用于索引优化，智能索引推荐技术已经在某些数据库产品中有落地，数据库参数智能优化也在不断探索



Sourced from Gartner, 2016



目录

- 数据存储技术的本质
- 数据存储技术的现状
- 数据存储技术发展规划

主见

新的概念和产品层出不穷，然而可能目前企业的现状还在使用头部企业N年前的技术，容易产生焦虑感和沮丧感。但任何新技术和产品都是为了解决某种场景的问题而诞生的，解决当前企业面临的问题用适合的技术就行

重点

可靠性，可用性，安全，成本这些永远是重点，其中对于数据库和存储而言前三点的重要程度是高于成本的，其次我们大概投入20%的时间和人力能解决80%的问题，剩下的20%问题要慎重评估后在确认是否投入



系统思维

系统思维是一种从整体、全局出发看问题的思维方式。

- 拆解，把系统问题拆解成各个元素
- 放大，把拆解后的元素放到整个系统里面来研究
- 控制，对每个元素单独控制



警惕AIOPS的陷阱

AIOPS定义

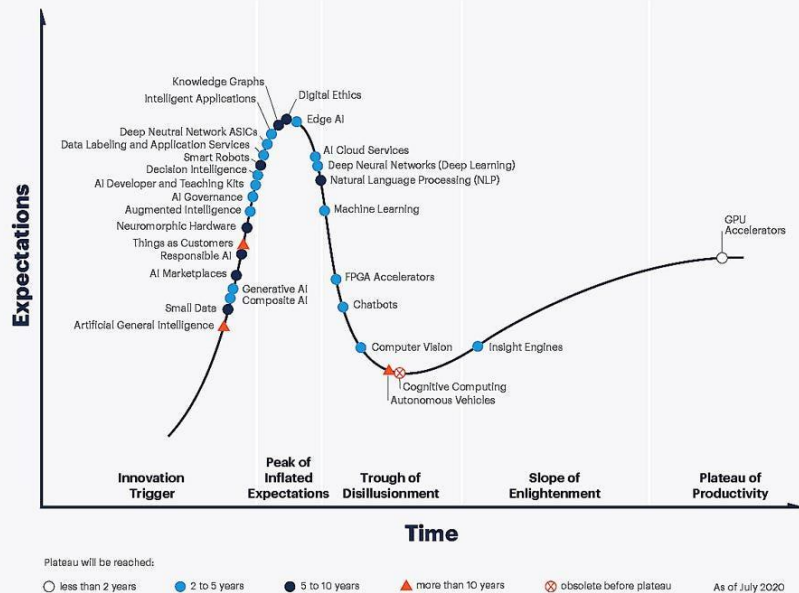
AIOPS是一种多层次的技术平台，包括如下两点内容

- 使用机器学习来分析IT运营系统的各类业务与系统数据，从而实现IT运营的自动化增强
- 能够实时的自动发现系统存在的问题且能自动实现故障自愈

AIOPS陷阱

- AIOPS \neq AI + OPS
- 对现阶段AIOPS发展水平认知陷阱，AIOPS目前正处于科技诞生的促动期进入过高期望的峰值的发展时期，而对其抱有超越现阶段的期望
- 技术选择上的陷阱，认为直接使用现成的深度学习或者机器学习算法可以很好的做出决策

Hype Cycle for Artificial Intelligence, 2020



gartner.com/SmarterWithGartner

Sources: Gartner
© 2020 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner and Hype Cycle are registered trademarks of Gartner, Inc. and its affiliates in the U.S.

Gartner.



数据库与存储平台规划

- 混合云统一数据存储管理解决方案
- 数据存储DevOps解决方案
- 数据变更稳定性解决方案
- 数据安全解决方案
- 数据存储产品可用性解决方案



数据库与存储产品规划

- 丰富数据产品矩阵与生态
- 培养数据存储产品研发能力
- 降本增效的需求一直在
- 安全的需求变的越来越重要
- 兼顾业务需求与运维需求



规划参考

vivo

数据存储管理服务平台

可用性

数据备份	服务监控	服务巡检	预案管理
高可用	故障管理	灾备管理	数据校验
故障自愈	数据闪回	从0恢复	服务熔断
服务降级	服务限流	智能诊断	数据校验

效率

账户管理	服务部署	数据变更
运营报表	数据查询	服务变更
慢日志管理	多环境管理	自动化服务
数据订阅	事件管理	智能客服

成本

预算管理	容量评估
资源分账	容量预测
弹性扩展	利用率管理
资源回收	资产管理

安全

安全审计	数据脱敏
数据防火墙	权限管控
数据治理	数据加密
GDPR合规	日志分析

MySQL

DRDBMS

RDS

RDBMS

OLAP

ElasticSearch

MongoDB

TSDB

GDB

NoSQL

TiDB

KV Storage

Table Store

Big Table

对象存储

通用存储服务

文件存储

块存储

IDC资源

公有云资源

元数据管理



全球敏捷运维峰会

Thank You!

