5G时代运营商数据和存储架构白皮书

HUAWEI

目录

CATALOGUE

1 5G 时代运营商的机会和 IT 基础设施的挑战 03-03

1.1 5G 进入规模部署阶段，带来无限可能 ---------- 03

1.2 5G 时代运营商 IT 基础设施的挑战 ---------- 03

2 5G 时代，运营商 IT 系统数据量将激增，数据存储效率要求更高 04-05

2.1 生产域数据增量和访问模式变化 ---------- 04

2.1.1 计费系统——话单数据量增长 7.5 倍，计费时延要求更低 ---------- 04

2.1.2 2C 业务——视频高清化带来数据 10 倍增长，数据读写带宽要求更大 ---------- 04

2.1.3 OSS 运维——高密度网元和自动化运维使数据量暴增 8 倍，运维效率要求更高 ---------- 05

2.2 分析域数据增量情况——大数据分析带来 5 倍数据增长，分析效率要求更高 ---------- 05

2.3 运营商总数据量增长 ---------- 05

3 5G 时代数据海量增加，需要重定义数据基础设施 06-09

3.1 5G 数据洪流对基础设施带来的挑战 ---------- 06

3.1.1 存不下——海量数据存储 ---------- 06

3.1.2 流不动——数据孤岛难题 ---------- 07

3.1.3 用不好——价值获取低效 ---------- 07

3.2 挑战应对方法 ---------- 08

3.2.1 改变 Silo 模式，重定义数据架构 ---------- 08

3.2.2 借力闪存升级，重定义数据性能和可靠性 ---------- 08

3.2.3 软硬双管齐下，重定义数据密度 ---------- 09

3.2.4 打破数据壁垒，重定义数据价值 ---------- 09

4 5G 时代运营商数据基础设施 10-20

4.1 5G 时代运营商的数据基础设施目标架构 ---------- 10

4.1.1 华为数据基础设施特性 ---------- 11

4.1.2 华为数据基础设施应对 5G 时代运营商挑战 ---------- 13

4.2 重定义数据架构——池—湖解决方案 ---------- 13

4.2.1 一池：面向生产域的融合资源池 ---------- 13

4.2.2 一湖：面向分析域的智能数据湖 ---------- 15

4.3 重定义性能和可靠性—全闪存存储 OceanStor Dorado ---------- 15

4.3.1 重定义性能：业界领先的性能，2,000 万 IOPS；0.1ms 时延 ---------- 15

4.3.2 重定义可靠性：高端存储可靠性新标杆 ---------- 17

4.4 重定义数据密度—业界超高密分布式存储标杆 ---------- 18

4.4.1 超高密度存储节点 ---------- 18

4.4.2 超高存储利用率 ---------- 18

4.4.3 智能散热，单 TB 功耗最优 ---------- 18

4.5 重定义数据价值—FusionData 解决方案，打破数据孤岛，释放数据价值 ---------- 19

4.5.1 存算分离实现计算存储资源按需灵活配置，整体 TCO 下降 30% ---------- 19

4.5.2 融合分析提升分析效率提升 50 倍 ---------- 19

4.5.3 数据虚拟化提升开发效率 3.5 倍 ---------- 20

4.6 智能数据，面向未来 ---------- 20

5 总结 22-22

1 5G 时代运营商的机会和 IT 基础设施的挑战

1.1 5G 进入规模部署阶段，带来无限可能

2019 年，全球 5G 建设已经如火如荼。全球几十家 Top 国家和国际化电信运营商纷纷宣布 5G 商用，其中包括 Vodafone，T-Mobile，TIM，OOredoo 等，在国内，2019 年 10 月 31 日，中国移动，中国电信和中国联通三大运营商公布 5G 商用套餐，并于 11 月 1 日正式商用，运营商用实际行动向世人宣告 5G 时代的到来。

5G 应用的主体不仅仅在消费市场，同时更在行业市场，其发展潜力是运营商投资未来的信心。5G 可以给运营商带来:

- 5G给消费者带来更多的业务和更好的体验。5G的高带宽使消费者更好地使用视频通信，使VR、AR等新业务成为可能，同时 5G 的低时延也给消费者带来了更好地的业务体验。5G 也激发了更多的业务创新，给消费者带来更多更好的业务。
- 5G 可以使能垂直行业。5G 的到来正让万物互联变得可能，5G 可以使能更多的垂直行业，如行业视频监控，车联网，智能制造，远程医疗等，更多的行业使能给运营商带来了新的业务增长点。
- 5G 可以带来新的商务模式转变。运营商可以通过 5G 带来新的业务和使能垂直行业的方式，通过创新的商业模式，有效改变运营商当前收入结构单一的问题。

1.2 5G 时代运营商 IT 基础设施的挑战

由于5G技术的发展和5G带来新业务的发展，运营商的数据量将爆发增长，运营商的传统IT基础设施将面临诸多挑战。

- 5G 的“高带宽”和“低时延”特性使得单位时间内产生的数据量急剧增长，海量原始数据将被采集。
- 5G 的“低时延”特性要求 IT 基础设施具有更好性能，更低的访问时延，从而满足客户极致体验的需求。
- 5G 的“全连接覆盖”特性使得万物互联成为可能。相比 4G 时代主要用于人与人之间的互联，5G 时代物联网将得到较大程度的发展，人与物、物与物之间的连接将急剧增多，数据采集渠道讲更加丰富，如联网汽车、可穿戴设备、机器人等，这也对数据存储与采集技术提出更高要求。
- 5G 时代下运营商的超高清业务、VR/AR 业务相比传统的高清业务需要更多的存储空间和更高的并发读写能力。
- 在 5G 时代下，运营商需要对现有的的 IT 基础设施进行升级来满足日益增长的业务诉求。

2 5G 时代，运营商 IT 系统数据量将激增，数据存储效率要求更高

运营商的 IT 系统和其他的行业相比更加的复杂。运营商的 IT 系统可以分为生产域和分析域。生产域包括 BSS（含 CRM，计费，2C 和 2B 的增值业务等）、OSS 和 MSS，分析域主要是面对生产域的大数据分析。

下面以典型 IT 系统的情况来说明 5G 时代下数据的增量以及数据访问模式的变化。

2.1 生产域数据增量和访问模式变化

2.1.1 计费系统——话单数据量增长 7.5 倍，计费时延要求更低

5G 网络，带宽增加，时延更低，给计费系统带来的变化是每用户的 DoU 增加，导致用户的话单数增加，计费库数据量增加，同时计费时延相比 4G 网络要求更低。对应到数据存储系统，要求 IOPS 更大，访问时延更低。

以百万用户为例，4G 百万用户到 5G 百万用户带来 7.5 倍的数据量增加，同时时延变为原来的 1/8，时延是业务 SLA 的保障关键。

比较项	4G 百万用户	5G 百万用户	改变
DoU 变化	9.8 G	55.9 G	增加 5.7 倍
话单数	1.2 亿	7.5 亿	增加 6.3 倍
每条记录大小	1 K	1.2 K*	增加 1.2 倍
计费库数据量	10.8 TB	81 TB	增加 7.5 倍
时延要求	4 ms	0.5 ms	减少 7/8

* 增加切片信息、QoS 信息导致每条记录大小变大

2.1.2　2C 业务——视频高清化带来数据 10 倍增长，数据读写带宽要求更大

5G 时代视频业务将成为主流，其收入也将成为运营商收入的重要来源之一。视频从高清 1080p 向超高清 4K 和 VR 演进。一个小时 1080p 视频需要约 5 GB 存储容量，4K 则需要 15GB 存储容量，而 VR 则需要 180 GB 存储容量。

以某视频业务规划为例，视频容量从 15 PB 增长到 152 PB，总数据量增长 10 倍以上。对应到数据存储系统，需要支持线性的存储扩展，需要支持更多的并发访问，需要支持更大的读写带宽。

项目	2018	2019	2020	2021	2022	2023
超高清比例	0	1%	5%	15%	25%	35%
视频数（万个）	1700	3450	5200	6950	8700	10450
视频时长（万小时）	314	612	936	1251	1566	1881
容量 (PB)	15	30	49	78	112	152

2.1.3 OSS 运维——高密度网元和自动化运维使数据量暴增 8 倍，运维效率要求更高

部署 5G 网络时需要对热点区域补站，网元密度为原来的 1.25 倍，每网元 KPI 数量是原来的 1.2 倍，因此 OSS 基础数据量是原来的 1.5 倍。

网元数增多，自动化运维成为必然，自动化网规网优需要更多数据。CHR 数据采样频率提高，每套网管数据流量从原来的 100 GB/ 小时提升到 275 GB/ 小时，同时分析从事后分析走向事前分析，保存时间从 7 天增加到 90 天。CHR 的数据量增加 35 倍（2.75 倍数据流量 *12.8 倍保存时间）。

以中国某省的 OSS 系统部署计算，OSS 的数据量将增加 8 倍。自动化运维提升了运维效率，运维效率提升需要数据存储系统提供更好的数据处理效率以支持运维任务的计算处理。

比较项	4G	5G	增加
基站数（万）	20	25	1.25 倍
OSS 基础数据量（PB）	11	16.5	6.3 倍
自动化网规网优 CHR 数据	2.7	95	35 倍
合计（PB）	13.7	111.5	8 倍

2.2 分析域数据增量情况——大数据分析带来 5 倍数据增长，分析效率要求更高

随着运营商 BSS 域、OSS 域、各业务系统以及外部第三方数据源的数据量的增多，参与大数据分析的数据量也随之增加。同时为了更多的挖掘数据潜在的价值，数据保存周期也会增长，比如说从分析的数据从原来的 1 个月增加到 3~6 个月，大数据分析需要保存的数据量会进一步增加。

以某运营商的大数据分析业务预计，2019 年大数据分析的数据量为 35 PB，预计到 2021 年，大数据分析的数据量将增加到 160 PB，数据增加近 5 倍数据量的增长。

传统的大数据分析一般采用通用服务器的方式，计算和存储耦合在一起，导致计算和存储不能独立扩展，存在计算资源或存储资源浪费的情况，分析效率不高。大数据分析需要采用计算和存储分离的方式以提高分析效率。

2.3 运营商总数据量增长

根据 IDC《数字化世界 – 从边缘到核心》的报告，全球运营商数据量预计从 2018 年的 3 ZB 扩展到 2025 年的 15.75 ZB，预计数据量增长超 5 倍。巨大的数据量增量对运营商的数据基础设施提出了巨大的挑战。

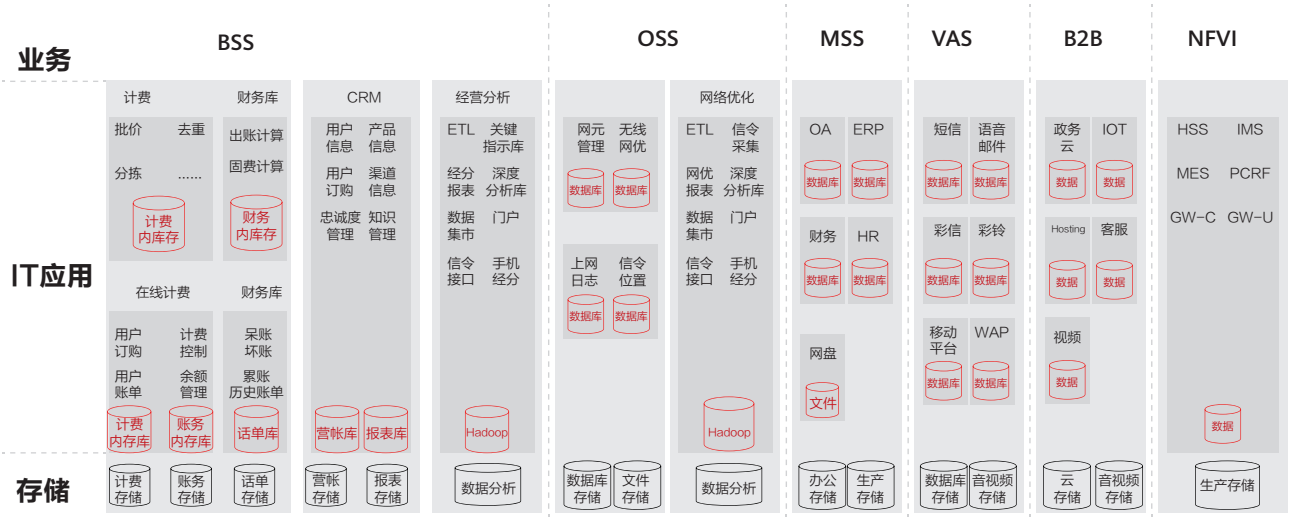


3 5G 时代数据从管理走向运营，需要重定义数据基础设施

3.1 5G 数据洪流对基础设施带来的挑战

海量数据蕴含巨大的价值，也给基础设施带来了前所未有的挑战。目前大多数运营商仍然采用 Silo 式建设，针对不同的业务域，甚至不同域的不同子业务采购不同的数据和存储，以满足不同应用不同客户对数据的各类存算取用需求。

如下为运营商的典型数据架构，经多代历史遗存后，设备多、版本多、垃圾数据多、功能复杂等问题不断放大，数据存不下、流不动、用不好成为了运营商数据应用最普遍的难题。



3.1.1 存不下——海量数据存储

传统架构已无法适应 5G 数据爆发性增长，体现在以下三点：

- 扩展性不足：传统集中式存储本身并不是为海量数据而设计，其扩展性有限，scale-out 能力尤其不足，即使高端存储也很难达到 PB 量级。运营商不得不采购多套存储满足日益增长的需求。
- 协议类型单一：随着视频、物联网等业务扩张，80% 的新增数据由各类音视频、日志等非结构化数据构成，非结构化数据逐步成运营商数据的主体。然而传统存储协议类型单一，无法同时满足块、对象、文件、大数据等多样性数据的存取需求，企业不得不为每一种新的数据类型新增一种存储设备，数据之间也难以互通共享。
- 成本高昂：出于合规等原因，运营商的日志类或大数据分析类数据必须长期保存，因为其存储量很大，也会导致存储成本极高。如 2017 年起，某中国的移动运营商因合规性要求，将日志文件的保存周期从 2 个月增加至 6 个月，这意味着其存储设备规模将增加至少 2 倍。因为该运营商采用 Hadoop 架构，也就意味着必须为此增加 2 倍的服务器。服务器因存储需求不断扩容，CPU 的使用率却始终处于较低的状态，资源得不到合理利用，无疑会对采购成本和维护成本造成更大的压力。企业往往因为存储成本原因而放弃大量宝贵数据。

3.1.2 流不动——数据孤岛难题

孤立的数据价值并不显著，只有当数据像水一样流动起来，才能打破“数据壁垒”，最大化释放其价值。然而当前运营商保存下来的数据，由于技术与流动性问题，只有 10% 的数据能得到分析，数据孤岛、多样性设备、业务难迁移成为数据“流不动”的主要瓶颈。

海量数据形成的“三类孤岛”

- 应用孤岛：不同应用产生的数据分别存放在不同的存储系统中，而且这些数据由于各自的特征，彼此之间是无法共享使用的，形成“应用孤岛”问题；
- 管理孤岛：为对生产数据加以保护和使用，会将生产数据的一个副本，拷贝到各个系统（如备份、容灾、归档、开发测试和分析系统）中进行管理和使用。即便是同一份数据，为实现不同目的，还需分别存储、管理和使用，即形成“管理孤岛”问题；
- 地理孤岛：由于企业的更新换代，将存在多套存储设备，比如生产环境、非生产环境、云环境和边缘环境，企业的数据将存放在不同的地方，形成“地理孤岛”问题。

产生上述问题的根本原因：运营商在建设数据基础设施时，从满足客户的诉求出发并考虑投资成本问题，会选择不同的计算资源、网络资源和存储资源来分别满足客户的不同诉求。

- 算力墙：各个存储系统的 CPU 能力，仅供本系统使用，无法将算力资源共享使用，形成各存储系统之间的“算力墙”；
- 网络墙：各个网络都有各自的协议，彼此之间无法互连互通，即各个网络之间形成“网络墙”；
- 介质墙：存储介质的性能、容量和成本各不相同，客户会选择合适的介质存储数据，这使得数据分别存储在不同系统的不同介质中，而且这些数据很难共享访问，即各个存储介质之间形成“介质墙”。

3.1.3 用不好——价值获取低效

海量的数据孕育了前所未有的机遇，也带来了巨大的挑战。很多客户都反映，从来不缺数据，数据多了反而成为一种负担。也有客户说，数据只是资源，而不是资产，很难产生价值。其根本原因是没有用好数据，数据没有释放价值。而影响数据价值释放的主要原因是数据供应不足，无法反馈业务本质，无法支持业务决策：

- 大量数据未存储：运营商每天会产生大量数据，但传统的数据录入需要预先的人工规划，这导致大量非结构化数据以及一些新型数据无法进入系统（例如 IoT 数据、视频数据、图片数据等）。数据的缺失会削弱对业务的感知，无法真实及时地反映出业务本质。
- 性能不足，可靠性低：传统的 IT 设施一般 3-5 年一换，但由于各种原因，很多运营商机房里 8 年甚至 10 年以上的设备仍然在使用。另外，部分项目采购时出于预算原因，仅考虑了当年的短期需求，没有对未来的业务增长做准备。因此部分设备存在性能差，可靠性低的问题，拉低了整个 IT 系统的运行效率，无法满足 5G 对生产系统的高性能低时延高可靠性的需求。
- 找不到数据：运营商通常通过数据表来管理和分析数据，但是数据表甚至可以达到数百万张，而且分散在各个业务系统中。如果没有统一数据目录和全局数据视图，要在上百万张报表中找到特定的数据，好比大海捞针，无法应对灵活多变的业务需求。
- 谁对数据负责：在大数据时代，一个典型的分析业务通常需要跨平台的数据协同。如果已经接入的数据无法满足分析需求，需要从前端多个业务系统获取新的数据，再加上缺乏统一的隐私与安全共享机制，数据就需要经过多部门间协调、拉通、核实才能获得。数据分析的链路冗长，一旦出现问题，往往需要“六方会谈”才能定位，无法保证数据供应的稳定和高可用，更无法实现高效的数据融合分析。

3.2 挑战应对方法

面对数据海量增长所带来的问题，华为认为，应对挑战的方法应该从重定义数据基础设施入手，使数据存得下，流得动，用得好，以实现每比特成本最优、每比特价值最大，每比特管理效率最高。

重定义数据基础设施包含 4 个方面：重定义数据架构、重定义数据性能和可靠性、重定义数据密度、重定义数据价值。

3.2.1 改变 Silo 模式，重定义数据架构

需要对现有的 IT 系统 Silo 架构进行重新梳理，对数据架构进行重新定义。建立高效高扩展架构，解决数据存不下问题；从数据生命周期视角融合企业数据，解决流不动问题；提升架构开放性，解决用不好问题。因此，重新定义的数据架构需要具备如下特性：

- 融合：由原来 Silo 式的 IT 系统融合成为资源池式的系统，资源池中的资源按照性能和可靠性进行不同的等级定义，根据不同的业务按需发放相关资源。
- 智能：统一建设智能的数据管理系统，充分利用机器学习 / 人工智能等技术，运维管理模式由人工操作向数据全生命周期自动化演进，以机器智能的方式控制数据的产生、使用、流动和终结；
- 开放：一是易集成，数据和存储系统需要具备多样性的生态接口，具备集成到多个系统的能力；二是易使用，需要从方便开发者的视角出发，数据分析系统基于开放的引擎，屏蔽多数据源、多数据接口差异，以一个目录，一个接口，一份数据，实现数据应用开发极简。

3.2.2 借力闪存升级，重定义数据性能和可靠性

5G 时代，运营商用户话单量和计费频次增加，需要更快的存储来应对。

用户对体验提升的需求是无止境的。更高的 IOPS 意味着单位时间内完成更高的任务数，更低的时延意味着单任务的更快处理时间。

根据我们对全球上百个 TOP 客户的性能需求长期分析，客户的最高性能层要求，已经从三年前的平均 3000 IOPS/TB 提升到 7500 IOPS/TB 以上；生产域的平均性能要求，从 400 IOPS/TB 提升到 1000 IOPS/TB 以上。

随着 5G 以及 AI、IOV 等新兴应用的发展，超低时延应用已经进入市场应用阶段。我们看到很多客户已经不满足于 1m 的时延保障，在快速地向 0.5ms 以下进发。

随着闪存的大量普及和闪存生命周期管理、重删压缩等闪存应用技术的成熟，全闪存的价格已经快速下降，单块 SSD 具备较 SAS 盘高百倍的性能，高 10 倍的价格 / 性能比。即使在 400 IOPS/TB 这样较低性能要求下，单从成本看，也已经具备直接替换 10K SAS 盘的能力。

计费业务系统要求 7/24 小时不间断运行，需要极致的存储可靠性保证。

数据量的暴增，意味着相同业务中断时间内，5G 时代的业务损失比 4G 时代大了 10 倍。这势必要求数据基础设施的可靠性提升 10 倍来抵消业务损失的代价，要求单设备可用性从 4G 时代的 99.999% 提升到 99.9999%（年业务中断时间 32 秒），方案级可用性从 99.9999% 提升到 99.99999%（年业务中断时间 3 秒）。要求即使在迁移、升级、维护时 QoS 也能得到完全的保障，做到业务不中断、性能不下降。

SSD 的年故障率较机械盘低 50~70%，但全闪存的可靠性提升并不仅仅依赖于 SSD。领先厂商已经抓住闪存化契机进行彻底的 Flash Native 改造，采用极高可靠架构满足客户的需求，这往往已经超过传统高端存储的要求。而且由于采用了全闪存，其整台阵列的价格反而更低，往往仅相当于传统高端存储的维保费用。

因此，借助闪存升级，既可以进行存储成本结构的优化，又可以解决运营商面向未来的性能和可靠性问题。

3.2.3 软硬双管齐下，重定义数据密度

重定义数据密度，使单位空间内能够存放更多的数据，以达到降低 TCO 的目的，解决数据存不下的问题。可以从如下方面提升数据密度：

- 硬件上采用高密硬件：通过结构性改造和高科技散热技术增加硬盘密度，从而增加单位空间的数据存储量。传统 3.5 寸机械硬盘的密度为每 U 高度 4 个盘位左右，通过高密的精细化设计，可以提升到每 U 高度 10 盘位以上。但高密硬件需要解决好三个问题：

  1）散热问题：高密硬盘框往往多个硬盘前后重叠放置，这会导致前部硬盘的散热会叠加到后面硬盘上，使后部硬盘温度过高，增加了后部硬盘的故障率；这也是部分高密存储设备故障率偏高的原因；

  2）重心问题：高密硬盘框往往采用抽屉式设计，需要将框体抽出机柜面进行维护，极易造成机柜重心不稳，往往需要对机柜进行特殊的加重加固，增加了机房负担；

  3）维护问题：一是线缆，因为盘框需要抽出柜面进行在线维护，这就需要与盘框连接的电源和数据线缆也能随之移动；业界往往采用专用的导线轨来解决这个问题，但导线轨体积较大，尤其会占用很大的有效深度，这也是部分高密存储要求非标机柜的原因。二是硬盘，硬盘维护时需要做到易插拔，而且不能影响其他硬盘的运行。

  一般的通用服务器存储无法满足高密的存储需求，需要专业的存储设备来实现高密的存储需求。

- 软件上采用 EC 技术。传统的分布式存储和大数据分析系统往往采用副本方式实现数据可靠性，这会导致存储利用率偏低，典型 Hadoop 的三副本的实际空间利用率仅有 33%。采用 EC (Erasure Coding) 来替代副本，N 份数据 +M 份校验的组合可以达到甚至超过 3 副本的可靠性，而空间利用率可以有 50% 甚至翻倍的提升。

在计算和存储融合部署（如大数据）的场景，建议使用存算分离的方案，计算资源采用高密的计算节点，存储资源采用高密节点的存储节点并使用 EC 技术，实现计算资源和存储资源的解耦和独立扩展，提高整体方案的密度，降低整体方案的 TCO。

3.2.4 打破数据壁垒，重定义数据价值

5G 时代数据爆发式增长，如果不能充分挖掘海量数据的价值，实现数据变现，数据资产反而成为数据负担。

传统的大数据分析工具包含数据仓库和 Hive、Spark 等分析平台，可选组件有上百个，接口更多得无法控制。大数据开发工程师将过多的精力耗散在组件选择和试用、耗散在对各类接口和工具的熟练使用上。

另外，大数据平台所需的计算和数据几乎无差异的均衡分布在服务器节点上。无论系统是因为分析性能不够，还是存储资源的不足，都必须增加节点。这势必造成计算或者存储资源的浪费。

因此我们建议，可以从如下方面重新定义数据价值，解决数据存不下，流不动和用不好的问题：

- 打破大数据与存储的壁垒，实现存储和计算分离：计算节点着重于分析能力和热数据的储存，温冷数据存放在专用的存储节点上，这样计算 / 存储资源按需灵活配置，整体 TCO 可获得大幅度降低。
- 打破数据库与大数据的壁垒，实现跨平台融合分析：从“数据跑路”（数据跨平台拷贝）到“算力跑路”（分析任务按需分发，数据不动），数据 0 搬迁，大幅提升分析效率；
- 打破数据应用与数据基础设施的壁垒，实现数据虚拟化：通过虚拟化能力将多源的数据转变为统一的接口和目录，简化应用开发门槛，实现秒级找数取数，开发效率的提升。

4 5G 时代运营商数据基础设施

4.1 5G 时代运营商的数据基础设施目标架构

如何整合运营商生产域和分析域的资源，实现数据存储、数据库和大数据的垂直融合。满足数据的全生命周期的需求，让数据的存储、管理、使用更加智能，让数据价值得到更大发挥。华为基于数据基础设施的一池一湖解决方案能够很好的帮助运营商应对 5G 时代的挑战。

华为的数据基础设施是传统 IT 基础设施的延伸，以数据为中心，服务于数据，最大化数据价值。华为的数据基础设施具有融合、开放和智能的特性。

华为的一池一湖解决方案是指：

- 一池：即存储资源池，满足生产域系统的全生命周期数据管理的需求。
- 一湖：即智能数据湖，满足分析域系统的全生命周期数据管理的需求。

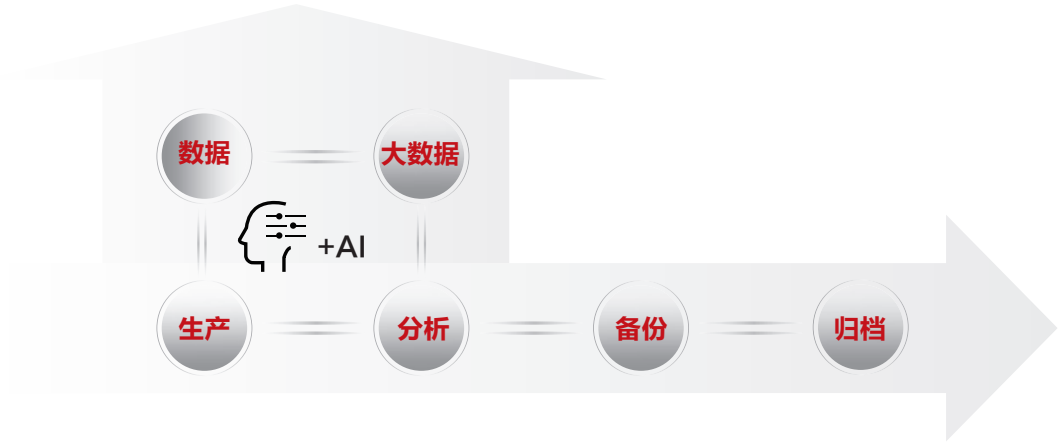


4.1.1 华为数据基础设施特性

华为围绕计算、存储和数据处理三个方面来构筑极简数据基础设施，将昇腾和鲲鹏处理器多样性算力引入到数据存储和数据处理，将数据存储和数据处理创新性地进行融合，并对数据全生命周期进行智能化管理，同时，通过数据虚拟化引擎将数据基础设施对外开放，搭建起数据基础设施同上层数据应用的桥梁。华为数据基础设施具备： 融合、智能、开放的特性。

- 融合：通过“打破数据处理与数据存储的边界”，实现数据高效共享和分析，降本增效。
- 智能: 通过“AI+ 存储 + 云”, 实现数据全生命周期内智能管理和智能运营, 让存储越用越快, 价值越来越大。
- 开放：通过“数据虚拟化引擎”，统一 SQL 语法，像使用数据库一样使用大数据。

4.1.1.1 融合



华为数据基础设施的融合体现在“一横一纵”两种方面。

- 横向融合是指数据全生命周期存储的融合。数据产生的第一环节是生产存储，以支撑交易型的数据处理；通过扩展至分析型存储来支撑核心的分析业务；备份存储进一步扩展分析场景；主存增加混合云备份、分级等特性，实现冷数据上云。通过对生产存储叠加轻量化备份和管理特性，拓展存储场景，实现从热到冷的数据全生命周期存储整合。
- 纵向融合是指数据处理与数据存储的垂直优化。包括交易型数据处理与生产存储相融合，提升性能，增强可靠性；数据分析与分析型存储融合，提高分析效率。在存储层，通过重定义存储架构，将块、文件、对象、HDFS 等多种存储服务融合，打通数据孤岛，解决多样性数据存储和共享问题；协议方面，通过多协议融合技术，实现一份数据同时支持数据库、大数据、AI 等多种业务的分析需求，节省数据无效流动时间，让分析更高效；算力层面，通过将数据库、大数据、AI 多引擎融合分析和多样性算力统一调度，降低海量数据处理难度，实现高效分析；管理层面，通过将 AI 融入存储全生命周期管理，从资源规划、业务发放、系统调优、风险预测、故障定位等方面实现智能运维，从容应对数千节点规模的复杂管理。

  通过横向融合和纵向融合，实现 TCO 降低 30% 以上，分析效率提升 100%。

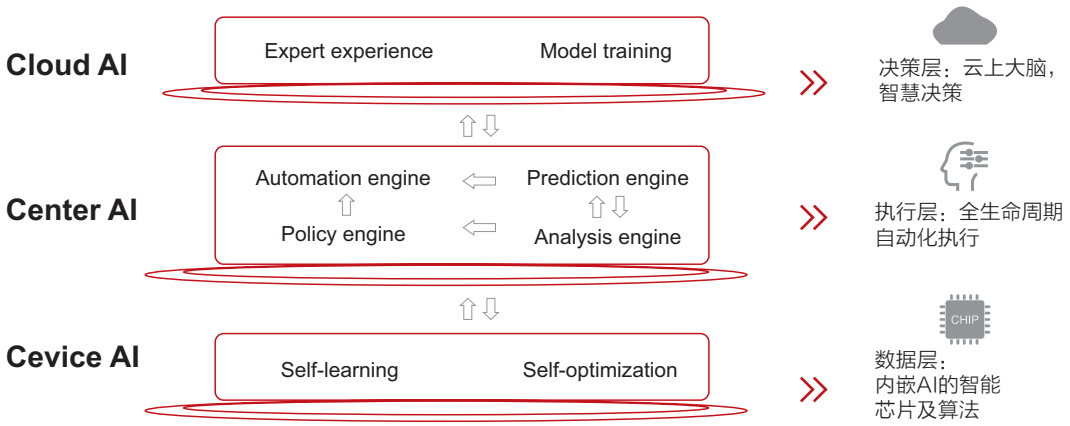
伴随着“一横一纵”的融合，底层的存储设施也向着“两化”的方向发展。面向关键业务的全闪存化和面向海量存储需求的分布化。

- 关键业务全闪存化：在 5G 时代对于运营商的核心业务比如 BOM 域的 Billing 和 CRM, B2B 的 Hot 业务，还有超快

业务比如包括 IOV（自动驾驶）、实时决策等，要求高性能低时延以及极高的可靠性。以 HDD 为主要存储介质的传统存储已经在性能和可靠性上无法满足 5G 时代新业务的需求。针对 HDD 的存储介质革命应运而生，引领介质革命的主角就是固态硬盘（Solid State Drive，简称 SSD）。相对于 HDD，SSD 具备更加优异的性能和更好的可靠性。因此面对 5G 时代的关键应用，存储设备正全面转向全闪存存储。

- 大容量需求分布式化：对于 5G 时代的其他业务需求，比如普通应用包括 OA、Web Service、AR、VR、邮件等，大容量应用包括视频、IOT 等，这些应用对时延和可靠性要求相对较低，但数据量会急剧膨胀。因此对大容量、易扩容、易运维等特性要求高。分布式存储能够做到：大容量，可横向扩展到 EB 级，性能与容量线性增长；易扩展，基于标准硬件按节点扩容，采购周期从月降到天，业务上线效率是传统存储的 10 倍以上；易管理，支持多种协议，一套存储可以应对各种需求，硬件换代数据免迁移，0 迁移费用，0 业务风险。因此分布式存储可以有效解决 5G 时代多业务、多场景、大容量的业务需求。

4.1.1.2 智能



过去，主要依靠人来完成设备配置和运维，运维人员经验和能力的高低决定了管理效率的优劣。华为基于设备级 AI、数据中心级 AI 和华为云 AI 的三层架构，通过云上训练和云下推理，让系统越用越快、越用越省。

- 我们在每一个设备里，内嵌昇腾 AI 芯片和相关算法，使我们的设备在毫秒级甚至更低级别时延上做到自愈自关系。比如在存储中依托昇腾 AI 能力，自动学习和识别 IO 流，提升 Cache 预取命中率，系统整体性能提升 20%。
- 我们在中间层有一个类似于网管的设置，称之为数据管理引擎（DME），DME 所管理的是一个数据中心或者相邻几个数据中心面向一个企业客户的网络运维设备。它可以进行分钟级的系统自动化运维。
- 云上的 AI，可以在天和月时间维度训练运维模型，并且它可以跨厂商、跨企业来进行训练。结合华为云自身大规模运维运营经验，当前可以实现提前 14 天预测硬盘故障，提前 60 天预测性能瓶颈，提前 365 天预测容量不足，其中 30% 的故障可以自我修复。

4.1.1.3 开放

华为的数据虚拟化引擎 -- 河图引擎，屏蔽了数据基础设施的复杂度，让开发者像使用数据库一样使用大数据，复用现有的生态、工具和技能，能够提升开发效率 2 到 10 倍。 河图引擎，就是要让数据治理简单，使用简单，更方便的挖掘数据价值。河图引擎能够提供一个目录、一个接口、一份数据，能够让用户像数据库一下使用大数据。

- 一个目录：通过元数据在线感知，构建 1000+ 异地异构数据源全局虚拟数据视图，打破数据孤岛，数据全局可视，解决企业找数难问题。

- 一个接口：通过开放的连接框架、5000 节点 SQL 引擎，实现 30 种异地异构数据源统一 SQL 访问，秒级获取，数据全局可得，解决企业取数难问题
- 一份数据：通过 CarbonData 技术，实现一份数据多场景分析，多应用共享，数据 0 搬迁，数据全局可用，解决企业用数难问题。

4.1.2 华为数据基础设施应对 5G 时代运营商挑战的方法

华为的 5G 数据基础设施从“重定义”四个维度来应对 5G 时代运营商海量数据处理和分析的挑战。

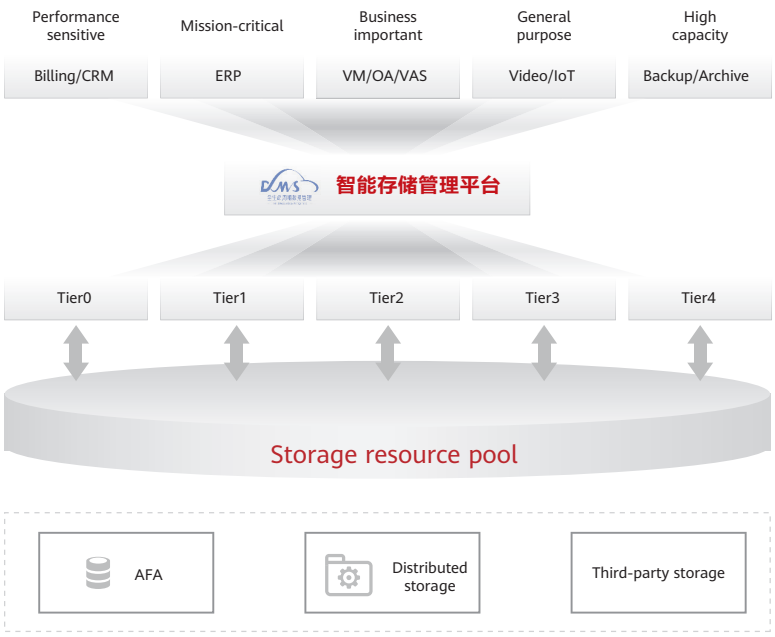
- 重定义数据架构：由原来 Silo 的架构演变成一池一湖的架构，并通过华为的全生命周期管理平台（DMS）进行统一管理，系统管理效率能够有 5 倍的提升，TCO 节省 30%。
- 重定义性能和可靠性：华为全闪存存储 Dorado 可提供 2000 万 IOPS，0.1ms 时延，有效应对 5G 时代业务对高性能的需求，例如：计费系统出账时长从 10 小时缩短到 3 小时，缩短 70%。另外在可靠性方面，华为全闪存最高容忍控制器 8 坏 7，业务不中断，可靠性达到业界最高的 7 个 9，保证全年业务 0 中断。
- 重定义数据密度：华为 OceanStor 分布式存储结合业界超高密度的硬件，相比传统存储，存储 150 PB 的视频数据，可以由 32 个机柜减少到 8 个机柜即可，空间减少 75%，能耗减少 65%。
- 重定义数据价值：通过大数据存算分离实现计算和存储资源按需灵活配置；通过华为的河图虚拟数据化引擎能够实现像使用数据库一样使用大数据平台，做到降低应用开发门槛，实现秒级找数、取数。

4.2 重定义架构 — 一池一湖解决方案

4.2.1 一池：面向生产域的融合资源池

面向生成域，构建融合、智能的存储资源池，实现数据生命周期的全自动化管理。传统的 Silo 存储建设方式演变成统一的存储资源池，资源池按照 SLA 的要求向上层应用提供存储资源。数据管理方面也从传统的人工管理、人工运维演变成在生命周期各阶段（规划、分配、运维和优化）自动化管理，大大提升存储效率。

4.2.1.1 融合：构建融合存储池，存储利用率提升到 70%



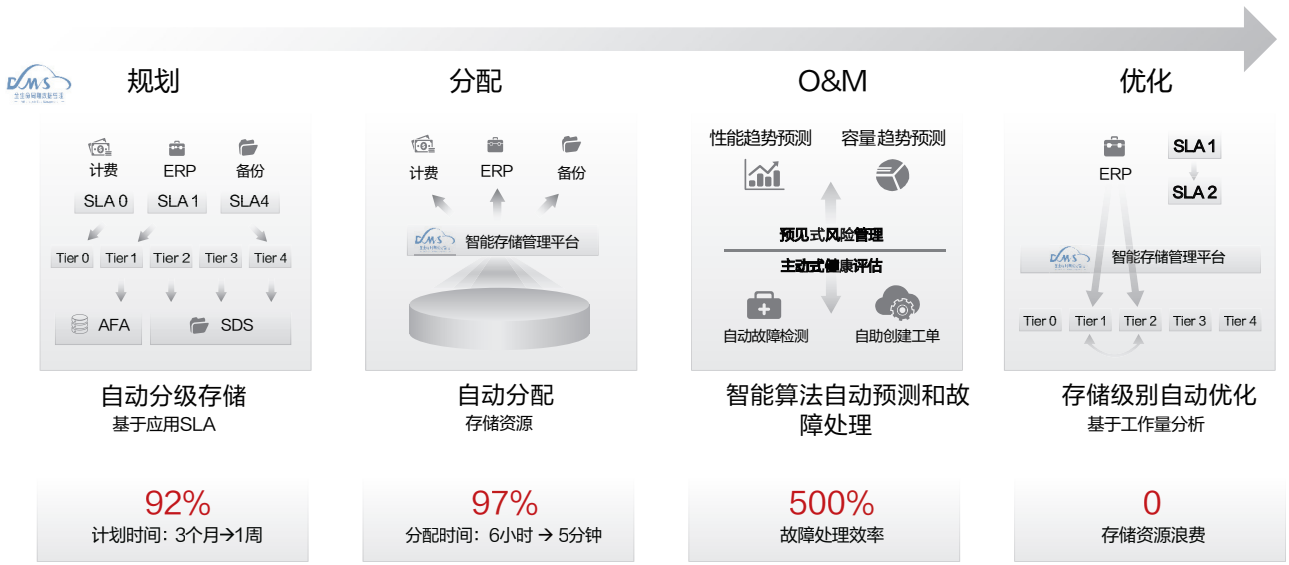
目前在运营商的 B 域、O 域和数字化业务开始实现云化或者虚拟化，目前运营商已经大规模采用云化和虚拟化的技术，但是相对于计算资源的云化，存储的部署模式依旧是采用 Silo 的部署方式。这种部署模式造成存储利用率 <30%，资源申请困难，难管理。

华为的统一存储资源池解决方案，基于自研的智能存储管理平台 DMS 整合众多的存储形态，例如华为的全闪存存储、分布式存储以及第三方厂商的存储，将这些不同来源和不同类型的存储整合成统一的资源池，并由华为智能存储管理平台统一根据不同业务的 SLA 自动发放，实现存储资源的按需灵活调度。

关键业务全闪存化和大容量需求分布式化可以大幅简化存储的种类和数量。比如像运营商的计费系统和 CRM 系统，这类业务对单位 TB 的 IOPS 要求高达到 4000 IOPS/TB 且对时延要求也比较苛刻 IO 响应必须小于 1 ms，另外对系统的可靠性要求也在 99.9999% 以上，像这类业务适合于部署在全闪存存储上。而其它的类型的业务比如像 VAS、视频、IOT 等对性能的要求不高，但普遍存在大容量需求，且系统扩容的频率也较高，这些业务的数据就完全可以用分布式存储来存储和管理。虽然运营商的业务多种多样，但我们完全可以用两种类型的存储简化底层的存储系统。

实际使用情况来看，通常情况下平均存储空间利用率从 30% 提升到 70%，整体 TCO 降低 30% 以上；并且存储资源申请可由业务需求方自助申请，自动发放，平均发放时间从 6 小时以上降低至 5 分钟，节省 90% 以上。

4.2.1.2 智能：构建全生命周期自动化，管理效率提升 5 倍



另外对于融合的存储资源池，构建全生命周期的自动化也是未来生产域演进的重要方向。全生命周期是指规划、分配、运维和优化四个阶段，我们认为只有实现了这四个阶段的高度自动化才能够称之为真正的全生命周期自动化：

- 规划阶段：用户需要针对其各种业务的 SLA 进行定义和建模，并完成针对不同 SLA 等级的存储类型和配置模型基线。这类工作通常由运营商的专家根据实际业务情况评估和分析得来，整体耗时经常需要 3 个月或更多，并且也很难完全评估准确。华为提供的工具可以自动分析业务的 SLA 状况并给出存储配置建议，实现自动化的规划。基于华为方案，用户可以大幅度的缩减规划过程的时间和工作量，通常 3 个月的评估可以降低到 1 周完成，节省 92%；
- 分配阶段：基于华为方案提供的智能存储管理平台，业务管理者能够根据自己的业务需求直接自助申请存储资源，而无需复杂的流程以及 IT 管理员的操作。智能存储管理平台能够根据业务所对应的 SLA 模型和存储配置模型自动从资源池中挑选最合适的存储资源，并自动的提供给应用。整体发放过程高度自动化，存储资源发放时间能够从 6 小时降低至 5 分钟，节省 97%；

- 运维阶段：华为方案能够提供主动式健康评估和预见式风险管理两种能力。在用户授权下，数据中心可以通过安全链接与华为公司的运维云连接。基于云端的大量经验模型，华为方案可以帮助用户主动的识别设备故障，并自动在云端建立维护工单，华为工程师能够根据工单尽快处理故障，整体故障解决效率提升 5 倍；另外，华为的 AI 能力能够帮助用户提前 60 天预测性能需求，提前 1 年预测容量需求，便于用户提前规划资源配置，解决使用上的风险。
- 优化阶段：实现了规划、分配和运维的自动化后，整体的统一存储资源池的配置关系依然是固化的，并不能根据业务的实际负载情况进行自动调优。基于智能存储管理平台，华为解决方案能够自动监控各种业务的实际运行情况，根据整体运行时间中各业务实际的负载变化来自动重新建立业务与存储的配置模型，并在用户的授权下能够自动根据业务负载的变化来自动在资源池不同的 Tier 中迁移数据，实现 TCO 和 SLA 需求之间的持续自动优化，做到资源 0 浪费。

基于全生命周期自动化能力，华为解决方案能够真正帮助客户显著降低存储 TCO、加快上线时间业务上线时间、并能够大幅度的降低使用的难度，整体的管理效率可以提升 5 倍。

4.2.2 一湖：面向分析域的智能数据湖

面向分析域，构建融合、开放的智能数据化，助力数字化转型，释放数据价值。

4.2.2.1 融合的数据处理平台：融合大数据分析平台，数据分析效率提升 10 倍

在传统的分析开发中，由于数据分散在不同的系统里，数据需要通过不同的 API 访问，而且存在数据在不同库之间导入加载的情况，数据分析开发效率低。

平台融合以后，通过融合的河图引擎，通过统一的 API 方式访问，数据可以在大数据和数据仓库中协同分析，不需要跨库进行数据迁移，协同效率可以大幅提高。典型场景下，数据的分析效率可以提升 10 倍。

4.2.2.2 开放的使能平台：兼容多种数据源接口，数据集成速度提升 10+ 倍

在现有的采集方式下，存在不同的数据类型和不同的数据采集工具，数据集成复杂，运维难度高。通过开放的数据采集平台，一套平台支持不同的数据类型，统一调度统一运维，大大提升数据的采集速度。典型场景下，数据集成速度可以提升 10+ 倍。

4.2.2.3 放的数据虚拟化引擎：统一的数据目录，秒级取数

采用统一的河图引擎，提供统一数据目录能力，解决原来的用不好、找数难、取数难的问题，实现一键式自动定位数据，秒级找数，实现取数与找数融合，所见即所得，秒级取数。

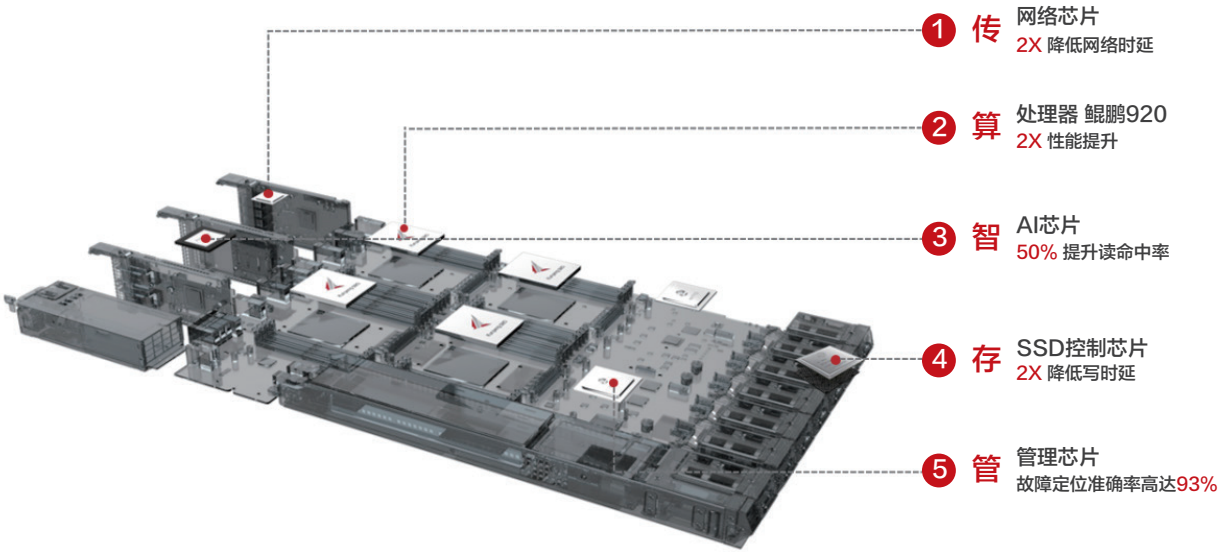
4.3 重定义性能和可靠性—全闪存存储 OceanStor Dorado

华为的全闪存 Dorado 通过重新定义性能和可靠性来满足 5G 时代高性能和高可靠性的诉求。

4.3.1 重定义性能：业界领先的性能，2,000 万 IOPS；0.1ms 时延

4.3.1.1 芯片引领、永快一步

华为通过在芯片领域不断的积累和持续投入，围绕 SSD 构建了最强性能全闪存存储产品 --OceanStor Dorado 新一代全闪存。在 OceanStor Dorado 新一代全闪存里，华为通过对传、算、智、存、管的垂直整合，构建了端到端的芯片平台，它包含 5 种华为芯片：



1. 智能多协议接口芯片承载了本应由通用 CPU 负责的协议解析工作，加速前端访问性能 2 倍；
2. 智能处理器芯片 鲲鹏 920 采用业界标杆性能的 CPU，算力优于业界水准 25%，单控制器性能提升 2 倍；
3. 智能 AI 芯片昇腾 310 基于机器学习框架，主动分析并掌握多个应用模型的 I/O 规律，实现内存空间的智能预取，读缓存命中率提升 50%；
4. 智能 SSD 控制芯片承载核心 FTL（FlashTranslation Layer）算法，加速 SSD 盘内数据访问，写时延降低 50%；
5. 智能 BMC 管理芯片内置华为存储故障基因库，加速部件故障诊断和定位，准确率达到 93%。

4.3.1.2 FlashLink，智能存储

业内大部分闪存厂商缺乏端到端的自研能力，无法充分发挥 SSD 的性能和存储的处理能力。华为 OceanStor Dorado 全闪存存储基于控制器、芯片和操作系统的全自研设计，开发了业内领先的 FlashLink 全闪存智能算法，能够让智能芯片和存储关键部件，同步协调 SSD 和存储控制器内的数据布局，采用大块顺序写和元数据独立分区技术保障性能平稳，采用 I/O 优先级调整保障读写 I/O 永远第一时间响应，采用业务分流算法减轻控制框负载，智能硬盘框主动承载数据重构业务。FlashLink 智能算法充分发挥闪存的潜力，帮助 OceanStor Dorado 全闪存存储实现无与伦比的性能，为用户提供流畅的业务员体验。

不可预测的业务增长要求存储提供更高可预测、更高性能和更平稳增长的基础架构，以适应不断变化的业务要求。OceanStor Dorado 全闪存存储 scale-out 架构支持线性扩展到 32 控，IOPS 随引擎数量线性增长至 20,000,000 IOPS，是友商 2 倍，以可预测的性能匹配未来不可预测的业务发展。

4.3.1.3 端到端 NVMe，致胜未来

传统的 SAS 协议是为机械硬盘设计，复杂的系统架构、过多的协议解析、有限的队列并发能力，这一切都成为限制 SSD 能力充分发挥的瓶颈。NVM Express 组织由此制定了 NVMe 协议标准，去掉了 SAS 系统中的 IO Scheduler 和 SCSI 等复杂的协议层，换成轻量级的 NVMe 协议。它实现 CPU 和 SSD 直接通信，缩短传输路径；提升并发数 65536 倍，性能更优；协议交互从 4 次减少为 2 次，写请求处理效率提升 1 倍。华为是目前唯一支持全系列端到端 NVMe 架构的全闪存存储厂商，OceanStor Dorado 全闪存前端采用业界领先的 32 G FC/100GE NVMe 协议，内部采用 PCIe 4.0 协议，后端采用 100 GE ROCE 协议，实现端到端数据加速，时延可达 0.1 ms，传输速率提升 5 倍（相对 SAS 全闪存）。

4.3.2 重定义可靠性：高端存储可靠性新标杆

架构是高端存储的灵魂所在，市场上主流的高端存储产品，多年来均未实现大的突破。那么我们先来看看为什么我们认为 OceanStor Dorado 全闪存高端存储重新树立了可靠性新标杆。从如下架构对比图上可以看出，华为的 SmartMatrix 全互联均衡架构不仅实现了控制器的全互联，而且还实现了前后端的全共享，使得硬件故障的容忍度极高，带来了极高的业务可靠性，当出现故障时，实现了上层业务“零”感知。

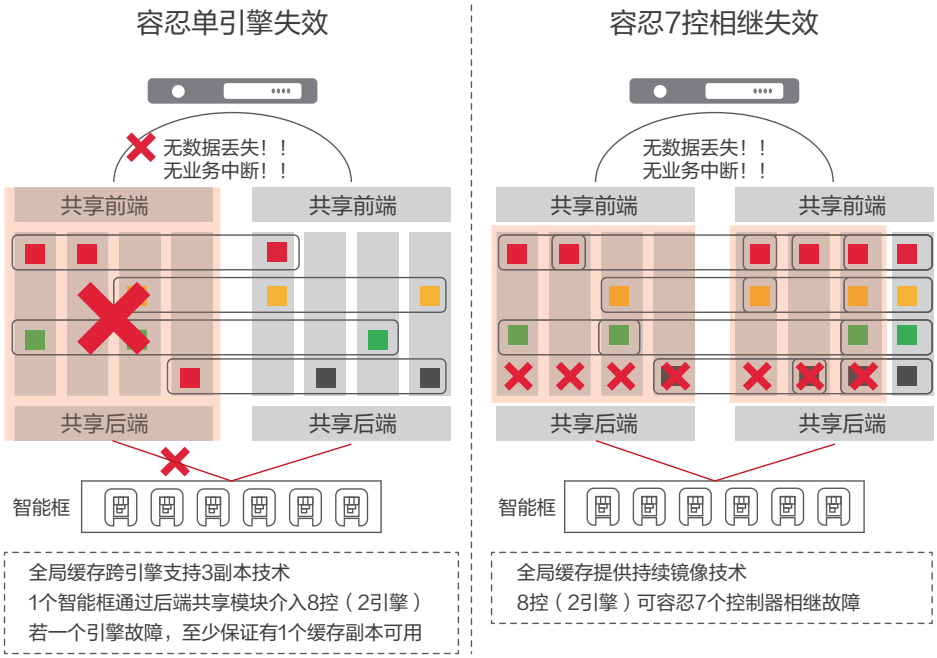
SmartMatrix 全互联均衡架构已从各个方面实现了对市场主流架构的突破和超越，使得华为 OceanStor Dorado 全闪存高端存储系统成为业界唯一实现控制器容忍 8 坏 7，引擎容忍 2 坏 1，控制器故障，主机业务 0 感知的高端存储，在可靠性方面树立了新的标杆。

4.3.2.1 可靠性标杆点一：控制器故障链路无切换，主机业务 0 感知

要实现控制器故障，主机业务 0 感知，关键在于前端接口卡。对于传统接口卡，前端业务 I/O 的下发和分发需要经过控制器预处理，存在不同控制器间的转发，并且在控制器故障场景需要主机多路径进行链路切换，倒换通常需要 5~30 秒。而 OceanStor Dorado 高端全闪存支持前端智能共享卡，智能共享卡对主机的 I/O 进行智能识别处理，并按特定规则分发，使得主机 I/O 无需控制器预处理就直接发送给最佳处理控制器，实现了主机 I/O 直通，避免主机 I/O 在控制器之间转发。当控制器意外故障时，智能共享卡端口芯片会感知到与控制器之间的 PCIe 链路断开，配合控制器内的业务倒换，智能共享接口卡把主机的请求重新分发到其他控制器，实现了控制器故障秒级切换，主机业务 0 感知。

4.3.2.2 可靠性标杆点二：容忍控制器 8 坏 7，容忍引擎 2 坏 1，业务 0 中断

传统高端存储系统，硬盘框不能做到被所有控制器访问，相当于硬盘框对于控制器来说存在归属关系，这样不仅链路上的可靠性不能得到保证，从控制器向硬盘框下发 I/O 也始终存在限制，始终存在部分硬盘属于某些控制器的“私人财产”，一旦这些控制器出现问题，这些硬盘就无法访问，整体还是存在可靠性方面的风险。而 OceanStor Dorado 高端全闪存采用后端智能共享接口卡实现硬盘框与 8 个控制器全互联。



硬盘框实现 8 控全互联后，结合内双活特性，一方面通过缓存 3 副本技术，即业界首创的 3 个缓存副本分布在 3 个不同的控制器，从而保证两个控制器同时故障，或者一个控制框（引擎）故障情况下，至少还有一份有效的写缓存副本，实现 2 个引擎任意坏 1 个引擎，业务 0 中断；另一方面还能通过业界领先的缓存持续镜像技术，即一个控制器失效，镜像控制器重建缓存镜像，从而实现 8 个控制器陆续坏 7 控，业务 0 中断。

OceanStor Dorado 高端全闪存采用的 SmartMatrix 全互联均衡架构打造了可靠性的新标杆，引领全闪存发展的新方向，把存储系统的可靠性推向一个新高度。

4.4 重定义数据密度—业界超高密分布式存储标杆

华为提供 OceanStor 分布式存储系统，通过构建可靠、高效和弹性按需的新型存储服务平台，助力企业轻松应对 5G 时代下大规模数据高效存取需求。

4.4.1 超高密度存储节点

2020 年华为将发布全新的 OceanStor 太平洋超高密存储节点，它在 5U 的空间内实现了创纪录的 120 盘位的超高密度。插盘数量相比普通高密节点提升 20%，配合大容量磁盘可实现 2 倍的裸磁盘容量。

太平洋高密存储节点突破了多个工程极限，使用了数十项专利技术，很好地解决了高密存储的三大难题：

- 散热上，华为充分利用框体空间进行特殊的风道设计，在热通道上不断进行新风补充，使盘框前后的温度均衡，有效降低了硬盘的工作温度；
- 重心问题上，华为采用特殊的蝶翼双飞设计，实现前向后向均可抽出盘框，每次维护仅需抽出一半框体即可，有效确保机柜重心均衡，无需加重加固；
- 维护问题上，华为采用了控制部分和盘部分分离的方案，内部采用特殊缆体将盘框和控制部分连接，做到框体维护时盘动控不动，因此外部电源和数据线缆无需跟随移动，将机框有效深度提升 30% 以上。盘框采用插秧式设计，单张硬盘维护对其他部件无任何影响。

4.4.2 超高存储利用率

Erasure Coding（纠删码，后文缩写为“EC”）是一种数据冗余保护机制，主要应用于分布式存储领域。EC 在分布式存储系统中使用 N 个数据块和 M 个校验块保证数据的可靠性，这 N+M 个数据块中有任意 M 个块数据损坏，都可以通过其他 N 个块上的数据恢复 M 个块的数据。

相比于副本存储方式，EC 数据冗余保护机制在提供高可靠性的同时也能够提供更高的硬盘利用率，从而降低成本。比如一个 4 MB 的 IO，在三副本存储方式下，共占有 12 MB 的硬盘空间，空间利用率 33%。而在 4+2 配比的 EC 存储方式下，4 个数据节点每个占用 1 MB 空间，2 个校验节点各占用 1 MB 空间，共 6 MB 空间，空间利用率可达到 67%。在提供相同可靠性的前提下，EC 的方式比三副本节省了 6 MB 硬盘空间。

OceanStor 分布式存储提供多种 M+N（或者 M+N:B）的冗余比配置，用户可根据业务需求在管理界面上进行配置。目前华为的 EC 算法可以提供业界最高 22+2 的 EC 数据冗余保护机制，存储空间利用率可以达到 91%，比业界普遍采用的 4+2 的 EC 模式，存储空间利用率提升 25%。

4.4.3 智能散热，单 TB 功耗最优

华为提供智能散热技术，可实现空闲磁盘休眠和智能风扇调速，做到单 TB 功能最优，使节点的整体能耗降低 50% 以上。

整体来看，采用华为的分布式存储解决方案结合高密硬件、华为的 EC 算法和智能散热技术，能够大幅帮助运营商节省 OPEX。例如：存储 150 PB 的视频数据，可以由 32 个机柜减少到 8 个机柜即可，可使空间减少 75%，能耗减少 65%。

4.5 重定义数据价值—FusionData 解决方案，打破数据孤岛，释放数据价值

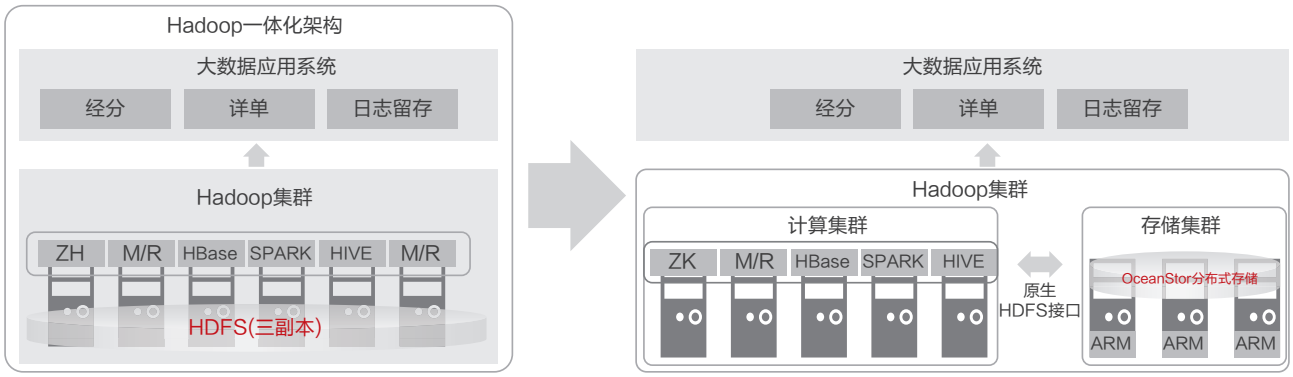
华为 FusionData 解决方案通过存算分离、融合分析和数据虚拟化能力，打破数据孤岛，释放数据价值。

4.5.1 存算分离实现计算 / 存储资源按需灵活配置，整体 TCO 下降 30%

目前运营商大数据平台建设模式一般计算和存储一体化融合部署，大数据服务器节点同时提供数据计算和数据存储，每个服务器配置固定，计算和存储同比扩展，存储通常使用 3 副本进行可靠性保障。这样将会造成以下主要问题

- 存储有效空间利用率低，存储使用 3 副本方案，1 份数据存 3 分，造成存储有效空间率不足 33%，节点数据成倍增加；
- 资源闲置，计算与存储等比例扩容，扩容配置相对固定。计算与存储资源不能按需扩容，扩容存储时计算也会跟着扩容，计算资源闲置；扩容计算时存储跟着扩容，存储可能闲置。

华为的计算存储分离解决方案能够让计算侧专注于计算任务和调度，不需要做数据管理，数据管理全部由存储来实现，计算和存储可灵活扩展。



使用存算分离方案的价值：

- 精准投资，计算与存储解耦，可独立扩容，计算资源不足扩容计算，存储资源不足扩容存储；
- 存储有效容量利用率提升，利用 EC 技术（Erasure Coding）替换 3 副本技术保障数据可靠性，利用率从 33% 提升到 91%，相同容量需求可减少 2.75 倍硬盘数量，降低成本；相同投资可以成本增加可使用容量；

通过计算与存储分离技术，打破系统烟囱式建设的顽疾。通过智能分布式存储的多协议融合技术，实现一份数据同时支持数据库、大数据、AI 等多种业务的分析需求，打造数据融合处理的基础。

4.5.2 融合分析提升分析效率提升 50 倍

运营商的数据往往跨越不同的数据平台和数据源，往往都需要大数据和数据库融合分析来用满足业务诉求。

比如在网规网优业务中，通过大数据平台的 Spark 处理引擎对 O 域的海量数据进行分析和处理，大数据平台在该场景主要用于数据挖掘和机器学习类批处理作业。分布式 MPP 架构数据仓库 GaussDB 在该场景支持深度挖掘和在线交互查询。做网规网优业务分析时有需要同时融合大数据平台和数据仓库平台中的数据来进行分析，如精准建站分析。

河图引擎的融合分析能力实现跨平台分析，从“数据跑路”（数据跨平台拷贝）到“算力跑路”（分析任务下发），数据 0 搬迁，传统的分析需要频繁加载数据分析时间往往需要数小时，现在采用融合分析的模式，大幅减少数据加载的时间，协同分析只需数十分钟就可以完成，分析效率可提升高达 50 倍。

4.5.3 数据虚拟化提升开发效率 3.5 倍

借助于河图引擎，我们希望能够帮助运营商打造一个数据平台。让我们的伙伴们能够像使用数据库一样方便的使用大数据。能够让传统的大数据业务开发效率提升 2—10 倍，新业务开发时间由 70 天能够提升到 20 天。华为的河图引擎具备一个目录、一个接口和一份数据的能力。

- 一个目录：秒级找数，提升业务处理效率

  对于运营商的大数据业务而言，在前期的数据准备阶段中，合作伙伴需要到各个数据库的数据目录中查找所需要的数据。对于不同的数据源还需要准备不同的数据查找工具，整个过程耗时耗力。而对于华为的虚拟化引擎，构建异地异构数据源全局虚拟数据视图，打破数据孤岛，数据全局可视，解决企业找数难问题。通过全局的统一目录，开发者能够快速的找到所需要的数据。

- 一个接口：开发效率提升 2~10 倍，降低开发成本

  另外一个，相对于传统的数据库平台，大数据组件多，每个组件的接口也不一致，这就造成大数据业务开发和数据获取困难，需要对每种接口都需要做相应的对接和转换，对大数据库的使用人员的也要求掌握多种开发工具和语言，从而导致开发和使用门槛变高，业务开发周期长。华为的河图引擎通过开放的连接框架，实现 30 种异地异构数据源统一 SQL 访问，秒级获取，数据全局可得，解决企业取数难问题。用户只需要熟悉标准的 SQL 语句，就能够像使用数据库一样使用大数据的各个组件。能够极大的提高大数据业务的开发效率。

- 一份数据：0 数据搬迁，减少数据加载时间

  传统的数据分析需要把数据在数据源之间进行数据的搬迁，往往还需要专门的 ETL 的工具做数据的加载。这个过程需要数周的时间完成跨部门的数据获取并完成数据的搬迁，另外也将会产生额外的数据量并占用宝贵的存储空间。华为的 Hetu 引擎能够在运营商实现一份数据多场景分析，多应用共享，数据 0 搬迁，数据全局可用，解决运营商用数难的问题。和传统的方式相比，河图引擎能够把数据获取的时长有数周降低到数天。

4.6 智能数据，面向未来

目前 5G 相关低时延标准还不成熟，到 uRLLC 阶段，运营商会开始低时延业务的部署，这会导致运营商的数据布局发生革命性变化。

出于降低 TCO 的需要，运营商会将时延不敏感数据上收集中到国家或跨国数据中心，部分数据还会进一步上到公有云。而时延敏感型数据，则会根据业务需要，尽量向终端用户靠拢。从而形成中心集中、边缘下沉的两极化分布。

未来 2C 的视频、云游戏、车联网，2B 的园区、面向不同行业的定制化边缘，以及实现对 5G 网络支撑的 NFVI，5G 的边缘必然百花齐放。业内外对 5G 低时延带来的业务革命寄予极大期许。但边缘下沉不仅意味着数据布局和业务架构有极大改变，也意味着巨大的成本开销，涉及选址、租赁、建站、布网等一系列操作，也是运营商无法承受之重。如果过早投资边缘，将会得不偿失；但如果投资过晚过慢，又会丧失商机。由于业务的不确定性和下沉的高成本，边缘下沉并不会一步到位，而是会分步骤逐步下沉。

因此，运营商急需寻找到一种成本低又能快速实现海量节点下沉的方式，在时机成熟时实现快速布局。建议运营商从两方面着手实现对多样边缘设备的整体控制：

- 提前准备边缘设备快速部署和业务迅速上线能力，把控下沉节奏，不冒进；
- 利用自身网络优势，提前规划并紧抓云网边 E2E SLA，实现 2B 和 2C 的 SLA 变现。

理想的边缘节点包含如下特性：

- 计算、存储、网络一体化：既方便管理又利于扩容。为确保灵活地支持业务承载，计算面需支持虚机和容器部署。
- 一柜承载所有业务：同一设备具备承载 B2B, B2C, NFVI 业务的能力，并提供物理和逻辑资源隔离的功能，以确保不同业务的质量保障需求。
- 免机房改造：在边缘逐步下沉期，边缘设备应满足 CT 机房机电标准，具备迅速下沉到 POP 和 CO(central office) 机房的能力，以节省机房选址、租赁和基建成本，加速边缘 IT 的普及。而在 5G 成熟期，近终端的不同的 2B/2C 边缘，其散热、供电等需根据办公、矿山等不同边缘环境进行整体设计，无需部署在专业或独立机房，以节约投资，加快 TTM。
- 运维免上站：站点统一接入运维管理中心，一个界面可以管理所有边缘设备。全图形化界面，所见即所得。可对设备进行远程管理与维护，软故障可远程处理与排除。
- 业务部署免上站: 中心提供应用仓库管理,将新业务上传到应用仓库,新业务可从中心批量下推到边缘,实现自动化部署。
- 多级灾备：采用本地、邻站与远程结合的多级灾备机制，以保证单站点灾难业务可迅速拉起。

5 总结

5G 已来。5G 的规模部署对运营商带来了机遇，也给 IT 基础设施带来了挑战。

5G 时代运营商数据爆发式的增长。在 BSS 域，BSS 域计费话单数据量增长 7.5 倍，OSS 域运营数据量增长 8 倍，2C 业务的高清视频业务增长超 10 倍，大数据分析数据量增长 5 倍。

海量数据蕴含巨大的价值，在带来更多机遇的同时，也给传统的 IT 基础设施带来了前所未有的挑战，数据存不下、流不动、用不好成为运营商数据应用最普遍的难题。应对难题的方法通常为：改变 Silo 模式，重定义数据架构；借力闪存升级，重定义数据性能和可靠性；软硬双管齐下，重定义数据密度；打破数据壁垒，重定义数据价值。

本白皮书给出了 5G 时代运营商数据基础设施的目标网架构。华为数据基础设施具备融合、智能和开放的特性。华为通过“一池一湖”重新定义运营商的数据基础设施架构，通过全闪存存储 OceanStor Dorado 重定义性能和可靠性，通过 OceanStor 分布式存储重新定义数据密度，通过 FusionData 解决方案重定义数据价值。

面向未来，对于时延敏感型数据，运营商的 IT 架构还会进一步向智能边缘演进，但是由于业务的不确定性和下沉的高成本，边缘下沉并不会一步到位，而是会分步骤逐步下沉。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编：518129

网址： https://www.huawei.com/

版权所有 © 华为技术有限公司 2020。 保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明

HUAWEI 和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。