

支撑关键应用上云

QingStor™ NeonSAN 性能评估报告

E 企研究院

目录

前言	2
<i>QingStor™ NeonSAN 架构解析.....</i>	4
<i>QingStor™ NeonSAN 逻辑架构图.....</i>	5
<i>E 企研究院验证：QingStor™ NeonSAN 性能评估</i>	7
基准性能测试：QingStor™ NeonSAN 契合分布式应用	7
Oracle RAC 性能测试：直面关键应用	10
<i>最佳实践 1：某零售旗舰借力 QingStor™ NeonSAN 助推关键业务上云.....</i>	13
<i>最佳实践 2：某 500 强保险集团利用 QingStor™ NeonSAN 打造核心业务存储引擎</i>	14

前言

云计算、大数据以及人工智能等新兴技术正使得企业数据发挥出更多价值，给企业业务带来强大的驱动力，企业用户越来越多地关注并希望利用新的 IT 技术以辅助商业决策。在这一趋势下，企业 IT 系统的角色由业务支持部门开始转向业务支撑部门，但传统基于 IOE（即以 IBM 小机、Oracle 数据库和 EMC 存储为代表）架构的数据中心已经很难适应数据化时代的需求，基于 x86 平台的云计算等新兴技术开始进入企业数据中心，促使企业 IT 开始内部革新。

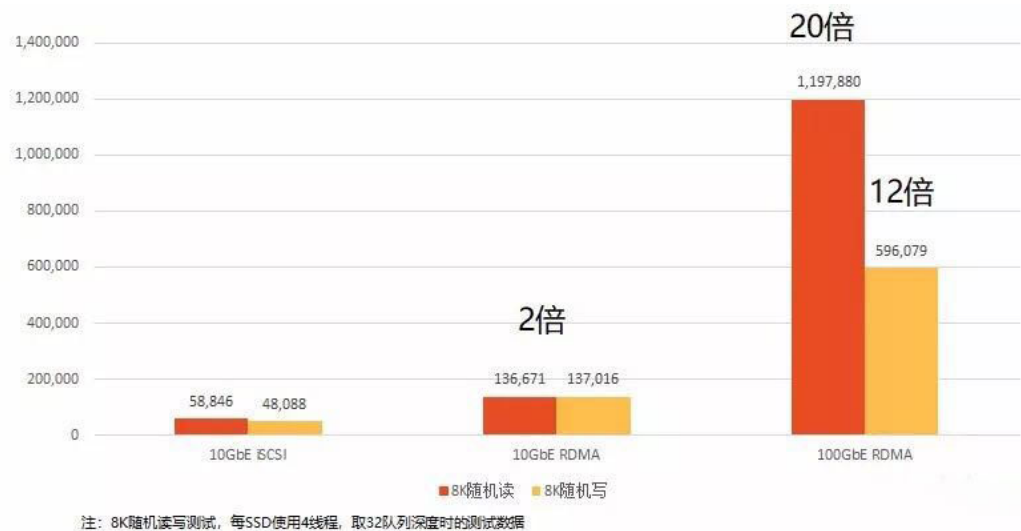
经过十几年的发展，x86 平台的计算性能和 RAS 特性（Reliability, Availability and Serviceability，可靠性、可用性与可维护性）都获得极高的成就，甚至超过了以 IBM Power 为代表的小机。2011 年的“去 IOE”风暴以及之后的企业实践，都证明了 x86 平台已经取代小机成为企业 IT 新的选择。

“E”所代表的传统 SAN 存储也正在被 SDS 的分支——Server SAN 取代。SAN 存储也有 RAS 特性，其中 S 除了可维护性（Serviceability）之外，还应该具有可扩展性（Scalability）。应用对存储的性能需求日趋增高，硬盘性能提升缓慢，传统 SAN 存储只能靠不断增加硬盘数量来满足，支撑关键应用的高端存储动辄支撑数千块硬盘都是很常见的配置。

但在 SSD 被引入企业级市场之后，SAN 存储的这种 Scale up（纵向扩展能力）几乎失去了实际意义——Scale up 扩展只能增加容量，而有助于性能；而 Server SAN 的 Scale out 扩展方式则让性能与容量线性增长。高端 SAN 存储具有一定的 Scale out 扩展能力，但控制器间所依赖的特定物理通路限制了其扩展规模——物理空间受限，且多控架构复杂。这也是为何高端存储通常只能最高支持 8 个控制器的原因之一。

SAN 存储的可靠性通常依赖特定硬件实现，例如双端口硬盘，双控制器是标配。而分布式架构的 Server SAN 软硬件解耦，通过多副本机制实现数据保护，并不依赖于特定硬件。软硬件解耦另一大好处就是，Server SAN 可以更快速地利用新的硬件技术，比如性能更好的 NVMe SSD；虽然 SAN 存储也可利用 NVMe 规范中的双端口功能以支持 NVMe SSD，但成熟应用需要时间，而这也意味着 Server SAN 的性能将有望超过 SAN 存储。

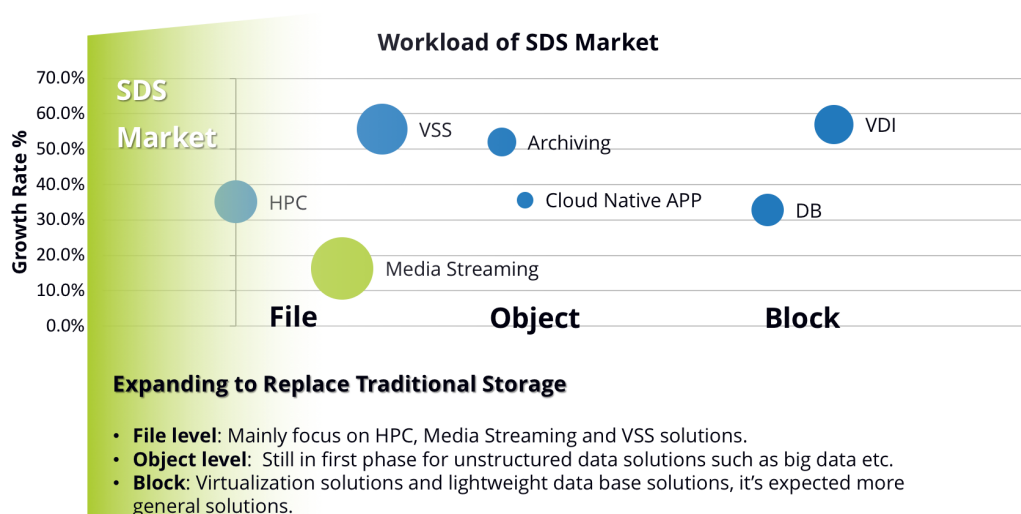
Server SAN 使用与计算同构的以太网络，这意味着能够随着通用以太网的升级而性能提升；而 SAN 存储则通常使用 Fibre Channel、InfiniBand 等专有网络。随着新一代支持 RDMA 的 25GbE（即 RoCE 或 iWARP）网络的普及，Server SAN 不仅在性能上提升迅速，并可便捷地升级到 100GbE、200GbE 乃至 400GbE 网络（常用 FC 网络规范要到 2020 年实现 128Gb/s），同时还能享受到以太网的规模经济效应。



相对于 10GbE iSCSI，支持 RDMA 的 100GbE 网络在存储应用上有近 20 倍的性能提升

正是基于 Server SAN 的架构优势以及潜在能力，让其逐步开始替代传统 SAN 存储。据 IDC 的统计数据显示，2017 年，VSS、流媒体、云原生应用、归档、VDI 以及 DB 等应用场景成为 SDS 产品的主要增长领域。其中 VDI 和 DB 都属于块数据存储场景，而块数据存储场景在以往通常是传统 SAN 存储的“天下”，但 SDS 在这两个场景的增长分别超过了 60%和 30%。这意味着 Server SAN 正在逐步取代传统的 SAN 存储。

China SDS Workload Analysis,2017

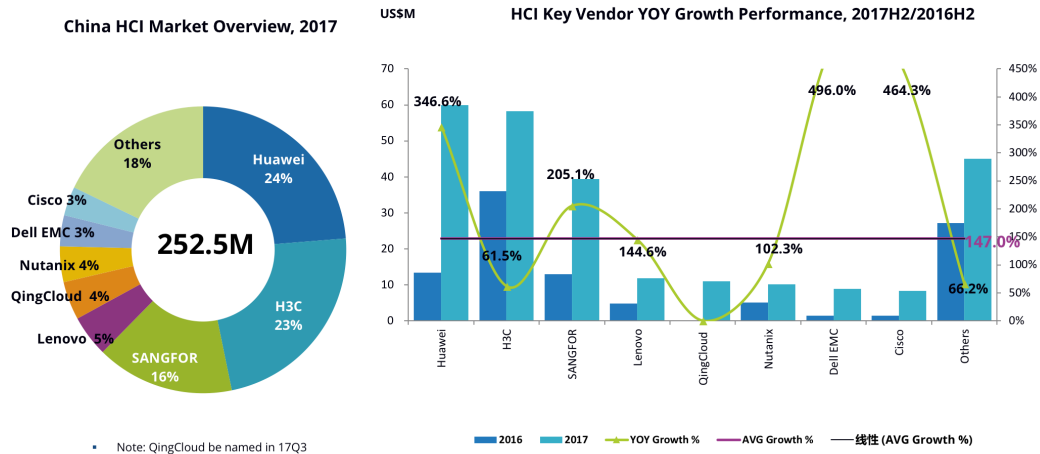


2017 年，SDS 在中国市场上的主要增长领域（数据来源：IDC）

旺盛的市场需求正促使诸多新兴供应商加入到 HCI 市场当中，HCI 以一体化交钥匙的方式交付给用户，大幅简化了部署实施过程。而 HCI 重存储的特点也让在 SDS 领域有经验积累的厂商能够迅速加入进来，其中较为典型的例子就是青云 QingCloud。青云 QingCloud 原本是一家公有

云服务提供商，在 2017 年提出了“全模云”概念，同时支持分布式和集中式业务架构的云端部署，希望根据不同用户不同应用场景下的不同需求提供灵活且适宜的产品或组合解决方案，推出了包括 Bare Metal(裸金属服务器)、QingStor™ NeonSAN 以及 RadonDB 等多种满足企业用户所需的产品及服务。其中 QingStor™ NeonSAN 是面向企业关键应用负载所推出的一款软件定义分布式块数据存储产品(Server SAN)，同时也可以作为其青立方超融合系统的底层多维存储选择之一。

China HCI Market Key Vendor Analysis, 2017H2



青立方超融合系统从 2017 年第三季度才开始被 IDC 纳入统计，近下半年两个季度，就占据了 4% 的中国地区市场份额，即使放到全年来看，青云 QingCloud 占据了 3% 左右的市场份额(数据来源：IDC)

QingStor™ NeonSAN 架构解析

青云 QingCloud 自主研发的 QingStor™ NeonSAN 是一款软件定义分布式块存储系统，其支持 HDD、SATA SSD、NVMe SSD 乃至 Intel Optane 等多种存储介质，并可选 10GbE 或新一代 25GbE 网络，通过不同的硬件组合，能够为不同应用场景下的不同存储性能需求提供灵活选择；同时也能根据应用特点融合部署在虚拟化环境（HCI），或者分离部署在商用 x86 服务器之上（Server SAN），Oracle RAC、SQL Server 等核心数据库集群、物理机和容器高可用架构、大数据分析计算等应用场景都有相匹配的 NeonSAN 产品。

经过 E 企研究院的测试结果显示：在采用 NVMe SSD+25GbE RDMA 技术后的 NeonSAN，其单卷（Volume）可提供不低于 10 万 IOPS 的随机读写性能，容量可扩展到 100TB，单个 NeonSAN 节点随机读写性能最高可达 50 万 IOPS。

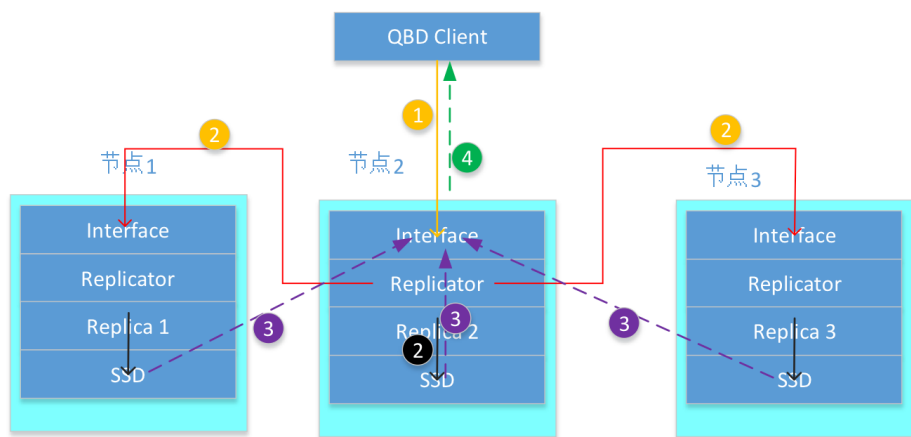


QingStor™ NeonSAN 逻辑架构图

QingStor™ NeonSAN 主要有四个大的功能角色组成：

- Zookeeper 用于集群管理，进行注册发现服务，新加入的 NeonSAN 节点会自动注册到 Zookeeper；
- 中心节点(Center Node)用于 NeonSAN 的控制节点，类似 SDS 中的控制平面(Control Plane)，通过一主多从模式来实现高可用，当主控制节点不可用时，通过 Zookeeper 可以自动产生新的主控制节点；
- 元数据信息存放在 MySQL Plus 数据库，MySQL Plus 是一个分布式数据库集群，可以分布到所有 NeonSAN 之上，主控制节点可自动访问 MySQL Plus 以获取元数据信息；
- Store Node 主要用于数据服务，类似 SDS 中的数据平面（Data Plane）。

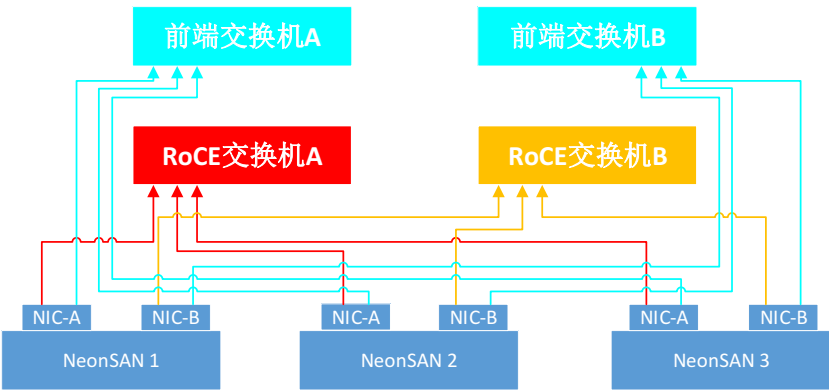
NeonSAN 通过多副本机制来保证数据的高可用，副本数量可以设置为一、二、三或更多（默认为三），接收到应用服务器的写请求之后，其在写入主副本的同时，Replicator 进程会同时将数据副本发送到其他 NeonSAN 节点，当主副本与从副本都写入成功之后，再返回应用服务器写入成功，这种机制保证了数据各副本之间的强一致性。



NeonSAN 数据写入流程图，在全闪存阵列中，NeonSAN 内部存储网络使用支持 RDMA 技术的新一代 25GbE 网络，利用 RDMA，数据本地写入与远程节点数据写入拥有几乎一致的响应时

间，这可大幅降低 NeonSAN 数据写入延迟

除了数据存储的高可用之外，存储系统还需要关注存储网络的高可用。与传统 SAN 存储的双控制器或多控制器来实现高可用不同，QingStor™ NeonSAN 的存储网络高可用是通过交换机和数据链路的冗余来实现高可用，如下图所示：



QingStor™ NeonSAN 网络架构拓扑图，以三节点全闪存 NeonSAN 集群为例

每个 NeonSAN 节点上都配备了 2 张双端口网卡，每个端口都连接到不同交换机上。其中互为冗余的两台前端交换机作为服务网络，即为前端应用服务器提供数据读取服务；两台支持 RDMA 技术（即 RoCE，RDMA over Converged Ethernet）的 25GbE 交换机作为数据网络，主要用于 NeonSAN 的多副本写入等数据交换操作。

作为支撑企业关键应用的数据存储解决方案，NeonSAN 在注重高可用性的同时，也致力于提供更高的性能。因此，在 NeonSAN 开发过程中，简洁是第一原则，所有存储软件堆栈都以“简单”为第一优先级以充分发挥 NVMe SSD 的高性能优势。

除了对 IO 路径进行优化之外，NeonSAN 还提供了诸多企业级存储增值特性，比如标准 iSCSI 接口支持，和最新的 NVMe over Fabric 规范支持，以及异步复制、数据加密、多路径故障切换、无中断的数据恢复、迁移和容量均衡等功能。

NeonSAN 通过多副本与多数据链路的机制达到数据存储与服务的高可用，利用 NVMe SSD+25GbE (RDMA) 来提升存储性能，辅以“简洁”的存储软件堆栈最大限度发挥 NVMe SSD 的高性能优势，并支持多种企业关键应用常用功能来满足不同应用场景下的存储功能需求，这一系列特点已经让 NeonSAN 具备了替代传统 SAN 存储的基础。目前，青云 QingCloud 正在引入 Intel SPDK 技术，以进一步提升 NeonSAN 性能。

E 企研究院验证：QingStor™ NeonSAN 性能评估

E 企研究院根据企业关键应用的特点，使用 x86 服务器、Intel NVMe SSD 和 25GbE 构建了 Oracle RAC 数据库环境，用以评估 QingStor™ NeonSAN 在 OLTP 中的性能表现。

压力客户端

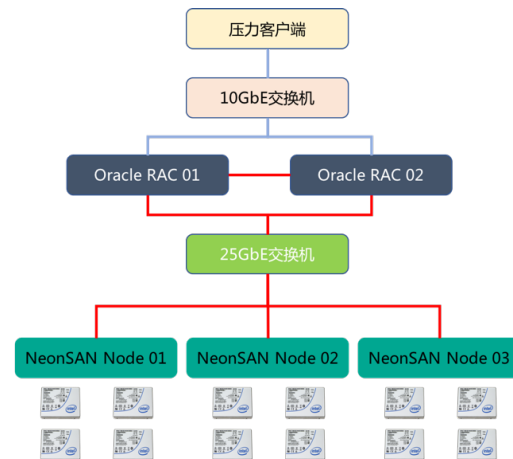
- Intel(R) Xeon CPU E5-2630 v4 @ 2.20GHz x2
- Samsung DDR4 2400MHz 16G x4
- Mellanox CX4 10G x2

Oracle RAC 配置

- Intel(R) Xeon CPU E5-2698 v4 @ 2.40GHz x2
- Micron 2133MHz 32Gx4
- Mellanox CX4 25G x 2

NeonSAN 配置

- Intel(R) Xeon SP Gold 6140 CPU @ 2.30GHz x2
- Samsung DDR4 2666 MHz 16Gx4
- Mellanox CX4 25G x 2
- Intel P4510 8T x 4



E 企研究院评估青云 QingStor™ NeonSAN 性能所构建的 Oracle RAC 架构及组件

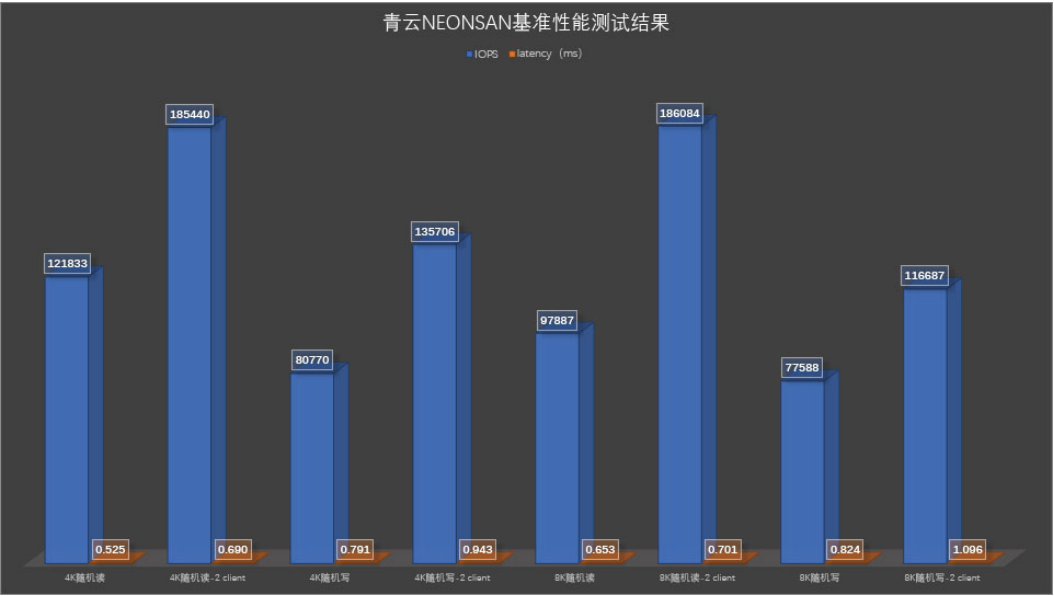
在此测试环境中，E 企研究院使用 2 台配备两颗 Intel Xeon E5-2698 v4 处理器的双路服务器构建 2 节点的 Oracle RAC 数据库应用环境，并安装 Oracle 12c 数据库软件（包括 Grid 和 Database）。

使用三台配备两颗 Intel Xeon SP Gold 6140 处理器的双路服务器作为 NeonSAN 存储节点，每节点上使用 4 片 8TB 容量的 Intel DC P4510 SSD 作为数据存储。Intel DC P4510 SSD 是 Intel 最新发布的采用 3D NAND 技术的 SSD 和新一代的闪存控制器的 SSD 家族，其最大容量可到 8TB，结合企业数据中心优化的固件，能够给企业关键应用存储带来极高的性能和稳定性。

基准性能测试：QingStor™ NeonSAN 契合分布式应用

在进行正式测试之前，E 企研究院先对 QingStor NeonSAN 的基准性能进行测试，以检验其是否正确安装，获得的测试结果也为后续 Oracle RAC 数据库性能测试提供参考。

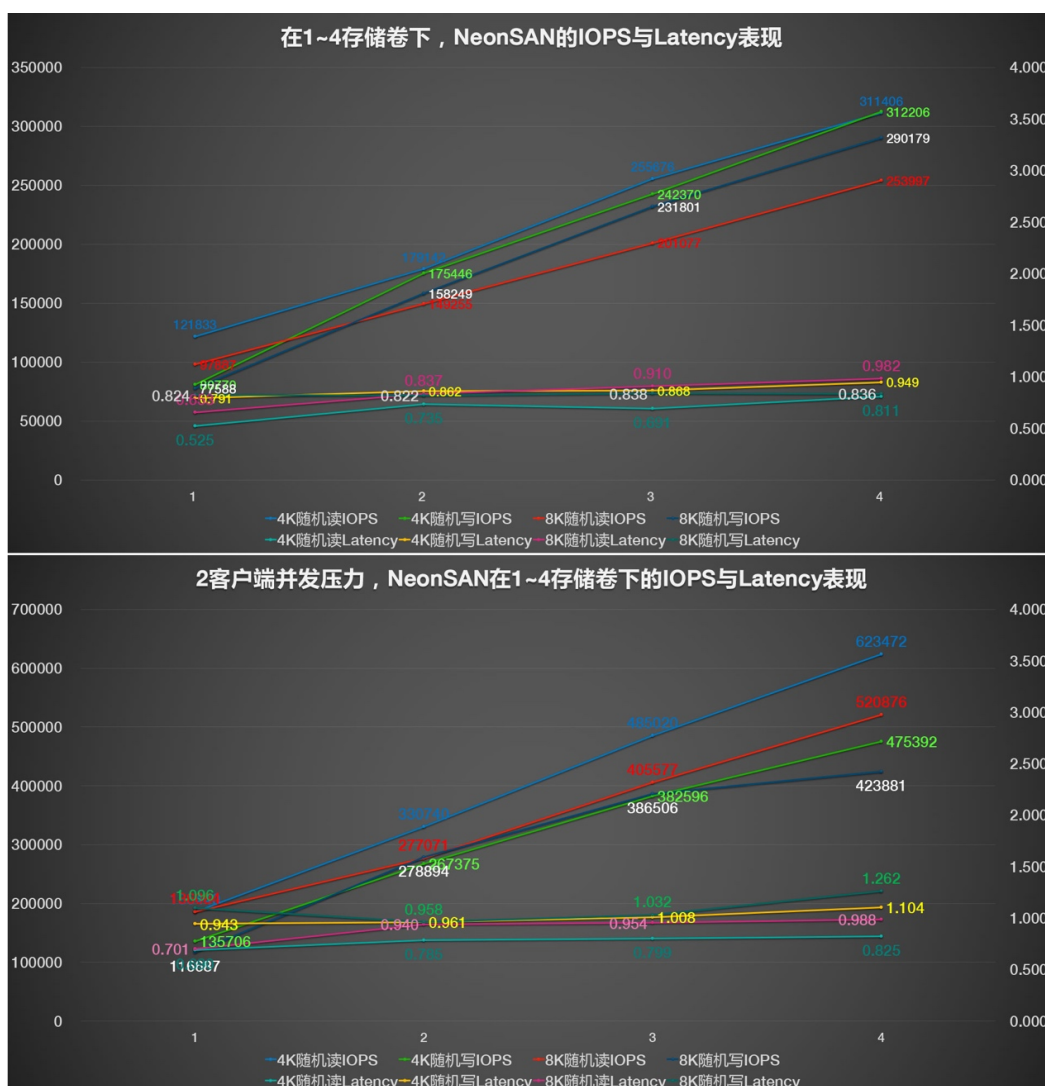
基准性能测试采用 4K 和 8K 两种尺寸的数据块进行随机读写测试，4K 的随机读写性能是业界用于评估存储性能的事实标准之一，8K 则是数据库类应用的常见数据块大小。并且，E 企研究院在评估了单一压力客户端下的存储性能之后，还增加了多台（2 台）压力客户端对同一个 NeonSAN 进行并发性能测试，以评估其在分布式应用环境下的存储性能表现，其结果如下图所示：



NeonSAN 在分布式环境下，4K、8K 随机读写 IOPS 与平均响应时间（Latency，ms）对比

NeonSAN 的 4K、8K 随机读写性能。从测试结果来看，在单台客户端下发压力情况下，4K 与 8K 随机读写性能都能接近或超过 10 万 IOPS，且平均响应时间在 0.5~0.8ms 之间，具有不错的性能。而在 2 台客户端并发压力下，其 4K 随机读写性能分别达到了 18.5 万和 13.6 万 IOPS，而 8K 随机读写性能也分达到了 18.6 万和 11.7 万 IOPS，相比单一客户端下，性能至少提升了 50%，而延迟增加则并不明显，仍保持较低水平。

同时，E 企研究院还评估了 NeonSAN 在多个存储卷下的性能增长情况，结果如下图：



在单个压力客户端（图上）和 2 个压力客户端（图下）测试环境中，QingStor NeonSAN 分别供应多个（1~4 个）存储卷下的性能（IOPS）和响应时间（Latency，ms）表现

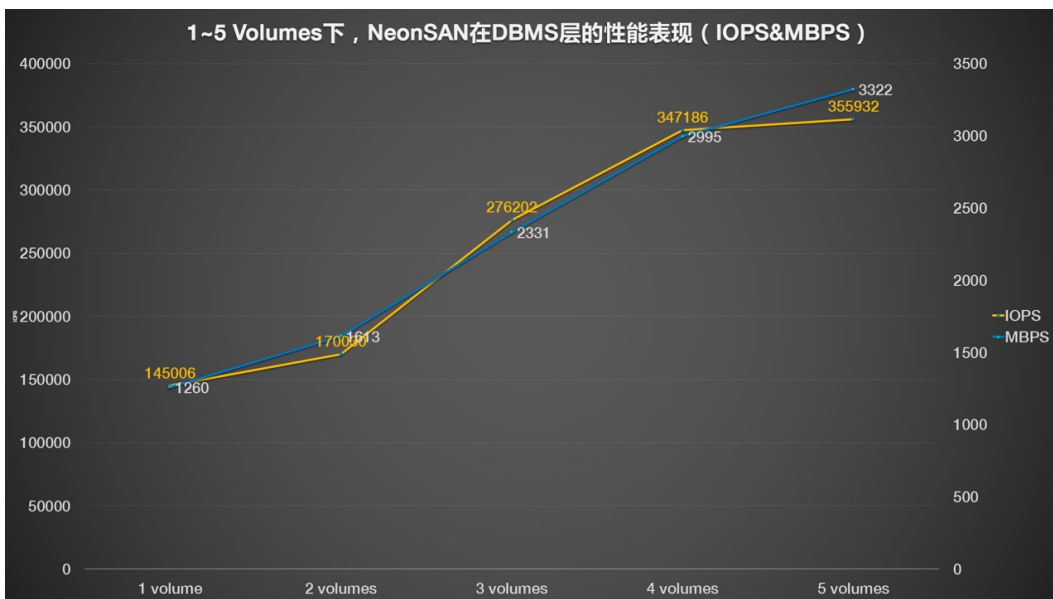
随着 NeonSAN 卷数量的增加，其性能线性增长，而平均响应时间的增长并不明显。在单台客户端性能测试中，在配置 4 个 NeonSAN 卷时，4K 随机读写和 8K 随机读性能均在 30 万 IOPS 左右，8K 随机写性能则能超过 25 万 IOPS，平均响应时间则均控制在 1ms 以内。

相比于单台压力客户端，在 2 台客户端并发压力下，4K 随机读写性能分别达到了 62 万和 52 万 IOPS，8K 随机读写性能则分别超过了 47 万和 42 万 IOPS。从测试结果可以看到，NeonSAN 在多个存储卷下均有 50% 的总体性能提升，这意味着 NeonSAN 能够很好地支持分布式应用环境。

Oracle RAC 性能测试：直面关键应用

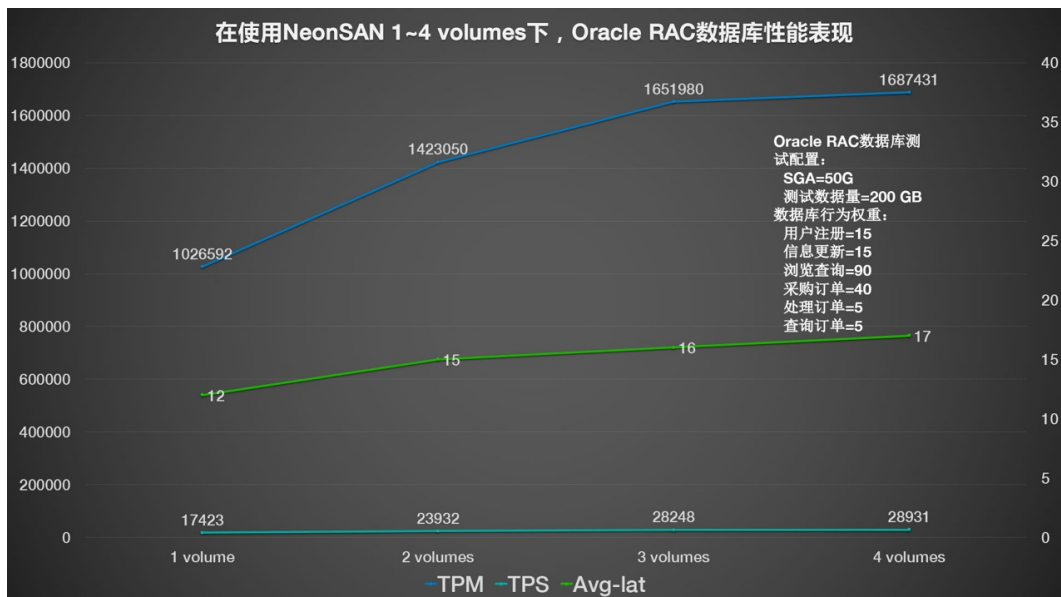
基准性能测试只能反应存储系统的性能表现 ,上述测试结果表明 QingStor™ NeonSAN 已经安装正常 ,可以接入已经构建好的 Oracle RAC 数据库应用环境中。但需要注意的是 ,在进行 Oracle RAC 数据库测试时 ,衡量的是包括 Oracle 数据库服务器、测试环境网络以及 NeonSAN 存储的综合性能表现 ,这三者任意子系统出现瓶颈 ,都会影响测试性能。

所以在正式测试之前 ,还需确认其中某些子系统的瓶颈 ,以明确可能影响 Oracle RAC 数据库性能的因素。E 企研究院利用 Oracle 数据库内嵌的一组命令来检验接入 Oracle RAC 环境后的 NeonSAN 性能。

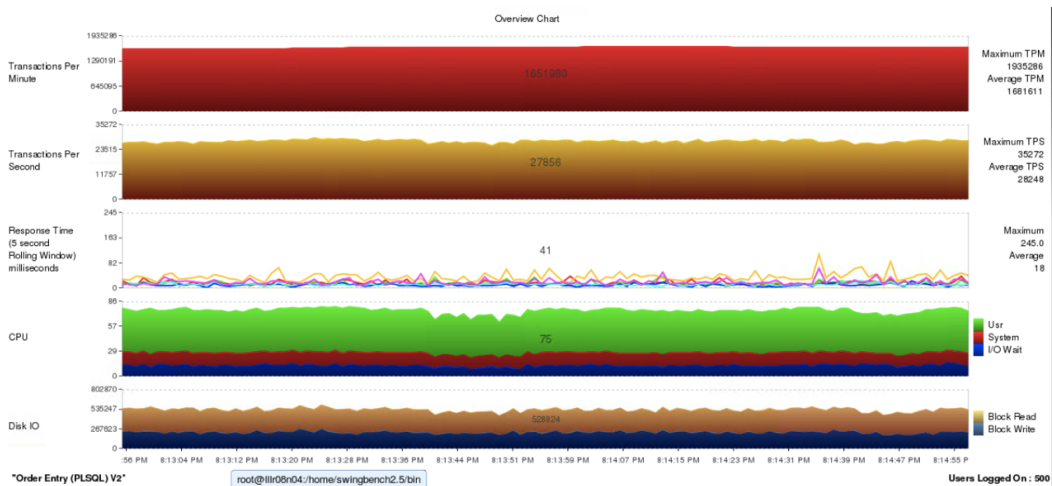


与之前基准性能测试不同的是 ,这一项测试是基于 RDBMS 层的测试 ,其通过模拟 Oracle 数据库的数据访问特点来对后端 NeonSAN 存储进行评估。上图显示在配置 4 个 NeonSAN 存储卷后获得最大接近 35 万 IOPS ,然后即使再添加 NeonSAN 存储 ,其性能几乎不变。而且顺序带宽则在 5 个 NeonSAN 卷时获得最高 3.3GB/s 吞吐量。

通过上述测试表现 ,在接入 2 节点的 Oracle RAC 环境之后 ,青云 NeonSAN 在提供 4 个存储卷时获得最大随机读取性能。因此 ,根据这一结果 ,E 企研究院为 Oracle RAC 数据库环境共分配 4 个 1TB 容量的存储卷 ,分别测试其在 1~4 个卷下的 Oracle 数据库性能表现 ,具体如下图 :



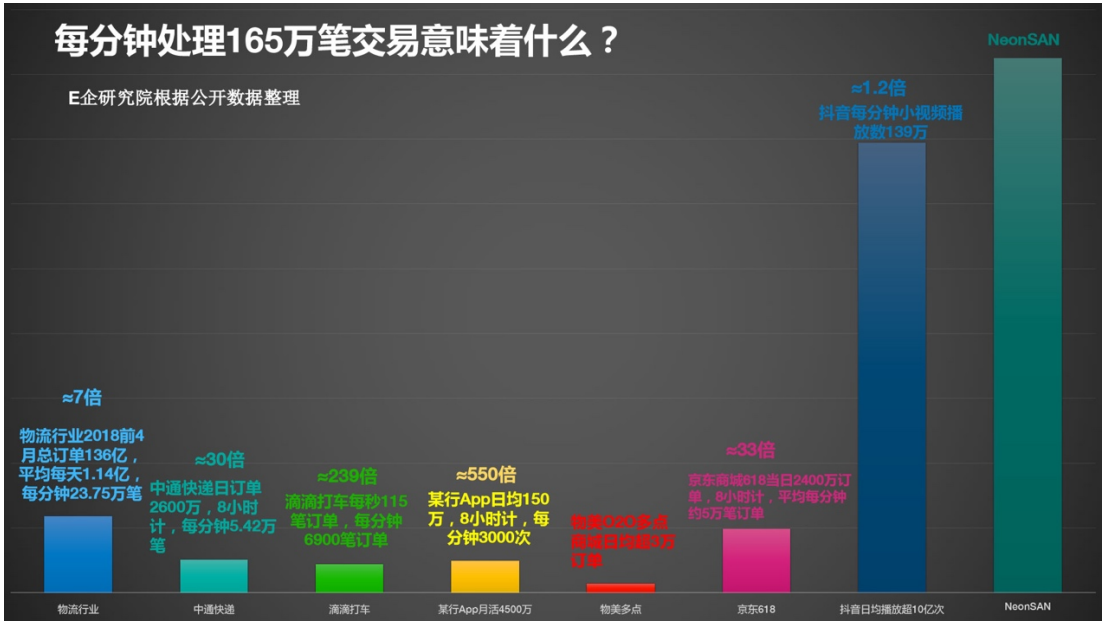
在 Oracle RAC 数据库测试中，在配置 3 个卷作为 Oracle 数据库存储时，获得超过 165 万 TPM（Transactions Per Minute，每分钟事务处理数），但再增加一个 NeonSAN 存储卷，数据库性能并没有增加，这意味着在 3 个 NeonSAN 卷时，Oracle RAC 应用的其他子系统已经出现瓶颈，导致尽管存储性能增加，但应用性能并没有增加。结合上一个测试结果，此时瓶颈可能出现在计算方面，即 Oracle 数据库服务器的计算能力达到瓶颈。



通过监控画面可以看到，Oracle 服务器的 CPU 平均利用率已经超过 75%。结合以往 E 企研究院的测试经验来看，通常情况下，服务器 CPU 利用率超过 70%就意味着已经达到最佳计算性能，即使再增加负载，CPU 占用率可能会进一步提升，但就应用来看，其性能并没有增加，反而延迟可能会出现不利影响。因此，从测试结果来看，在配备 3 个 NeonSAN 卷做 Oracle 数据库存储时，整个 Oracle RAC 环境就已经达到最高性能，而在此环境中，Oracle 服务器配备的是 Intel Xeon E5-2698 v4 处理器，要提升计算性能就只能更换到最新一代 Xeon SP 系列中的顶级处理器，或者使用配备更多 CPU 的四路及以上服务器。

即使受到 Oracle 数据库服务器计算性能的影响，在此 Oracle RAC 数据库环境中，也取得了超过 165 万 TPM 的性能，即每分钟能够完成超过 165 万笔事务处理，这之中包括新增商品的添加、用户的浏览、订单的处理、修改以及订单查询、追踪等等常见操作。平均到每秒的事务处理量为 3 万笔，且每笔事务处理的响应时间不超过 15ms。对于企业核心数据库而言，NeonSAN 不仅具有超高的性能，同时还具备极低的延迟，能够为需求苛刻的关键业务应用提供强有力的支撑。

每分钟 165 万事务处理量意味着什么？E 企研究院收集了目前较为热门的相关行业或新秀公司，根据其公开资料/数据整理，以与 NeonSAN 的 Oracle RAC 数据库性能对比，具体如下图所示：



常见应用 App 平均日订单量一览

虽然上图列出的订单量与数据库性能之间并不能划等号，一笔订单背后可能会有包括用户登录、用户信息修改、产生订单以及查询等多个操作，但仍能提供部分参考建议。比如据国家邮政局的最新统计数据显示，今年（2018年）前4个月全国快递业务量累计完成136.7亿件，平均到每天的快递业务量约为1.14亿件，按8小时计算，相当于每分钟会产生23.75万件快递，相当于23.75万笔订单，每个订单背后按5个数据库操作计算（即1订单需要5数据库性能作为支撑），那么大约需要120万的数据库性能，即120万TPM。

而通过 E 企研究院搜集的数据简单计算的话，三节点（全 NVMe SSD+25GbE）配置的 NeonSAN 存储集群能支撑 Oracle RAC 数据库达到 165 万 TPM 性能，均能支撑上述应用数据库所需的存储性能，并有富余，可应对更高的业务峰值。

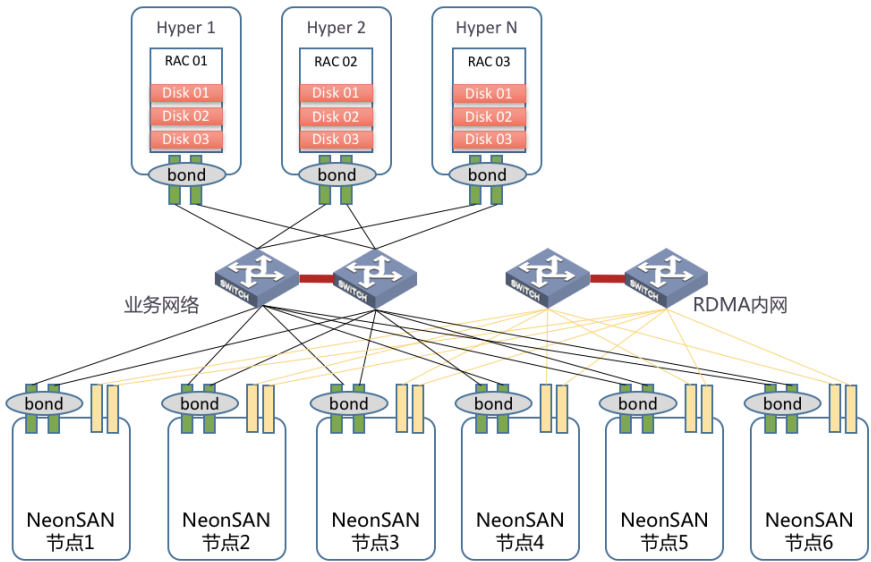
尽管因为具体到应用可能有不同的流程，会对后端存储产生不同程度的性能需求。但 E 企研究院认为，QingStor™ NeonSAN 完全有能力满足绝大部分企业应用对块存储性能的需求。测试表明，QingStor™ NeonSAN 不仅是一款优秀的 Server SAN 产品，符合 SDS（Software Define Storage，软件定义存储）的软硬件解耦、高可用、高可靠、弹性以及无厂商锁定等特点，同时借

助 NVMe SSD+25GbE 网络技术，能够提供极高的性能和极低的 I/O 响应时间，完全有能力满足企业关键应用负载提出的苛刻性能需求，助力企业向混合云迈进，为数字化转型奠定坚实基础。

最佳实践 1：某零售旗舰借力 QingStor™ NeonSAN 助推关键业务上云

在互联网+的背景下，数字化已经成为新零售的核心驱动力。某零售公司基于青云 QingCloud 全栈云计算解决方案构建了集团云平台，为旗下各业务单元提供敏捷、高效、简化管理、成本可控云计算服务，包括测试云与生产云，IaaS 层组件与 PaaS 与 SaaS 层服务。

在将大量企业核心 ERP、Oracle RAC 业务在向云端迁移时选择了 QingStor™ NeonSAN 作为其后端存储，现有平台无须做任何改造，可平滑添加 NeonSAN 存储，并通过云平台进行统一管理。同时，原有业务环境的 Oracle RAC，仅需添加 NeonSAN 作为共享盘，进行业务数据的迁移，无须其他复杂的配置工作。



某零售核心 Oracle RAC 业务架构拓扑图，后端采用 6 节点 NeonSAN 存储集群

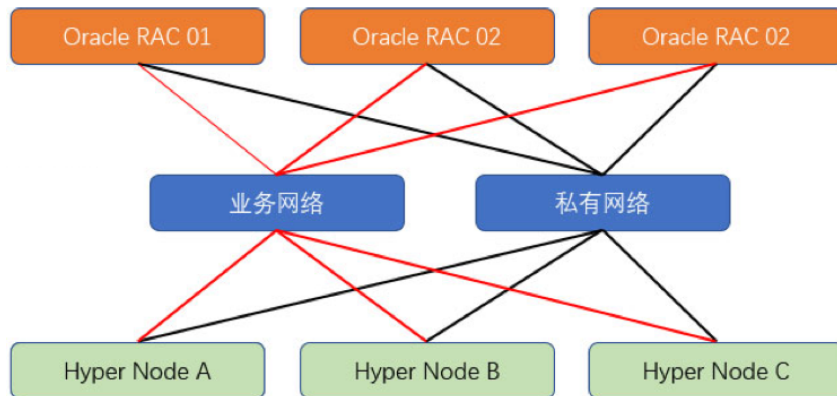
客户收益

- 通过企业级分布式存储方案，推进核心 ERP 系统向云计算架构的演进，实现私有云环境下的一体化运营与管理，在保障业务连续性的同时大幅降低了采购和运营成本。
- NeonSAN 为高级别的可用性提供坚实的存储平台，可避免组件故障、网络故障或者人为错误造成的业务中断现象；与 Oracle RAC 无缝集成，实现应用程序和数据的长期稳定运行，让客户安枕无忧。
- 通用开放的 X86 平台取代专用的存储设备，使得存储系统的建设和扩容周期从几个月缩短至一周，有效满足业务数据量激增的扩容需求，推动业务系统的平滑快速地发展。

最佳实践 2：某 500 强保险集团利用 QingStor™ NeonSAN 打造核心业务存储引擎

区块链，人工智能，大数据等技术的迅猛发展给互联网保险带来消费场景、产品形态和营销方式三大维度的颠覆，推动互联网保险向新技术密集型经营模式转型。在互联网保险业务中，IT 系统已经深入到产品开发、销售、服务等流程的各个环节，成为业务发展的核心驱动力，需要具备高效的业务支撑能力与高投资回报率。

某保险企业自开展互联网业务以来，在线保险业务年均增长 100%以上，原有数据库一体机的物理硬件资源趋于饱和，不再支持扩容与升级。为了匹配将来的业务发展，该保险企业决定采用数据库自 Oracle 一体机向 x86 平台迁移的方案：将 Oracle RAC 数据库迁移到物理机，使用 NeonSAN 集群作为后端存储，提供数据库支持服务，配置三副本的数据保护机制。



某保险 Oracle RAC 数据库应用架构图，后端使用 3 节点 NeonSAN 集群，为部署在 3 台物理机上的 Oracle RAC 数据库提供存储服务

客户收益

- 经过业务场景实测，基于 NeonSAN 的复杂视图查询响应时间从 20 分钟以上缩小到 2 分钟，精算准备金复杂 SQL 执行效率从分钟级缩小到秒级，实现效率 100% 的提升，可确保所有类型的数据处理都能够实现卓越的性能。更快的查询速度和更快的报表运行速度，帮助泰康推动业务决策的制定、提高准确度并获得高品质成果。
- NeonSAN 为最高级别的可用性奠定坚实的存储平台，可避免组件故障、网络故障或者人为错误造成的业务中断现象。与 Oracle RAC 搭配，应用程序和数据可实现最长的正常运行时间，让客户安枕无忧。
- NeonSAN 分布式架构，使存储扩容时间从几个月提升至几天，提升速度高达 10 倍，有效满足业务数据量激增的扩容需求，推动业务系统的平滑快速地发展。