分布式存储全共享架构(Shared-Everything)分析与研究

《百易存储研究院-少数派报告》（1）：

分布式存储全共享架构（Shared-Everything）分析与研究

全共享架构（Shared-Everything）设计

全共享架构（Shared Everything）设计是分布式存储在软件系统架构设计方面的最新进展，2023年11月由XSKY星辰天合对外发布，并命名为星海（XSEA，eXtreme Shared-Everything Architecture，极速全共享架构））架构，率先在星飞（XINFINI）全闪分布式存储产品中应用。

全共享架构是软件架构中，每个元数据控制器节点上的ChunkServer（数据持久层服务），都可以直接访问所有存储节点上的SSD，从而提高了数据访问速度和灵活性。

在此，我们将全共享架构称为新的架构，而将此前的架构称为旧架构；将采用新架构的分布式存储产品称为新的产品，而将采用旧架构的分布式存储产品称为旧产品。

旧架构产品以通用x86服务器为基础，一台x86服务器就是一个存储节点，新旧架构都是如此。不同之处在于其元数据管理器中的ChunkServer设计，旧的架构，ChunkServer与存储节点紧耦合，只能够管理调度本地盘，对于他存储节点本地盘的访问，需要通过其他存储节点上的元数据管理器进行。ChunkServer不了解非本地盘的情况，不能够直接进行管理。

相比旧的架构，新架构最大的变化在于ChunkServer，每个存储节点的ChunkServer都可以管理全局管理所有的存储盘，从而超越了本地管理的局限。在新的架构中，每个存储节点元数据管理中的ChunkServer的功能，或者说角色，都是完全相同的，对等的。

新的产品可以高效解决数据和缓存一致性问题，减少以往旧机构存储节点之间的cross talk影响，从而极大降低了系统复杂度，提供系统的响应速度。

这里需要交代的技术背景是，新的架构之所以能够成立，完全有赖于NVMe SSD，RDMA网络、硬件压缩、NVMe OF等新技术的出现，以及商业化程度的完善，特别是成本、价格上的成熟，为新架构的诞生，奠定了物质的基础。

新架构带来的主要价值

考虑到技术上复杂性，特别是对于广大的企业级用户而言，元数据管理器、ChunkServer、NVMe SSD、RDMA、NVMe OF等技术理解还是有一定的门槛，特别谁对非常专业人员而言，理解分布式存储都有一定的困难。

但是这也无妨。我们只要了解，分布式存储是由通用x86为基础构建起来的存储，有新、旧架构之分，其中，新架构指的就是采用全共享架构（Shared Everything）设计的分布式存储新产品。

我们可以不了解新架构新在什么地方，但是一定要了解新架构，解决了旧架构哪些不能够解决的问题，及其带来的创新价值。

总结起来，新架构产品的价值有以下几点：

首先是新架构产品能够将存储访问控制到100微秒超低延迟的水平（旧架构产品中，少数优秀产品业能够达到这个水平，因此，非新架构独有。）

其次，新架构解决了旧架构普遍存在的慢盘、亚健康网络的问题，也就是所谓P95、P99问题，简单说，就是分布式存储系统在95%、99%的状况下可以达到设计指标，但是总有5%或者1%的情况下，会达不到设计性能，出现所谓的性能抖动。

这些性能抖动就是由慢盘、亚健康网络状况所造成的，具有不可预测性。所谓慢盘、亚健康网络并非故障，而是随时都有可能出现，也有可能随时消失，这种不确定性，对于很多关键业务应用是完全没有办法出现的。

新的架构完美解决了p95、P99的难题，将存储系统响应的下限，稳定控制在100 毫秒内的水平。也就是说，无论何时、何地，能够确保存储系统的响应延迟控制在100 毫秒～100微秒的区间。

最后，通过全局 EC （如 20+4，）和数据压缩技术可以实现超过 100% 的存储系统得盘率（可用容量/裸容量）。这个表述意味着，相同的性能、容量设计，新架构的产品在成本上将及其具有竞争力（考虑到商业因素，这里没有办法给出定量的数据，用户可以进行比对，但是可以肯定的是，新架构带来的竞争力是及其显著的）。

未来市场前景分析和预测

新架构带来的竞争优势是显而易见的。

但是新架构技术要转化为市场竞争优势会存在很大的不确定性。因为市场不是完全由技术来决定的，会受市场推广、品牌、成熟度等很多因素的影响。分布式存储全共享架构（Shared Everything）也是如此。

从技术上说，新的架构绝对领先，突破了以往结构性的一些难题，并且在成本上优势明显。但要把这些优势转化为市场上的胜势，还会有很多不确定的因素，所谓谋事在人，成事在天。但是尽管如此，我们还是看好新架构的潜力和能力，看好其颠覆旧架构，引领未来的可能性！

一切皆有可能！

百易存储研究院出品

2024年1月