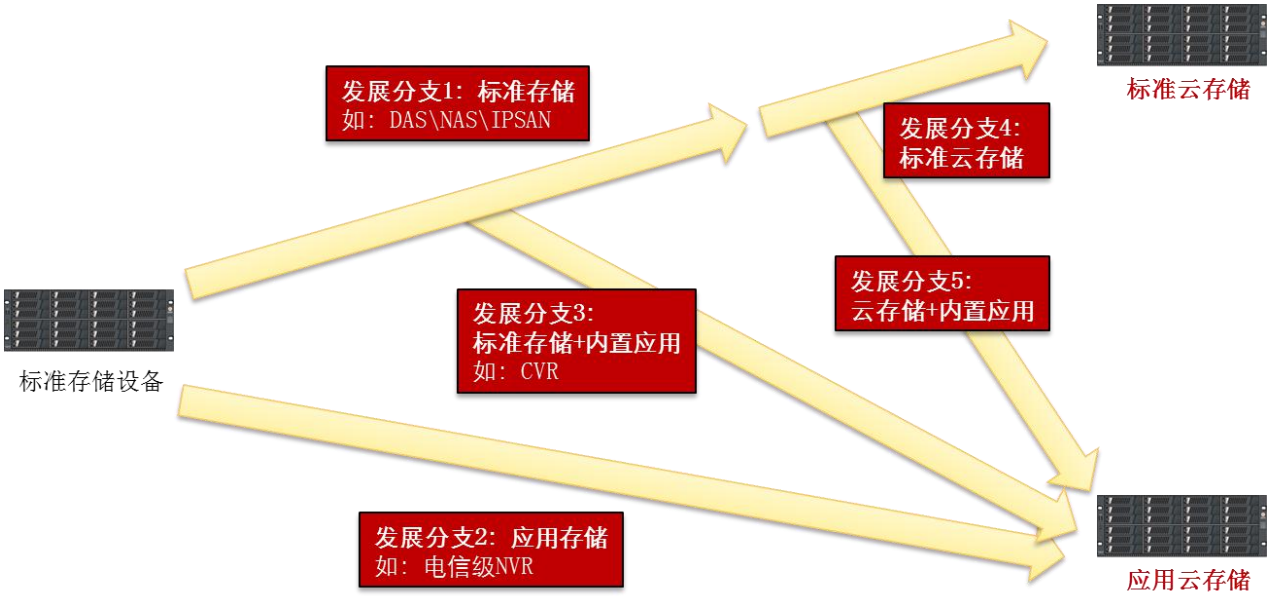
视频监控应用云存储节点

高清监控的浪潮正向人类社会席卷而来。高清晰度图像对传统视频监控系统产生了巨大的冲击，采集、传输、存储与计算的数据量都出现了爆炸似的增长。其中，存储作为视频监控必不可少的组成部分，在平安城市建设如火如荼的当今受到了前所未有的重视。

纵观全球安防企业，几乎都不拥有存储设备的研发和定制能力。而拥有存储设备研发与定制能力的大型 IT 企业却鲜有进入安防行业，仅是以提供标准产品设备的方式参与其中。所以，在行业内一直没有出现针对视频监控业务特性来设计的专业存储产品。针对不同的行业，在通用产品的基础上设计出有行业特性的产品，是行业需求的必然导向，也是所有企业能够维持持续发展的必经之路。从下图可以看出，随着 IT 存储技术的发展，在不同的标准数据存储中添加行业属性，是大势所趋。



图中最主要的概念就是应用存储，在这里也就是指为视频监控业务定制的存储。

业务连续性

视频监控存储与传统数据存储最大的区别就是业务连续性。我们通过两个简单的例子来体现它们之间的区别。

应用场景一：银行业务数据

如果数据发生了故障，储户的业务是必须中止的。至到数据完全正确无误的被恢复后，储户才能操作账户进行存取操作。日常生活中，我们经常见到左图中的“暂停办理业务”的告示牌。有两种情况，一种是柜员人数不够，第二种也就是说，系统数据可能出现了错误。当然，银行是一个比较极端的应用，其数据通常会保存几份。因为数据的可靠性和完整性的重要级别是第一的，用户的业务是可以中断的。这就是在存储领域中的“标准数据存储”。

应用场景二：在家看电影

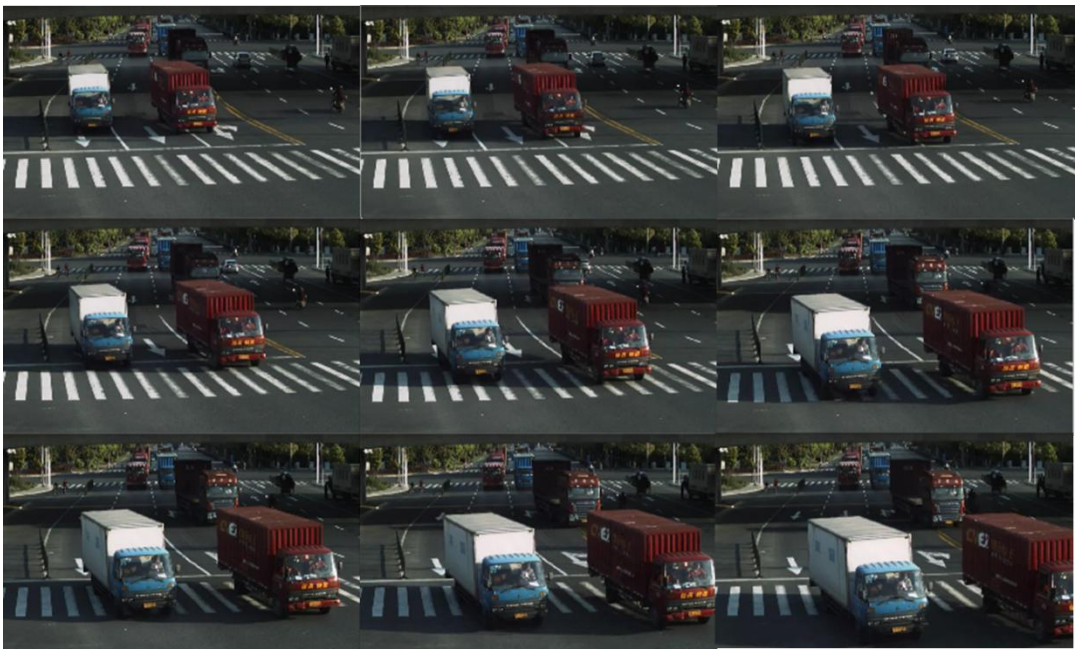
试想一下我们正在家里的家庭影院中观看一部好莱坞大片，突然间电影画面花了一下或者卡了一下，我们是把播放器关掉不看了还是继续观看呢？相信所有人都会做出同样的决定，容忍那一段错误，继续观看。

这是一个典型的不同于银行场景中标准数据存储的的应用场景。华为以存储厂商的视角对视频监控行业进行了长达 6 年的研究，发现传统方案中，因为几块硬盘的故障损坏就使整个视频监控系统停止工作，这是完全不符合客户使用场景的。针对这类保持业务连续性为第一重要级别，数据可以容忍少量错误的场景，华为推出了拥有视频监控业务特性的“应用存储”。当然，保障业务连续性仅仅是华为视频监控应用云存储的特点之一，后面将会详细阐述几大特点。

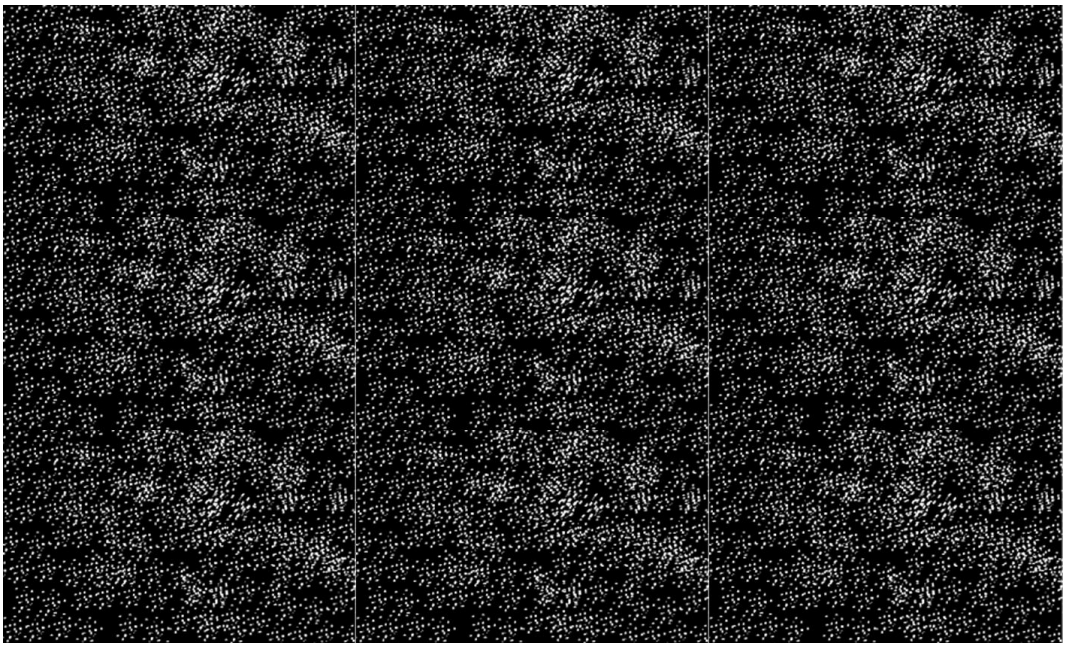
为了保障上述的业务连续性，华为在存储设备的底层做出了一些特殊的设计。在传统存储设备中，多块磁盘需要组成 RAID 磁盘组以增加数据的可靠性。例如 RAID5 算法可以支持在一块磁盘故障损坏后数据不丢失，可以通过检验数据还原业务数据。通常来说，在一台 16 盘位的磁盘阵列中只会建立 1 个 RAID 组，那么可用的数据空间为 14 盘位。按照主流的 3TB 磁盘来计算，将可以保存 70 路标清摄像机 1 个月的视频图像。此时，如果在一个 RAID 磁盘组中有两块以上磁盘损坏，那么很不幸，整个 RAID 组的数据将全部丢失。

这是一个问题现象，经走访得知，大量公安体系的用户对此表现出极大的担忧。问题的原因是因为采用的 RAID 技术是传统存储厂商为数据可靠性为首的应用场景设计的。标准数据存储主要是设计用来保存结构化数据的。以上述 RAID 组为例，在存储时有可能会将文件切片为 14 片，每块磁盘上存储 1 片（实际可能每盘 N 片）。此时如果有几片数据丢失，那么整个文件就无法再恢复了，因为文件的特殊性，有缺失的部分，文件就无法使用了。

下图我们可以看到一个视频图像片段序列。

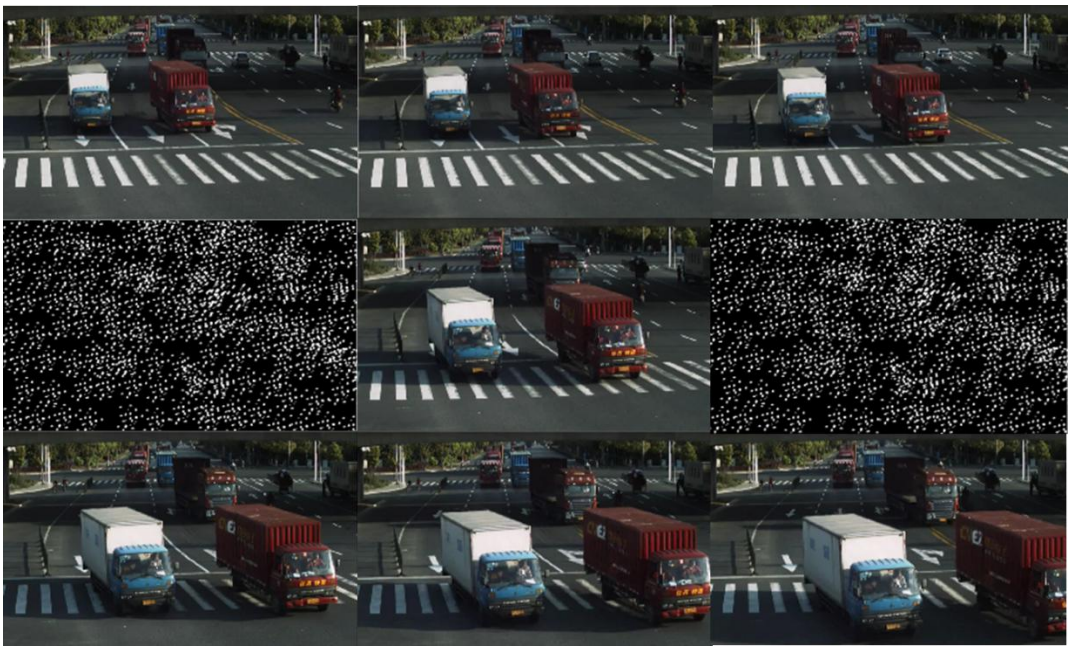


在传统的 RAID5 磁盘阵列上，两块以上硬盘损坏后，RAID 组内所有图像全部丢失。



目前大量的视频监控客户均反馈对于这种现象无可奈何。通过调研总结发现，在视频监控的应用场景中，客户可以容忍少许的图像数据丢失，但是监控的业务必须可以继续进行。希望未损坏的磁盘至少还可以读取数据，尽可能保证数据的完整性。

于是华为针对这种现象对 RAID 算法进行了一些定制和改造。其结果表现为同一个 RAID 组内，无论多少块硬盘故障，只要还剩余一块无故障硬盘，那么其上面的视频仍可提供读取服务。用户的体验是一段正在播放的视频，突然卡一下，时间向后跳了一下（遇到故障硬盘）。



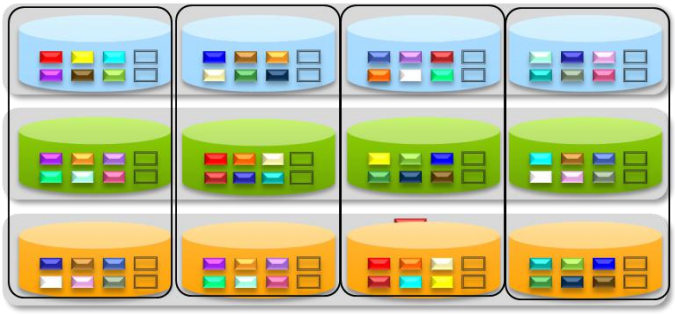
从上面的几副图的对比来看，结论很明显。目前华为将这种技术称为 SafeVideo 技术，在全线视频监控专用存储中已经应用。视频监控的业务不但不会因为多块硬盘发生故障而导致系统暂停，还最大限度的的保护了用户的数据资源。大量的使用者，特别是公安用户反馈，此技术的诞生，说明了华为公司是真正第一家为安防行业定制标准 IT 设备的厂商。

SafeVideo 技术并未就此停下脚步，当维护人员将新的硬盘替换上线后，无需做任何配置，RAID 组将会自动重新组建并上线工作。此举更是进一步的彰显了华为公司的技术能力与在行业中钻研的精神，提升了在行业用户心目中的形象。

数据可靠性

虽然在研调需求的时候，发现客户的需求是“业务连续性为第一重要级别，数据可以容忍少量错误”，但是华为并未将“可容忍少量错误”当做一种借口和满足于现状的理由，而是利用在 ICT 技术上多年的积累，尽最大可能保护用户的数据安全以及数据的完整性。前面讲到的 SafeVideo 技术也是向着这个目标出发，从而设计出了“RAID 组内多块磁盘故障后，RAID 组不失效，还能提供读取服务”的特性，尽可能的减少了数据的损失。

然而 SafeVideo 技术仅仅是华为应用云存储的多级可靠性设计中的一环。从文章的标题可以看到，本篇文章内容的主要硬件载体不是传统的磁盘阵列或者标准数据云存储，而是视频流媒体数据专用的云存储。这里又出现了两个概念，我们来解释一下。



从文章前面的内容可以了解到，标准数据存储和视频监控行业应用存储的区别，是由于视频监控的业务特殊性而产生的。在理论上，一个标准的云存储的实现是将数据切片并且复制多份后随一定均衡算法存储在多个云存储结点中。从左图可以看到，一种色块代表一个文件的切片，它们随机的部署在不同硬件结点的不同硬盘上。

这种方式在视频监控业务场景中应用时会遇到几个问题。

1、 网络带宽的问题。

特别是在平安城市这种大型项目中，若全网部署为一套云存储系统，在现阶段必然会遇到网络带宽的瓶颈。

所以可行的解决方案是仅在派出所或者分局来部署多个小云，然后级联由上级统一管理。

2、 多结点读写带来的问题。

多台设备共同分担同一份数据的读和写，这无疑是提升读写性能的最优方式。特别是在部署大型数据库的应用场景中，这种高并发的读写能够带来性能的最大提升。但是在视频监控应用中，用户查询一份录像，往往一个结点设备内一个 RAID 组的读写能力就可以满足，而不需要调动所有的云结点设备和其中的硬盘部件。从能耗的角度来说，所有设备都在工作，无疑是对能源的巨大浪费。另一方面，因为数据切片的位置不固定，磁头的寻道工作负担加重，这也会带来磁盘寿命的降低。

这种情况下，最可行的解决方案就是“就近存储”。一个派出所或者分局职能辖区内所有的摄像机图像均“就近”存储在本地的云中。而单路摄像机的图像则尽可能的“就近”存储在一个云结点内的一个 RAID 组里面。

3、 多盘失效数据丢失的问题。

在云存储中有两种数据部署方式。一种是数据的多重拷贝，这样数据的可靠性非常的高，但是会导致真实的可用容量仅为实际部署容量的几分之一。对于视频这种非高价值数据来说，这样的投入产出比太低。

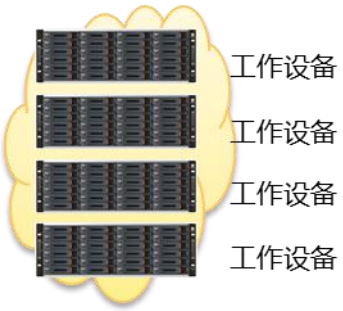
在另一种高利用率的部署模式中，所有结点的硬盘全部拉通组建 RAID 组，可以设值多份冗余数据。这一点和上一小节中 SafeVideo 的内容相同。当一个 RAID 组中的几块硬盘故障后，RAID 所有数据全部丢失，这是不可接受的。

华为公司拥有强大的云计算与云存储研发能力，在调研安防行业的特性后，针对这些问题，对云存储系统进行定制和上述的一些优化。并提出了“视频监控应用云存储”的概念。其中，高可靠性则是视频监控应用去存储的主打特性。

首先，我们一直在提到“应用云存储”，那就是将视频监控业务中所有能抽象出来的特性都内嵌到云存储中了，比如摄像机的接入、存储、转发、点播、智能分析等等。作为摄像机与业务平台之间的中间件的方式，单结点设备即可为用户提供高效并稳定可靠的业务能力支撑。

然后，所有应用云存储结点设备完全对等部署，无差异。经过虚拟化后，所有的应用云存储结点会虚拟化为一台巨大的应用云存储设备。在肉眼上，人们看到的是机柜中整齐划一的单一品种设备。在逻辑上，成千上万的摄像机，以及各种各样的后台软件，看到的只是一台设备。这台大设备拥有海量的接入、存储以及转发等能力。

作为高靠可性的技术代表，华为推出了“N+0”的概念。无论是传统的磁盘阵列方案还是新颖的数据云存储方案，无可避免的采用了“N+M”的部署方式。如右图所示，N 即是工作设备，M 则是备份冗余设备，当某台工作设备故障倒下后，冗余设备可以顶替上去工作。



而华为 N+0 云化模式不再需要冗余设备，如左图所示，所有设备都是工作设备，而且均对等部署没有主次之分。当其中一台设备故障时，云内其余结点设备共同分担故障设备的业务工作。这种设计不仅仅是简单的节省了冗余设备的投资与运维，而是拥有对安防行业业务的深刻理解才做出的设计。故障设备上的摄像机会立即切换到其它正常设备上，视频的实时预览和录像以及上墙等业务均可得到保障，最大限度地实现了设备故障后业务不中断的根本需求。

另外一方面，传统的基于磁盘阵列的方案架构中，会部署一到两台数据库服务器，用以标识磁盘中 SCSI 数据块的意义。如果这个数据库损坏或者数据库服务器故障，那后果不堪想象，所有数据全部“丢失”。因为磁盘阵列只知道这是一个数据块，但是是什么数据，它不知道。

这个时候，应用存储的出现解决了这一个问题。因为其在存储设备内嵌了应用，也部署了一个小型的数据库。也就是说，每台应用云存储结点都知道自己肚子里面装的数据是什么。不管身边的谁故障倒下了，自己的那片小天地不会倒下。

下图描绘了当一台工作设备故障后，视频业务会被其它工作设备接管的过程。当中，如果摄像机与视频监控应用云存储之间的网络链接断开了，摄像机会启动内部 SD 存储进行录像。当

网络恢复后，将实时图像与历史录像同时上传至应用云存储中。经过多级可靠性保证，目前视频监控应用云存储方案可将视频图像的数据完整性提升至业界空前的 99.95%！

并行计算高性能

目前在平安城市的建设中，客户遇到的最大的困扰就是：破案效率低下。这个痛点的直接支撑技术原因就是搜索定位的效率低下。在前面的小节中已经提到，传统方案中会部署数据库服务器，即使是标准数据云存储也会要部署 MDS 元数据服务器。各大方案厂家都会在数据库服务器中保存视频数据的索引，如基于图像帧的索引或者基于时间的索引，无外乎是这两种方式。

当然，数据库也多种多样，有 MS SQLSERVER，有 ORACLE，有 MySQL，PGSQL 等等等等。但无论采用哪种数据库，无论采用哪种索引，都会遇到同一个问题。那就是，随着数据量的增加，搜索查询性能会越来越慢。数据库就是这个瓶颈。

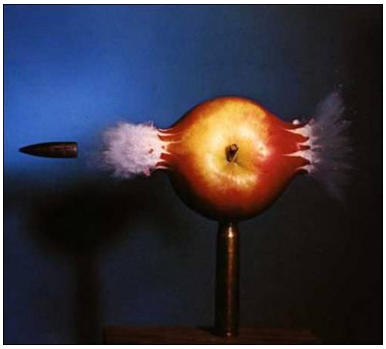
一般来说，视频是由 25 帧或 30 帧来组成的。若我们把每一帧都理解为一张图片，那么视频就是由大量图片组成的了，如下图所示。如果把这些图片快速轮替着放映，就成为了视频。这是传统的动画片的制作方式。如果每一张图张都在数据库里面记录一条索引，那这就是帧索引。

我们用国内某平安城市的实际案例举例。26,000 个 25 帧网络摄像机存储 1 个月时间，如果按照标准的帧索引来计算的话，记录数量会达到 16,848,000,000,000 条。数不清楚？这是 1 万

6848 亿条记录。如果放在 SQL SERVER 这种大型关系型数据库里面，执行搜索命令定位到其中 1 条，实验室里面的数据是耗时 1 小时 50 分钟。当然，经过大量的优化后，定位时间还是保持在 30 分钟以上。

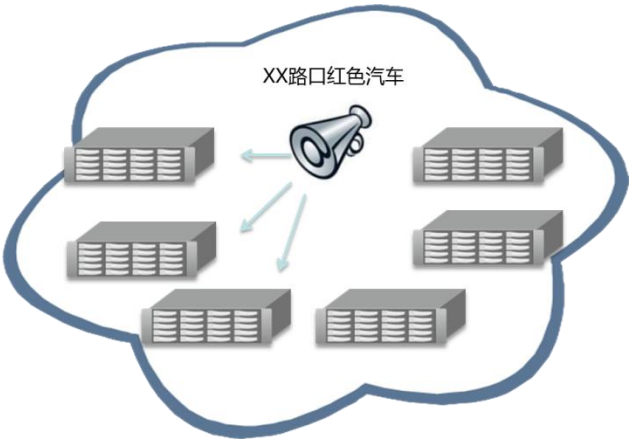
这个性能数据带来的影响是非常可怕的。一个真实的案例，在某大型城市里发生了一起汽车肇事的案件。幸运的是，案件发生后马上接到了报警。公安在城市治安卡口里面搜索汽车牌照号码，半个小时过去了，结果还没有反馈出来。这次事故在公安体系里面都知道。

那为什么有的厂商都宣称自己能实现秒级搜索呢。没有错，的确是秒级，系统部署后确实是秒级。但后面的苦只有客户自己清楚。一个月后，半年后，还是不是秒级搜索就不得而知了。



在重大事件发生时，或者未来视频数据向公众开放的时候，高并发搜索的性能需求就会浮现出来。当华为公司了解到这个客户的痛点以后，进行了细致的分析，最后推出了基于视频监控应用云存储架构的“分布式并行搜索”特性。

在数据写入的时候，视频监控应用云存储结点保存了一份分布式索引，这是一个基于秒偏移技术的视频索引专用散列算法。这个算法也是为安防行业中视频监控的应用特别设计的。在一台云结点设备内部，定位到某一帧在磁盘上的位置仅仅需要不到 10 个微秒。1000000 微秒才等于 1 秒。右图是互联网可以经常看到的一张图片，子弹穿过苹果的瞬间，照相机快门的曝光时间为 3 微秒。可见，应用云结点内的视频帧地址定位速度在人体所能感受到的时间刻度上来说，是瞬间。

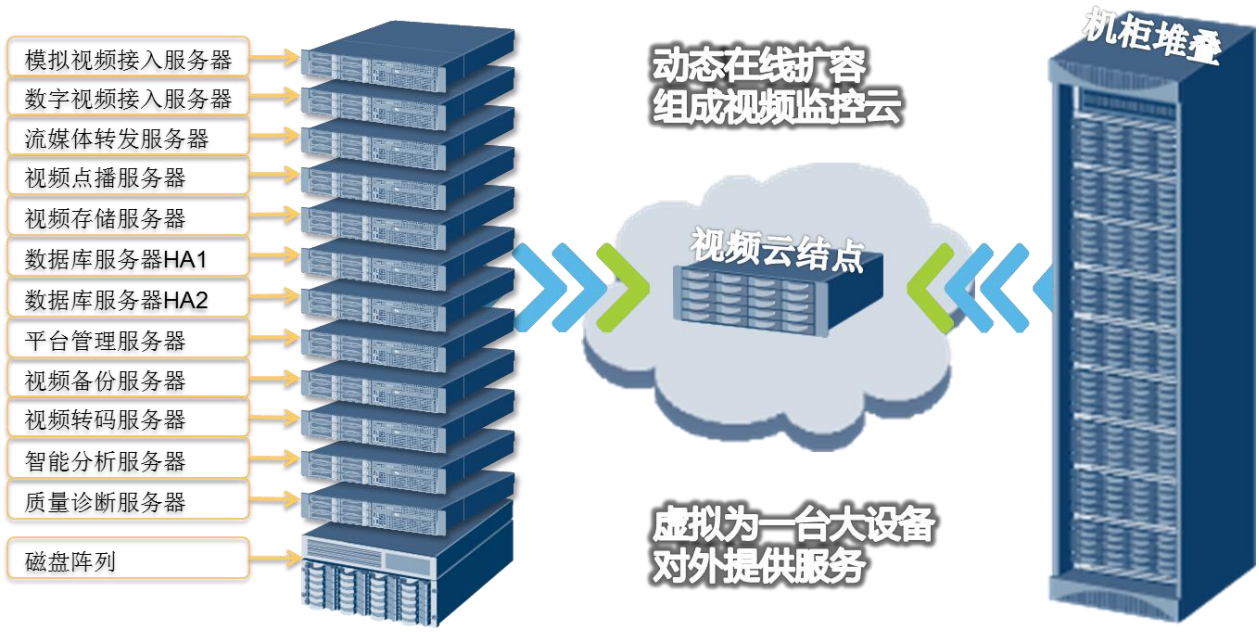


这仅仅是单结点设备的性能。在日常的业务应用中，用户搜索一条指令“在 XX 路口一辆红车的汽车”。当确认键被按下时，搜索服务会拿出一个广播，对着 XX 路口附近所有的应用云结点设备一声大喊：“我要找在 XX 路口的红色汽车”。这个区域的应用云结点设备听到后，均埋头在自己内部空间里面寻找（事先已经被智能分析系统打好标记了）。因为每个结点的内部搜索速度都很快，而且所有听到广播的结点都是并发的在找，谁也不打扰谁。于是乎，100 毫秒左右，搜索服务器就能获得所有结点的反馈了。

对于一个大规模的平安城市来说，视频云存储的搜索模型不会随着平安城市的扩容、存储时间延长等导致数据量的增加而变慢。当然，仅搜索速度快也不能完全解决破案效率的问题。还需要如智能分析、视频摘要、案件归档、串并案分析等多种手段配合才能使客户使用视频监控系统的效率得到提升。

低 TCO（总体拥有成本）设计

本文一开头就提到了应用存储与传统标准数据存储的最主要的区别，就是将行业应用定制安装在了存储设备的内部。仅这样一步，就能节省大量的服务器设备，我们来看一个比较夸张的示例。



这是视频云结点的设计理念示意图。我们可以看到左侧有大量的各种各样的应用服务器和磁盘阵列。从现在开始，都不再需要了，单台视频云结点设备内部就可以提供完整的应用服务。一眼看过去就能看到它节省了大量的服务器。对成本的描述一定要通过数字才够直观，这个帐很好算。我们不要这么复杂，举一个简单点的全集中存储的例子。

	传统存储模型	云存储节点模型		传统存储模型	云存储节点模型
条件	720P高清，H.264格式，4Mbps码流，800路，30天实时存储，Raid5配置，每磁盘框1快热备盘；		条件	720P高清，H.264格式，4Mbps码流，800路，30天实时存储，Raid5配置，每磁盘框1快热备盘	
管理服务器	2	2	管理服务器（2U）	2	2
接入服务器	1	0	接入服务器（2U）	1	0
分发服务器	5	0	分发服务器（2U）	5	0
存储服务器	6	0	存储服务器（2U）	6	0
存储阵列	39	28	存储阵列	39（3U/16盘位）	28（4U/24盘位）
总量	14	2	总量	470	343
结论	每节点，每800路720P全实时图像 节约服务器12套		结论	每节点，每800路720P全实时图像 节约机房空间127U（约3个机柜）	

	传统存储模型	云存储节点模型
条件	720P高清，H.264格式，4Mbps码流，800路，30天实时存储，Raid5配置，每磁盘框1快热备盘；	
管理服务器（460W）	1	1
接入服务器（460W）	1	0
分发服务器（460W）	10	0
存储服务器（460W）	14	0
存储阵列	98（500W）	68（480W）
总量	60960W	33100W
结论	每节点，每800路720P全实时图像,节约功耗28KW，按照PUE=2.2,年度节电54万度	

	传统存储模型	云存储节点模型
条件	720P高清，H.264格式，4Mbps码流，800路，30天实时存储，Raid5配置，每磁盘框1快热备盘；	
管理服务器（4H）	1	1
接入服务器（4H）	1	0
分发服务器（4H）	10	0
存储服务器（4H）	14	0
存储阵列（4H）	98	68
摄像机空间分配	0.08H/IPCAM	0
配置总量	576H	276H
结论	每节点，每800路720P全实时图像节约部署时间300小时，约37.5人天	

这里只是简单计算了节省服务器所带来的 CAPEX 建设成本，当一个项目建成时还有漫长的 OPEX 运维成本需要计算。例如，节省了服务器的发热，就节省了空调的制冷；节省了了机柜空间，使空调制冷集中，效率更高等等。在大量的项目案例中，统计发现，采用华为视频监控云存储方案将为客户节省运维成本 30%以上。

文章一共从可靠性、性能与成本这三个方面来展开。从这里已经可以看出，文章开篇所提到的发展趋势图是正确的。所有通用的标准 IT 技术必须为专业特性进行定制，才能设计出更符合客户需求的产品。

从 2006 年开始，华为技术一直在对安防行业进行研究。一边积累专业的行业人才资源，一边积累行业专用的技术能力。在 2012 年中国国际安防博览会上，以 SafeVideo 与 N+0 应用云存储技术一炮打响。吸引了大量的客户、媒体以及友商。

后续在安防行业中，华为将会深耕细挖行业特性和需求，用最低的总体拥有成本带给客户最好的用户体验！