

Nexsan 数据归档解决方案

数据归档应用分析

人类已经步入一个数字化的信息时代，信息在社会生活的各个领域中正处于前所未有的关键地位。据统计，现在全人类每年要产生 10 亿~20 亿 GB 的新信息，这一信息数量相当于全球每人每年要产生 2.5GB 的信息。另据美国加州大学 Berkeley 分校及 Nexsan 公司所做的调查显示，在全社会每年产生的所有信息中，有超过 85% 的信息是属于固定内容 (Fixed Content) 的数据，也就是说，这些信息一旦生成就不再变更，类似于光存储中的 WORM (Write Once Read Many) 盘片，一次写入，多次读取。这类固定内容的信息存储主要包括法律条文、标准和规范的电子文档以及数字化医学信息、电子邮件及附件、支票图像、卫星图像、音频/视频信息等。

| 介质 | TB/年 | 增长率 |
|------|---------|-----|
| 纸质印刷 | 240 | 2% |
| 胶片 | 427216 | 4% |
| 光盘 | 83 | 30% |
| 磁盘 | 1693000 | 55% |

如上表所示，纸质印刷、胶片、光盘和磁盘是信息存储的四种主要存储介质。在全人类每年产生的新信息中，磁存储无疑是这些信息最大的存储介质，也是发展最为迅速的存储技术，而所有印刷文档仅仅占每年新产生的总信息数量的 0.003%！采用这四种存储介质的信息数量的粗略对比如表 1 所示。不仅如此，磁存储的成本也在急剧降低。据预计，磁存储的成本将由 2000 年的大约 10 美元/GB 跌至 2012 年的 50 美元/TB。

由此可以看出，固定内容信息数量的急剧增长与磁存储技术的不断发展是联系在一起的，这意味着新的网络存储技术空间出现——尤其是当传统网络存储技术无法适应大量的固定内容存储时。虽然采用块或文件访问方式的传统网络存储系统在一定程度上可以满足几十、几百 TB 的网络存储应用需求，但这类存储系统的可扩展性、管理效率以及扩展整个存储系统后的性价比远远不能满足动辄几百 TB (Terabyte, 1TB=1000GB) 乃至 PB (Petabyte, 1PB=1000TB) 的大规模面向固定内容的网络存储需求。不仅如此，对数据的安全访问和长久保存要求苛刻的新存储系统亦对传统网络存储系统解决方案提出了严峻的挑战。

关于固定内容的数据对象

所谓固定内容 (Fixed Content)，简言之就是具有长时间保存价值的不可更改的数据对象。在人类社会的信息化过程中，固定内容的信息数量正急剧增长，同时它们所

需保存的时间也比以前大为延长。

固定内容的数据对象可以分为以下三类：

(1) 电子文档，主要包括研究开发资料，电子图书、合同契约、电子邮件及附件，以及诸如工程绘图的 CAD/CAM 等。

(2) 数字记录，主要包括医学检查图像、图片、地震/油气藏/天文/卫星地球物理图像信息等。

(3) 多媒体，主要包括音频、视频等。

如何对如此庞大的固定内容信息进行高效的存储、管理、检索，这使公司在归档方面面临的挑战如下：

确保遵守规范。国外及国际的一些规范，例如 SEC 17a-4、NAND 3110、HIPAA、CFR Part 11 以及 SOX，要求很多不同产业的公司都需要它们的特殊商业记录得到正确的捕获并且通过一种安全的方式得到保持，该种安全方式可以确保记录的完整性、可靠性以及可访问性，并且有一些方式则有求存储中使用的介质不能被重写。一些规范则保证了信息的保持力，包括对特殊时期的索引以及信息存储。满足遵守规范需求的核心就在于对于审核所需要的信息记录可以有及时并快速的提供能力，一般是在 24 到 48 小时之内。另外近几年我国也对于企业信息归档不断提出更高的要求，制定了各种相关的规定，相信很快会有更多的相关法律法规出台。

减少电子检索 (ediscovery) 的风险。最近美国联邦法案的修订对电子检索规定了更多的步骤和控制，而且，各个洲法庭以及联邦法庭的规范对于信息破坏以及篡改做了更严厉的惩罚规定。在当今诉讼请求不断增加的商务环境下，公司必须在法律下遵守职责以保护用于某种目的的电子信息。在有电子检索的请求时，采用传统的基于序列的归档技术——比如磁带，磁盘，来查找所归档的文档时间上的代价是非常昂贵的，并且往往缺乏效率。因此针对电子发现的对于授权的索引信息的检索在当今的环境中是绝对必须的。

有效管理完全不同的知识库、归档以及相关技术。当今，公司正在寻找对不同应用——包括商业以及公司发展的应用的内容进行归档的办法，以形成一个单一的归档架构。但是对于目前的公司来讲，已经拥有不同的内容资源、资源本身是不同的物理归档并且采用不同的归档技术是非常普遍的现象。对这些资源的管理、培训以及不同系统的技术所需要的资金是巨大的，并且这种完全不同的架构会限制相应的信息管理服务，比如集中的搜索/获取、数据安全、审核以及数据完整性。

减少管理和资金成本。信息技术公司在进行对归档架构投资保护时寻求相应操作效率的提高。管理完全不同归档所导致的结果和寻求提高信息管理操作效率并在公司内部共享知识的目标是完全不一致的。信息技术公司需要保护他们在技术、应用以及信息技

术策略上的投资。如果所采用的技术完全铺开，并且公司在技术层次上仍然不能和其它组件相结合，则会限制公司投资最大化的能力。因此，软件、系统以及存储，无论是旧的还是新的技术，都应该一起工作以形成一个集成的、整体的管理办法，进而让公司进行优化，保证对原有技术投资的扩展保护，懂得这一点是非常重要的。

确保灾难恢复的能力。灾难恢复方法和技术对于业务以及法律上非常有价值的数据是十分关键的。专注于保证公司信息得到保护的规范，灾难恢复计划以及方法对于主要的和归档数据都非常重要。当发生物理上灾难时，归档数据要能够得到恢复，并且那些静态的数据也可以被访问得到。因此，当前很多公司已经开始依赖于远程复制功能来作为灾难恢复的方案，当然还有一些公司使用低成本的可移动介质方案来实施离线存储。

上面所述的各种问题无疑对网络存储方案服务商提出了全新的更高的要求，主要包括：

- ◆ **存取性** 要保证在任何时间、任何地点都能对这些固定内容信息快速、便捷地访问。
- ◆ **可靠性和完整性** 要确保对信息内容不得有任何修改，且所存取的任何固定内容信息都必须符合一定的规范或标准。
- ◆ **持久性和可扩展性** 能够持续地对整个固定内容存储系统数据进行访问和可用，整个存储系统具有较强的可扩展能力，可以非常容易地升级至 PB 级，必须保证避免因可能的设备更新而产生的数据遗失问题。
- ◆ **可管理性和可维护性** 必须满足商业应用的连续性和灾难恢复需求，需要最小化的存储管理，自动化程度尽可能地高。
- ◆ **位置无关性** 对任何固定内容信息的访问必须与存储数据的物理地址/逻辑地址无关，而仅与数据内容有关。
- ◆ **可预防重大灾难性** 可满足用户异地数据保存的要求，当重大灾难发生时，可保证异地数据的完整性和可恢复性。

由上述固定内容的存储需求可以看出，传统的网络存储解决方案很难同时满足这些需求，传统上的 DAS 以及 NAS 和 SAN 网络存储技术都是基于文件数据或块数据且面向事务处理应用而设计的，而非专门针对固定内容存储需求定身而制。NAS、SAN 和 CAS 之间的技术特性对比如下表所示。

| 存储类型 | SAN 存储 | NAS 存储 | CAS 存储 |
|------|-----------|-----------|--------|
| 传输方式 | IP/光纤通道 | IP/光纤通道 | IP |
| 数据类型 | 块 (block) | 文件 (file) | 固定内容 |

| | | | |
|------|-----------------|------------|------------|
| 关键需求 | 确定性的性能 | 多协议共享 | 长久，确保完整性 |
| 典型应用 | OLTP，数据仓库，ERP 等 | 中间数据，文件服务器 | 长期，内容管理 |
| 信息周期 | 所存储内容创建并积极共享 | | 所存储内容固定并保存 |
| 寻址类型 | 基于位置 | | 基于内容 |
| 数据内容 | 可变，有些总在不断变化 | | 固定 |

另一方面，归档与用户现有的备份系统也是完全不同的两种应用类型，所以拥有了备份并不意味着就完成了数据的归档，我们可通过下表看到两者的本质区别：

| | 备份 | 归档 |
|--------------|-------------------------------------|--------------------------------|
| 主要的工作方式及使用方式 | 复制主系统中的数据用于短期的数据恢复 | 移动主系统中的数据到归档存储用于长期的数据再利用 |
| 保存策略 | 数据基于固定的时间周期被复制到备份介质上，一个保存策略应用于所有文件。 | 数据基于策略或事件被迁移到归档存储上，基于文件级的保存策略。 |
| 针对的数据类型 | 不断变化的动态数据 | 长期不变的固定内容的数据 |
| 用户是否可直接访问 | 不可，需要恢复后再访问 | 可直接访问 |
| 是否易于查找 | 不易查找 | 易于查找 |
| 自动化记录清除 | 不能自动清除 | 可自动清除 |

如上所述，对于拥有大量重要固定内容数据的用户，一个好的归档存储系统是至关重要的。

Nexsan Assureon 数据归档存储介绍

Nexsan 结合 13 年对行业 IT 需求的深层理解，推出了满足企业客户存储归档需求的解决方案——Assureon。Nexsan's Assureon™ 是一个智能、安全的网络存储系统，用于以一种经济高效的方式，存储、管理、保护、处理固定内容 / 监管达标信息，并实现对这些



信息的快速而安全的访问。一个组织内的数据分为两类：动态数据 — 一般位于数据库内，经常更改；静态数据 — 为参考或固定内容数据，很少更改，一般拥有很长的生命和很高的价值。参考数据有许多种形式，例如办公室文档、电子邮件、合同、员工信息、医疗记录与图像、监控录像和媒体内容等。

对于面临存储和管理不断增长的参考信息这个难题、并希望为此找到一种经济有效的方式的组织，Assureon 是一个合适的解决方案。Assureon 可以简化参考信息的长期保留过程，并降低相关的人员和管理成本。

Assureon 适用于需要遵守现有法规或与信息保护、管理及处置有关的新法律的企业。Assureon 可以实现一些业务措施，降低与电子查找请求和潜在诉讼有关的风险。

对于实施企业管理措施，用于对其数字资产进行长期管理的组织，Assureon 具有的前所未有的灵活性和安全性可以使其如愿以偿。使用 Assureon 提供的、处于严格管理之下的“硬化”信息存储资源，每家企业都可以节省大量总体拥有成本。

Assureon 功能特征

- ◆ 一个安全的磁盘存储设备，用于固定内容信息的长期保留和管理。
- ◆ 多层安全与数据保护，通过采用安全 AES 加密、严格的可访问性和高可用集群化硬件体系结构而实现。
- ◆ 一个充分集成的完整硬件 / 软件解决方案。
- ◆ 简化的兼容性问题 — Assureon 强大的接口消除了对复杂的 API 的需求。
- ◆ 高伸缩性系统 — 性能与容量可以独立伸缩。
- ◆ 满足对数据记录管理有影响的主要法规（《SEC 17-4》、《HIPAA》、《CFR 21 第 11 部分 FDA》、《Sarbanes-Oxley》、《加拿大 PIPEDA》、《英国数据保护法案》等）的要求。
- ◆ 无故障技术，可以保持内容的真实性与完整性，同时确保无数据丢失。
- ◆ 存储成本的降低
- ◆ 实现信息生命周期管理 (ILM) 过程的实施。
- ◆ 自管理与自配置功能，可以减少管理大量信息所需的人员。
- ◆ Assureon 的 CAS 技术消除了同一文件存在多个副本的现象。
- ◆ 自诊断与自恢复功能可以保证信息的安全性；Assureon 带来了无单点故障的绝对冗余性、持续性的数据完整性检查、RAID 6 硬件保护与总部呼叫功能。

Nexsan Assureon: 新一代安全归档产品

Assureon 基于 CAS（内容可编址存储）技术和先进的 256 位加密及其他安全与数据保护技术。CAS 根据所有对象的内容，为其生成唯一文件标识符。Assureon 的第二代对象存储模型可以将文件、图像、记录、网页等作为带有安全、唯一通用标识符的对象进行存储。这种方法不同于传统文件系统模型。在后者中，应用程序或用户对文件进行命名，然后将其放入一个层次化文件系统中。传统文件系统模型的缺点是缺乏伸缩性和管理成本高。

在对象模型中，已归档的物理对象与逻辑位置或应用程序完全分离。现在，管理员可以通过 Assureon，从格式化文件系统、创建卷和绑定 LUN 等常见存储“家务活儿”中摆脱出来。

Assureon 带来了以下高价值优点：

- ◆ **单实例存储：** Assureon 的 CAS 技术实现了单实例存储，因此只需保留数据对象的一个副本。Assureon 可以根据对象的 CAS 地址，消除对象的重复副本，仅存储一个副本，从而减少了所需的存储空间，降低了网络带宽和总体成本。
- ◆ **身份验证与完整性：** Assureon 技术可以验证和报告所有资产和资产元数据是否被篡改或毁坏。
- ◆ **永恒性：** 处于 Assureon 管理之下的对象直到保留期结束之前，都是无法更改或删除的。
- ◆ **WORM 功能：** Assureon 带有非可擦除、非可覆盖功能
- ◆ **访问速度：** 与传统的离线存储不同的是，Assureon 可以实现对参考数据的即时在线访问。
- ◆ **信息保证：** Assureon 技术可以确保信息的真实性、在需要的位置和时候的可用性，并确保信息仅在授权的情况下接受访问。
- ◆ **执行文件级的保留策略：** Assureon 可以执行对象级的保留策略，允许根据文件对一个组织的重要性，对其采取不同的保留策略。
- ◆ **保留期限的指定：** Assureon 可以根据第三方应用程序（例如文档管理或电子邮件管理软件包）指定保留期限，也可以根据自己的文件监控程序采用合适的保留策略。
- ◆ **旨在满足法规要求的审查痕迹：** Assureon 可以满足无数法律和隐私法规的种种要求，包括：
 - 数字时间戳记录 — 一条 FDA 法规要求
 - 序列化 — 一条 SEC 法规要求
 - 用户访问痕迹 — 一条 HIPAA 要求

- ◆ **军事级的删除/处置：**在保留期限结束时，对象在经过有关管理员的授权后，通过一个包含七个步骤的过程擦除。此外，已删除的加密数据的密钥也会被“擦除”，带来了“双重”删除功能。
- ◆ **安全与数据保护：**可以对处于 Assureon 管理之下的全部或选定资产采用 256 位 AES 加密。资产的防篡改时间戳增加了另外一层安全性。
- ◆ **业务持续性保护：**Assureon 支持通过 WAN，对资产进行远程复制，并且可以经过配置，在远程站点保留复制文件。
- ◆ **介质独立性：**除了磁盘之外，Assureon 还支持磁带和光学磁盘等离线介质，让用户既能受益于磁带上的加密技术带来的高安全性，也能受益于能够降低存储消耗与成本的 CAS 单实例存储技术。
- ◆ **离线介质上的选择性处置：**此前，备份到磁带或光学磁盘等离线介质的对象还无法在单个文件级别进行删除。Assureon 通过提供这些文件的加密密钥的删除功能，带来了更高的安全性。Assureon 可以在数据保留期结束时、或根据处理安全破坏或攻击的授权，擦除在线加密文件，及其存储在多个冗余位置的相关加密密钥。密钥毁坏后，在线或可删除介质上的文件信息就无法读取。
- ◆ **实现信息生命周期管理 (ILM)：**Assureon 是一个强大的工具，可以自动执行组织内部参考数据的生命周期过程的操作步骤（分类、放置、保护、归档、访问、到期）。
- ◆ **CAS 无故障技术：**Assureon 采用双加密过程（哈希算法），创建自己的唯一 CAS 地址，确保在可能的数据碰撞过程中，保证文件完整性，而且无信息丢失。
- ◆ **自诊断与修复：**Assureon 通过扫描管理中的文件，将其当前唯一标识符与原始标识符进行比较，定期验证其完整性。如果检测到差异，Assureon 会生成一条出错消息，记录错误，并用它管理的镜像副本自动替换文件。
- ◆ **搜索与访问：**Assureon 提供文件元数据的一个可搜索数据库，为找到文件提供帮助。通过 Assureon 的完全自动化的密钥管理，还实现了对长期系统访问的支持。
- ◆ **容错：**Assureon 还可以在无数据丢失或不中断数据访问的情况下，经受住单起或多起故障。如果某节点发生故障，处理工作会被重定向到另外一个节点。如果某个存储位置无法使用，处理工作将被重定向到第二和第三个位置。如果所有位置都无法使用，Assureon 将记录新文件和元数据，直到有可以使用的为止。
- ◆ **可伸缩性与负载平衡：**Assureon 的体系结构可以让您独立添加处理能力，以提高性能或增加存储容量。Assureon 可以自动标识增加的节点或存储资源，并且进行负载平衡，实现最大的吞吐量。
- ◆ **终极 D2D 备份：**Assureon 的强大功能提高了 D2D 备份与磁盘归档的水准，并且能够对移动计算机及分支办公室提供支持。