

VERITAS™

数据基因指数

一份探究存储环境真实结构的报告, 由非结构化数据专家倾力奉献

2016



数据基因项目是 Veritas 自主发起的一项计划,旨在改变人们对数据管理的看法。Veritas 发起该计划,希望为志同道合的数据科学家、行业专家和思想领导者创建一个共同的讨论平台,从而深入了解企业日常创建、存储和管理的非结构化数据的真正本质。作为基于实际存储环境组成的基准测评报告,首份《数据基因指数》报告便是此项计划的首个研究成果。

Veritas 为 86% 的财富 500 强企业备份、归档或分析数 EB 数据。凭借自身的独特优势, Veritas 能够轻松收集企业环境的本质特征。当前, Veritas 研究的重点对象是元数据的特征。通过 Veritas 的文件分析产品,我们可以利用从各个客户汇集的元数据,准确呈现企业实际数据环境结构的各个细节。

当今数据环境现状

过去，企业数据环境中，

41%

的数据 **3年内** 从未修改过



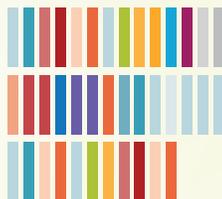
排名前几位的文件类型出乎意料。



传统的办公文件使数据环境不堪重负。



孤立信息



孤立信息所占比重过大，且过于陈旧。

系统归档以下文件类型



减少存储成本

50%

企业应对自己的删除或归档策略进行优先级划分，以节约巨额的存储空间成本。



首份 Veritas 数据基因指数报告

为了进一步了解客户数据环境的真正结构, Veritas 在 2015 年分析了来自众多客户非结构化数据环境的数百亿份文件及其属性。本次分析涵盖了 8,000 多种最常见的文件类型扩展名。因此, 报告中的数据基本能够代表客户文件系统环境的数据构成。

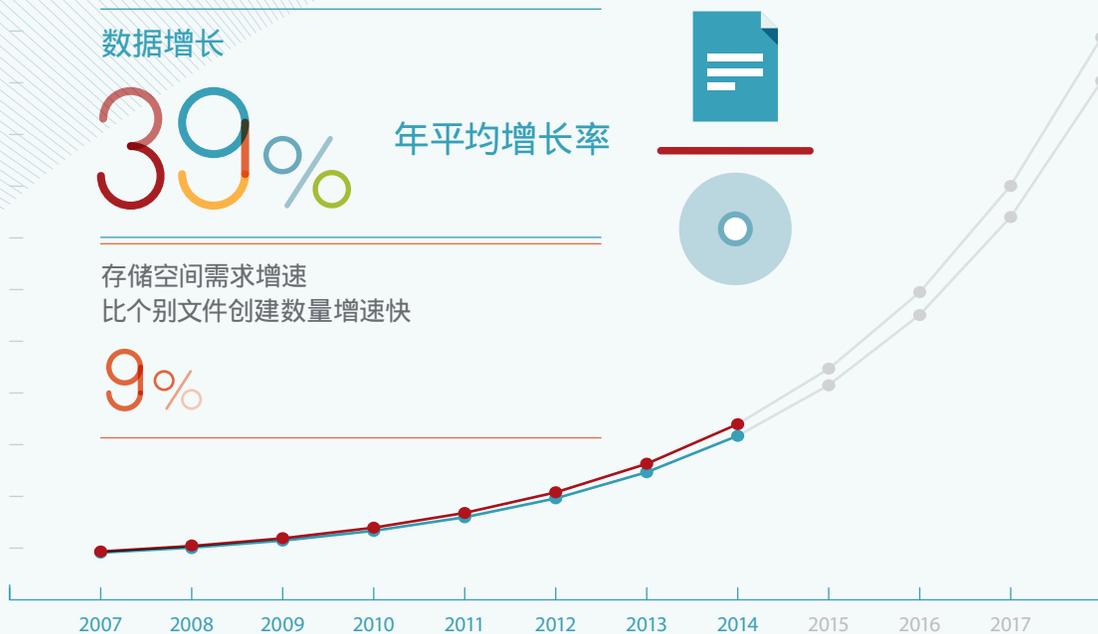
数据增长

39%

年平均增长率

存储空间需求增速
比个别文件创建数量增速快

9%



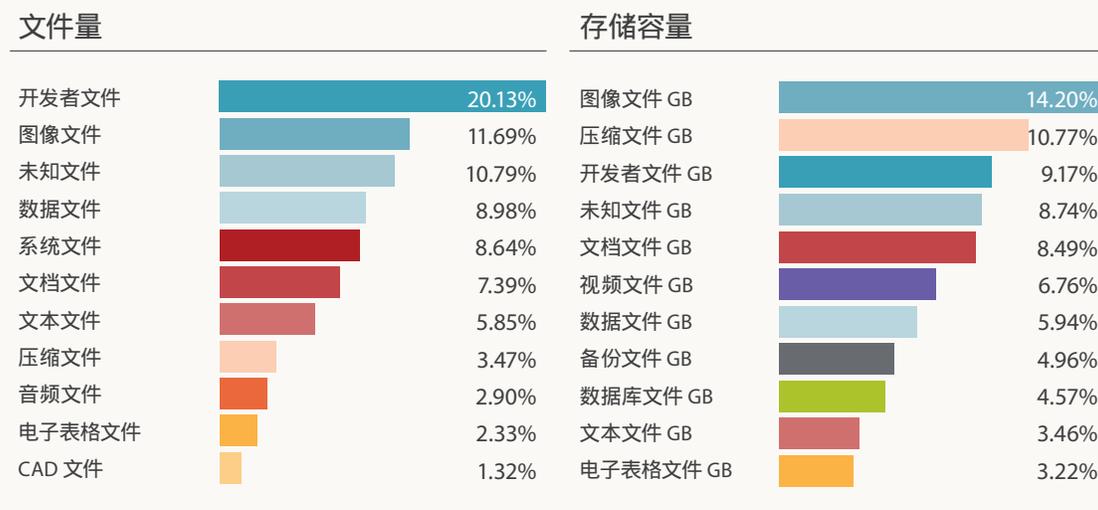
数据呈爆炸式增长

从文件级别来看，在过去七年中，数据的实际增长速度为每年平均增长39.2481189%，而相应的存储空间需求增长比创建单个文件增长快9%，因此，尽管操作层面上的改变可以一定程度地控制某些方面的增长，但根本依然是存储管理问题。

控制存储空间并不仅仅是存储问题。现在，存储环境杂乱无章，平均1 PB的信息包含了2,312,000,000个文件。

增速最快的文件类型：图像和开发者文件

下面的企业数据总量表明，数量最多的文件类型和最占空间的文件类型具有明显差异。企业可以将下方图表作为存储环境与成本对照表。



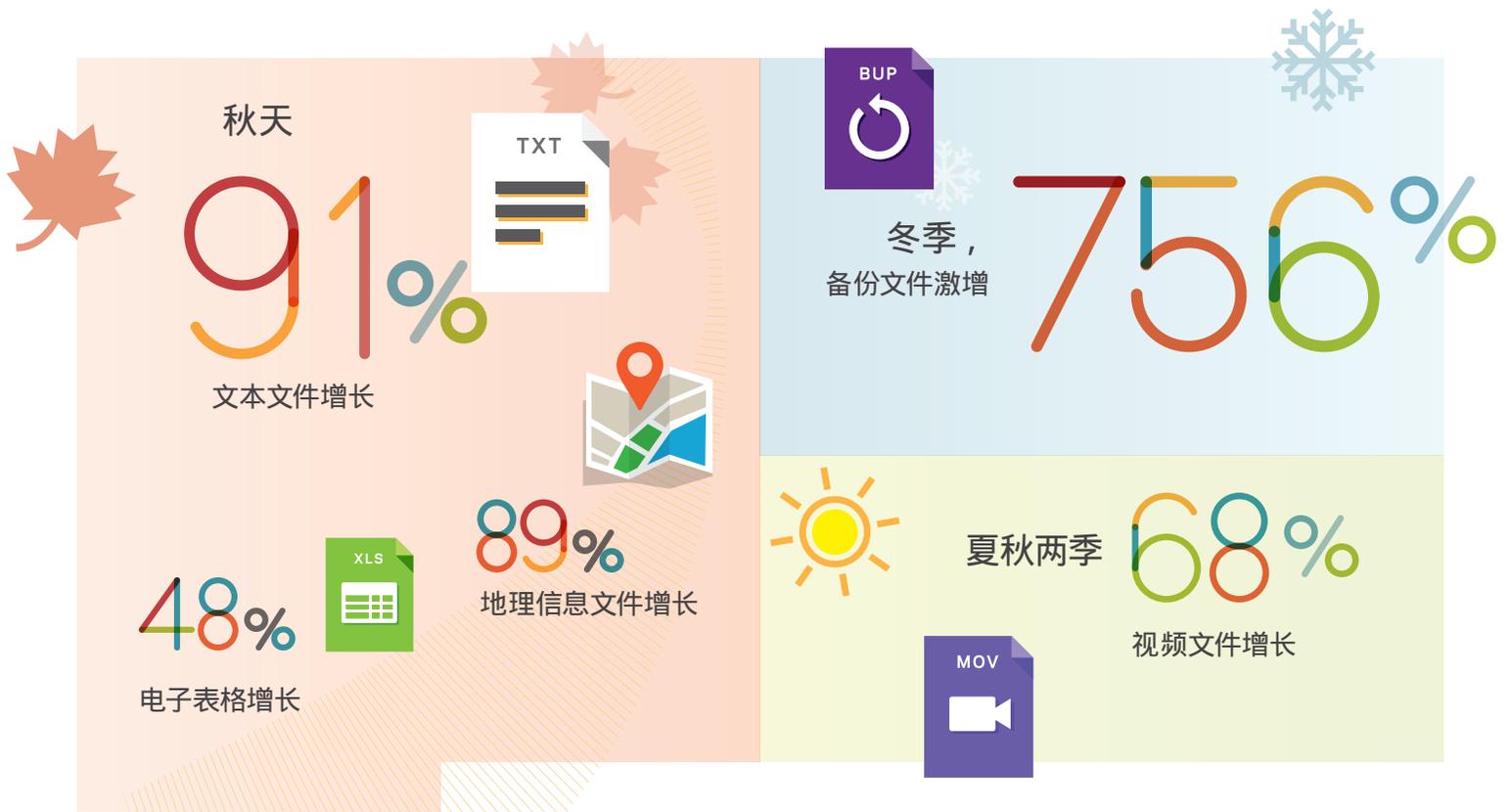
10年前 VS 现今：企业数据的变化

随着时间的推移，数据结构已经发生了巨大的变化。过去十年间，相较其他文件类型，变化最大的文件类型为：



在存储空间方面，变化最大的文件类型为：





不同季节, 企业的文件增长情况

秋季是文件创建的旺季。文本文件的增长率尤为突出, 高达 91%, 其次是地理和信息系统文件, 增长率为89%; 电子表格的增长率为 48%。

只有备份和文档文件会在秋冬季激增。在企业完成年度备份后, 备份文件将激增 756%。但图像文件的创建量明显减少, 降幅达 63%。此外, 68%的视频都创建于夏秋两季, 电子邮件 (pst) 的情况可预测性高, 各季节间只有 0.7% 的标准偏差。

过去，企业数据环境中，

41%

的数据在过去 3 年内从未修改过

12% = 陈旧文件
(7 年内无人修改)

不到 3 年

至少 3 年

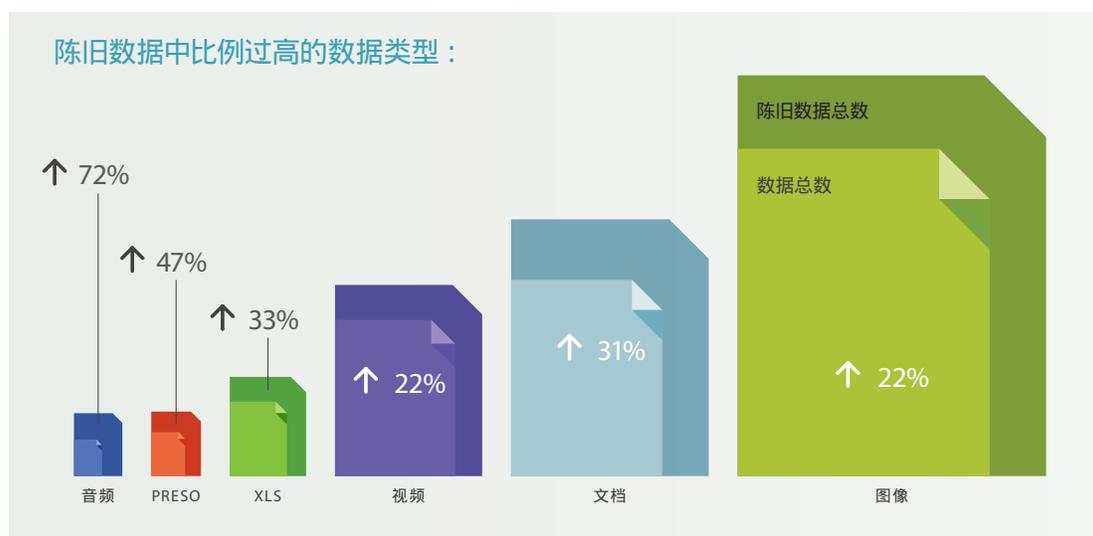
超过 7 年

为何企业还在保留这些数据？

信息是当今企业的关键所在，但由于信息创建速度过快，企业中的有价值信息转瞬即逝。

治理：数量过多的文件类型

面对海量的陈旧数据，以及在可以执行多个可行处理决策时，为信息管理“决策预算”的重点划分优先级，无疑能够帮助企业选择文件管理的入手点。陈旧数据和总数据中数量最多的文件类型中，传统“office”文件是企业巨大的负担：



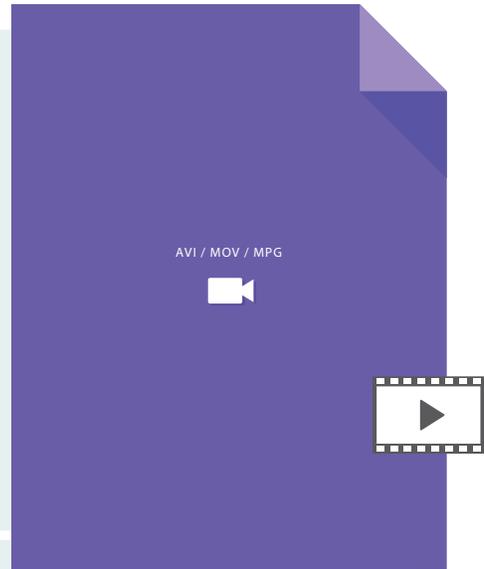
如果企业希望最大限度地节约存储空间成本，但又无法确定优先处理哪些文件类型，那么企业可以重点对以下五大类型文件进行处理，每一类型都能够帮助企业找回GB级的存储空间。

1. 虚拟机文件
2. 安全文件
3. 游戏文件
4. 科技文件
5. 地理位置信息系统文件



在陈旧数据存储空间总量中，
视频文件所占比例最高，超过其所在
陈旧文件比例的：

15.8 倍



7.3x



6.4x



2.2x



文件数量和空间比例失衡

如果企业希望优先处理特定文件类型，只需观察哪些文件类型的数量和空间不成比例。例如，视频文件在陈旧数据存储空间总量中的比例，比其在陈旧文件总量中的比例高 15.8 倍。虚拟机文件所占空间为 7.3 倍，演示文件是 6.4 倍，电子邮件为 2.2 倍，这些类型的文件都是企业优先处理的最佳选择。

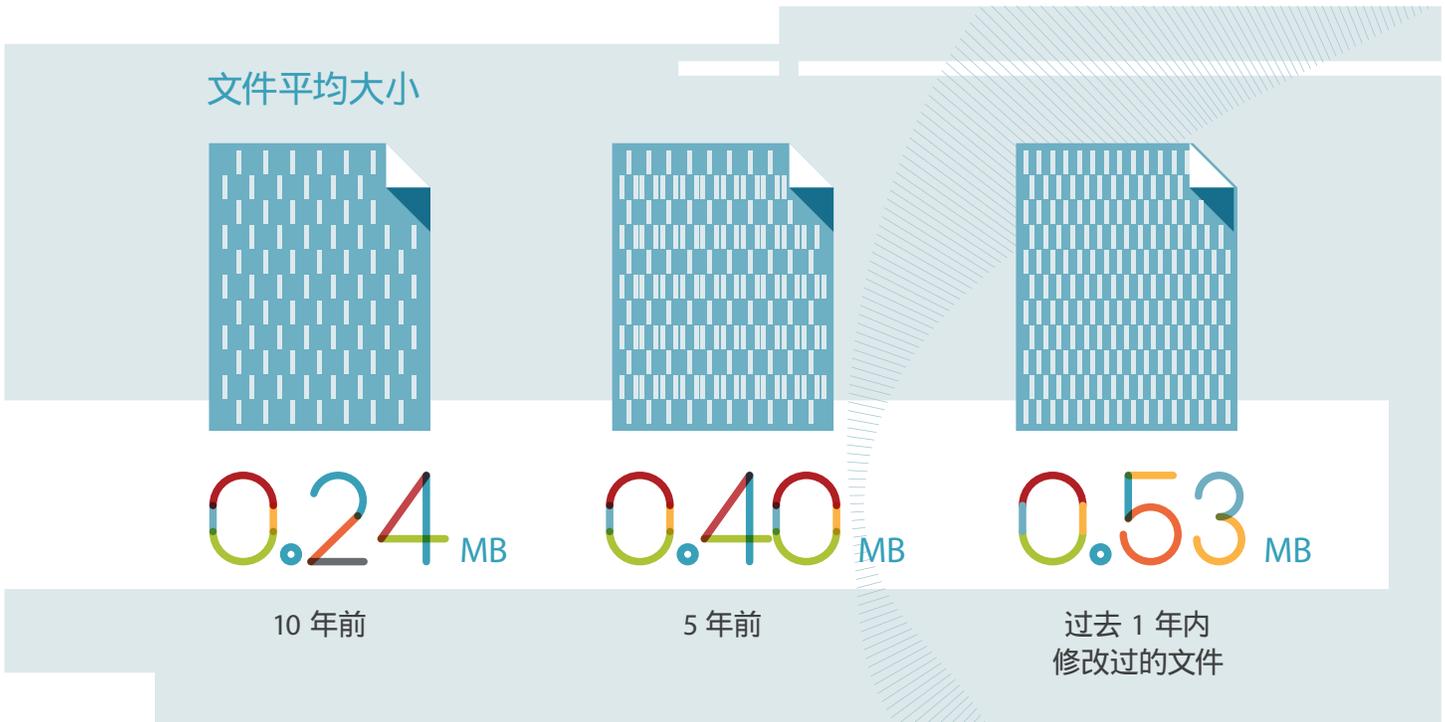


当员工离职后, 却留下了数据残局

当数据失去其所有者, 即会成为孤立数据。由于职位变更、员工离职以及常用活动目录混乱等原因, 企业很难追踪数居环境的传承关系; 此外, 追踪数据传承需要巨大的资金支持。

孤立数据会占据企业高额成本, 一方面是因为其占据了过量的存储空间。即便孤立数据仅占文件总量的 1.6%, 但它占据的存储空间却达到了 5.1%。不仅如此, 大多数孤立数据是内容丰富的数据类型, 这类数据所占据的空间都远远超正常比例, 如图像文件超出正常存储空间的 88%, 而视频和演示文稿分别超出 165% 和 229%。

报告调查发现, 企业人员流动趋势对存储环境具有一定的影响。孤立文件的大小是平均文件的 222%。数据管理人员可能认为, 文件越大, 其内容越重要, 因此在员工离职后倾向于继续保留这类密集文件。如果企业希望恢复更多存储空间, 从孤立数据下手无疑是首选。



通过存储密度来判断文件是否有用并不可靠

如今, 企业创建大密度的内容并不奇怪, 但令人吃惊的是, 过去 7 年来, 这类内容的涨幅只达到了10.3%, 增长相对缓慢。文件的平均大小为:

过去 10年或更长时间内使用过的文件的平均大小:	0.24MB
过去 5 年内使用过的文件的平均大小:	0.40MB
过去 1 年内修改过的文件的平均大小:	0.53MB

被分类为陈旧文件, 要比去年一年内修改的文件小 33%。

现在, 企业该如何应对?

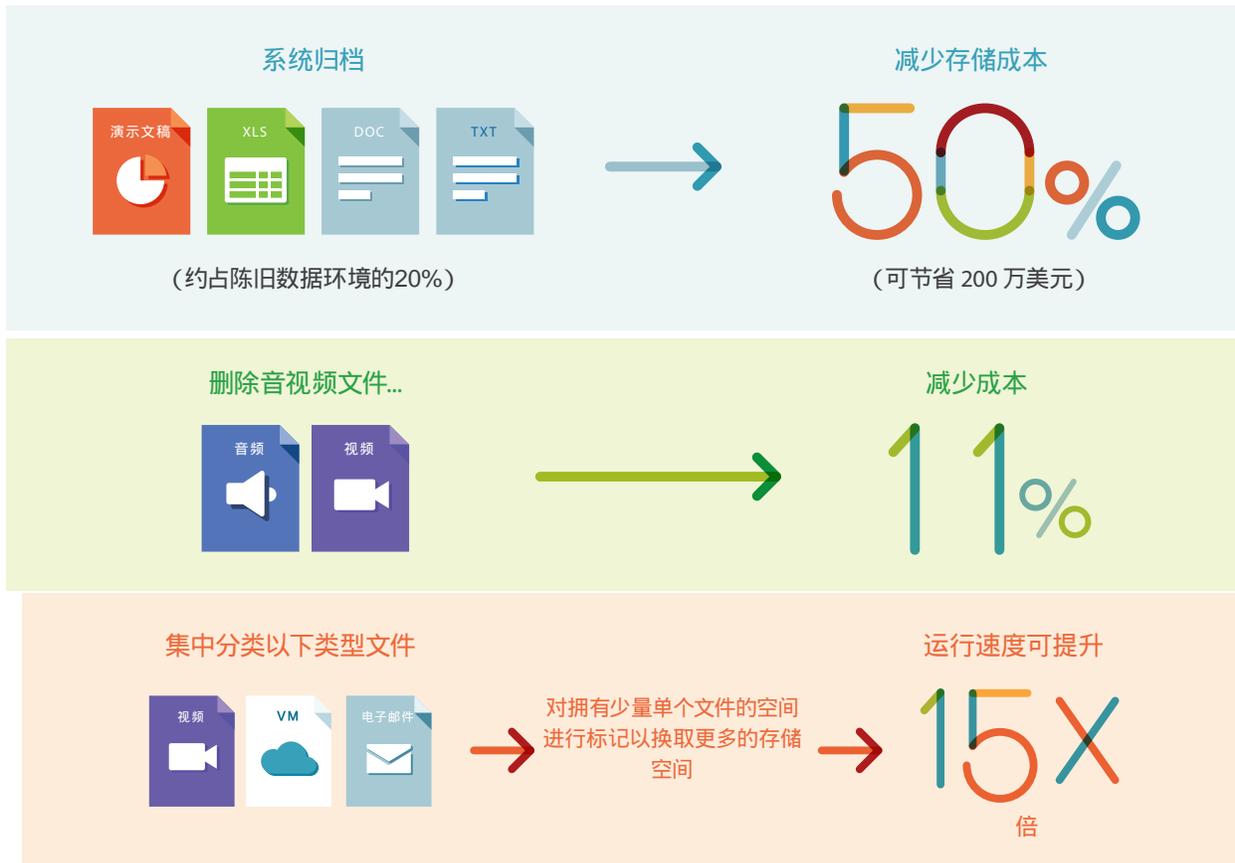
如果企业用户的企业存储环境与我们分析的环境类似, 那么企业用户将有很多机会来改变自身存储现状。以 10 PB 为普通环境举例...



如果企业数据环境中 41% 的数据为陈旧数据, 那么每年, 企业需要投入 2050 万美元来管理3年来无人问津的数据。然而, 清除陈旧数据非常困难。企业整理4.1 PB 的数据, 便需要对 94.79亿个个体文件进行分类、删除或归档的决定。



企业用户不得不划分优先级...



演示文稿、电子表格、文档和文本文件等内容丰富的文件占陈旧数据的 20%，开展一项专门处理这些文件的归档项目，能够降低企业至少 50%，相当于 200 多万美元的存储成本。删除陈旧数据中音频和视频文件，能够帮助企业降低 11% 的成本。此外，图像文件占据企业陈旧数据 18% 的存储空间，它们甚至在长达7年或更久的时间里从未受到修改。

集中处理拥有少量单个文件的空间，对其进行标记以换取更多的存储空间，例如视频、虚拟机文件和电子邮件等。这不但能够回收大量的存储空间，同时运行速度可迅速提升 15 倍。当员工离职或职位调整时，对其遗留的数据进行评估，可帮助企业节省 5%，大约近百万的成本...

无论企业现状如何, Veritas洞察分析都能够帮助企业对抗过度增长曲线, 重新掌控企业环境。

VERITAS™

Veritas Technologies 旨在帮助企业驾驭信息的力量, 通过解决方案帮助企业应对全球最大且最为复杂的异构环境。我们服务于全球各类规模的企业, 目前, 86% 的世界财富 500 强企业都在使用Veritas所提供的解决方案。此外, Veritas 已经连续12年荣列 Gartner “企业级备份软件和集成式一体机”¹ 以及 “企业级信息归档”² 魔力象限领导者。

Veritas拥有经验丰富的合作伙伴, 全力帮助企业增强数据可用性, 并通过对信息的洞察来提升竞争优势。Veritas 市值高达 25 亿美元, 在全球 58 个国家/地区拥有 7,800 多名员工。同时, Veritas与众多大型科技领导品牌建立了合作伙伴关系, 其中包括 Amazon、Cisco、Fujitsu、Google、Hitachi、HP、IBM、Microsoft、NetApp、OpenStack 以及赛门铁克等。

从传统数据中心到私有云、公共云和混合云, 无论企业 IT 环境多么复杂, Veritas 都能在合作伙伴的支持下, 通过智能信息管理解决方案, 帮助企业保护、识别和管理数据。选择 Veritas, 企业将全面掌握信息洞察与信息可用性, 以此了解自身拥有哪些信息, 如何妥善保护这些信息, 以及哪些信息理应删除。

如需了解服务详情, 请访问 veritas.com。

法律资料:

¹资料来源: Gartner, 《企业级备份软件和集成式一体机魔力象限》(Magic Quadrant for Enterprise Backup Software and Integrated Appliances), 分析师: Dave Russell、Pushan Rinnen、Robert Rhame, 2015 年 6 月 15 日

²资料来源: Gartner, 《企业级信息归档魔力象限》(Magic Quadrant for Enterprise Information Archiving), 分析师: Alan Dayley、Garth Landers、Anthony Kros, 2015 年 10 月 29 日注意, 2005-2009 年此项研究名为 “Magic Quadrant for Email Active Archiving (电子邮件有效归档魔力象限)”。

Gartner 并不推荐其研究出版物中描述的任何厂商、产品或服务, 也不建议技术用户只选择评价最高或拥有其他荣誉称号的那些厂商。Gartner 研究出版物仅代表 Gartner 研究机构的观点, 不应视为客观事实陈述。Gartner 对此项研究不作任何明示或暗示担保, 包括对于特定用途的适销性或适用性的任何担保。

© 2016 Veritas Technologies LLC. © 2016 年 Veritas Technologies LLC 版权所有。All rights reserved. 保留所有权利。Veritas、Veritas 标识是 Veritas Technologies LLC 或其附属机构在美国和其他国家/地区的商标或注册商标。