

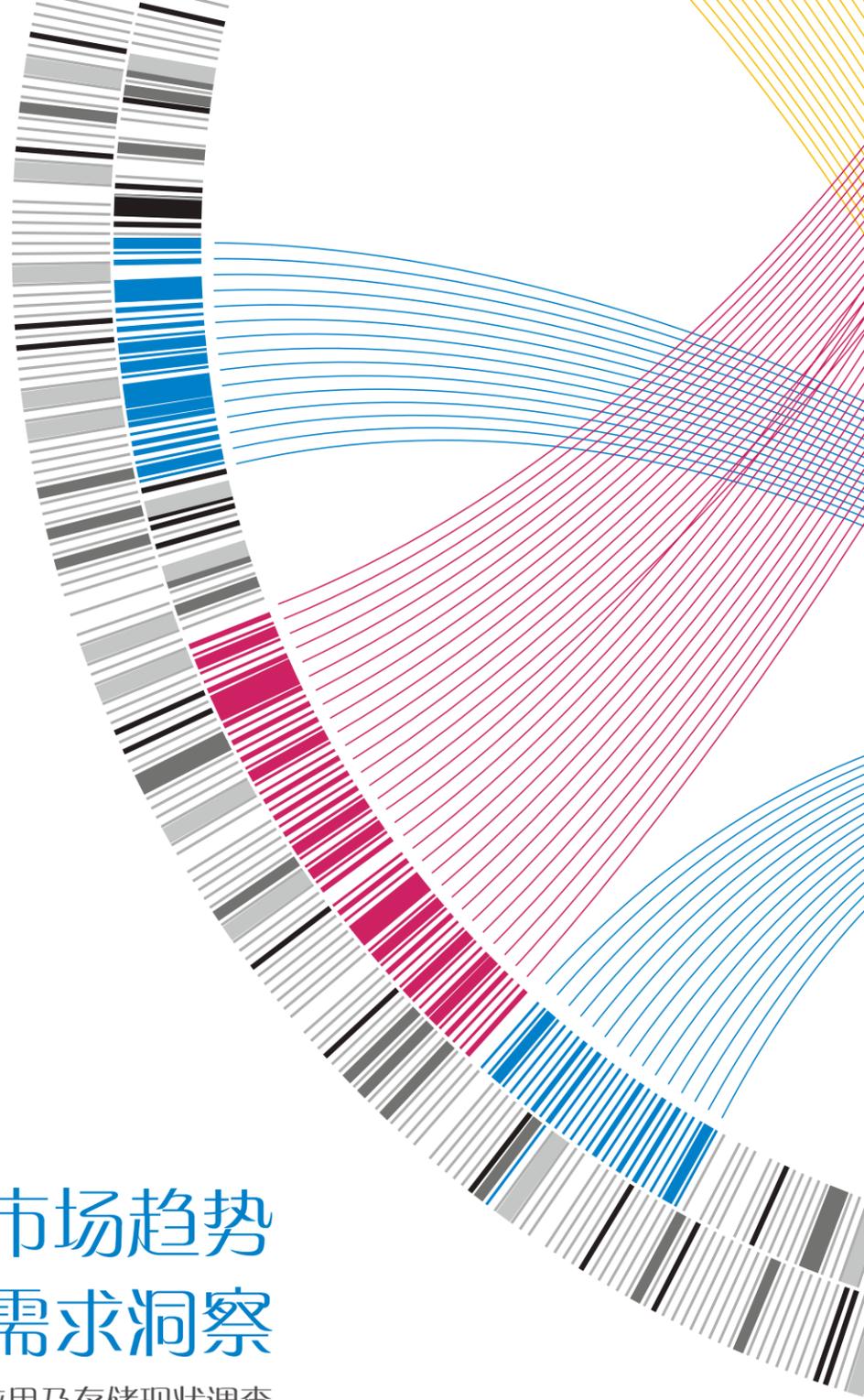


作者简介

王丛 (Kim Wang) 女士, 中桥调研咨询的创始者兼首席分析师, 有 23 年在欧洲、南亚和北美的管理和咨询经验, 其中有 14 年时间专注于数据中心领域, 对存储、服务器、网络、客户端和数据中心管理软件等领域有着深入的了解和研究。2012 年, 王丛在东南亚及中国就数据中心技术和市场做了 60 多场现场和网络在线演讲, 从全球视角为中国用户提供可信参考依据, 加快中国市场对新技术的评估和接受能力。

马艳, 中桥调研咨询的调研分析师, 对中国数据中心细分技术有着深入的了解。2006 年加入中桥调研咨询, 负责数据中心领域的调查和评测, 以及各种项目的策划和执行。主持完成了多个调查项目的规划、调查和深访问题设计, 并撰写了相关使用评测报告、白皮书和调查报告。

杨凌霄, 中桥调研咨询助理分析师, 对数据分析及中国数据中心相关技术有一定的了解。2013 年加入中桥调研咨询, 主要负责中小企业数据中心领域相关技术市场和趋势的分析。



中国大数据市场趋势 及存储管理需求洞察

暨 2013 中国大数据市场应用及存储现状调查



关于中桥调研咨询

中桥国际调研咨询有限公司 (Sino-Bridges Research and Consulting Ltd.) 成立于 2006 年, 是一家专注于数据中心领域的咨询和调研公司, 致力于从全球视角结合调查数据和市场技术, 为 IT 厂商和 IT 专业人士提供前瞻性、可信的市场和技术趋势参考以及网上学习 & 提高平台 (www.webinars-china.com)。其主要服务和研究领域集中在数据中心相关技术, 如存储、服务器、网络、客户端设施、商业智能和数据中心架构管理软件等。主要研究课题包括: 虚拟化、云、大数据、数据保护、IT 架构和应用趋势等。

中桥调研咨询的分析师对欧美和中国数据中心技术和市场的调查和咨询有着多年的经验。此外, 中桥调研咨询拥有数万最终用户数据和调研会员, 通过强化最终用户的互动, 来透彻了解中国用户的需求、挑战和困扰。中桥调研咨询主要服务形式包括调研报告、产品和使用评测、分析报告和技术白皮书等。2008~2012 年, 中桥调研咨询与全球十大咨询公司之一 ESG 形成联合品牌 ESG-Sino (中桥), 在中国提供技术和市场咨询服务。中桥国际调研咨询在美国西雅图和北京、武汉等地设有办事处。客户包括 IBM、Dell、HP、EMC 等及国内厂商华为、联想、浪潮等。

关于调查报告主体报告, 请联系 contact@sino-bridges.com。

* 戴尔的常规条款和条件在此适用, 并可网上获得或致函索取。戴尔会尽力排查报价错误或其他错误, 但由于我们的疏漏, 某些错误仍可能发生。戴尔有权不接受任何包含错误信息的订单。以上图片仅供参考。戴尔、戴尔标志、Venue 是 Dell Inc. 的注册商标或商标。超极本, 赛扬, Celeron Inside, Core Inside, 英特尔, 英特尔标识, 英特尔凌动, Intel Atom Inside, 英特尔酷睿, Intel Inside, Intel Inside 标识, 英特尔欢跃, 英特尔博锐, 安腾, Itanium Inside, 奔腾, Pentium Inside, vPro Inside, 至强, 至强 Phi 和 Xeon Inside 均是英特尔公司在美国或其他国家的商标。微软、微软标识和 Windows 是微软公司在美国和其他国家的商标 / 注册商标。文中提及的其他商标或商品名称均指拥有该商标或名称的机构或其产品。戴尔不拥有其他机构的商标和商品名称的相关权益。版权: ©2014 Dell Inc. 版权所有。请浏览以下网站获得更多信息: www.dell.com.cn/product_information。



作者: 王丛 中桥调研咨询总经理兼首席分析师
马艳 中桥调研咨询调研分析师
杨凌霄 中桥调研咨询助理分析师

摘要

中桥调研咨询（以下或简称“中桥”）结合2013年就中国市场**455名最终用户**的大数据调查结果，简要介绍了中国大数据市场和技术趋势；并结合5个典型大数据应用场景，浅析了存储在大数据演进过程中的价值。

洞察

1. 大数据近实时分析及实时分析助力业务创新渐成趋势
2. 未来2年中国大数据市场演进
3. 大数据管理复杂性上升
4. 大数据存储新需求：高并行和低延迟



如何通过大数据创造业务价值

大数据创造价值 充分利用大数据：

根据麦肯锡预测，可以使零售利润提高**60%**
制造业研发和组装成本下降**50-70%**

美国健康医疗——
每年减少3000亿美金开支



3000亿

全球定位系统——
1000亿美金 + 服务商销售额



1000
亿美金

Tesco 通过大数据——
每年降低耗能2千万英镑



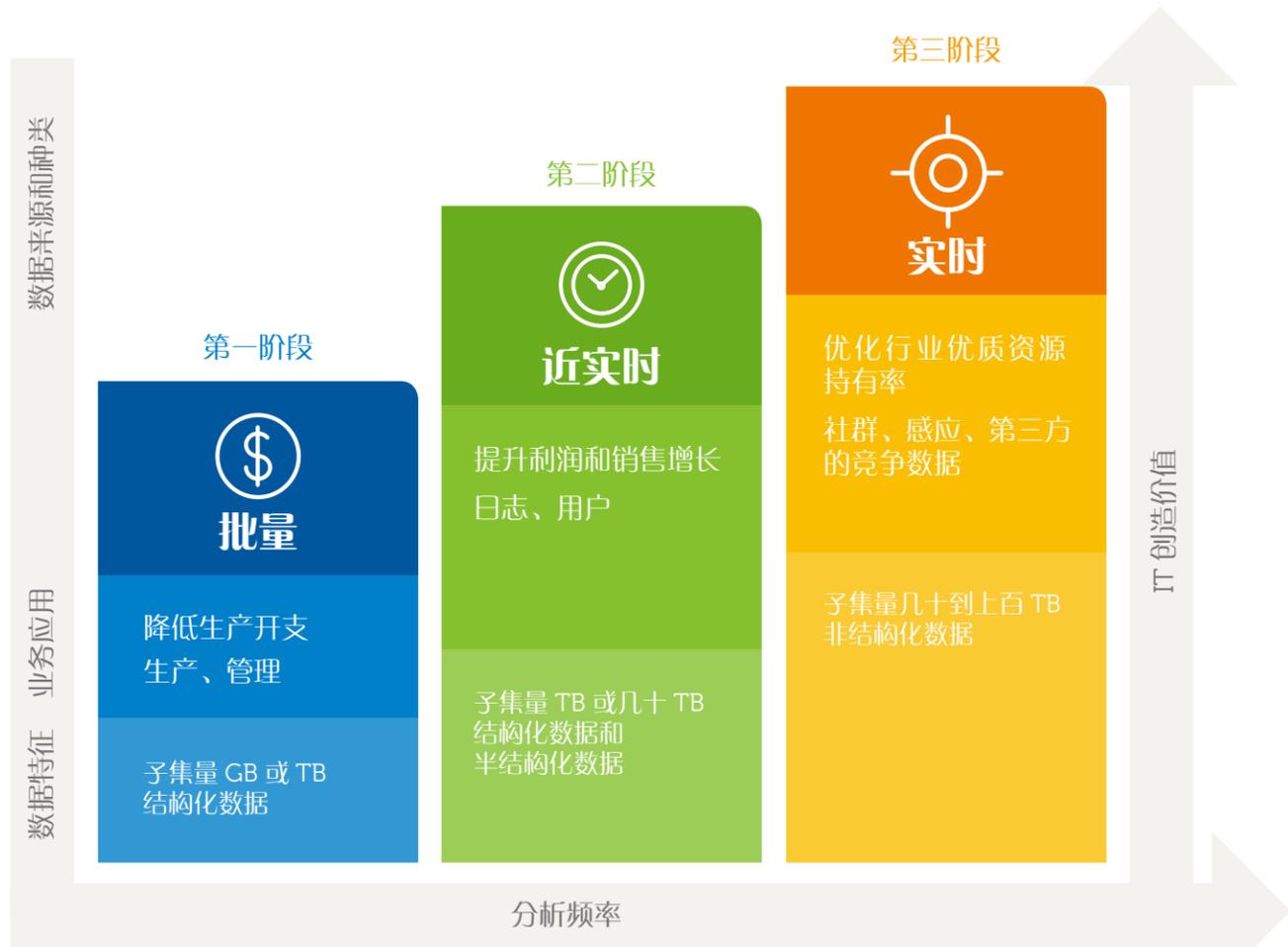
降低耗能
2千万英镑

- 可以实时监控产销链各环节潜在风险
- 提高从资金链到物流链的生产效率
- 为企业提供诸多预见，降低业务风险，更快创建更具细分优势的产品和服务
- 提高优质用户持有率和用户体验
- 提升营销活动投资回报和竞争优势

有助于企业从多维度判断全球市场潜在商机，通过业务模式的突破创新，实现跳跃式快速发展。

大数据分析不同阶段的特点

大数据通过 IT 创造价值的主要维度是 1. 数据分析频率；2. 数据来源和种类；3. 大数据分析效率；大数据是一个演进过程，主要分为三个阶段。



第一阶段

以批量分析为主。数据主要来自企业内部结构化数据（如生产、管理等数据等）。数据分析的主要目的是降低生产开支，提高资金周转和物流效率，提高业务智能决策能力。用户通过提升传统数据分析的数据种类和速度，逐步向大数据分析架构转变。数据分析子集量通常在 GB 或 TB 范围。

第二阶段

以近实时分析为主。数据分析类型逐步从结构化演进到包括非结构化（音视频、社群等）和半结构化数据（系统日志、客户信息等）。提升利润和销售增长、优质客源获取和持有率成为主要目的。用户对从分析、处理到呈现的时效性更强，提高了对数据处理能力和分析速度的要求。数据分析子集量在 TB 到几十 TB 甚至几十 TB 范围。

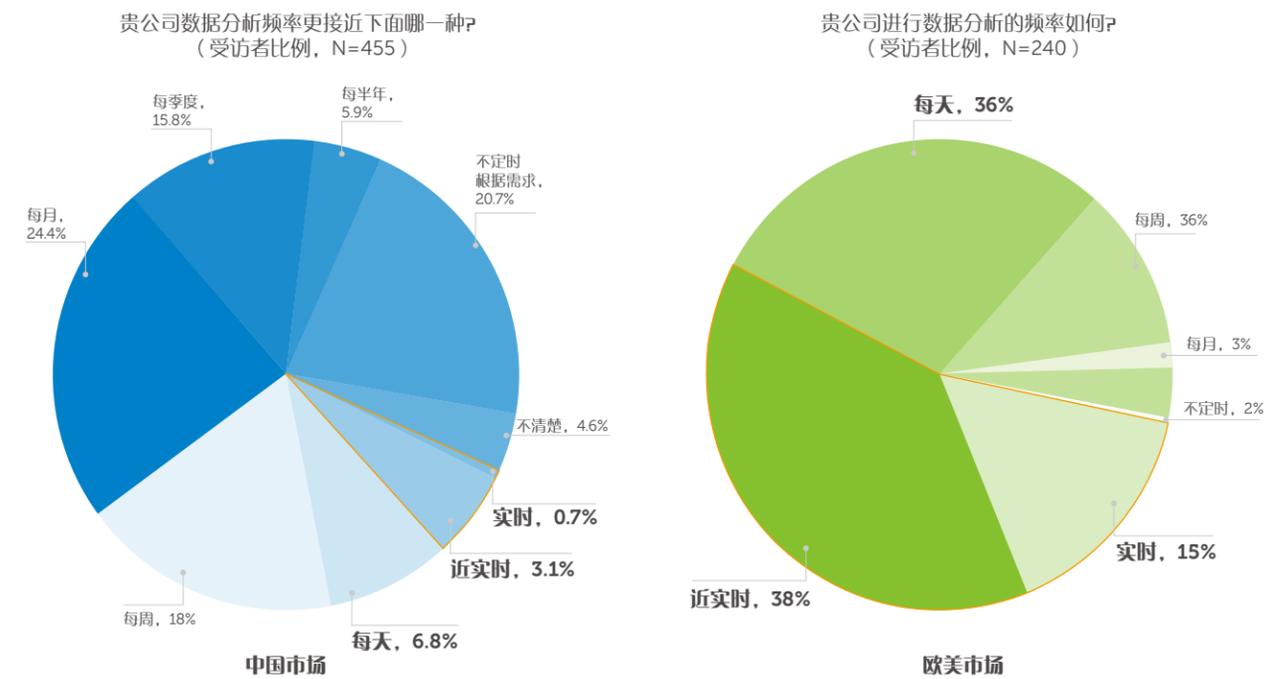
第三阶段

以实时分析为主。数据来源和种类更加丰富，不仅限于企业内部的生产数据、用户数据和社群网站、监控、感应数据，同时会纳入来自第三方的竞争数据，以非结构化数据为主。主要目的是实现业务突破创新，通过系统驱动实时“行动”，提升企业在全市场核心竞争力，优化企业优质资源持有率。实时分析对分析速度要求更为苛刻，大大提高了对计算、网络、数据存储容量和性能、动态资源配置能力的要求。数据分析子集量在几十到上百 TB 范围。



中国市场大数据分析频率

欧美 53% 的企业在大数据分析的第二和第三阶段，而这一数字在中国仅为 3.8%。

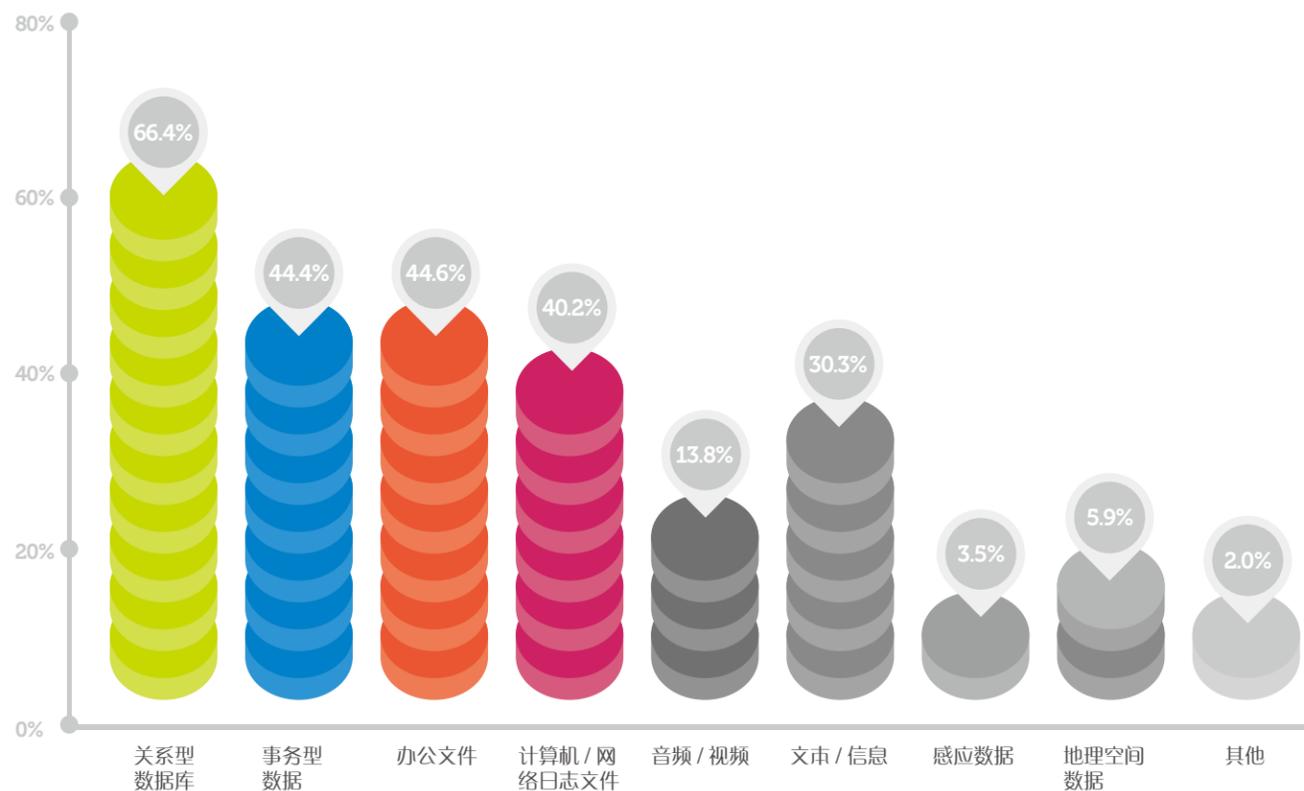


从图 2 的中国市场大数据分析频率调查结果来看，中国市场 90% 以上的用户处于第一阶段，其中近 85% 的用户分析频率以每周、每月、每季度、每半年或按需分析为主；第二阶段和第三阶段的用户占比总和为 3.8%；日分析占比仅为 6.8%。相对而言，欧美市场 53% 的用户已经采用近实时和实时分析；日分析占 36%。

在未来 24 个月，中国市场将快速从大数据分析的第一阶段向第二阶段演进，从大数据分析频率来看，正从批量分析快速向近实时分析演进。在从第一阶段到第二阶段的演进过程中，用户首先要考虑如何提高数据采集和管理效率。这将刺激数据存储容量的增加。此外，大数据分析子集量，以及大数据分析性能，都给传统 IT 提出了更大挑战。

中国市场大数据分析类型和来源

问题：贵公司数据分析（商业智能）包含下列哪些数据类型？（可多选，最多选三项）（受访者比例，N=455）



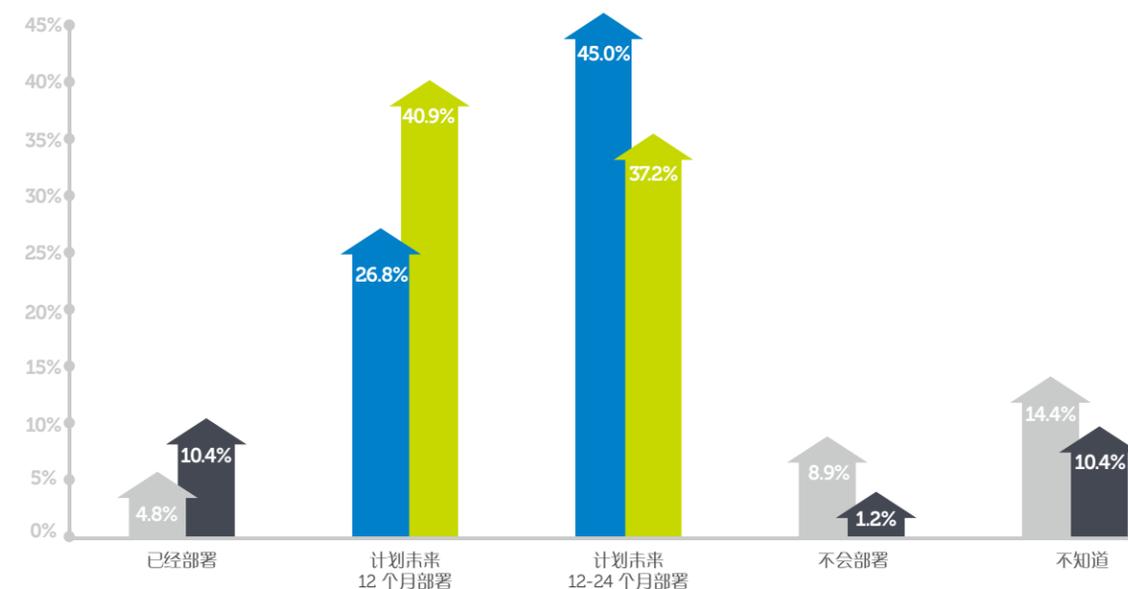
目前中国市场大数据分析主要围绕企业内部的结构化数据，如数据库或事务型数据。而半结构化和非结构化数据在中国大数据分析的占比在未来 24 个月将会快速增长。如何有效应对半结构化、非结构化数据的有效管理，保护大数据 IT 创造价值的数源，将是决定企业未来 5 年是否能够通过 IT 提高竞争力的重要因素。

大数据分析市场趋势

未来 24 个月，大数据分析 IT 投资呈强增长。

企业级
提高生产效率，决策精准度；
提高用户体验，降低优质客户流失

中小企业
实现业务突破创新，以“小”搏“大”，以“速度”搏“规模”



目前，中国用户在大数据通过 IT 创造价值的过程中，亟待快速提高数据分析频率、增加数据来源、提升数据整合和分析速度。数据是大数据时代通过 IT 创造价值的“种子”。大数据分析的四个重要环节：数据采集和存储；ETL（Extraction-Transformation-Loading）数据的抽取、转换、加载；数据分析；分析呈现。



如何保证大数据演进过程的存储容量、性能、业务连续性，通过提升资源利用率降低存储开支，是选择大数据存储的重要考虑因素。



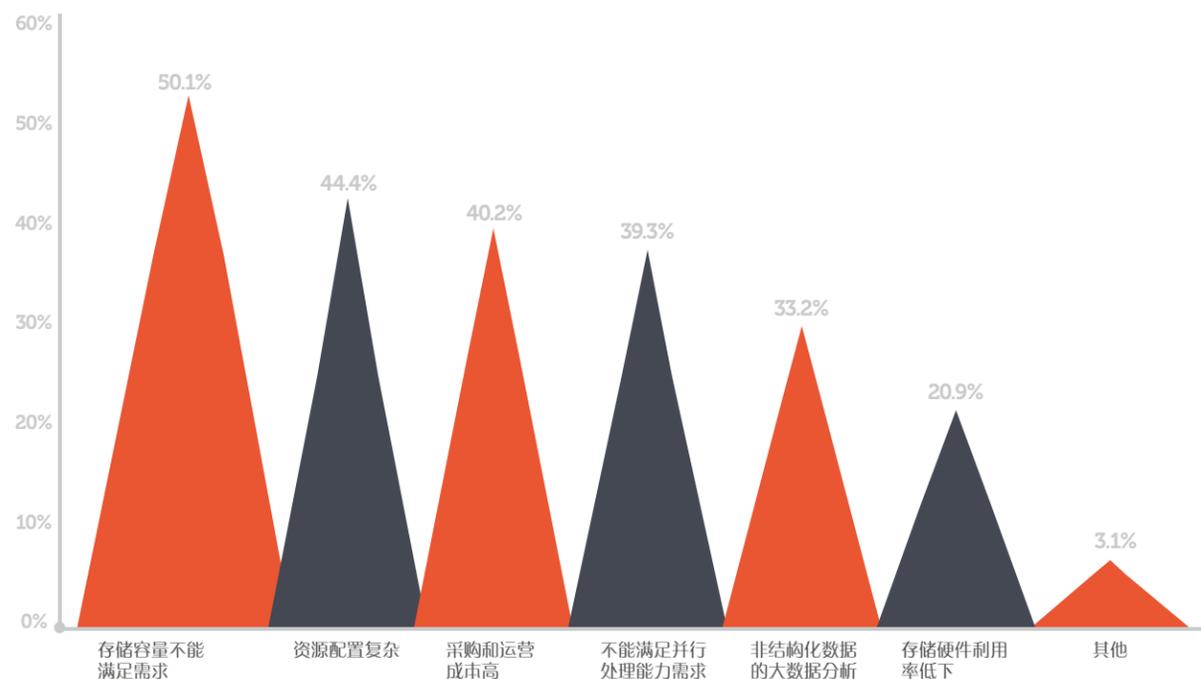
大数据 存储挑战



戴尔 流动数据



问题：支持数据分析和 / 或进程活动方面，存储面临哪些难题？（可多选，最多选三项）（受访者比例，N=455）



大数据时代，在数据量飞速增长的情况下，企业的存储容量面临的压力最大，而应用的多元化和 IT 的分层管理，带来了系统 IT 资源配置和管理的复杂；要满足如此海量数据和多种应用的需求，企业的存储采购和运营成本也居高不下；而多个应用的同时运行，则对系统的并行处理能力提出了更高要求；非结构化数据的大数据分析则进一步反映了企业数据类型的多样性和复杂性。



为了满足大数据时代的 IT 需求，企业需要大容量、高性能、保证数据生命周期高性价比的存储来满足大数据存储、数据保护和业务连续性需求。

戴尔流动数据之道 “精耕细作·物尽其用”

大数据环境中，随着数据量、应用数量和应用使用者数量的增加，为了保证应用性能，企业的 IT 系统不仅要充分发挥各种存储、网络 and 架构技术的作用，而且通过“精耕细作”实现工作负载层的自动化配置和资源优化，要求存储能通过革新架构，结合高智能，细粒度地实现“物尽其用”，即动态地将最好的资源配置给所需负载，以满足业务性能需求，这是满足大数据性能需求的关键。

软件定义的优化存储

戴尔流动数据架构还结合一系列的存储容量优化技术（自动精简、数据压缩和重复数据删除），使系统对网络带宽资源的需求降低 90% 以上，大大减少了资源浪费。集群、闪存技术（SLC 和 MLC）和固态硬盘分层技术的结合，有效提高了 IOPS，降低了时间延迟，实现了大数据对 OLTP 和 OLAP 的高要求，以及存储容量可扩展和资源优化的目的。

高利用率

结合一系列的存储容量优化技术，降低大数据生命周期的存储、网络资源需求达 90%。

自动精简

数据压缩

重复删除

只根据写入数据配置空间，可以大幅度提高存储资源的利用率，大大降低存储新增采购的需求。

数据压缩和重复数据删除：大幅度降低了对网络 and 存储资源的需求，降低了总存储拥有成本。

戴尔流动数据： 适应大数据“四高”存储需求的新生代架构

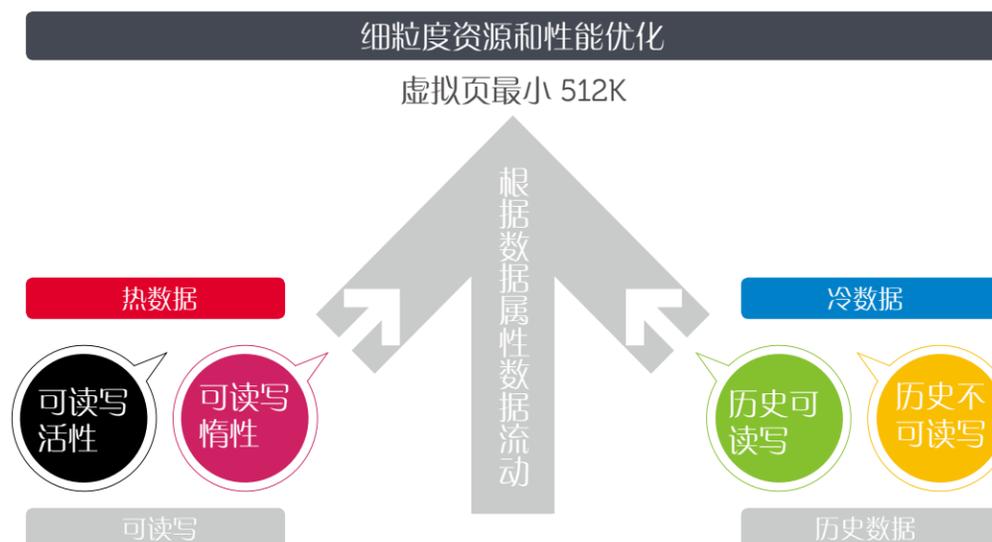


戴尔流动数据架构针对这种市场需求，所有产品以模块化的形式，实现了存储容量和性能升级扩展的简单快捷；横向扩展可以跨不同节点扩展，保证了在容量扩展时性能不会衰减，以及容量和性能的独立升级线性扩展；同时跨不同节点实现了无断代技术升级，避免了叉车式断代升级带来的管理难度和对业务连续性的影响，确保了存储生命周期的高可扩展性。

感知冷热，自如“流动”

- 数据活性标记
- 自动化跨层迁移
- 业界最细粒度存储块

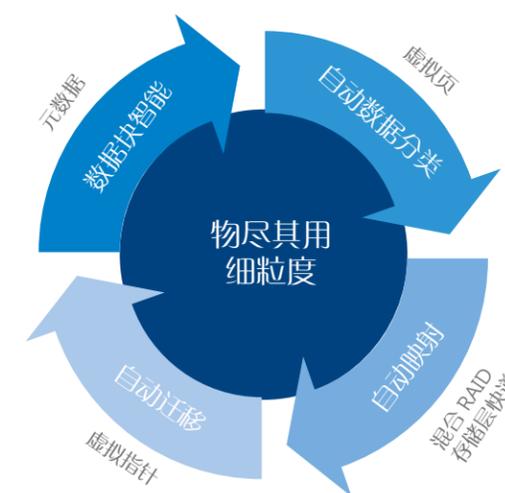
戴尔流动数据架构的数据块级智能和动态资源配置，可以根据数据属性，灵活地对数据块进行自动化分类，根据数据类别实现自动映射和数据自动分层迁移。



“数据”为中心的智能管理

根据数据属性动态配置资源，自动迁移、细粒度优化，满足对存储效率和性能的要求

戴尔通过一个“智能”（数据块级智能）和三个“自动化”（数据分类自动化、映射自动化、迁移自动化）、流动数据的分层结合、不同 RAID 级别和存储层，满足不同工作负载对高可用、高性能和容量的需求，以最微粒化的维度提高大数据分析过程中的资源利用率。



典型大数据应用（真实行业案例解析）

在大数据演进过程，大数据分析效率可以提升企业的市场反应速度和突破创新能力；同时，提升大数据演进过程业务处理能力、业务可扩展性，决定着业务发展空间和利润空间。下面，我们对 5 个场景做了简单分析：

1. SQL 大数据近实时分析
2. SAP HANA 大数据近实时和实时分析
3. Hadoop 大数据实时分析
4. Oracle 数据库性能及扩展性提升
5. “启动风暴” VDI 性能提升



挑战

英国某卫生行业云计算提供商，为 82% 的国家健康服务机构提供产品或服务。它通过对来自各种医疗机构的运营信息（判断类似病例的数量、范围和威胁性）、就诊病患信息（通过历史数据提取、及时更新数据实现跨地域名医实时会诊）、急救信息（判断死亡率和危害性）、社群网站和手机短信（增加数据来源，提高采样涵盖率）、卫生监控信息等多种大数据实现商业智能分析，针对任何可能威胁社会健康的疾病，为医疗机构提供近实时监控和预测信息，提高健康突发事件的应对能力。该提供商对于客户各种需求的满足，加大了对数据存储容量和性能、大数据分析频率和速度的要求。

价值

医疗卫生是大数据演进过程通过 IT 创造价值受益最大的行业。据麦肯锡预测，充分利用大数据分析，美国医疗卫生行业每年开支可以**节省 3000 亿美元**。实现这个目的，医疗卫生行业需要完整、准确地收集数据，近实时和实时分析，以及根据个性化需求呈现分析结果。



存储需求

扩展和存储利用率：

欧美医疗行业医疗档案需要终生保存，大量图像（如 X 片、CT 影像、PACS 信息）信息导致医疗卫生的数据存储总量持续快速增长，对存储的高可扩展性提出高需求。各种文件生命周期阶段读取性能大相径庭，对数据生命周期智能管理提出更高要求。为了满足 HIS（医院信息系统）和 PACs（影像系统）的存储需求，统一存储既要满足大数据演进过程事务处理性能需求，又要满足非结构化和半结构化数据的容量需求。

近实时分析性能：

通过 SQL 实现近实时大数据分析服务，对 OLTP 和 OLAP 都提出更高要求。同时，相对于批量分析，近实时分析的数据子集量更大。通过分级存储可以提高资源利用率，但传统通过卷迁移实现的分级会导致迁移到底层的数据读写性能难以满足近实时分析的需求。

大数据近实时分析应用



该提供商结合 Microsoft SQL Server 2012 Enterprise 软件和商业智能工具，通过 Windows Azure HDInsight Service 的混合云平台，为各种医疗机构和医生提供用户可选择的商业智能分析服务。HDInsight 的大数据功能可以处理大量结构化、半结构化和非结构化数据，并对上述数据进行实时或近实时分析。

解决方案

横向扩展、无断代升级：

戴尔流动数据架构可以实现控制节点和容量节点的独立升级和线性扩展，保证容量高可扩展性的同时，不会出现性能瓶颈。同时，通过增加 NAS 节点，单一域名可以管理 **PB 级数据**。此外，通过微粒度可以提高存储资源利用率，无断代升级存储架构可以降低存储新增采购和存储管理成本，实现降低存储 TCO 的目的。

高性能、低延迟：

戴尔流动数据架构结合全闪存和固态硬盘分层等技术，通过 SQL 提供大数据近实时和实时分析所需的 OLTP 和 OLAP，并满足低延迟需求。流动数据通过数据自动沉降，让用户快速将最优质存储资源让位给活性数据。结合 MLC 和 SLC 固态硬盘技术，戴尔使分层存储在性价比上最具有竞争力。

分层数据高性能：

戴尔流动数据架构从根本上解决了传统以卷迁移为机制的数据分层带来的数据分析性能问题。以卷迁移实现自动分层，不仅很难实现细粒度化（512KB 单元），同时，迁移到较低性能存储层的数据读写性能很难满足大数据近实时和实时分析需求。戴尔流动数据架构，通过其具有专利的数据自动沉降、虚拟化指针（参见“戴尔流动数据”部分）技术，保证分层数据不降低数据读取性能，在实现存储资源优化的同时，满足大数据分析的性能需求。



SAP HANA 大数据近实时 和实时分析

电信行业
大数据应用



存储需求



快速部署

在部署和管理过程，如何快速部署 SAP HANA 方案，实现通过 IT 创造价值，给 SAP HANA 潜在用户带来巨大挑战。



近实时和实时分析性能

SAP HANA 作为近实时和实时大数据分析平台，对存储容量、存储性能、数据高可用性，以及数据保护提出了高要求。



子量级容量和性能

近实时和实时分析数据分析子集量大。如何有效利用一层存储性能满足子集容量和性能要求，这就给传统存储带来了挑战。

价值

优质客源获取和持有率直接决定着电信行业服务商的业务增长和利润。

对于电信供应商，降低 1% 的客源流失，就会增加 6% 的销售额。



挑战

美国某电信供应商，其面临的问题是，**如何通过大数据分析，将不同来源的海量数据进行快速近实时分析，通过零售、客服中心等多种渠道为 2100 万用户提供所需服务。大数据分析的速度、准确性和实时性，直接决定着该公司业务增长和利润空间。**随着数据量的快速增长，该电信供应商的 IT 系统已经难以满足大数据过程业务对分析速度的要求。这直接导致优质客源快速流失，产品和服务接受力度差，服务对市场响应速度慢。由于分析资源有限，该公司无法将 Facebook、Twitter 等社群网站信息，以及产品传感信息纳入大数据分析，进一步降低了其市场竞争力和服务水平。

应用

2012 年，该公司部署了 SAP HANA，对多种维度的数据进行大数据近实时和实时分析，

- 数据分析速度提高了 **56 倍**
- 可以在 **3 个小时**内完成来自 **2100 万客户**的 **20 亿记录**。
- 使防止丢失客户的单位成本下降了 **10-25 美元**。
- 每年通过降低优质客户流失而产生的销售额达 **几十亿美金**。
- 大幅度提高了个性化产品和服务推出的速度和投资回报，**提升了整体核心竞争力**。

戴尔 SAP HANA 存储和解决方案

戴尔提供了 SAP HANA 认证整体解决方案：

戴尔高性能服务器采用英特尔® 至强™ E7 多核处理器，结合 512GB DDR3 缓存，提供了 In-Memory 所需的处理性能。结合 PCIe SSD 和戴尔全闪存以及 SSD 分级技术，满足了高性能需求。通过将高性能和高可扩展计算、网络和存储节点与 SAP HANA 整合，以及管理整合，最大限度降低了 SAP HANA 部署、调优和管理所需的资源。

数据子集性能和容量：

戴尔全闪存结合 MLC 和 SLC 性能，大幅度提高了分层存储的性能和容量，为 SAP HANA 大数据近实时和实时分析提供了最具高性价比的存储，满足了子集性能和容量的需求。通过与硬盘形成整体虚拟池，以及无断代升级扩展，保证了整体存储容量扩展性的需求。

戴尔流动数据架构的细粒度分层：

优化 SAP HANA 应用的存储资源利用率。结合数据块级智能，以及自动分类、自动映射和自动迁移，可以根据业务需求自动调整资源，满足性能和可用性需求，大幅度提高 SAP HANA 环境存储配置和管理效率。同时，降低了人为误操作可能带来的业务风险。戴尔流动数据架构通过精简快照、精简复制、Live Volume 等技术，提高了数据保护级别。同时，保证硬盘、系统，甚至于站点故障不影响业务运营，实现数据的高可用性。



Hadoop 大数据近实时 和实时分析

云计算服务
大数据应用



存储需求



如何有效存储扫描、监控和分析的海量数据，既保证实时分析所需要的性能，也能优化存储资源利用率，是保证 SLA 服务水平和控制存储成本的关键。

随着客户量的增加，数据量快速攀升，如何保证存储高可扩展性、数据高可用性，以及存储动态资源配置能力，是选择存储的关键。

价值

信息安全成为业务安全最重要的防线。SecureWorks 作为信息安全服务商，帮助从政府到企业的各种用户通过 Web 应用、网络、日志等扫描及全天候安全监控，提高信息安全性。SecureWorks 结合 Hadoop、Cloudera 和戴尔 Crowbar，利用云计算构建组元的戴尔 C 系列（包括服务器、存储和网络）搭建大数据分析技术平台，提供信息安全服务。通过大数据实时分析，**SecureWorks 提高信息安全服务水平，通过优质客源的积累，实现业务快速增长。**

挑战

- Hadoop 作为一个开源软件，通常需要用户具有较强的再开发能力，以及技术整合和运维能力。
- Hadoop 环境资源配置复杂，耗时且人为误操作风险高，给业务的稳定安全性带来挑战。
- Hadoop 的节点架构，可能导致单点故障，增加业务风险。
- Hadoop 对存储容量需求大，如果不能对数据进行优化（压缩和重复数据删除），会带来大量存储资源和网络资源浪费。



应用

- SecureWorks 采用了 Crowbar 和 Cloudera 简化 Hadoop 部署，提高了在线资源的配置效率以及应用配置效率。利用 Hadoop 实现对海量扫描和监控数据的实时分析。
- 戴尔 Crowbar 让 SecureWorks 实现 IT 资源配置和信息服务标准化、资源管理流程化和资源监控实时化，以及资源和性能管理自动化。不仅降低了 IT 运维资源，同时，最大限度降低了服务运营层面的风险。
- 戴尔云计算构建组元的 C 系列服务器，能够根据业务需求快速以插板模式增加计算、内存、网络 and 存储资源。通过冗余和故障切换，消除单点故障可能对业务的影响。

戴尔流动数据方案



1 戴尔 Crowbar 结合 Cloudera Hadoop 方案：简化了 Hadoop 的方案整合，提高了存储性能，降低了业务风险。为从 IT 架构管理，向 IT 服务交付提供了标准化、规范化、自动化的 IT 管理平台。



2 结合戴尔最近发布的全闪存技术：结合 SLC 和 MLC 技术，可以经济地实现 IOPs 可达 30 万以上，满足 OLTP 和 OLAP 的性能要求。同时，综合 SLC 和 MLC 技术，使 IOPs 性能提升的开支降低 **1/3**。



3 戴尔的 C 系列云计算构建组元：包括 Force10 高性能网络插板、EqualLogic 横向扩展存储。同时，计算节点可以选择 GPU 模块，满足高性能计算所需要的并行处理能力。



4 戴尔流动数据架构在单一域名下可以支撑 PB 级数据量：节点集群架构和自动分级，让 SecureWorks 可以保证容量扩展过程稳定的业务性能。EqualLogic 可以跨 16 个节点和不同的云计算构建组元通过 Peer 架构，实现高可扩展。无断代技术升级保证了存储的长期投资回报。同时，重复数据删除和数据压缩大幅度降低了存储容量和网络带宽的资源消耗。



Oracle 数据库性能及扩展性提升

教育科研行业
大数据应用



存储需求



该服务商数据库容量和课件的非结构化文件存储容量快速攀升。就数据库而言，活性数据仅占数据总量不到 10%。如何提高数据库存储资源利用率，是控制存储成本和满足各种应用所需性能的关键。在管理这样庞大的数据库集群过程中，只有细粒度化资源优化、自动化数据流动，才能满足高自动化 SLA 服务交付的需求以及批量分析所需的 OLAP 性能。

数据库是业务的灵魂。保证数据库所需要的高 OLTP 和 OLAP，是确保 SLA 服务质量的关键，这就对存储的 IOPS 和低延迟提出更高要求。数据库数据的高可用和数据保护，成为保证该服务商服务的必要条件。这就对高可用性、异地容灾、高效数据保护提出更高要求。

此外，该服务商需要能实现对结构化和非结构化（课件）存储的集中统一管理，降低存储管理强度。确保课件所需要的容量和吞吐量，以及课件的在线响应速度。

价值

无论在中国市场还是全球市场，大数据时代，教育和科研通过云计算，降低 IT 的 CapEX 和 IT 管理成本，让更多的 IT 资源用于提高教育多元化和科研社会化效率，是教育科研的 IT 发展趋势。纽约某教育信息交换中心支持 64 所不同的高等教育机构，为用户提供网上推广、ERP 等多种应用；同时，为 24 所高教机构管理 140 个 Oracle 数据库、50 多种 Oracle 应用，对各种数据进行批量分析。**该信息交换中心通过为高教和科研院所提供高质量的数据库、分层应用和数据分析服务，业务得到快速发展。**

挑战

- 对于服务于教育行业的云计算服务商，如何快速增减数据库节点，有效满足各种用户不同业务周期的需求，是该云计算服务商提高用户体验、提升利润的关键。
- 保证大量数据库、ERP 应用，以及网上应用所需性能，要求系统动态满足混合负载性能的同时，还要满足低延迟的需求。
- 教育系统的各种数据在进行批量分析时，还要求系统在提升存储利用率的同时保证分析速度。
- 如何保证在学生业务峰期，激增流量不会影响系统响应速度，保证万人同时读取的交易响应速度在**微秒级**，使该服务商在大数据演进过程中面临的挑战进一步升级。

Oracle 大数据分析应用

为了提高 Oracle 高可扩展性和高可用性，该服务商部署了 Oracle 网络方案实现数据库和应用整合，结合 Oracle RAC 应用，让该服务商可以快速根据业务需求增减应用实例，不仅提高了应用部署效率，而且大大提高了应用可扩展性和高可用性。保证各种业务周期业务稳定性，决定着用户体验和服务质量。整合后的 ERP 可让**17 万学生同时生成交易处理，1 万 1 千学生同时进行各种课程注册，而且一小时内可以同时运行 7 万课程，处理容量比以前提高了 37 倍**。同时，保证了激增流量（学生注册期）并发查询速度、交易处理速度，以及近实时分析速度。

戴尔流动数据架构

戴尔流动数据架构让用户可以根据业务需求，将活性数据保留在分层存储全闪存，将非活性或归档数据存储 1500rpm、1000rpm 或 7200rpm 磁盘。结合 SLC 和 MLC 全闪存技术分级，可以满足较为极端的 OLTP 和低延迟以及批量分析所需的 OLAP 性能需求。同时，以 512KB 或 2MB、4MB 为单位实现存储分级和数据流动，这种细粒度化数据分级提高了整体存储资源利用率。结合着流动架构的全闪存技术、SSD 分级和各种不同存储技术，降低了整体存储开支。

通过数据块级智能，提高利用率：

结合数据块级智能和虚拟页技术，戴尔流动数据架构让用户制定灵活的 RAID 级别和存储层，以及数据存储的盘道位置，通过各种资源整合效应，确保业务关键型工作负载性和高可用性需求，同时保证分层存储不影响大数据批量分析的性能要求。与诸多传统存储不同，RAID 级别可以根据业务需求或数据生命周期价值实现动态调整，最大限度提升了数据利用率。

结合精简快照、精简复制，戴尔流动数据架构降低了数据保护存储和网络的资源需求：

其 Live Volume 既可以作为数据库容灾方案，也可以作为跨站点 Oracle 存储节点。

流动数据架构通过 NAS 节点，可以管理 PB 级存储容量：

同时，横向扩展保证容量扩展过程的性能满足课件所需要吞吐量，保证课件在所有时段的响应速度。



“启动风暴” VDI 性能提升

保险行业
大数据应用



大数据存储挑战

如何保证启动风暴所需要的 IOPs，同时控制桌面虚拟化存储成本，满足存储容量需求，是该保险用户提高桌面虚拟化投资回报的最大挑战。

价值

大数据时代使保险行业竞争白热化。如何通过增加代理扩大服务地域覆盖率；如何根据市场需求调整服务项目和服务水平；如何通过移动办公提高服务效率和用户满意度，直接决定着保险行业的业务竞争力。因此可以说，大数据时代刺激了桌面虚拟化在保险行业需求的快速增长。

挑战

大数据演进过程，随着桌面终端数量、虚拟应用数据量，以及各种应用非结构化数据量的快速增加，虚拟桌面用户遇到最大的问题是桌面虚拟化启动风暴。启动风暴对业务影响的典型特征是在大量终端启动或关机过程，使用者长期等待，导致生产效率下降。针对某一应用并发使用率大时，出现应用反应速度慢，用户使用体验差等现象。应对桌面虚拟化启动风暴，逐步成为评估大数据演进过程桌面虚拟化投资回报的重要指标。

应用

某保险公司有 450 个保险代理、50 万用户，部署了 1300 个虚拟终端。随着虚拟终端数量以及应用的增多，启动风暴导致每天上班时启动时间在 15-30 分钟。随着越来越多的应用部署，以及查询数量的快速增加，桌面虚拟化性能不仅影响着工作效率，也直接影响着优质客源持有率、业务增长速度和利润空间。同时，桌面虚拟化启动过程带来的资源争用，也影响着其核心应用的性能。

戴尔全桌面虚拟化方案

戴尔 12 代服务器通过 GPU 技术，提高了对于桌面虚拟化图形文件的处理能力，大大降低保险行业处理高清影像文件对终端显卡的开支。同时，通过 C 系列和 VRTX，为保险行业用户提供整体 IT 架构解决方案，降低桌面虚拟化部署周期和管理难度。

决定桌面虚拟化启动风暴的因素包括计算能力、存储网络性能、存储 IOPs 和低延迟。戴尔通过 12G 保证桌面虚拟化所需的计算能力。通过 Force10 和 16GB 光纤通道网络技术，让用户可以在 iSCSI 和光纤技术之间做出灵活选择的前提下，保证网络性能。此外，流动数据架构满足存储性能需求。同时，戴尔各种技术可以通过 VMware、Citrix 或微软的桌面虚拟化管理平面实现集中统一管理，简化部署的同时提高管理效率。

戴尔流动数据架构

戴尔全闪存结合 SLC 和 MLC 让用户经济有效地提升大数据演进过程桌面虚拟化的性能问题，解决启动风暴 IOPs 瓶颈。作为业界最具性价比优势的全闪存方案，戴尔全闪存经济高效地满足了启动风暴所需要的 IOPs 和低延迟需求。通过全闪存技术，存储 IOPs 可以从传统的 7000 提升到 12000。同时，时间延迟可以从传统存储的 8-10 微秒降低到 **1-3 微秒**，提高虚拟化应用的使用体验和虚拟化整体生产效率。

大数据正在改变着各行各业的产业格局和价值链分配

对于中国企业级用户，大数据直接决定着企业在全球化经济环境的竞争力。对于中小企业，大数据带来巨大的通过业务实现创新突破的商业机会。这也是欧美各种规模企业和机构将大数据纳入业务战略，而不只是 IT 发展路线的重要原因。

整体来看，目前，大数据分析尚未成为中国用户 IT 战略重点

目前，中国用户的大数据分析发展阶段远落后于欧美同行。无论是大数据分析频率，还是半结构化和非结构化数据的分析能力，中国 90% 以上的用户尚处于通过大数据分析创造价值的第一阶段，也就是从传统数据分析逐步开始向大数据过渡的阶段。相对而言，欧美市场 53% 的用户已经进入近实时（第二阶段）和实时分析（第三阶段）阶段。

未来 24 个月，中国市场在大数据分析的新增投入将快速增加

即使在第一阶段，中国市场日分析占比仅为 6.8%，欧美市场日分析占比为 36%。中国企业用户已经意识到这种差距给业务带来的风险。40.9% 的企业级用户在未来 12 个月将加大数据分析解决方案投入，通过大数据分析提高企业竞争力。相对而言，中小企业大数据分析投入增长峰期在未来 12-24 个月。45% 的中小企业在这一阶段计划加大数据分析投入，通过大数据实现创新突破，实现以“小”搏“大”，以“迅捷”搏“规模”。

提高数据采集完整性和数据存储管理效率，是大数据分析的起点

大数据时代通过 IT 创造价值将会是一个循序渐进的过程。充分了解大数据对 IT 架构的需求，是企业经济地利用已有 IT 资源实现大数据分析演进的基本点。通过大数据创造价值的过程包括四个重要环节：数据采集管理、ETL（抽取、转换、加载）、数据分析和分析结果呈现。只有拥有完整的数据，才能使数据分析创造价值变为可能。没有数据“种子”，就难以筹划大数据“丰收”。

大数据时代需要高动态、高扩展、高度自动化存储

传统存储在大数据演进过程，无论在容量、性能，还是业务连续性和数据生命周期管理效率，都面临着很大挑战。大数据时代，如何动态满足工作负载性能需求，如何通过智能和自动化最大限度简化存储资源配置和管理，如何保证容量扩展不会带来性能瓶颈，以及实现存储设备生命周期使用管理效率，是保证大数据存储长期高性价比的关键所在。

戴尔流动数据架构，细粒度优化存储利用率，动态满足大数据工作负载性能需求，横向高可扩展为数据高可用和业务连续性提供保证

通过对不同行业大数据演进过程的应用场景分析，戴尔流动数据架构通过数据块级智能和数据自动分类、自动映射、自动分层，实现细粒度化资源优化，动态为不同负载配置合理存储资源。此外，结合着 RAID 和存储分层，保证用户可以为不同应用动态设置子逻辑层实现 RAID 级别。通过灵活的闪存技术选择，结合着 SLC 和 MLC 分层，以及各种磁盘技术（15000rpm、10000rpm、7200rpm），满足大数据演进过程对 OLTP 和 OLAP 的需求，提高存储整体资源使用效率。此外，结合着压缩、重复删除、自动精简和 Live Volume，提高数据保护效率，保证硬盘和存储系统故障不影响业务运营。不仅通过“智能”和“自动”简化大数据环境存储管理强度，降低人为误操作给业务的风险，同时降低数据保护对存储和网络资源需求，降低大数据存储的总拥有成本。