

混合数据仓库：流动、灵活且强大

2016年6月

→ Michael Lock, 分析与商业智能副总裁兼首席分析师



报告要点

p2

大多数 IT 和数据专业人士将数据孤岛视为现阶段面临的最严峻挑战。

p5

顶尖企业采用混合数据仓库架构的几率较其他企业几乎高出一倍。

p7

采用混合数据仓库的企业的信息速度满意度相当于其他企业的 2.3 倍。

p8

混合数据仓库用户的年收入增长率较其他企业几乎高出一倍。

作为大部分关键业务决策的基石，当今数据环境不仅是 IT 基础架构的重要组成部分，也是企业战略的关键环节。然而，随着用户群规模的不断壮大及数据逐步走向多元化，我们存储、管理和组织数据的方法也必须持续调整。本报告深入探索了一种新型数据仓库，不但可以在传统的本地系统环境下运行，还能利用尖端技术和部署方式的潜在优势。

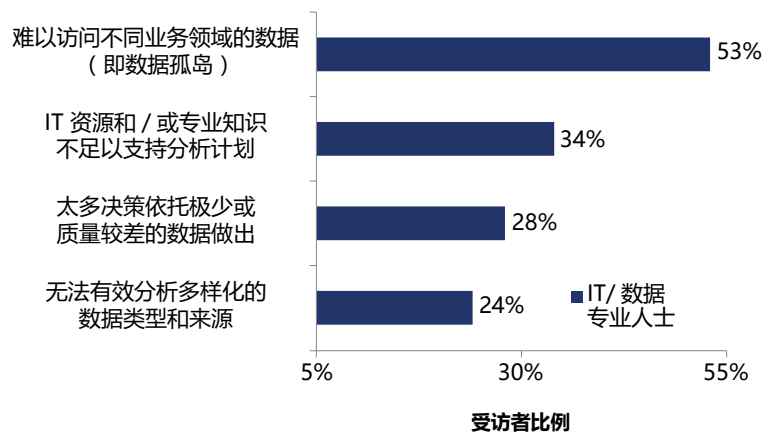
2

目前很多企业仍会受到旧系统和 IT 官僚主义的掣肘，此类企业往往面临一大挑战：分析需求增长往往超出它们的服务能力。

持续满足分析需求

众所周知，分析和数据驱动决策在商业领域日渐盛行。大批相关数据不断涌入，新一代工具利于商业运作且便于使用，为更多非技术用户自行研究数据及发掘洞察铺平了道路。目前很多企业仍会受到传统系统和 IT 官僚主义的掣肘，此类企业往往面临一大挑战：分析需求增长往往超出它们的服务能力。此外，Aberdeen 的研究表明，随着分析活动水平的攀升，支持分析活动的 IT 和数据专业人员在处理数据时面临多重艰巨挑战（图 1）。

图 1：当今数据专业人员面临的重大挑战



n = 70, 来源：Aberdeen Group, 2016 年 6 月

虽然分析需求攀升，但不一定要与交付更多仪表盘或报告的能力挂钩。实际上，用于推动分析活动的数据库基础才是人们面临的首要挑战。其中，数据孤岛现象首当其冲。随着越来越多的企业用户开始积极参与分析活动，好奇心自然会不断加重。

→ 相关研究，“通过云或混合方法执行精益分析”

→ 相关研究“视觉边缘：交互式发现与传统商业智能”

3

为什么东北地区的销售额暴跌？客户对近期推出的产品反响如何？这些都是值得思考的宝贵问题，但为了妥善解答这些问题，通常需要从公司的多个领域广泛采集数据，营销数据、财务数据、销售数据、产品数据、供应商数据等。一旦数据被锁定到这些职能孤岛，前端分析将崩溃，留下用户独自做出假设驱动决策，而非数据驱动决策。

此外，随着新数据源的上线，还要做到双管齐下，因而必然也面临一定的挑战。企业渴望利用不同形式的数据（例如，非结构化社交媒体数据、基于位置的数据），但基础架构往往没有能力安置这些不同类型的数据并保证数据便于访问。此外，伴随新数据源的不断涌现，还会出现人为错误以及数据质量和可用性方面的问题。面对所有这些问题，以下挑战日渐突显：严重缺乏应对所有这些数据复杂困境的 IT 人手和经验，而且无法提供未来企业蓬勃发展所需的分析经验。

一流的数据管理

鉴于这些挑战，领先企业纷纷在各种不同领域采取行动。Aberdeen 近期发布的[数据准备报告](#)展现了一流企业的若干本质特征。从组织角度而言，企业希望加强数据易用性，培养内部分析技能（见侧边栏）。从技术角度而言，企业希望大力培养以下三大领域的的能力：

- **快速检索。**目前，数据紧迫性是影响决策者的一项重大挑战。运营和战略决策者在需要信息的时候，都很难获得他们所需的信息。为抓住大好时机，首先要从数据基础架构入手。一流企业纷纷利用尖端数据仓库技术，此类技术支持在大型复杂数据环境下执行复杂查询及快速检索数据。

实事快报

一流的企业具备以下特征：

- 跨业务职能部门安全共享数据的概率高出 88%
- 部署数据治理策略的概率高出 45%
- 内部培养分析技能的概率高出 68%

相较于所有其他企业

4

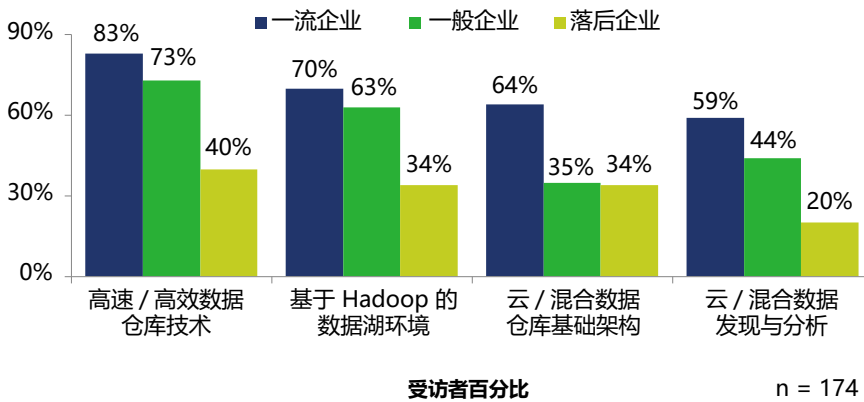
这种技术和基础架构方面的差异极大地推动了混合架构需求。混合架构灵活多样，可以提供多种类型的分析工具及部署选项。

- **灵活部署。**大部分具有分析倾向的企业，特别是大型企业，多年来一直坚持开展技术投资。不仅投资打造各种数据处理引擎和分析工具，还投资开发部署方法（例如，SaaS、公有云、私有云或传统本地解决方案）。这种技术和基础架构方面的差异极大地推动了混合架构需求。混合架构灵活多样，可以提供多种类型的分析工具及部署选项。
- **可扩展性。**数据增长并不新鲜，但多种不同数据（例如，非结构化数据、机器/传感器生成的数据、基于位置的数据）的爆炸式增长再次印证，亟需部署可扩展式数据环境。很多企业同时运用多种开源技术（如 Apache Hadoop，允许大规模管理各种不同数据）满足可扩展性和灵活性需求。

研究表明，一流企业纷纷利用技术改善以上三个方面（图 2）。

5

图 2：适当的技术，适当的背景



来源：Aberdeen Group，2016年6月

值得注意的是，一流企业十分注重混合效率，包括数据基础架构和分析工具。鉴于企业投入大量资金打造 IT 环境，混合方法让企业不仅可以充分利用投资，同时还能探索更灵活、可扩展性更出众的云解决方案。

现代数据仓库的特征

研究依托混合数据仓库架构概念，展现了这些企业呈现出的各种显著的共同特征。首先，采用此类现代数据仓库方法的企业不仅可以利用不断增长的现有数据，还能应对持续增多的数据类型（图 3）。

实事快报

最佳数据准备活动 / 技术
(受访者百分比)

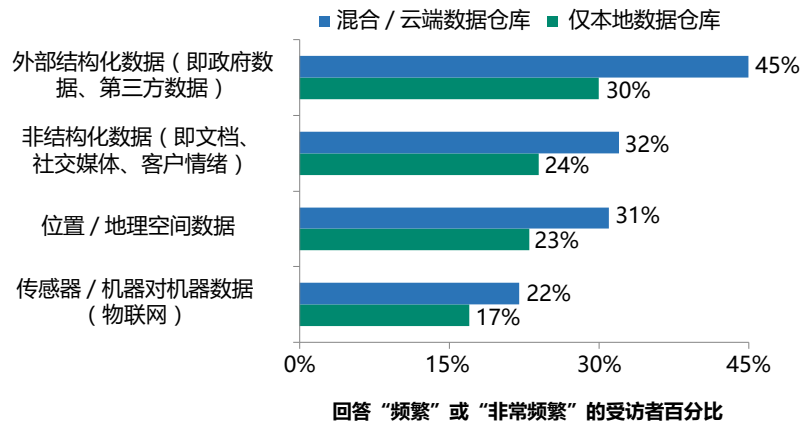
- 分析 – 79%
- 清理 – 64%
- 转换 – 57%
- 建模 – 57%
- 去重 – 43%
- 整理 – 43%

6

→ 相关研究 “数据准备与分析演变”

→ 相关研究 “来自上级的商业智能福音：管理层参与的影响”

图 3：使用的数据广度



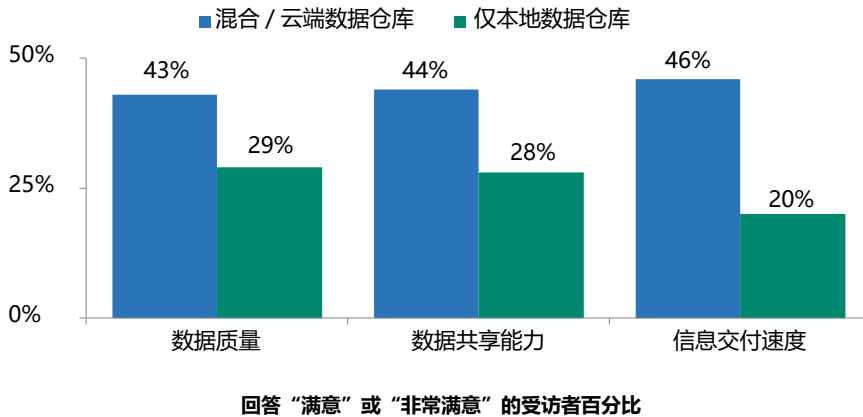
n = 115, 来源：AberdeenGroup, 2016年6月

来自业务应用及其他事务系统的传统结构化数据仍然包含当今企业使用的大量信息，但大多数企业还设法运用不同形式的的数据（如来自文档或社交媒体渠道的非结构化数据以及基于位置的地理空间信息）来丰富分析。综合运用传统数据与其他形式的信息将可揭示重大关联，有助于做出关键决策。例如，结合社交媒体渠道的客户情绪数据与内部结构化产品数据可以发掘洞察，帮助制定产品路线图及传达市场讯息。运用地理空间数据丰富供应链分析，揭示提高运营效率及降低成本的大好良机。采用混合数据仓库方法后，企业将可以抓住机会丰富分析及发掘新洞察。

采用现代数据仓库环境的企业还有另外一项重要特征：它们能够理解并满足用户获取快速流动数据的需求。研究表明，此类企业更倾向于部署流程来收集并满足用户的分析需求。另外，此类企业更倾向于招聘并培养能够利用规模不断扩大的数据的分析人才。或许更重要的是，此类企业更愿意从数据环境的多个关键领域报告用户满意度（图 4）。

7

图 4：快速流动的数据



n = 115, 来源：Aberdeen Group, 2016年6月

数据流动性是一个有趣的概念，也是现阶段很多数据专业人员关注的焦点。流动与流畅、轻松和优雅同义。究其本质，流动数据是指无（最低）障碍数据移动和交换。数据质量差是一大障碍，因为这会引发分析流程耗费更多的时间和精力。数据孤岛是企业内部交流思想和洞察的重要障碍。从根本上而言，所有这些障碍都将导致企业数据传播速度减慢，继而引发信息传递延迟乃至错失良机。采用现代数据仓库方法后，企业将可以对数据环境的以上三大关键领域做出改进，减轻信息流障碍，支持用户顺利完成流程。

实事快报

采用云或混合数据仓库架构的企业具备以下特征：

- 使用基于云的传统 BI/ 外包报告的概率高出 **3.8 倍**
- 采用基于云的数据发现 / 交互可视化的概率高出 **4.3 倍**
- 采用基于云的预测性分析的概率高出 **3.2 倍**

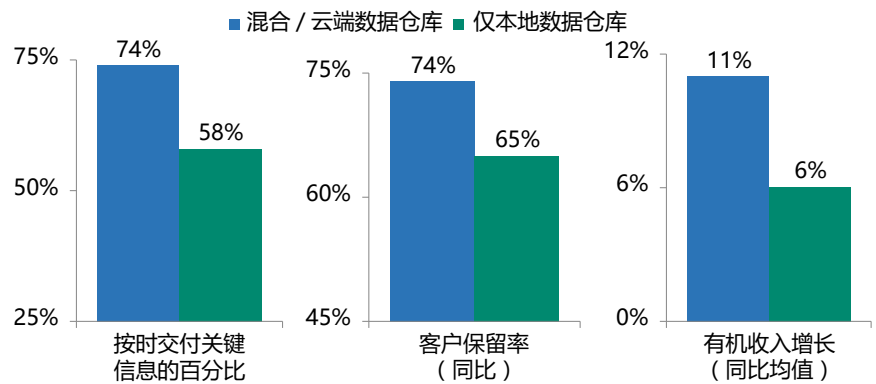
8

结合高效数据环境与强大分析活动，为客户开展交叉销售和追加销售、提高保留率及增加收入创造机会。

推动实现业务成果

究其根本，若不能取得明显成效，所有这些努力都将徒劳无功。研究表明，除了组织成熟度和技术使用广度提高以外，这些企业的以下三大关键指标也有所改善（图 5）。

图 5：将数据转化为成果



n = 115，来源：Aberdeen Group，2016年6月

专注提高数据流动性，以便更快、更清晰地交换信息，从而更及时地提供关键洞察。随后，将可利用这些洞察满足客户群的紧急需求，帮助提高客户满意度和保留率。各种因素均有助于增加收入，但或许没有哪种方式比深入了解现有客户更直接了。结合高效数据环境与强大分析活动，为客户开展交叉销售和追加销售、提高保留率及增加收入创造机会。

关键点：

如果说一流的数据驱动决策和颠覆性洞察是当今企业开展分析的关键原因，那么清洁、可靠而又及时的数据则是强大后盾。鉴于分析技术和支持基础架构不断演进，很多企业的复杂传统硬件

9

和软件系统网络与现代云架构松散耦合，这种局面不足为奇。倘若企业在数据仓库环境中采用混合方法，将会发现数据与基础架构挑战不足为惧。研究结果，人们需要仔细留意以下三大概念：

- **数据孤岛是一个重大问题（也是一次发展机遇）。**无论技术用户还是非技术用户，均渴望运用更丰富、更多样的数据来支持分析。因此，数据孤岛效应成为现阶段影响企业的首要数据相关挑战，这一点不足为奇。另一方面，若能有效应对这一挑战，提高数据的可访问性和可移植性，企业必将获得大量收益。领先企业纷纷采取措施管理并控制数据使用，同时加强整个企业的信息流动。这些领先企业不仅数据流动性更强，而且用户满意度也更高。
- **领先企业探索尖端数据仓库技术。**与上述数据孤岛效应一样，数据环境中的复杂性和差异性挑战也是一把双刃剑。在探索新技术的同时管理陈旧系统并从中攫取价值显然困难重重，但同样会带来很多机会。倘若企业可以利用混合数据仓库环境和尖端技术提高速度并增强可扩展性，那么不仅可以从容应对这些挑战，还能更有效地利用数据。

10

→ **成效显著而且可以实现。**整理无懈可击的 IT 实施 ROI 案例绝非易事。但现在，人们可以比以往任何时候更轻松地利用高效分析来提高管理水平。强大而又灵活的数据环境可以发掘更丰富、更及时的洞察，从而提高决策流程效率。接着，这些数据驱动洞察又可以为促进业务增长及提高效率创造大量机会，最终切实提高收益。

要了解有关本研究主题或其他研究主题的更多信息，请访问 www.aberdeen.com。

相关研究

[通过云或混合方法运行精益分析](#)；2016 年 4 月

[视觉边缘：交互式发现与传统商业智能](#)；2016 年 4 月

[数据准备与分析演变](#)；2016 年 3 月

[来自商机的商业智能福音：管理层参与的影响](#)；2016 年 3 月

[数据质量计划的三大 ROI 级别](#)；2016 年 3 月

[商业智能在中型市场表现抢眼：敏捷版企业](#)；2016 年 2 月

作者：Michael Lock，分析与商业智能副总裁兼首席分析师
(michael.lock@aberdeen.com)

11

关于 Aberdeen Group

自 1988 年以来，Aberdeen Group 一直坚持发表调研报告来帮助全球各地的企业提高业绩。我们的分析师使用专有分析框架来提取有证可循、与供应商无关的深刻洞察，通过对行业从业者开展一手调研来识别出优秀企业。由此而产生的调研报告被成千上万的商业专业人士用于推动更明智的决策和改进业务战略。Aberdeen Group 总部位于马萨诸塞州波士顿。

这份报告是 Aberdeen Group 开展的主要研究的成果，代表了该报告发表之时可用的最佳分析成果。除非另有说明，否则本出版物的全部内容的版权归 Aberdeen Group 所有，未经 Aberdeen Group 事先书面许可，不得以任何格式或以任何方法复制、分发、归档或传递本出版物。