



大数据发展趋势

作者： 彭雅芳 中桥调研咨询分析师

马艳 中桥调研咨询分析师

日期： 2014年6月

概要：互联网金融、移动终端支付的出现和壮大加剧了金融企业对与大数据利用的紧迫性。中桥调研指出，金融企业已经开始利用大数据分析去挖掘数据背后的价值，但是在这之中金融企业还面临着挑战。

概况

随着数据收集能力逐步提高，金融企业将形成时间连续、动态变化的面板数据，其中不仅包括用户的交易数据，也包括用户的行为数据。对于金融企业来说，简单的数据收集是远远不够的，还需要对大数据进行深度挖掘。只有对金融数据进行复杂分析，才能快速匹配供需双方的金融产品交易需求，才能发现趋势和隐藏的信息，才能让金融企业洞察和发现商机。如何对多源数据实现快速高效的海量数据处理？如何应对互联网金融产生的碎片化数据、快速响应需求引发的风险问题？如何充分利用数据分析、挖掘来获取更大的经济效益？是金融企业在进行大数据分析时面对的几大挑战。中桥咨询调研（以下简称“中桥”）就大数据分析在金融行业中的应用和发展情况进行了调研，以此帮助用户洞悉整个行业现状。

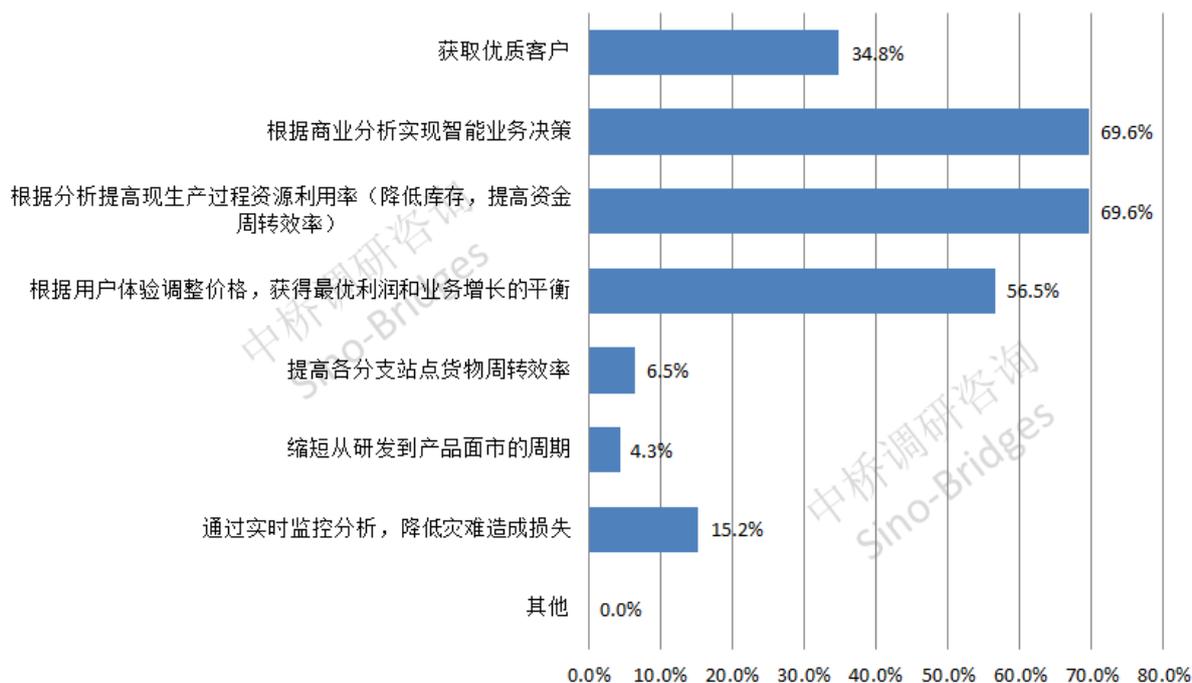
金融行业大数据分析应用阶段现状

大数据的价值

一直以来，金融行业对数据的重视程度非常高。随着移动互联网发展，各种金融业务和服务的多样化和金融市场的整体规模扩大。对于大数据分析带来的主要业务价值，参加调研的金融企业表示，大数据分析的价值是他们可以根据商业分析实现更加智能的业务决策，让决策的制定更加理性和有据。依靠有前瞻性的决策，实现生产过程中资源更优化的分配，能够根据市场变化迅速做出调整，提高用户体验以及资金周转率，降低库存积压的风险，从而获取更高的利润。

图1. 大数据的业务价值

贵公司认为大数据分析的主要业务价值？（可多选，最多选三项）



数据来源：中桥调研咨询公司

此外，据中桥对“金融企业未来重要的 IT 举措”调查中还可以看出，金融行业视数据分析为其重要的 IT 举措之一。其中，47.8%受访者表示数据分析举措是他们未来 12 个月的最重要 IT 举措，另有 50.0%的受访者表示数据分析是他们的重要 IT 举措。通过数据分析获取多维信息，是加速决策的制定，提升企业核心竞争力的关键因素之一。并且，随着移动互联网的发展，不仅带来更加繁多的数据，而且互联网企业“跨界金融”也对传统金融行业带来一定的冲击。传统金融业要进入这个时代，要实现“以客户为中心”的精准营销、风险控制，提升行业竞争力，也到了一个不得不拼数据的时代。数据分析是金融行业在今后大力投入的一个重要方面。

金融行业大数据分析频率

大数据是一个演进过程。大数据通过 IT 创造价值的主要纬度是数据分析频率、数据来源和种类，以及大数据分析效率，主要分为三个阶段：

阶段 1：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）。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，逐步向打大数据分析架构转变。数据分析子集量通常在 GB 或 TB 范围。

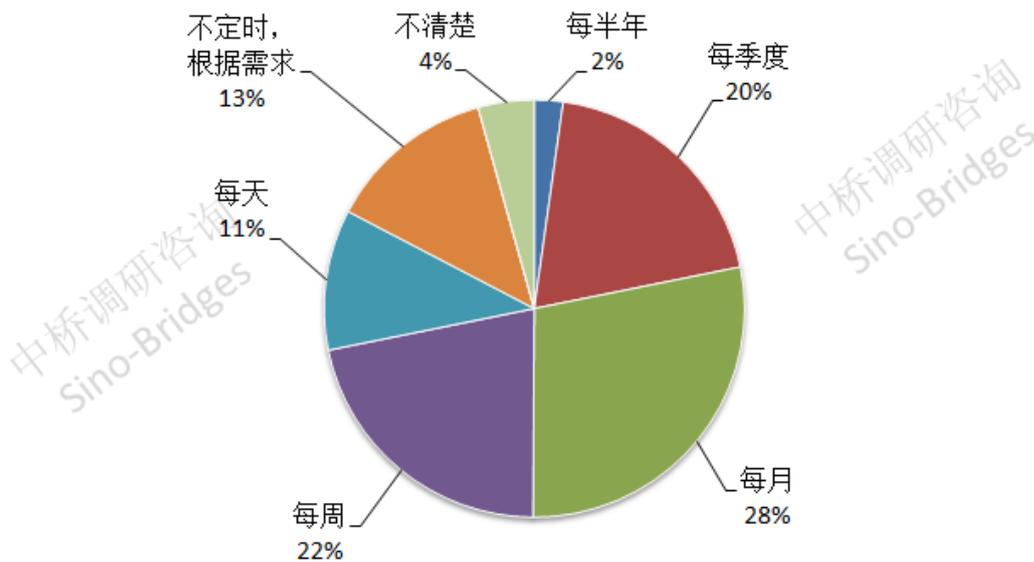
阶段 2：以近实时分析为主。数据分析类型逐步从结构化演进到包括非结构化和半结构化数据。提升利润和销售增长、优质客源获取和持有率成为主要目的。用户对从分析、处理到呈现的时效性更强，提高了对数据处理你能给力和分析速度的要求。数据分析子集量在 TB 到十几 TB 甚至几十 TB 范围。

阶段 3：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”，提升企业在全世界市场上的竞争力，优化企业资源持有率。实时分析对分析速度要求更为苛刻，大大提高对计算、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甚至上百 TB 范围。

那么金融行业在对大数据的利用处于那一个阶段呢？从数据分析频率上来看（图 2），70%受访金融行业对大数据分析频率主要以每周、每月、每季度、每半年为主，而日分析只占到 11%。可以看出，尽管在谈到大数据的时候，金融行业是应用大数据的重点行业，但是在分析频率上，大部分受访金融企业还没有进入近实时或实时分析阶段，也就是说，金融行业在大数据分析上还处于第一阶段。

图2. 大数据分析数据来源

目前，贵公司商业智能（BI）分析频率更接近下面哪一种？



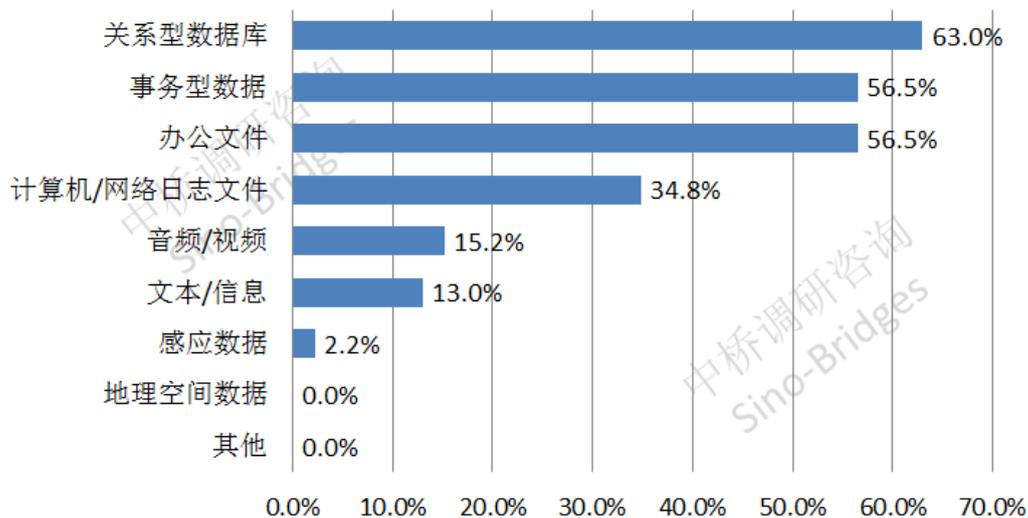
数据来源：中桥调研咨询公司

金融行业大数据来源

目前金融行业主要是通过数据分析来提高整个企业的运营效率，降低运营成本。既然金融企业首肯大数据对于他们可以带来更多的利益。那么金融行业大数据来源有哪些？中桥（图3）对数据类型的调查结果来看，目前，关系型数据库、事务型数据、办公文件数据是最主要的数据来源。从另一个方面来说，就是金融企业在目前进行的大数据分析中是以结构化数据为主，非结构化数据的应用比例相对来说较少。但是随着互联网金融来袭，金融业也需要利用网络的数据，来自社交媒体、传感器、地理信息、音视频等信息。互联网或者大型应用系统中，日志的产生和记录是非常重要的事情。日志分析则是进行数据挖掘进而推进下一步工作的基础。计算机/网络日志文件、音频/视频等也逐渐成为金融企业的数据来源。

图3. 大数据的来源

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）



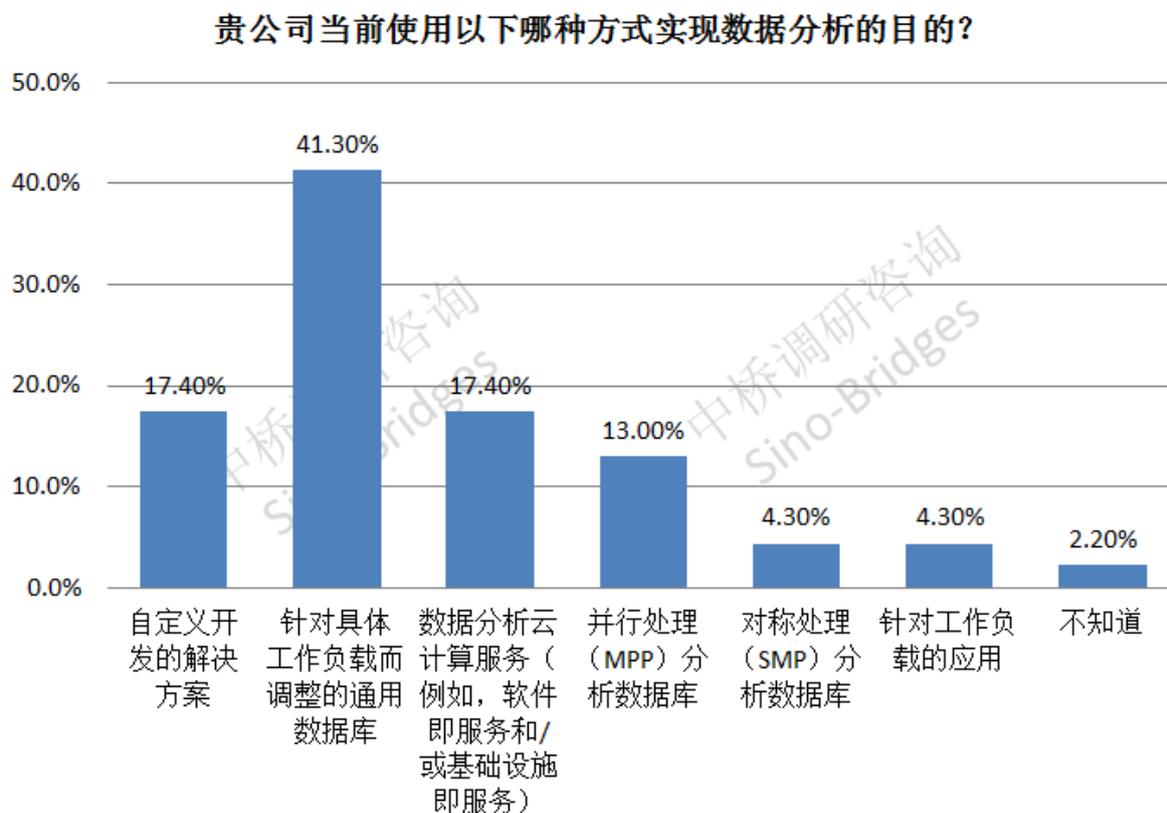
数据来源：中桥调研咨询公司

大数据分析方法的选择

在了解了企业大数据的来源和种类之后，如何采取有效方式对这些数据进行分析，从而最大程度攫取数据价值，转化为

最明智的商业决策以利于企业业务运营，是金融企业对大数据进行分析的目的所在。从目前大数据分析的分析方法来看（图4），有41.3%的企业选择针对具体工作负载来调整通用数据库；17.4%的受访企业选择数据分析云计算服务（如软件即服务和/或基础设施即服务）和企业选择自定义开发的解决方案；13%的用户使用了并行处理（MPP）分析数据库，仅4.3%使用了对称处理（SMP）分析数据库。这一结果表明，大多数的金融企业用户仍处于数据分析的第一阶段。而且，目前他们大多采用通用数据库、云计算或自定义开发的解决方案和数据库工具作为大数据分析方法，而没有选择去购买数据分析的软件。

图4. 大数据分析方法



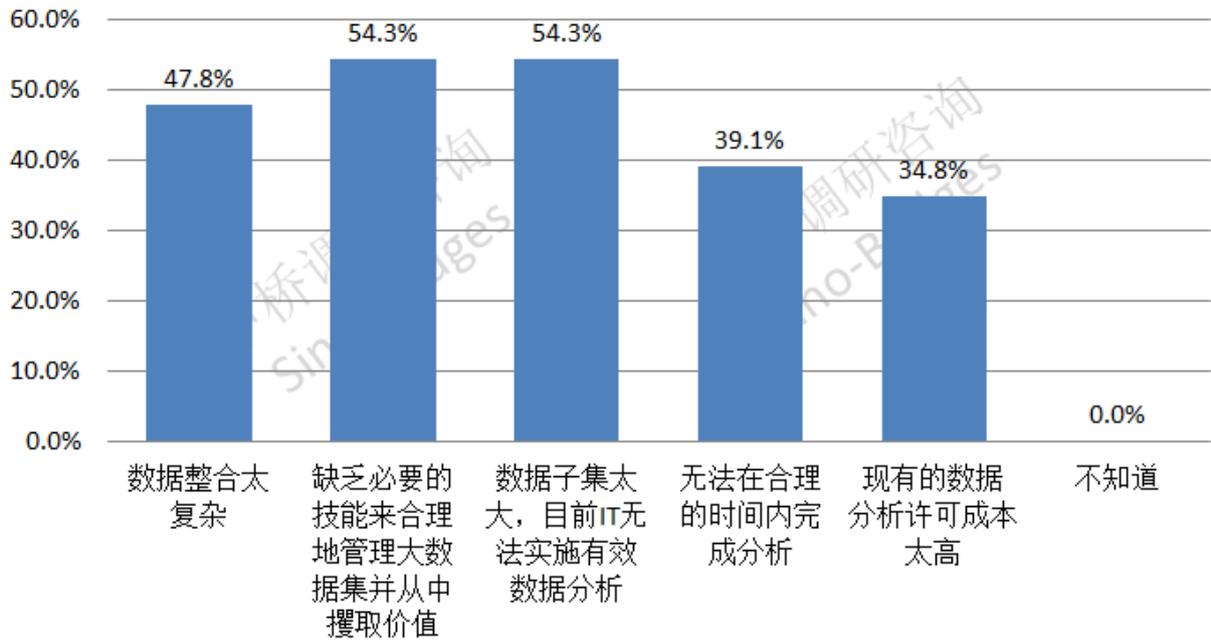
数据来源：中桥调研咨询公司

大数据分析难题

数据类型的多样化和数据量的急剧增长使得传统的关系型数据库难以应对。那么在数据分析过程中，金融企业会遇到哪些难题？从下图5的调研结果得知，缺乏必要的技能来合理地管理大数据集并从中攫取价值、数据集太大，目前IT无法实施有效数据分析、数据整合太复杂、无法在合理的时间内完成分析、现有的数据分析许可成本太高等都是企业在数据分析中遇到的难题。在大数据时代，数据分析的手段是关键，但是传统的数据分析能力，无法处理这么大量的数据。平常分析上千万的数据量的时候，都会花费几十个小时的时间才能得到结果。当数据量达到十亿的时候，软件就根本运行不下去，有的时候会花上好几十天，这个速度是无法接受的。这成为阻碍金融企业从大数据挖掘价值的最大障碍。

图5. 数据分析难题

贵公司经历了以下哪些数据分析难题？（可多选，最多选三项）



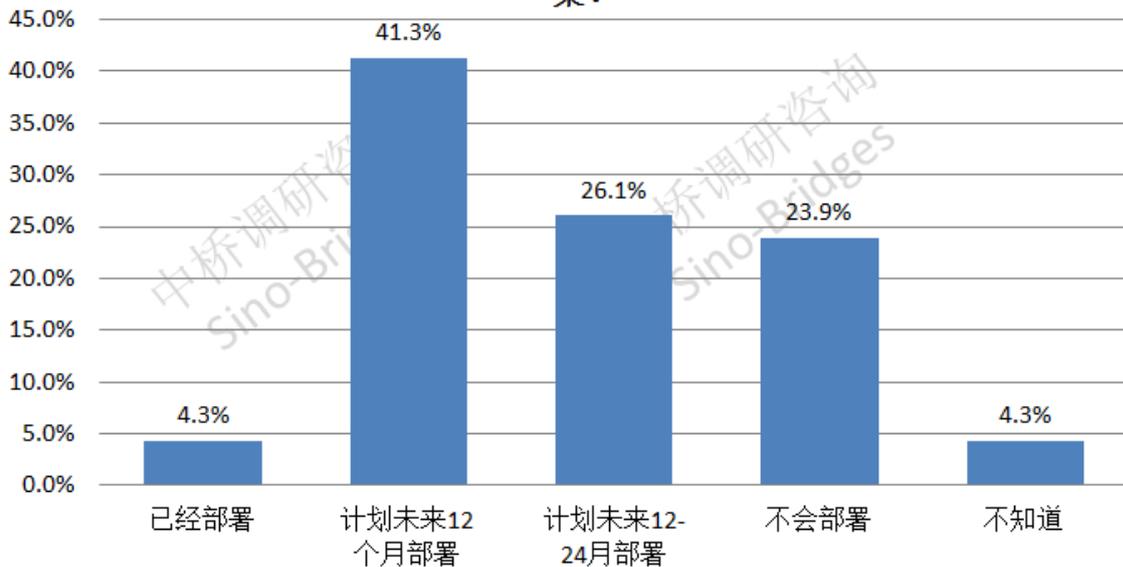
数据来源：中桥调研咨询公司

大数据分析发展趋势

数据分析是发掘数据背后价值的关键一环。虽然身处于大数据分析的第一阶段，但对于金融企业来说，数据分析已经是他们关注的重点。在数据分析解决方案部署上（图6），67.4%的企业表示他们在未来24个月内将会部署新的数据分析平台，来提升商业智能效率、资金和物流周转效率以及优质资源持有率。其中更有41.3%的受访者表示他们会在12个月内着手新数据分析平台的部署。这表示，随着数据的积累以及移动互联网的发展，现有的数据分析解决方案已经不能满足金融企业用户时效性的需求。加大对于数据价值的挖掘，提升数据分析的价值以及分析效率是他们今后的发展重点。

图6. 大数据分析发展趋势

贵企业是否计划部署新的数据分析平台以支持其快速增长的数据集？



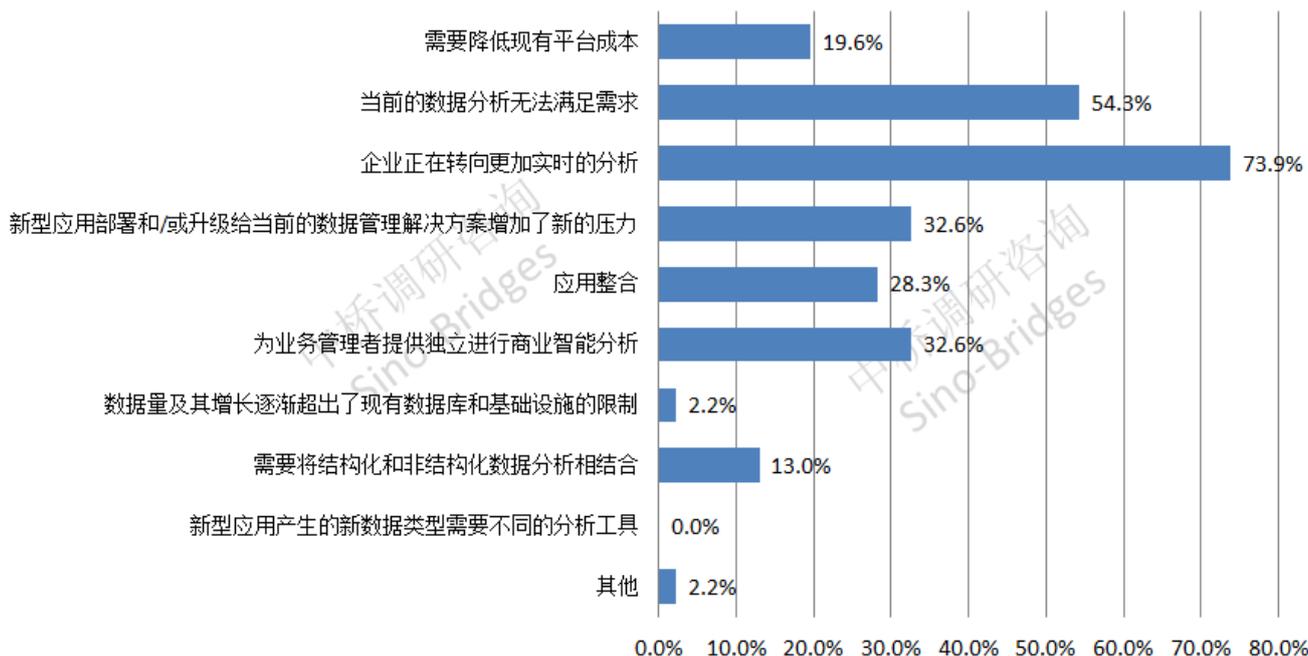
数据来源：中桥调研咨询公司

那么驱动金融企业选择新的大数据分析解决方案的因素有哪些？由图7可知，前5大驱动因素包括：1、企业正在转向更

加实时的分析（73.9%）；2、当前数据分析无法满足需求（54.3%）；3、新型应用部署和/或升级给当前的数据管理解决方案增加了压力（32.6%）；4、为业务管理者提供独立进行商业智能分析（32.6%）；5、应用整合（28.3%）。这说明，大数据意味着如何获取、存储、保护安全数据；下一步，如何保护你的数据，如何清洁、发现相关的数据，如何将其他的数据与其连接起来；最后，如何在数据中获得洞察力。根据中桥调研的结果，整个中国市场在大数据分析阶段上正快速从第一阶段向二、三阶段演进，对于实时或近实时的数据分析需求也会随之得到增长。而金融行业在这方面的需求会更加明显，例如高频交易就是通过速度换取利益。金融企业对于近实时或实时数据分析需求的快速增长，导致他们对于分析效率愈加的关注，从而推动企业去评估新的数据分析解决方案。

图7. 大数据分析解决方案的驱动因素

是什么原因促使贵公司评估新的数据分析解决方案？（可多选，最多选三项）

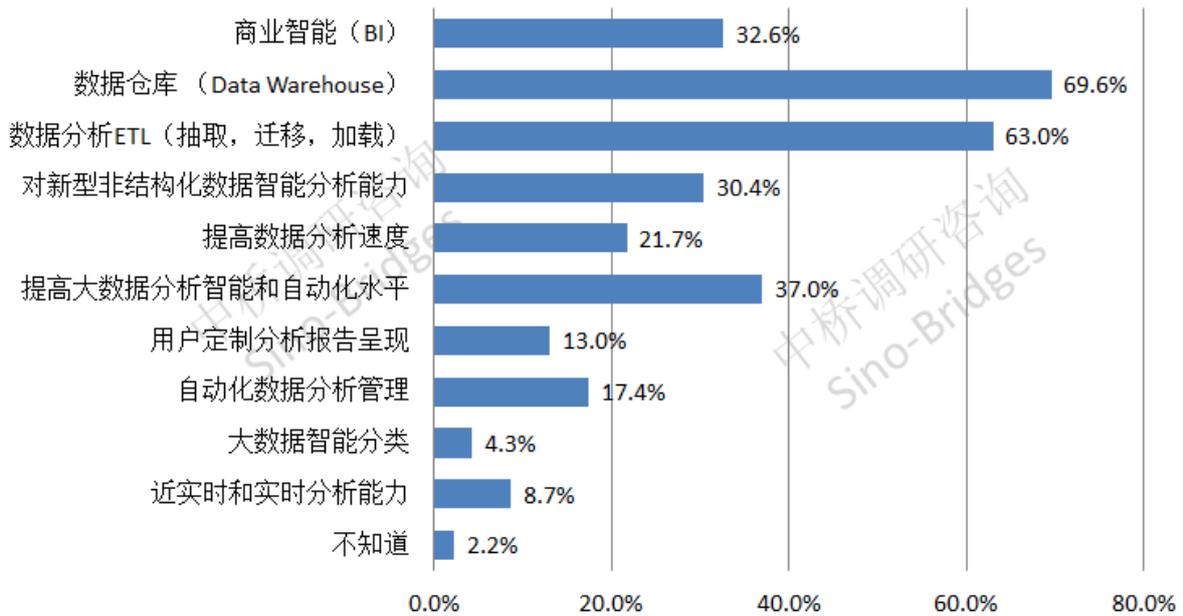


数据来源：中桥调研咨询公司

各种规模和类型的金融企业，大数据分析中都遭遇缺少必要技能困境。因此它们将优先投资于能敏捷分析和强有力的处理技术。对内部和外部数据进行预测分析将使得金融服务企业可以更好、更主动地管理和解决各种可能遇到的信贷和经营风险、欺诈和声誉风险、客户忠诚度和盈利能力等问题。据中桥调研结果可知（图8），金融企业用户在未来12个月会在数据仓库、数据分析ETL（抽取、迁移、加载）和商业智能（BI）上进行投入。传统的以数据库为中心的数据组织模式，由于业务处理能力的局限性，无法在满足决策分析系统对数据的要求，数据仓库技术是在数据库的基础上，通过联机分析处理（OLAP）和数据挖掘技术的综合运用，进行实时查询、访问以及知识的发现，为决策者提供完整、及时、准确的决策信息。另外，对于将在未来加大对ETL和BI的投入，在这里也是十分吻合前文关于金融企业用户所处大数据分析初级阶段的描述。总结来说，数据的商业智能和商业效率是大数据时代企业的焦点，也直接影响到企业的业务发展和数据的价值挖掘。

图8. 大数据投入

未来12个月，贵公司会在下面哪三种“大数据”相关领域做IT投入？



数据来源：中桥调研咨询公司

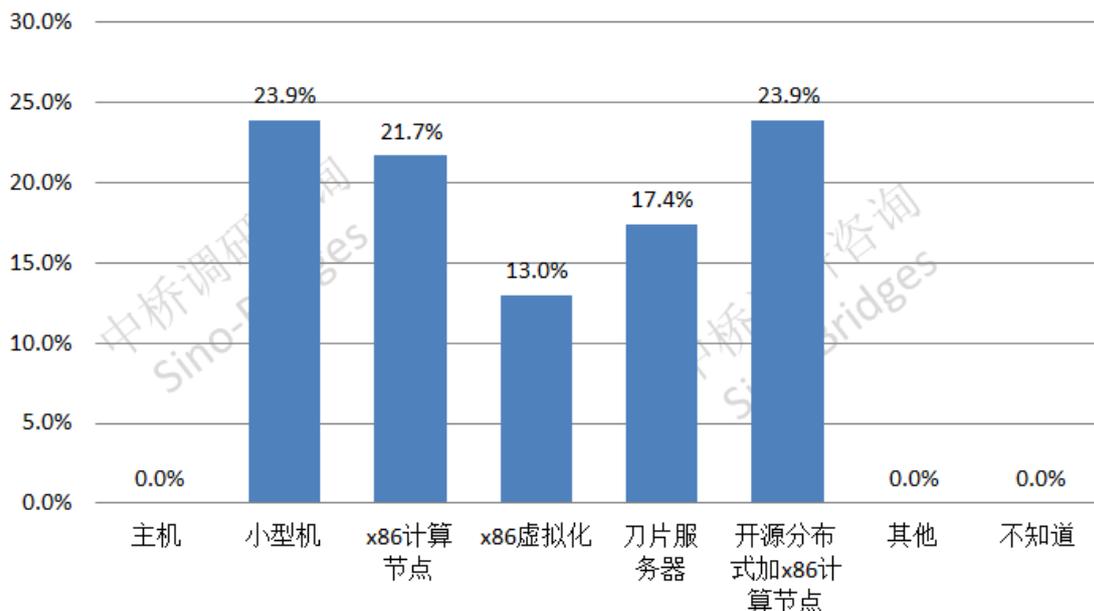
大数据分析中的需求

大数据分析对计算能力的需求

首先从大数据分析的计算方式来看（图 9），选择小型机和开源分布式加 x86 计算节点的金融企业用户比例相同，均为 23.9%；另外，还有 21.7% 的用户表示他们会选择 x86 计算节点。从这组数据可以看出，以往传统来看，小型机可以算作称霸金融行业。2012 年，金融、电信等主要行业“去小型机化”的趋势趋于明显，并且 x86 处理器性能和可靠性的不断提升，推动了 x86 服务器在这些行业的增长。

图9. 大数据分析对计算能力的需求

部署大数据分析方案，会考虑下面哪种计算方式？



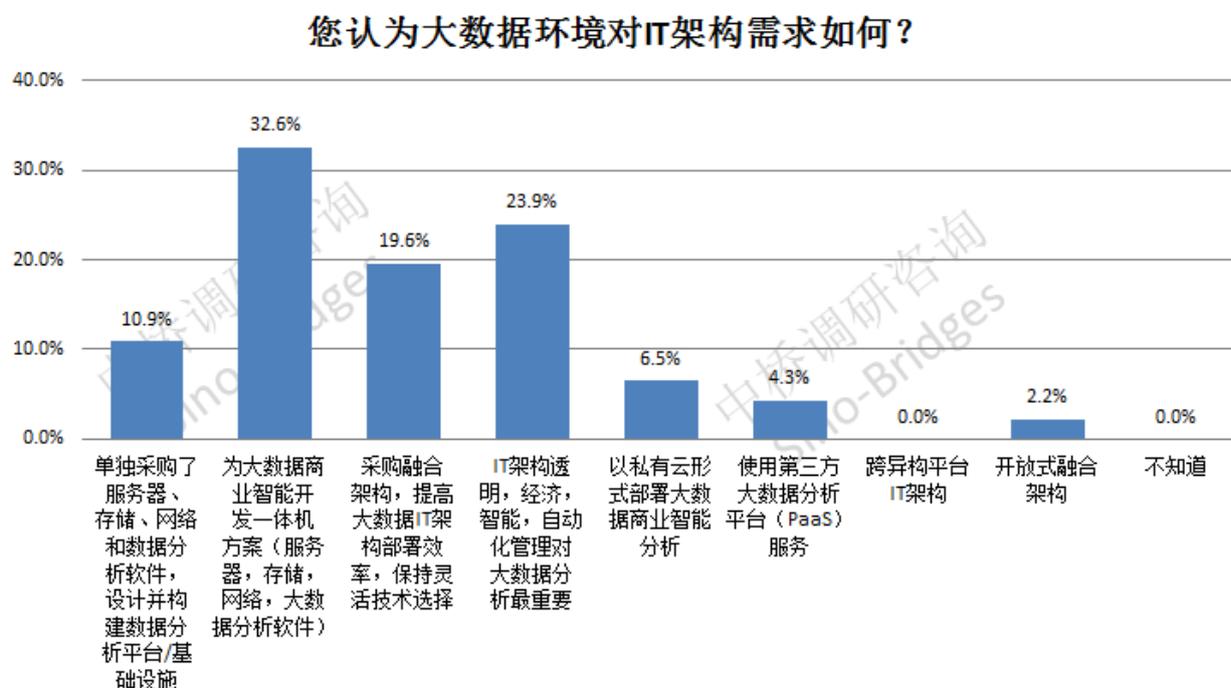
数据来源：中桥调研咨询公司

大数据分析对 IT 架构的需求

在大数据时代，随着数据存储量的爆炸性增长以及分层网络架构的出现，IT 复杂性达到了前所未有的高度，而大数据分

析使得传统 IT 架构更是不堪重负。那么从企业角度来看，他们的大数据环境需要怎样的 IT 架构呢？中桥调查结果表明（图 10），金融企业第一选择的是“大数据一体机”架构（32.6%），其次选择“透明、经济、智能、自动化”的 IT 架构（23.9%）；第三是选择“融合架构形式，提高大数据 IT 架构部署效率以及保持灵活技术选择”（19.6%）。这表明，金融企业用户对于大数据环境下的 IT 架构，倾向于灵活、智能、经济、易管理的 IT 架构。

图10. 大数据对IT架构的需求



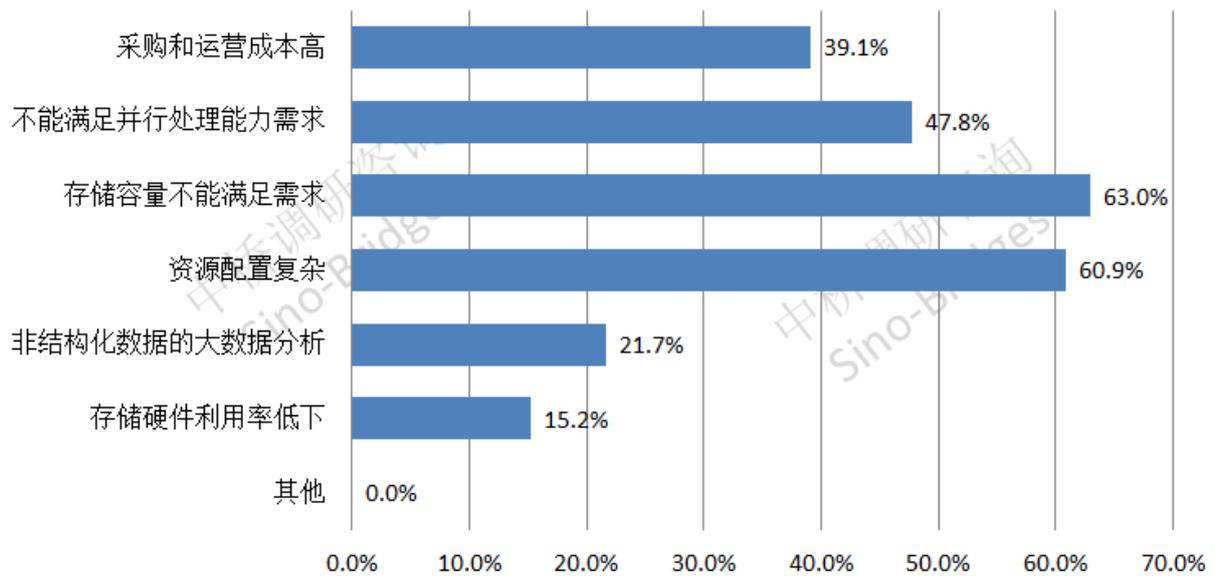
数据来源：中桥调研咨询公司

大数据分析对存储的需求

从大数据分析的第一个环节——数据收集和存储来看，大数据时代应用数量、应用数据量和使用用户数量的增长，对存储 IOPS 以及 OLTP 和 OLAP 的要求越来越高。对于大数据分析中，存储面临的哪些难题？图 11 调研结果显示，63.0%的企业认为存储容量在大数据时代下是一个大问题，此外还有诸如存储资源配置复杂（60.9%）、不能满足并行处理能力需求（47.8%）、采购和运营成本高（39.1%）、非结构化数据的大数据分析（21.7%）、存储硬件利用率低下（15.2%）等存储问题是需要金融企业着手解决的。这一调查结果也体现了，能够整合更多各种格式的数据源，并能高效快速地输出分析结果，对于金融企业用户来说愈加重要。而这些在存储设备上，就表现为对于容量可扩展性、并行处理能力、管理上的简易、资源利用率和性价比的需求。

图11. 大数据分析下存储难题

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）

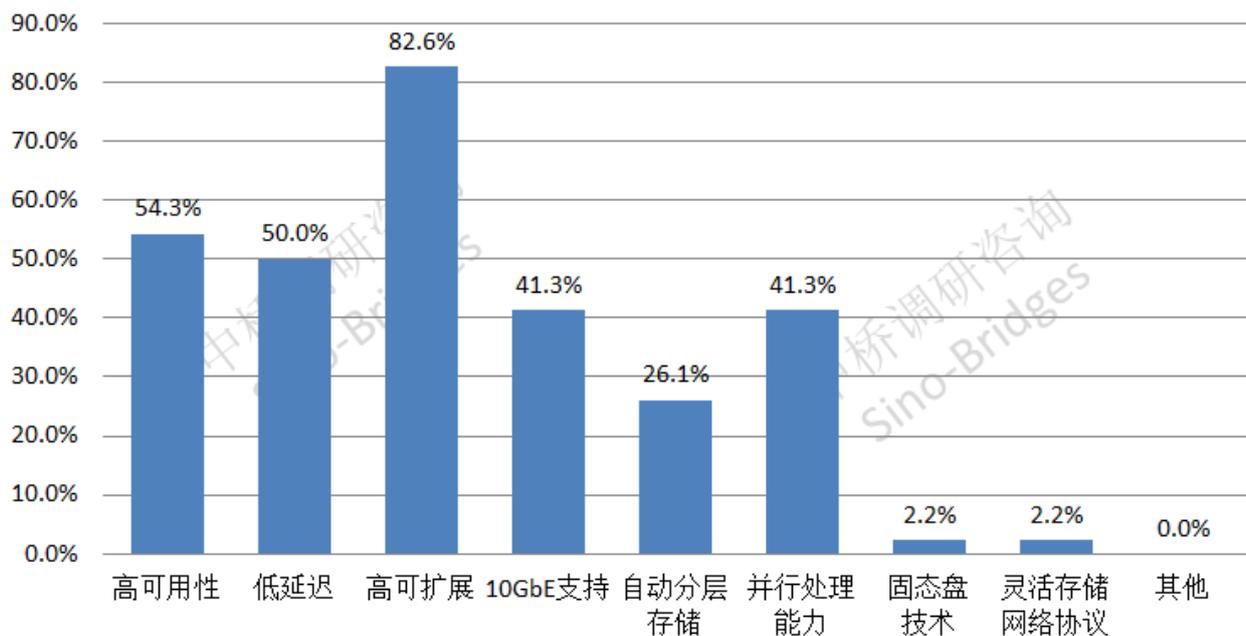


数据来源：中桥调研咨询公司

那么所选择的新型存储技术应该以什么样的指标来权衡，才能确保整个大数据分析流程平稳、高效运行？中桥对企业的调查结果显示（图 12），存储高可扩展性、高可用性和并行处理能力是企业评估大数据存储最重要的三个因素。高可扩展性可以确保企业的 IT 能够随着数据量的增长和性能需求进行扩展，以满足海量数据的存储和处理需求；高可用性则能够保证大数据分析过程的平稳、无间断运行，确保了业务连续性；高并行处理能力则能够确保在大数据处理过程中同时进行更多数据的处理，高效地完成数据分析，从而将分析结果转化为业务决策，加快产品或技术的面市周期。此外，低延迟、自动分层存储以及 10GbE 支持等也是用户评估大数据存储的重要考核因素。

图12. 评估存储的重要指标

评估数据分析存储技术，最重要3个指标是什么？



数据来源：中桥调研咨询公司

大数据分析工具认知度

最后，来看下大数据分析工具认知度。目前市面上比较主流的大数据分析工具：

1. 开源大数据分析工具

Hadoop HDFS、Hadoop MapReduce、HBase、Hive 渐次诞生，开源生态圈活跃，并免费。早期 Hadoop 生态圈逐步形成。

2. 商用大数据分析工具

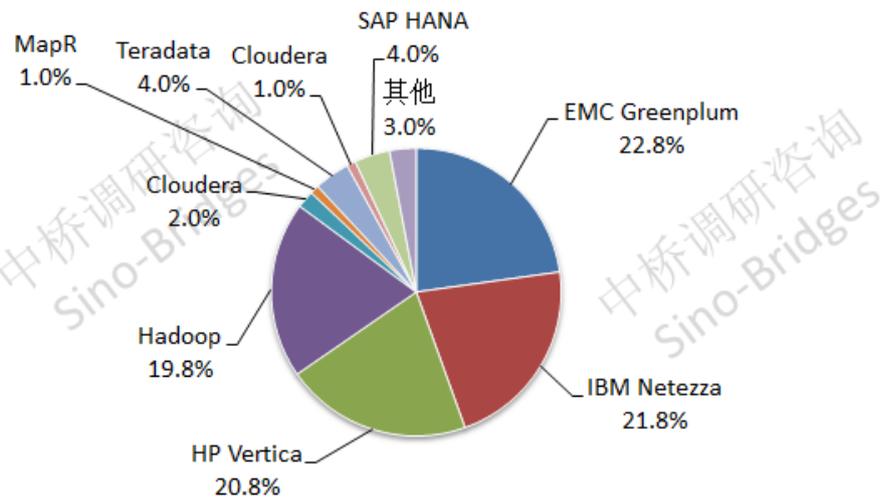
一体机数据库/数据仓库（费用很高）：IBM PureData(Netezza), Oracle Exadata, SAP Hana 等等。

数据仓库（费用较高）：Teradata AsterData, EMC GreenPlum, HP Vertica 等等。

在中国金融行业，用户使用最多的数据分析工具还是商用大数据工具(图 13)，其中以 EMC GreePlum（22.8%）、IBM Netezza（21.8%）和 HP Vertica（20.8%）三大工具占据市场比例最高。

图13. 大数据分析工具认知度

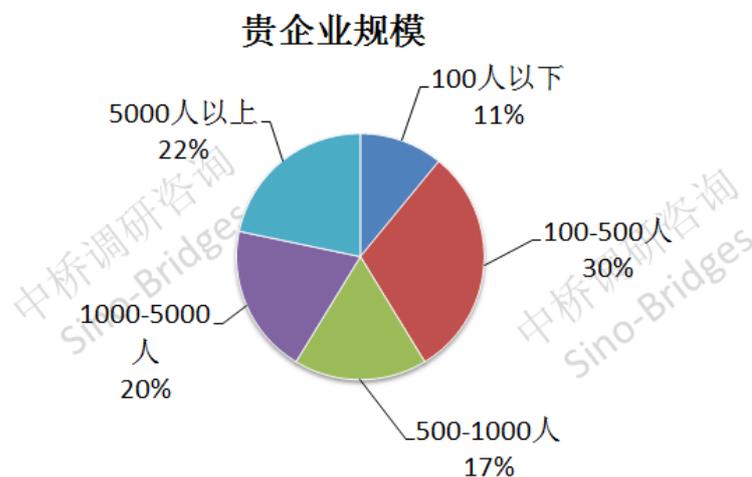
您对下面哪种商业智能、大数据分析技术比较熟悉？（可多选，最多选三项）



数据来源：中桥调研咨询公司

最后，附上本次参与调查的金融企业规模分布

图14. 金融企业规模分布



数据来源：中桥调研咨询公司

结论

- 尽管在谈到大数据的时候，金融行业是应用大数据的重点行业，然而就目前分析频率上的表现来看，金融行业在大数据分析上还处于第一阶段。但是，随着移动互联网的发展和数据量的不断增长，数据分析是金融行业在今后大力投入的一个重要方面。
- 在大数据时代，数据分析的手段是关键，大数据分析对于计算方式选择有两种趋势：小型机和分布式加 X86 计算节点形式。X86 服务器在金融企业中接受度也越来越高。
- 此外，数据存储量的爆炸性增长以及分层网络架构的出现也使得传统 IT 架构不堪重负。在大数据时代，随着金融企业用户对于大数据环境下的 IT 架构，倾向于灵活、智能、经济、易管理的 IT 架构。
- 金融企业用户对于大数据分析的实时性要求越来越高，再加上应用数量、数据量和使用者数量的增长，对存储 IOPS 以及 OLTP 和 OLAP 的要求越来越高，从而对于存储容量可扩展性、并行处理能力、管理上的简易、资源利用率和性价比提出更高的要求。存储高可扩展性、高可用性和并行处理能力是企业评估大数据存储最重要的三个因素。
- 金融企业所应用的大数据分析工具中，还是以商用品牌为主，EMC、IBM 和 HP 这三家占据的份额最高。随着 Hadoop 开源工具发展和推广，也逐渐成为金融市场占有率最高的开源工具。

