

# 大数据时代的商业银行数据治理研究

中国银行股份有限公司管理信息部 王玮



王玮：现任中国银行总行管理信息部综合分析团队主管，清华大学、香港中文大学博士后，曾任中国银行总行管理信息中心信息标准团队主管。

随着金融创新的快速发展和信息科技的日新月异，商业银行积累的数据量呈现几何倍数的增长，数据来源也已从传统的结构化数据逐渐扩展到以网络日志、社交媒体为代表的半结构化和非结构化数据，商业银行逐步进入大数据时代。在大数据时代如何有效治理结构化、半结构化和非结构化的海量数据，

挖掘分析其中潜在的业务价值，已经成为商业银行面临的新课题。

## 一、大数据的特征

大数据是指无法在一定时间内用常规软件工具对其进行获取、管理和处理的数据集合。大数据是一种海量的、高增长的和多样化的数据资产。大数据通常具有“4V”（Volume、Variety、Velocity和Value）特征。

（1）巨量性（Volume）：数据量巨大是大数据最显著的特征，商业银行积累的数据量快速增长，从TB级别跃升到PB级别甚至ZB级别，传统的集中存储/计算已经无法适应大数据的处理。

（2）多样性（Variety）：商业银行业务发展涉及的数据类型已从以二维表结构方式表达的结构化数据，扩展到包括日志、微博、视频、图片等半结构化和非结构化数据。

（3）快速性（Velocity）：商

业银行的数据创建、存储、处理和分析的速度在持续加快，某些数据必须实时地进行分析，才能及时、有效地对业务管理产生价值。

（4）低密性（Value）：价值密度的高低与数据总量的大小成反比。商业银行如何借助强大的信息技术更迅速地完成海量数据价值的提升，是大数据时代需要解决的问题。

## 二、大数据时代的机遇和挑战

大数据时代，商业银行面临良好的发展机遇，数据量的快速增加为商业银行提供了精准营销的基础（即利用数据获知客户群体和个体网络行为模式），充分利用这些信息，商业银行就可以为客户制定个性化、智能化的服务模式，也可以通过对大数据的挖掘分析，设计开发出更贴近用户需求的新产品。大数据正逐渐改变商业银行经营决策和业务发展的模式，促进商业银行甚至整个金融行业的转型。

(1) 大数据为商业银行提供了全新的沟通渠道和营销手段。一方面, 社交媒体的兴起为商业银行提供了全新的与客户接触的渠道, 利用社交媒体, 商业银行可以进一步树立社会化形象, 拉近与客户的距离; 另一方面, 通过全面分析商业银行内部数据和外部的社会化数据, 可以获得更为完整的客户全貌, 避免因客户信息不全面而导致的错误结论, 从而进行更为精准的营销和管理。

(2) 商业银行可以基于大数据的挖掘分析来提升自身的竞争能力。随着大数据时代的来临, 商业银行可以运用科学分析手段对海量数据进行分析 and 挖掘, 更好地了解客户的消费习惯和行为特征, 优化运营流程, 提高风险模型的精确度, 研究和预测市场营销和公关活动的效果, 从每一个经营环节中挖掘数据的价值, 从而使得商业银行进入全新的科学分析时代。

(3) 商业银行可以利用大数据获得全面准确的决策支持。商业银行拥有大量的客户数据和交易数据, 基于海量数据已能够统计得出部分具有指导意义的结论, 结合政府、评级机构、第三方支付等合作机构的外部数据源, 可以有效对经济形势和市场趋势进行分析预测, 指导商业银行进行业务创新, 或为经营管理提供更加全面及时的决策支持信息。

大数据时代, 商业银行既面临良好发展机遇, 也遭遇以下严峻的挑战。

(1) 大数据催生新型金融形态。大量的数据来源和强大的数据分析工具可产生出很多新的金融形态。例如, 利用社交网络将客户的信息碎片关联起来、通过对海量客户的交易数据分析来预测客户风险, 从而作为其信贷服务的评估依据, 这将直接影响银行的信贷市场, 对现有的金融行业格局产生深远影响。

(2) 大数据改变金融服务形态。随着数据化和网络化的深入发展, 金融服务形态将逐渐向虚拟化方向发展: 产品的虚拟化、电子货币等数字化金融产品的发展空间巨大; 服务的虚拟化, 即通过移动互联网等科技手段, 银行可以通过虚拟渠道向客户提供金融服务; 流程的虚拟化, 即银行业务流程中各类单据、凭证等可以用数字文件的形式出现和处理, 从而提高服务的便利性和效率。

(3) 大数据考验银行的数据处理能力。由于大数据的数据量庞大、数据格式复杂多样、对数据存储的要求高、数据价值密度低, 传统的数据仓库技术已经不适用于大数据的处理, 商业银行需要采用诸如云计算等新技术去处理大数据, 才能发挥其真正价值。

大数据不仅要求商业银行具备海量数据的存储能力, 更要求商业

银行能够对这些数据进行专业化管理。如果将大数据比作一种产业, 那么这种产业实现盈利的关键, 在于提高数据的治理能力, 以实现数据资产的有效增值, 完成数据向信息的转换。

### 三、大数据时代的数据治理面临的问题

数据治理是一套完整的制度、指引和规范, 用于统筹人员、流程以及技术等要素, 使得商业银行能将数据作为企业的重要资产而有效利用。完善的数据治理可以确保商业银行数据的可用性、可获取性、完整性以及一致性。

数据治理的核心内容是统一数据标准, 其关键任务是提升数据质量, 其服务目标是提供决策支持信息。目前商业银行在数据标准和数据质量方面都或多或少地存在如下问题。

(1) 定义缺失。即部分业务源系统和外部数据源缺少关键业务元素定义。由于商业银行系统开发建设的出发点大多以满足客户业务交易为主要目标, 对于统计分析涉及的数据要素项的定义不够关注, 因此造成部分业务源系统和外部数据源的数据定义不完备, 使得不同业务部门有不同理解, 对同一字段的理解产生歧义。

(2) 标准各异。由于商业银行早期建设的业务源系统较为分散,

缺乏从全行的角度构建统一的基础信息平台,因此形成了一个信息孤岛,造成相关系统的指标口径、会计科目的不一致。标准各异使得系统中的信息无法有效共享,造成了商业银行信息资源利用的低效率。大数据时代,由于半结构化和非结构化数据的引入,商业银行企业级数据标准的统一面临更大挑战。

(3) 数据异常。部分业务源系统和外部数据源的个别字段出现了异常信息,包括取值错误、格式错误、多余字符、乱码等。有些虽然在系统表中已经设计了相关字段,但是在实际使用中存在很多记录没有收集该字段的信息,或者出现信息不准确,信息重复登记等情况。

(4) 代码问题。结构化、半结构化和非结构化数据之间相同用途代码的编码不一致,如仅采用文字存储,而非将信息代码化,存在商业银行信息存储得不少,但这些信息不能被充分地分析使用的情况。

(5) 更新滞后。商业银行为满足各系统和外部数据的内部访问,提高访问效率,减少数据传输,相同信息经常会在不同系统进行冗余存放,但是由于更新滞后,导致冗余数据不一致,产生数据质量问题。

#### 四、大数据时代的数据治理措施

##### 1. 建立健全数据标准体系

根据业务管理中涉及的产品、渠道、合约、分类、事件等数据要

素,商业银行已逐步建立起机构标准类、产品标准类、员工标准类、客户标准类、渠道标准类、营销标准类、账户标准类、合约标准类、交易标准类、公共标准类等十大类数据标准体系框架。未来将结合大数据的特点和应用,不断丰富和完善数据标准体系,指导商业银行科学规范地使用数据。

##### 2. 制定科学的数据标准流程

数据标准的核心是要建立统一的企业级规范,促进全行范围的数据有机共享,有效提高数据的管理和使用水平。大数据时代,商业银行要根据结构化、半结构化以及非结构化数据的不同特点,从加强内外部数据共享的角度为大数据制定统一的业务解释和标准,加强协调商业银行相关部门对数据标准的不同需求,建立适用的统一模型和数据共享为基础的数据标准流程。

##### 3. 制定前瞻性的数据质量规划

制定前瞻性的数据质量规划是做好数据质量管理的前提和基础。大数据的管理和应用要纳入商业银行全行统一的数据质量规划范畴。对于新建项目的业务范围和系统规划必须完全执行已正式发布的数据标准;对于已经建设完成的系统需在数据质量规划的要求下酌情进行适应性改造。

##### 4. 形成有效的数据质量评估机制

商业银行要加强数据质量的完

整性、一致性和准确性评估。

(1) 完整性:必需的数据必须存在,数据关联关系必须存在,不存在违反数据标准和数据质量需求的数据。

(2) 一致性:数据值提供的信息不应存在冲突,不存在重复记录。

(3) 准确性:数据必须反映商业银行真实信息,数据的提供时间和数据的生命周期必须符合商业银行的业务需要。

##### 5. 构建闭环的数据质量管理模式

商业银行通过对数据质量的持续监控,要做到发现问题及时解决、及时纠正,从源头强化数据质量的控制。通过定期发布数据质量报告、不定期对基础数据进行质量检测 and 抽查、把外部监管报送数据和内部经营管理数据的准确性和稳定性指标纳入绩效考核等措施手段,不断提升数据质量问题的解决效率并跟踪实际效果,构建闭环的数据质量管理模式。

大数据在商业银行经营管理中正发挥着越来越重要的作用,有效利用大数据的价值将成为商业银行提升自身核心竞争力的重要手段,只有不断完善商业银行的数据治理,建立统一的数据标准,提升数据质量,改善服务水平,才能使商业银行在激烈的市场竞争中立于不败之地。FCC