



毕马威数据湖 数据管控平台

毕马威中国

毕马威企业咨询（中国）有限公司
2020年3月

目录

CONTENTS

1. 毕马威智慧之光 - 创新服务	2
2. 数据湖产品介绍	3
3. 公共数据服务-毕马威智听	4
4. 数据湖组件介绍	5
5. 全方位合作模式	10
6. 附录 – 基于数据湖的应用案例	11

毕马威智慧之光-创新服务



智慧之光简介

毕马威智慧之光（Lighthouse）是毕马威全球的大数据和人工智能创新中心，包含顶尖的技术人才和创新解决方案。通过深入一线业务场景，识别各行各业的数字化转型需求，为客户设计和研发在营销、风险、财务和运营等领域的数字化解决方案。

国际化引进，本地化落地和原创研发是智慧之光的特色。

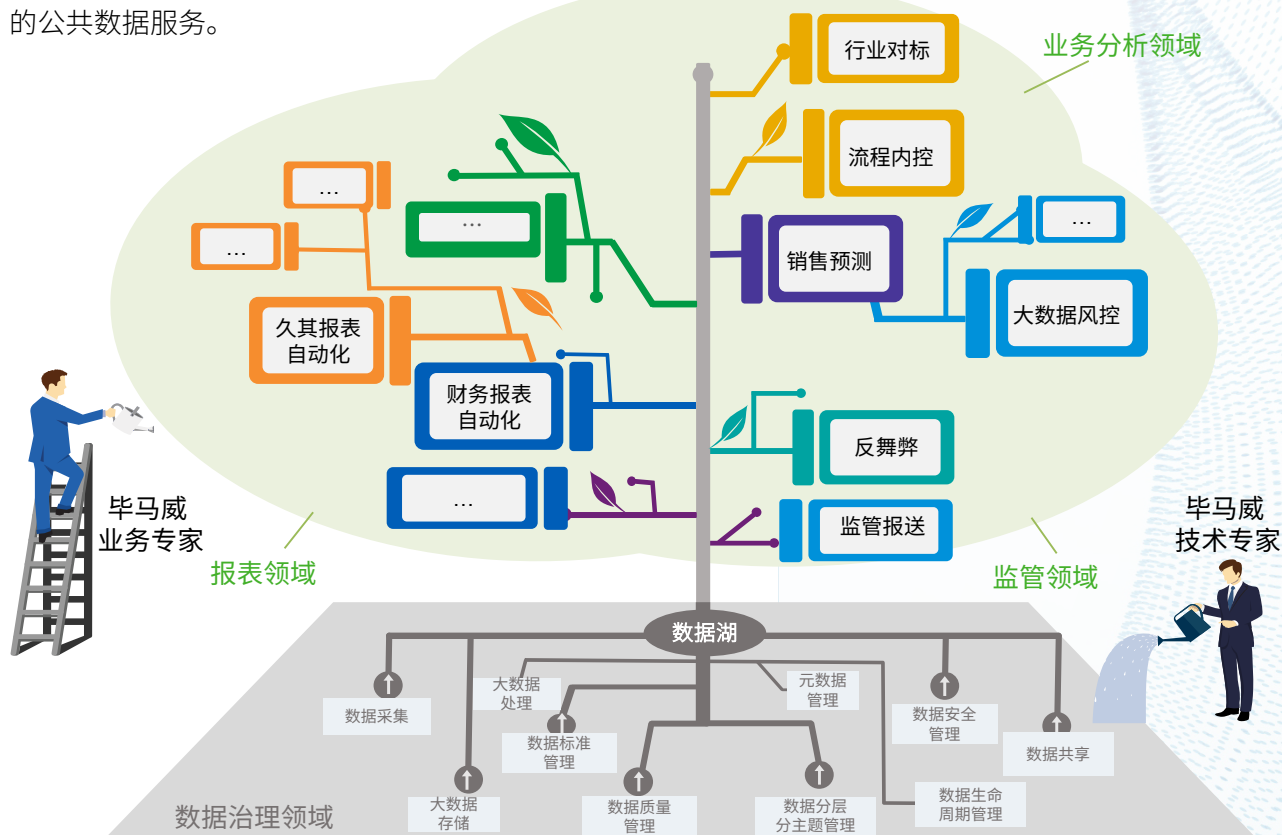


毕马威AI工厂赋能业务创新

越来越多的企业在实际业务应用中发现了数据地重要性，如何从公司内外部采集全面的数据，并有效使用这些数据是各个企业面临的问题。数据湖平台是毕马威根据多年的数据治理经验总结出来的一套标准的数据管理平台，可以帮助采集内部各个ERP系统数据及外部的公开数据，制定统一的数据标准、数据治理和数据处理规范，并利用大数据技术对搜集的数据进行加工挖掘，从而支撑企业内部的经营决策、报表分析、监管报送等业务，帮助企业提升数据使用价值。

同时，为解决企业公共数据搜集不全面，数据来源不合法等问题，我们也推出了公共数据服务毕马威智听解决方案，为企业提供合法、可靠、稳定的公共数据服务。

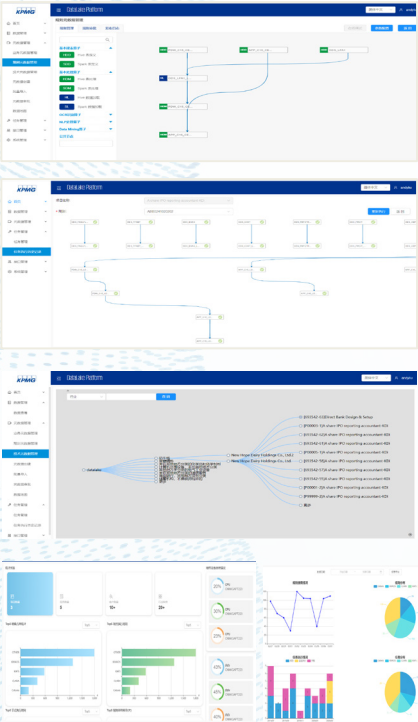
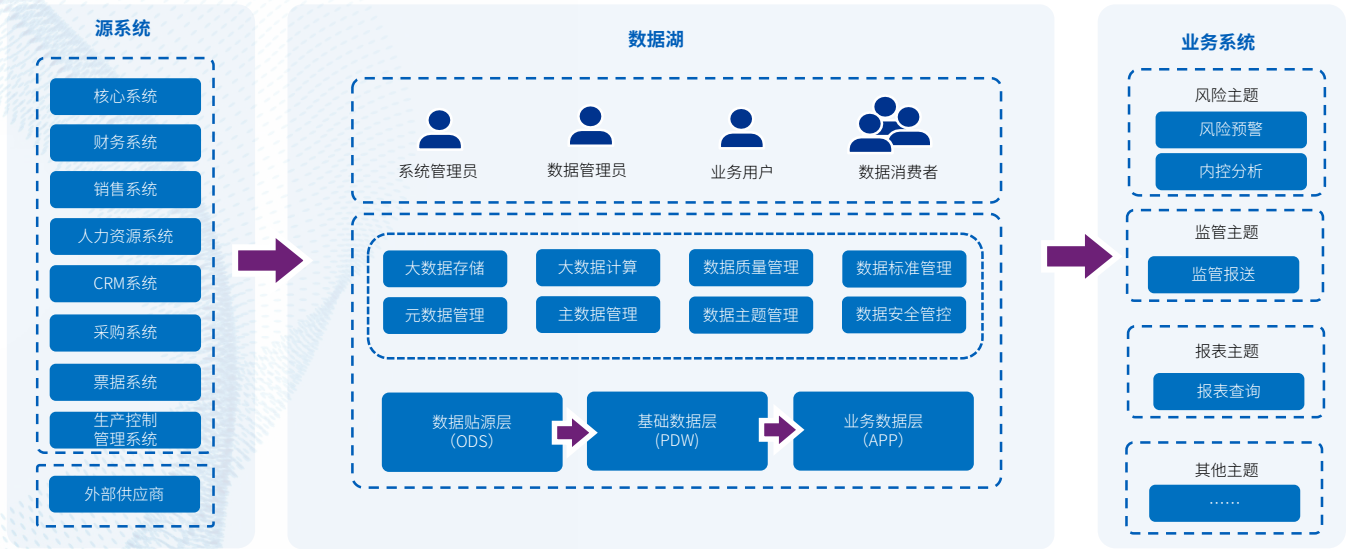
- ① 数据湖产品介绍
- ② 公共数据服务-毕马威智听
- ③ 数据湖组件介绍
- ④ 全方位合作模式
- ⑤ 基于数据湖的应用案例



数据湖产品介绍

毕马威数据湖是基于大数据技术构建的一体化数据平台，可帮助企业集中存储及管理内部和外部的各种数据，消除信息孤岛，辅助公司领导进行各种业务决策。

该平台集成了数据采集工具，同时也支持其他多种方式进行采集，并提供了大数据存储、大数据计算、数据标准管理、数据质量管理、元数据管理、主数据管理、数据生命周期管理、数据安全管控等一系列功能。



功能优势

- 能力全：涵盖了数据管控和数据处理的相关功能。
- 组件化：所有的数据管控组件和数据平台组件都可以独立部署。
- 易操作：通过可视化拖拽就可以实现业务数据处理编排。
- 可扩展：提供了各种开发接口，方便进行二次开发。

技术优势

- 基于Hadoop大数据开源技术，支持Spark, Flink, Hive扩展。
- 基于分布式微服务架构建设。
- 各个服务组件包解耦，可插拔式部署。
- 业务处理规则配置化，可动态扩展。

公共数据服务 – 毕马威智听

为满足越来越多的企业业务对外部数据的需求，毕马威智慧之光开发了一个公共数据平台 毕马威智听，能够安全、持久、稳定、合规地从外部搜集各类公共数据，这些数据包括合法获取的公开财务数据、关系数据、风险数据、舆情数据、宏观经济数据、舆情数据等，也包含了毕马威长期总结的行业经验数据等，这些数据结合企业内部数据等，作为咨询方案地补充，有效支撑企业的业务分析，提升业务决策的能力。



方案特点

- **合法合规：**所有的数据搜集渠道经过法务严格的合规评审，保证合法合规。
- **可靠稳定：**数据来源可靠稳定，保证可持续性的数据服务供应。
- **方便接入：**提供标准的API数据服务，可按照主题选择需要的数据服务范围。
- **数据安全：**数据使用过程安全可靠，传输通道加密。



数据类型

毕马威行业
经验知识库

公开财务
数据

企业关系
数据

企业股东
数据

风险数据

舆情数据

宏观经济
数据

年报财报
数据

上市企业
股票数据

司法数据

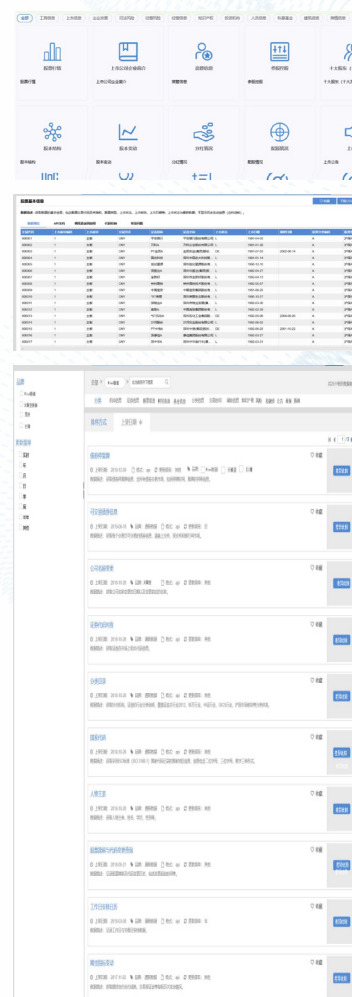
政策法规
解读数据

天气数据



业务收益

可以帮助客户有效地补全外部数据，并基于公司内部的数据进行最大价值的数据分析，支撑业务价值提升。



数据质量管理

数据湖组件介绍

场景说明

- 通过构建数据质量管理平台，制定数据检核规则和检核策略，可以有效地帮助企业发现和暴露各种数据质量问题，生成各种关键的数据质量报告，把控公司内部各个系统真实的数据质量状况，并针对各种问题提出改进建议，制定改进计划，逐步消除质量问题，从而为各个业务提供准确有效的决策支撑依据。

产品特征

可视化规则编辑	灵活的检核任务编排
提供可视化的质量规则检核算子，简单易用，不需要较强的专业技能，业务人员也可以进行质量规则编辑。	通过简单地拖拽就可以实现整个数据检核流程任务地编排，直观、简单、易用。
丰富的数据质量广告牌	可定制化的质量报告
提供丰富的数据质量广告牌，可以从各个维度查看数据质量情况，了解各个系统真实的质量状态。	除内置的各种质量报告模板外，也可以根据需要进行定制化的质量报告开发。
质量举证和质量问题核销	实时的质量监控能力
提供业务部门对已有的质量问题举证，同时支持对非质量问题进行核销。	提供实时的数据质量监控能力，让发现的各种数据质量问题第一时间通知到管理员。

关键价值

- 为企业内部落实统一数据质量标准、数据质量管理体系、执行规范及落地措施，提供了有力地平台支撑。
- 简化数据质量管理实施难度，让业务人员也能够参与到质量规则管理中去。
- 通过各种维度的质量分析和质量广告牌，能够全方位把控系统的数据质量问题。
- 质量问题及时通知，让管理员可以第一时间响应，减少可能出现的业务和决策风险。
- 快速定位质量问题产生的原因及影响范围，有效控制和减少因为数据质量问题带来的影响。

服务特色

- 可提供各种形式的数据质量平台定制化建设服务，构建真正满足企业业务场景的数据质量管理平台。

数据标准管理

数据湖组件介绍

场景说明

- 公司内部存在各业务部门数据标准不统一，部门之间数据标准不一致的情况，导致部门间数据交换，数据共享比较困难。建立统一的数据标准有助于对数据进行统一规范的管理，消除各部门间的数据壁垒，方便数据共享，同时有助于业务流程的规范化。

产品特征

可定制化的标准模型

提供可定制的数据标准模型，可以根据业务需要定义标准的业务属性、技术属性和管理属性。

灵活的标准定义方式

支持在线标准定义，也支持离线批量创建。

严格的标准审核机制

提供了严格的标准审核发布机制，标准版本控制及生效机制。

多维度的标准统计

提供了各个维度的标准统计，方便查看标准的状态、分布及引用情况等。

自定义标准主题分组

可以按照业务需求，灵活地定义标准主题分组。

强大的标准关联能力

提供强大的标准关联能力，能够让数据处理、表创建、数据质量检核等功能便利地引用到需要的标准。

关键价值

- 为企业内部落实统一数据标准管理体系、执行规范及落地措施，提供了有力的平台支撑。
- 为企业内部所有数据相关地处理、检核操作等提供了有效的标准依据和关联方式。
- 为企业内部各种标准的制定情况，贯彻落实情况，引用情况等提供全方位地掌控。

服务特色

- 可提供各种形式地数据标准管理平台定制化建设服务，构建真正满足企业业务场景的数据标准管理平台。

元数据管理

数据湖组件介绍

场景说明

- 企业内部的业务系统中大量数据如何存放、数据之间存在关系、数据的来源、数据的影响范围、数据变化趋势、系统和系统的关系、各种组织关系等越来越多问题，要求企业必须打造一个强大的元数据管理平台来帮助管理这些内容。

产品特征

灵活的元数据采集 提供多种灵活的数据采集方式，支持自动采集，也支持在线或者批量导入的方式添加元数据信息。	可自定义的主题分组 可以根据业务需要灵活定义不同的主题分组。
多种维度的数据地图展示 提供各种维度的数据地图展示能力，可以轻松查看所有业务系统中的数据情况，了解数据全貌。	数据关联性分析 可以对任意的数据进行关联分析，找出数据间存在的各种联系。
数据全链路分析 通过数据全链分析，可以查看数据的整个链路环节。	数据影响性分析 提供数据影响分析能力，可以看到相关数据所有的关联影响。
数据血缘关系追踪 提供表级和字段级的血缘关系追踪，可以追踪数据的来龙去脉。	数据历史轨迹分析 支持查看所有业务系统中数据的历史变化趋势。

关键价值

- 有效地帮助企业了解内部各个业务系统的数据情况，数据的变化趋势，从而更好地发挥数据的价值。
- 让企业了解所有数据的来源、用途、关联关系、影响性分析，从而有效地支撑各个业务部门使用数据。

服务特色

- 可提供各种形式的元数据管理平台定制化建设服务，构建真正满足业务需要的元数据管理平台。

智能大数据处理

数据湖组件介绍

场景说明

- 当企业需要搜集的数据资产达到一定的量级，不可避免的会遇到数据如何存储，如何保证数据处理及时性、可靠性、稳定性及高容错，以及各种类型的数据如何处理及AI技术的应用问题。毕马威数据湖基于开源大数据hadoop平台打造，能够有效地帮助企业解决数据存储、处理、计算、治理、安全、可靠性、稳定性、高容错等一系列问题，依托于毕马威长期的咨询实施经验，该平台引入了大量的AI处理技术，可以很好地帮助企业处理内外部数据，能够有效地支撑业务需求。

产品特征

大数据存储和计算

基于hadoop平台打造，同时支持HDFS, Hive, Spark等多种大数据组件，可横向无限扩展。可支持任意类型的数据。

可视化的数据处理任务编排

支持通过拖拽式的可视化任务编排功能，客户可以根据业务需求便捷地定义业务处理全流程。

AI能力支持

支持各种AI相关的算子，例如：发票识别算子、身份证识别算子、护照识别算子、年报解读算子、报告审阅算子等。

可视化任务监控

支持整个数据处理全流程任务的可视化监控，日志分析，在线调试，数据查看等。

数据共享管理

支持系统内及系统外的数据共享标准流程管理，支持数据通过OpenAPI、文件等方式共享，同时也支持数据文件推送。

可扩展的规则处理算子

支持各种类型的数据处理算子，例如HDFS文件处理算子、Hive数据处理算子、Spark数据处理算子，同时也支持各种自定义算子。

数据治理规范化

该平台具备规范化的数据治理能力，支持不同数据分层的数据质量管理，支持元数据管理，支持数据标准管理及主数据定义，数据敏感性分级，数据生命周期管理等。

数据分层分主题管理

可以按照业务地要求定义数据分层及不同分层下的数据主题定义。

关键价值

- 帮助企业快速打造一个大数据平台，处理内部海量数据及相关的数据治理、数据管控、数据共享、数据安全等一系列问题。
- 让企业了解所有数据地来源、用途、关联关系、影响性分析，从而有效地支撑各个业务部门使用数据。

服务特色

- 可提供各种形式的大数据平台定制化建设服务，构建真正满足业务需要的大数据平台。

数据洞察

数据湖组件介绍

场景说明

- 企业内部在构建完数据平台后，不可避免地面临的问题是，如何向各个业务部门说明系统中有哪些数据，如何使用这些数据，同时需要配合业务部门，将数据以需要进行呈现及分析。毕马威数据洞察平台可以有效地帮助企业解决数据呈现、数据检索以及多种维度的在线数据分析的问题。

产品特征

数据目录管理 提供灵活的数据目录管理及可视化的数据目录呈现，方便业务部门查看系统中有哪些数据。	数据内容检索 提供不同维度的数据内容检索能力，例如：公司数据检索，客户数据检索，行业数据检索等。
文件内容检索 采用AI技术，提供对文件内容地检索能力，例如：上市公司年报内容检索，合同内容检索等。	自定义报告生成 可根据业务需要自定义需要生成报表的格式、内容和生成策略。
功能自定义 业务部门可以根据需要对页面展示的内容进行自定义，做到“千人千面”。	功能及数据权限管理 提供了对功能和权限的细化管理，可以根据实际需要，为业务部门分配相关的功能权限和数据权限。
在线数据分析 提供强大地在线数据能力，例如：提供统一行业下不同公司的财务指标对比，同一公司下不同业务指标的同比和环比等分析。	历史轨迹分析 详细记录各个业务人员的访问操作历史记录，并可以根据历史记录进行轨迹分析，实现智能数据推送，功能推荐等。

关键价值

- 有效地帮助企业内部各个业务部门了解数据情况，帮助其进行数据分析，有效地发挥数据的价值。
- 帮助各个部门提高数据使用效率，提升数据分析和数据使用能力。
- 通过分析各个部门使用数据的历史趋势，实现智能数据推送，业务推荐等，实现数据使用智能化。

服务特色

- 可提供各种形式的元数据管理平台定制化建设服务，构建真正满足业务需要的元数据管理平台。

全方位合作模式



数据湖平台整体输出

- 服务目标：以各种形式帮助客户打造属于自己的各种数据平台，提升对数据的管控能力。
- 预期收益：可开展多种合作模式，打造最适合客户的数据平台。
- 部署方式：客户本地化部署或公有云部署。



数据咨询服务

- 服务目标：帮助客户基于数据湖平台实施各种咨询业务。
- 预期收益：提升业务价值点和业务处理效率。
- 部署方式：客户本地化部署或公有云部署。



徐琅朗

毕马威智慧之光
主管合伙人
毕马威中国

电话：+86 (10) 8508 5432
邮箱：langlang.xu@kpmg.com



开云

管理咨询大数据分析
合伙人
毕马威中国

电话：+86 (21) 2212 2369
邮箱：yun.kai@kpmg.com



陈立节

管理咨询
合伙人
毕马威中国

电话：+86 (10) 8508 5433
邮箱：felix.chen@kpmg.com



蔡冠斌

毕马威智慧之光
总监
毕马威中国

电话：+852 3927 3121
邮箱：robert.choi@kpmg.com



杨晗

管理咨询
总监
毕马威中国

电话：+86 (10) 8508 5465
邮箱：vivian.yang@kpmg.com



王亚军

管理咨询
总监
毕马威中国

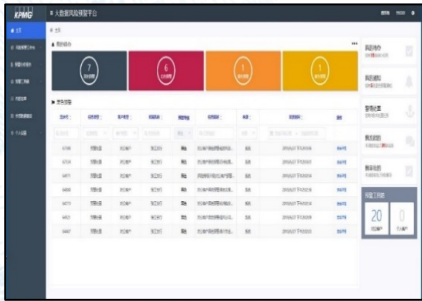
电话：+86 (21) 2212 3672
邮箱：echo.y.wang@kpmg.com

● 联系方式：●

cnfmmclighthousedata@kpmg.com

基于数据湖的应用-大数据风控

附录



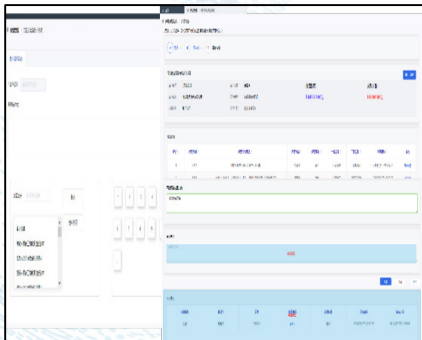
应用场景

- 通过汇集银行内部信贷系统，核心系统，数据仓库等经营数据，以及征信，司法，工商等外部数据，基于毕马威成熟信贷风控模型，有效识别银行客户的信用风险，全面实现对信贷风险地识别、预警、处置、解除的全流程、全生命周期地发现与管控。



方案亮点

- 多年国内外金融机构风险项目的实施经验总结。
- 广泛的大数据应用。
- 全面覆盖各种风险场景的预警模型。
- 关联关系分析和风险传导预警。
- 灵活的预警监控和处置功能。



应用部门

- 合规部、风险部、评审部、法务部等。

业务收益

- 可有效地帮助企业提前识别相关风险，及时处置，避免或减少可能出现的损失。

基于数据湖的应用-反舞弊

附录

应用场景

- 企业内部控制流程存在漏洞时会引发舞弊风险，舞弊风险有可能导致声誉风险和法务风险。毕马威反舞弊数字化解决方案采用自动化的流程、大数据分析模型、可视化的界面，可以帮助客户高效迅速地开展反舞弊调查，调查和预防企业内部的舞弊行为或风险，协助客户做出更好的行动或决策。

内控点测试

(基于重要内控点及风险点设计的测试)



方案亮点

- 简化人工程序，减少人工错误，提高识别地全面性和准确性。
- 自动检测舞弊，节省人力及成本，提高效率。
- 检测潜在的舞弊风险，及时止损。
- 加强合规控制，避免风险扩散。

关联分析

(结合外部数据：工商注册、地理位置等信息)

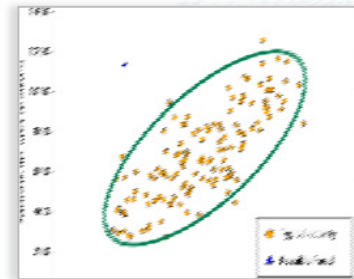


应用部门

- 合规部、风险部、评审部、法务部，销售部，采购部等。

异常值发现

(与正常行为不一致的情况)



业务收益

- 可有效地帮助各个企业识别可能出现的舞弊行为，及时处置，避免或减少可能出现的损失。



毕马威数据湖 让您的数据更智慧！

毕马威智慧之光

kpmg.com/cn/socialmedia



如需获取毕马威中国各办公室信息，请扫描二维码或登陆我们的网站：
<https://home.kpmg.com/cn/en/home/about/offices.html>

本刊物所载资料仅供一般参考用，并非针对任何个人或团体的个别情况而提供。虽然本所已致力提供准确和及时的数据，但本所不能保证这些数据在阁下收取本刊物时或日后仍然准确。任何人士不应在没有详细考虑相关的情况及获取适当的专业意见下依据本刊物所载资料行事。

© 2020 毕马威华振会计师事务所(特殊普通合伙)、毕马威企业咨询(中国)有限公司及毕马威会计师事务所，均是与瑞士实体—毕马威国际合作组织(“毕马威国际”)相关联的独立成员所网络中的成员。毕马威华振会计师事务所(特殊普通合伙)为一所中国合伙制会计师事务所；毕马威企业咨询(中国)有限公司为一所中国外商独资企业；毕马威会计师事务所为一所香港合伙制事务所。版权所有，不得转载。在中国印刷。

毕马威的名称和标识均属于毕马威国际的商标或注册商标。

刊物编号：CN-MARKETS18-SD-0002c

二零二零年三月