



关于数据湖的五个误区

数据及其开启新机会和收入模式的能力非常重要,具备极高的价值;随着企业对这一点的认识越来越高,他们正被那些提供“最佳”方法来管理所有数据和推动洞察分析的技术所淹没。

得益于其具备的种种优势,数据湖已成为这样一种备受关注的参考架构。然而,企业和 IT 经理在研究各类选项的时候,有关数据湖的一些误解会让大家浪费很多宝贵的时间。

1.

数据湖只能部署在单个云中 → 错

数据湖不限于单个位置或单个云,也不局限于本地部署。您可以使用混合选项跨多云构建数据湖。数据湖是一种独立于技术的参考架构。这是组织用来将数据置于其运营核心的一种方法,便于大规模访问各种数据类型并为用户提供自助分析。

2.

Hadoop 是唯一的数据湖 → 错

即使数据湖这个术语经常与 Hadoop 或面向 Hadoop 的对象存储等关联,但数据湖也可以在无 Hadoop 环境中有效开发和运用。例如,一个有效的数据湖可以构建于不同的关系数据库管理系统之上。数据湖融合多种技术来建立洞察分析系统,为数据科学家提供敏捷的数据探索,以应对业务需求。

3.

您可以使用数据湖转储任何数据——无需治理 → 错

虽然软件和硬件是数据湖解决方案的关键组成部分,但数据编目、数据质量以及数据治理和管理流程等都同等重要。

正如一些数据仓库已经成为海量数据永远无法逃脱的巨大黑洞一样,如果没有采用良好的治理策略,数据湖也可能会变成数据沼泽。

4.

数据湖的成功是通过提供访问保障来衡量的 → 错

将数据转储到中心位置并不是真正的分析解决方案。数据湖的目标通过运行数据分析,从而产生富有意义的业务洞察成果,以及发现新的营收来源、客户留存模型或产品拓展方式。

但这些数据必须是可信的、相关的,并且可供所有数据使用者访问。数据湖需要一个智能元数据目录,该目录可以关联业务术语、移动加密编码数据并可确保数据在上下文中更易于理解。它还将根据数据的来源和质量将结构化和非结构化的信息资产以及治理架构分配给数据,以确保信息受到保护和有效管理、符合标准化要求,并值得信赖。

5.

数据湖是数据仓库的替代品 → 错

数据湖可以纳入多个企业数据仓库 (EDW), 以及其他数据源, 例如来自社交媒体或物联网的数据源。这些都汇聚在可嵌入治理模式的数据湖中, 从而简化组织内所有用户采取可信方式探索数据的过程。

因此, 数据湖增强了 EDW 环境, 为数据科学家和分析师赋能, 支持并助力大家轻松探索数据并发现新的视角、洞察成果和机会, 从而加速创新和业务增长。



多云数据湖的优势

不会被错误信息所困扰;受治理的数据湖可以在整个企业范围内提供对各种结构化和非结构化数据的访问,同时可确保在任何地方都能获得可靠安全的数据。

针对业务需求进行优化时,受治理的数据湖可加速分析并提高洞察成果的准确性,因为:



数据存储于安全可靠的基础架构设施中。



受控的数据源输送可靠的信息填充数据湖,然后记录信息资产、元数据和业务环境,从而为数据湖提供实时数据流。



数据的质量、来源和谱系清晰明了。



数据以商业语言呈现,让数据科学家能够立即开始工作,而不是纠结于术语的晦涩含义。



数据经过适当的分类、保护和治理。

了解数据湖的真实情况

IBM 数据管理组合成功帮助客户避开了有关数据湖的常见陷阱和误区,有助于客户确定关键执行步骤,从而能顺利部署受治理的数据湖方案。

探索 IBM 数据湖解决方案

运用来自更多数据源的更多数据类型,促成更明智的决策。



阅读电子书

探索受治理的数据湖如何创造机会,从而获取关键业务洞察成果。

