


# 大数据项目的 隐形杀手








如何避免代价惨重的重大大数据错误

错误越大，  
摔得越重。



越来越多的大公司  
意识到大数据存在  
巨大机遇。

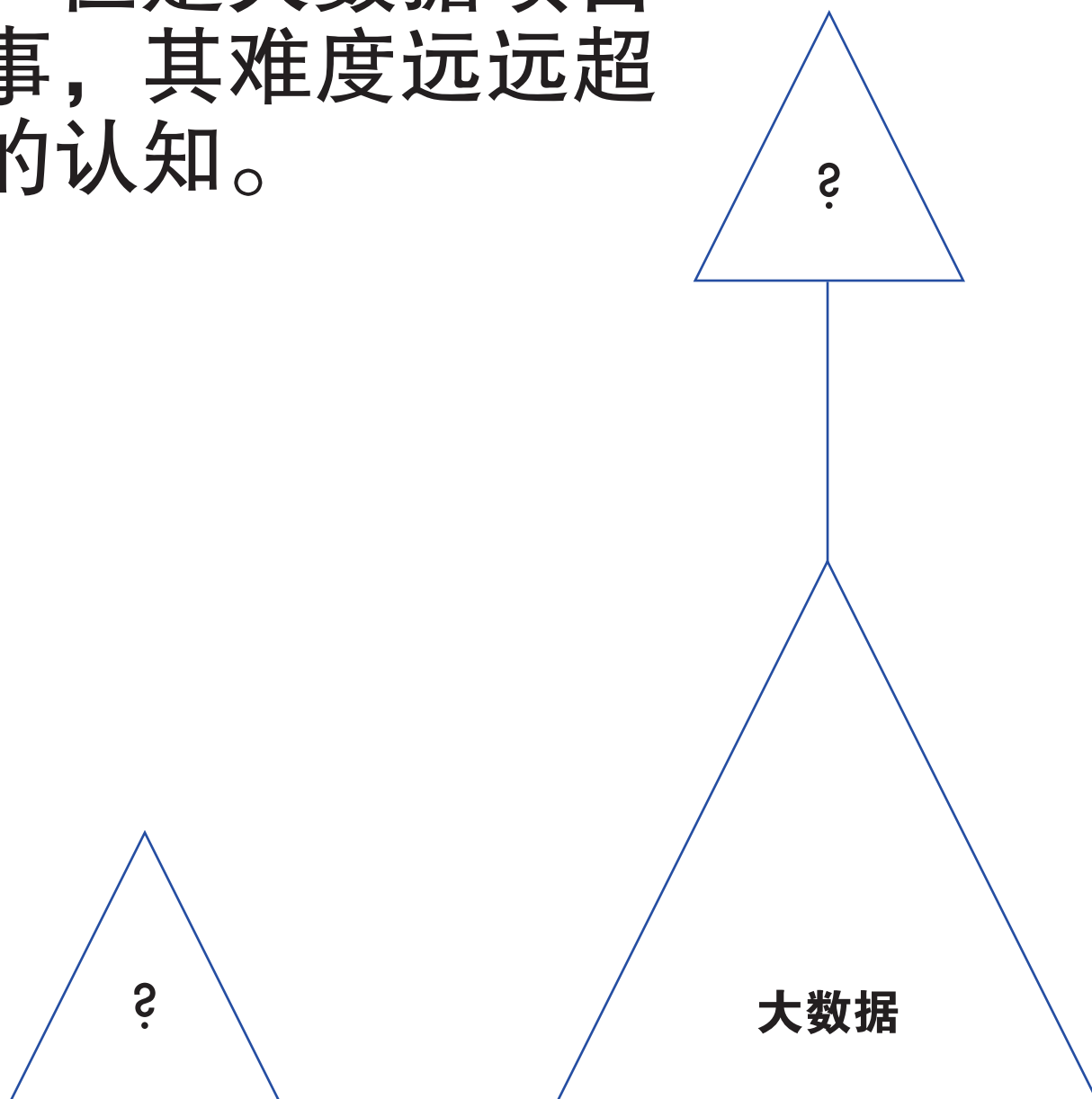
# 机遇：

-  了解客户及其与公司的关系。
-  更好地向客户推荐产品。
-  了解客户如何通过不同渠道进行交互。
-  为风险经理提供实时分析仪表盘。
-  优化管理库存的方式。
-  更加有效地管理供应商关系。
-  基于实时需求分发产品。



大数据所带来的  
实实在在的机遇  
不胜枚举。

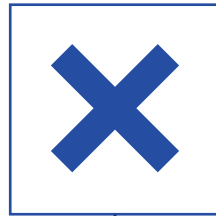
虽然大数据可以解答现代企业面临的部分重要问题，但是大数据项目执行起来并非易事，其难度远远超过了大多数企业的认知。



实际上，

—半

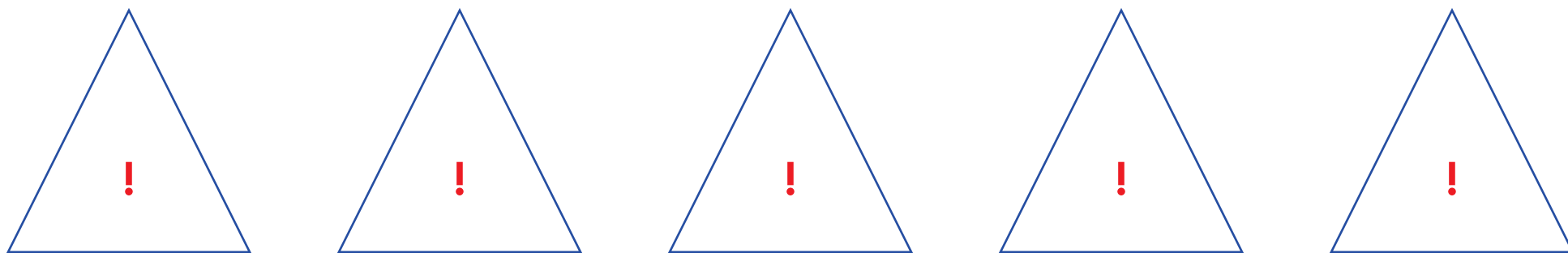
以上的大数据项目均已失败告终。<sup>1</sup>



究其原因，不外乎成本超支、  
过度拖延、目标不明确和缺乏  
数据人才，等等。<sup>2</sup>



但是，对于大数据项目，我们往往会忽略隐藏在所有上述原因表面之下的最大威胁：



# 坏数据



# 重复、不一致、不完整、 包含错误、不可靠的数据



重复原因 — 数据来自  
多个源系统。



不一致原因 — 数据  
存储格式不同。



不完整原因 — 数据输  
入未经验证。



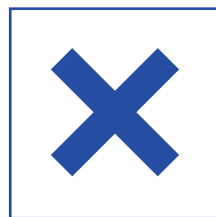
包含错误原因 — 从未  
清理过数据。



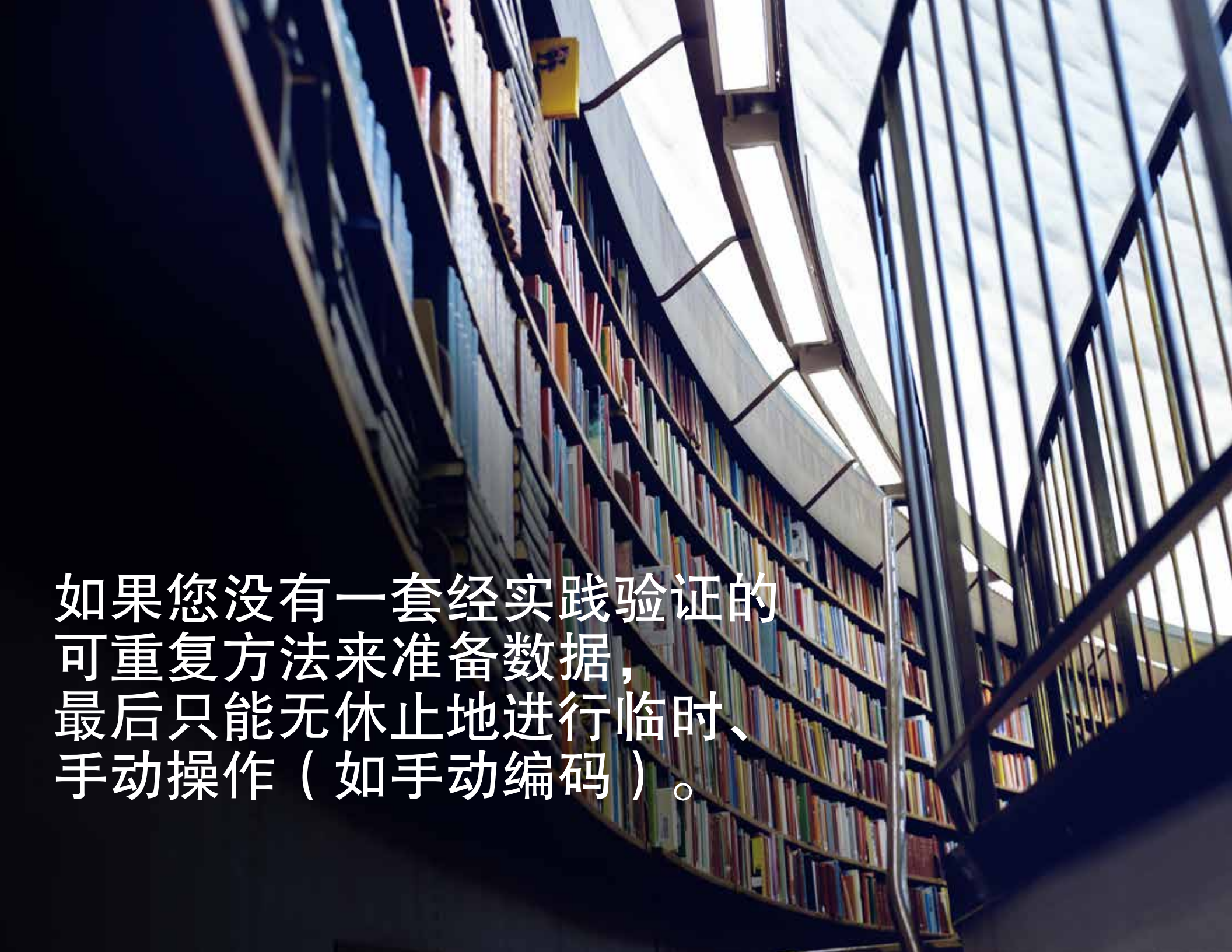
不可靠原因 — 使用  
真实性相冲突的最新  
或非最新版本。

问题在于，众多组织在将数据注入其分析和应用系统之前，低估了其数据准备不当的程度。





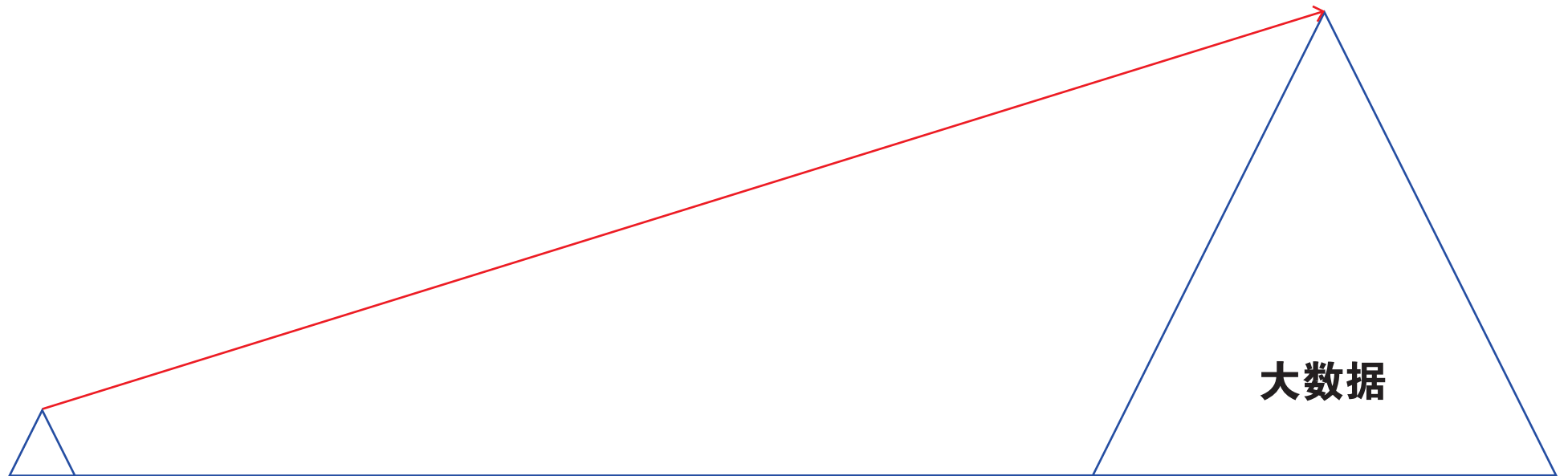
因此，如果相关基本数据准备不当，无论分析技术如何高端精密，也难逃失败厄运。



如果您没有一套经实践验证的  
可重复方法来准备数据，  
最后只能无休止地进行临时、  
手动操作（如手动编码）。

尽管在小规模数据中使用您的那一套策略时可能不会遇到什么问题，但将此策略应用于规模宏大且极其复杂的大数据时，工作会变得异常繁重。

结果总是惨不忍睹。



# 1

您高价聘请的稀缺、高端数据科学家不得不将大部分时间浪费在数据处理这种低级工作上。<sup>3</sup>

您聘请数据科学家的目的是提供洞察，但最终他们在此方面花费的时间只占了工作总时间的一小部分。



# 2

您的主管不得不根据接收到的不可靠数据来制定决策。但是，如果他们无法信任这些数据，那么他们将不会信任分析，当然也不会信任这些“洞察”。

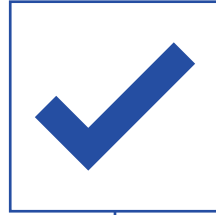
因此，他们投入大量资金进行的数据驱动型决策制定，到头来不过是数据所带来的失望。

# 3

手动准备大数据不仅会影响项目进度、花费过多成本，而且还出了名地缺乏可持续性。这些大数据无法扩展、包含有错误，而且耗时费力难以重复。

突然，您的大数据项目开始让人感到耗财费时而且不堪一击，变成公司的一种不可靠负担。

告诉您一个好  
消息：



有更好的途径来处理大数据。

# 首先需要 主数据管理 (简称 MDM)

MDM

如果您听说过此技术，则可能想要了解它如何推动大数据决策制定。

[阅读电子书](#)

# MDM 通过一些重要元素为 您的大数据项目提供支持：



使您能够全方位了解与客户、产品、  
供应商和站点相关的重要数据。

# MDM 通过一些重要元素为您的大数据项目提供支持：



在丰富的“主”配置文件中管理所有数据，自动协调重复数据。因此，它知道需要将有关“J. Robinson”和“Jodie Robinson”的记录合并到一个事实版本中。

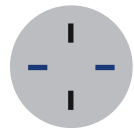
# MDM 通过一些重要元素为您的大数据项目提供支持：



自动保持所有数据干净、精确且处于最新状态，甚至会将所有质量良好的数据重新注入分析和应用系统中，从而改善整个企业内的流程和决策制定。



# MDM 通过一些重要元素为您的大数据项目提供支持：

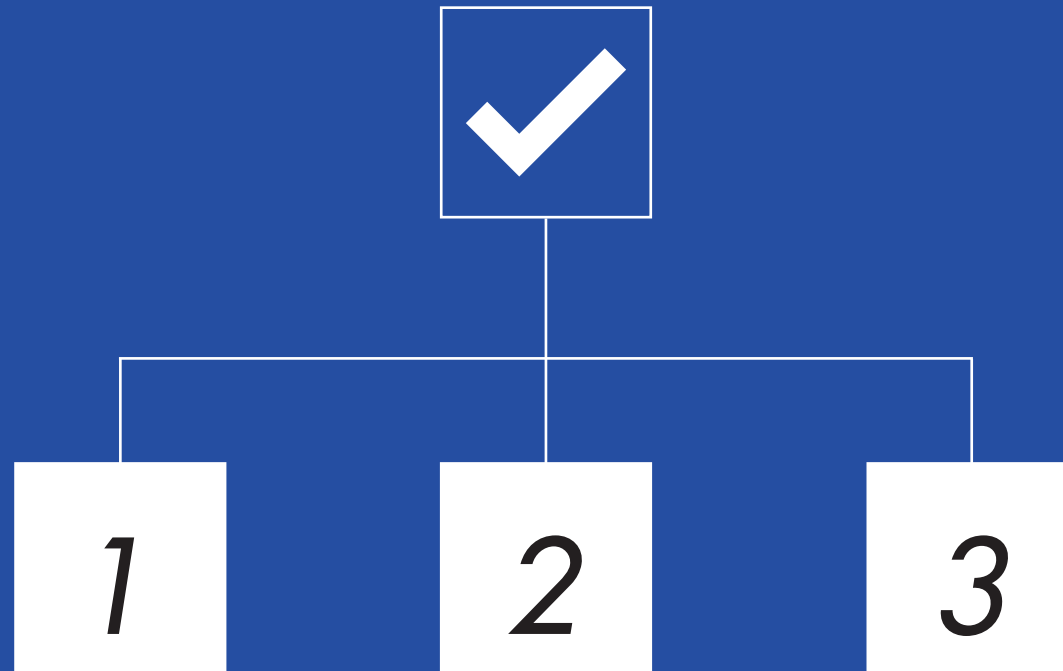


由于它负责管理这些主配置文件，因此它可以向您展示其之间的关系。这样，您就可以详细了解 Jodie 与其中某款产品的所有交互。或者，发现 Jodie 与您的另一位客户 George 是夫妻关系。

这是促就大数据  
决策就绪的最有  
效途径。

意义重大

- 存在 3 大原因:



1

您高价聘请的稀缺、高端数据科学家可以专注于提供您所期望的洞察（而无需将时间浪费在琐事和数据清理上）。

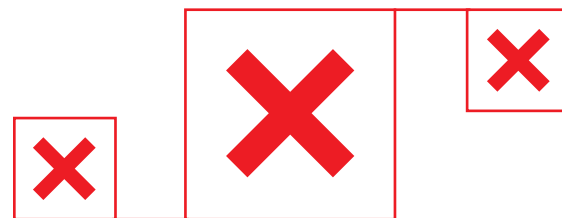
2

您的主管可以根据值得信赖的数据制定决策（而不是十指交叉，祈祷这些图表不会出错）。

# 3

您的大数据项目具有所需的可靠性和可扩展性，以提供组织所需的洞察（而不是每次根据需求花一大笔钱并浪费数月时间进行扩展）。

简而言之，它为您摆脱了  
大数据项目的头号杀手：



坏数据





并且，它还为您提供  
了所有干净、互联的  
重要数据。



这样，您便可通过掌控大数据来大干一场。

这样，您便可实现

决策就绪

# 补充阅读

要了解 MDM 提供了哪些大数据处理方式，请阅读：  
大数据时代的主数据管理。  
让企业决策就绪。



[立即阅读](#)

# 关于 *Informatica*

Informatica 致力于帮助各种规模的企业处理大数据以实现决策就绪。借助我们的 MDM 解决方案，您可以全方位了解客户、产品、供应商和位置，并为业务用户提供支持。

A large, thin red circle is centered in the lower half of the page.

**请联系我们**

# 资料来源

1. <http://www.analytics-magazine.org/july-august-2014/1074-the-data-economy-why-do-so-many-analytics-projects-fail>
2. <http://blogs.gartner.com/svetlana-sicular?s=0lbid>
3. [http://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html?\\_r=0lbid](http://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html?_r=0lbid)