



# 恒天软件大数据白皮书

## Contacts

### 目录

- 01 大数据的概念
- 02 大数据的应用
- 03 企业实施大数据面临的挑战
- 04 恒天软件的大数据解决方案



# 大数据的概念

大数据最初是指由于数据量太大，以至于不能用传统的软件工具来搜集、存储、处理以及分析的数据集。这个从技术角度定义的大数据范畴，往往是指用诸如Hadoop MapReduce、Spark等分布式计算模式，NoSQL和内存数据库等新兴的存储方式，以及最新的数据压缩、大数据可视化等前沿技术进行分析处理的数据集。

然而，随着计算机硬件处理速度和存储能力的快速增强，以及大数据技术的日渐成熟，原本对大多数企业和机构来说无力累积、处理和分析的数据集，如海量的历史数据、互联网数据、日志数据等，已经可以通过一定的成本付出，转化成源源不断的价值。至此，*大数据的概念被更多地理解为利用数据的处理和分析为企业和机构带来价值的模式，特别是涉及到利用和融合传统业务数据以外的数据集带来的创新性的价值提炼手段。*

对企业来说，这种非技术型的大数据定义更为贴切。一般来说，大数据在企业的成功实施需要具备三个条件。第一是技术条件，包括内部数据的集成、外部数据的采集、海量数据的存储、处理和挖掘、以及分析结果的应用和展示等支撑整个数据生命周期的能力；第二是业务条件，即是否能在充分理解数据潜在价值的基础上，从易到难地梳理出企业相关的业务需求和问题，从而清晰地制定大数据项目的业务目标；第三是业务目标与技术能力的无缝结合，即采用合适的、可扩展的技术方案，以逐步实现业务目标的节奏展开，并保持持续的生命力。从实际情况看来，第三条往往是阻碍企业大数据战略顺利发展的最大因素。



# 大数据的应用

从应用层面来看，互联网是大数据应用的发源地和领跑者。然而向更广阔的金融、零售、电信、公共管理、医疗卫生等传统领域的渗透，才是大数据爆发更大影响力的所在。目前大数据应用的发展也确实呈现出往传统行业蔓延的趋势。

## 大部分传统企业通常通过两个方面获取数据价值：

1. 整合行业和企业内部的已有的各种数据源，通过集成、分析和挖掘，为业务提供支持和服务。如百货公司可以利用大数据平台整合支付系统、CRM系统、ERP系统等的数据，打通用户数据，基于用户的基本属性和行为属性进行用户价值分类、交叉销售分析、客户流失分析等。
2. 借助外部数据，包括第三方数据和互联网数据，来实现相关的应用。尤其是互联网数据的方便获取和低成本，逐渐显示出了它的优势，如证券分析机构通过微博舆论进行的股价预测，零售企业借助对网络新闻数据、社交数据的分析来做舆情的监控和危机公关等。

企业对数据的重视已经到达了前所未有的高度。越来越多的企业家已经认识到，数据是企业最重要的资产之一。按照世界经济论坛报告的看法，“大数据为新财富，价值堪比石油”。大数据之父维克托则乐观预测，数据列入企业资产负债表只是时间问题。数据的保存、备份、安全，作为资产的存在被企业所重视。具体来说，大数据的成功运用可以在以下方面给企业带来显著价值：

### ● *产品创新*

一方面，可以通过大数据进行市场分析，发现新的市场需求，研发新的产品；另一方面，对已有产品，可以量化地监测产品功能的使用频率、使用方式和使用途径等，从而改进产品的功能和设计，为产品的发展提供有力的支持；

### ● *渠道创新*

利用对产品或服务的相似性以及用户相似性的分析，可以针对既有客户进行精准营销，推荐用户可能喜欢的产品或服务，从而降低渠道获取的成本，提高销售额；

### ● *优化决策*

大数据分析对正在发生的事实的展现以及对未来趋势的预测，可以有效地辅助企业的决策。同时在执行决策或者活动期间，可以通过监控和量化分析数据，来考察决策的效果以做出实时调整。

### ● *服务创新*

对客户提供服务的公司可以利用大数据分析不同个体的消费习惯，通过用户的基本属性和行为属性，如年龄、性别、消费时间、购买量、消费习惯等的组合来划分客户，更精准地给客户群体画像，从而给不同客户提供不同形式、内容和数量的服务。

### ● *风险防范*

利用大数据对企业风险指标进行更全面更实时地监控，进一步降低企业风险。如零售行业通过对社交网络进行舆情监控，及时发现品牌负面信息，从而采取有效的应对措施。



# 实施大数据面临的挑战

- **数据策略的制定**

中国仍然处于企业信息化的时代，许多企业虽然对数据资产的认识和数据处理的框架还处于起步阶段。但为了占据先机，不得不开始策划大数据战略。然而由于大数据的实施涉及到整个公司对数据的收集、管理、授权、分发等策略，以及实施周期和投入产出比等问题，往往要求公司在启动之前制定比较明确的大数据策略，以保证项目的成功。可借鉴案例的缺失使得大部分企业很难制定出有效的真正适合自己的数据策略。

- **跨业务系统的数据整合**

外部数据越来越重要，而内部业务产生的数据始终都是影响企业决策的重要部分。随着企业的发展，会催生越来越多的业务部门，而每个业务部门产生的数据采用不同逻辑、不同的单位，甚至不同的方式存储在各个业务系统中。往往从单一部门数据分析得出信息的价值不如各个业务数据关联后分析产生的价值高，这就对企业统一分析、挖掘提出了一个新的要求，如何合理、客观的整合这些数据并且尽可能减少数据的丢失提出了挑战。

- **互联网数据的采集**

进入互联网时代后，越来越多的数据在日常的互联网行为中产生，例如互联网新闻，微博，论坛帖子等。与此同时，越来越多的企业不满足于企业内部所产生的相关数据，他们需要通过对外界数据的分析获取对自己业务有帮助的信息。企业数据采集的传统方式往往是一些市场调研，问卷调查等。在互联网时代中这些方法虽然适用，但是不够高效和精准。在互联网时代背景下，数据采集要求能够快速、精准、合法地从互联网中抓取大量的数据以供企业的后续分析所使用。

- **深入的数据挖掘方法**

传统的数据分析大多基于小样本的数据以及经典的统计学模型而来，但是互联网时代数据量的飞速增长让传统的统计学方式遇到了性能、精度、过拟合等问题。作为统计与计算机技术相结合的数据挖掘逐渐成为企业从海量数据中提取有价值知识的主要方法。面对TB级甚至PB级的数据量，如何借助关联分析、分类、聚类等深入的机器学习算法，从数据中挖掘出有意义的知识是企业面临的又一难题。

- **快速的业务迭代**

如何展现数据挖掘的结果，将其直观的表达给决策层或客户是一个重要的过程。企业内部最常用的就是报表。传统的报表工具依赖于BI，但是其生成方式却不是非常友好，需要业务人员和计算机开发人员协同工作才能生成一定格式的报表。而在当下的敏捷时代信息过剩，快速的变化让这样的方式越来越难以适应，如何高效地迭代报表开发的流程，是大数据背景下值得研究的问题。



## 恒天软件大数据解决方案

恒天软件数据集成与分析解决方案提供从数据集成到数据分析的一站式解决方案，包括提供咨询、制定数据策略、可定制化的数据集成方案、数据基础设施的实施、数据挖掘服务、报表和可视化开发等。恒天软件的方案是基于开源大数据技术，可横向扩展的一站式解决方案，可支持结构化与非结构化数据的管理与分析，同时实现了可插拔和可定制化的高级数据挖掘算法支持。该解决方案同时具备了高级可视化分析功能，统一报表门户平台可实现多元化的报表展示，高度定制化和丰富数据源支持。



图2 恒天软件数据分析平台