

导读：本文介绍数据治理有关的名词和概念。当然，与数据治理相关的概念非常多，以下仅罗列几个常见的。

作者：用友平台与数据智能团队

来源：华章科技

| | |
|------|--|
| 标识符 | DE03.01.001 |
| 中文名称 | 药品通用名称 |
| 短名 | YPTYMC |
| 定义 | 国家药品标准或者国家药典委员会《中国药品通用名称》或其增补本收录的药品通用名称。 |
| 数据类型 | 字符型 |
| 表示格式 | YPTYMC |

▲图1-1 食品药品监管信息基础数据元标准

02 元数据

1. 名词解释

元数据是描述数据的数据或关于数据的结构化数据。

你是不是看了这个定义依然一头雾水？那我们来举个例子。一本书的封面和目录向我们展示了这样的元数据信息：图书名称、作者姓名、出版商和版权细节、图书的提纲、标题、页码等。

2. 主要作用

在数据治理中，元数据是对数据的描述，存储着数据的描述信息。我们可以通过元数据管理和检索我们想要的“书”。可见元数据是用来描述数据的数据，让数据更容易理解、查找、管理和使用。

3. 应用举例

元数据是业界公认的数据治理中的核心要素，做好元数据管理，能够更容易地对数据进行检索、定位、管理和评估。用哲学的思维理解元数据的话，元数据其实解决的是我是谁、我在哪里、我从哪里来、我要到哪里去的问题。

元数据是建设数据仓库的基础，是构建企业数据资源全景视图的基础，清晰的血缘分析、影响分析、差异分析、关联分析、指标一致性分析等是数据资产管理的重要一环。

如果说数据是物料，那么元数据就是仓库里的物料卡片；如果说数据是文件夹，那么元数据就是夹子的标签；如果说数据是书，那么元数据就是图书馆中的图书卡。

| 交易号 | 商品 | 数量 | 价格 | 销售日期 |
|-----|----|----|------|----------|
| 1 | G1 | 10 | 3.00 | 1999-5-1 |
| 1 | G2 | 5 | 2.00 | 1999-5-1 |
| 1 | G3 | 2 | 4.00 | 1999-5-1 |
| 1 | G4 | 5 | 2.00 | 1999-5-1 |
| 1 | G5 | 2 | 3.00 | 1999-5-1 |
| 3 | G1 | 15 | 3.00 | 1999-5-1 |

表1-1 某企业的产品销售记录

05 主题数据

1. 名词解释

主题数据是根据数据分析的需要，按照业务主题对数据所做的一种组织和管理方式，其本质是为了进行面向主题的分析或加速主题应用的数据。

主题数据是分析型数据，是按照一定的业务主题域组织的，服务于人们在决策时所关心的重点方面。一个主题数据可以由多个主数据和交易数据组成。主题数据一般是汇总的、不可更新的、用于读的数据。

2. 主要作用

主题数据是按照一定的业务主题域组织的，服务于各种数据分析或应用开发。

3. 应用举例

主题数据与行业或领域有较大的关系，不同行业关注的主题是不一样的。即使是同一行业，不同企业也有不同的主题数据定义。

例如，某生产制造企业定义了12大主题数据，包括综合服务、人力资源、财务管理、质量管理、生产管理、工艺管理、库存管理、销售管理、采购管理、设备管理、能源管理和安全环保。

06 数据仓库

1. 名词解释

数据仓库 (Data Warehouse , DW) 是一个面向主题的、集成的、相对稳定的、反映历史变化的数据集，用于支持管理决策。数据仓库是数据库的一种概念上的升级，可以说是为满足新需求而设计的一种新数据库，需要容纳更加庞大的数据集。本质上，数据仓库与数据库并没有什么区别。

2. 主要作用

数据仓库是企业所有级别的决策制定过程提供所有类型数据支撑的战略集合，有以下三个主要作用。

- 数据仓库是对企业数据的汇聚和集成，数据仓库内的数据来源于不同的业务处理系统，包含主数据和业务数据。数据仓库的作用就是帮助我们利用这些宝贵的数据做出最明智的商业决策。
- 数据仓库支持多维分析。多维分析通过把一个实体的属性定义成维度，使用户能方便地从多个维度汇总、计算数据，增强了用户的数据分析处理能力，而通过对不同维度数据的比较和分析，用户的数据处理能力得到进一步增强。
- 数据仓库是数据挖掘技术的关键和基础。数据挖掘技术是在已有数据的基础上，帮助用户理解现有的信息，并对未来的企业状况做出预测。在数据仓库的基础上进行数据挖掘，可以对整个企业的发展状况和未来前景做出较为完整、合理、准确的分析和预测。

3. 应用举例

数据仓库是数据分析和数据可视化的基础，通过将来自不同业务系统的数据汇集到一起，并按照一定的主题进行编号、归类、分组，方便用户快速定位数据源，为数

据分析提供支撑。

为了提升数据仓库的数据质量，确保数据分析的准确性，数据仓库的建设需要实施数据治理的策略。很多企业的[数据治理项目](#)实施的动因都是解决数据仓库中的数据质量问题，以便获得更准确的分析决策。

07 数据湖

1. 名词解释

根据维

基百科的定义

，数据湖是一个以原始格式存储数据的存储库或系统。

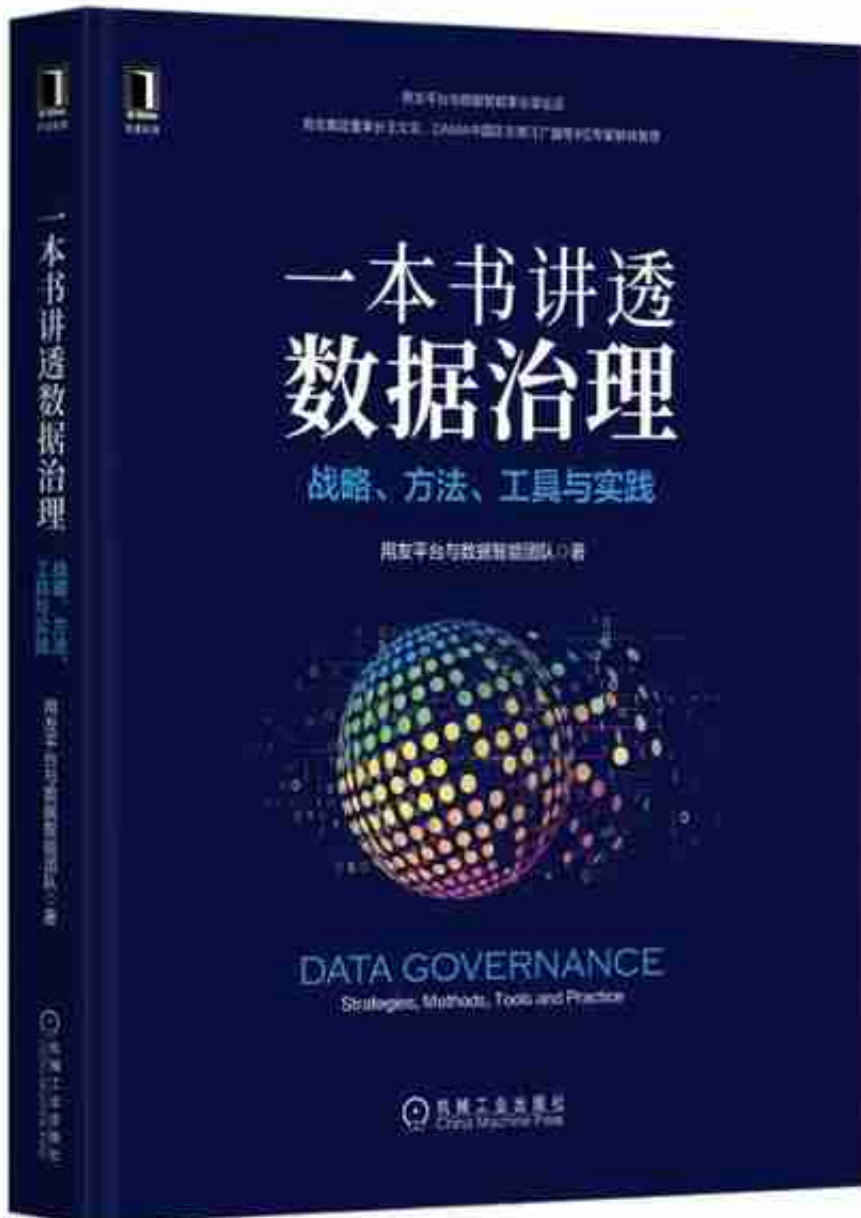
它按原样存储数据，而无须事先对数据进行结构化处理。

数据湖可以存储结构化数据（如关系型数据库中的表）、半结构化数据（如CSV、日志、XML、JSON）、非结构化数据（如电子邮件、文档、PDF）和二进制数据（如图形、音频、视频）。

数据湖可以更方便、以更低的成本解决不同数据结构的统一存储问题，同时还能够为机器学习提供全局数据。我们可以将数据湖理解为一个融合了大数据集成、存储、处理、机器学习、数据挖掘的解决方案。

2. 主要作用

关于数据湖的作用，AWS将它与数据仓库进行了类比，如表1-2所示。



《一本讲透数据治理》

推荐语：

这是一本能为数字化转型企业的数据治理提供全面指导的著作。用友集团官方出品，基于国际主流的数据治理框架和用友多年的数据治理经验，从道、法、术、器4个维度全面、深入展开，不仅有数据治理在战略层面的顶层设计，还有数据治理在执行层面的实施方法，既可以作为数据治理的纲领性指南，又可作为数据治理的实操手册。