# 数据库与数据仓库的区别

## 从定义方面来看

数据仓库，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合，用于支持管理决策。是为企业所有级别的决策制定过程，提供所有类型数据支持的战略集合。它是单个数据存储，出于分析性报告和决策支持目的而创建。 为需要业务智能的企业，提供指导业务流程改进、监视时间、成本、质量以及控制。而数据库则是“按照数据结构来组织、存储和管理数据的仓库”。是一个长期存储在计算机内的、有组织的、可共享的、统一管理的大量数据的集合。

## 从目的来看

构建数据仓库的主要目的在于为企业提供一个决策分析用的工具，帮助决策人员更好地制定企业策略，或找出企业的潜在问题，提高客户满意度，最终提高企业竞争力。主要是为分析数据设计。而数据库的目的主要是高效且条理分明地存储数据，使人们能够更加迅速和方便地管理数据，主要体现在以下几个方面：1) 数据库可以结构化存储大量的数据信息，方便用户进行有效的检索和访问。2) 数据库可以有效地保持数据信息的一致性、完整性、降低数据冗余。3) 数据库可以满足应用的共享和安全方面的要求，把数据放在数据库中在很多情况下也是出于安全的考虑。4) 数据库技术能够方便智能化地分析，产生新的有用信息。

## 从模型来看

数据仓库的模型可分为：星型模式、雪花模式、事实或星座模式。数据库模型可以分为1概念模型、2层次模型、3网状模型、4关系模型、5面向对象模型。而且分的模型的分类原则不同，模型基础的分类对象也不同。

## 从设计原则来看

数据仓库模型的设计原则是“自顶向下、逐步求精”。设计分为三个阶段：1）概念模型，对业务进行系统划分，从整体上进行抽象和概括，也就是主题划分。2）逻辑模型，进一步细化概念模型。定义实体和实体间的关系，和每个实体的属性。即定义具体表的作用，表与表的约束，表的字段。形成ER图。这些实体的设计都是基于业务规则，可以说，这一阶段主要面对的是业务。3）物理模型， 具体实现。依照逻辑模型，在数据库中进行建表、索引等。数据仓库，为了满足高性能的需求，可以增加冗余、隐藏表之间的约束等反第三范式操作。数据库模型的设计原则：1)一致性原则：对数据来源进行统一、系统的分析与设计，协调好各种数据源，保证数据的一致性和有效性。2)完整性原则：数据库的完整性是指数据的正确性和相容性。要防止合法用户使用数据库时向数据库加入不合语义的数据。对输入到数据库中的数据要有审核和约束机制。3)安全性原则：数据库的安全性是指保护数据，防止非法用户使用数据库或合法用户非法使用数据库造成数据泄露、更改或破坏。要有认证和授权机制。4)可伸缩性与可扩展性原则：数据库结构的设计应充分考虑发展的需要、移植的需要，具有良好的扩展性、伸缩性和适度冗余。5)规范化原则：数据库的设计应遵循规范化理论。规范化的数据库设计，可以减少数据库插入、删除、修改等操作时的异常和错误，降低数据冗余度等。

## 总结如下：

1. 主要区别在于数据结构，数据库中的建模一般遵循三范式，而数据仓库的建模有特定的方式，一般采用维度建模，使用这些建模方式的原因是便于OLAP建立，增加统计查询效率等。

2. 数据仓库中数据通常来源于多个不同的业务系统数据库（存储多年数据），数据量较大，一般作为企业数据中心用。

3. 数据库一般存储在线交易数据，数据仓库存储的一般是历史数据。

4. 数据库设计是尽量避免冗余，一般采用符合范式的规则来设计，数据仓库在设计是有意引入冗余，采用反范式的方式来设计。

5. 数据库是为捕获数据而设计，数据仓库是为分析数据而设计，它的两个基本的元素是维表和事实表。维是看问题的角度，比如时间，部门，维表放的就是这些东西的定义，事实表里放着要查询的数据，同时有维的ID。

6. 数据仓库，是在数据库已经大量存在的情况下，为了进一步挖掘数据资源、为了决策需要而产生的，它决不是所谓的“大型数据库”。