

# OLTP vs. OLAP 区别总结

特性	OLTP (On-Line Transactional Processing)	OLAP (On-Line Analytical Processing)
功能	面向客户：日常操作、事务和查询处理	面向市场：复杂查询、数据分析、决策支持
用户	文员、客户、数据库管理员、IT 专业人员	知识工作者：经理、执行者、分析师
数据内容	当前、实时数据，过于详细，不易用于决策	大量历史数据，汇总、集成，不同粒度的信息
数据库设计	面向应用：实体-关系 (ER) 数据模型	面向主题：星型或雪花模型
视图范围	企业或部门内部	多版本数据库模式，跨组织数据
数据访问	数据写入：读/写，主键索引/哈希	信息读取：大量只读扫描
数据规模	数十条记录；GB 到高阶 GB；数千用户	数百万条记录；TB；数百用户

## 为什么需要专门的 Data Warehouse

原因	解释
高性能	数据库管理系统 (DBMS) 优化用于 OLTP，而数据仓库优化用于 OLAP，支持复杂查询和 多维分析。
数据和功能的差异	数据仓库结构化用于分析，具有标准化的模式和整合的历史数据；而操作型数据库处理 频繁的事务和更新。
独立性	数据仓库是独立的存储，避免了操作型数据库的性能瓶颈。
可扩展性和可维护性	数据仓库支持大规模数据分析，具有更好的可扩展性和维护性。

## 什么是 Data Cube

Data Cube 是一种基于多维数据模型的表示形式，用于支持数据仓库中的多维分析。它将数据组织为多个维度和度量的组合，便于用户从不同的角度观察和分析数据。

- **维度 (Dimensions):** 描述数据的不同视角，例如时间 (time)、产品 (product)、地点 (location) 等。
- **度量 (Measures):** 数值型数据，用于分析，例如销售额 (sales)、数量 (quantity) 等。
- **基本概念:**
  - **Base Cuboid (基础立方体):** 包含所有维度的最详细数据。
  - **Apex Cuboid (顶点立方体):** 包含最高级别汇总数据的立方体。
  - **多维立方体 (n-Dimensional Cube):** 包含从基础到顶点的所有立方体。

Data Cube 支持多种 OLAP 操作，例如 Roll-up、Drill-down、Slice、Dice 和 Pivot，用于灵活地探索和分析数据。

## Data Cube 示例

以下是一个包含 0-D、1-D、2-D 和 3-D（Base Cuboid）的 Data Cube 示例，基于销售数据：

数据维度

- 时间 (Time): 年 (Year)
- 产品 (Product): 类别 (Category)
- 地区 (Region): 国家 (Country)

示例数据

Time	Product	Region	Sales (\$)
2023	Electronics	USA	1000
2023	Furniture	USA	500
2023	Electronics	Canada	700
2023	Furniture	Canada	300

Cuboid 层次

1. 0-D (Apex Cuboid): 汇总所有维度的数据。

- 总销售额: \$2500

2. 1-D Cuboid: 按单一维度汇总数据。

- 按时间汇总:

Time	Sales (\$)
2023	2500

- 按产品汇总:

Product	Sales (\$)
Electronics	1700
Furniture	800

- 按地区汇总:

Region	Sales (\$)
USA	1500
Canada	1000

3. 2-D Cuboid: 按两个维度汇总数据。

- 按时间和产品汇总:

Time	Product	Sales (\$)
2023	Electronics	1700

Time	Product	Sales (\$)
2023	Furniture	800

- 按时间和地区汇总:

Time	Region	Sales (\$)
2023	USA	1500
2023	Canada	1000

4. **3-D Cuboid (Base Cuboid):** 包含所有维度的详细数据。

- 见示例数据表。

通过这些 Cuboid，可以灵活地进行 OLAP 操作，如 Roll-up、Drill-down、Slice 和 Dice.