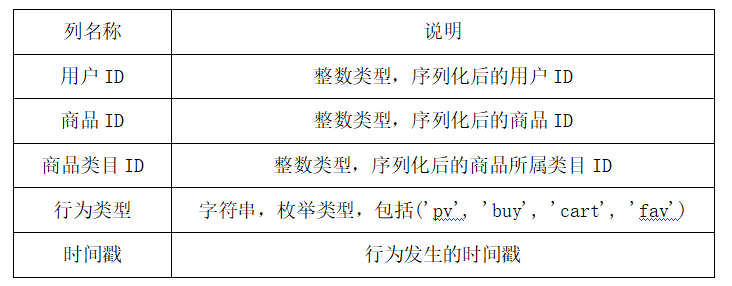
# 淘宝交易数据处理与分析

**一、题目简介**

本题目是根据所学的大数据基础的相关知识对提供的淘宝数据进行分析，并对分析结果进行可视化展示。包括用户行业统计、商品销售统计、用户购买商品数量统计、商品销售统计图等。

1. **数据集说明**

该数据集包含了2017年11月25日至2017年12月3日之间，有行为的约一百万随机用户的所有行为fav（行为包括点击、购买、加购、喜欢）。数据集的每一行表示一条用户行为，由用户ID、商品ID、商品类目ID、行为类型和时间戳组成，并以逗号分隔。关于数据集中每一列的详细描述如下：

[](https://dblab.xmu.edu.cn/blog/wp-content/uploads/2022/09/LSH5XNSXLTW7EAGLFSX.png)

关于数据集大小的一些说明如下：

[](https://dblab.xmu.edu.cn/blog/wp-content/uploads/2022/09/SKGCAUID8ZMCOSK36EZWXW.png)

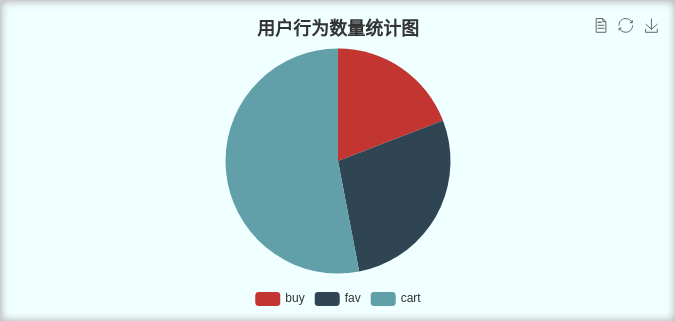
**三、步骤概述**

完成课题，大致包含以下5个步骤：  
（1）建立环境。（Hadop、HBase、Hive、MySQL、Sqoop)  
（2）创建数据源读取数据,并将数据存储进MySQL。  
（3）对数据集指定转换操作，保存到HBase。

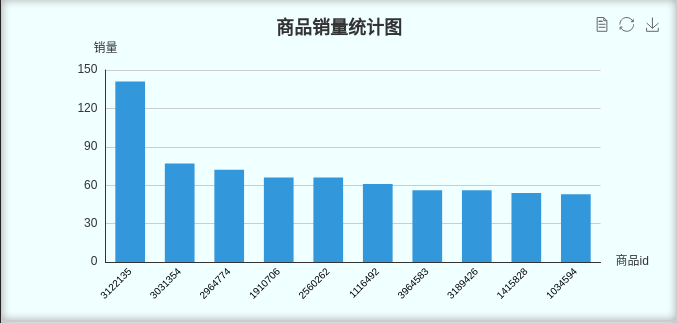
（4）利用Hive实数据进行分析，并将结果保存到MySQL。  
（5）对结果进行可视化。

**四、分析结果样例**

#### （1）用户行为数据统计



#### （2）商品销售统计



#### (3)用户购买商品数量统计



#### (4)商品销售统计图

