# 初识Hadoop

HDFS实现数据的存储，MapReduce实现数据的分析和处理。

MapReduce是一个批量查询处理器，能够在合理的时间范围内处理针对整个数据集的动态查询。

MapReduce比较适合以批处理方式处理需要分析整个数据集的问题，尤其是动态分析。RDBMS适用于点查询和更新，数据集被索引之后，数据库系统能够提供低延迟的数据检索和快速的少量数据更新。MapReduce适合一次写入多次读取数据的应用，关系型数据库则更适合持续更新的数据集。

MapReduce和关系型数据库之间的另一个区别在于它们所操作的数据集的结构化程度。结构化数据是具有既定格式的实体化数据，如XML文档或满足特定预定义格式的数据库表。半结构化数据比较松散。非结构化数据，例如纯文本或图像数据。MapReduce对非结构化数据或本结构化数据非常有效，因为它是在处理数据时才对数据进行解释。

数据本地化特性是MapReduce的核心特征，并因此而获得良好的性能。