

大数据的基本特征：

容量大：每年全世界创造的数据量高达ZB级别(1021B)；

速度快：数据增长、更新换代速度快；

多样性：数据形式多样化，包括传统的结构化、半结构化、非结构化的数据；

价值高：部分信息高价值、低密度；

真实性：数据质量的真实性。

大数据时代面临的问题及方法：

1. 如今随着数据指数级爆炸，数据采集是很大的调整。需要考虑超大规模、分布式的、多种类的数据采集方式；
2. 采集到数据之后，我们还面临超大规模的数据存储的问题。我们可以不同于传统的信息论方法，通过尽可能压缩信息冗余度以获取数据压缩，我们在此基础上，根据重要性分配存储资源，进一步舍弃不重要的信息进行压缩；
3. 拥有如此庞大的数据量之后，如何传输也是一个挑战。我们可以采用NIM信息传输处理等方式，考虑到不同事件的传输差错的代价不同，采用基于信息重要性的加权重建误差；
4. 我们还需要基于用户个性化需求做出数据挖掘的决策。需要我们探索更先进的数据挖掘、推荐系统等方面的知识。