1. **（1）简述大数据**

大数据的定义：

目前业界对大数据还没有一个统一的定义。常见的研究机构基于不同的角度给出如下定义。

麦肯锡：大数据是指大小超出常规的数据库工具获取、存储、管理和分析能力的数据集

维基百科：大数据是指无法在一定时间内 用常规软件工具对其内容进行抓取、管理和处理的数据集。

大数据的特征：

现在普遍用5v来描述大数据，其反应了大数据在5个方面的特点：

Volume(巨量性)：数据量巨大。这是大数据的显著特征，数据集合的规模不断扩大，已从GB到TB再到PB级，甚至开始以EB和ZB来计数。

Variety(多样性)：数据类型复杂多样。以往生产或者处理的数据部分是结构化数据，如今，还包含大量半结构化和非结构化数据。

Velocity(高速性)：数据具有高速性。数据产生、处理和分析的速度持续在加快、数据流量大。

Veracity(准确性)：数据准确性。该特性体现了大数据的数据质量。较为典型的应用是垃圾邮件，它们给社交网络带来了严重的困扰

Value(高价值、低价值密度)：数据具有潜在价值。大数据由于数据体量不断扩大，单位数据的价值密度不断降低，而数据的整体价值在提高

1. **简述大数据思维**

维克托.尔耶.舍恩伯格在《大数据时代：生活、工作与思维的大变革》中最具洞见之处在于，他明确指出，大数据时代最大的转变就是，放弃对因果关系的渴求，而取而代之关注相关关系。也就是说只需要知道是什么，而不需要知道为什么。大数据不仅是一种资源，也是一种方法，伴随大数据产生数据密集型科学，是继实验科学、理论科学和计算科学之后的第四种科学研究模模式，这一研究模式的特点表象为不在意数据的杂乱，但强调数据的量；不要求数据精准，但看重其代表性；不刻意追求因果关系，但重视规律总结。这一模式不仅用于科学研究，更多的会用到各行各业，成为从复杂现象中透视本质的有用工具。

在无法确定因果关系时，数据为我们提供了解决问题的新方法，数据中所包含的信息可以帮助我们消除不确定性，而数据之间的相关性在某种程度上可以取代原来的因果关系，帮助我们得到我们想知道的答案，这便是大数据思维的核心。大数据时代，计算模式也发生了转变，从“流程”核心转变为“数据”核心。Hadoop体系的分布式计算框架已经是“数据”为核心的范式。非结构化数据及分析需求，将改变IT系统的升级方式：从简单增量到架构变化。大数据下的新思维——计算模式的转变。这主要包括如下几个转变。

1）由功能为价值转变为由数据为价值

2）从抽样转变为需要全部数据样本

3）由精确度转变为效率

4）由因果关系转变为相关性

5）从不可预测到可预测

6）从人找信息转变为信息找人

7）由人懂机器到机器更懂人

1. **谈谈你对大数据的理解与展望**

**理解：**

大数据是一个庞大的数据集，在当今社会扮演着重要角色，与人们的生活息息相关。大数据的本质就是一推结构化和非结构化的数据集合，由于数据量太大，所以没有办法全部使用这些数据，这时我们就需要从中抓取我们需要的数据并进行分析。随着大数据的普及，我们的生活方式发生了很大的改变，并且大数据在无形中影响着我们的生活。比如最近流行的学生人才画像，我们大学生的大部分信息都被记录，例如去图书馆的时间与频率，离开宿舍的时间、考试分数等等，这些都属于大数据。然后通过特定的算法从这些数据中选择合适的数据来进行分析，可以较为准确的得出学生的各项能力如独立思考能力、学习能力、创新能力等。这些结果可以让招聘公司能够立即了解到学生的大致情况，从而可以提高招聘的对口率以及提升招聘的效率。最让我们体会到大数据就在身边的例子就是淘宝了。我们往往会发现淘宝会自动推荐给我们一些我们想要的东西，并且货源就在附近。这就是很明显的应用了大数据。比如通过不同人群的使用习惯，它可以统计出南北方地区不同省份人群的特点，比如：不同省份人员的网购量，哪个性别，那个年龄段购买最多，哪些地方哪类商品销售最好。当你在淘宝上搜索奶瓶时，往往会推荐你另外一些婴儿用品，这就是大数据的强大之处。大数据在医疗和科技发展方面也发挥了重大作用。当然大数据可能也会有一些弊端，例如我们在某些视频平台观看视频，平台会给我们大量推送相同类型的视频，因此会缩小我们的社交圈，限制我们的发展。总的来说大数据无处不在，已经深深的走进了我们的生活。

**展望：**

大数据在未来是人类社会离不开的技术，并且会发展的非常完善，运用场景非常多。在医疗方面：大数据能够结合过去已经获取的疾病信息来预测未来可能出现的疾病，并大大加快抗病毒药物的研发速度，这是非常重要的应用。因为疾病是人类生命最大的威胁，并且在短时间内很难找出应对新型疾病的治疗方法。然而通过大数据可以让我们未雨绸缪，通过对各种疾病的分析可以让我们察觉到病毒的演变规律从而提前做好防范，并且生产疫苗或特效药来阻断疾病的传播。在教育方面：教育资源分配不平衡是大家非常关心的问题，但是目前并没有特别有效的手段来规避这些问题，由于不同地区有不同的实际情况所以教育资源很难评估。大数据的出现使的这成为可能。大数据可以整合不同地区的教育资源使用情况，并根据当地的实际情况给出客观的评价，从而使得教育资源可视化，为分配问题提供有力的参考依据。在环境保护方面：科学家已经预测了很多未来的气候变化例如温度升高的趋势以及冰川融化的时间，但是这些都是长期的预测，像地震、洪水、海啸这些突发性灾难至今还没有有效的预测方法，往往这些突发情况都会造成巨大的损失。通过对大数据的研究的不断深入，这些突发事件在未来都可以通过大数据来预测。这将会大大降低自然灾害对人们造成的损失。并且大数据通过对环境各项指标的评估还可以提醒人们保护环境，这些由大数据所得出来的数据会更让人们信服。大数据在未来是人们所不可或缺的，对人类的发展起了很大的推动作用，甚至会彻底颠覆人类对某些领域的认知。