1. 何为大数据？

* 根据wiki：

“Big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools”

这个定义中给出了大数据的关键特点：数据过于复杂，无法用常规数据库管理工具处理

* 根据IDC报告：

“Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis.”

这个定义关注与大数据技术，指出大数据具有的“very large volumes”和”wide variety” 的特点

* 根据IBM定义：

Big data analytics is the use of advanced analytic techniques against very large, diverse data sets that include structured, semi-structured and unstructured data, from different sources, and in different sizes from terabytes to zettabytes.

这个定义关注于大数据分析，指出大数据包括了来自各个来源不同大小的结构化、半结构化、非结构化数据。

1. 大数据有哪几个特征？

大数据具有5v特征：

Volume：大体量，可以从几百TB到EB的规模，数据量很大

Variety：多样性，大数据包括了各种各样的数据，数据的来源多、数据的类型多、数据的关联性强

Velocity：时效性，很多大数据需要在一定时间内得到处理，否则就不再有意义

Veracity：准确性，大数据可以为我们提供更准确和全面的信息

Value：价值大，大数据是新时代的新石油，里面包含了大量的信息和价值等待挖掘。但是其价值密度也较低。

1. 请解释数据和信息的区别和关系。

数据指的是在物理世界中收集到的全部数据集，它可能是数字、图像、语音等等。根据维基百科，数据是指未经过处理的原始记录，是关于事件之一组离散且客观的事实描述，是构成消息和知识的原始材料。而信息的经典的哲学定义有北京邮电大学钟义信教授：“信息是事物运动的状态及其改变方式”。从信息科学的角度来说，信息是反映事物的形成、关系和差别的东西，是可以直接被利用的资源。

因此，数据中可以包含信息，对数据进行一定处理之后可以得到我们想要的可以利用的信息。这是一个从物理世界到逻辑世界的转换过程，也是一个数据的处理和信息的提取过程。

4. 请简述金融数据的特征。

* 数据量大：金融数据往往在交易等金融操作中产生，具有数据量大而杂的特点
* 时效性高：金融数据来自于真实的市场情况，而市场是在不断变化的，因此金融数据往往具有“流数据”的特点，需要在短时间内快速处理，否则其有效性将快速下降。
* 复杂度高：金融数据产生于日常经济活动中，包括各种类型的数据，还可能包括大量非结构化数据，处理难度较大
* 可靠性要求更高：金融数据影响到经济参与者的经济利益，甚至可能影响社会和国家利益，因此对于可靠性和安全性的要求更高