1. **Describe a business situation in which Hadoop would be a better choice for storing the data than a relational database and explain why (give at least three reasons)**

Social sentiment analysis is a popular business case of using Hadoop and advanced text

analysis tools to analyze unstructured text from social media and social networks (including Tweets and Facebook) to determine the user's emotions for a particular company, brand, or product.

Why use hadoop:

1. Hadoop is a cost-effective and efficient way to handle and analyze a large number of unstructured and semi-structured data, as Hadoop clusters can be extended to PB-level or even EB-level data, companies no longer have to rely on sample datasets and can process and analyze all relevant data. (Traditional data warehouse’s capacity is rarely more than a few terabytes of data, because large amounts of data can consume data warehouse resources and reduce performance.)
2. Hadoop cluster runs on cheap commercial hardware, so there is no financial pressure on hardware expansion. On the other hand, the expansion of traditional enterprise data warehouses to adapt to potential petabytes of data requires significant investment in new dedicated hardware.
3. Hadoop's default file storage system is the Hadoop distributed file system. HDFS File system is good at storing large amounts of unstructured and semi-structured data because they do not need to organize data into relational rows and columns. This feature distinguishes it from "traditional" business data.
4. **Describe a business situation in which a relational database would be the better option and explain why (give at least three reasons)**

OLTP is known as transaction-oriented processing system in that the customer can immediately transfer the original data to the computing center for processing, and in a very short period of time to give the results of the process. The OLTP database is designed to allow the transactional application to write only the required data in order to process a single transaction as soon as possible. OLTP system is popular used in bank and financial organizations where they consider reliability, adaptability, scalability, predictability, and manageability as the most important factors to implementing the OLTP transaction system. In this case RDBMS system consider to be better choice such as Real-time requirements are high, the integrity and security of the transaction are strict, the data consistency is stable and data update overhead is very small.

1. **Describe a business situation where MongoDB would a better choice than either and again give at least three reasons**

Mongodb is mainly used to store billing data and log data. For example, when using MongoDB in O2O Express application, we can store courier rider, courier business information (including location information) in MongoDB, and then through the MongoDB geographical location query, we can easier locate nearby businesses, riders and other functions etc.

Advantages using Mongodb:

Mongodb supports clustering and sharding, support space retrieval, sharding fragmentation function, easy to query.

Mongodb can do real-time insert, update and query, and the site real-time data storage required for replication.

Mongodb’s high scalability is very suitable for dozens or hundreds of servers composed of the database.

Mongodb can also be used to store objects and JSON data. MongoDB's BSON data format is well suited for document formatted storage and queries applications

1. **What is the point in having and using Pig if we have Hive so can use SQL (again, three advantages)?**
2. Pig is relatively lightweight compared to Hive, Computationally efficient, when your query has a lot of join and filter, pig is more appropriate. Hive is more for analysis; generally prefer to use Hive to generate reports.
3. Pig can efficiently handle structured and unstructured data. However, Hive can efficiently handle structured data. Hive is more suitable for data warehouse tasks; Hive is mainly used for static structure and the need for frequent analysis of the work.
4. Pig gives developers more flexibility in the field of large data sets and allows the development of concise scripts for converting data streams for embedding into larger applications.