Q). Why Spark is preferred over Java Program for transferring large amounts of data?

Spark is a popular big data processing framework that is often used instead of Java programs to process large amounts of data for several reasons:

1. Distributed Processing: Spark is designed to handle distributed processing, which means that it can process large amounts of data by breaking it into smaller chunks and processing them on a cluster of computers. Java programs can also handle distributed processing, but it requires more code and management.
2. In-Memory Computing: Spark uses in-memory computing, which means that it stores data in memory rather than on disk. This allows Spark to process data much faster than Java programs, which have to read and write data to disk.
3. Ease of Use: Spark provides a user-friendly interface that makes it easy for developers to work with big data. It has a simpler programming model than Java, which can be complex for big data processing.
4. Scalability: Spark is highly scalable and can handle large amounts of data without compromising on performance. Java programs can also be scalable, but it requires more effort to design and implement.
5. Support for Multiple Languages: Spark supports multiple programming languages such as Java, Scala, and Python, which makes it easier for developers to work with the language they are most comfortable with.
6. Integration with Other Technologies: Spark integrates well with other big data technologies such as Hadoop, Cassandra, and Kafka. This makes it easier to build data processing pipelines that span multiple technologies.

Overall, Spark offers several advantages over Java programs for processing large-scale data, including faster processing speed, simpler programming model, and better scalability. While Java programs can also handle big data processing, Spark provides a more efficient and user-friendly solution.