**A discussion on how big data and storage issues could be relevant for your case, if you had big data sets, for instance.**

Let's begin by understanding the three characteristics of big data- volume, velocity, and variety. The volume concerns with the size of data, velocity refers to how fast the data is generated and processed, and variety talks about different data types/categories. Each of these characteristics brings its own set of challenges. If we had big data sets, then we would naturally think of the complexities because of big data and make a plan for it.

If we had big data sets, then one of the biggest concerns would be data storage infrastructure (storage). When we talk about data storage infrastructure, we must consider factors like data compression, data search, short-term storage, long-term storage and data movement. We can build such infrastructure right from scratch to meet our goal or go for a solution that is already available in the market.

If we decide to build such architecture by ourselves, we need to understand our data types, data volume, and velocity among other. The type of data and business goal will help to decide on what kind of database management system we need. Some of the possible database solutions we have are a relational database, NoSQL, and graph database. A proper data storage infrastructure can be quite expensive and difficult to maintain.

We could also use the open-source software framework like Hadoop as an alternative to building storage infrastructure from scratch. Hadoop is an open-source software framework for storing data and running applications on clusters of commodity hardware.

Apart from the data storage issue, there are other big data concerns like processing and accessibility. Processing is about applying algorithms that help us to make sense out of data. The challenging part is algorithms can't be made standard as each data set is unique. The processing of data has a direct impact on the performance of the system.

Accessibility deals with access and display of the data intended for the end-users. The biggest challenge in this area is how should data be displayed to the end-users. Since we are dealing with a large set of data set to decide which is the relevant data required a lot of effort in understanding the data and its attributes.

**A short reflection on ethical and legal issues that might be relevant for your business case**

We have used a data set that contains information related to sales. It holds information about the products and their sales each week. This data set may also include customer information, which could be used to study, for example, a customer's purchasing behavior.

Since we are handling customer's data we must protect their privacy and identity. We will follow the standard law and guidelines in practice to ensure that the data remains safe and will be used for specified, explicit and legitimate purposes. Besides, we need to inform customers about the use of their data when we acquire them. Besides, they will always have control over what information they want to share with us and they are allowed to withdraw their consent at any time. From our side, we will make sure that their data is processed fairly and in a transparent manner.