**Question 1: Research and describe the factors why Big Data has been attracting so much attention.**

There are many reasons why Big Data has been attracting attention nowadays. Here are some key factors contributing to its importance:

- **Technology Improvement & Data Growth:** According to Statista, the total amount of data created in 2025 is forecast to reach more than 180 zettabytes, showing that there are a lot of advantages that Big Data can bring and companies & organizations nowadays need to invest in Big Data to take advantage of it. In addition, the evolution of technical tools such as cloud storage, distributed computing and analytics tools allow people to easily access, store and process large data sets to improve many aspects of businesses and human life.
- **Competitive advantage:** Countries and organizations realize that the more data in their hand, the more competitive advantage over other competitors. Companies can make valuable data-driven decisions, improve efficiency, and gain a competitive edge in the market. They also can create new products & services, satisfying their customers and making breakthrough development worldwide. For example, ChatGPT is a leading innovation nowadays, thanks to the available data and data from Human feedback during its development. Furthermore, governments take use of Big Data in improving their public services, policymaking, military innovation, and overall governance. Countries with more efficient Big Data management methods will have more competitive advantages than other countries in economy, military, society, etc.
- **Societal Impact:** Big Data helps solve challenging problems in many industries such as Healthcare, Education, Environment, etc. It also creates new job opportunities as the demands on data experts increases.
- **The Thread of Big Data:** While mentioning the positive impacts of Big Data, its threads also make it attract attention. Data is created everyday and increased rapidly based on the increase of digital tools such as social media. Therefore, the privacy concerns (how to collect, store and analyze data ethically) are one of the biggest problems in Big Data management, including: avoid the misuse of personal information, protect the data from cyberattack, maintain data quality and accuracy, data ownership and control, etc.

**Question 2: Explain the V's of Big Data with your own words and provide examples for each of the characteristics.**

V's of Big Data mentioned to Volume, Velocity, Value, Veracity and Variety with these detailed explanation:

- **Volume:** This concept refers to the size of data being generated & collected or analyzed & processed, requiring distinct processing techniques which are more advanced than traditional data management systems. For example, bank A handles more than 2 million customer transactions (3 petabytes of data) every day, so 2 million & 3 petabytes are volumes of daily data at bank A.

- **Velocity:** This concept refers to the speed data is generated & collected or analyzed & processed. Velocity is very important because Big Data requires real-time or near-time processing to provide accurate insights, helping organizations in decision-making process. For example, a stock company captures every click of mouse on its trading platform, using high-frequency trading algorithms to process and analyze data in real-time to make accurate trading decisions.
- **Value:** This concept refers to the insights, information and pattern recognition which are discovered from Big Data (the statistical factors or correlations between features). Value is essential as it helps organizations make better data-driven decisions for positive changes and development. For example, company A uses the insights when analyzing sales data to predict yearly sales to adjust strategy in product development, marketing and human resource.
- **Veracity:** This concept refers to the trustworthiness and authenticity of the data as the data collecting process can create missing data or error. The accuracy of data is very important when it decides the effectiveness of data-driven decisions. For example, company A uses inaccurate data sets of drug tests so company's executives decide to put a defective product on the market, causing customer allergies and serious losses for the company.
- **Variety:** This concept refers to the diversity of data types such as unstructured, semi-structured, structured data. For example, customer reviews on an ecommerce platform are unstructured data while sale data on that platform are collected as spreadsheets which is considered as structured data with tables.

**Question 3: Elaborate on how different fields/industries can utilize big data technologies.**
The use of Big Data is variable as all industries/fields use it as a sharp knife to move forward:
- **Healthcare:** Big Data is used to analyze electronic health records (EHRs), historical data, medical research data, etc. for better disease prediction, personalized treatment, new methods for the treatment of fatal disease cure, hospital readmissions reduction, etc.
- **Retail & Ecommerce:** This industry can use Big Data to analyze customer behaviors, market trends, competitive environment, supply chain management, etc. to make better decisions in pricing, marketing, customer satisfaction, product R&D, sales, and human resource strategy.
- **Manufacturing:** Big Data is widely used in this industry to optimize production processes, quality control, product R&D, inventory management, etc.
- **Transportation and Supply Chain:** Big Data is used to optimize route planning, fleet operations, vehicle management and inventory management, helping companies in this industry save time, fuel, cost and vehicles.
- **Energy:** Big Data can help companies in this industry forecast demands, improve grid reliability and enhance efficiency of energy. It is also used in equipment maintenance and energy generation optimization to save operation cost.

- **Agriculture:** The use of Big Data in this industry is to analyze data of weather, season, soil, water, cattle, poultry, seeds factors to optimize irrigation fertilization, pest & livestock epidemic control and enable agriculture precision.
- **Public sector:** Governments take use of Big Data in improving their public services, policymaking, military innovation, and overall governance.
- **Financial services:** Financial services organizations take use of Big Data to analyze customer data, transaction data, trading data, etc. to predict account opening, detect fraud, prevent risk, and get real-time stock market insights, etc.
- **Technology:** Organizations in this industry take use of Big Data to create innovation with AI and Machine Learning.
- **Education:** Big Data serves academic research to provide more knowledge for teaching. In addition, big data also helps improve job search opportunities by helping students better understand the job market from analyzed insights.

## Question 4: Explain GPU vs. CPU and GPU's relationship with Big Data.

Central Processing Unit (CPU) is the control center of a computer which is used to handle a wide variety of tasks and has a few cores optimized for sequential processing that require complex executing instructions. Graphic Processing Unit (GPU) is a specialized processor, which is used to handle graphic-related works such as rendering graphics, and image processing and has smaller cores optimized for parallel processing and high-throughput computation.

Because GPU is used for tasks that are able to broken down, GPUs can be used in many aspects of Big Data, such as:
- **Data Analytics:** GPU decides the speed in data processing (sorting, filtering and aggregation) by leveraging the parallel processing capabilities. In addition, GPU relates to graphic-related works, helping data visualization process smoother and faster . Therefore, GPU helps improve the productivity of data analytics.
- **Machine Learning and Deep Learning:** GPU performs parallel computations which help reduce time in training complex ML models and increase the accuracy of these models.

## Question 5: Research and describe the relationship between AI and Big Data.

What is Artificial Intelligence (AI)? According to John McCarthy ([in 2004 paper](#)), "It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable." The definition of AI mentions the understanding of human intelligence to relate similar tasks, meaning that the training process requires a large amount of data to help AI models perform better over time. In detail, the relation can be explained in two primary ideas:

- **The impacts of Big Data on AI:** As the definition of AI, AI algorithms need data to recognize the patterns and make predictions. All the AI innovations rely on the data which help models accurate over time, perform human tasks better.
- **The impact of AI on Big Data:** While AI relies on Big Data, Big Data also benefits from AI. AI-powered tools and techniques such as NLP, CNN, predictive analytics are used to extract information, insights for the data collection process. The development of AI enriches the data worldwide and makes data analytics more easier to access.

In summary, the relation between AI and Big Data is mutual growth and development based on each other. For example, in the finance sector, AI helps detect fraud in transactions, ensuring bank security. The algorithms to detect are based on training data from previous fraud transactions. When applying AI, data is also automatically collected for training to improve algorithms and make AI innovation in fraud detection more efficient and accurate.

**Question 6: Explain the steps to deploy a Big Data solution.**
The steps to deploy a Big Data solution depends on the specific business requirements and industries. In general, the process includes these primary steps:
- **Define objectives and requirements:** Before deploying a Big Data solution, the identification of objectives and requirements is very important as all the methods, processes and results are based on this step.
- **Choose the method:** Based on the objectives and requirements, we can choose the most suitable Big Data techniques and platforms. For example, bank A requires to process and analyze real-time transaction data, the tools bank A should use are Apache Kafka and Apache Flink as they can handle large streams of data in real-time. In another example, supply chain department at company B needs to use Big Data for reporting, the tool to use is Google BigQuery as it is a serverless data warehouse that supports SQL queries.
- **Design the architecture:** This step mentions the model creation to represent the storage, processing, and analysis progresses. This step also maps out the data quality management strategy to maintain data privacy.
- **Develop and test the solution (processing, analytics and validation):** Use the chosen method to implement analytics and processing steps by building the required big data components. At the same time, the quality checks are deployed to validate the solution performance. This step prevents security and other issues that help algorithms and solutions be more accurate and safe.
- **Apply the solution:** Perform data ingestion and data integration from other sources and check the data quality. Then, run through the testing process to validate the solution when applying on entire data to avoid errors.
- **Visualization and reporting:** Extracting the results after deploying the solution with the right platform such as Microsoft Excel, Power BI or SQL Server Reporting Services.

- **Maintain, optimize and evolve the solution:** Set up the maintenance process to ensure the solution is still fit with the organization's objectives and error-free. Make sure that the solution is scalable to meet the increased demand when handling larger data.

**Question 7: Elaborate on ethical issues regarding Big Data (e.g., security, privacy).**

Big Data 's impacts on today's world is undeniable, even so, the ethical issues surrounding Big Data are controversial and somehow organizations and governments are finding ways to address them to protect privacy rights and avoid security vulnerabilities.

- **The privacy violation and data misuse:** Big Data is applied to help industries innovate, for example, in the marketing industry, individual web browser data was collected to help companies make better recommendations on digital, intended to encourage customers to buy more products these companies advertise. However, some advertisements are created to attack the vulnerabilities and fears of customers which hurt them. Therefore, these targeted ads are used intently for bad purposes which harm people. This example is one of many cases in which privacy is violated and fallen into wrong hands. However, in some countries where the laws stand on the side of profit and power, not of the people, resulting in negative impacts on people.
- **Data security:** Even when data is collected, handled properly and ethically, it is important to keep this data secure. For example, Bank A exposing customer account information leads to thousands of fraud transactions and loss of customer savings. The disclosure of this information not only seriously affects the bank but also the customer when their personal information is leaked to the outside and in the wrong hands. Therefore, ensuring information security is extremely important and also in the ethical issues surrounding Big Data.
- **The algorithmic bias:** *Artea (D): Discrimination through Algorithmic Bias in Targeting* - a Harvard case study discusses this ethical issue when pointing out unfair treatment when targeting customers based on biased algorithms they create in marketing analytics campaigns, which leads to the potential discrimination in Artea's customer targeting efforts, therefore, customers are treated unfairly based on their race, gender, or age. The case study reminds that companies nowadays have to be aware of maintaining the trust in data-driven processes and evaluate algorithms to keep it error-free and more importantly consider the bias as an important ethical issue which can harm the company and customers.

Addressing ethical issues regarding Big Data is important as its impact on the rights of people and the sustainable development of any organizations. It requires policymakers, executives, and experts to establish and follow rules and regulations strictly to develop in the benefits of both organization and people.