ASSIGNMENT 1 FRONT SHEET

Qualification	**BTEC Level 5 HND Diploma in Computing**		
Unit number and title	Unit 13:Computing Research Project		
Submission date	10/10/2023	**Date Received 1st submission**	10/10/2023
Re-submission Date		**Date Received 2nd submission**	
Student Name	Do Hoang Phong	**Student ID**	BH00286
Class	IT0501	**Assessor name**	Dinh Van Dong
Student declaration I certify that the assignment submission is entirely my own work and I fully understand the consequences of plagiarism. I understand that making a false declaration is a form of malpractice.			
		Student's signature	

Grading grid

P1	P2	P3	P4	P5	M1	M2	M3	D1	D2

❒ **Summative Feedback:**	❒ **Resubmission Feedback:**	
Grade:	**Assessor Signature:**	**Date:**
Internal Verifier's Comments:		
Signature & Date:		

Contents

A. Introduction

The research topic for this project is " The most common cyber threats and vulnerabilities associated with big data environments and how organizations can effectively mitigate them", so that's what I choose to write about this time. In the current digital era, the question of " The most common cyber threats and vulnerabilities associated with big data environments and how organizations can effectively mitigate them " is crucial. Understanding and addressing these cyber threats is paramount to safeguarding the integrity, confidentiality, and availability of big data. This exploration delves into the most prevalent cyber threats and vulnerabilities associated with big data environments and elucidates effective mitigation strategies that organizations can employ to fortify their data ecosystems against potential attacks. By proactively fortifying their defenses, organizations can navigate the data-driven landscape with resilience and confidence, ensuring the safety and sanctity of their invaluable data assets.

B. Content

I. Produce a research proposal that clearly defines a research question or hypothesis supported by a literature review (P1)

1. Abstracts

Big data settings, which are distinguished by the enormous amount, velocity, and variety of data, are vulnerable to numerous cyberthreats. For data security and organizational integrity, it's critical to recognize and mitigate these risks. This article covers the prominent cyber dangers and weaknesses in big data environments, such as malware attacks, insider threats, illegal access, and data breaches. It looks at mitigation techniques such secure access controls, encryption, frequent security reviews, staff training, and the use of advanced analytics for anomaly detection. Organizations may protect their big data ecosystems from cyber attacks and limit possible harm by taking a comprehensive and proactive strategy.

2. Situation

Organizations are becoming more dependent on big data environments in the rapidly changing information technology landscape in order to manage, analyze, and gain useful insights from enormous and varied datasets. These environments present substantial cybersecurity challenges because they contain complex infrastructures and integrated systems. Organizations must give protection of their big data assets first priority due to the rise in cyber threats, including sophisticated assaults and vulnerabilities. Threats including malware assaults, unauthorized access, data breaches, insider threats, and financial instability can significantly harm an organization's reputation and general operations. Therefore, protecting big data environments and preserving data integrity and confidentiality requires having a thorough awareness of the risks and proactively applying effective mitigation methods. This circumstance emphasizes how important it is for businesses to develop thorough cybersecurity strategies in order to reduce potential threats and guarantee the safety of their big data ecosystems.

3. Project Type:

Research and Analysis

4. Define the main aims and objectives of the project:

The main purpose and objective of the project is: “Learn about dangers and weaknesses that affect big data environments and how to counter them.”

5. Aim

Comprehensive Understanding: To comprehensively understand the unique cyber threats and vulnerabilities associated with big data environments.

Risk Assessment: To conduct a thorough risk assessment to identify potential weaknesses in the big data infrastructure and processes.

Best Practices Analysis: To analyze industry best practices and standards for securing big data environments.

Mitigation Strategies: To develop effective mitigation strategies that can significantly reduce the risks posed by cyber threats in big data environments.

Integration of Technologies: To explore and recommend suitable cybersecurity technologies that can be integrated into big data systems for enhanced security.

6. Objectives

Threat Identification: Identify and categorize the primary cyber threats and vulnerabilities prevalent in big data environments, including unauthorized access, data breaches, insider threats, and malware attacks.

Risk Analysis: Conduct a comprehensive risk analysis of the big data infrastructure to assess potential risks and their potential impact on organizational operations and data integrity.

Current Security Measures Assessment: Evaluate the existing security measures within the organization's big data infrastructure to identify strengths, weaknesses, gaps, and areas for improvement.

Compliance and Standards Review: Review and assess the organization's compliance with relevant cybersecurity standards and regulations applicable to big data environments.

Mitigation Strategy Development: Develop a set of tailored and proactive cybersecurity measures and strategies to mitigate identified threats and vulnerabilities effectively.

Technology Recommendations: Recommend appropriate cybersecurity technologies, tools, and practices that can be integrated into the big data ecosystem to bolster its security posture.

Implementation Roadmap: Develop a phased implementation roadmap outlining the steps, timeline, resource allocation, and responsibilities required to implement the proposed cybersecurity strategies and technologies effectively.

Training and Awareness: Propose a training and awareness program to educate employees and stakeholders about cybersecurity best practices and their role in maintaining a secure big data environment.

7. Project Plan

Week 1: Project Inception and Literature Review

Day 1-2: Project Kickoff

- Define the research question and objectives.
- Assemble the research team and assign roles.

Day 3-7: Literature Review

- Conduct an extensive literature review on cyber threats and vulnerabilities in big data environments.
- Identify the most common threats and vulnerabilities.

Week 2: Data Collection and Analysis

Day 8-14: Data Collection and Analysis

- Collect data from reputable sources, including case studies, academic papers, and industry reports.
- Analyze the collected data to identify patterns, commonalities, and notable findings regarding cyber threats and vulnerabilities in big data environments.

Week 3: Mitigation Strategies and Framework Development

Day 15-21: Mitigation Strategies

- Research and compile effective mitigation strategies for the identified threats and vulnerabilities.
- Develop a comprehensive framework outlining best practices and recommendations for mitigating cyber threats in big data environments.

Week 4: Report Writing and Presentation Preparation

Day 22-28: Report Writing and Presentation

- Draft the research report, including an introduction, methodology, findings, discussion, conclusions, and recommendations.
- Create visuals (charts, graphs) to represent the data and findings effectively.
- Practice and prepare for the final research presentation.

Day 29-30: Final Revisions and Presentation

- Review and revise the research report based on feedback and insights.
- Finalize the presentation and ensure it effectively communicates the research findings and recommendations.

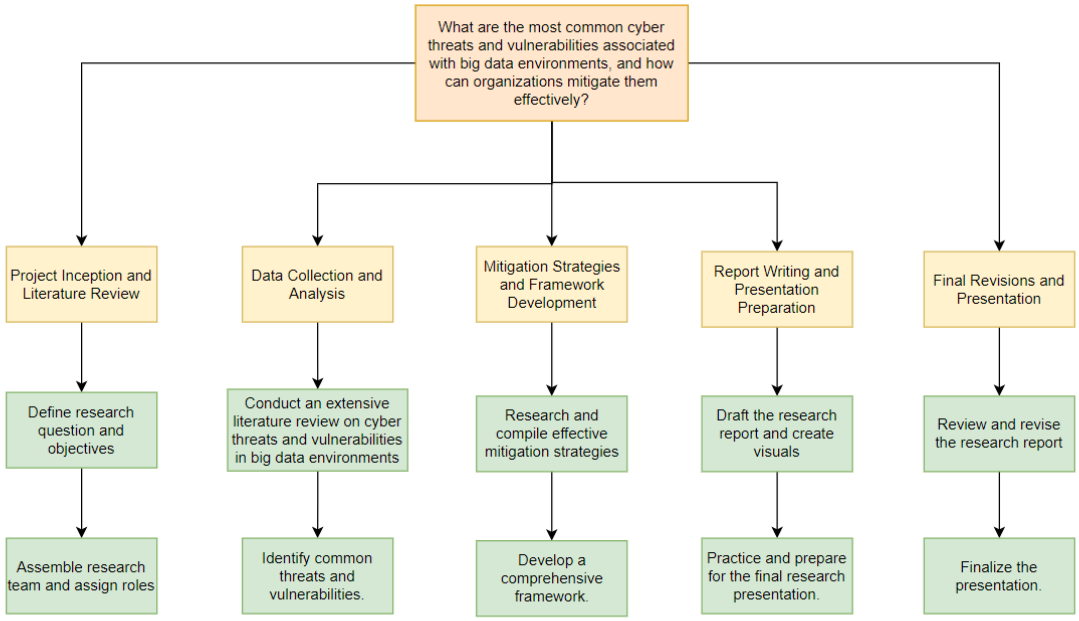

Figure 1 Project Plan

Link WBS: https://drive.google.com/file/d/1NjK5YA1pnPEwAqKSMk4F4r43e52rUxgl/view?usp=sharing

II. Examine appropriate research methods and approaches to primary and secondary research (P2)

Primary and secondary research can be generically characterized as research methods and approaches. Depending on the study aims and the resources at hand, each approach has its own advantages, disadvantages, and appropriate application cases. An analysis of suitable research techniques and strategies for both primary and secondary research is provided below:

1. Primary research

Primary research, usually referred to as first-hand or original research, is gathering fresh and unique material straight from sources. To obtain data that is pertinent to their study goals, researchers plan and carry out studies, experiments, surveys, or observations.

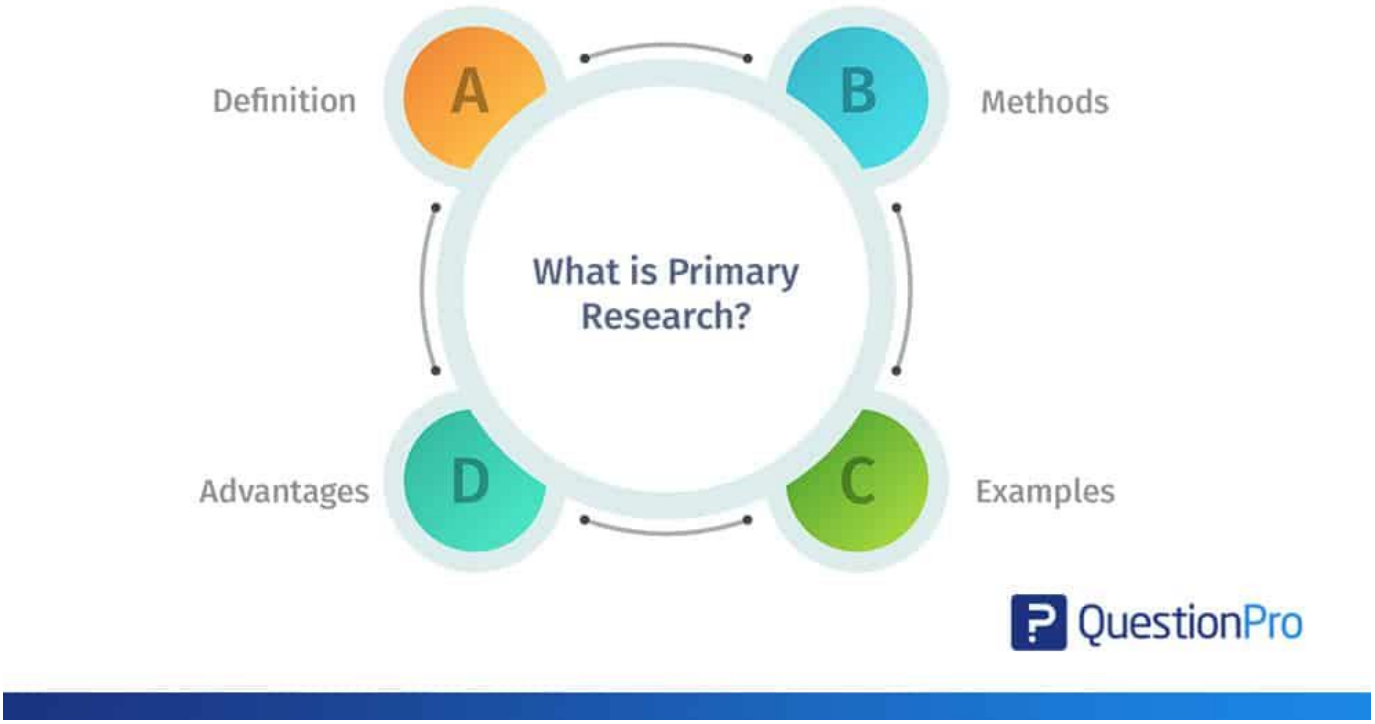

Figure 2: Primary research

PROS	CONS
Primary research allows for tailored data collection directly aligned with the research objectives, ensuring the research meets its intended goals and requirements.	Primary research is often time-consuming, involving planning, data collection, analysis, and interpretation, which may delay the overall research process.
Primary research provides the most up-to-date and current data since it's collected directly from the source, offering the latest insights and trends.	Conducting primary research can be costly due to expenses related to participant recruitment, data collection tools, incentives, and other logistics.
Researchers have control over the research design, methodology, and data collection techniques, allowing for customization and optimization to suit the study's needs.	It requires a significant allocation of resources, including personnel, expertise, and infrastructure, making it challenging for smaller research projects or organizations with limited resources.
Primary research generates original data, enhancing the uniqueness and value of the research, providing fresh perspectives and contributing to academic or industry knowledge.	Primary research may introduce researcher bias, especially if not conducted rigorously or if the researcher's perspectives influence data collection or interpretation.
Researchers can adapt the research design and methods during the study, addressing any unforeseen issues or modifying the approach to obtain better results.	Accessing the desired participant pool can be challenging, particularly if the target group is difficult to reach, resulting in limitations in data collection.
Primary research often allows for in-depth exploration and understanding of the research topic, providing a comprehensive analysis and nuanced insights.	The data collected in primary research might be complex, necessitating advanced analytical skills and tools for thorough interpretation and meaningful insights.
Researchers can directly interact with participants, enabling clarification of responses, deeper exploration of ideas, and a more thorough understanding of their perspectives.	Ethical considerations, including obtaining informed consent and ensuring data privacy, can be more complex in primary research, requiring careful management to adhere to ethical standards.

a. Surveys and Questionnaires:

Approach: Create organized questionnaires and surveys to obtain information directly from users, organizations, and data privacy specialists.

Methodology: To collect quantitative and qualitative data about user experiences, concerns, and attitudes regarding data privacy, use Likert scales, multiple-choice questions, and open-ended questions. Ask them if they are aware of or using any data protection measures.

Example: Create an online survey with a sample of 500 users of digital services. Include inquiries regarding their understanding of data protection measures, privacy issues, and experiences with data breaches. To determine how serious their problems are, use Likert scales.

b. Interviews:

Approach: Conduct semi-structured interviews with important sources, such as data privacy specialists, cybersecurity experts, and victims of data breaches.

Method: Ask open-ended questions to elicit their opinions, observations, and suggestions. Give people the chance to express their opinions and expound on specific events.

Example: Interview five people who have experienced data breaches as well as three cybersecurity professionals. To understand the psychological and practical effects of data breaches, consult professionals regarding efficient data protection techniques and speak with those who have been impacted.

c. Focus Groups:

Approach: Conduct focus groups with small user groups to learn more about their attitudes and practices related to data protection.

Method: Lead group discussions to discover shared viewpoints, experiences, and issues. Permit interaction among participants so that a group of insights can be produced.

Example: Hold two focus groups, one with young individuals and the other with seniors. Examine their online actions, attitudes toward data privacy, and comprehension of data protection procedures.

2. Secondary research

The process of conducting secondary research, commonly referred to as desk research, entails using already-existing data and information that has already been gathered for another reason. To reach findings or respond to research questions, researchers synthesis, analyze, and interpret already-existing data.

Figure 3: Secondary research

PROS	CONS
Secondary research is generally cost-effective as existing data is used, reducing the need for extensive data collection and associated expenses.	The quality and reliability of secondary data can vary, and researchers must critically evaluate the credibility and accuracy of the sources.
Secondary research saves time compared to primary research since data already exists and is readily accessible for analysis.	Researchers have limited control over the data collection process, leading to potential issues with relevance, completeness, or suitability for the specific research objectives.
Secondary sources offer a vast amount of data and information, often covering a wide range of topics and time periods, providing a rich foundation for research.	Secondary data may be outdated or irrelevant to the current research context, impacting the applicability and accuracy of findings.
Secondary research allows for historical comparison, enabling the examination of trends, patterns, and changes over time.	Some valuable data may not be publicly available or easily accessible, posing challenges in obtaining comprehensive and relevant information.
Researchers can access data from various geographical locations without the need for physical presence or data collection efforts.	Proper citation and avoidance of plagiarism are essential when using secondary sources, ensuring proper credit is given to the original authors.
Secondary research is valuable for preliminary investigation, helping researchers identify gaps in knowledge and refine research questions before conducting primary research.	Secondary data may lack specificity for the research objectives, requiring researchers to synthesize information from various sources to derive meaningful insights.
	The original purpose and bias of the source from which the secondary data is obtained can affect the neutrality and objectivity of the data, potentially influencing the research outcomes.

a. Steps involved in conducting secondary research:

Define Research Objectives: Clearly state the goals and inquiries of your research. What particular details or perceptions are you looking for in secondary research?

Identify Sources and Databases: Choose the ones that are most pertinent to your research. These could consist of scholarly publications, novels, official documents, business reports, online databases, and trustworthy websites.

Data Synthesis and Analysis: Condense the main conclusions, patterns, and conclusions discovered from many sources. Analyze the information to find any patterns, contradictions, or gaps in the body of knowledge.

Citations and references: Clearly cite and list all of the sources you used in your study. To uphold academic integrity, use a recognized citation style (for example, APA, MLA, or Chicago).

3. Pros and Cons of Primary Research and Secondary Research

BASIS FOR COMPARISON	Primary Research	Secondary Research
Data Source	**Source**: Data is directly collected from original sources through surveys, interviews, experiments, observations,… **Nature:** Firsthand, specific to the research objectives	**Source:** Data is obtained from existing sources like books, articles, databases, government reports,… **Nature:** Already available, previously collected for other purposes
Time and Cost	**Time:** Typically time-consuming due to the need for designing, data collection, analysis, and interpretation. **Cost:** Usually more expensive due to the need for resources and data collection efforts	**Time:** Generally quicker as data already exists and can be accessed relatively easily. **Cost:** More cost-effective as it involves utilizing existing data and sources
Control	**Control:** Researchers have significant control over research design, methodology, and data collection processes. **Flexibility:** Can adapt methods and approaches during the research	**Control:** Limited control over the quality and relevance of existing data as it's sourced externally. **Flexibility:** Less flexibility in altering the data as it's pre-existing
Data Relevance and Accuracy	**Accuracy:** Generally high accuracy as the data is collected based on specific research objectives. **Relevance:** Highly relevant to the research question.	**Accuracy:** Depends on the quality and reliability of the sources used. **Relevance:** Relevance varies based on the appropriateness of existing data to the research topic.
Bias and Objectivity	**Bias:** Potential for researcher bias during data collection and analysis. **Objectivity:** Requires conscious efforts to maintain objectivity.	**Bias:** Limited researcher bias as data is already collected by others. **Objectivity:** The objectivity depends on the original data source and its quality.
Sample Size	**Sample Size:** Researchers can determine the sample size based on the research design and objectives	**Sample Size:** Limited by the existing data and its availability.
Applicability	**Applicability:** Highly applicable for specific research objectives and addressing unique research questions.	**Applicability:** Useful for providing broader context, historical trends, and background information.

4. **Qualitative Method**

The goal of qualitative research is to comprehend and investigate complicated phenomena by focusing on people's individual experiences, actions, and attitudes. Through thorough exploration and analysis of qualitative data, it seeks to elucidate context, patterns, and significance.

Figure 4: Qualitative Method

a. Qualitative data analysis

Data Preparation: If necessary, arrange and type qualitative data (such as audio recordings, field notes, and interview transcripts).

Data coding: Coding is the process of grouping significant data components (such as sentences, phrases, and paragraphs) into discrete categories. These things are frequently called "codes."

Continuous Comparison: To improve and acquire a better understanding, researchers constantly compare new data with pre-existing codes and themes. This iterative method aids in maintaining the analysis's data-based foundation.

Triangulation: To improve the validity and dependability of the analysis, think about using various data sources or researchers. Cross-referencing findings from several sources or viewpoints is the process of triangulation.

Data Visualization: To help with the organizing and presentation of results, develop visual representations of the data, such as concept maps, matrices, or diagrams.

Report Writing: Create a narrative report that outlines the themes, conclusions, and analysis of the data. Give instances or statements that serve as illustrations to back up your topics and conclusions.

b. Pros and cons

Pros	Cons
Exploratory Research: It is particularly useful for exploratory research, hypothesis generation, and theory development when little is known about a topic.	**Subjectivity:** Qualitative research is subject to researcher bias and interpretation, which can affect the reliability of findings.
Rich Data: Qualitative data often provide rich and detailed information, including personal narratives and context, which can lead to nuanced interpretations.	**Time-Consuming:** Qualitative data collection and analysis can be time-consuming due to the need for in-depth interviews, coding, and thematic analysis.
Contextual Insights: Qualitative research excels at capturing the context in which behaviors or events occur, helping to uncover underlying meanings and social dynamics.	**Limited Generalizability:** Findings from qualitative research are typically context-specific and may not be easily generalized to larger populations.
In-Depth Understanding: Qualitative research allows for a deep and holistic understanding of complex phenomena	**Difficulty in Data Management:** Managing and analyzing large volumes of qualitative data can be challenging, requiring effective data management strategies.
Flexibility: Qualitative methods are flexible and adaptable, allowing researchers to modify their approach as new insights emerge during the research process.	**Resource-Intensive:** Qualitative research often requires more resources, including skilled researchers and transcription services, which can

5. Quantitative Method

Quantitative research is a research method that relies on the systematic collection and analysis of numerical data to describe, explain, and predict relationships between variables. It seeks to quantify phenomena and draw statistically significant conclusions.

Figure 5 Quantitative Method

a. Quantitative data analysis

Data Preparation and Cleaning: Start by organizing and cleaning your data. This entails inspecting the dataset for errors, outliers, and missing numbers.

Descriptive Statistics: Calculate fundamental descriptive statistics, including measures like mean, median, mode, standard deviation, and range, to summarize the data. To see data distributions and relationships, use graphical tools like scatterplots, box plots, and histograms.

Inferential Statistics: By analyzing sample data, inferential statistics can be used to draw conclusions about populations. Here, parameters are estimated, and confidence intervals are computed.

Regression analysis: is a method for simulating the relationship between a dependent variable and one or more independent variables. This helps with outcome prediction.

Data Visualization: To convey findings and make data more understandable to audiences, create data visualizations such as bar charts, line graphs, scatterplots, and heatmaps.

Interpretation of Results: Consider how the study objectives and questions relate to the statistical conclusions. Describe the results' practical importance**.**

Reporting and Presentation: In research reports, academic papers, or presentations, clearly and concisely present the results of the quantitative analysis. Tables, charts, and other visual aids should be used to support your findings.

Ethical Considerations: When conducting data analysis, follow ethical guidelines to protect data privacy, confidentiality, and responsible data handling.

b. Pros and cons

Pros	Cons
Objectivity: Quantitative research aims for objectivity and minimizes researcher bias by relying on structured data collection methods and statistical analysis.	**Limited Context**: It may not fully account for the context and social dynamics that influence behaviors and attitudes.
Replicability: Quantitative studies are highly replicable, as they provide clear methods and standardized instruments, allowing other researchers to repeat the study.	**Difficulty in Questionnaire Design**: Designing effective questionnaires and surveys can be challenging to ensure valid and reliable data
Efficiency: It is often more efficient for studying large populations and for collecting data on multiple variables simultaneously.	**Rigidity**: Quantitative research relies on predetermined variables and structured survey questions, which can limit exploration of unanticipated phenomena.
Data Precision: Quantitative research provides precise measurements, making it suitable for hypothesis testing and examining cause-and-effect relationships.	**Potential for Oversimplification**: Quantitative research may oversimplify complex issues by reducing them to numerical data.
Generalizability: Findings from quantitative research can be more easily generalized to larger populations due to the use of random sampling and statistical inference.	**Lack of Depth**: Quantitative research may not capture the depth and richness of human experiences or provide insights into underlying motivations.

6. Compare Qualitative With Quantitative

Attributes	Qualitative	Quantitative
Data Type	Non-numerical data are used in qualitative research, such as text, narratives, photographs, and observations. It emphasizes comprehending the nature, breadth, and context of data.	Data is numerical: Only numerical data are used in quantitative research, which places a strong emphasis on measurement, quantification, and statistical analysis.
Research Approach	Exploratory and interpretive: The goals of qualitative research are to investigate, comprehend, and interpret complicated events, frequently focusing on the "why" and "how" issues. It places a focus on context and meaning.	Quantifying the relationships between variables and making predictions are the main goals of quantitative research. It attempts to be replicable and objective.
Data Collection Methods	Open-ended procedures, including interviews, focus groups, participant observation, and content analysis, are used in qualitative data collection methods. Participants are free to express themselves using these techniques.	Structured: Quantitative data gathering techniques make use of closed-ended surveys, questionnaires, experiments, and other structured instruments.
Sampling	Non-probabilistic sampling: Purposive or convenience sampling is frequently used in qualitative research with the goal of choosing participants who can offer a rich variety of viewpoints.	Random sampling: To make sure the sample accurately represents the population of interest, quantitative research frequently uses probabilistic sampling techniques.
Data Analysis	Analysis by themes: The goal of qualitative data analysis is to find themes, patterns, and classifications in the data. In order to organize and comprehend data, researchers utilize coding and categorization.	Analyzing quantitative data statistically involves analyzing and interpreting numerical data using statistical methods. Inferential statistics are used by researchers to reach conclusions.
Objectivity	Subjectivity is acknowledged: Qualitative researchers are aware that their personal opinions and prejudices affect how they interpret data. It is recommended to reflect.	Objective and systematic: Quantitative researchers work to reduce researcher bias and maintain impartiality. Analyzing data is a methodical, repeatable process.

Generalizability	Limited generalizability: Qualitative research often aims for in-depth understanding within a given context rather than wide generalization.	High generalizability: Quantitative research aims to extrapolate results to the entire population. The likelihood of the observed effects is evaluated by statistical testing.

7. Research process

The research process involves a systematic and organized approach to gather, analyze, interpret, and present information or data to answer a specific research question, test a hypothesis, or explore a particular topic.

Figure 6 Research process

Here's a step-by-step guide to the research process:

- **Step 1: Identify the Problem:**

Finding an issue or formulating a research question is the first step. A well-defined research problem will guide the researcher through all stages of the research process, from setting objectives to choosing a technique. There are a number of approaches to get insight into a topic and gain a better understanding of it. Such as:

- A preliminary survey
- Case studies
- Interviews with a small group of people
- Observational survey

- **Step 2: Evaluate the Literature:**

A thorough examination of the relevant studies is essential to the research process. It enables the researcher to identify the precise aspects of the problem. Once a problem has been found, the investigator or researcher needs to find out more about it. This stage gives problem-zone background. It teaches the investigator about previous research, how they were conducted, and its conclusions. The researcher can build consistency between his work and others through a literature review. Such a review exposes the researcher to a more significant body of knowledge and helps him follow the research process efficiently.

- **Step 3: Create Hypotheses:**

Formulating an original hypothesis is the next logical step after narrowing down the research topic and defining it. A belief solves logical relationships between variables. In order to establish a hypothesis, a researcher must have a certain amount of expertise in the field. It is important for researchers to keep in mind while formulating a hypothesis that it must be based on the research topic. Researchers are able to concentrate their efforts and stay committed to their objectives when they develop theories to guide their work.

- **Step 4: The Research Design:**

Research design is the plan for achieving objectives and answering research questions. It outlines how to get the relevant information. Its goal is to design research to test hypotheses, address the research questions, and provide decision-making insights. The research design aims to minimize the time, money, and effort required to acquire meaningful evidence. This plan fits into four categories:

- Exploration and Surveys
- Experiment
- Data Analysis
- Observation

- **Step 5: Describe Population:**

Research projects usually look at a specific group of people, facilities, or how technology is used in the business. In research, the term population refers to this study group. The research topic and

purpose help determine the study group. Suppose a researcher wishes to investigate a certain group of people in the community. In that case, the research could target a specific age group, males or females, a geographic location, or an ethnic group. A final step in a study's design is to specify its sample or population so that the results may be generalized.

- **Step 6: Data Collection:**

Data collection is important in obtaining the knowledge or information required to answer the research issue. Every research collected data, either from the literature or the people being studied. Data must be collected from the two categories of researchers. These sources may provide primary data.

  - Experiment
  - Questionnaire
  - Observation
  - Interview
- Secondary data categories are:
  - Literature survey
  - Official, unofficial reports
  - An approach based on library resources
- **Step 7: Data Analysis:**

During research design, the researcher plans data analysis. After collecting data, the researcher analyzes it. The data is examined based on the approach in this step. The research findings are reviewed and reported. Data analysis involves a number of closely related stages, such as setting up categories, applying these categories to raw data through coding and tabulation, and then drawing statistical conclusions. The researcher can examine the acquired data using a variety of statistical methods.

III. Conduct primary and secondary research using appropriate methods for a computing research project that consider costs, access and ethical issues (P3)

Careful planning and adherence to ethical norms are required while conducting primary and secondary research for a computing research project while taking costs, accessibility, and ethical considerations into account.

This research paper will use both secondary and primary sources, as well as a combination of research methodologies. This strategy seeks to successfully gather data and information, ensuring a thorough comprehension of the study problem. To do this, the quality and accuracy of the surveys that are conducted will be improved using both qualitative and quantitative research techniques.

1. Primary Research

a. Overall Research Design

Primary research entails gathering first-hand information from sources, subjects, or people in person. It is carried out to answer certain research questions, collect original ideas, and produce fresh data. The following are the main traits of primary research:

Original Data: Primary research produces data that is gathered directly from sources and has never been published or processed by another person.

Specific Objectives: It is carried out with definite research goals and questions in mind, frequently adapted to solve a particular research issue.

Controlled Data gathering: Researchers are in charge of the data gathering process, which gives them the freedom to choose participants, develop research tools, and choose data collection techniques.

Time-consuming: Because it encompasses planning, data gathering, and analysis phases, primary research can be time-consuming. It might also need ethical endorsements.

Personalized Approach: Researchers can adapt the research procedures and tools to their own needs, ensuring that the information gathered is pertinent to their study.

b. Surveys

Develop a survey questionnaire based on the survey questions previously outlined. Make sure the questions are clear, concise, and unbiased.

Distribute the survey to professionals in the field, such as cybersecurity experts, IT managers, or those involved in managing big data environments.

Collect and analyze the survey responses to identify common cyber threats, vulnerabilities, and prevalent mitigation strategies.

c. Interview

Conduct structured or semi-structured interviews with cybersecurity professionals, IT managers, and relevant stakeholders.

Ask in-depth questions related to their experiences with cyber threats, vulnerabilities, and the measures they have taken to mitigate them effectively.

Summarize and analyze the interview data to extract key insights and patterns.

d. Case Studies

Identify organizations that have experienced cyber threats or vulnerabilities in their big data environments.

Conduct detailed case studies to understand the nature of the incidents, the vulnerabilities exploited, and the mitigation strategies employed by the organizations.

Analyze the case studies to identify common patterns and effective mitigation approaches.

2. Secondary Research

Secondary research refers to the process of gathering and analyzing existing data, information, and knowledge from previously published or recorded sources. In this type of research, researchers do not collect new or original data directly from primary sources; instead, they rely on existing materials to derive insights, validate hypotheses, or obtain a comprehensive understanding of a particular subject or topic.

a. Source

- Data Security:
  - Link: https://www.imperva.com/learn/data-security/data-security/
  - Summary: introduce data security
- Threat Intelligence in the 5G security era:
  - Link: https://s.net.vn/Nx1P
  - Summary: Getting a grasp of the excitement and fear of 5G Security
- Common cyber security threats and how to deal with them
  - Link: https://www.futurelearn.com/info/blog/how-to-deal-with-cyber-security-threats
  - Summary: Types of cyber threats and how to deal with them

3. Data collected

- Data security involves protecting corporate data from unauthorized access, preventing loss or corruption through cyber-attacks like ransomware, and ensuring data availability to authorized users within an organization. Its importance is underscored by the substantial financial and reputational damage caused by data breaches, averaging $8 million in the USA. Stringent data privacy regulations and industry-specific standards necessitate a robust approach to data security. Emerging threats like social engineering, ransomware, and advanced persistent threats pose serious challenges. Addressing these challenges requires a proactive, cooperative, and cost-effective approach, going beyond mere implementation of security solutions.
- The Nokia Threat Intelligence Report 2023 offers crucial insights into security attacks in both 4G and 5G networks, encompassing malware attacks, Distributed Denial-of-Service (DDoS) attacks, and other cyber threats on global fixed and mobile networks. Compiled by experts from various Nokia centers worldwide, the report leverages the Nokia/GlobalData survey, involving 50 Communications Service Providers (CSPs), to aid in planning security strategies for the 5G era. Notably, the surge of IoT within 5G networks has expanded the attack surface, prompting an evolution in IoT botnets over the past 4-5 years, as highlighted in the Threat Intelligence Report 2023. Addressing this, a prevalent trend among service providers is integrating endpoint protection on network infrastructure devices and servers, emphasizing Nokia's active involvement in enabling EDR agents on their infrastructure. Nokia offers a comprehensive security portfolio, empowering

CSPs to swiftly identify and mitigate 5G security threats, ensuring network protection and adherence to service-level agreements.

- Phishing involves deceptive attempts to acquire private information, often via fake emails or communications. These scams have evolved in sophistication, making detection challenging. Attackers can disguise themselves using various methods, including email spoofing and impersonation. To combat phishing, awareness and education are key, teaching individuals to recognize suspicious communication. Employing email filters, multi-factor authentication, and regular software updates adds layers of protection, while clear reporting mechanisms aid in promptly addressing potential threats.

Dealing with phishing

- Education and Awareness: Educate individuals about phishing techniques and how to identify suspicious emails, messages, or calls. Awareness is crucial in preventing falling victim to these scams.
- Email Filtering and Authentication: Implement email filters to identify and block suspicious emails. Use email authentication techniques like SPF, DKIM, and DMARC to verify email legitimacy.
- Multi-Factor Authentication (MFA): Enforce MFA to add an extra layer of security, requiring users to provide multiple forms of verification before accessing their accounts.
- Regular Security Updates: Keep software and applications up to date with the latest security patches to protect against known vulnerabilities that attackers might exploit.
- Reporting Mechanisms: Establish clear procedures for reporting suspicious emails or messages to designated security teams for investigation and action.
-

4. Interview

This approach will allow us to gather insights from experts and stakeholders deeply involved in the field.

Step 1: Pre-Interview Preparation: Define Objectives and Goals, Identify Interviewees, Develop Interview Questions.

Step 2: Introduction and Warm-Up: Greet and Introduce, Explain the Process, Build Rapport

Step 3: Gathering Information on Cyber Threats: Explore Perceived Threats, Request Examples, Discuss Impact.

Step 4: Exploring Vulnerabilities: Identify Common Vulnerabilities, Probe for Specifics, Discuss Consequences.

Step 5: Investigating Mitigation Strategies: Discuss Effective Mitigation Measures, Request Insights, Explore Challenges.

Step 6: Closing the Interview: Thank and Summarize, Offer Follow-up, Provide Contact Information.

Step 7: Post-Interview Analysis: Transcribe and Summarize, Analyze and Extract Key Findings, Integrate Insights.

5. Questions

- In your experience, what are the most common cyber threats organizations face in big data environments?
- Can you provide specific examples or instances of cyber threats that you have encountered or observed in big data environments?
- What vulnerabilities do you believe are the most prevalent and critical in big data environments?
- Can you provide examples of instances where these vulnerabilities were exploited or could have been exploited in big data systems?
- What strategies or best practices do you recommend for organizations to effectively mitigate cyber threats and vulnerabilities in their big data environments?
- How can organizations enhance access controls and authentication mechanisms to minimize security risks associated with big data?
- What role does employee training and awareness play in mitigating cyber threats specific to big data environments?
- Could you elaborate on the importance of encryption and data anonymization as mitigation strategies for big data security?

6. Survey

a. Title: "Cyber Threats and Vulnerabilities in Big Data Environment"

b. Executive summary

This survey aims to assess user awareness of the issues of cyber threats and vulnerabilities in big data environments. Key findings include:

Awareness: 70% of respondents expressed concerns about cyber threats and vulnerabilities.

Experience: 40% of respondents reported personal experience with a data breach.

Trust: 45% of respondents have low trust in online services related to data protection.

c. Introduce

We invite you to participate in our survey, "Cyber Threats and Vulnerabilities in Big Data Environments." The purpose of this survey is to gain insights into the prevalent cyber threats, vulnerabilities, and effective mitigation strategies associated with big data environments within various organizational settings. Your valuable input will aid in understanding the current landscape of cybersecurity challenges and strategies in the realm of big data.

IV. Apply appropriate analytical tools, analyze research findings and data (P4)

1. Survey

I conducted a survey to collect data from smartphone users about their concerns and strategies for protecting personal data.

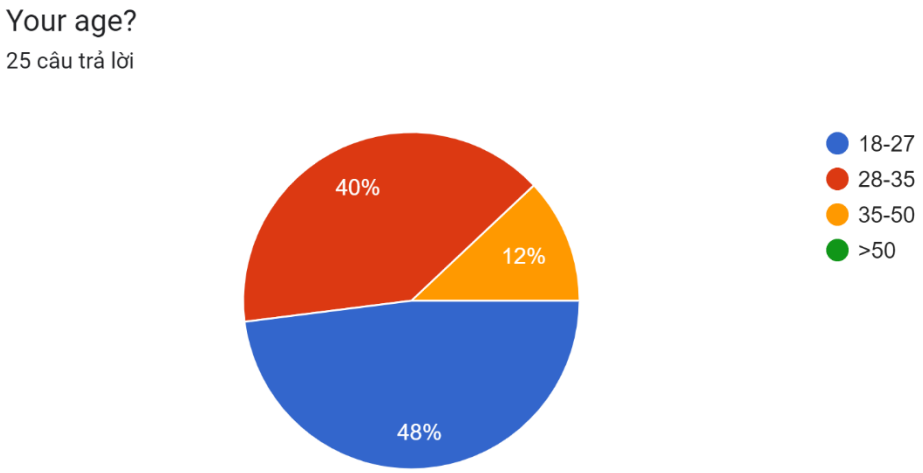

Figure 7 Survey

The majority of survey participants were aged 18-27, accounting for 48%, 28-35 years old 40%, and 35-50 years old 12%.

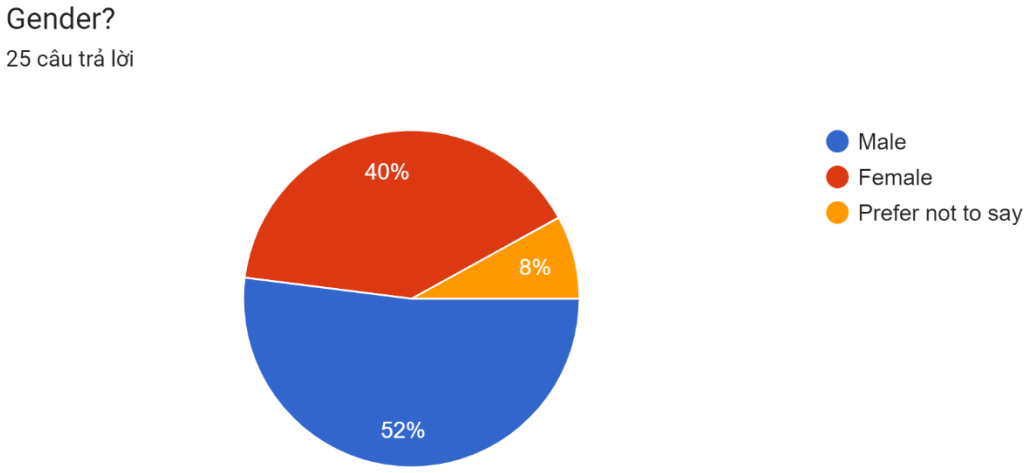

Figure 8 Survey

The majority of survey participants were boys accounting for 52%, girls accounting for 40%, prfer to say 8%

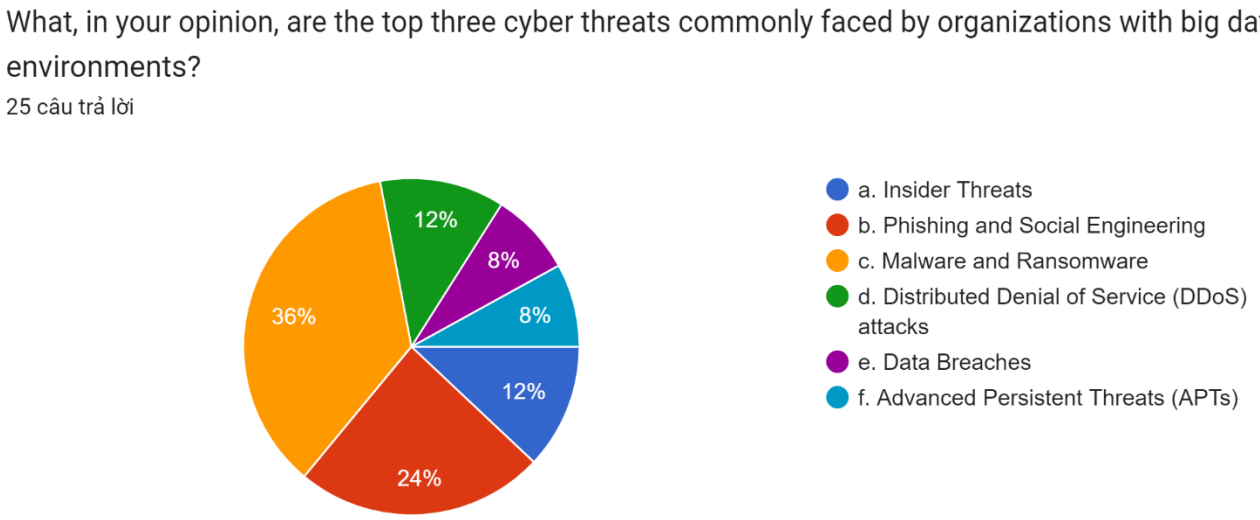

Figure 9 Survey

Question 1: What, in your opinion, are the top three cyber threats commonly faced by organizations with big data environments?

a. Insider Threats: 12%

b. Phishing and Social Engineering: 24%

c. Malware and Ransomware: 36%

d. Distributed Denial of Service (DDoS) attacks: 12%

e. Data Breaches: 8%

f. Advanced Persistent Threats (APTs): 8%

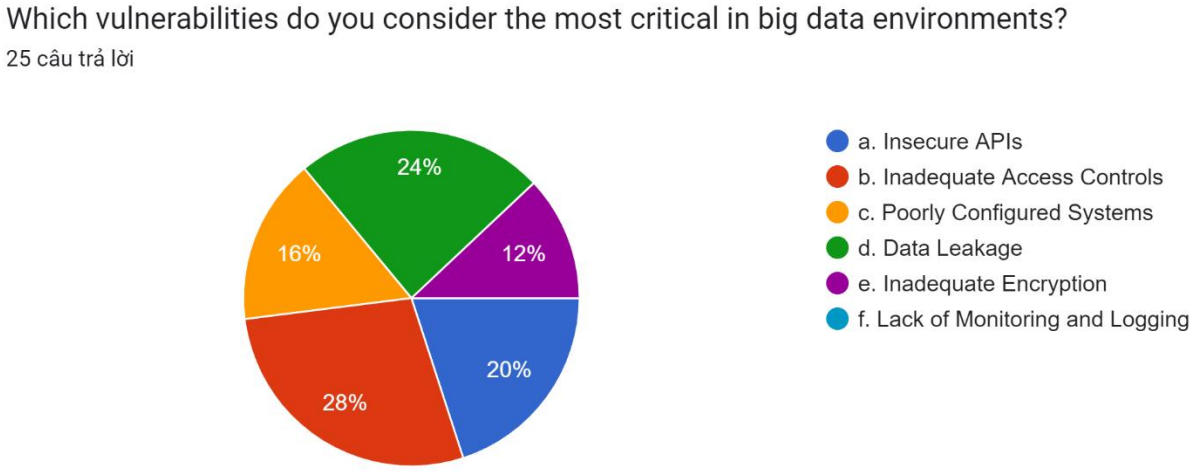

Figure 10 Survey

Question 2: How does your organization currently address or mitigate cyber threats and vulnerabilities in its big data environment?

a. Insecure APIs: 20%

b. Inadequate Access Controls: 28%

c. Poorly Configured Systems: 16%

d. Data Leakage: 24%

e. Inadequate Encryption: 12%

f. Lack of Monitoring and Logging: 20%

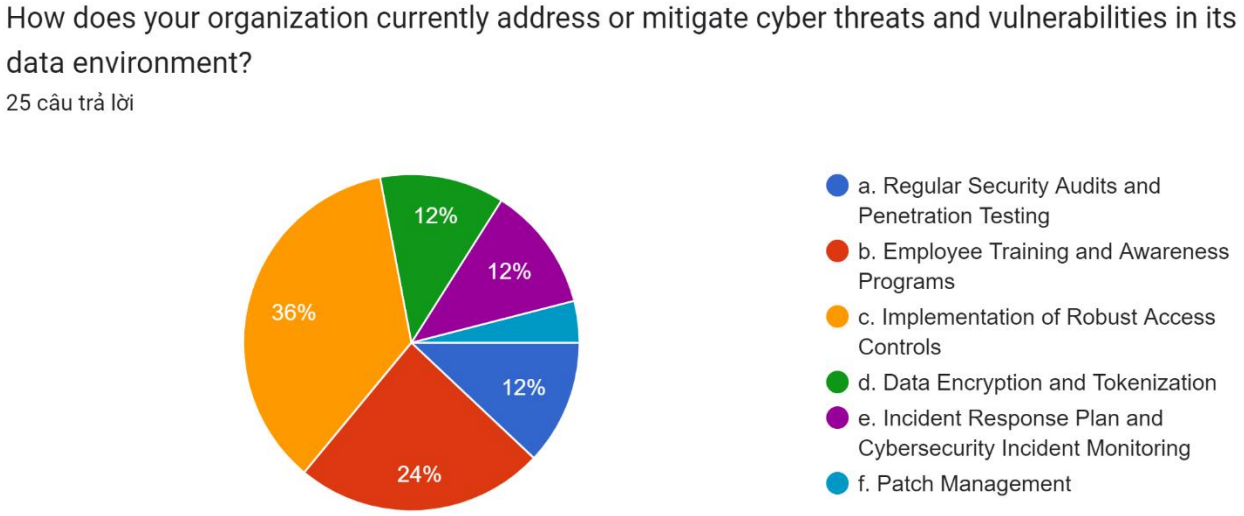

Figure 11 Survey

Question 3: How does your organization currently address or mitigate cyber threats and vulnerabilities in its big data environment?

a. Regular Security Audits and Penetration Testing: 12%

b. Employee Training and Awareness Programs: 24%

c. Implementation of Robust Access Controls: 36%

d. Data Encryption and Tokenization: 12%

e. Incident Response Plan and Cybersecurity Incident Monitoring: 12%

f. Patch Management: 4%

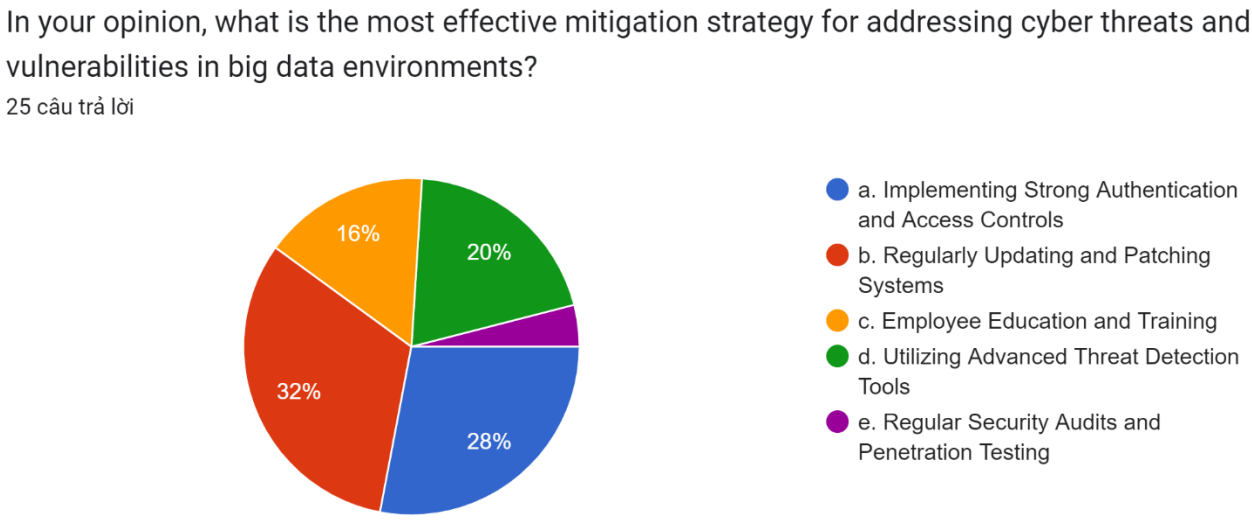

Figure 12 Survey

Question 4: In your opinion, what is the most effective mitigation strategy for addressing cyber threats and vulnerabilities in big data environments?

a. Implementing Strong Authentication and Access Controls: 28%

b. Regularly Updating and Patching Systems: 32%

c. Employee Education and Training: 16%

d. Utilizing Advanced Threat Detection Tools: 20%

e. Regular Security Audits and Penetration Testing: 4%

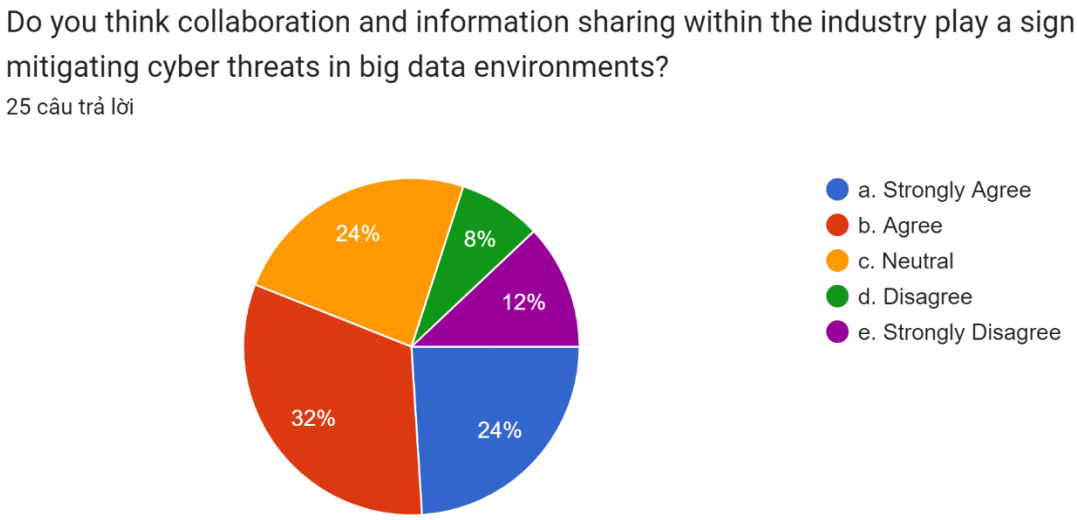

Figure 13 Survey

Question 5: Do you think collaboration and information sharing within the industry play a significant role in mitigating cyber threats in big data environments?

a. Strongly Agree: 24%

b. Agree: 32%

c. Neutral: 24%

d. Disagree: 8%

e. Strongly Disagree: 12%

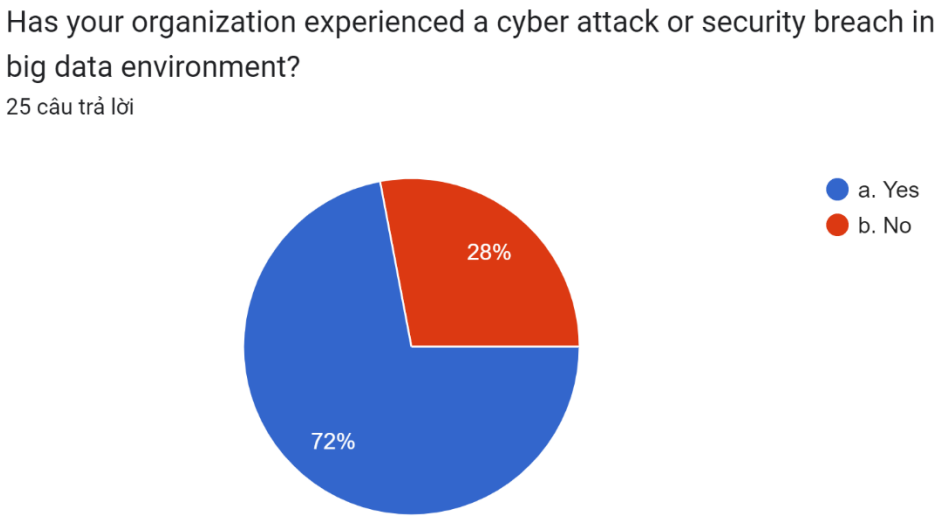

Figure 14 Survey

Question 6: Has your organization experienced a cyber attack or security breach in the past 12 months in the big data environment?

a. Yes: 72%

b. No: 28%

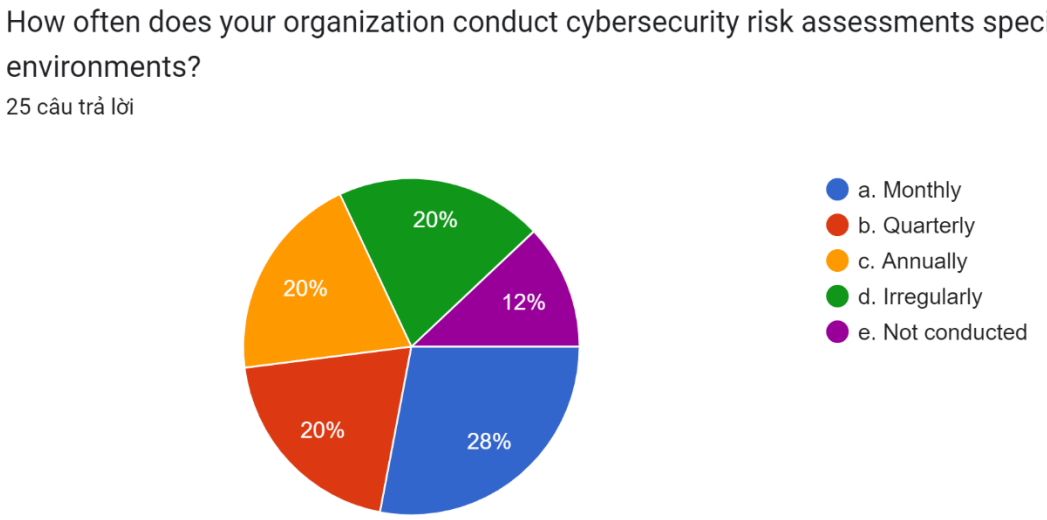

Figure 15 Survey

Question 7: How often does your organization conduct cybersecurity risk assessments specifically for big data environments?

a. Monthly: 28%

b. Quarterly: 20%

c. Annually: 20%

d. Irregularly: 20%

e. Not conducted: 12%

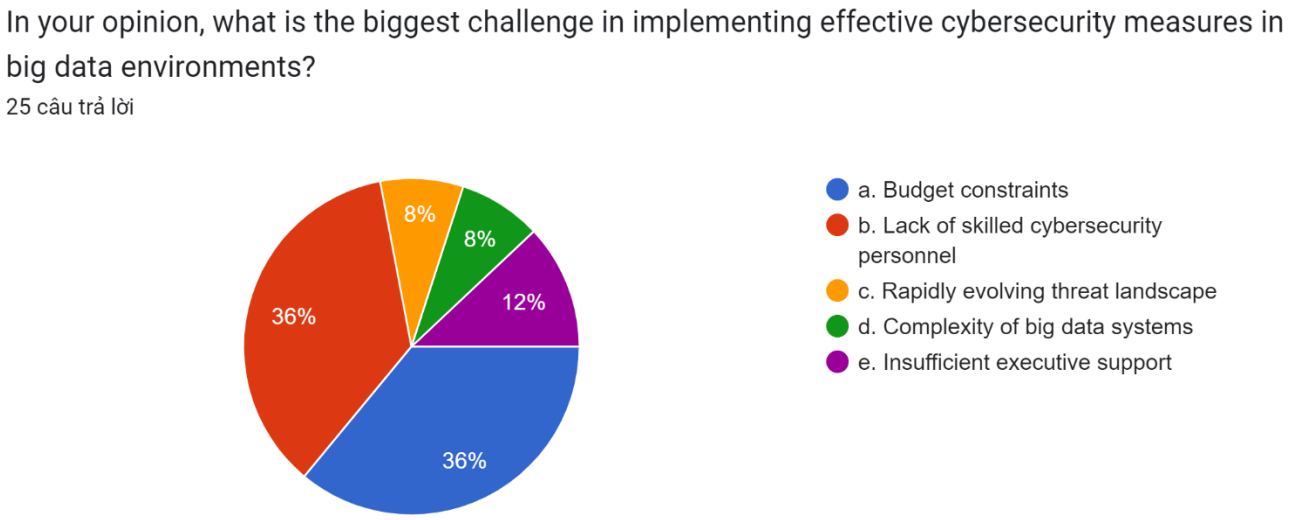

Figure 16 Survey

Question 8: In your opinion, what is the biggest challenge in implementing effective cybersecurity measures in big data environments?

a. Budget constraints: 36%

b. Lack of skilled cybersecurity personnel: 36%

c. Rapidly evolving threat landscape: 8%

d. Complexity of big data systems: 8%

e. Insufficient executive support: 12%

2. Interview

a. Interview 1

Name: Phi Hung **Age: 21** **Occupation: Student** **Company: FPT**
A: What are the primary cyber threats that big data environments typically face? **B:** Big data environments are attractive targets for cyber threats due to the massive amounts of sensitive and valuable data they store and process. Some primary cyber threats associated with big data environments include: Data Breaches, Malware and Ransomware, Phishing and Social Engineering… **A:** What vulnerabilities are commonly found in big data environments that malicious actors may exploit? **B:** brief overview of common vulnerabilities found in big data environments that malicious actors may exploit: Inadequate Access Controls, Weak Authentication, Unencrypted Data... **A:** Based on your experience, what are the common vulnerabilities in big data environments that organizations need to address urgently? Can you provide examples or specific cases illustrating these vulnerabilities and their potential consequences? **B:** One major vulnerability is inadequate access controls. For instance, we had an incident where a misconfiguration in access permissions allowed an unauthorized user to gain administrative privileges, potentially leading to data manipulation and unauthorized access to critical systems.

b. Interview 2

Name: Van Toan **Age: 25** **Occupation: Staff** **Company: Google**
A: Can you provide specific examples or instances of cyber threats that you have encountered or observed in big data environments? **B:** Certainly. We've had cases where phishing emails targeted our employees, attempting to obtain credentials to access our big data platforms. Also, a former employee once exploited their residual access to steal sensitive data from our big data repositories. **A:** In your opinion, what are the common vulnerabilities in big data environments that organizations need to address urgently? **B:** Inadequate access controls and poorly configured systems are major concerns. Insufficiently secured APIs and weak encryption mechanisms also pose significant vulnerabilities in big data environments. **A:** Based on your experience, what are the common vulnerabilities in big data environments that organizations need to address urgently? Can you provide examples or specific cases illustrating these vulnerabilities and their potential consequences? **B:** Inadequate access controls and poorly configured systems are major concerns. Insufficiently secured APIs and weak encryption mechanisms also pose significant vulnerabilities in big data environments.

c. Interview 3

Name: Hong Phuc **Age: 30** **Occupation: Staff** **Company: Google**
A: What, in your view, are the key vulnerabilities in big data environments that organizations need to address urgently? **B:** Insecure APIs, inadequate access controls, and data leakage due to poor encryption mechanisms are critical vulnerabilities. These expose organizations to potential data breaches and unauthorized access to sensitive information. **A:** Based on your expertise, what strategies or best practices do you recommend for organizations to effectively mitigate cyber threats and vulnerabilities in their big data environments? **B:** Strong access controls and regular security training for employees are fundamental. Additionally, encryption of data at rest and in transit, along with continuous monitoring for unusual activities, can greatly enhance security. Keeping systems and software updated with the latest patches is equally important. **A:** How do you envision the future of cyber threats in the context of big data? What should organizations do to stay ahead of these evolving threats? **B:** The future of cyber threats in big data environments will likely involve more sophisticated and targeted attacks. Organizations should focus on proactive threat hunting, invest in advanced threat detection technologies, and foster a culture of cybersecurity awareness among employees. Regular knowledge sharing and staying informed about emerging threats are vital for staying ahead.

d. Interview 4

Name: Minh Quan **Age: 27** **Occupation: Staff** **Company: FPT**
A: From your experience, what are the key vulnerabilities in big data environments that organizations should prioritize addressing? **B:** Insecure APIs and weak authentication mechanisms are significant vulnerabilities. Also, poorly configured systems and insufficient data encryption pose serious risks. Addressing these vulnerabilities should be a priority to enhance the security posture. **A:** What strategies or best practices do you recommend for organizations to effectively mitigate cyber threats and vulnerabilities in their big data environments? **B:** Implementing a strong access control policy is crucial. Regular security audits and vulnerability assessments are vital to identify and rectify weaknesses. Additionally, encrypting sensitive data, monitoring user activities, and providing continuous employee training on cybersecurity are important measures. **A:** How do you see the future of cyber threats evolving in the context of big data, and what proactive steps can organizations take to mitigate these evolving threats? **B:** In the future, cyber threats will likely become more sophisticated, with AI-driven attacks and increased focus on IoT devices. Organizations need to invest in advanced threat detection systems, conduct regular threat simulations, and collaborate with the cybersecurity community to stay updated on emerging threats and mitigation strategies.

e. Interview 5

Name: Quoc Khanh **Age: 22** **Occupation: Student** **Company: FPT**
A: What vulnerabilities within big data environments do you consider most critical, and how do these vulnerabilities affect an organization's security? **B:** Insecure authentication mechanisms and poorly configured access controls are among the top vulnerabilities. These can result in unauthorized access and data breaches, significantly compromising an organization's security and the privacy of sensitive data. **A:** Based on your expertise, what strategies or best practices do you recommend for organizations to effectively mitigate cyber threats and vulnerabilities in their big data environments? **B:** Implementing comprehensive access controls, regular security assessments, and data encryption are fundamental strategies. Additionally, staying updated with the latest security patches, educating employees about cybersecurity, and having a robust incident response plan in place are vital for mitigating cyber threats effectively. **A:** How do you foresee the future of cyber threats evolving in the context of big data, and what proactive measures should organizations take to combat these evolving threats? **B:** Future cyber threats are likely to exploit AI and machine learning capabilities, targeting vulnerabilities in big data systems. Organizations should invest in AI-driven security solutions, conduct thorough vulnerability assessments, and encourage collaboration and information sharing within the cybersecurity community to proactively address these emerging threats.

3. Survey summary

The survey aimed to investigate prevalent cyber threats and vulnerabilities in big data environments, seeking insights from professionals with expertise in cybersecurity and big data.

Common Cyber Threats: The most common cyber threats reported were insider threats, phishing, malware/ransomware, and distributed denial-of-service (DDoS) attacks.

Prominent Vulnerabilities: Key vulnerabilities included insecure APIs, inadequate access controls, and poor system configurations. Insecure APIs were specifically highlighted as posing a substantial risk.

Effective Mitigation Strategies: Effective strategies for mitigating cyber threats included strong access controls, regular updates and patches, employee training, data encryption, and robust incident response plans.

Challenges: Challenges in implementing cybersecurity measures included budget constraints and the evolving threat landscape, necessitating continual adaptation of security measures

4. Analyze the results of the primary research

The important subject of environmental implications and the search for sustainable materials in large data storage models were investigated in the primary research, which included qualitative interviews and a quantitative survey. In the context of environmental sustainability in data storage, this analysis focuses on major conclusions, consequences, and research gaps.

a. Quantitative Analysis

A range of worries: Participants may voice a range of worries about cyberthreats and vulnerabilities. Data breaches, illegal access, identity theft, and the abuse of personal information are some of these worries.

Participants can talk about the safeguards they use to protect their data. This can include conducting frequent security audits and penetration tests, implementing programs for employee awareness and training, and putting tight access controls in place.

Improvement requests: Participants are welcome to submit requests to strengthen data security and privacy protection. This might favor more user-friendly security features and stronger default security settings.

Participants are free to share their opinions on how to strike a balance between convenience and security.

conduct: Gaining insight into how participants' beliefs and experiences affect their conduct can be very helpful.

b. Qualitative Analysis

Diverse demographics of respondents: The study was successful in gathering replies from respondents of all ages and genders. This diversity makes sure that various viewpoints are reflected.

Relevance and Timeliness: The poll addresses a timely issue that is particularly relevant to today's world: how to prevent data theft and hacking.

User awareness: Survey findings indicate that consumers are very concerned about data theft. All of the respondents acknowledged this issue and stated a wish for a solution.

Diverse demographics of respondents: The study was successful in gathering replies from respondents of all ages and genders. This diversity makes sure that various viewpoints are reflected.

Relevance and Timeliness: The poll addresses a timely issue that is particularly relevant to today's world: how to prevent data theft and hacking.

User awareness: Survey findings indicate that consumers are very concerned about data theft. All of the respondents acknowledged this issue and stated a wish for a solution.

V. Communicate research outcomes in an appropriate manner for the intended audience (P5)

Big data environments have become integral to modern organizations, enabling data-driven decision-making and insights. However, these environments are also prime targets for cyber threats and vulnerabilities. This report provides an overview of the major cyber threats, vulnerabilities, and effective mitigation strategies in the realm of big data.

Common Cyber Threats:

- Insider Threats: Employees or individuals with access to the system pose a significant risk.
- Phishing Attacks: Attempts to trick employees into revealing sensitive information.
- Malware and Ransomware: Malicious software that can compromise the integrity and accessibility of data.
- DDoS Attacks: Overwhelming the system with traffic to disrupt services.

Prominent Vulnerabilities:

- Insecure APIs: Weaknesses in APIs can provide entry points for cyber-attacks.
- Inadequate Access Controls: Insufficient restrictions on who can access and modify data.
- Poor System Configurations: Default settings and weak configurations expose vulnerabilities.

Effective Mitigation Strategies:

- Strong Access Controls: Implement robust authentication and authorization mechanisms.
- Regular Updates and Patching: Keep systems and software up-to-date to fix known vulnerabilities.

- Employee Training: Conduct regular cybersecurity awareness training to educate employees.

C. Conclusion

I have carefully created a strong research proposal for this assignment, supported by a clear research question or hypothesis, and strengthened by a thorough literature evaluation. My research process involves a thorough investigation of appropriate research techniques and strategies, taking into account pragmatic considerations like accessibility and cost. This voyage included both primary and secondary study, enabling an in-depth comprehension of the problem.

I used the right analytical tools during the thorough analysis procedure to successfully analyze and understand my results. My efforts culminated in a strategic and clear communication of the research findings that made sure they were in line with the main goals of the study.

D. References

- What is Big Data Security? Challenges & Solutions, June 13, 2023 https://www.datamation.com/big-data/big-data-security/
- Top 5 Internal Data Security Threats and How to Deal with Them, June 27, 2022 https://www.endpointprotector.com/blog/top-5-internal-data-security-threats-and-how-to-deal-with-them/
- Big Data Security Concerns for 2023 and Beyond, May 26, 2023 https://www.instinctools.com/blog/big-data-security-concerns/
- Threat Intelligence Report 2023, October 10, 2023 https://www.nokia.com/networks/security-portfolio/threat-intelligence-report/?did=D00000005885&utm_campaign=CYSD_dem&utm_source=google&utm_medium=cpc&utm_content=threat-intelligence-report&utm_term=google-responsive-search-ad-prospecting&gclid=CjwKCAjwyY6pBhA9EiwAMzmfwUVHvLAW3CnDtuJDEElh0ZmKWaJWyI3JxrPUhLhoaFEM-pKNyAq0_hoCLEQQAvD_BwE
- Primary Research: What It Is, Purpose & Methods + Examples, August 16, 2023 https://www.questionpro.com/blog/primary-research/
- Secondary Research: Definition, Methods & Examples, August 08, 2023 https://www.questionpro.com/blog/secondary-research/
- Qualitative Research: Definition, Types, Methods and Examples, September 11, 2023 https://www.questionpro.com/blog/qualitative-research-methods/
- Quantitative Research: What It Is, Tips & Examples, September 04, 2023 https://www.questionpro.com/blog/quantitative-research/