UNIVERSITY OF ALBERTA

Master of Science in Internetworking
Edmonton, Canada

Project Title:

Research on IoT Threats & Implementation of AI/ML to Address Emerging Cybersecurity Issues in IoT with Cloud Computing

Presented by:
Glory Dhayanidhi

Supervisor:
Juned Noonari

Fall 2021 - Winter 2022

DECLARATION

Glory Dhayanidhi solemnly declares that the project "**Research on IoT threats and implementation of AI/ML to address emerging cybersecurity issues in IoT with cloud computing"** in the Department of Computing Science, Master's in internetworking, University of Alberta, is based on my work. I guarantee that this report has not been published at any other university.

ACKNOWLEDGMENT

This project's success and outcome required a lot of guidance and assistance from many people, and I am incredibly fortunate to have got this all along with the completion of my project work. It is obligatory to record gratitude to them.

I heartily thank Mr. Juned Noonari for his constant guidance and suggestions, which made me complete the project on time.

I owe my profound gratitude to our Program director Dr. Mike MacGregor who gave me this excellent opportunity to experience this project.

ABSTRACT

Internet of Things (IoT) has become one of the progressive innovations and inviting space of interest for the research world and financially captivating for the business world. Integrating different devices and associating devices with humans requires artificial intelligence/ machine learning (AI/ML) to break down the data stored in the cloud framework. These IoT devices utilize their unique identifiers and the embedded sensor with every device to impart to one another and trade data among them using the web and cloud-based network infrastructure. We live in the period of big data where the need to apply AI/ML has been fundamental to the cycle to analyze the gathered cloud-based big data quickly and precisely. IoT security concerns incorporate significant difficulties and vulnerabilities such as cyber-attacks, data fraud, remote access, and hacking. Notwithstanding, although AI is now assuming a more important part in working on conventional cyber security, both cloud vulnerability and networking of IoT devices are substantial threats.

Inadequately secured IoT devices pose the risk of utilization by DDoS (Distributed Denial of Service) attacks. These attacks expose security vulnerabilities and cause service disruption that negatively affects consumers and business productivity. In addition, the vast majority of remotely connected IoT devices conveyed on a public organization are likewise under ongoing cyber threats. It is impossible to rely on humans to control the network, so AI and ML are beneficial and necessary to our future networks. IoT security needs AI/ML as a security tool to combat the challenges mentioned above. This research proposes a hybrid detection model as a solution by using AI/ML in a cloud computing environment both at the host-based and network level. AI/ML can improve the security services to elevate them to an advanced level.

The proposed method of this research approaches the security issues in cloud network infrastructure by creating predictive analysis to prevent future attacks. Subsequently, this research uses 3GPP, MIMO, and datasets taken from CIADA and Packt to create the right AI/ML application. The application of AI components has been critical and taken advantage of, conveying a more feasible Intrusion Detection System (IDS) and Intrusion Prevention System (IPS) abilities. Due to its highly adaptable cyber-physical system nature, AI comprises many interconnected devices, so data movement and analytics occur in a complex-wide area network. This research uses and applies required datasets, client information, and endpoint log records for host base detection and network data for a network-level location to execute the algorithm. This research approach resolves the security issues by learning, analyzing, and identifying cyber attacks so our AI/ML solutions outfit with the history of attacks, recognize future attacks, and walled protection from zero-day attacks.

The expected outcome of this research is as follows.

1. The research work follows the techniques of AI, ML, IoT, cloud computing to tackle cybersecurity issues.
2. Software implementation of the proposed method, how AI/ML is used to combat emerging cybersecurity issues.

3. Predictive analysis using AI/ML to prevent future attacks.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 What is IoT?

The Internet of Things is a system where billions of embedded devices are connected to the Internet to collect and exchange data. The Internet of Things (IoT) means a global infrastructure for the information society, enabling progressed benefits by interconnecting network devices based on advancing interoperable information and communication technologies (ICT) [1].

Internet connection gives us all sorts of benefits that were not possible before. Benefits that compound each other can be put into three categories collecting and sending information, receiving and acting on the data, and doing both (Goal of an IoT system). The fundamental characteristics of IoT are:

1) Interconnectivity - Devices can be interconnected with the global information and communication infrastructure
2) Heterogeneity - Devices can interact with each other or service platforms based on different hardware platforms
3) Networks and dynamic changes - The number of devices and state of devices change dynamically [2].

The IoT systems integrate four distinct components

1. sensors/devices
2. connectivity
3. data processing
4. user interface

It is referred to as sensors or devices because multiple sensors can be tied together. Sensors are found everywhere in our workplaces, shopping centers, homes, hospitals. Sensors not only detect but also respond to changes in the environment. Inputs can come from light, temperature, motion, and pressure sources. For example, a cell phone has numerous sensors such as a camera, accelerometer, and GPS. Sensors can perform in various ways and provides valuable information when connected to a network. Sensors share data with management systems and other connected devices. The information transferred to the cloud requires a path to reach. Sensors are of various shapes and sizes. Sensors are interconnected to the cloud through WiFi, cellular, satellite, and Bluetooth. Each method has trade-offs between power consumption, range, and bandwidth.

IoT sensors critically improve operational efficiency, enhance workers' safety, and reduce costs. The Industrial Internet of things brings the utilization of sensors to a new level. Once the data reaches the cloud (data processor), the software processes it. For example, checking if the temperature reading is within an acceptable range or checking the intruders on a property on video. Alternatively, it could also be very complex, such as capturing video to identify intruders on a property. When there is an intruder on a property?

A user comes in at this point. Further, information is made applicable to the end-user via an alert (as an email, text, or notification). For instance, a user can check the video feeds on various properties via a phone app or a web browser. The user may also act and affect the system. Moreover, some actions perform automatically. Rather than making a call to alert about an intruder, the IoT system could also automatically notify security teams or relevant authorities via predefined rules. A user interface allows users to input or check in on the system. The system sends the user's adjustments or actions in the opposite direction. An IoT system works at a high level from the user interface to the cloud and back to the sensors/devices to change [2].

1.2 Evolution of IoT

The Internet has been in a continuous state of evolution since its inception. ARPANET is a US military intelligence network created in the 1960s. It was used for the development of various intelligence programs. The evolution of the Internet of Things (IoT) has been widely celebrated, with its ability to enhance the living experience by enabling various devices and services. RFID has become the main component of the Internet of Things, including various devices such as Amazon Alexa, Google Assistant, and Apple Siri. The initial hype about RFID as the next generation of tech has stalled. In 2015, the authors proposed that RFID has the tracking capabilities needed to identify objects in real-time. The concept of RFID as the founding technology of the Internet of Things was not well-received.

During the pre-Internet era, people commonly used SMS to communicate. The World Wide Web, which is not yet synonymous with the Internet, emerged in March 1989 as a virtual platform where files are shared between computers. By the mid-1990s, the web had become a significant communication infrastructure. In March 1989, the World Wide Web (WWW), which is not inseparable from the Internet, arose as a virtual stage that pre-owned hypertext moves convention (HTTP) to move records on the web. By the mid-1990s, the 'www,' generally known as the 'web,' developed as a significant correspondence framework during the website agitation.

The web was utilized for the most part for email interchanges or for obtaining logically rich data. The early development of the web was a tremendous assortment of hypertext markup language (HTML) archives that were hyperlinked (electronically associated) through three sorts of conventions: HTML, HTTP, and uniform asset finder (URL). HTML is the language in which website pages are written, while HTTP is the most well-known protocol, grown explicitly for the World Wide Web and known for its ease of use and speed. The URL is the location where the electronic document dwells on the web. Web pages were made when hypertext was joined with the Internet, giving the internet browser its name, the World Wide Web. From 2003 to 2004, JavaScript paved the way for the web to be less static, and Web 2.0 was created, known as the more interactive and dynamic variant of its predecessor. Before long, social media was recognized as a bunch of Internet-based applications shown by the well-known affordances of Web 2.0, specifically Skype (2003), Facebook (2004), YouTube (2005), and Twitter (2006).

Finally, the global economy entered another wave of development with the advent of machine-to-machine (M2M) communication that paved IoT's new period of insight. IoT saw many potential devices and applications equipped with identification, tracking, monitoring, sensing, metering, automation, and processing capabilities that ultimately enabled these IoT devices to be pervasive, context-aware, and ambient intelligent. With the ever-increasing number of intelligent technologies given noteworthy visibility, the gathering of IoT devices has enlarged individuals' overall impression of IoT as an improved living empowering influence on that of vulnerabilities for cybercriminals. Over time, the far-reaching reception of IoT prospered out of the progressions made in Internet connectivity, wireless networking technologies, cloud computing, the reduced expense of sensors, memory, and the significant number of associated devices.

1.3 IoT applications

According to PwC's 6th annual Digital IQ survey, IoT is being enabled by advances in miniaturization, wireless connectivity, and increased data storage capacity. The applications for IoT extend across a wide variety of use cases and verticals [1].

Wearables

Wearable technology is a sign of IoT applications and probably is one of the earliest industries to have deployed the IoT in its administration. We see Fit Bits, heart rate monitors, and smartwatches everywhere. One of the least known wearables includes the Guardian glucose monitoring device. The purpose of the device is to aid individuals who have diabetes. It identifies glucose levels in the body using a tiny electrode called a glucose sensor placed under the skin and transfers the data using Radio Frequency to a monitoring device [3].

Smart Home Applications

For example, *Jarvis* is the AI home automation employed by Mark Zuckerberg. Allen Pan's Home Automation System performs functions in the house incited using a string of musical notes.

Health Care

Reactive medical-based systems are being converted to proactive wellness-based systems. The current medical research lacks critical real-world information. It mainly utilizes leftover data, controlled environments, and volunteers for medical examination. IoT opens ways to an ocean of valuable data through analysis, real-time field data, and testing.

Smart Cities

The optimized traffic system is one of the many aspects of a smart city. The smart city idea is that it is particular to a city. The problems faced in one city differ from those in other cities. Each city is affected differently because of global issues, like finite clean drinking water, deteriorating air quality, and increasing urban density, occurring in different intensities across cities.

Agriculture

Innovative Greenhouse enhances the growth of crops by *controlling environmental parameters.* A greenhouse with embedded devices is easy to be monitored and enables us to control the climate. Statistics estimate that the world population may reach nearly 10 billion by 2040. Greenhouse parameters are measured and sent to the cloud. Data is processed to apply a control action according to the requirement.

Industrial Automation

Return on Investment (ROI) is based on product quality and faster development. IoT applications are game-changing solutions for the following domains to deliver better cost and customer experience [3].

- Factory Digitalization
- Product flow Monitoring
- Inventory Management
- Safety and Security
- Quality Control
- Packaging optimization

1.4 Artificial Intelligence (AI)

What does intelligence mean? It is possible to portray intelligence as the capacity to understand, perform, and adapt to various techniques suitable to the situation to solve problems and accomplish goals appropriate to the circumstances in an evolving world. Intelligence can be cultivated by experience, analyzing past outcomes, or selecting the desired environment [4]. Artificial Intelligence (AI) focuses on machines. AI is capable of mind or acquiring and applying knowledge and skills, whereas ML does not require explicit programming. Unlike natural intelligence like human Intelligence, AI is a type of intelligence displayed by machines, especially in the current context of computer systems [5].

In 1956 AI was first introduced by John McCarthy at an academic conference. Before that, the concept of machine intelligence and its capacity to advance rationale was an understudy. In 1950 Alan Turing fostered an empirical test of AI, later known as the "Turing Test." The test was to examine an artificial entity is intelligent or not by contrasting the answering capability of the human being involved in the test [6].

The rise of AI started with research that involved modeling the neurons in a human brain. The research was based on a binary variable representation of artificial neuron signals, which can be switched on or off. According to the programming concept, these signals were represented as 0's and 1's, which further helped Donald Hebb in 1949 to develop Hebbian Learning for neural networks [5].

The first neural network computer was built in 1951 by Marvin Minsky and Dean Edmonds, named as Stochastic Neural Analog Reinforcement Calculator (SNARC). In the forthcoming years, AI has emerged as a new discipline in the computer and network technology industry whose objective was to create computer systems that could learn, react, and make decisions in a complex and changing environment [5] [6].

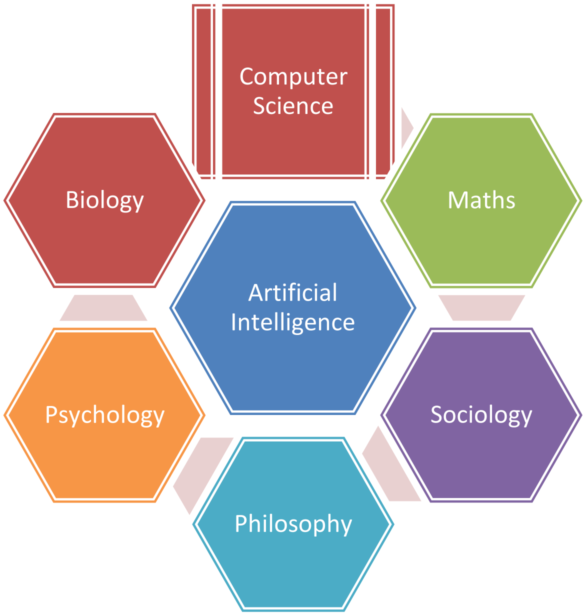


Figure 1: Disciplines Contributing to AI *[7]*

The factors contributing to AI are always valid. Those factors continuously change based on the development of new technologies and how they are inherited into AI techniques. Figure 1. represents the science and technology disciplines such as Mathematics, Computer Science, Biology, sociology, philosophy, and Psychology are the widely known contributed disciplines for AI. These areas assist in developing intelligent systems that can control the AI for performing human intelligence-related machine features, such as reasoning, understanding, and problem-solving [8].

1.4.1 History of Artificial Intelligence

Year	Milestone/Innovation
1923	Karel Capek's play Rossum's Universal Robots (RUR) opened in London, which marked the first English use of the term "robot."
1943	Foundations for neural networks laid.
1945	Isaac Asimov, a Columbia University alumnus, coined the term Robotics.
1950	To define Intelligence, Alan Turing implemented the Turing Test and published Computing Machinery and Intelligence. Claude Shannon published a detailed Study of Chess Playing as a Search.
1956	John McCarthy invented the word Artificial Intelligence. Demonstration of Carnegie Mellon University's first AI program running.
1958	John McCarthy invents LISP programming language for Artificial intelligence.
1964	The dissertation by Danny Bobrow at MIT showed that machines could comprehend natural language well enough to address algebra word problems correctly.
1965	ELIZA, an interactive problem that conveys a conversation in English, was developed by Joseph Weizenbaum at MIT.

1969	Stanford Research Institute scientists have developed Shakey, a robot fitted with locomotion, vision, and problem solving
1973	Freddy, the Popular Scottish Robot, was designed by the Assembly Robotics group at Edinburgh University, capable of using vision to find and build models.
1979	The first autonomous computer-controlled vehicle, the Stanford Cart, was constructed.
1985	Aaron, the drawing program
1990	Necessary Machine Learning simulations, Case-based reasoning, Multi-agent planning, Scheduling, Data mining, Web Crawler, Comprehension of natural languages and translation, Vision, Virtual Reality, Games
1997	The Deep Blue Chess Program would defeat Garry Kasparov, the then world chess champion.
2000	Interactive robot pets are being available commercially. MIT features Kismet, a robot with a face that conveys thoughts. The Nomad robot visits Antarctica's isolated areas and locates meteorites.

Table 1: History of AI during the 20th century

1.4.2 Symbolic Artificial Intelligence

Symbolic artificial intelligence is the earliest and most common representation of AI. The intelligent system can be explicitly described in Symbolic AI. Knowledge is expressed symbolically, and intellectual operations can be identified as formal operations over symbolic expression and structures. The symbolic AI is sub-grouped into two generic models of knowledge representation and intelligent operations, where the AI approach like cognitive stimulation and logic-based reasoning is defined. The second sub-group concentrates on specific applications and is based on representations of domain knowledge. The rule-based representation, structural knowledge representation, and the mathematical linguistics approach of AI are defined in this part of symbolic AI [7].

1.4.3 Computational Intelligence

Computational Intelligence is another group of methods in AI. The standard features include numeric information is fundamental in a knowledge representation, and numeric computation is used to process the information. Unlike symbolic AI, the information is not represented straightforwardly. Not all computational intelligence models include these features; model like the Bayes network model does not fulfill all the characteristics. The models are presented in a generic way to avoid this misunderstanding and grouped under connectionist, mathematics-based, and biology-based models [7].

1.4.4 Weak Artificial Intelligence (AI)

It is also known as Narrow AI, the primary and most common type of AI we use in our computer systems, smartphones, or the Internet. It is non-sentient machine intelligence with not-so-complex programming instructed to perform narrow tasks with less complexity. These limited tasks include facial recognition internet searches with given keywords [9].

1.4.5 Artificial General Intelligence (AGI)

Before briefing about AGI, it is essential to understand the concept of general Intelligence. General Intelligence is the ability to accomplish various objectives and initiate various tasks in various environments and platforms. These systems should deal with issues and circumstances that are very different from those predicted by their developers [9] [10]. "Artificial General Intelligence" has emerged as a synonym for "narrow AI" to relate to systems with solid generalization ability. The AGI method views 'general intelligence' as a primarily specific property from the task or problem-specific capability and mainly focuses on knowing this particular property and developing display systems. AGI can also close the gap between the narrow AI and the advanced AI programs, including the complex AI programming and functionality seen in robots [9] [10].

1.4.6 Superintelligence

Artificial Superintelligence or Superintelligence is considered the most advanced type of AI ever developed/developing. Most researchers view it as a futuristic approach to building advanced multi-functional, complex AI. *Superintelligence* is defined as any form of intelligence greater than current general intelligence and exceeds humans' cognitive performance across all operations platforms; that is pretty ambiguous now. Under this concept, various systems with disparate performance attributes may classify as superintelligences. It is beneficial to disaggregate the basic notion of superintelligence. It can be separated into various bundles of intellectual super-capabilities to advance the study. There are several ways that such decomposition can be accomplished. Thus, we can classify superintelligence as speed superintelligence, collective superintelligence, and quality superintelligence [11].

Speed superintelligence can process information much faster than a human. Collective superintelligence achieves superior performance by aggregating large numbers of more negligible intelligence. Quality superintelligence processes information fast and is vastly qualitatively more intelligent [11].

1.5 Machine Learning

The influence of machine learning technologies has dominated the 20th-century Information Technology infrastructure. It has a promising result that can efficiently integrate big data algorithms with the knowledge-based artificial intelligence techniques in various computer systems platforms. Many attributes and techniques are combined to represent machine learning technology. However, it is essential first to understand the proper definition of machine learning and how it has a massive impact on the current computer and network infrastructure. So, what is machine learning? Many definitions and explanations were introduced over the years on machine learning. Machine learning is the area in computer science based on understanding pattern analysis and the principle of machine learning in Artificial Intelligence. Machine learning deals with developing algorithms and studying algorithms that can understand a particular human and non-human behavior and predict the data generated. These algorithms use the available data to derive predictions and decisions from standard programming, as standard programming comes to conclusions based on static programming instructions. In contrast, machine learning algorithms are based on data-driven methodology [12].

In 1959, Arthur Samuel defined *machine learning* as a "Field of study that gives computers the ability to learn without being explicitly programmed." Tom M. Mitchell provided a widely quoted, more formal definition: "A computer program is said to learn from experience E concerning some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" [12].

Machine learning is generally defined as improvements in artificial intelligence-related systems. This task includes information gathering, analysis, recognition, diagnosis, planning, robot control, prediction, and decision making. Figure 2. represents a simple artificial intelligence system known as a typical AI agent [13].

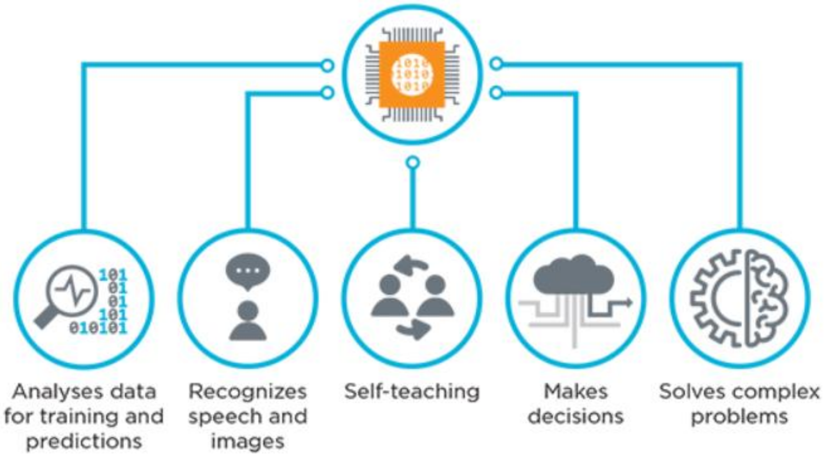

Analyses data for training and predictions
Recognizes speech and images
Self-teaching
Makes decisions
Solves complex problems

Figure 2: AI System *[13]*

This AI system builds its environment and computes necessary actions based on the results accumulated from each of these components shown in figure 2. The decisions are made in anticipation of their effects on the results. The changes made based on the previous results can be counted as learning, and the AI systems use various machine learning mechanisms depending on which subsystem is being changed [13].

1.5.1 Importance of Machine Learning

Machine learning is essential, and the reasons below explain its importance.

- The importance of establishing a relationship between input and output data for a more significant set of sample inputs, machine learning algorithms can reorganize the internal structure to generate accurate outputs. Thus, the functionality is extended to connect inputs and outputs.
- Machine learning can perform data mining, which helps to extract meaningful relationships and correlations hidden in a large pile of data

- Machine learning methods help adapt to the changing working environment quickly and avoid the inefficiency that can cause a system in a production environment before and after that system is designed and commissioned.
- The volume of information about these tasks can be too high for explicit encoding by humans. Machines that ultimately obtain this information would capture more than people would like to markdown.
- Machine learning methods can help the AI systems adapt to the redesign based on new knowledge gained and new technologies associated with the systems [13].

1.5.2 Machine Learning Algorithms

Machine learning is applied in various fields like computer games, surveillance, security, data mining, virtual personal assistants like Amazon "Alexa" Google home & personal assistants' programs like Nest mini, and other areas like share market and search engines, medical diagnosis. Various machine learning algorithms are deployed to different areas of interest based on the approach it needs to take, the input and output data type, and the complexity of the algorithm's problem. Techniques like supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning depend upon the availability of styles and categories of training data. Below are the most commonly used machine learning algorithms [14].

- **Decision Tree Learning**

  Decision tree learning participates in a decision tree as a predictive model, which plots an item's expectations to predict its target value. A decision tree is a learning model used for classification. [15]

  This machine-learning algorithm uses a tree structure technique to divide data into groups and represent the flowchart outcomes. This model classifies knowledge in a dataset by flowing through a request system from the root before reaching the leaf, which implies one class. The root is the grouping feature that plays a primary role, and the leaf determines the class [16] [17] [18].

- **Rule-Based Learning**

  Rule-based learning is another algorithm used in machine learning that is almost similar to a decision tree. Hence, it can be converted without much complexity. Identifying and utilizing a set of relational rules that collectively describe the information captured by the system is the defining characteristic of a rule-based machine learning environment. Unlike other machine learning algorithms, this is generally defined by a singular model that can be uniformly applied to any example to make a prediction [18] [18].

- **Artificial Neural Networks**

  An artificial neural network algorithm is a machine learning algorithm inspired by biological neural networks' configuration and functional characteristics, widely referred to as the 'neural network.' Simulations are structured using a connectionist approach to estimating and processing an interactive group of artificial neurons. Modern neural networks are computational data processing tools that are non-linear. Typically, they are used to model dynamic relationships between inputs and outputs, locate data correlations, or capture the statistical structure of an undefined joint probability distribution function between variables observed [16] [18].

- **Inductive Logic Programming**

Inductive logic programming is a rule-learning method that uses logic programming as a consistent representation for input cases, context knowledge, and hypotheses. The known history encoding of an inductive logic programming scheme can derive a hypothesized logic program that includes both positive and negative examples, with information and a collection of examples described as a rational database of evidence. *Inductive programming* is a similar discipline that considers the expression of theories in some form of programming language, such as functional programs [18].

- **Support Vector Machines**

Support Vector Machines (SVMs) can be related to supervised learning methods used for classification and regression. The algorithm produces a model that predicts whether a new norm falls under one or another group, presenting a collection of training instances, each marked as belonging to one of two classes, an SVM training course [18] [19]. It is possible to classify SVM models into four distinct groups, namely:

a. Classification SVM Type or C-SVM classification
b. Classification SVM Type 2 or nu-SVM classification
c. Regression SVM Type 1 or epsilon-SVM regression
d. Regression SVM Type 2 or nu-SVM regression

- **Clustering**

The clustering algorithm of machine learning analyzes a cluster. It allocates information into subsets, so that information within the same cluster is identical according to some pre-designated parameters or criteria. In contrast, observations from different clusters are dissimilar. Other clustering methods make various assumptions about the data structure, often described by a metric of similarity and evaluated, for example, by internal compactness and separation. It is between multiple clusters. Other approaches are based on the approximate density and connectivity of graphs. Clustering is a tool for unsupervised learning and a popular method for evaluating statistical results [18] [19].

- **Bayesian Networks**

A Bayesian network is a stochastic graphic model relating a series of random variables and their conditional independence through a directed acyclic graph, also known as a belief network or directed acyclic graphical model. For instance, a Bayesian network may describe the probabilistic interactions between diseases and symptoms. The network can measure the probabilities of different illnesses' occurrence and provide the signs [18].

- **Reinforcement Learning**

  It is concerned with how an agent believes in taking action in an environment to optimize any notion of long-term reward. Reinforcement learning algorithms focus on finding a policy that guides the agent's actions to the states of the universe in those states. Reinforcement learning varies from supervised learning because sub-optimal action neither presents nor specifically corrects good input/output pairs [18].

- **Representation Learning**

  Several learning algorithms, primarily unsupervised learning algorithms, seek to find better representations of the inputs provided during training. Typical examples include analysis of critical components and cluster analysis. Representation learning algorithm attempts to conserve the data in their input but transforms it to make it usable. It is often transformed as a pre-processing step before classification. Predictions enable the information from the uncertain distribution of data generation. Predictions are reconstructed to unplausible configurations under that distribution [18].

- **Genetic Algorithms**

  Genetic algorithms are more stable and used for different optimization problems. Unlike other machine learning and AI algorithms, these algorithms do not deviate readily in the face of noise. Genetic algorithms are used in large-area or multimodal space searches. They are algorithms that rely on the evolutionary principle of natural and hereditary preference. Genetic algorithms are adaptive heuristic search algorithms, i.e., the algorithms adopt an iterative pattern that differs over time. It is a form of reinforcement learning where it is essential to provide feedback without specifying the correct direction. The input may be either positive or negative [20].

1.5.3 Applications

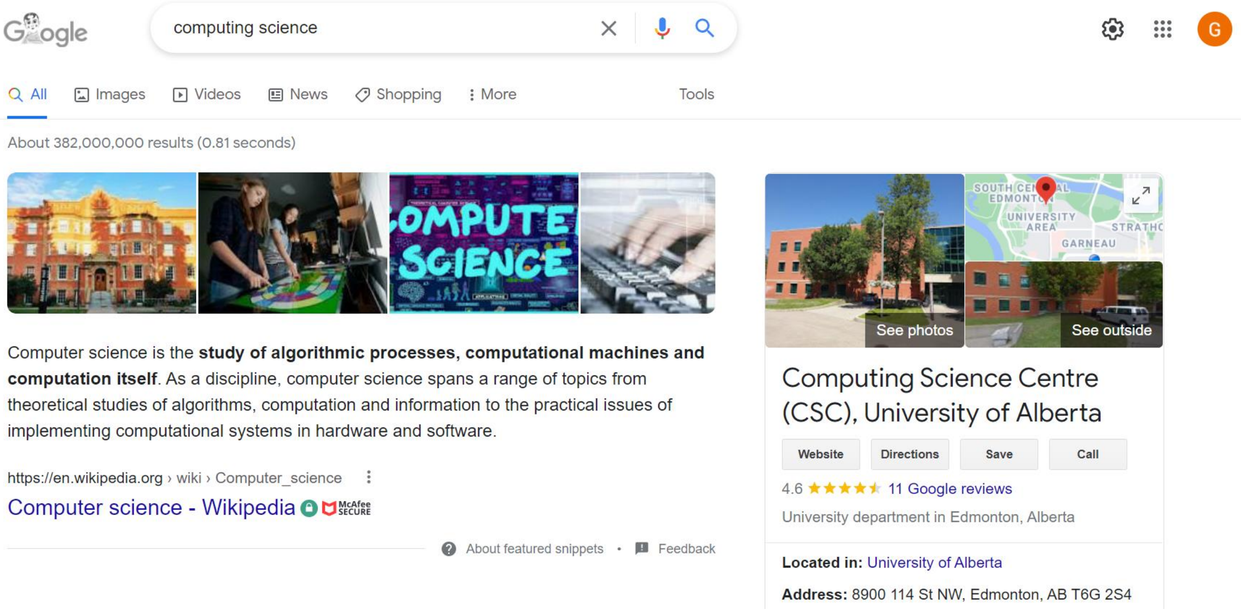
Machine learning is one of the most discussed and applied technologies. There is always progress in combining artificial intelligence algorithms, machine learning, and increasing the efficiency of many well-known computer technologies. Hence, machine learning has many uses and areas of interest in platforms. Here are some of them

- **Web Page Ranking**

  One of the most common functions on the Internet is search engines. A wide range of online search engines is available based on user needs. Web pages are loaded on the search engine based on the query provided in the search engine, which then compares it to the web, various domains in the search engine database with keywords included in the query. Figure 3. is an example of these search engines (in this case, the Google search engine is for the original word "Computing Science"). As search engines produce these related sites and related resources? Search engines use search algorithms for searching. Different website structures analyze them with many subfactors, such as user geographic location, the original word, and the number of times that the websites are displayed. After running through various algorithms and tools, the final result is displayed mainly due to the user's geographic location and frequent visits to the University of Alberta and academic buildings (based on statistical data for Google Maps) [14].

- **Collaborative Filtering**

  Another critical but similar category of web pages is shared data filtering. This data is intended for anyone of any type or level of complexity. Collaborative filtering is the process of filtering a result based on an analysis of the behavior of previous users or computer systems. For example, users typically see the list of suggestions on their Netflix page, comics on the Amazon eCommerce site, Facebook status, and friend suggestions based on mutual friends. Various ML and AL techniques support the participatory filtering process with different degrees of complexity [14].



About 382,000,000 results (0.81 seconds)

Computer science is the **study of algorithmic processes, computational machines and computation itself**. As a discipline, computer science spans a range of topics from theoretical studies of algorithms, computation and information to the practical issues of implementing computational systems in hardware and software.

https://en.wikipedia.org › wiki › Computer_science

Computer science - Wikipedia

Computing Science Centre (CSC), University of Alberta

4.6 ★★★★★ 11 Google reviews

University department in Edmonton, Alberta

Located in: University of Alberta

Address: 8900 114 St NW, Edmonton, AB T6G 2S4

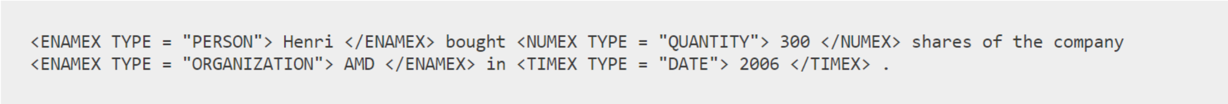
Figure 3: Web Page Ranking on a Search Engine

- **Facial and Named Entity Recognition**

  One of the main applications where machine learning is very effective is facial recognition technology. This machine learning application plays a vital role in different layers of security in different government and non-governmental organizations and individual skills. Several security applications, such as access control use this technology. This technique categorizes people's faces and determines their security clearance. It divides the face circumference into complex mathematical values based on eye relief and face shape. This process has different degrees of complexity and the relative frequency with which artificial intelligence techniques are adapted to machine learning algorithms [14].

- **Specific entity recognition**

It is an application used in face recognition. Identifying a particular entity deals with the problem of identifying a particular substance or entity. These definitions are crucial for data mining, automated extraction, and document understanding. For example, some modern applications such as Apple Mail use this technology to automatically identify and complete delivery addresses and complete them in the address book. While systems that use manually designed rules can produce satisfactory results, coding document samples to automatically identify these dependencies is more effective, especially when implementing the framework in multiple languages. Figure 4 is an example of identifying a named entity using LingPipe to tag a person's name. Relevant information such as location and organization is marked to extract further information [14].



```
<ENAMEX TYPE = "PERSON"> Henri </ENAMEX> bought <NUMEX TYPE = "QUANTITY"> 300 </NUMEX> shares of the company
<ENAMEX TYPE = "ORGANIZATION"> AMD </ENAMEX> in <TIMEX TYPE = "DATE"> 2006 </TIMEX> .
```

Figure 4: Named entity tagging on a news article *[14]*

- **Automated Hacking**

Automated hacking is another important machine learning application with many positive and negative effects. The adaptation of machine learning technology and artificial intelligence technology to computer hacking has made a massive difference in dealing with such machine hackers, both offensively and defensively. Machine learning detects intruders, random and continuous injections, and distributed denial of service (DDoS) attacks. This research paper also discusses these two aspects of offensive and defensive methods in machine hacking. Aside from those applications, here is a list of other technologies that use machine learning as an essential method of implementation and processing.

a. Affective computing
b. Bioinformatics
c. Brain-machine interfaces
d. Cheminformatics
e. Classifying DNA sequences
f. Detecting credit fraud
g. Machine perception
h. Robot locomotion
i. Stock market analysis
j. Medical diagnosis
k. Software engineering
l. Optimization and metaheuristic
m. Speech and handwriting recognition

Chapter 2: Cybersecurity Issues

2.1 What is Cybersecurity?

It refers to the tools, practices, and protocols that protect data assets from malicious agents. Cybersecurity is an upgrade in the traditional concept of information security. InfoSec views security as a corporate issue rather than a problem influencing consumers and national security on some level. Cybersecurity comprises three core elements: people, process, and technology [21].

People

Almost all cyberattacks target individuals, even if they are allegedly the victim of a government company or organization. Hackers steal people's personal information or disrupt people's lives. Hence, people should be aware of their role in reducing and preventing threats [21].

Process

Organizations need to develop policies and processes that reduce threats. They need to adapt to the changing cybersecurity landscape. Regulations are part of that mix. The General Data Protection Regulation (GDPR) law and data breach notification statutes help protect individuals and businesses from cyberattacks [21].

Technology

People and companies need to invest in technologies that protect them from cyber-attacks. Attackers use technology, so defenders should do the same that including things like firewalls, encryption, intrusion detection, and more [21].

IoT devices like industrial sensors are vulnerable to various cyber threats that include hackers who take over the device as part of a DDoS attack and gain unauthorized access to the data collected by the device. IoT devices are the primary target of malicious agents due to their number, geographic distribution, and often outdated operating systems. A data breach is information theft by a malicious agent. Motives for data breaches include crime (e.g., identity theft), wanting to embarrass an organization (like Edward Snowden or hacking DNCs), and espionage. Cybersecurity threats refer to infiltrations of infrastructure and data breaches, spear phishing, and brute force [21].

2.2 What is a data breach?

A data breach is a cyber security incident where a malicious agent gains unauthorized access to private data. This attack leads to the theft of private and primarily confidential data. The nature and approach of data breach attacks vary widely, but the result is almost always the same. People or organizations who do not have access to our data can see it and, in most cases, steal it. There have been several widespread and highly disruptive data breaches in recent years. It destroys public trust in well-known brands and public institutions and threatens to harass millions of consumers, the victims of fraud. Data breach protection is improving, although much remains to protect data from unauthorized and unlawful security breaches [21].

How can a data breach happen?

The word "breach" refers to a gap, a break in a defensive wall. Hackers use various methods to reach a target's defense system and steal their information. Usually, the attacker penetrates the target system via a remote access point. Many similar options are offered but accessed from a remote location with stolen funds. A talented hacker can impersonate IT personnel and gain information about an actual system user [21].

Spear Phishing

In spear phishing, the attacker pretends to be a colleague, friend, or ex-employee to steal login credentials. Most people admit that hackers use well-known standard logins (factory settings) that have not changed since the deployed system [21].

Identifying vulnerabilities

Sometimes, hacking is more technically sophisticated. Hackers look for vulnerabilities in networks to find minor holes in defenses that allow them to slip away unnoticed. Login credentials are stored in the cache, and when the cache is uncleared, an attacker could steal credentials and use them in a breach. IT managers fail to apply security patches to systems and devices. Then, attackers cause known but unknown exploits to infiltrate the system [21].

Eavesdropping

Data breaches can also occur during eavesdropping. Suppose the hacker can insert himself in the middle of the message stream on the target network. It is possible to collect data from the message traffic with external communication links (e.g., the Internet) [21].

2.3 Phases of a data breach

Data breach includes three different phases examination, break-in, and exfiltration.

Phase 1: Examination

The attacker examines the target initially, which usually means mapping the network and system infrastructure. For example, before launching an attack, a hacker wants to know which programs, operating systems, and databases the target uses. The techniques used to hack Microsoft SQL Server databases are different from the methods used to crack Oracle databases on the Linux platform. Reconnaissance also includes identifying individuals responsible for securing and managing data. Hackers use social engineering and public and semi-public mechanisms like Facebook and LinkedIn. The publicly available personal information enables hackers to impersonate the users as required to break in [21].

Phase 2: Break-in

In this phase, the actual break-in takes place. Once inside, the attacker infiltrates the database itself. It does not matter unless the attacker does not want to be detected. Multiple data breaches occur over the months with the aim of not knowing that an attacker is on their network and copying and exfiltrating terabytes of confidential information. The attacker usually needs "root" or super administrator access to the target system to achieve this. Attackers can create a fake account. A skilled hacker can also use root access to hide their activity [21].

Phase 3: Exfiltration

Eventually, unauthorized extraction or copying of data occurs. The hacker can send the stolen information offline in an almost virtually invisible state by encrypting the stolen data. Data that does not seem important can be valuable to someone. Information is stolen with strategic value but low monetary worth. The critical sectors at risk for data breaches include Business, medical, Government, banking, and education. [21]

Malware in mobile applications is just as vulnerable to malware attacks as other computing devices. Attackers could include malware while downloading applications, mobile websites, or phishing emails and text messages. When a mobile phone is hacked, it can give the malicious person access to personal information, location data, and financial accounts.

2.4 Types of Cybersecurity

While cybersecurity encompasses a wide range of tools and techniques, cybersecurity is divided into three general categories

- **Data Security**

  Hackers are often behind the data. They want to see or steal off-limits information. In some cases, the hacker steals information, such as a credit card number, to sell on the black market (The "dark web"). Information thieves embarrass the target by exposing private conversations or spying on a geopolitical enemy. Data security includes protecting data from unauthorized access, including data encryption, data access control technologies, and policies [21].

- **Network Security**

  A hacker must first access the target network under almost all circumstances for a cyberattack to work. Network security is one of the most dangerous areas of cybersecurity and typically the center of significant investments. Network security includes firewalls, bastion hosts, hardware extensions, intrusion detection systems (IDS), security incident and event management systems (SIEM), and more [21].

- **Application Security**

  Hackers also like to use software programs such as Enterprise Resource Planning (ERP), CRM, and email servers. Backstage inside an app is a great way to spy on a target or disrupt their activities. Application security comes in many forms, but it typically combines policies such as who has "back end" access to the application and management and control over the Application Programming Interfaces (APIs) that allow other software programs to access the app [21].

2.5 Types of Cybersecurity threats

Cybersecurity threats refer to infiltrations of infrastructure, data breaches, spear phishing, and brute force. With the popularity and outstanding capabilities of IoT devices, more and more companies are keen to develop IoT products that support their business operations. Companies build their IoT architecture to implement products that suit them. In general, the IoT architecture consists of three layers according to its function, namely the sensor and operator layer (information acquisition), the network layer (data transmission), the information processing and application layer, and an in-depth overview of 4 IoT architecture [22].

The sensor and actuator layer consists of different sensors, actuators, and ports, referred to as edge devices. We receive a lot of physical parameters, sensory information, or other desired information such as gas sensors, proximity sensors, infrared sensors, smoke detectors, GPS, and cameras. Therefore, these devices are smaller and have little processing power and memory. Therefore, they lack a robust encryption mechanism or a sophisticated algorithm for data backup. In addition, providing physical security is a hassle because these devices function in different environments. Attacks like vandalism, eavesdropping, RF jamming, simulation, forgery, and DDoS always appear on the horizon [22].

We can transfer data through sensors and motors to the cloud via the Internet or wireless networks with the network layer. The attacks as a man in the middle attack, DOS, DDO, the kidnapping period, Sibel, the selective deviation can prevent the hello attack in this layer. Finally, the data processing layer includes cloud technologies to store a large amount of data collected in real-time through many analytical algorithms. Delivery of specific applications for users through the human user interface, for example, the intelligent fire system, home equipment, creative production, are emerging. This layer is also vulnerable to attacks such as injection of malicious code, script via the site, SQL injection, phishing, social engineering, and DDoS. [22].

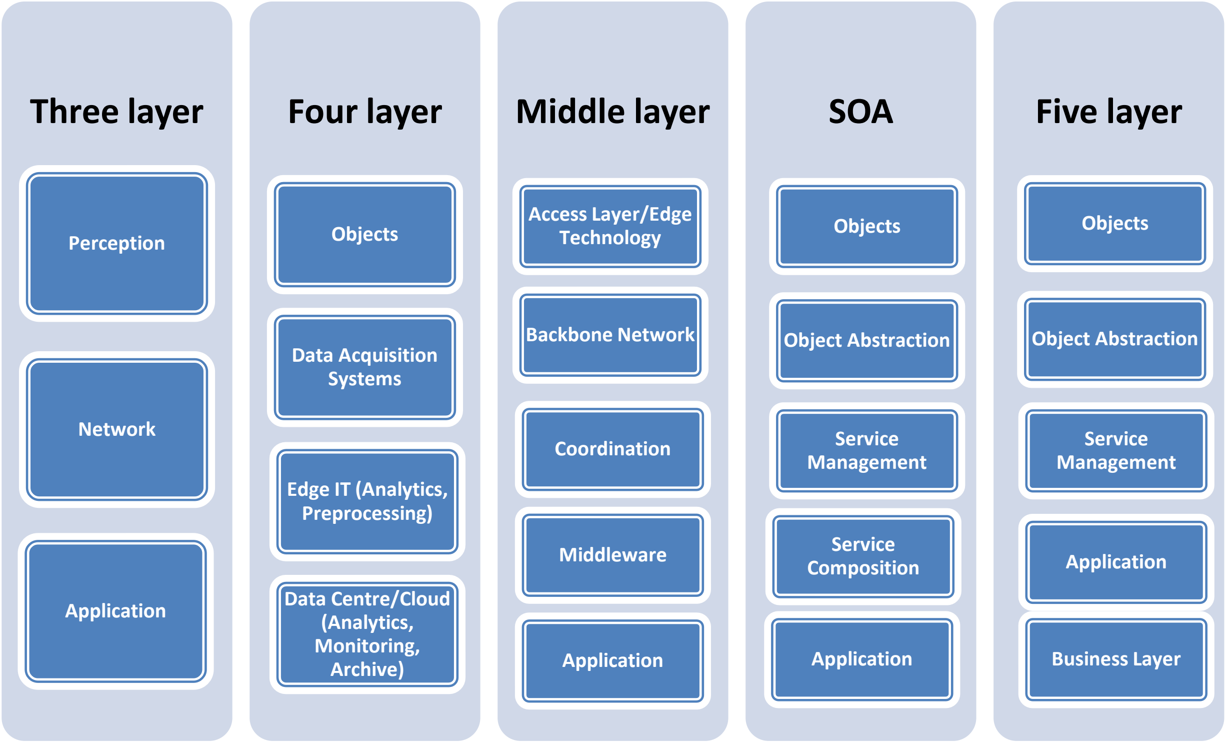


Figure 5: A typical IoT Architecture *[22]*

Description of different attacks in IoT Networks

Cyber threats are one way to attack a data source that is not an actual attack. It is like a plan of attack. There are currently hundreds of millions of cyber threats and are generally divided into the following categories

Virus/Malware

Software that does malicious work on a device or network, such as corrupting data or taking over a system. The virus is malicious software installed on a device. Once implanted, a virus can do several destructive things, including blocking the system, stealing data, or even hijacking a device for criminal purposes such as extracting digital currency without permission. For example, "crypto-jacking" [22].

Identity theft

A hacker steals enough personal information (date of birth, SSN, and address) to reveal your identity. A hacker could potentially steal money from a bank account, open credit card accounts on our behalf, and more [22].

Phishing

An email attack tricks an email recipient into disclosing sensitive information or downloading malware by clicking a link in a message. As mentioned above, a more sophisticated form of attack is known as spear phishing. It involves an attacker forging a friend or colleague to identify themselves, usually to share the account balance [22].

Password Attacks

If a hacker has our password, they can access our accounts. Password attacks use special software to guess the password and often try thousands of possibilities before finding the right word [22].

Man in the Middle Attack (MitM)

An attacker creates and intercepts a location between the sender and recipient of an email and can change it during transmission. The sender and receiver believe that they are in direct contact. MitM attack can be used in the army to confuse the enemy [22].

Trojan horses

In ancient Greek history, the name Trojan horse is malware that penetrates a target system, e.g., a standard piece of software, but executes malicious code on the host system. It is a cyber-attack that infiltrates the target network under pretenses. For example, a hacker could embed a virus in a PDF document and send it to us as an email attachment. The virus file is inserted into our system when we open a PDF file while the document is opened in Acrobat Reader [22].

Ransomware

An attack encrypts data on the target system and demands a ransom to allow the user to reaccess the data. These attacks range from minor disruptions to major incidents, like the city of Atlanta completely shutting down municipal data in 2018. A type of malware encrypts our data and forces us to unblock the ransomware, usually Bitcoin [22].

Denial-of-service attack or distributed service denial attack (DDoS)

An attacker takes possession of many (possibly thousands) devices and uses them to call up the functions of the target system, for example, a website, and to crash them due to high demand [22].

Advanced Continuous Threat (APT)

APTs are arguably the most effective cyber threat. APTs are the product of national intelligence agencies. An access point sender and receiver are designed to stealthily infiltrate our network for months and then hack it unnoticed. It moves horizontally and locks onto different parts of our infrastructure repeatedly until activated. It can then do incredible damage [22].

How can effective cybersecurity practices be maintained?

Achieving and maintaining cybersecurity can be challenging but is not a compression process. No single element makes everything, but weaknesses in one area can be catastrophic for anyone. For a larger organization, a more complex cybersecurity program is necessary.

Although security is technical, it depends on the security policy. These rules and regulations define how cybersecurity is implemented and maintained. For example, a security policy can require passwords to be of a specific length and contain different characters. Security policies can determine who can access which system and approve or deny access requests. A lawyer should also ensure that the operation complies with the relevant regulations.

Security is also an organizational and popular topic. Security managers oversee policy definition and implementation. Managing security activities typically includes monitoring the system and incident response process. The critical element is the strategic cybersecurity technology suite, starting with a strong foundation, guidelines, and people. There are many options here, and the right option largely depends on the company's size. However, robust network security is a must. As mentioned earlier, network attack detection and monitoring are beneficial [22].

Steps to be taken after a data breach

The victim of a data breach can follow the steps to protect the data by limiting the breach's impact on life. A small comforting idea that came up after the Equifax attack involves its attribution. Most data breaches involve selling stolen data on the 'Dark Web,' an online black market. Anecdotal evidence suggests that the Equifax breach did not result in the expected increase in identity fraud from dark web sales. The attacker may have been a national government like Russia or China that needed the data for espionage purposes [22].

1. Be prepared

A company like Exactis and Facebook can own data. Beware of unusual notifications, such as messages that look like unsolicited emails but contain essential information about the status of our personal information.

2. Fraud alert with banks and credit bureaus

If hackers steal someone's personal information, they can try to open credit accounts on their behalf. Creating fraud alerts makes it difficult for hackers to open such accounts. If anybody tries to open an account in another name, the bank or credit bureau contacts the account holder to confirm the request is legitimate. It is a prevention control to block the dire consequences of a data breach.

3. Monitor online bank and finances

When someone fraudulently uses the banking information (e.g., a debit/credit card number) to perform unauthorized charges from the account, the bank usually provides a refund if contacted immediately.

4. Monitor credit

Credit bureaus like Equifax and Transunion allow us to check our credit scores. When someone commits identity theft and runs into debt using the card on their behalf, their credit score may reduce. Being aware of these activities as soon as possible can solve this problem.

5. Block credit card

Bank allows us to block accounts so that nobody can use them or change the criteria. E.g., if a hacker gets information and then contacts a bank that claims to be us and asks us to add our "spouse" (another hacker) as a cosignatory for the account, the frozen account may block it from happening. Most banks and credit card providers allow us to set up regular account inquiries and announcements of individual activities.

6. Subscribe to portals

This website tracks unauthorized use of email accounts. If an email is hacked, attackers can discover sensitive information. If notified of an email hack, it is best to change the account password or close the account. For example, http://www.haveibeenpwnd.com/

7. Watch out for phone scams

When hackers have our personal information, they can trick us into giving more information or sending them money over the phone on pretenses. For example, someone who claims to be from the IRS can contact us over the phone. They can ask us to pay the tax filing fee with a credit card. That way, they give enough information about us to make it seem like it comes from the IRS. Note that the IRS never contacts anyone. They only use the mail. Caution helps with such calls. One recommended method is to hang up the phone and call the institution they claim to be, for example, a bank. Ask them to contact back.

Most companies prepare data preparedness plans to reduce a data breach's financial and visual risks. These plans are not always bulletproof. However, they represent efforts to maintain data security. Associated plans and activities include steps such as

- Data encryption
- Implement robust data access controls
- Manage network access control
- Application monitoring and program management

Data breaches may not end anytime soon and may increase over the years, but their impact on consumers can diminish if adequately regulated. [22]

2.6 Sources of Cybersecurity threats

Cyber threats come from different places, people, and contexts. The malicious actors include the following

- Individuals use their software tools to create attack vectors
- Criminal organizations run like companies, with many employees developing attack vectors and carrying out attacks.
- Nation-states
- Terrorists
- Industrial spies
- Organized crime groups
- Unfortunate insiders
- Hackers
- Business competitors

Nation-states are the source of many of the most dangerous attacks. There are several versions of national cyber threats. Some of them are simple espionage, an attempt to uncover the national secrets of another country. Others are aiming at disruption. These are so-called "cyber weapons" with which the power supply can be terminated from enemy territory during the war. In some countries, the lines between criminal organizations and national intelligence agencies are blurring, and criminals are doing the real work of cyber espionage. Many cyber threats are traded on the "Dark Web," an illegal but ubiquitous part of the Internet. Avid hackers can buy ransomware, malware, compromised systems, and more in this online marketplace. The dark web acts as a threat amplifier where a hacker can sell his work repeatedly [23].

Best tools for cyber defense

The best organizational practices for countering cyber threats include basic but essential countermeasures, such as patching systems. When a technology vendor finds that their product has a security vulnerability, they usually write code that fixes or "patches" the problem. If Microsoft discovers that a hacker can gain access to the Windows server using a code exploit, it may issue a patch and distribute it to all Windows server licensees. They do this at least once a month, among other things. Many attacks fail if the IT departments import all security patches quickly. A wide range of new technologies and services is being launched, making it easy to build strong defenses against cyber threats [23].

1. Outsourcing of security services

Many companies have high-security skills and specialize in corporate security. Outsourcing security services does not mean moving to the cloud. Companies across a wide range of industries are turning to cloud computing to reduce their support staff, reduce costs and provide services that are not far away. Organizations evaluate the best cloud sourcing options due to the high costs to maintain hardware, programs, and staff required to provide hardware services. However, security services are not overwhelming. Fortunately, any organizational solution was unfound in outsourcing the public cloud. The management of the Managed Security Service (MSSP) and the use of MSSPs to manage vulnerabilities, security rules, and events management (SIEM), penetration detection and virtual networks (VPNs), and other items are discussed below [22] [23].

- **What is meant by Managed Security Services?**

MSS now offers the level of management in any security service we can imagine. The growth of high-speed band networks is very high to limit service to physical networks and move them partially or entirely to the cloud. Some services provided by MSS are the following

Firewalls and VPNs

These network security services are the most popular outsourcing to MSS. Depending on the complexity of our environment, they quickly approach the raw material state. Unless we have a variable range, Outsourcing firewall management and VPN services can be a quick way to reduce security management in our staff [22].

Content filtering

Organizations that use content filters are often signed in the blocklist provided from abroad from known sites that contain harmful and abusive content. MSS now offers full service in the cloud, using shared proxy servers or secure DNS services [22].

DDoS Protection

DDoS attacks have a significant risk for targeted organizations, with the ability to quickly consume all bandwidth and project management. DDoS security services, provided by both ISPS and independent suggestions, traffic filters before reaching its limitations, blocking applications related to attack patterns [22].

Security monitoring

The safety records of maintenance and review of one of the tasks take longer and cause fatigue to security analysts. MSS currently provides simple records of the most advanced analysis services, including total security management and events management [22].

Vulnerability scanning

It offers one published MSSP option and a dual-division area for adult joint services. MSSPs can offer an external scan operation from a third-party perspective and internal scan devices set up on our network. The results are usually collected in an administrative controller for central evaluation, prioritization, and treatment [22].

2. Threat Detection Tools

Threat detection tools, also known as Extended Detection Response (XDR), are an integral part of the company's cybersecurity technology portfolio that is a first-level option or first response to send up a flare when something suspicious is found on the corporate network [21].

3. Crowdsourced attack simulation/vulnerability testing tools

Some well-researched companies that offer group security services are professional hackers who can identify a company's weaknesses and report them to the security team. Two great companies that offer these services are Bugcrowd and Hackerone [21].

4. Point solutions for device management

There are several unique solutions for managing our devices. Of course, we believe that our team is the best in Prey. Eliminates the weaknesses in managing disparate devices in organizations of all sizes, with services ranging from device tracking software to remote scanning and hard disk encryption. Prey is an integrated security solution [21].

2.7 Threat modeling

Threat modeling is a structured process with the following objectives: identifying security requirements, identifying potential security threats and vulnerabilities, identifying the importance of threats and vulnerabilities, and prioritizing corrective actions. The listed procedures are followed when producing a threat model [24].

1. Document how data flows to determine the location of the attack on the system.
2. Document potential system threats as well as possible.
3. Document security controls that can be applied to reduce the likelihood or impact of a potential threat.

Threat modeling identifies threat factors that harm an application or a computer system. It takes the eyes of malicious hackers to see how much damage they can do. In threat modeling, organizations comprehensively analyze software structure, business context, and other artifacts such as performance profiles and user documentation. This process enables a deep understanding and discovery of essential aspects of the system. Organizations typically perform threat modeling when designing new applications to help developers identify vulnerabilities and understand the security implications of design, guidance, and configuration decisions [24].

Identifying the gaps

The following are the three components necessary for making a threat model

a. framework to identify the threats
b. a definition
c. concrete research questions inside the field.

The gap analysis

Gap analysis concerns threat modeling and IoT research based on three dimensions.

1. The gap between threat modeling frameworks and the IoT
2. The gap between threat modeling frameworks and security research
3. The gap between security research and the IoT

Threat modeling is performed by developers using the below steps:

Data flow diagram

It represents the flow of information in our system. Specifies a location where data enters or exits a processor subsystem that applies wherever data is stored in the system, temporarily or for a more extended period [24].

Identify threat

A threat agent is a person or group capable of carrying out a specific threat. It must be determined who wants to exploit the company's assets and whether they can use them against the company. Some threats require more experience or resources, increasing the level of actors required. However, with cloud computing and the proliferation of cyberattacks, other threats with relatively low skill levels and resources [24].

Mitigate

Mitigation measures reduce the likelihood or impact of a threat without necessarily preventing it entirely. For example, if we store user passwords as a hash in a database, two users with the same password may have the same hash. So if an attacker has access to hashed passwords and can set the password for a hash, they can search for the same hash and find all other users using the same password. However, if we add salts to each user's passwords, the cost of this particular attack increases exponentially as the attacker has to crack every single password. Rising costs reduce the chances and thus the severity of the attack [24].

Validate

Controls are preventive or countermeasures to avoid, identify, manage, or reduce potential threats to our information, systems, or other assets. Preventions are controls that can completely prevent a particular attack. For example, if we spot a threat where users' personal information can be identified by registering a particular application and deciding to obliterate it, we can prevent that particular threat [24].

Trust boundary

It is a point on the data flow diagram where data changes the level of trust. There is usually a trust boundary where data is transferred between two processes. When our program reads a file from the hard drive, there is a trust boundary between the program and the file because external processes and users can modify the data in the file. When our application calls an external process or when an external process calls our application, these are trust boundaries. When reading data from a database, reliability is usually limited because other processes can modify the database data. Any place that accepts user input is always a trust boundary [24].

Best practices of threat modeling

The excellent application of threat modeling improves security perception across the team. Creating security is the first step for everyone. Conceptually, threat modeling is a simple process. When building or updating a threat model, consider these five methods

1. **Determine the scope and depth of the analysis**

   Define the scope with stakeholders and then determine the depth of analysis for individual development teams to endanger the software model.

2. **Gain a visual understanding of the threat models**

   Graph the system's main components such as the application server, data warehouse, bold client, and database and the interactions between these components.

3. **Model of attack possibilities**

   Identify software resources, security controls, and threats and map their location to create a system security model. Once the system is designed, methods like STRID are used to identify what could go wrong. For example, threats

4. **Identify threats**

   To create a list of possible attacks, check if there are many ways a threat agent can go into an asset without investigating? Can a threat agent bypass this security check? What should the threat agent do to crack this control?

5. **Create a traceability array of weak security controls**

   Be aware of the threat factors and follow their control paths. Getting to the program's roots without a security clearance is a potential attack. When we give up control, consider whether the threat agent is stopping us or whether there are ways to work around it [24].

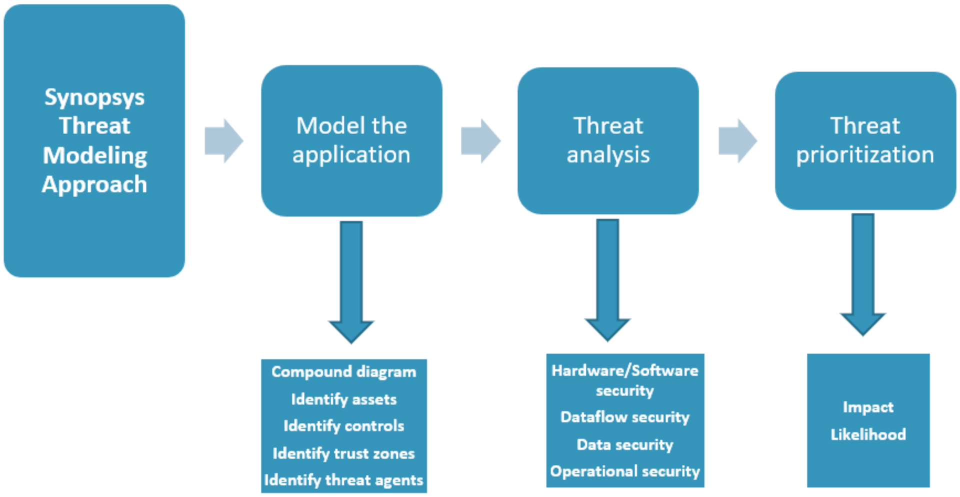


Figure 6: Synopsys Threat Modeling Approach

Synopsy's high-level approach to threat modeling involves the following steps.

a. Model the system
b. Conduct a threat analysis
c. Prioritize the threats

Model the system

System modeling consists of two parts. Create a component diagram with a control flow graph, and identify assets, security controls, trust zones, and threat agents.

Conduct a threat analysis

The critical activity in threat modeling is threat detection. The two categories are Checklist-Based Approaches and Non-checklist Based Approaches.

1. **Checklist-Based Approaches**

   Many threat modeling techniques involve checklists or templates. E.g., STRIDE recommends considering six types of threats - spoofing, tampering, denial, information disclosure, denial of service, and escalation of privilege for any over-authorized data flow.

2. **Non-checklist Based Approaches**

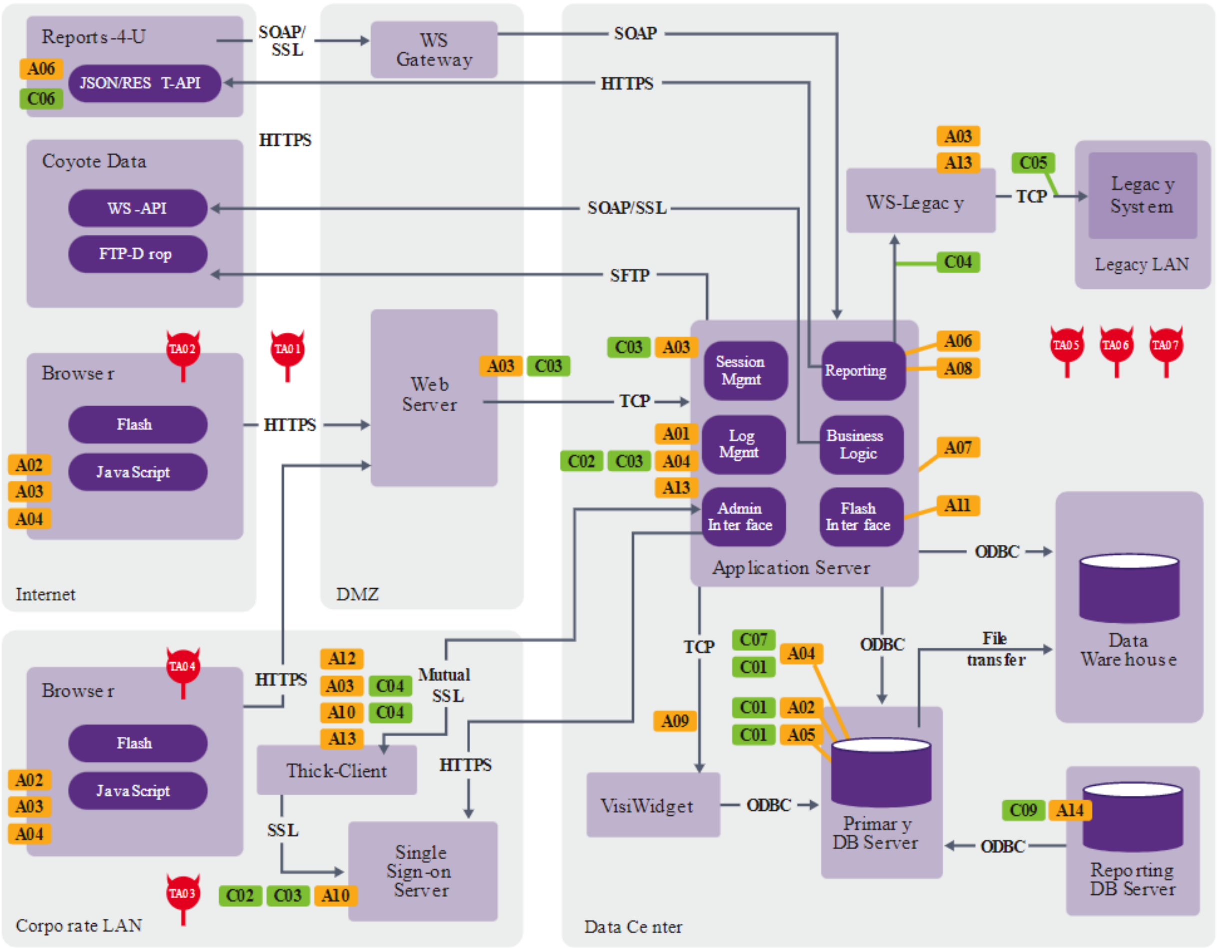
   These methods typically use creative methods such as brainstorming to identify attacks. Synopsis threat analysis uses a quasi-checklist approach. It uses a template to guide fundamental analysis but still leaves room for creative analysis. Synopsys uses pre-built application protocol threat analysis for popular application layer protocols like OAuth, SAML, OIDC, Kerberos, and password-based authentication. It is a non-exhaustive list, but it allows us to think about areas of interest to be analyzed. [24]

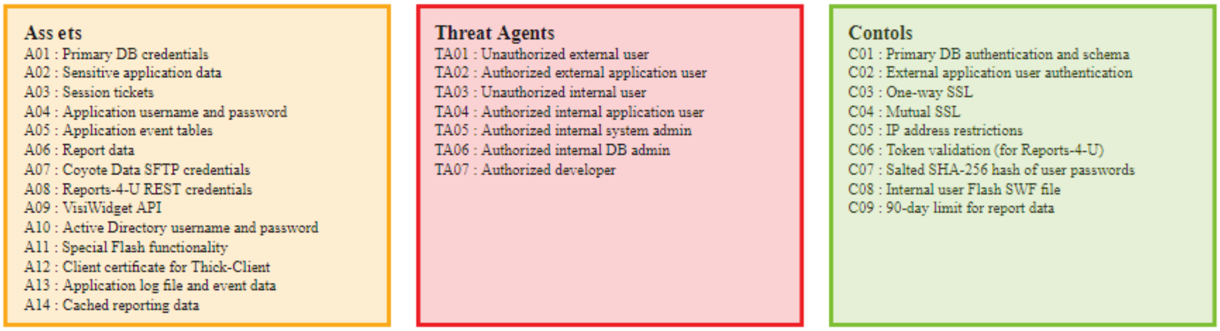
Prioritize the threats

After modeling the system and analyzing the threats, a list of threats is created. Now is the time to prioritize them. At Synopsys, the NIST approach prioritizes threats, and guidelines determine each threat's likelihood and impact on achieving severity.

2.8 Design and security challenges in IoT

Manufacturers compete on who would get the latest device in the hands of consumers first. These are the most significant security and privacy challenges currently plaguing the field of IoT-connected devices.





Ass ets
A01 : Primary DB credentials
A02 : Sensitive application data
A03 : Session tickets
A04 : Application username and password
A05 : Application event tables
A06 : Report data
A07 : Coyote Data SFTP credentials
A08 : Reports-4-U REST credentials
A09 : VisiWidget API
A10 : Active Directory username and password
A11 : Special Flash functionality
A12 : Client certificate for Thick-Client
A13 : Application log file and event data
A14 : Cached reporting data

Threat Agents
TA01 : Unauthorized external user
TA02 : Authorized external application user
TA03 : Unauthorized internal user
TA04 : Authorized internal application user
TA05 : Authorized internal system admin
TA06 : Authorized internal DB admin
TA07 : Authorized developer

Contols
C01 : Primary DB authentication and schema
C02 : External application user authentication
C03 : One-way SSL
C04 : Mutual SSL
C05 : IP address restrictions
C06 : Token validation (for Reports-4-U)
C07 : Salted SHA-256 hash of user passwords
C08 : Internal user Flash SWF file
C09 : 90-day limit for report data

Figure 7: Security model of a system *[24]*

1. Insufficient testing and updating

More than 35 billion devices worldwide are connected to the Internet of Things, and more than 60 billion by 2025. This massive wave of new tools is not available for free. One of the biggest problems for the technology companies that make these devices is that they are very careless about the security risks associated with the devices. Most of these IoT devices and products are not receiving adequate updates. In contrast, some do not receive any significant security updates, which means that the device was initially considered secure by customers when it was purchased insecure and eventually becomes vulnerable to hackers and other security breaches. Older computer systems had the same problem solved by automatic updates. However, IoT manufacturers are more eager to build and deliver their devices as quickly as possible without worrying too much about security. Unfortunately, most developers only offer middleware updates for a short time to stop when they start working on the following tool that gets the headline. It is even worse when they use old, unsupported legacy Linux kernels, making their trustworthy customers vulnerable to possible attacks due to outdated hardware and software. Every device must be thoroughly tested before being released, and companies must regularly update it to protect their customers from such attacks. Failure to do so is bad for businesses and consumers alike, as it only takes one large-scale consumer information breach to destroy a business. [24]

2. Brute-forcing and the issue of default passwords

It is used in some of the most significant and malicious DDoS attacks; the Mirai botnet is the best example of problems with shipping devices and default passwords that consumers are supposed to change after receiving them. Some government reports advise manufacturers not to sell IoT devices with poor security credentials, such as using "admin" as username or password. However, these are now more than just guidelines, and there are no legal ramifications for getting manufacturers to abandon this dangerous practice. Bad passwords and credentials make almost all IoT devices vulnerable to password hacking, especially brute-force manipulation. The Mirai malware succeeds because it identifies vulnerable IoT devices and uses standard passwords to log in and infect them. As a result, any business that uses standard factory credentials on their devices exposes their business, assets, customers, and valuable information to a brute-force attack. It is recommended to use various SSH security features to make the user inaccessible to prevent brute force attacks. Avoid weak or similar passwords on multiple devices. Instead, use a lengthy password or complex password or captcha to avoid sensitive privacy issues. We may need to set a login limit for a specific IP address or domain or create unique login URLs as part of our security strategy [24]

3. A gap in IoT skills

Most companies are arguing that there is a significant gap in the skills of IoT security professionals. This skill gap prevents companies from realizing the full potential of their employees. There should be training programs and skills development to solve this issue. Insightful workshops, hands-on newsletters, and brochures can make a huge difference. The better and more willing the team members are to use IoT security solutions, the stronger the IoT [24]. More information about the IoT security gap is mentioned in threat modeling (refer to 2.7)

4. Poor management of IoT devices

Devices enable IoT and the Internet of Medical Things (IoMT) in healthcare, retail, manufacturing, and life sciences. It shows many vulnerabilities in excellent types of connected devices. Computed Tomography machines (CT) and Magnetic Resonance Imaging (MRI) devices are primarily responsible for poor security issues in IoT devices. The combination of traditionally connected devices and legacy systems like ventilators, patient monitors, lighting, infusion pumps, and thermostats with poor security features makes them vulnerable to hacking attacks. It includes financial loss, disruption, compromised customer data, reputational damage, and safety. The IoT mentioned above security threats can be significantly reduced by implementing IoT security solutions. It addresses the requirements of customer-wide solutions and the fundamental security challenges of the device addressed by the device manager. These platforms can improve asset provisioning firmware updates, reduce vulnerabilities, and provide alerts and reports on specific metrics related to IoT assets [24].

5. IoT malware and ransomware

As the number of IoT-connected devices grows in the coming years, so will the number of malware and ransomware used to exploit them. Traditional ransomware relies on encryption to prevent users from using different devices and platforms. A combination of malware and ransomware strains is afoot that aims to combine different types of attacks. Ransom attacks can restrict or disable the device functionality while user information is compromised. For example, an IP camera is ideal for capturing sensitive information in various locations, including our home, work, or even a local gas station. The webcam is locked, and footage is sent to an infected web address. Sensitive data is extracted using the malware access point and demanding a ransom to unlock the device and restore the data. An increasing number of IoT devices may lead to unpredictable unauthorized access or theft in the future [24].

6. IoT botnets aiming at cryptocurrency

A recent surge in the value of the cryptocurrency has been very enticing to hackers trying to monetize this crypto craze. While most people find blockchain resistant to hacking, the number of attacks in the blockchain field seems to be increasing. The most significant vulnerability is not the blockchain itself but the development of blockchain applications. Social engineering extracts usernames, passwords, and private keys, and more of these may use to hack blockchain-based applications in the future.

The Monero, an open-source cryptocurrency, is one of several digital currencies currently being mined using IoT devices. Some hackers have even reused IP and video cameras to extract cryptocurrency. The blockchain breaches, IoT botnet miners, and manipulation of data integrity pose a significant risk to the flood of the open crypto market and disrupt the unstable value and structure of cryptocurrencies. IoT applications, architectures, and platforms based on blockchain technology must be coordinated, monitored, and constantly updated to prevent future misuse of cryptocurrencies. [24]

7. Data security and privacy concerns

Data privacy and security continue to be the most significant issues in the current world. Data is constantly harnessed, transmitted, stored, and processed on the web by large companies. IoT devices such as smart TVs, speakers and lighting systems, connected printers, HVAC systems, and smart thermostats are being monitored. The appropriate user data is shared between companies or even sold to many companies, violates our privacy and data security rights, and leads to public distrust. We need to govern compliance and privacy rules that determine sensitive information before storing and eliminating the IoT data payloads used to identify users. Cache and data that are not required anymore must be deleted securely. The biggest challenge is consistent with various legal and regulatory structures if the data is stored. This practice must be employed using programs and applications through the mobile, web, and cloud to access, manage and process the data associated with IoT devices [24].

8. Security Problems In Device Update Management

Firmware is one of the most significant sources influencing software security. Still, the manufacturer can provide the latest product updates with the devices they sell. These updates may result in security breaches.

When an automatic update occurs, the device also sends it back to the cloud. As a result, the device experiences a shorter downtime. If the connection is not encrypted or the files are not secure, the hacker is more likely to steal sensitive information [24].

9. Inadequate data protection

Insufficient data protection can be a critical IoT security concern. This problem can occur because of hazardous communication or data storage. In IoT security, high-end devices can access confidential data, and the necessity of secure data storage and network segregation was never evident. The potential of cryptography overcomes these data protection challenges. By encrypting confidential data, we can prevent unofficial access or data theft. In addition, data decryption can help us to protect the confidentiality and privacy of data. Furthermore, cryptography is an effective solution for thwarting eavesdropping attacks used in industrial espionage or sniffing attacks. The hacker can have passive access to ICS data (Industrial controls systems) that is received or received sent over the network. Cryptography is the standard defense against man-in-the-middle attacks, where the hacker intercepts important messages and injects new ones. [24]

10. Insecure interfaces

Every IoT device processes and sends data. They need applications, services, and protocols to communicate, and many IoT security patches come from insecure interfaces. The most common interface issues are inadequate device authentication and poor or no encryption. Device authentication prevents unauthorized access to the connected device and the data generated by it only for authorized persons. Get help from digital certificates that enable a digital entity to transmit data securely. Apply strict standards, best practices, and guidelines from reliable sources. [24]

11. Hacking IoT devices

Ransomware is one of the most dangerous types of malware. It prevents access to our confidential files through encryption. Next, the scammer demands a ransom to decrypt the confidential files. Wearables, smart homes, healthcare devices, and other ecosystems could be at risk in the future. Sometimes malware completely shuts down the device. Set to the maximum when the car cannot be turned on unless we owe a ransom while the thermostat is on. [24]

12. IoT security risks

IoT-enabled devices face many security challenges for their users. While the IoT offers excellent connectivity to devices, common IoT security vulnerabilities are nothing new. In addition to that, there are many malicious IoT risks like minimal computing power, sharing of network access, incompatible security standards, and lack of operating system updates. Security is essential for the Internet of Things. ESIM is soldered directly onto circuit boards to make it harder for hackers to damage. [24]

13. Minor IoT attacks by avoiding detection

Mirai Botnet was the largest IoT-based botnet two years ago. In 2017, the Reaper was a much more dangerous botnet than the Mirai. While large-scale attacks are acute, small attacks that escape detection are also a concern. Micro breaches might cross the security net in the years to come. Instead of using big guns, hackers may likely use a classified attack small enough to reveal information that can reach millions of records at once. [24]

14. AI and automation

Since IoT devices continue to conquer our everyday life, companies eventually deal with hundreds of thousands of people, if not millions of IoT devices. This vast amount of user data is hard to manage from the data collection and network perspective. AI tools and automation filter large amounts of data. AI tools help network administrators and security officers enforce data-specific rules and detect anomalous data and traffic patterns.

The use of autonomous systems to implement autonomous decisions that affect millions of functions in large infrastructures like healthcare, power, and transportation seems risky. The autonomous systems keep IoT secure against attacks and user data secure against theft. These challenges are mitigated with strict legal and regulatory frameworks aimed at manufacturers, with hefty fines and working constrictions used for those who do not follow said frameworks. [24]

15. **Home invasions**

Perhaps one of the terrifying threats to the Internet of Things is a home invasion. IoT devices are used in many homes and offices today, leading to the advent of home automation. The security of these devices can expose IP addresses that can identify whereabouts as a residential address. Hackers can use this vital information. In addition, if we use IoT devices in our security systems, they are likely to put us and our homes at potential risk. [24]

16. **Remote vehicle access**

In addition to home invasion, hijacking the car is also one of the IoT threats. Smart cars are becoming a reality with the help of IoT-connected devices. However, due to the Internet of Things connection, there is also a higher risk of car hijacking. A skilled hacker can remotely access our smart car, where anyone can control the car, leaving us vulnerable to deadly crimes. [24]

17. **Unreliable connections**

Many IoT devices send messages to the network without any encryption, which is one of the most significant security flaws in the IoT. It is time for all companies to ensure the highest level of encryption between their cloud services and their devices. The best way to prevent this security threat is to use shipping encryption and security standards like TLS. Another option is to use different networks that isolate different devices. Private communication ensures that the data on the web is safe and confidential. [24]

2.8.1 Design framework of AI & ML

From robots to Google Siri and now with the new Google Duplex, Artificial Intelligence seems to have made great strides toward becoming more and more human. The demand for machine learning and artificial intelligence has grown exponentially. In addition, it has grown society itself, which has led to the development of some artificial intelligence frameworks that make learning about artificial intelligence much easier. [25]

What are AI frameworks?

The AI framework makes it easier and faster to build AI applications, including machine learning, deep learning, neural networks, and NLP (natural language processing) solutions.

Framework	Language	Open source?	Features of Architecture
Tensor Flow	C++ or Python	Yes	Uses data structures
Microsoft CNTK	C++	Yes	GPU/CPU based. It supports RNN, GNN, and CNN.
Caffe	C++	Yes	Its architecture supports CNN
Theano	Python	Yes	Flexible architecture allowing it to deploy in any GPU or CPU
Amazon Machine Learning	Multiple languages	Yes	Hailing from Amazon, it uses AWS
Torch	Lua	Yes	Its architecture allows powerful computations
Accord.Net	C#	Yes	Capable of scientific computations and pattern recognition
Apache Mahout	Java, Scala	Yes	Capable of making machines learn without having to program
Spark MLib	R, Scala, Java, and Python	Yes	Drivers and executors run in their processors (horizontal or vertical clusters)

Table 2: AI Framework comparison *[25]*

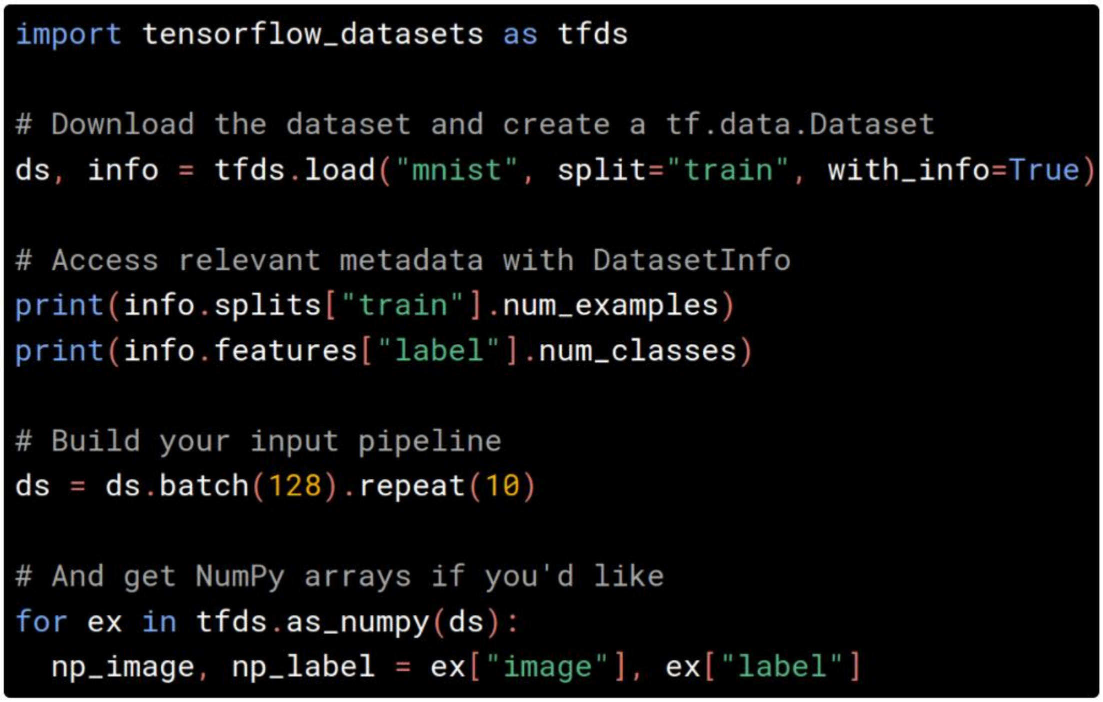
1. Tensor Flow

Tensor Flow hails from the Google family, proven to be a robust open-source framework that supports deep learning and is even accessible via a mobile device. Tensorflow is a tool for statistical program development that provides distributed training at any level of abstraction preferred by the user. The following are the features of Tensor Flow [25].

- Several scalable multi programming interfaces for easy programming
- Powerful growth drivers with a robust open-source community
- Provides comprehensive and documented manuals to the public

Drawbacks:

- The framework traverses the input data through several nodes for decision making or prediction
- It is time-consuming
- It lacks many pre-trained AI models



```
import tensorflow_datasets as tfds

# Download the dataset and create a tf.data.Dataset
ds, info = tfds.load("mnist", split="train", with_info=True)

# Access relevant metadata with DatasetInfo
print(info.splits["train"].num_examples)
print(info.features["label"].num_classes)

# Build your input pipeline
ds = ds.batch(128).repeat(10)

# And get NumPy arrays if you'd like
for ex in tfds.as_numpy(ds):
  np_image, np_label = ex["image"], ex["label"]
```

Figure 8: Tensor Flow data sets *[26]*

2. Microsoft Cognitive Toolkit

Microsoft CNTK is a faster and more versatile open source based on neural networks that support text, messaging, and voice remodeling. It provides an efficient scaling environment through a faster, comprehensive evaluation of machine models and takes care of accuracy. Microsoft CNTK integrates into a vast dataset, making it the leading choice of big players like Skype and Cortana with its expressive architecture and ease of use. The following are the features of Microsoft CNTK [25].

- Highly optimized for performance, scalability, speed, and high-level integrations
- It has built-in components such as hyperparameter tuning, guided and reinforced learning models, CNN, and RNN.
- Resources to deliver the best efficiency
- Private networks are efficient as full APIs at high and low levels

Drawback:

- It has no visualization panel and ARM Mobile support.

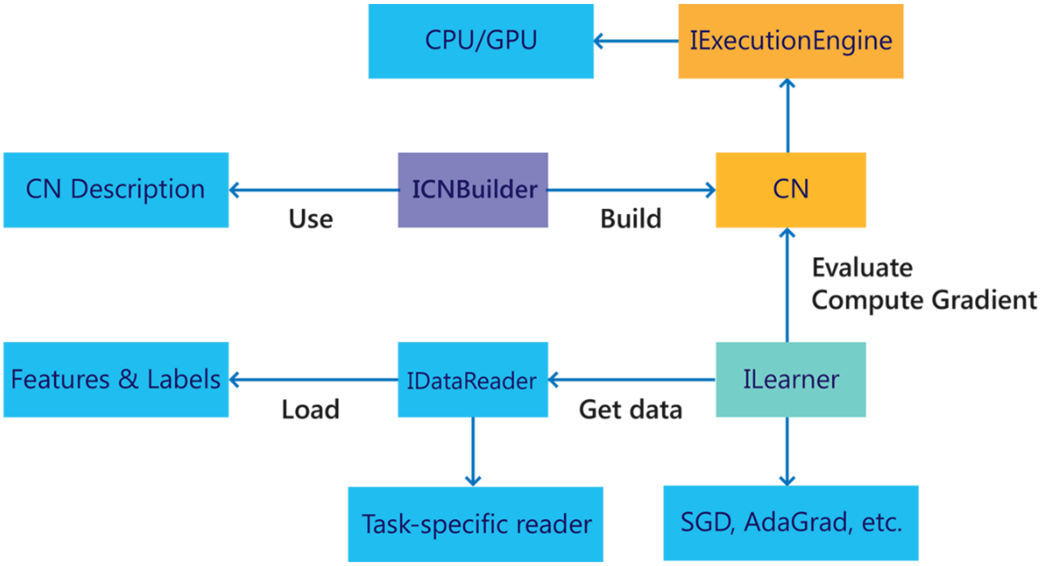


Figure 9: Microsoft Cognitive Toolkit on neural networks

3. Caffe

Caffe is a deep learning network with several preloaded sets of trained neural networks that should be our first choice as our deadline approaches. This framework is known for its image processing functions and extended MATLAB support. The features are as follows [25]

- All models are written in plain text schemas.
- It is preloaded to offer high speed and very efficient work.
- An active open source community for collaboration of code

Drawbacks:

- Caffe cannot process complex data but is relatively quick at handling visual image processing.

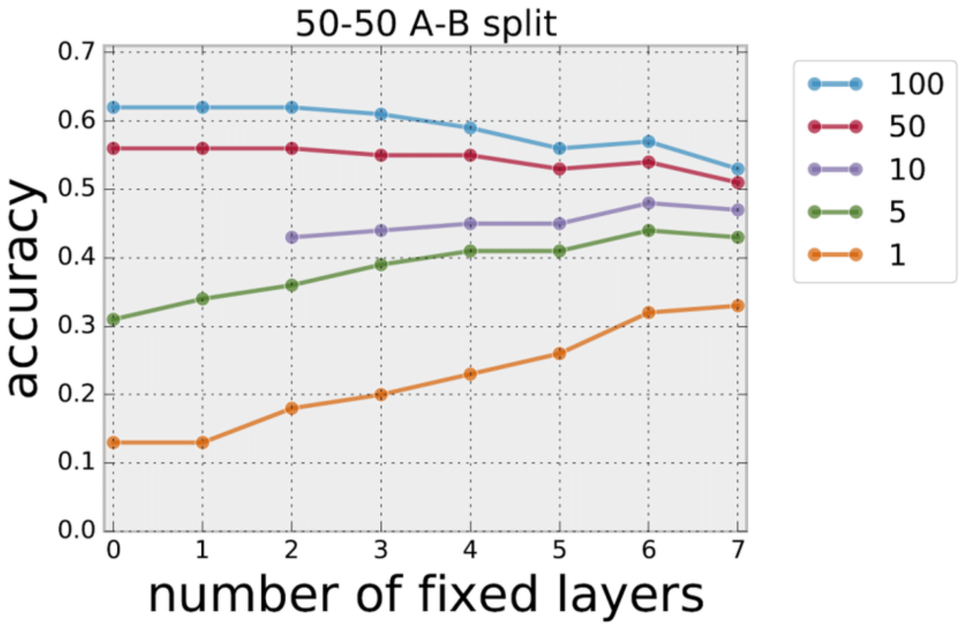


Figure 10: Accuracy of Caffe on ImageNet-1000 with varying number of fixed layers *[27]*

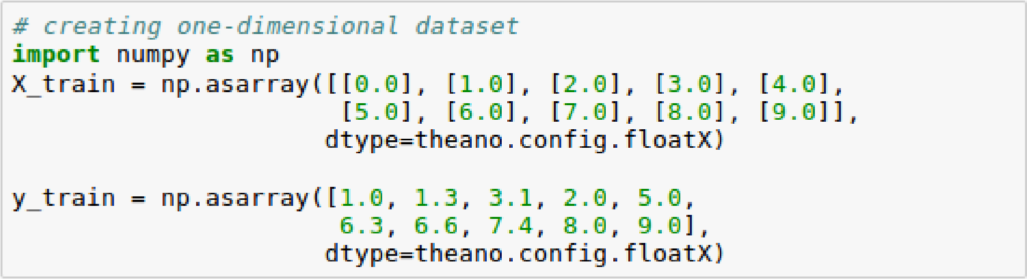
4. Theano

By using GPUs instead of CPUs, this framework supports in-depth learning research and can provide reliability and accuracy to networks that require high computing power. Theano is based on Python, a proven programming language for faster processing and response. The following are the features of Theano [25]

- Evaluation of expression is faster due to the generation of dynamic code
- It provides an excellent accuracy ratio even with minimal values
- Unit testing is an essential feature of Theano as it allows the user to self-verify the code and to find and spot bugs easily

Drawback:

- There are no updates or add ons to the existing version.



```
# creating one-dimensional dataset
import numpy as np
X_train = np.asarray([[0.0], [1.0], [2.0], [3.0], [4.0],
                      [5.0], [6.0], [7.0], [8.0], [9.0]],
                     dtype=theano.config.floatX)

y_train = np.asarray([1.0, 1.3, 3.1, 2.0, 5.0,
                      6.3, 6.6, 7.4, 8.0, 9.0],
                     dtype=theano.config.floatX)
```

Figure 11: Linear regression with Theano (Creating a One-dimensional dataset) *[28]*

5. Amazon ML

As a popular participant in the AI community, Amazon Machine Learning offers expanded support for developing self-learning tools. The framework currently has user bases in its various AWS, S3, and Amazon Redshift. It is a service operated by Amazon and comprises three operations performed on the model: data analysis, model training, and evaluation. The following are the features of Amazon machine learning [25]

- There are tailored tools for every AWS experience, even if they are a beginner, data scientist, or developer.
- Security comes first, so all data is encrypted
- Provides comprehensive tools for data analysis and understanding
- Integration with all-important datasets

Drawbacks:

- Since the framework is entirely abstracted, it lacks flexibility. So we cannot choose a specific compromise or machine learning algorithm
- It lacks data visualization.

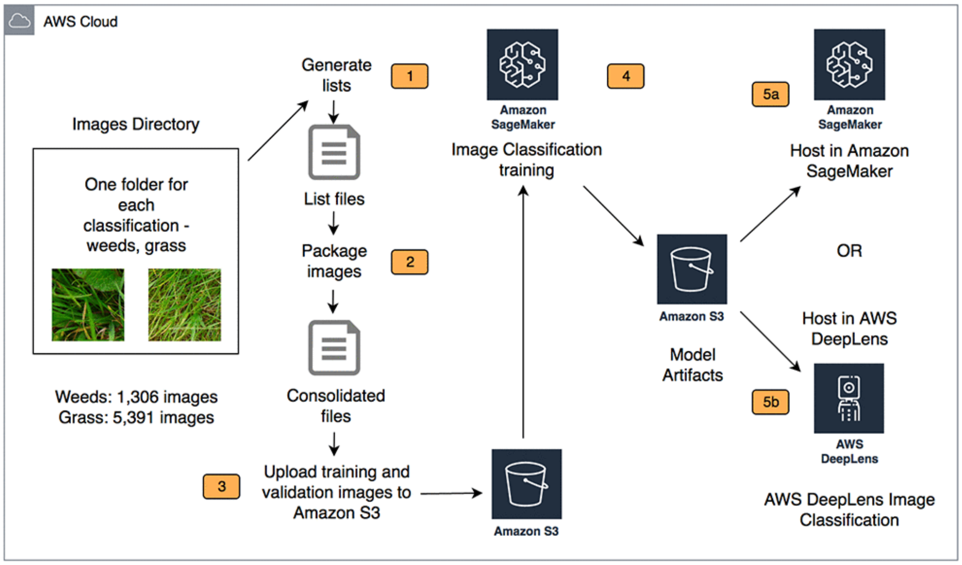


Figure 12: AWS Machine Learning and IoT services *[29]*

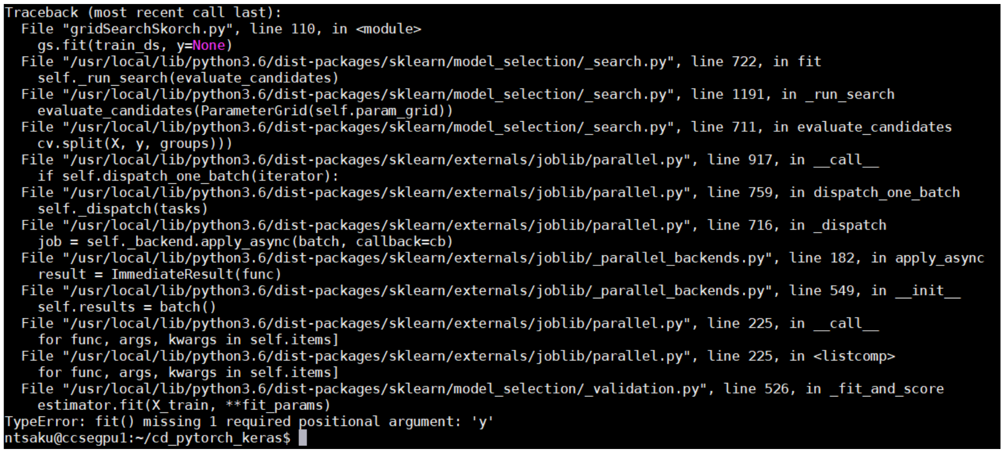
6. Torch

Torch is an open-source framework that can support scalar operations. It offers many algorithms for developing deep learning networks faster. It is widely used in Facebook and Twitter AI labs. A Python-based framework called PyTorch has proven to be more straightforward and reliable. The following are the features of Torch [25]

- It has many indexing, partitioning, and switching techniques using the N-dimensional array model
- Optimization routines are mainly based on numbers with neural network models
- Highly efficient GPU support
- Easy integration with iOS and Andriod

Drawbacks:

- Documentation is not very easy for users, so it leads to a much faster learning curve
- Lack of code for immediate use, so it may take time
- It is originally based on a programming language called Lua that many people do not know.



```
Traceback (most recent call last):
  File "gridSearchSkorch.py", line 110, in <module>
    gs.fit(train_ds, y=None)
  File "/usr/local/lib/python3.6/dist-packages/sklearn/model_selection/_search.py", line 722, in fit
    self._run_search(evaluate_candidates)
  File "/usr/local/lib/python3.6/dist-packages/sklearn/model_selection/_search.py", line 1191, in _run_search
    evaluate_candidates(ParameterGrid(self.param_grid))
  File "/usr/local/lib/python3.6/dist-packages/sklearn/model_selection/_search.py", line 711, in evaluate_candidates
    cv.split(X, y, groups)))
  File "/usr/local/lib/python3.6/dist-packages/sklearn/externals/joblib/parallel.py", line 917, in __call__
    if self.dispatch_one_batch(iterator):
  File "/usr/local/lib/python3.6/dist-packages/sklearn/externals/joblib/parallel.py", line 759, in dispatch_one_batch
    self._dispatch(tasks)
  File "/usr/local/lib/python3.6/dist-packages/sklearn/externals/joblib/parallel.py", line 716, in _dispatch
    job = self._backend.apply_async(batch, callback=cb)
  File "/usr/local/lib/python3.6/dist-packages/sklearn/externals/joblib/_parallel_backends.py", line 182, in apply_async
    result = ImmediateResult(func)
  File "/usr/local/lib/python3.6/dist-packages/sklearn/externals/joblib/_parallel_backends.py", line 549, in __init__
    self.results = batch()
  File "/usr/local/lib/python3.6/dist-packages/sklearn/externals/joblib/parallel.py", line 225, in __call__
    for func, args, kwargs in self.items]
  File "/usr/local/lib/python3.6/dist-packages/sklearn/externals/joblib/parallel.py", line 225, in <listcomp>
    for func, args, kwargs in self.items]
  File "/usr/local/lib/python3.6/dist-packages/sklearn/model_selection/_validation.py", line 526, in _fit_and_score
    estimator.fit(X_train, **fit_params)
TypeError: fit() missing 1 required positional argument: 'y'
ntsaku@ccsegpu1:~/cd_pytorch_keras$
```

Figure 13: Creating a Skorch dataset from a Torch dataset *[30]*

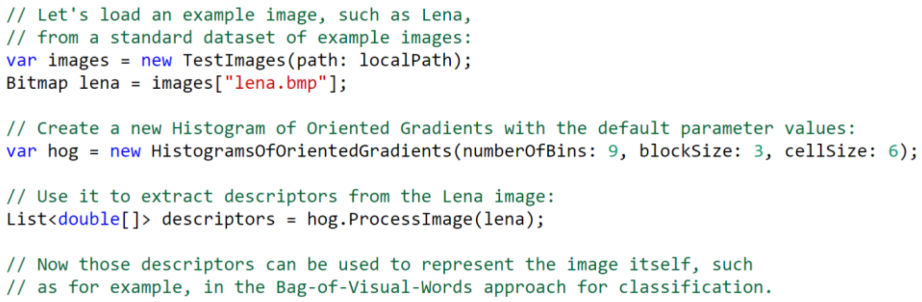
7. **Accord.Net**

Accord.net is a C # based framework that helps develop neural networks for audio and image processing. Applications can also be used commercially to produce computer vision applications, signal processing, and statistical applications. The features are as follows [25]

- The thoroughly tested mature codebase was introduced in 2012
- It provides a comprehensive set of sample models and datasets to get us started with our program quickly

Drawbacks:

- It is unknown compared to other frameworks
- The performance is slower than other frameworks.



```
// Let's load an example image, such as Lena,
// from a standard dataset of example images:
var images = new TestImages(path: localPath);
Bitmap lena = images["lena.bmp"];

// Create a new Histogram of Oriented Gradients with the default parameter values:
var hog = new HistogramsOfOrientedGradients(numberOfBins: 9, blockSize: 3, cellSize: 6);

// Use it to extract descriptors from the Lena image:
List<double[]> descriptors = hog.ProcessImage(lena);

// Now those descriptors can be used to represent the image itself, such
// as for example, in the Bag-of-Visual-Words approach for classification.
```

Figure 14: Accord.Net dataset

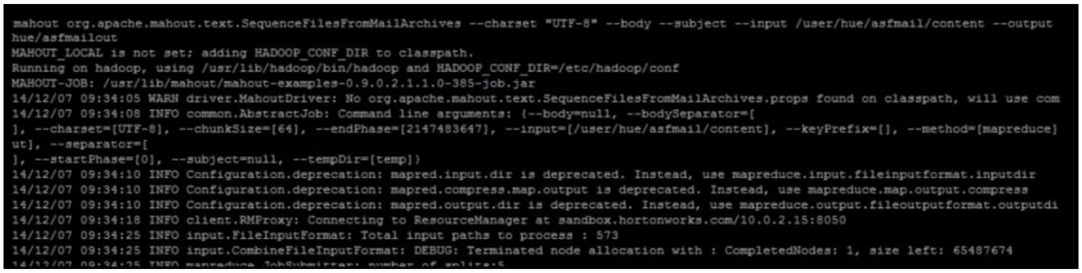
8. Apache Mahout

Apache Mahout, an open-source framework, aims to develop scalable frameworks for machine learning. It does not account for APIs that way, but it does help data scientists and engineers implement new machine learning algorithms. The following are the features [25]

- It is known for its mathematically very expressive Scala DSL
- Supports multiple distributed backends

Drawbacks:

- Python libraries are not as compatible with this framework as Java libraries
- Calculations are slower than Spark MLib.



```
mahout org.apache.mahout.text.SequenceFilesFromMailArchives --charset "UTF-8" --body --subject --input /user/hue/asfmail/content --output
hue/asfmailout
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using /usr/lib/hadoop/bin/hadoop and HADOOP_CONF_DIR=/etc/hadoop/conf
MAHOUT-JOB: /usr/lib/mahout/mahout-examples-0.9.0.2.1.1.0-385-job.jar
14/12/07 09:34:05 WARN driver.MahoutDriver: No org.apache.mahout.text.SequenceFilesFromMailArchives.props found on classpath, will use com
14/12/07 09:34:08 INFO common.AbstractJob: Command line arguments: {--body=null, --bodySeparator=[
], --charset=[UTF-8], --chunkSize=[64], --endPhase=[2147483647], --input=[/user/hue/asfmail/content], --keyPrefix=[], --method=[mapreduce]
ut], --separator=[
], --startPhase=[0], --subject=null, --tempDir=[temp]}
14/12/07 09:34:10 INFO Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.inputdir
14/12/07 09:34:10 INFO Configuration.deprecation: mapred.compress.map.output is deprecated. Instead, use mapreduce.map.output.compress
14/12/07 09:34:10 INFO Configuration.deprecation: mapred.output.dir is deprecated. Instead, use mapreduce.output.fileoutputformat.outputdi
14/12/07 09:34:18 INFO client.RMProxy: Connecting to ResourceManager at sandbox.hortonworks.com/10.0.2.15:8050
14/12/07 09:34:25 INFO input.FileInputFormat: Total input paths to process : 573
14/12/07 09:34:25 INFO input.CombineFileInputFormat: DEBUG: Terminated node allocation with : CompletedNodes: 1, size left: 65487674
14/12/07 09:34:25 INFO mapreduce.JobSubmitter: number of splits:5
```

Figure 15: Apache Mahout dataset

9. Spark MLib

The Apache Spark MLib framework is supported by R, Scala, Java, and Python. It can be loaded with Hadoop workflows to provide machine learning algorithms such as classification, regression, and clustering. In addition to Hadoop, it can be integrated into the cloud, Apache, or even standalone systems. The features of Spark MLib are as follows [25]

- High performance is a crucial element and is said to be 100 times faster than MapReduce
- Spark is incredibly versatile and works in multiple computing environments

Drawbacks:

- It can only be connected to Hadoop
- It is difficult to understand the mechanism of this framework without extensive work.

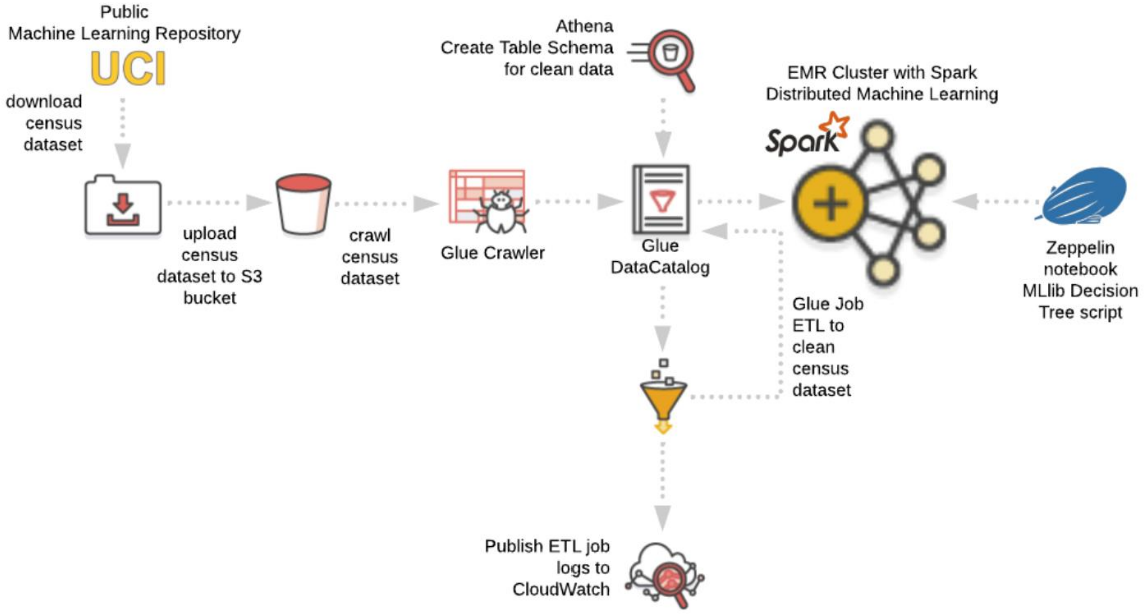

Public
Machine Learning Repository
UCI
download census dataset
upload census dataset to S3 bucket
crawl census dataset
Glue Crawler
Athena
Create Table Schema for clean data
Glue DataCatalog
EMR Cluster with Spark
Distributed Machine Learning
Spark
Zeppelin notebook MLlib Decision Tree script
Glue Job ETL to clean census dataset
Publish ETL job logs to CloudWatch

Figure 16: Spark Mlib

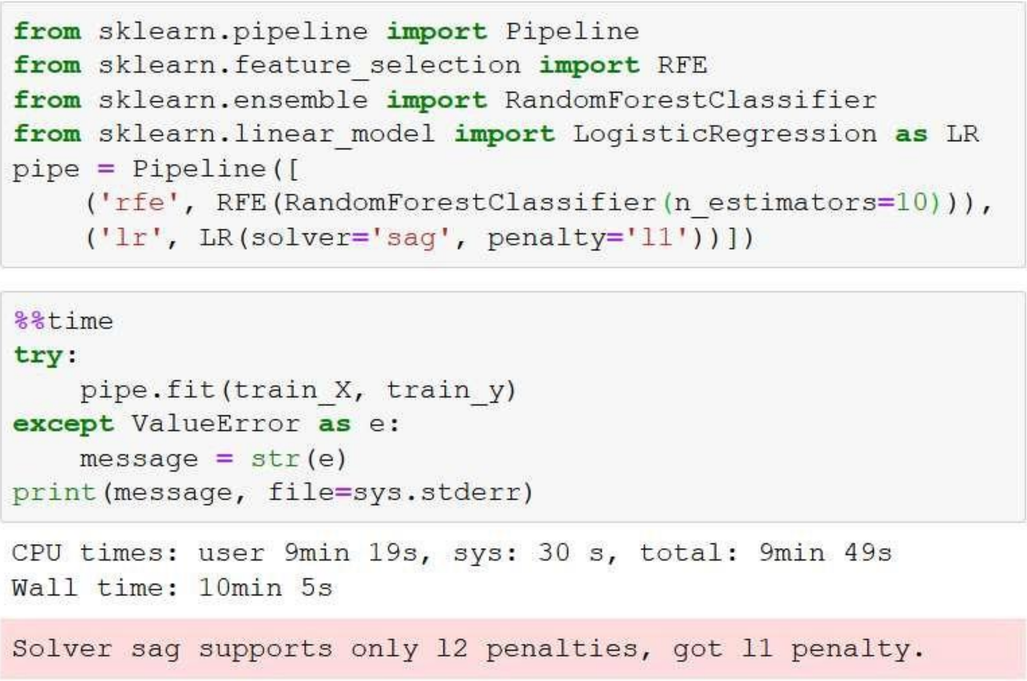
What are ML frameworks?

A machine learning framework is an interface that enables developers to quickly build machine learning models without digging deeper into the underlying algorithms. The machine learning frameworks are as follows [31]

1. Scikit-learn
2. H20
3. Google Cloud ML engine
4. Azure ML studio

Scikit-learn

It is one of the most popular ML libraries out there. It is preferred for administered and unsupervised learning calculations. Precedents perform direct and calculated relapses, choice trees, bunching, and k-implies. The framework includes many computations for regular AI and data mining tasks, including bunching, relapse, and order. [31]



```
from sklearn.pipeline import Pipeline
from sklearn.feature_selection import RFE
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import LogisticRegression as LR
pipe = Pipeline([
    ('rfe', RFE(RandomForestClassifier(n_estimators=10))),
    ('lr', LR(solver='sag', penalty='l1'))])
```

```
%%time
try:
    pipe.fit(train_X, train_y)
except ValueError as e:
    message = str(e)
print(message, file=sys.stderr)
```

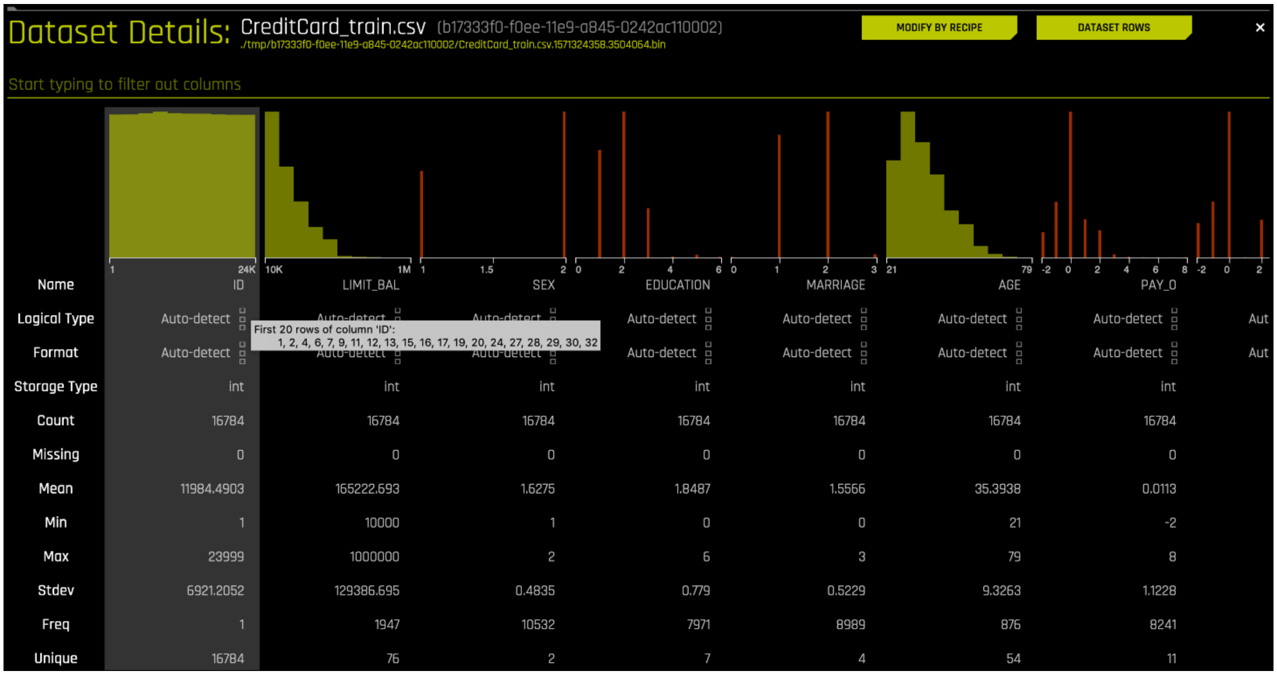
```
CPU times: user 9min 19s, sys: 30 s, total: 9min 49s
Wall time: 10min 5s
```

```
Solver sag supports only l2 penalties, got l1 penalty.
```

Figure 17: Sample Scikit learn error checking *[32]*

H20

It is an open-source machine learning platform. It is a business-oriented artificial intelligence tool that aids in data-driven decision-making and enables users to extract ideas. It is used for predictive modeling, risk and fraud analysis, insurance analytics, advertising technology, healthcare, and customer intelligence. [31]



Dataset Details: CreditCard_train.csv (b17333f0-f0ee-11e9-a845-0242ac110002)
./tmp/b17333f0-f0ee-11e9-a845-0242ac110002/CreditCard_train.csv.1571324358.3504064.bin

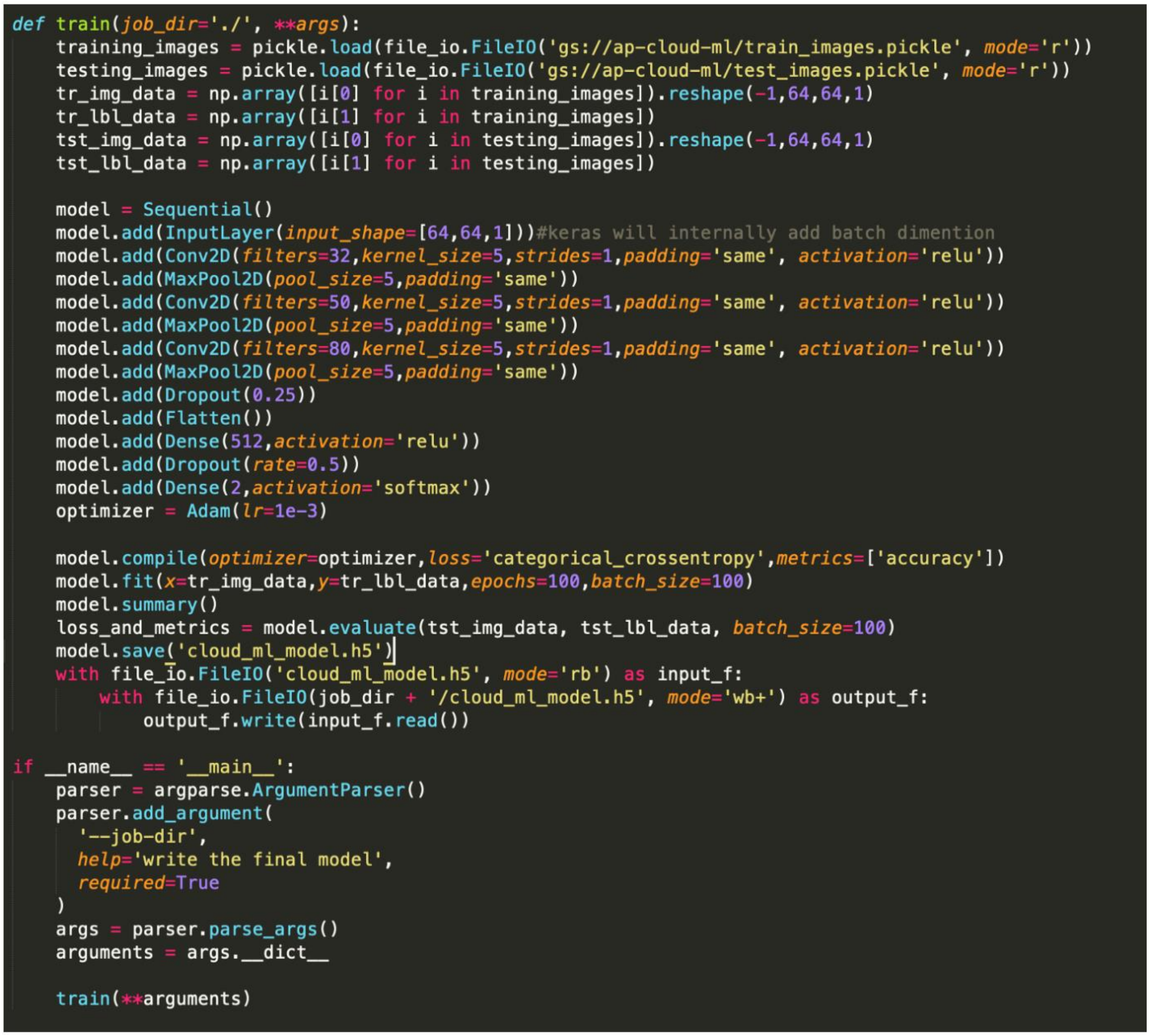
First 20 rows of column 'ID':
1, 2, 4, 6, 7, 9, 11, 12, 13, 15, 16, 17, 19, 20, 24, 27, 28, 29, 30, 32

Name	ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	
Logical Type	Auto-detect	Auto-detect	Auto-detect	Auto-detect	Auto-detect	Auto-detect	Auto-detect	Aut
Format	Auto-detect	Auto-detect	Auto-detect	Auto-detect	Auto-detect	Auto-detect	Auto-detect	Aut
Storage Type	int	int	int	int	int	int	int	
Count	16784	16784	16784	16784	16784	16784	16784	
Missing	0	0	0	0	0	0	0	
Mean	11984.4903	165222.693	1.6275	1.8487	1.5566	35.3938	0.0113	
Min	1	10000	1	0	0	21	-2	
Max	23999	1000000	2	6	3	79	8	
Stdev	6921.2052	129386.695	0.4835	0.779	0.5229	9.3263	1.1228	
Freq	1	1947	10532	7971	8989	876	8241	
Unique	16784	76	2	7	4	54	11	

Figure 18: Dataset details on H20.AI *[33]*

Google Cloud ML Engine

It is a managed service that helps developers and data scientists build and deploy superior machine learning models. Provides learning and prediction services that can be used together or separately. Used by businesses to solve problems like ensuring food safety, clouds in satellite images, and answering customer emails at 4x the speed. [31]

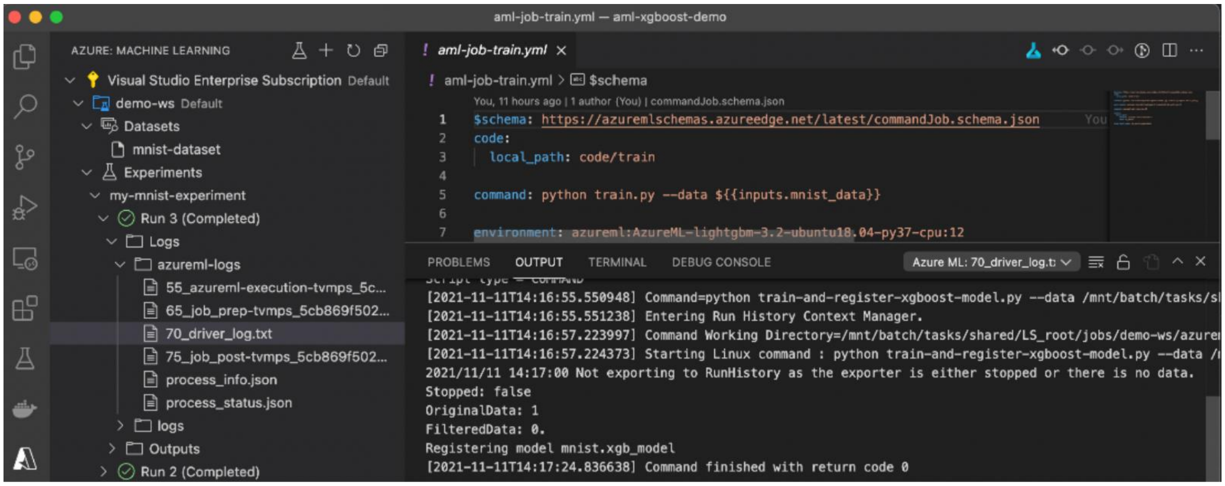


```python
def train(job_dir='./', **args):
    training_images = pickle.load(file_io.FileIO('gs://ap-cloud-ml/train_images.pickle', mode='r'))
    testing_images = pickle.load(file_io.FileIO('gs://ap-cloud-ml/test_images.pickle', mode='r'))
    tr_img_data = np.array([i[0] for i in training_images]).reshape(-1,64,64,1)
    tr_lbl_data = np.array([i[1] for i in training_images])
    tst_img_data = np.array([i[0] for i in testing_images]).reshape(-1,64,64,1)
    tst_lbl_data = np.array([i[1] for i in testing_images])

    model = Sequential()
    model.add(InputLayer(input_shape=[64,64,1]))#keras will internally add batch dimention
    model.add(Conv2D(filters=32,kernel_size=5,strides=1,padding='same', activation='relu'))
    model.add(MaxPool2D(pool_size=5,padding='same'))
    model.add(Conv2D(filters=50,kernel_size=5,strides=1,padding='same', activation='relu'))
    model.add(MaxPool2D(pool_size=5,padding='same'))
    model.add(Conv2D(filters=80,kernel_size=5,strides=1,padding='same', activation='relu'))
    model.add(MaxPool2D(pool_size=5,padding='same'))
    model.add(Dropout(0.25))
    model.add(Flatten())
    model.add(Dense(512,activation='relu'))
    model.add(Dropout(rate=0.5))
    model.add(Dense(2,activation='softmax'))
    optimizer = Adam(lr=1e-3)

    model.compile(optimizer=optimizer,loss='categorical_crossentropy',metrics=['accuracy'])
    model.fit(x=tr_img_data,y=tr_lbl_data,epochs=100,batch_size=100)
    model.summary()
    loss_and_metrics = model.evaluate(tst_img_data, tst_lbl_data, batch_size=100)
    model.save('cloud_ml_model.h5')
    with file_io.FileIO('cloud_ml_model.h5', mode='rb') as input_f:
        with file_io.FileIO(job_dir + '/cloud_ml_model.h5', mode='wb+') as output_f:
            output_f.write(input_f.read())

if __name__ == '__main__':
    parser = argparse.ArgumentParser()
    parser.add_argument(
      '--job-dir',
      help='write the final model',
      required=True
    )
    args = parser.parse_args()
    arguments = args.__dict__

    train(**arguments)
```

Figure 19: Google Cloud ML Engine

Azure ML Studio

This framework enables Microsoft Azure users to build and train models and convert them into APIs that other services can use. We can also connect our Azure storage to offer larger models. We do not need an account to use Azure ML Studio to try this service. We can register anonymously and use Azure ML Studio for up to eight hours. [31]


aml-job-train.yml — aml-xgboost-demo
AZURE: MACHINE LEARNING
Visual Studio Enterprise Subscription Default
demo-ws Default
Datasets
mnist-dataset
Experiments
my-mnist-experiment
Run 3 (Completed)
Logs
azureml-logs
55_azureml-execution-tvmps_5c...
65_job_prep-tvmps_5cb869f502...
70_driver_log.txt
75_job_post-tvmps_5cb869f502...
process_info.json
process_status.json
logs
Outputs
Run 2 (Completed)
aml-job-train.yml
aml-job-train.yml > $schema
You, 11 hours ago | 1 author (You) | commandJob.schema.json
PROBLEMS
OUTPUT
TERMINAL
DEBUG CONSOLE
Azure ML: 70_driver_log.t

```yaml
$schema: https://azuremlschemas.azureedge.net/latest/commandJob.schema.json
code:
  local_path: code/train

command: python train.py --data ${{inputs.mnist_data}}

environment: azureml:AzureML-lightgbm-3.2-ubuntu18.04-py37-cpu:12
```

```
[2021-11-11T14:16:55.550948] Command=python train-and-register-xgboost-model.py --data /mnt/batch/tasks/s
[2021-11-11T14:16:55.551238] Entering Run History Context Manager.
[2021-11-11T14:16:57.223997] Command Working Directory=/mnt/batch/tasks/shared/LS_root/jobs/demo-ws/azure
[2021-11-11T14:16:57.224373] Starting Linux command : python train-and-register-xgboost-model.py --data /
2021/11/11 14:17:00 Not exporting to RunHistory as the exporter is either stopped or there is no data.
Stopped: false
OriginalData: 1
FilteredData: 0.
Registering model mnist.xgb_model
[2021-11-11T14:17:24.836638] Command finished with return code 0
```

Figure 20: Python on Azure ML

2.8.2 Countermeasures

Several layers of administrative, technical, and physical controls protect organizational assets against dangers. It creates a strong defense. Senior management's participation and support are essential for the successful establishment and a continuous information security structure. IoT's important features require management attention. [34]

Manufacturers and vendors should include the security design process. The most effective strategy for secure IoT is based on fundamental principles. IoT device manufacturers, IoT application and platform developers, IoT connectivity architects, IoT service developers, and experienced IoT designers should work together. Anyone involved in the IoT must be essential to add safety features in its solution development design stage. The best attempt to prevent attacks includes designing the security and embedding firewall functions to add a defense layer to provide encryption capabilities that include tamper detection functions. [34]

When manufacturers do not test their devices, consumer trust and safety can be at risk. It is vital to ensure that security is purpose-built in any aspect of the ecosystem is guaranteed that performs the IoT product or a specific device. Vendors should always aim for good practice and focus on confidentiality, integrity, and availability (Central Intelligence Agency triad). The main difference in IoT security is compared to traditional safety, the number of devices, purpose, and physical conditions. The most critical problem is that IoT manufacturers still do not consider their devices like computers. [34]

Testing must be provided to confirm that the device and protocols can handle the IoT ecosystem by developing the test specification accepted in the market that helps introduce the time required to receive the tested product or protocol, which helps the devices that can work with other IoT objects. Improved security configuration requires IoT web interface management, IoT traffic review, and the need for physical ports, authentication assessment, and communication with cloud and mobile applications [34].

The IoT fragmentation increases network security. Therefore, the development of IoT protocols works together and ensures security and privacy. Unused services/ports must be closed because these ports/services expose the device to additional attack vectors. It is essential to cancel unnecessary services that can be unpleasant to be used as a target or result of attacks. It is also necessary to create authentication between devices so that reliable devices can only exchange data. A solid password management tool is required to manage multiple IoT passwords [34]. This training promotes awareness to users and consumers about the vulnerabilities experienced by their devices. When choosing an appropriate IoT device, consumers must require vendors to defend the device against common attacks.

User data should be processed and encrypted to stay safe. The complete communication channel from the sensors should be secure. The huge gap in security includes ensuring confidentiality by providing encrypted communication streams and ensuring integrity by providing encrypted data storage and using hash integrity checkers, and providing authentication methods to communicate with known and trusted entities. Providing security updates continue in the form of patches and bug fixes. [34]

Regulations force manufacturers and vendors to prioritize security and provide developers and IoT manufacturers guidelines. The IoT regulation provides transparency to consumers or packaging that can reflect the device's security level. Creating a proper legal framework and developing basic technology and privacy in mind is necessary. Regulations force manufacturers to upgrade and secure their products. IoT applications should be considered for the protection of General Information EU (GDPR). GDPR implemented a mandatory notification regime in the event of personal data breaches. Data controllers are asked to report personal data breaches to their supervisory authorities no later than 72 hours after such a violation. In some cases, such breaches are reported to affected individuals. Data controllers that use the IoT must ensure that they are in a position to identify and respond to security breaches in such a way as to comply with the requirements of GDPR. [34]

Regular firmware updates help protect the ecosystem and IOT from dealing with almost all functional processes. It must be possible to get the firmware updates, the operating system, or specialized logic on stationary and mobile IoT devices at all costs. These maintenance interfaces require access to the application runtime environment and security settings for the applications themselves. [34]

It is essential to place monitoring systems when an event occurs. When the event is discovered, a responsive action must be triggered to prevent malicious device use. A backend application must have a function in which this deformation in the data received must be lodged separately. Supervision and software maintenance reduce device downtime due to software bugs or potential problems. [34]

A practical IoT framework must provide guidelines for managing the IoT risk organizations face.

1. Enable security and control in each design right from the beginning.
2. Build security in the life cycle of software development in IoT devices.
3. Enable IoT hardening, access management, log management, and patch management.
4. Enable audit controls for data collection, privacy, storage, handling, and disposal.

5. Enable controls in network protocols for remote access, session management, and access control.
6. Enable test control and search vulnerabilities by creating and testing use cases and misuse cases.
7. Practice program effectiveness of IoT monitoring controls.
8. Build a watchdog protocol to monitor the connectivity continuously and detect connection loss by optimizing resources. A watchdog on every activity may track IoT devices, handling any events on time.
9. Emphasize the importance of security with performance.
10. Build and improve security skills and ensure information technology to ensure cyber security, risks, and benefits.
11. Align IT function and usage of IoT business.
12. Plan system acquisition, development, and maintenance
13. Organize trust between IoT devices.
14. Maintain asset inventory and dispose of IoT devices when necessary.
15. Practice governance on IoT initiatives.
16. Design devices with security in mind.
17. Build malware software in IoT programs.
18. Audit the IoT environment and determine the data stream in the IoT environment.
19. Build a vulnerability management program. Include vulnerability assessment and penetration testing.
20. Develop IoT threat modeling. Implement governance and accountability.

It is essential to develop a framework for identifying and assessing safety risks in IoT to ensure confidentiality, integrity, and availability in future work. [34]

Chapter 3: Cloud Computing

3.1 What is cloud computing?

Cloud computing provides computing services, including servers, storage, databases, networking, software, analytics, and intelligence over the Internet to enable faster innovation, flexible resources, and economies of scale. We usually pay for the cloud services that allow us to reduce operating costs, run our infrastructure more efficiently, and scale our business by changing our business needs. Information and data stored on physical or virtual servers are maintained and managed by a cloud computing provider such as Amazon and their AWS products. [35]

Cloud computing has speared employment opportunities worldwide with technological giants such as Amazon, Google, and Microsoft by recruiting people for their cloud infrastructure. Before the advent of cloud computing, companies and businesses had to set up their data centers and allocate resources and other IT professionals by increasing costs. The rapid development of the cloud leads to more flexibility, downsizing, and scalability. The cloud computing market may increase to USD 832.1 billion by 2025. Cloud computing quickly evolves and gradually understands its business and more researchers, scholars, computer scientists, and practitioners. The formation of different cloud techniques constitutes cloud computing. [35] [36].

Figure 21: Cloud computing services *[36]*

3.2 Importance of Cloud computing

Before cloud computing, companies had to store their information and software on their hard drives and servers. The greater the business, the more storage space is required. This method of data processing is not scalable in terms of speed. Cloud computing is suitable for individuals and corporations. Regardless, the cloud has also changed our lives. Many of us use cloud services daily. We use cloud hosting apps to update our social media profile, enjoy new subscription series or check our bank balance. Instead of being stored on our hard drives or devices, we can access these applications online. Cloud technology enables businesses to rapidly expand and adapt, accelerate innovation, increase business agility, streamline operations, and reduce costs. Cloud technology also helps companies cope with the current situation and costs. Global-scale, performance, productivity, reliability, security, and speed are the top benefits of cloud computing. Depending on the cloud services, cloud computing helps do the following: [35] [36]

1. **Lower IT costs**

   With the cloud, we can completely or partially eliminate the costs and effort of buying, installing, configuring, and managing our internal infrastructure

2. **Optimize speed and valuable time**

   With the cloud, a company can start using commercial applications in minutes instead of waiting weeks or months for their IT department to receive inquiries, purchases, and configuration of support hardware and installation of software. By leveraging the cloud, specific users such as developers and data scientists can help themselves with software and infrastructure support

3. **Easier and cheaper to scale**

   Cloud provides flexibility; instead of buying additional idle capacity during slow periods, we can scroll up and down to increase or decrease traffic. We can also use the global network of cloud service providers to distribute our applications to worldwide users.

Virtualization enables cloud providers to get the most out of their data center resources. Unsurprisingly, many organizations adopt their in-house cloud delivery model to maximize usage and cost savings over traditional IT infrastructure while providing end-users the same self-service and flexibility. [35] [36]

3.3 Cloud computing services

IaaS (Infrastructure as a Service), PaaS (Platform as a Service), and SaaS (Software as a Service) are three popular cloud service models, and it is not uncommon for a company to use all three. However, there is often confusion between all three and what is in each

SaaS (Software as a Service):

It allows users to access software applications without downloading, installing, or storing the software and its various components on their device or hard drive. Most cloud computing programs rely on this subscription for an annual or monthly fee. Instead, users get integrated solutions and functionality without the need for hardware as they get stuck installing updates or other maintenance tasks. When it was discovered, Salesforce was one of the earliest cloud computing and SaaS companies. Sales Cloud, Marketing Cloud, and Service Cloud are all cloud-based software applications. [37]

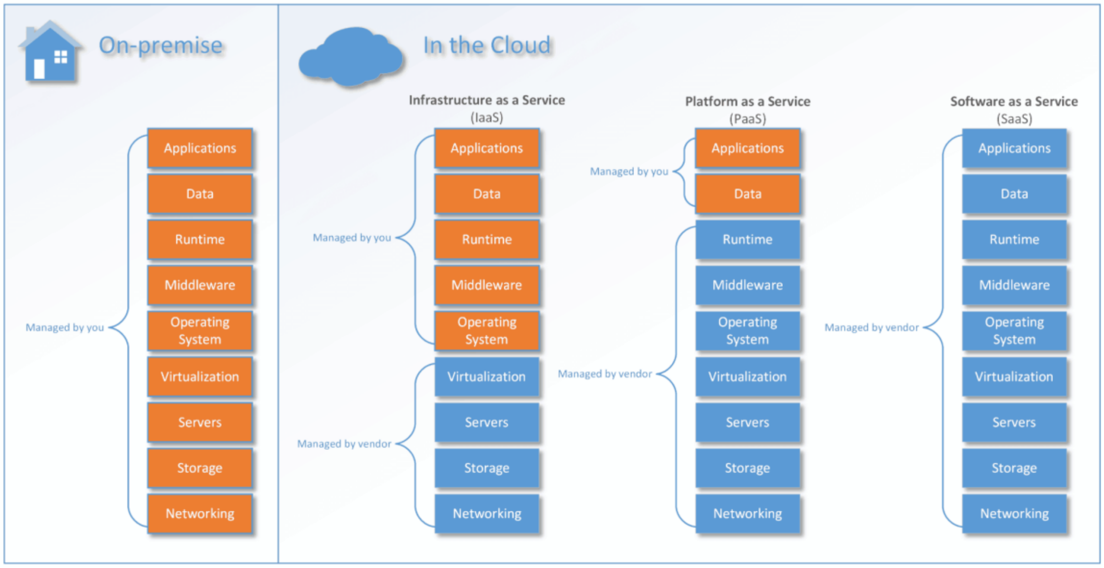


Figure 22: Saas (Software as a service)

PaaS (Platform as a Service):

The Salesforce platform is the world's leading solution as a service (PaaS). It is the most powerful way to quickly build our app and make it available to users, all with cloud power. Designing an application is a difficult task, and user interaction, experience, graphic design, and development need to be considered. Everyone should know the purpose of the application and the business's goals start it. That is where PaaS comes in. PaaS enables users to develop, run, and manage applications without the infrastructure required to build applications. As PaaS is Salesforce's No.1 platform worldwide, At the same time, Salesforce maintains the infrastructure to execute the job. PaaS offers users the flexibility they need with reliable and scalable hardware and software. [37]

IaaS (Infrastructure as a Service):

IaaS provides on-demand access to critical computing resources such as physical and virtual servers, networks, and storage over the Internet on a service fee basis. IaaS enables end-users to scale and reduce resources as needed, eliminates high initial capital costs or unnecessary internal or "abbreviated" infrastructure, and saves many resources to adapt to growing usage periods. Unlike SaaS and PaaS (and even newer PaaS computing models like containers and serverless), IaaS gives users the least control over computing resources. IaaS was popular in the early 2010s. While the cloud model continues for many types of workloads, the adoption of SaaS and PaaS is growing much faster. [37]

Serverless computing:

Simply serverless offloads backend infrastructure management, tasks provisioning, scaling, scheduling, and patching to the cloud provider. It helps the developers to focus on the coding time and effort and business logic specific to their applications. In addition, the server supports the program code based on every request and infrastructure and automatically supports the answer to the number. With the serverless, customers pay only for resources when executed and never pay for idle capacity.

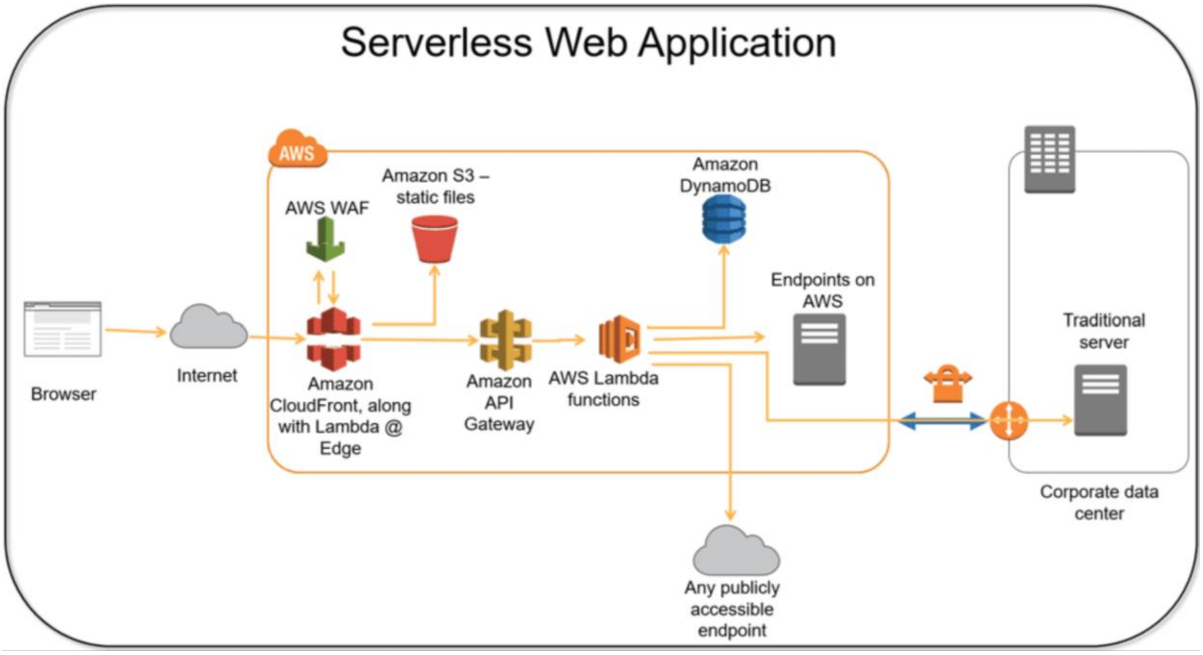


Figure 23: Serverless computing structure

Faas (Function as a service)

It is often misunderstood with serverless computing as a subset of serverless. Faas allows developers to execute specific parts of application code called functions in response to special events. Everything from the code includes physical hardware, virtual machine operating system, and web server software management. It is provisioned automatically through real-time cloud service providers in which the code runs as soon as the implementation is completed. Billing starts when execution begins and stops when execution stops. [37]

3.4 Types of cloud computing

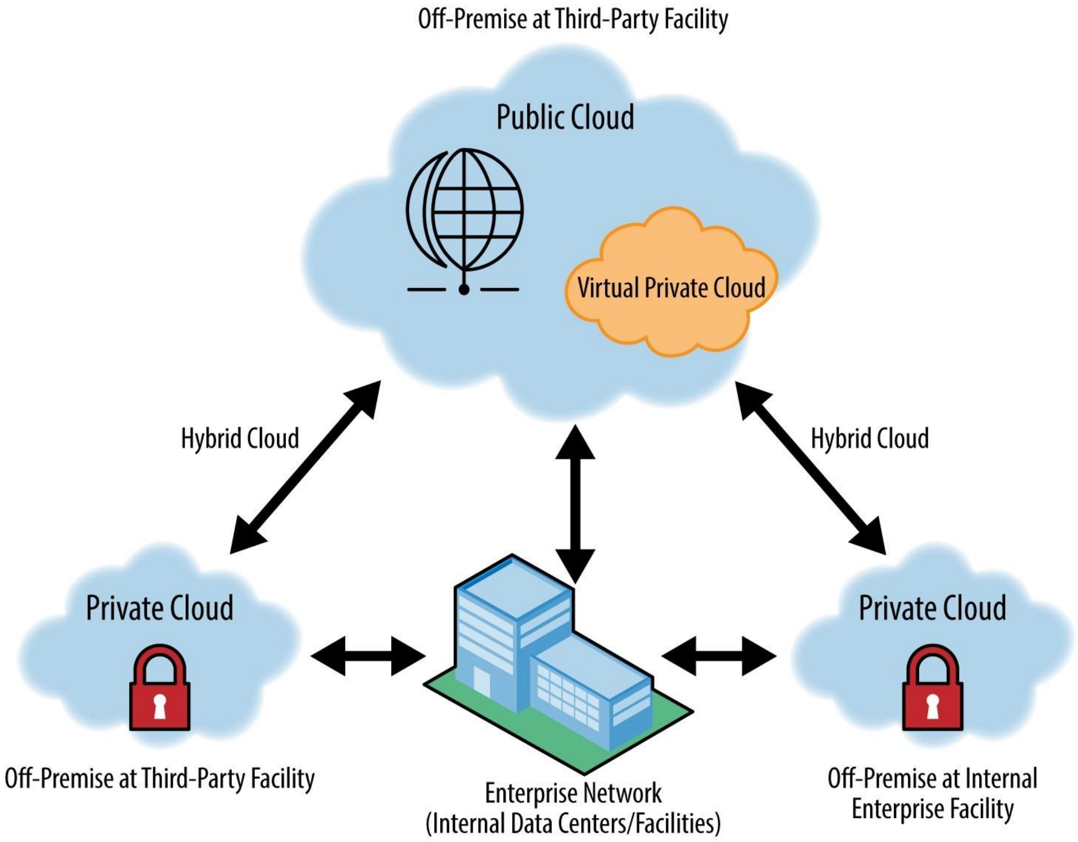


Figure 24: Private, Public, and hybrid cloud integration *[38]*

Public Cloud

The public cloud service provides everything from SaaS programs to individual virtual machines (VMs) for bare-metal computers to complete the development-based infrastructure available for public online users. These resources can be accessed for free or sold according to the price models based on subscription or payment for each use. The general cloud provider is managed and takes responsibility for all data centers, devices, and infrastructure. Its customer's workload usually has a high bandwidth network connection to ensure high performance and quick access to Provides programs and data. The public cloud is a multi-rural demolition to all public cloud customers share the cloud data center infrastructure. Amazon (AWS), Google Cloud, IBM Cloud, Microsoft Azure, and Oracle Cloud are numbered millions in general management web services. The global cloud computing market has proliferated in recent years, and analysts predict this process; Gartner Industry analysts expect overall cloud income worldwide to reach the end of 2022 $ 330 (link beyond IBM). Different settings from our computer infrastructure go to the public cloud because public cloud services are flexible, easily compressed, and flexible to meet the interaction instructions. Others are attracted with more promises and less than lost sources because customers pay only for what they are using. Others still try to cut the costs of devices and buildings. [37] [39]

Private cloud

In a private cloud, all cloud infrastructure and computing resources are assigned and used by a single customer. A private cloud combines the benefits of cloud computing, including flexibility, scalability, and serviceability, with access control, security, and resource allocation for the on-site infrastructure. The private cloud is usually hosted locally in the customer data center. A private cloud is hosted on the infrastructure of an independent cloud provider or built on the rental infrastructure of an external data center. Many companies prefer the private cloud to the public cloud because it is the easiest (or only) way to meet their compliance needs. Others choose a private cloud because their workload involves confidential documents, intellectual property, personal identification information (PII), medical, financial, or other sensitive information. Building a private cloud based on indigenous cloud principles gives a company the flexibility to move workloads to run them in a hybrid cloud environment when they are ready [37] [39].

Hybrid cloud

A hybrid cloud is precisely a combination of public and private cloud environments. In particular and ideally, a hybrid cloud combines corporate, private services, and public clouds into a single, flexible infrastructure for the operation of business applications and workloads. The purpose of a hybrid cloud is to create a combination of public and private cloud resources, with a level of orchestration between them that gives the company the flexibility to choose the optimal cloud or workload and move the workload freely between these two clouds as circumstances change. The cloud movement enables the company to achieve its technical and business goals more effectively and cheaply than the public or private cloud alone [37] [39].

Multicloud and hybrid multi-cloud

Multicloud uses two or more clouds from two or more different cloud providers. A multi-cloud environment is simple as using SaaS email from one provider and hosting SaaS from another. However, when organizations talk about multiple clouds, they usually use multiple cloud services from two or more leading public cloud providers, including SaaS, PaaS, and IaaS. In one study, 85% of companies used multi-cloud environments.

Hybrid Multicloud combines the use of two or more public clouds with a private cloud environment. Organizations choose multicolored to prevent providers from being locked in, have more services, and access more innovation. However, the more clouds we use, each with its management tools, data transfer speeds, and security protocols, the more difficult it becomes to manage our environment. It provides visibility across multiple provider clouds from a central dashboard where development teams can view their projects and deployments, operations teams can monitor clusters and nodes, and cybersecurity personnel can monitor threats. [37] [39]

3.5 Cloud security

Traditionally, security concerns have been the biggest obstacle for companies considering cloud services, especially public cloud services. However, in response to demand, the security provided by cloud service providers has consistently overtaken internal security solutions. According to security software provider McAfee, 52% of companies today have better cloud security than on-premises. Gartner predicts that infrastructure as a Service (IaaS) may have 60% fewer cloud workloads by 2022 than traditional security incidents in data centers. However, the security of the cloud requires different methods and skills from employees than in older IT environments. Some of the best cloud security practices are [37]

1. **Shared responsibility for security**

   The cloud provider is responsible for securing the cloud infrastructure, and the customer is responsible for protecting their data in the cloud. Still, it is essential to clearly define and share ownership of the data between private and public third parties.

2. **Data encryption**

   Data must be encrypted while at rest, in transit, and use. Customers must retain complete control of the security keys and hardware security module.

3. **User Identity and Access Management**

   Customers and IT teams need to fully understand and access the network, device, application, and data.

4. **Collaborative management**

   Collaborative management ensures IT operations, security teams, integrated, secure, and sustainable cloud integration processes.

5. **Security and compliance monitoring**

   It starts with understanding all legal compliance standards applicable in our industry and monitoring all connected systems and cloud-based services to ensure the transparency of all data exchange in public, private, and hybrid cloud environments.

3.6 Cloud computing and monitoring tools

Cloud monitoring uses manual tools to manage the cloud computing architecture, services, and infrastructure. These tools help us understand how the cloud is managed. The general cloud management strategy enables managers to use cloud-based resources. We can avoid minor problems with the help of these cloud computing tools, which help identify and correct errors and problems immediately. The following is the list of cloud monitoring or cloud tools. [40]

1. **Cloudwatch**

   Amazon Web Services is one of the best cloud computing management systems. Amazon services can be performed with cloud resources and applications. We can track and display instances that give us a brief overview of the general health and performance of the system. This system can be used in a unique way to improve business operations. The best thing about cloud services is that we do not have to install any software or invest in massive installation services. That is the most fantastic thing about multi-cloud management strategies. Amazon provides all security services in any form or in the event of problems. [40]

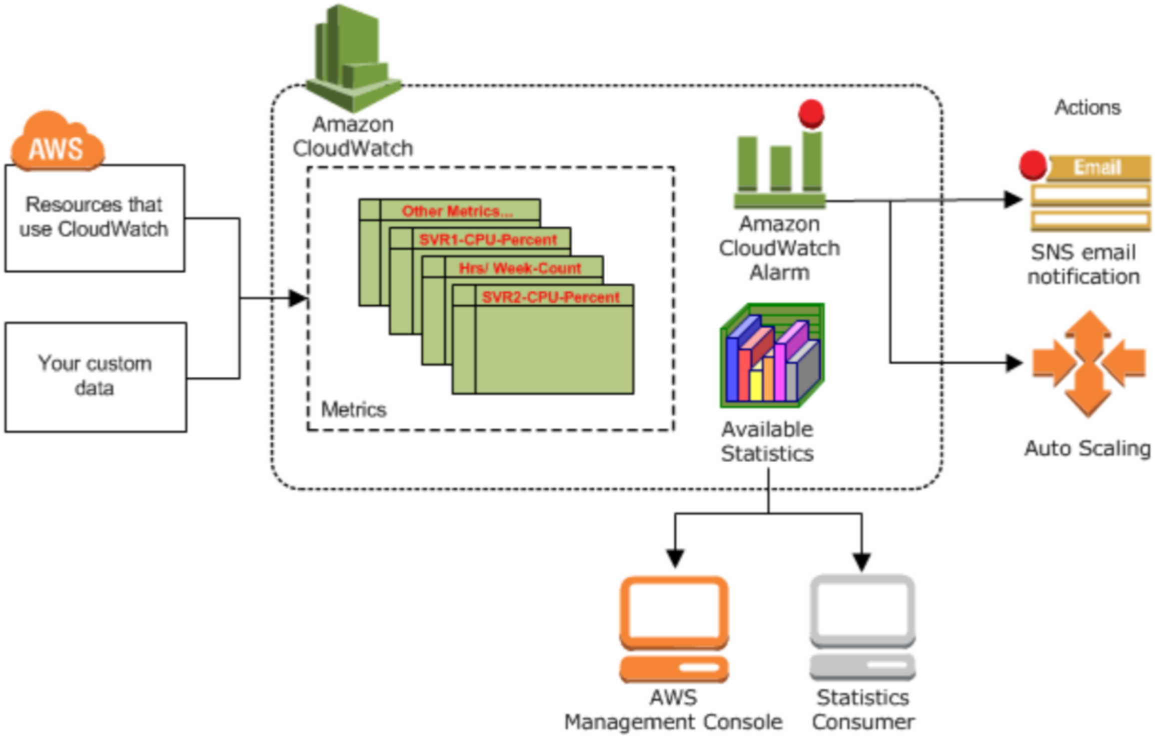
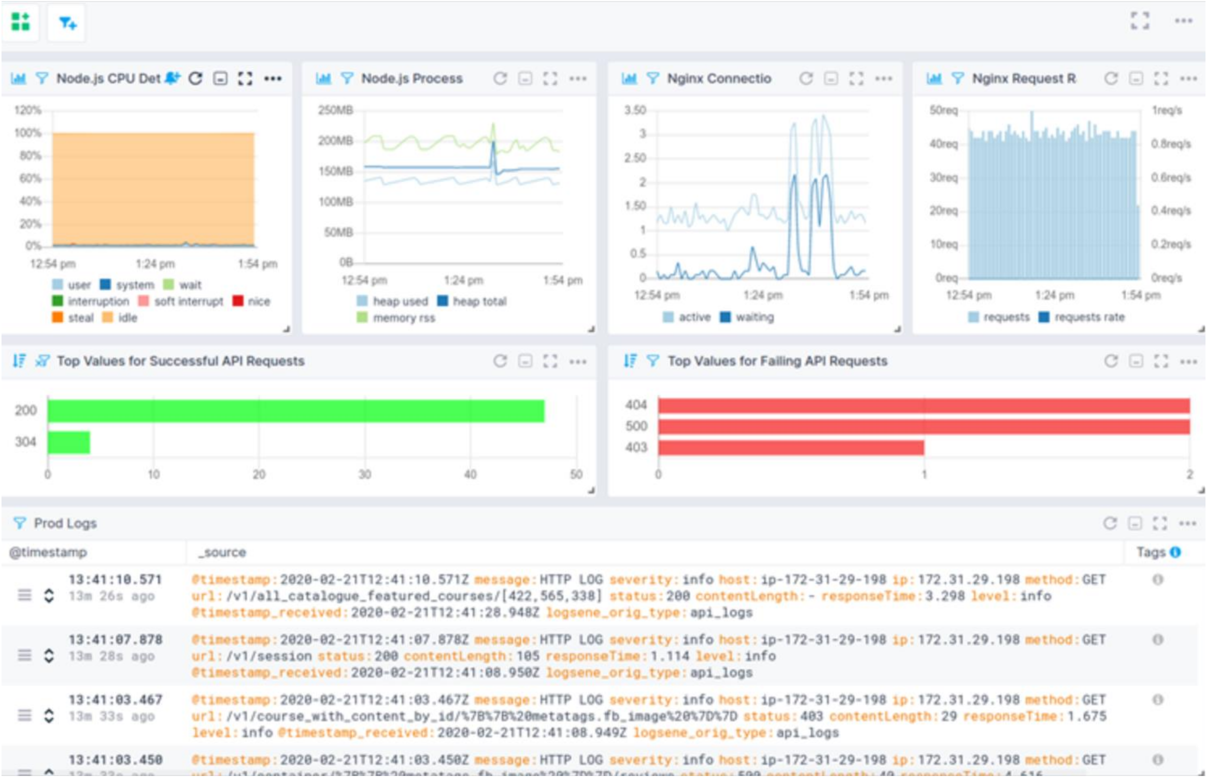


Figure 25: Cloudwatch

2. Cloud monitoring tool

The popular Microsoft Cloud Monitoring Tool that runs on Azure is a cloud virtualization tool that monitors our workload. It gives us good visibility into our workload to monitor and analyze reports and identify security risks. The main advantage of cloud monitoring tools is their ease of installation and transparency. The main similarity of these cloud monitoring tools is that they do not require us to install additional software for an extra charge. [40]



```
13:41:10.571   @timestamp:2020-02-21T12:41:10.571Z message:HTTP LOG severity:info host:ip-172-31-29-198 ip:172.31.29.198 method:GET
13m 26s ago    url:/v1/all_catalogue_featured_courses/[422,565,338] status:200 contentLength:- responseTime:3.298 level:info
               @timestamp_received:2020-02-21T12:41:28.948Z logsene_orig_type:api_logs
13:41:07.878   @timestamp:2020-02-21T12:41:07.878Z message:HTTP LOG severity:info host:ip-172-31-29-198 ip:172.31.29.198 method:GET
13m 28s ago    url:/v1/session status:200 contentLength:105 responseTime:1.114 level:info
               @timestamp_received:2020-02-21T12:41:08.950Z logsene_orig_type:api_logs
13:41:03.467   @timestamp:2020-02-21T12:41:03.467Z message:HTTP LOG severity:info host:ip-172-31-29-198 ip:172.31.29.198 method:GET
13m 33s ago    url:/v1/course_with_content_by_id/%7B%7B%20metatags.fb_image%20%7D%7D status:403 contentLength:29 responseTime:1.675
               level:info @timestamp_received:2020-02-21T12:41:08.949Z logsene_orig_type:api_logs
13:41:03.450   @timestamp:2020-02-21T12:41:03.450Z message:HTTP LOG severity:info host:ip-172-31-29-198 ip:172.31.29.198 method:GET
```

Figure 26: Cloud monitoring tool

3. Infrastructure operator

DX Infrastructure Manager is a management platform that provides intelligent analytics to monitor infrastructure. It is a great tool that is fully active in troubleshooting and resolving issues affecting cloud infrastructure performance. It is an excellent tool for managing storage services, networks, servers. Displays specific patterns or analysis by identifying the latest trends, simplifying the troubleshooting process, and reporting various activities. It is an easy-to-use platform with customizable dashboards for better visualization and monitors every aspect of the cloud ecosystem. It can convert DX IM into a convenient incident management tool that enhances monitoring capabilities. [40]

4. AppDynamics

Cisco acquired it in 2017 as a leading cloud development tool that provides cloud management to speed up applications. It allows users to understand the actual state of cloud applications and helps with business transactions and coding levels. It is an environmentally adaptable system that has excellent features. This tool is handy for monitoring performance in the cloud. [40]

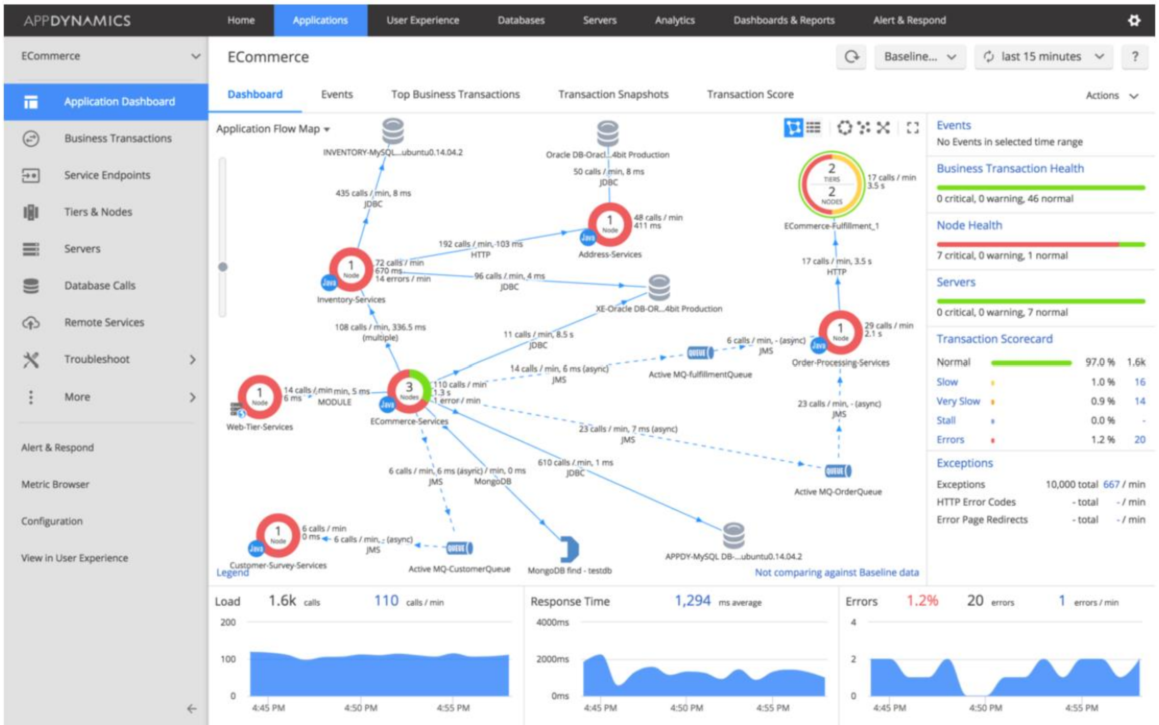


Figure 27: AppDynamics

5. Relic

New Relic is a valuable cloud computing tool for managing complex, ever-changing applications. It helps the server run in real-time, taking care of the problem and solving it quickly. It also lets us measure our operation with usage. It reviews various application implementation and upgrades, whether a mobile application or a web-based application. It helps place all data in a dashboard in one place, giving us a clear idea of each part of cloud computing. [40]

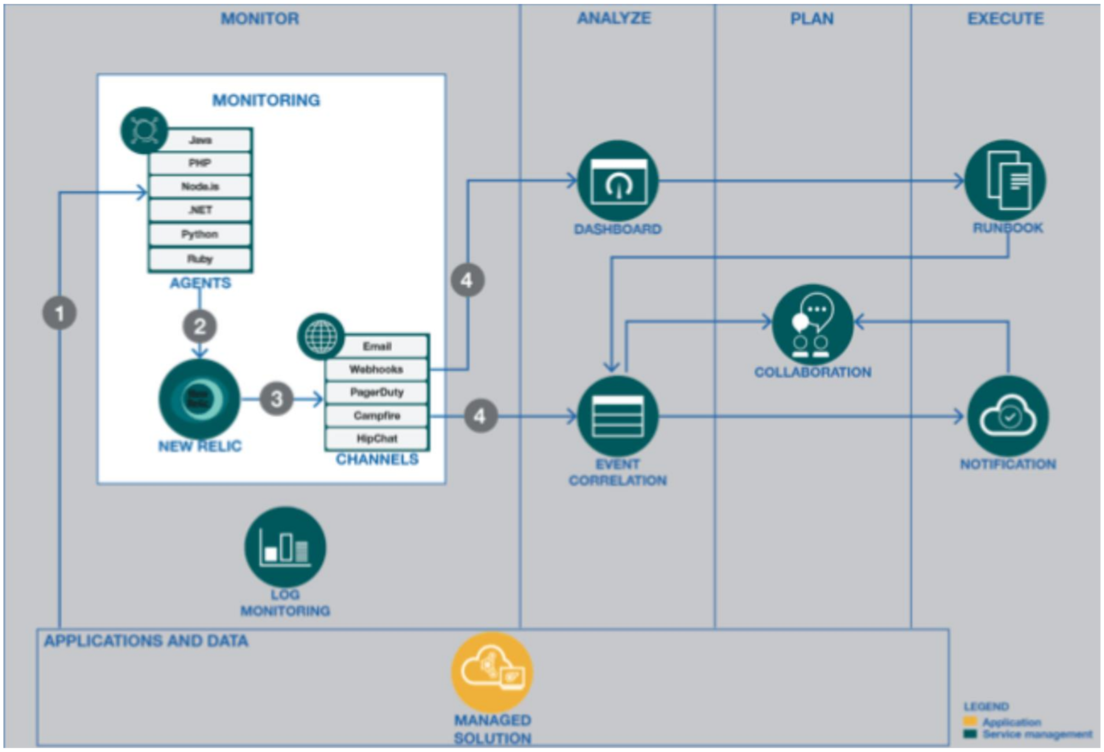
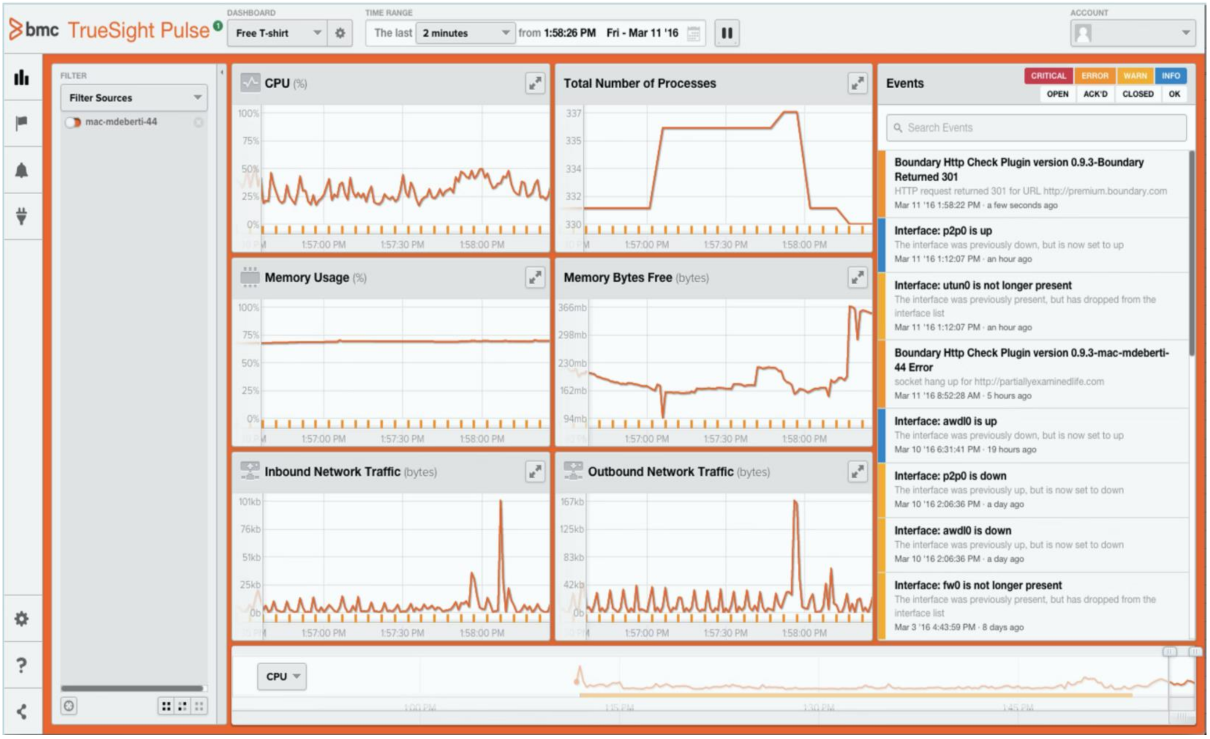


Figure 28: Relic structure

6. True Sight Pulse

BMC true sight pulse is another cloud monitoring tool that helps improve performance, operational control, and cost management settings. It gives us the end-user experience to solve problems and track resource issues. It makes it possible to build management tools for cloud operations. It also allows us to manage, improve, and control costs efficiently. It helps break down costs that companies can use later and invest in business needs. [40]



Boundary Http Check Plugin version 0.9.3-Boundary Returned 301
HTTP request returned 301 for URL http://premium.boundary.com
Mar 11 '16 1:58:22 PM · a few seconds ago

Interface: p2p0 is up
The interface was previously down, but is now set to up
Mar 11 '16 1:12:07 PM · an hour ago

Interface: utun0 is not longer present
The interface was previously present, but has dropped from the interface list
Mar 11 '16 1:12:07 PM · an hour ago

Boundary Http Check Plugin version 0.9.3-mac-mdeberti-44 Error
socket hang up for http://partiallyexaminedlife.com
Mar 11 '16 8:52:28 AM · 5 hours ago

Interface: awdl0 is up
The interface was previously down, but is now set to up
Mar 10 '16 6:31:41 PM · 19 hours ago

Interface: p2p0 is down
The interface was previously up, but is now set to down
Mar 10 '16 2:06:36 PM · a day ago

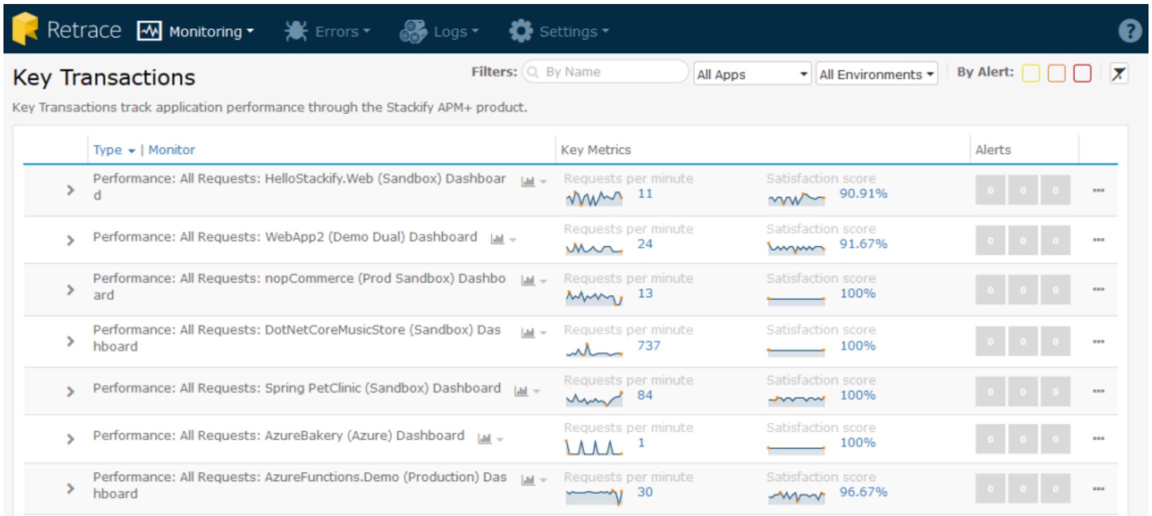
Interface: awdl0 is down
The interface was previously up, but is now set to down
Mar 10 '16 2:06:36 PM · a day ago

Interface: fw0 is not longer present
The interface was previously present, but has dropped from the interface list
Mar 3 '16 4:43:59 PM · 8 days ago

Figure 29: True Sight Pulse

7. Solar winds

This monitoring tool provides cloud monitoring services, networking, and database solutions. This cloud management platform allows us to monitor applications, servers, and the performance and health of virtual machines. It is encrypted with the management tool to control the cloud environment. [40]



Key Transactions

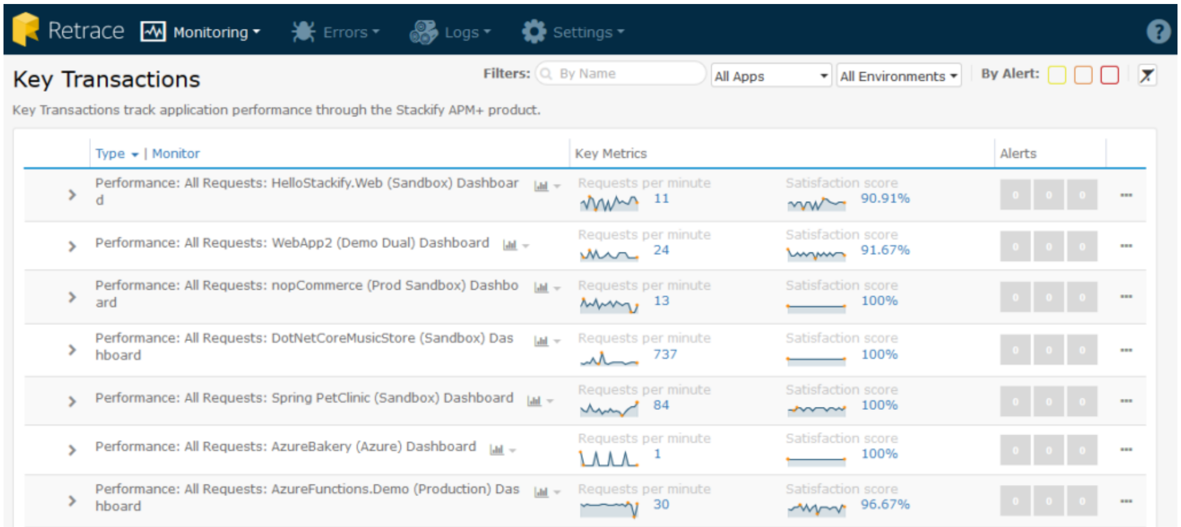
Key Transactions track application performance through the Stackify APM+ product.

Type ▾ | Monitor	Key Metrics		Alerts
Performance: All Requests: HelloStackify.Web (Sandbox) Dashboard	Requests per minute 11	Satisfaction score 90.91%	0 0 0
Performance: All Requests: WebApp2 (Demo Dual) Dashboard	Requests per minute 24	Satisfaction score 91.67%	0 0 0
Performance: All Requests: nopCommerce (Prod Sandbox) Dashboard	Requests per minute 13	Satisfaction score 100%	0 0 0
Performance: All Requests: DotNetCoreMusicStore (Sandbox) Dashboard	Requests per minute 737	Satisfaction score 100%	0 0 0
Performance: All Requests: Spring PetClinic (Sandbox) Dashboard	Requests per minute 84	Satisfaction score 100%	0 0 0
Performance: All Requests: AzureBakery (Azure) Dashboard	Requests per minute 1	Satisfaction score 100%	0 0 0
Performance: All Requests: AzureFunctions.Demo (Production) Dashboard	Requests per minute 30	Satisfaction score 96.67%	0 0 0

Figure 30: Solarwind real-time cloud monitoring

8. Retrace

It is considered the top cloud computing tool for developers and designers to help with design development and code. It helps to track execution and ensures developers make advanced coding at any time. These tools aim to make developers more productive with fewer complications and make the developer and designer more manageable with faster work. It is suitable for small and medium-sized enterprises. [40]



Key Transactions

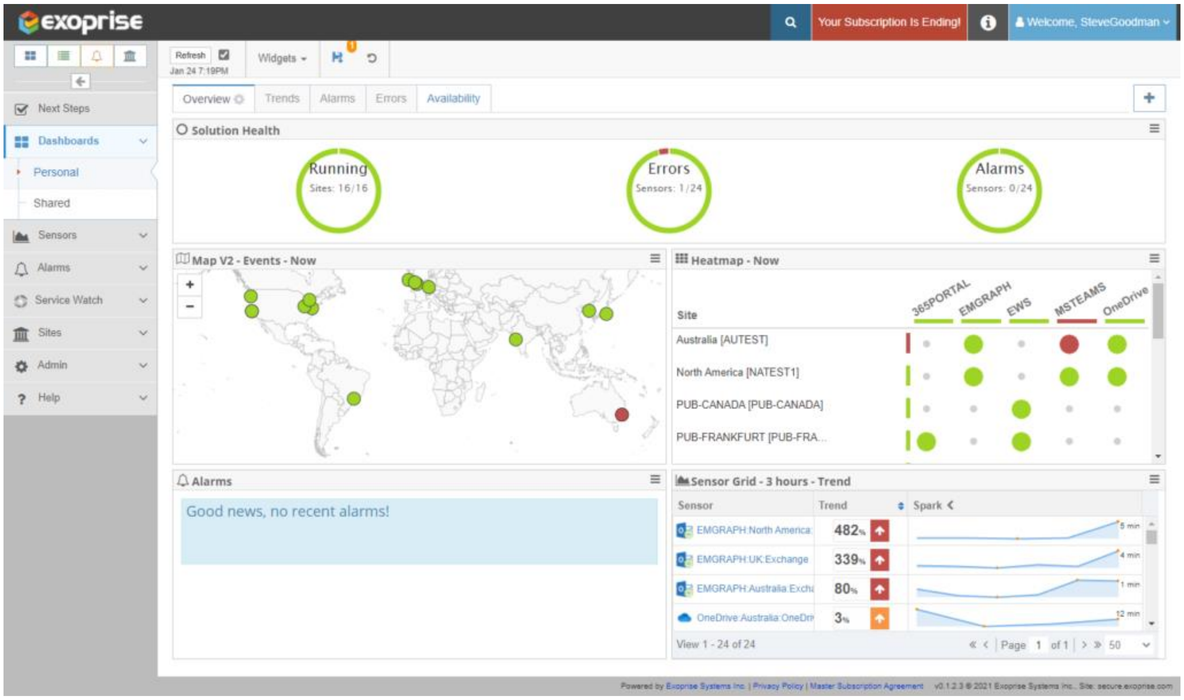
Key Transactions track application performance through the Stackify APM+ product.

Type | Monitor	Key Metrics		Alerts
Performance: All Requests: HelloStackify.Web (Sandbox) Dashboard	Requests per minute 11	Satisfaction score 90.91%	0 0 0
Performance: All Requests: WebApp2 (Demo Dual) Dashboard	Requests per minute 24	Satisfaction score 91.67%	0 0 0
Performance: All Requests: nopCommerce (Prod Sandbox) Dashboard	Requests per minute 13	Satisfaction score 100%	0 0 0
Performance: All Requests: DotNetCoreMusicStore (Sandbox) Dashboard	Requests per minute 737	Satisfaction score 100%	0 0 0
Performance: All Requests: Spring PetClinic (Sandbox) Dashboard	Requests per minute 84	Satisfaction score 100%	0 0 0
Performance: All Requests: AzureBakery (Azure) Dashboard	Requests per minute 1	Satisfaction score 100%	0 0 0
Performance: All Requests: AzureFunctions.Demo (Production) Dashboard	Requests per minute 30	Satisfaction score 96.67%	0 0 0

Figure 31: Retrace tool by Stackify

9. Exoprice

The SaaS monitoring tools or services provide security and service optimization of cloud applications. The application tools like Dropbox and Salesforce watch and manage all the tools with troubleshooting and fixing problems for devices that may directly or indirectly affect business, business platform, and company platforms such as Starbucks, PayPal, and P&G. [40]



Site	365PORTAL	EMGRAPH	EWS	MSTEAMS	OneDrive
Australia [AUTEST]	·	🟢	·	🔴	🟢
North America [NATEST1]	·	🟢	·	🟢	🟢
PUB-CANADA [PUB-CANADA]	·	·	🟢	·	·
PUB-FRANKFURT [PUB-FRA...	🟢	·	🟢	·	·

Good news, no recent alarms!

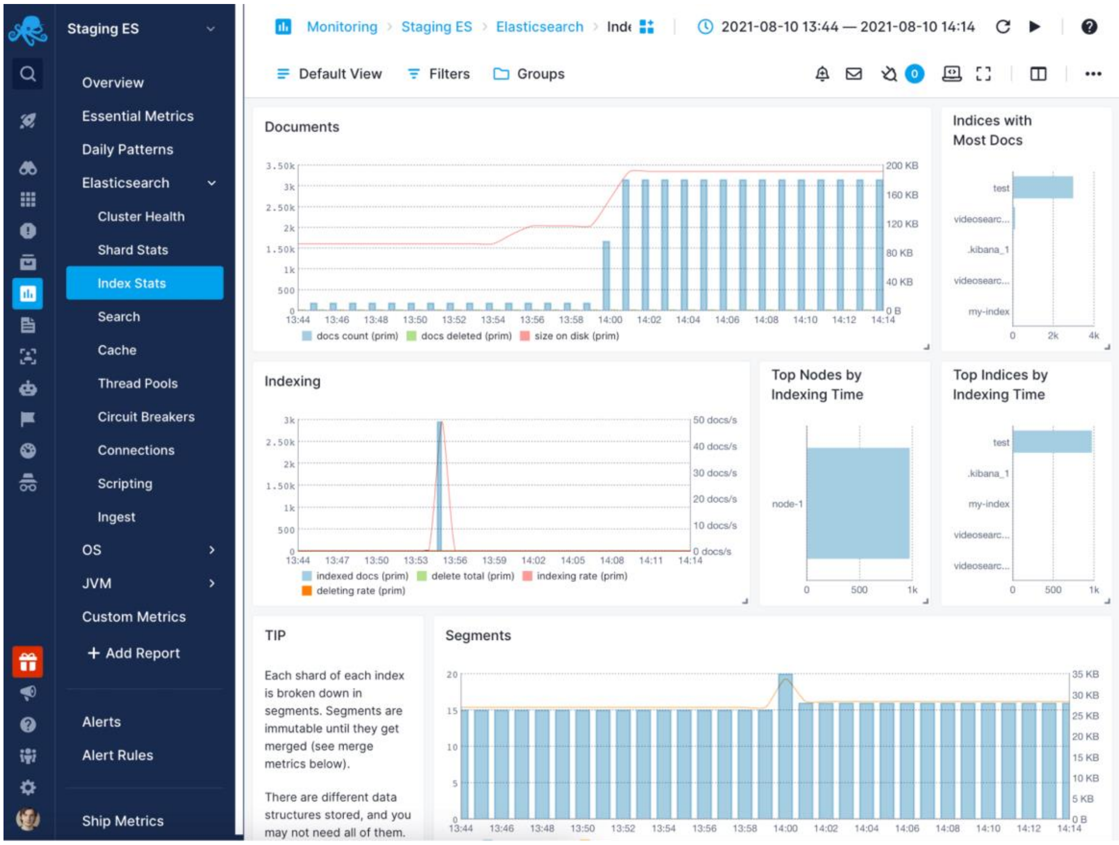
Sensor	Trend	Spark
EMGRAPH:North America:	482% ↑	5 min
EMGRAPH:UK:Exchange	339% ↑	4 min
EMGRAPH:Australia:Exch	80% ↑	1 min
OneDrive:Australia:OneDri	3% ↑	12 min

View 1 - 24 of 24 Page 1 of 1 50

Figure 32: Exoprise Saas cloud monitoring

10. Sematext

It is a valuable tool for cloud computing that collectively manages performance and solutions available in the cloud and on-premises. It allows users to diagnose and effectively solve performance issues and the latest trends for the end-user. [40]



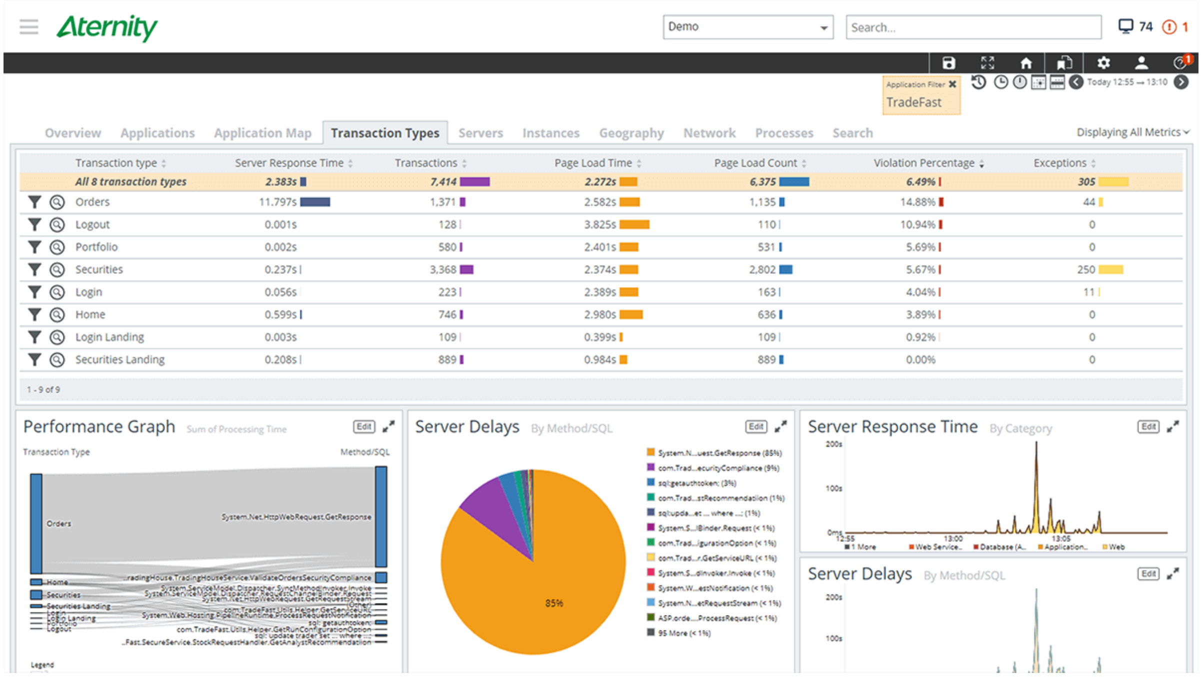
Each shard of each index is broken down in segments. Segments are immutable until they get merged (see merge metrics below).

There are different data structures stored, and you may not need all of them.

Figure 33: Sematext cloud

11. Aternity

This cloud is created specifically for the end-user. It ranks number one in the virtual screen system with the mobile end-user experience. Various tests are performed to operate such a system and verify the proper functionality. It determines the loading time of the site along with any traffic problems. It also improves customer interaction. It provides a comprehensive list of tools that most effectively improve user experience. [40]



Transaction type	Server Response Time	Transactions	Page Load Time	Page Load Count	Violation Percentage	Exceptions
All 8 transaction types	*2.383s*	*7,414*	*2.272s*	*6,375*	*6.49%*	*305*
Orders	11.797s	1,371	2.582s	1,135	14.88%	44
Logout	0.001s	128	3.825s	110	10.94%	0
Portfolio	0.002s	580	2.401s	531	5.69%	0
Securities	0.237s	3,368	2.374s	2,802	5.67%	250
Login	0.056s	223	2.389s	163	4.04%	11
Home	0.599s	746	2.980s	636	3.89%	0
Login Landing	0.003s	109	0.399s	109	0.92%	0
Securities Landing	0.208s	889	0.984s	889	0.00%	0

1 - 9 of 9

Performance Graph Sum of Processing Time
Edit
Transaction Type
Method/SQL
Orders
System.Net.HttpWebRequest.GetResponse
Home
Securities
Securities Landing
Login
Login Landing
Portfolio
Logout
Legend
Server Delays By Method/SQL
Edit
85%
System.N...uest.GetResponse (85%)
com.Trad...ecurityCompliance (9%)
sql:getauthtoken; (3%)
com.Trad...stRecommendation (1%)
sql:upda...et ... where ...; (1%)
System.S...lBinder.Request (< 1%)
com.Trad...igurationOption (< 1%)
com.Trad...r.GetServiceURL (< 1%)
System.S...dInvoker.Invoke (< 1%)
System.W...estNotification (< 1%)
System.N...etRequestStream (< 1%)
ASP.orde....ProcessRequest (< 1%)
95 More (< 1%)
Server Response Time By Category
Edit
200s
100s
0ms
12:55
13:00
13:05
1 More
Web Service...
Database (A...
Application...
Web
Server Delays By Method/SQL
Edit
200s
100s

Figure 34: Application performance monitoring

3.6.1 Cloud Use cases

With 25% of organizations planning to change the application to the cloud, the cloud computing use cases are limitless. Disaster recovery and business continuity are natural for the cloud. The cloud provides cost-effective redundancy to protect data against system failures and the physical distance required to recover data and applications in local outages or disasters. The primary public cloud provides Disaster Recovery as a Service (DRaaS). The target of cloud computing is anything that involves storing and processing huge volumes of data at high speeds requires vast storage and computing capacity that more companies want to purchase and deploy on-premises. [37]

Big data analytics, IoT, AI, and ML are examples and targets for cloud computing services. For teams adopting Agile or DevOps (or DevSecOps) to streamline development, the cloud offers the on-demand end-user self-service that keeps operations tasks such as spinning up development and test servers from becoming development bottlenecks. [37]

3.7 Cloud Computing Trends

Cloud allows businesses to expand and adapt quickly. Cloud computing accelerates innovation, drives business agility, streamlines operations, and lowers costs. It can help firms get through the present situation and enhance long-term growth. [35]

1. **Cloud AI**

One of the latest trends to access massive datasets is the AI-enabled cloud. Machine learning can use these data to improve their crucial capability. Artificial intelligence is used to tackle the problems associated with cloud-related challenges. The only way to scale up AI systems that create enormous data is through cloud computing services. Cloud computing enables AI to perform quicker computations and better resource management. [35]

2. **Improving Saas operations**

   Specialist solutions such as BetterCloud, Cloud Manager, and others are developing to handle migrations and operations as Saas is very popular. It enables the administration of whole solutions suites such as Google G Suite, Microsoft Office 365, and other known SaaS solutions. [35]

3. **Containerization by industry giants**

   Soon, the industries may embrace containerization technologies quickly and successfully. The cloud computing sector may begin working on containerization efforts asap. Big corporations like Amazon and Microsoft invest in their containerization software suites. [35]

4. **Enhanced Security**

   Security is the most significant issue of every organization, and they do not want to rely on a third-party security solution. Companies cannot afford or do not have enough resources to implement in-house security solutions, and therefore they have to rely on third parties somehow. Cloud service providers have better security choices for their clients in the upcoming years. [35]

5. **Kubernetes**

   It is a management solution developed by Google which has a massive impact on the worldwide economy. It is embraced by many corporate sectors to overcome the limitations of the cloud. Kubernetes enhances enthusiasm and makes industry aspirations a reality through its unique offering from cloud infrastructure providers. [35]

6. **Intelligent SaaS**

   Machine learning is the most widely used technology in all SaaS and IT processes. AI and ML have changed how people function as it has beneficial qualities. No other product would be regarded as clever in the future. [35]

7. **Kubernetes Supremacy**

   Kubernetes is gaining attention across the board, so it is critical for cloud infrastructure providers to keep up. Any business that considers Kubernetes competes in the marketplace. Microsoft has purchased Deis to extend its Kubernetes toolset; Net app bought StackPointCloud. [35]

8. **Quantum computing**

The performance of a computer gets better in the future years. However, to make this viable, quantum computing is required so that computers can transfer information at a quicker rate. In this aspect, cloud computing is equally crucial for increasing processing power and performance. [35]

9. Multi-cloud to Omni cloud

Multilevel cloud in recent years has strived much traction in the commercial world. Many businesses started using Cloud computing in their operation. Portable applications enable smoother communication for the multi-cloud to become omni cloud in the upcoming years. [35]

10. Integrated Blockchain Technology

It is a technology that allows companies to track the phases of the product's lifecycle. When the cloud is combined with Blockchain technology, companies can understand the associated characteristics like information such as product pollution and weather delay. [35]

11. The appearance of private cloud

Earlier companies approached the public without apparent purpose or guidance. However, after the company identified the security risks and expenses of using the public cloud, it became less attractive. Rebuilding private cloud infrastructure drives the demand for improved security and efficiency. [35]

12. Cloud-native applications

The business must provide products through the use of the cloud because of the evolving cloud-based solutions. The growth of cloud-native apps accelerates open source project management platforms. [35]

Chapter 4: 5G, MIMO, 3GPP

4.1 Leveraging AI and ML for 5G

The heterogeneous nature of wireless networks includes multiple access networks, frequency bands, and cells with overlapping coverage areas, current wireless operators, network planning, and deployment challenges. ML and AI can help wireless operators meet these challenges by analyzing geographic information, technical parameters, and historical data. AI and ML with 5G multi-access edge computing (MEC) allow wireless operators to provide

- High levels of automation with distributed ML and AI architectures at the edge of the network
- Application-based traffic management and aggregation based on programs on different access networks
- Dynamic network slicing to handle different use cases with different QoS requirements
- ML/AI as a Service is offered to end-users [41]

AI and ML for Beamforming

5G has beam-based cell coverage deployed with the mm waves, while 4G is based on sector-based coverage. A machine-learned algorithm assists the 5G cell site in computing a good set of candidate beams, beginning either from the serving or its neighboring cell site. An ideal set contains fewer beams and is highly likely to contain the best beam. The best beam has the highest signal strength or RSRP. When the activated beams are present more, the probability is higher to find the best beam. Although higher the number of activated beams increases the consumption of system resources.

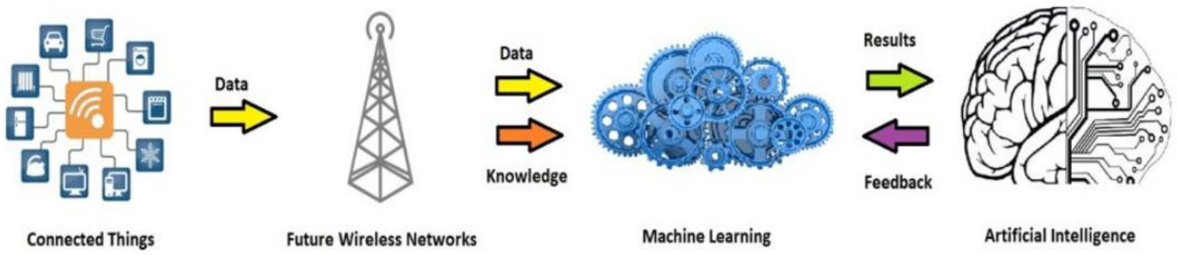


Figure 35: 5G enabling simultaneous connections to multiple IoT devices *[41]*

The user equipment (UE) measures all candidates' beams to the serving cell site. The UE is handed off to the neighboring cell site that candidates the beam. The UE reports that the Beam State Information (BSI) is based on the measurements of Beam Reference Signal (BRS), including parameters such as Beam Index (BI) and Beam Reference Signal Received Power (BRSRP). The best beam is calculated using BRSRP. It leads to a multi-target regression (MRT) problem, but finding the best beam using BI leads to a multi-class classification (MCC) problem.

AI and ML assist in finding the best beam by considering instantaneous values updated at every UE measurement of the parameters listed.

1. Beam Index (BI)
2. Beam Reference Signal Received Power (BRSRP)
3. Distance (of UE to serving cell site)
4. Position (GPS location of UE)
5. Speed (UE mobility)
6. Channel quality indicator (CQI)
7. Historical values

Historical values are based on past events and measurements. They include previous serving beam information, time spent on each serving beam, and the distance trends. Once the UE identifies the best beam, the random access procedure connects to the beam using timing and angular information. Once the UE connects to the beam, the data session begins on the dedicated UE-specific beam. [41]

AI and ML for Massive MIMO

Massive MIMO is a key to 5G technology that comprises many antennas like 32 or more logical antenna ports in the base station of the antenna array. Massive MIMO helps improve a significant user experience center by significantly increasing throughput, network capacity, and coverage with reduced interference that can be done by serving multiple spatially separated users with an antenna array at the same time and with frequent resources.

Serving some users with beamforming steers a narrow beam with high gain. Radio signals and information are directly sent to the device instead of broadcasting across the entire cell by reducing radio interference. The weights of antenna elements for a massive MIMO 5G cell site are critical for maximizing the beamforming effect. AI and ML are used dynamically to identify changes and forecast user distribution by analyzing historical data. Historical data optimizes the weights of antenna elements. Adaptive weights can be performed for specific use cases with unique user distribution. Thereby improves the coverage in a multicell scenario by considering the intersite interference between multiple 5G massive MIMO cell sites. [41]

AI and ML for Network Slicing

Most of the resources of the wireless networks are still limited and not optimistic for high bandwidth and low latency scenes. To use the available network resources fixed resource assignment for diverse applications with differential requirements is not an efficient approach. Network slicing helps create multiple dedicated virtual networks using a shared physical infrastructure. Each network slice is independently managed and orchestrated. Embedding ML and AI into 5G networks increases automation and adaptability and enables efficient network slice orchestration and dynamic provisioning. ML and AI collect real-time information for multidimensional analysis and construct a panoramic data map of each network based on User subscription, Quality of service (QoS), Network performance, Events, and logs. Different aspects of AI and ML can be leveraged, including [41]

1. Predicts and forecasts the network resources enable wireless operators to anticipate network outages, equipment failures, and performance degradation

2. Cognitive scaling assists wireless operators in dynamically modifying network resources. Capacity requirements are based on the predictive analysis and forecasted results
3. Predicting UE mobility in 5G networks allows Access and Mobility Management Function (AMF) to update mobility patterns based on user subscription, historic statistics, and instantaneous radio conditions for optimization and seamless transition to ensure a better quality of service.
4. Enhancing the security in 5G networks prevents attacks and fraud by recognizing user patterns and tagging certain events to prevent similar attacks.
5. Heterogeneous wireless networks are implemented with different technologies that address different use cases, providing connections for millions of users who require customization per slice and service. It involves many KPIs to maintain and support ML and AI to grow future wireless operators. [41]

Deploying AL and ML into Wireless Networks

Wireless operators can deploy AI in the following three methods.

1. Embedding ML and AI algorithms in edge devices to reduce the computational capability
2. Lightweight ML and AI engines perform multi-access edge computing (MEC) for real-time computation and dynamic decision-making suitable for low-latency IoT services addressing various use cases.
3. ML and AI platforms are built within the system for centralized deployment. High computation, storage, and projections are performed. [41]

4.2 Benefits of Leveraging AI and ML in 5G

The application of ML & AI in wireless networks is currently in infancy but gradually increases in the upcoming years to create more intelligent networks. Network topology, network design, and propagation model with users' mobility and usage patterns in 5G are complex. AI & ML helps build intelligence in 5G systems and allows for a shift from managing networks to services. AI & ML plays a significant role in assisting wireless operators in deploying, operating, and managing the 5G networks with the proliferation of IoT devices. AI & ML addresses several use cases to help wireless operators transition from human management to self-driven automatic management by transforming the network operations and maintenance process. There are high synergies between AI, ML, and 5G. The sensing and processing of data are time-sensitive as they address low latency use cases. The use cases include self-driving autonomous vehicles, time-critical industry automation, and remote healthcare. 5G offers ultra-reliable low latency, and it is ten times faster than 4G. A paradigm shift is required from the current centralized and virtualized cloud-based AI. Lower latencies enable event-driven analysis, real-time processing, and decision-making toward a distributed AI architecture. [41]

4.3 AI & ML in cybersecurity Use Cases

4.3.1 CRQ and residual risk calculation

With this solution, decision-makers can precisely and objectively measure their company's cyber risk. The solution also enables companies to take a holistic, technology-based approach to business to prioritize cybersecurity investments and evaluate return on investment. In this model, cyber threats can be neutralized at different levels of strategic complexity. Notable aspects of CRQ solutions based on AI and ML are

a. **Sophisticated cyber risk scoring**

- Creates strong, empirically determined scores that are a forward-looking indicator of security risks
- It provides a unified assessment of the company's overall cybersecurity posture. This score applies to the subgroup level (geography, technology), macro-level (on all technology devices), and micro-level (score for each IP address) [41]

b. **Exhaustive assessment**

- It enables a comprehensive and almost instantaneous security assessment of all critical parameters, including technology stack, cyber-scale risk signals, topology, threat level, business priorities, legal obligations, internet asset time-series observations, and historical insights. It helps to reach the cyber risk posture. [41]

c. **Advanced modeling techniques**

- Uses industry-leading probabilistic ML security modeling techniques that use layered intelligence. They derive the relationship between a company's vulnerabilities, security configurations, and behavioral factors.
- Technologies such as hidden Markov models, random forest, Naive Bayes, support vector machines (SVMs), linear models, fuzzy c-means clustering (FCM), artificial neural networks (ANN), and k-nearest neighbor algorithms (KNN), Recursive Neural Networks (RNN) are used. [41]

d. **Advanced dashboards and reports**

- Customizable dashboards and reports are enabled for technical and non-technical stakeholders, including the CEO, CISO, CRO, CXO, and security team.
- This robust classification provides comparisons and benchmarks for cyber risk postures across the portfolio and against competitors.
- It offers other advanced features such as advanced heat mapping and intelligent drill-downs. [42]

Stakeholders and benefits of CRQ solutions based on AI and ML

Firms

a. Assess the effectiveness of their cyber risk programs and prioritize and improve cyber risk investments.
b. Optimize cyber insurance coverage

c. Assess and mitigate cyber risks, including third parties, and focus cyber concentration risks on all partners. [42]

M&A Institutions

Identify cyber risks and investment requests for potential target companies. This insight is crucial for price negotiations. [42]

Cyber insurers

Cyber insurance underwriting process and public portfolio risk management. [42]

4.3.2 Network intrusion detection and prevention

New and unexpected intruders and threats, which traditional signature-based systems cannot detect, can be identified using artificial intelligence and machine learning solutions. This new approach includes

1. Monitor inbound and outbound network traffic to detect suspicious activity and classify the type of threat.
2. Identify malware in large corporate networks
3. Enable robust network security through Ethernet, wireless, supervisory control and data acquisition (SCADA), and software-defined networks (SDN).
4. Use ML-based anomaly detection capabilities to identify and classify enterprise-wide network threats, including botnet detection and domain generation algorithms (DGA).
5. Use ML-powered network traffic analysis. It can use supervised and unsupervised learning algorithms to classify and cluster the attacks based on packet headers and data flow information, such as protocols, number of bytes, rates, and counters.
6. Offering ML techniques to help classify IP traffic. [42]

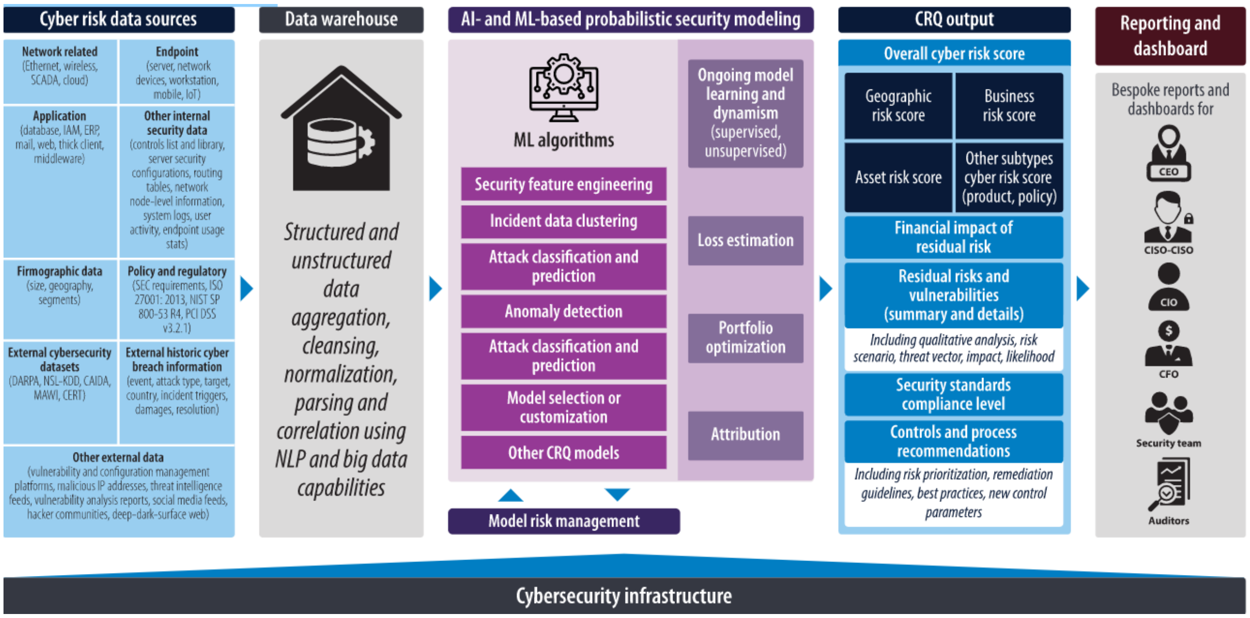

Cyber risk data sources
Network related (Ethernet, wireless, SCADA, cloud)
Endpoint (server, network devices, workstation, mobile, IoT)
Application (database, IAM, ERP, mail, web, thick client, middleware)
Other internal security data (controls list and library, server security configurations, routing tables, network node-level information, system logs, user activity, endpoint usage stats)
Firmographic data (size, geography, segments)
Policy and regulatory (SEC requirements, ISO 27001: 2013, NIST SP 800-53 R4, PCI DSS v3.2.1)
External cybersecurity datasets (DARPA, NSL-KDD, CAIDA, MAWI, CERT)
External historic cyber breach information (event, attack type, target, country, incident triggers, damages, resolution)
Other external data (vulnerability and configuration management platforms, malicious IP addresses, threat intelligence feeds, vulnerability analysis reports, social media feeds, hacker communities, deep-dark-surface web)
Data warehouse
Structured and unstructured data aggregation, cleansing, normalization, parsing and correlation using NLP and big data capabilities
AI- and ML-based probabilistic security modeling
ML algorithms
Security feature engineering
Incident data clustering
Attack classification and prediction
Anomaly detection
Attack classification and prediction
Model selection or customization
Other CRQ models
Ongoing model learning and dynamism (supervised, unsupervised)
Loss estimation
Portfolio optimization
Attribution
Model risk management
CRQ output
Overall cyber risk score
Geographic risk score
Business risk score
Asset risk score
Other subtypes cyber risk score (product, policy)
Financial impact of residual risk
Residual risks and vulnerabilities (summary and details)
Including qualitative analysis, risk scenario, threat vector, impact, likelihood
Security standards compliance level
Controls and process recommendations
Including risk prioritization, remediation guidelines, best practices, new control parameters
Reporting and dashboard
Bespoke reports and dashboards for
CEO
CISO-CISO
CIO
CFO
Security team
Auditors
Cybersecurity infrastructure

Figure 36: Functional Architecture of CRQ and residual risk calculation solution *[42]*

The following AI and ML approaches are explored to detect and prevent network intrusion

1. Unclassified examples and supervised learning algorithms are used to improve classification performance, for example, by reducing false alarms.
2. A combination of extreme learning machines and SVMs with a set of k-means clustering as an Intrusion Detection System (IDS) model is used.
3. Use the KDD'99 dataset to improve accuracy and reduce false alarms.
4. IDS is based on the SVM least-squares sample.
5. Fuzz-based semi-supervised learning.
6. Nonsymmetric deep autoencoder (NDAE) is a new deep learning-based method for network intrusion detection.
7. Genetic algorithms (GA) and fuzzy logic detect network intrusion. It uses gloss analysis to create a digital signature for the network segment. It can also predict network traffic behavior over time and assess anomalies.
8. Ant tree miner classification is a new decision tree method
9. The IDS uses binary KNN and particle swarm optimization (PSO), including feature selection and classification steps
10. PSO quick learning network. [42]

4.3.3 Case review optimization through document digitalization

Digitizing documents with a traditional system requires rule-based methods to identify fields of extracting optical character recognition (OCR) from fixed positions. These traditional solutions do not always work optimally. [42]

As the number of documents like case management, policy, and cyber incident documentation used in cybersecurity risk management increases, companies need to improve the efficiency and effectiveness of the process. Such a robust solution improves delivery times by converting such documents into machine-readable formats. An AI or ML solution improves the case review optimization by digitizing documents. [42]

a. Optimized document data ingestion

Automatically converts a physical document into structured and machine-readable data

b. Automatic data extraction

The solution can automatically use ML and Natural Language Processing (NLP) to analyze data with a machine-readable document. Robotic Process Automation (RPA) enters data into storage

c. Handle most document types

NLP, ML models and ANN can accurately handle handwritten text and cursive writing

d. Detecting Fraud or Criminal Intent

Associative rule learning and sentiment analysis can identify potential criminal activity on payroll statements and identification documents.

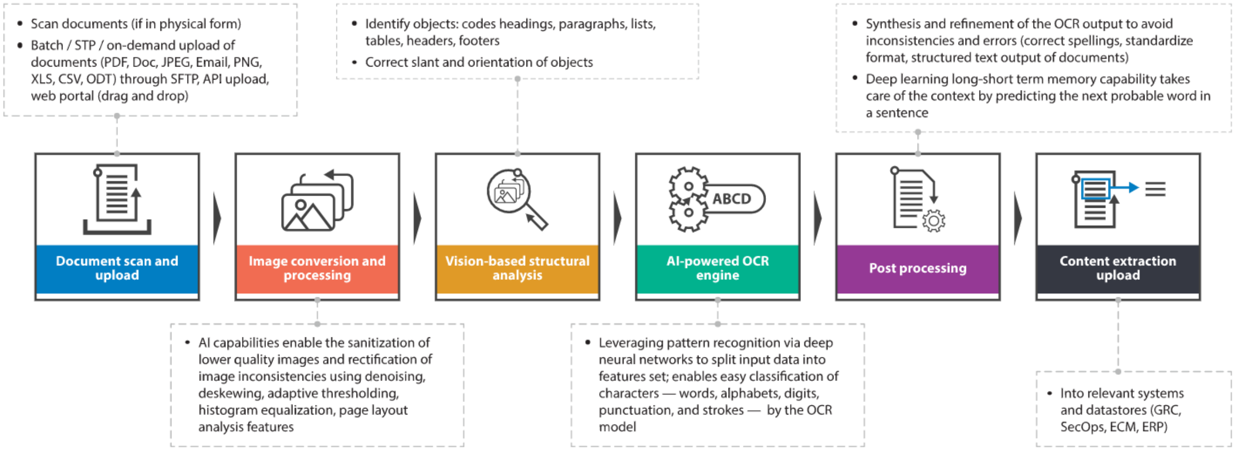

• Scan documents (if in physical form)
• Batch / STP / on-demand upload of documents (PDF, Doc, JPEG, Email, PNG, XLS, CSV, ODT) through SFTP, API upload, web portal (drag and drop)
• Identify objects: codes headings, paragraphs, lists, tables, headers, footers
• Correct slant and orientation of objects
• Synthesis and refinement of the OCR output to avoid inconsistencies and errors (correct spellings, standardize format, structured text output of documents)
• Deep learning long-short term memory capability takes care of the context by predicting the next probable word in a sentence
Document scan and upload
Image conversion and processing
Vision-based structural analysis
ABCD
AI-powered OCR engine
Post processing
Content extraction upload
• AI capabilities enable the sanitization of lower quality images and rectification of image inconsistencies using denoising, deskewing, adaptive thresholding, histogram equalization, page layout analysis features
• Leveraging pattern recognition via deep neural networks to split input data into features set; enables easy classification of characters — words, alphabets, digits, punctuation, and strokes — by the OCR model
• Into relevant systems and datastores (GRC, SecOps, ECM, ERP)

Figure 37: High-level workflow for AI-based document digitalization *[42]*

4.3.4 Automation of cyber security controls

AI capabilities, particularly RPA, ML, and NLP, can help intelligently automate duplicate manual cybersecurity tasks, including cybersecurity controls and control functions. The use of AI in this method improves the efficiency and effectiveness of using cyber controls, process simplification, enhanced control visibility, and lowered audit costs. Cybersecurity control solutions that support AI & ML can enable [42]

1. Automation of cyber security controls
2. Standardization of framework automation controls
3. Automated scan and inspection across network logs for applications and device inventory.
4. Automated controls to discover, extract, classify, and aggregate data.
5. Automated detection to monitor data loss.
6. Security validation and remediation.
7. Automatic updates and patches.
8. Automated backup for cybersecurity processes.
9. Ensure compliance with the European Unions General Data Protection Regulation, the California Consumer Privacy Act (CCPA), and the Payment Card Industry Data Security Standards (PCI DSS).
10. Organize threat intelligence triaging with automatic breach notification.
11. The automated gate checks monitor security activities throughout the software development lifecycle.
12. Realizing identity and access fulfillment, including role-based secure access and revoking unauthorized access. [42]

4.3.5 Alert investigation and qualification

AI or ML can help Internet teams to improve the efficacy and effectiveness of research and rehabilitation. Platform examines millions of records each day, filters data, and transfers them to a human analyst that reduces the number of alerts to almost 100 per day. Along with a significant reduction in false positives, the platform also helps to improve the attack detection rate. The following elements must be embedded in AI or ML for alert investigation and qualification solutions. [42]

1. Enriched incident context: The use of clustering and classification features on threat indicators to improve background to prioritize the alarm.
2. Sophisticated threat alert research and analysis: Use of NLP for semantic analysis, Deep nets for automatic information, and insights are collected on the incident, forensics, behavior, and time series analysis
3. False-positive and noise reduction: Using the principal component analysis, RNNs, deep convolutional neural networks (CNN) based transfer learning methods
4. Automation of repetitive alert management tasks: such as triaging of low-risk alerts or tedious alert data enrichment tasks with RPA, NLP, and ML
5. Planning, simulations, and recommendations: Use decision matrices, observational learning, and association rule learning modes. [42]

4.3.6 Malware detection, analysis, and prevention

AI or ML solution enables effective malware detection and analysis. It can detect, analyze, prevent novel malware variants and evolving malware. The malware like viruses, trojan horses, worms, exploitation, retroviruses, botnets, malvertising, and ransomware are detected. Analyze accessed fields on the disk, accessed APIs, consumed bandwidth, processor power, data volume transmitted over the internet, accessed products such as keyboards and cameras to identify and analyze malware. They leverage inference techniques by utilizing the predefined malignant malware properties to predict future threats that signature-based approaches cannot detect. Use ML models to perform a behavior, advanced static analysis, detect and categorize malware before executing. Use unsupervised future learning with automated encoders to discover the features of malware samples. Some AI and ML methods are considered for malware detection and analysis. ML enables hardware-aided malware discovery using virtual memory access patterns. The ML device learns through logistics regression, SVM, and random forest classifiers. Use operational codes, KNN, and SVM as ML classifiers for malware classification. [42]

Deep learning detects smart malware using automatic encoding and multi-layer restricted Boltzmann machines. Rotation forest and a new ML algorithm. ANN along with the raw sequence of API models to detect Android malware. The hybrid model depends on CNN and deep autoencoder to identify large-scale android malware detection. Bio-inspired methods for the feature to improve functions, malware classification, and classifier parameters optimization using PSO or GA. [42]

4.3.7 Phishing, spam detection, and filtering

The traditional anti-phishing and spamming approaches involve simple word filtering, IP blocklists, content filtering, and sender reputation mapping are not always practical. Advanced solutions instead of advanced ML models provide hundreds of input functions. It uses NLP to check grammar and automatically applies visual analytics to detect and classify phishing and spam emails automatically. In addition, these methods can identify the content that organizations want to ban. The following methods are used in industries. [42]

Anti-phishing deception methods use different ML algorithms and functions to diagnose the leading phishing websites. Reinforcement learning applications and neural networks to identify phishing websites. These methods use the principles of risk reduction and Monte Carlo algorithms. Real-time anti-phishing systems using different classification algorithms and NLP-based functions. Stacking models combine XGBoost, gradient boosted decision trees, and LightGBM using HTML and URL functions to classify phishing web pages. Combined, Naive Bayes and SVMS develop spam filter systems. Spam classification using modified cuckoo search to promote random spam classification. Cuckoo search algorithm extracts functions, and SVM is used for classification. The PSO algorithm is used for feature selection, and decision tree SVMs handle classification. A hybrid approach to identify spam profiles on Twitter through bio-inspired computer and social media analytics. Spammer detection uses k-means integrated with the Levy-flight firefly algorithm and chaotic maps. Email spam detection and identification systems based on random weight networks and GAs. [42]

4.3.8 Countering advanced persistent threats

APTs use sophisticated methods to exploit sensitive data without being detected. APT attackers often target valuable ones, such as governments or the security division of a large company. The intent is often to steal long-term information. AI or ML solutions are used to mitigate against APT attacks, including

1. The decision tree for creating IDS to detect APT attacks from the start and respond quickly

2. Frameworks based on multiple parallel classifiers
3. Deep Neural Networks (DNNs) that use dynamic analysis capabilities to set up APT

4. Use the self-organizing feature map and machine activity metrics to differentiate between legal software and malicious software

5. MLAPT is an ML-based approach to the identification and prediction of APTs. [42]

4.3.9 Identifying domain names generated by domain generated algorithms (DGA)

DGAs are algorithms that periodically generate many fake random domain names. These names hide the command and control server from the operator. AI or ML solutions can accurately identify domain names created by DGAs. One solution could deploy ML models to look for DNS logs for DGAs associated with anonymous malware. Some AI or ML solutions include

1. Adopt an RNN or CNN based architecture
2. Use generalized likelihood ratio tests
3. Novel LSTM based algorithms to solve imbalance problems in multiple categories through DGA malware detection
4. ML-based DNS converts the DGA detection system. One such system uses a specificity score, improved term frequency-inverse document frequency, and other algorithms to detect malicious domain names
5. Identification of word-based DGAs by leveraging word frequency distribution and ensemble classifier developed using additional trees, Naive Bayes, and logistic regression. [42]

4.3.10 Prevention of zero-day attacks

A zero-Day attack is a software security vulnerability that is unknown or has not been fixed by the parties responsible for resolving the problem. Until these security vulnerabilities are closed, hackers can exploit them for criminal purposes. Zero-day vulnerabilities cannot be detected with conventional methods such as anti-malware or IDS-IPS because no signatures have been created yet. An AI or ML solution is used to close zero-day vulnerabilities. With these insights, organizations can fix software vulnerabilities and exploit exploits before leading to security breaches. More precisely, solutions based on artificial intelligence or machine learning

1. Speed up the sandbox method that experts use to analyze zero-day threats. Sandbox mode installs suspicious files or programs on a virtual machine and then examines its behavior
2. This process can take several minutes if done manually, while AI and machine learning models can assess in milliseconds
3. Use unsupervised and behavioral learning techniques to analyze user interactions with software and any background information to predict malicious actions
4. Use the capabilities of a guided graph model, such as the Galois network, to classify and analyze actions to keep track of the relationship between events [42]

4.3.11 Cyberthreat hunting and penetration testing

Cyber threat hunting is active and repeated by duplicate networks to find advanced threats that can evade existing solutions. This management approach is for traditional threats such as firewalls, IDS, security information, and event management systems, which usually require research into the data after a warning is issued. [42]

Penetration testing

It is also known as pen testing or ethical hacking, which involves probing a network, web application, and computer system to determine the security vulnerabilities that the attackers can use. AI-based threat hunting and pen testing solutions enable companies to remember and reproduce complex cyber security weaknesses effectively. These solutions provide

1. A detailed, reliable, comprehensive coverage and possible testing of attack vectors
2. Allow automated, intelligence, and guided hunting of undetermined cyberthreats
3. Use advanced ML technologies, such as reinforcement learning and Markov's partially observed decision process, to determine vulnerabilities
4. Integration with industrial pen-testing frameworks and diverse data sources of Internet data for practical vulnerability assessment
5. Use pre-guided threat libraries consisting of models, questions, data functions, and playbooks to support a wide range of threats
6. Using the link analysis and chaining capabilities, we can automatically connect all events that are linked to an incident and provide a full context
7. We can even predict the next attack with a specific threat scenario and information about damaged devices and use them to activate pre-emptive remediation. [42]

4.3.12 AI-based antivirus software

Traditional antivirus software scans files on a company's network and determines whether they match the signatures of known viruses or malware. It works well with exposed threats to a public signature, and new threats are difficult to identify. This software is slow and does not provide real-time threat detection as updates are required to detect new viruses. AI-powered antivirus software can help companies meet these challenges. Instead of matching signatures, AI uses anomaly detection functions to monitor program behavior and detect abnormal actions. [42]

4.3.13 Behavioral modeling and analysis

In some cyber security breaches, a cyber attacker steals a user's credential (without the user's knowledge) to gain access to the organization's network using legal, technical means. Hence, it is difficult to detect and stop this type of attack. An AI-based solution can use behavior modeling and analysis capabilities to prevent such attacks. The following are the solutions

1. Data acquisition by software installed on customer workstations and sensors located in network segments
2. Enter the above data for the behavior analysis and threat intelligence engines that use ML to identify anomalies, abnormal endpoints, and network behavior
3. The system deactivates these connections when they occur. Use ML regression analysis functions to analyze raw network traffic data to determine the underlying behavior of each user and device

4. Use ML classification techniques for different group users, analyze a peer group, group individual user groups, and identify outliers
5. Identify significant deviations from basic behaviors, such as a login at unusual hours, and alert the company to potential cyber threats. [42]

4.3.14 Combating AI threats

Cybercriminals can use artificial intelligence and machine learning technologies. Such threats share the ordinary value propositions of behavior analysis, rapid scalability, and customization technologies.

1. Create new types of older malware
2. Create new phishing and spam content based on training sets from previous successful campaigns
3. Help phishers and spammers spot duplicate patterns in malicious content
4. Identify vulnerabilities in corporate networks and use this information to gain access to spyware, phishing, or DDoS attacks
5. Build smart malware or artificial hackers to carry out personalized attacks
6. Improve malware targeting by indexing potential victims using publicly available, aggregated, or extracted data
7. Finds new zero-day vulnerabilities by entering unexpected, invalid, or random data as inputs
8. It enables the ML algorithm to learn which procedure is required to identify new vulnerabilities
9. Let the robot nodes learn and share information to determine the most effective form of attack
10. Cybersecurity vendors and companies can enable artificial intelligence or machine learning solutions to react and adapt to new scenarios continuously
11. These solutions use sophisticated ML algorithms to detect attacks by other ML algorithms.
12. Advanced countermeasures can detect these cyber attacks and security vulnerabilities and provide recommendations in real-time
13. Artificial intelligence and machine learning solutions can also help identify individual devices that are inactive or anomalous in botnets. [42]

4.4 AI and ML cybersecurity in the real world

The below firms provide real-world AI and ML cybersecurity risk management solutions.

Company	Use Case	Details
Absolute	Endpoint resilience	This firm enables adaptive, intelligent, and self-healing endpoint security with constant visibility of data, devices, applications, and users regardless of whether the endpoints are on/off. Absolute information identifies data through asset and security analysis that allows security managers to ask all relevant questions

Blue Hexagon	Network Intrusion Detection & Prevention	It offers real-time network threat protection. The firm uses AI to create malware on the dark web and global threat data to test its systems and enhance security capabilities.
Callsign	Identify verification & validation for fraud prevention	Its solution uses AI and ML capabilities to validate a person's identity based on the number of keystrokes, touch-screen swipes, and several locations. Callsign uses multifactor authentication and fraud analytics powered by deep learning technology to combat identity fraud, SMS phishing, and other cybercrimes.
CrowdStrike	Threat hunting	It provides clients with visibility and protection on their enterprise networks. The platform focuses on averting endpoint attacks and offers actionable threat intelligence and managed threat hunting.
Cyberwrite	CRQ and residual risk mitigation platform	The advanced algorithms and simple reporting systems are accessible to businesses of any size. It supports cyber profiling. It assesses cybersecurity threats and exposure for proactively making data-driven cybersecurity improvements. The solution automatically gathers internet data related to the dark web, attack surfaces, proprietary cybersecurity risk, and digital risk-based data connected to each entity. Such data are then used as classifiers in the solution's exclusive machine learning models and transformed into benchmarked risk scores compared to industry peers that allow visibility of the firm's risks Beyond customer-specific data, the solution's cyber risk models also consider the inherent and external risk factors, including sector geography and operational risks. The solution also leverages data returned to current regulatory fines, historical cyberattack damages, and risks that firms must manage.
Cylance	Antivirus	Cylance's innovative antivirus product uses AI to predict, identify and respond to threats. The solution identifies patterns that indicate malicious programs.
Darktrace	Network Intrusion Detection and Prevention	Darktrace has Enterprise Immune System and Darktrace Antigena platforms to identify various cyber threats in their initial stages. The solutions can be integrated into financial institutions' networks and offer the Darktrace Threat Visualizer, a real-time dashboard monitoring cyber threats.
Deep Instinct	Prevent zero-day threats and APT attacks	Deep instinct protects enterprise endpoints and mobile devices against cybersecurity threats. It applies deep learning to safeguard against zero-day threats and APT attacks.

FICO	CRQ and residual risk mitigation platform	FICO released its ML-based cyber risk score on the AWS marketplace. The algorithm utilizes new globally gathered micro signal data that enhances the capability to quantify an organization's cyber risk. The score is delivered via applications tailored to various use cases, including self-assessment, vendor risk management, and cyber insurance underwriting.
Kount	Online fraud prevention	Kount's next-generation adaptive AI solution uses supervised and unsupervised ML capabilities to prevent transaction fraud.
SECURITI.ai	PrivacyOps	SECURITI.ai is a leader in AI-powered privacy. Its platform simplifies privacy-related compliance using an RPA and NLP interface.
ShieldX Networks	Security policy identification	This cloud-native security platform constantly ascertains workloads and identifies risks in a multi-cloud environment ShieldXuses AI to expedite determining which cybersecurity policies are applicable for each application.
Tessian	Email monitoring for phishing	Tessian offers an AI-based email monitoring solution that helps financial institutions prevent phishing attacks and data breaches.
Vade Secure	Email Security	Vade secure deploys AI and ML to protect more than 600 million mailboxes in 76 countries from threats, including ransomware, spear-phishing, and malware.
Versive	Network Intrusion Detection and Prevention	The eSentire-owned Versive offers an AI-based enterprise cybersecurity solution, VSE Versive Security Engine, that uses dissonant detection to find network security vulnerabilities. The engine uses DNS, proxy data, and the bank NetFlow network protocol to gather IP traffic information and monitor network traffic.
Zero Networks	Zero trust security	The zero networks access orchestrator uses AI to enable a zero-trust network model. The platform watches how users and machines usually communicate and automatically defines and enforces a zero-trust network model across an enterprise.

Table 3: AI and ML cybersecurity in the real-world *[42]*

Chapter 5: Software Implementation to combat cybersecurity issues

5.1 Cybersecurity software implementation

Specific ways are vital to maintaining adequate cybersecurity, i.e., updating software or systems, conducting security audits from top to bottom, social engineering audits, regular data backups at work, and maintaining physical security & industry compliance. While opting for a Cybersecurity tool, Cyber Resilience should also be considered. Cyber Resilience means making every attempt to block the threat and working on minimizing the effect of a successful attack. With this attribute, business and email communication can be continued without any disruption. [43]

CyberSecurity Software is categorized into different types:

- Network Security Monitoring tools
- Web Vulnerability Scanning tools
- Encryption Tools
- Network Defence Wireless Tools
- Packet Sniffers
- Firewall
- PKI Services
- Antivirus Software
- Managed Detection Services
- Penetration Testing

The below-listed cybersecurity software is used worldwide

Software	Category	Features
SolarWinds Security Event Manager	Cloud-based tool for SIEM	SIEM security & monitoring, threat intelligence, log correlation, network and host intrusion detection
Intruder	Cloud-based vulnerability scanner	Over 9000 security vulnerabilities, checks for web application flaws, Emerging threat notifications, smart recon, network view, PCI ASV scans are available
Syxsense	Cybersecurity Software	A vulnerability scanner, patch management, device quarantine, threat monitoring and alerting, live device location maps, device history, drag and drop workflow designer.
Acunetix	On-premise & cloud-based Web Application Security Scanner	Dashboards, access controls, and multiple scan engines.

NetSparker	Cloud-based on-premise web application security	DAST+IAST approach
Vipre	Cloud-based email, endpoint security solutions & anti-virus	Endpoint security, email security, and network security
LifeLock	Identity Theft Protection	Block cyber threats, detect & alert, restore & reimburse
Bitdefender Total Security	Cybersecurity software	Multi-layer ransomware Network threat protection.
Malwarebytes	Cybersecurity for home and business	Multi-layered protection, prevention of threats in real-time
Mimecast	Email Security & Compliance Platform.	Cyber Resilience for Email, email web security, cybersecurity training
CIS	Cybersecurity tools	Securing organization and a specific platform, tracking specific threats
Snort	Network intrusion prevention system.	Real-time packet analysis, packet logging
Wireshark	Network protocol analyzer	The decryption of various protocols, Output in XML, PostScript, CSV, or Plain Text, Inspection of hundreds of platforms
Webroot	Cybersecurity for endpoints and networks	Real-time protection, Multi-vector protection, Predictive threat intelligence

Table 4: Comparison of Cybersecurity software

1. **Solarwinds security event manager**

   It is a network and host intrusion detection system, an all-in-one cloud-based scalable solution for managed service providers of SIEM tools. It performs real-time monitoring, responding, and reporting security threats. It has primarily listed log search capabilities. It can be used for small to large businesses. [43]

2. **Intruder**

   The intruder is a cloud-based network vulnerability scanner to find the weaknesses in exposed systems to avoid costly data breaches by saving time. The intruder is a one-stop solution for every cybersecurity need. It can be used for small to large businesses. [43]

3. **Syxsense**

It provides security scanning, patch management, and remediation in one console from the cloud, allowing IT and security teams to collaborate automatically in a single console to terminate breaches with one endpoint security solution and close attack vectors. It can be used for small to large businesses. [43]

4. **Acunetix**

Acunetix is the solution to secure our websites, web applications, and APIs. This application identifies over 7K vulnerabilities and scans all pages, web apps, and complex applications. It has built-in vulnerability management functionality and has on-premise and on-demand deployment options. It is an intuitive and easy-to-use solution and performs lightning-fast scanning. Acunetix seamlessly integrates into our current systems. It can be used for small businesses, enterprise customers, pen-testers, and web professionals. [43]

5. **Netsparker**

It is an application security testing solution for enterprises. It has the features and functionalities for automating the security testing throughout the SDLC. It also has automation, visibility, accuracy, scalability, and security capabilities. Netsparker web application security solution offers a complete picture of our application security by providing onboarding assistance and training. The unique DAST + IAST approach offers increased visibility for deeper scans. [43]

6. **Vipre**

It provides cybersecurity solutions and protects against computer viruses, ransomware, and identity theft. It can provide comprehensive email & endpoint security & privacy and real-time threat intelligence for business protection. This application also offers layered protection to businesses and partners as it supports Windows and Mac platforms. Vipre is easy to install and use. It can provide cybersecurity protection with DLP and business VPN and security awareness training for comprehensive protection against emerging threats. [43]

7. **LifeLock**

LifeLock is a tool to monitor identity theft and threats. Norton 360 antivirus software with LifeLock provides all-in-one protection to our identity, devices, and online privacy. It is the platform to block cyber threats, detect and restore by resolving ID theft issues with identity restoration agents. The stolen funds can be reimbursed because of Id theft. LifeLock blocks the information on public Wi-Fi through a secure VPN. It provides alerts through text, email, or mobile apps and 24/7 live member support. [43]

8. **Bitdefender total security**

Bitdefender Total Security provides online privacy and personal information, features of file shredder, social network protection, privacy firewall, vulnerability assessment, and safe online banking. It comprises the features for Anti-Phishing and Anti-Theft by providing 24/7 comprehensive support. It supports Windows, Mac, Android, and iOS devices as anti-malware software. [43]

9. Malwarebytes

It offers cybersecurity solutions and protects against malware, ransomware, malicious websites, and advanced real-time online threats not detected by antivirus software. It supports Windows, Mac, Android, iOS, Chromebook devices. Malwarebytes provides a cybersecurity solution for homes and businesses. [43]

10. Mimecast

It is a cloud-based platform that provides email security and cyber resilience. It provides services like email security with threat protection, Information protection, Web security, Cloud Archiving, spam detection, and URL security. [43]

11. CIS - Center for Internet Security

It provides various cybersecurity tools, services, memberships and provides CIS SecureSuite for commercial use. CIS Security suite includes CIS controls and Benchmarks. [43]

12. Snort

Snort is an open-source platform and an application for network intrusion prevention. It supports FreeBSD, Fedora, Centos, and Windows platforms and performs the task of watching network packets and streaming data to our screen. Snort acts as the second level of defense as it is located behind the firewall and compares the traffic against the rules. It can be used for small and medium-sized businesses. [43]

13. Webroot

Webroot is a cloud-based platform solution for home use, offices, businesses, and partners. It can protect PCs, Mac, and mobile devices and supports Windows, Mac, Android, and iOS platforms. Webroot provides DNS protection, Endpoint Protection, and threat intelligence for businesses. [43]

14. GnuPG

GnuPG is a free tool for encryption, signing of data, and communications. It supports Windows, Mac, and Linux platforms. GnuPG has features like key management and access to public key directories. [43]

15. Norton Security

Norton provides a one-stop solution through Norton 360 with LifeLock. The company offers a variety of cybersecurity software solutions such as Antivirus, Virus Removal, Malware Protection, Cloud Backup, Password Manager, and Secure VPN. [43]

16. BluVector

It is based on Artificial Intelligence, Machine Learning, and speculative code execution by providing real-time advanced threat detection. BluVector Cortex is an AI-driven security platform. It has flexible deployment options and provides 100% network coverage. [43]

17. Nmap

NMap is a port scanning tool used for network discovery, security auditing, Network Inventory, and managing service upgrade schedules. It also monitors host or service uptime. It can be used for scanning large networks as well as single hosts. [43]

18. Sparta Antivirus

It provides a full range of security for total protection. Sparta's system is designed with the latest technology of AI that keeps our environment clean from all possible threats. Keep all our online data safe from malware, viruses, trojans, and phishing websites. It can be used for Removing Malware and Fixing our PC or Mac with One Click. [43]

5.2 Penetration Testing

Penetration Testing is also known as pen testing security pen-testing. It describes the purposeful launching of simulated cyberattacks by “ white hat” penetration testers using strategies and tools to access or exploit computer systems, networks, websites, and applications. Even though the main idea of pen-testing is to identify exploitable issues to enforce adequate security controls, security professionals can also use penetration testing ways. Along with technical testing tools, to test the robustness of a company's security policies, its nonsupervisory regulatory compliance, its workers' security mindfulness, and its capability to identify and respond to security issues and incidents similar to unauthorized access as they do occur. [44]

Penetration testing aims to improve the security bugs present in the app so that hackers cannot take advantage. During Web App pen testing, the tested program is a web application stored on a remote server that clients can access across the internet. Web applications are more prone to automated attacks [45] [46]. An organization that conducts penetration testing can be divided into three distinct penetration testers.

- **Grey hat** can be modified into a custom test plan (a hybrid solution) for the below two types [45].
- **Black hat** is primarily used to identify how the undesired attack is being dealt with by the employees of a particular organization [45].
- **White hat** is an ethical hacking following the organization's guidelines, policies, and procedures for research and development purposes

Penetration testing is performed during a system or application security evaluation. Most organizations perform this test during their production release or a product upgrade. It is essential to conduct penetration testing when installing new software, after policy modification, security patches update, or after installing new infrastructure to the overall system network topology [45].

5.2.1 Common pen-testing strategies

The commonly used penetration testing strategies are as follows based on an organization's strategy

- **External testing**

  It involves attacks on an organization's network perimeter and procedures performed outside the organization, e.g., Extranet and Internet. These attacks are carried out from outside the organization and require web applications to monitor the internet. Testers pretend like hackers who are not familiar with the internal structure. Testers are provided with the IP of the target systems to simulate these attacks, and no further information is provided. It helps to understand and identify how deep a malicious attacker can penetrate the network and the business effect of a successful attack. The scope of any vulnerabilities found is also measured. [45]

- **Internal testing**

  It is performed within the organization's environment to understand what happens if the network perimeter is successfully penetrated or when an authorized user could penetrate specific information resources within the organization's network. This testing process helps estimate the damage of an attack created within an organization's infrastructure. It focuses on the attack caused by internal sources who have quit/withdrawn from the company but are aware of codes, security protocols, attacks on social engineering, attacks exploiting user rights or misuse of an unlocked terminal, and simulations of phishing attacks. [45]

- **Blind testing**

  It is performed when a tester tries to simulate the actions of a real hacker. The testing team has no clue about the organization. Still, it can only rely on public information such as the corporate website and domain name registry to collect information about the target and conduct its penetration tests. [44]

- **Double-blind testing**

  Very few people within the organization are made aware of the testing process. The IT and security staff is neither notified nor informed, and they are "blind" to the planned testing activities. It helps to test an organization's security monitoring, incident identification processes, escalation, and response procedures. [44]

- **Targeted testing**

  It involves both IT and penetration testing. Testing activities information concerning the target and the network design are made aware. Targeted testing consumes less time and effort when compared to a blind test. Still, it typically does not provide an accurate picture of an organization's security vulnerabilities and response capabilities as other testing strategies. [44]

5.2.2 Pen-testing tools

The following are the pen testing tools to fortify defenses and provide security analysts services.

1. Port scanners
2. Vulnerability scanners
3. Application scanners
4. Web application assessment proxies

Network Mapper

www.nmap.org

Nmap is a free open source tool used at any stage to identify vulnerabilities in a network environment. NMAP takes raw data packets that are just created and uses them to determine the hosts available on a particular network trunk or segment. It also determines information about the host services, type of operating system used, versions, and types of data packet filters/firewalls used by any particular host. Organizations create a virtual map of the network segment and pinpoint the areas of vulnerabilities that an attacker could potentially penetrate. [44]

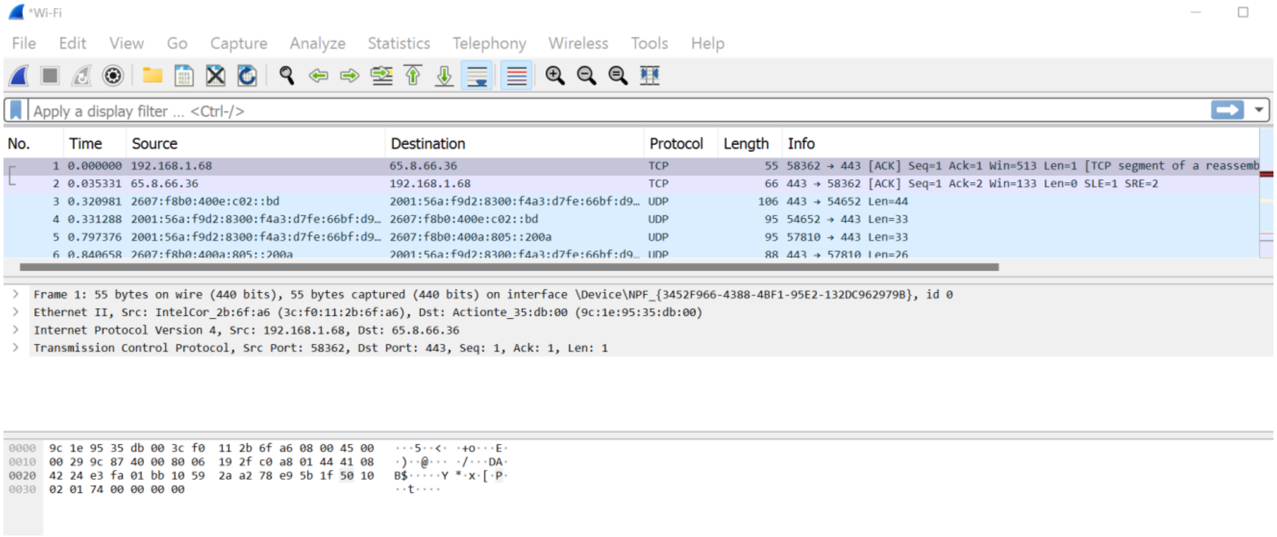
Metasploit

It provides a package of various pen-testing tools. Metasploit is a framework that constantly evolves to match today’s ethical hackers. It is powered by the PERL platform, which comes with an entire host of built-in exploits used to execute different kinds of pen tests that are even customizable. [44]

Wireshark

It is an actual network protocol, and the data packet analyzer points out and assesses traffic for vulnerabilities in real-time. It reviews connection-level information and constituents of data packets to highlight their characteristics, origin, destination, and more. When a potential vulnerability is flagged, a penetration testing tool is still required to exploit them. [44]

Wireshark is the most preferred ethical hacking tool used by hacker communities across the globe. It is an easily accessible, free, open-source network packet analyzing tool. It allows us to examine network packets from a live network where we can interactively browse and capture network data. Its efficiency in detecting and solving network security problems makes it a unique sniffing tool many ethical hackers use. The tool's ability to intercept and read the network packets in a human-readable format makes it easier to identify serious threats, vulnerabilities, and low latency problems. [47]


*Wi-Fi
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
Apply a display filter ... <Ctrl-/>

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	192.168.1.68	65.8.66.36	TCP	55	58362 → 443 [ACK] Seq=1 Ack=1 Win=513 Len=1 [TCP segment of a reassemb
2	0.035331	65.8.66.36	192.168.1.68	TCP	66	443 → 58362 [ACK] Seq=1 Ack=2 Win=133 Len=0 SLE=1 SRE=2
3	0.320981	2607:f8b0:400e:c02::bd	2001:56a:f9d2:8300:f4a3:d7fe:66bf:d9…	UDP	106	443 → 54652 Len=44
4	0.331288	2001:56a:f9d2:8300:f4a3:d7fe:66bf:d9…	2607:f8b0:400e:c02::bd	UDP	95	54652 → 443 Len=33
5	0.797376	2001:56a:f9d2:8300:f4a3:d7fe:66bf:d9…	2607:f8b0:400a:805::200a	UDP	95	57810 → 443 Len=33
6	0.840658	2607:f8b0:400a:805::200a	2001:56a:f9d2:8300:f4a3:d7fe:66bf:d9…	UDP	88	443 → 57810 Len=26

```
> Frame 1: 55 bytes on wire (440 bits), 55 bytes captured (440 bits) on interface \Device\NPF_{3452F966-4388-4BF1-95E2-132DC962979B}, id 0
> Ethernet II, Src: IntelCor_2b:6f:a6 (3c:f0:11:2b:6f:a6), Dst: Actionte_35:db:00 (9c:1e:95:35:db:00)
> Internet Protocol Version 4, Src: 192.168.1.68, Dst: 65.8.66.36
> Transmission Control Protocol, Src Port: 58362, Dst Port: 443, Seq: 1, Ack: 1, Len: 1
```

```
0000  9c 1e 95 35 db 00 3c f0  11 2b 6f a6 08 00 45 00   ···5··<· ·+o···E·
0010  00 29 9c 87 40 00 80 06  19 2f c0 a8 01 44 41 08   ·)··@··· ·/···DA·
0020  42 24 e3 fa 01 bb 10 59  2a a2 78 e9 5b 1f 50 10   B$·····Y *·x·[·P·
0030  02 01 74 00 00 00 00                               ··t····
```

Figure 38: Wireshark traffic from home wifi

What does it provide?

Wireshark intercepts the network packets and converts them into human-readable formats. It extensively supports thousands of network protocols but majorly ICMP, TCP, and UDP. It has added filtering features that make it quite useful to extract information related to the router's IP address related to the network. Moreover, this tool highlights color coding's group results by filter types, making it more beneficial to identify information. Regular expressions are also supported by the tool for packet filtering. The best way to analyze network traffic is to capture everything, and Wireshark is one of the best tools available. [48]

How does it work?

Wireshark's robust GUI is for network packet capturing and analysis. It has packet analyzers that allow users to view and capture network traffic using network interface controllers (NICs). The packets captured through Wireshark can be filtered or distinguished into paths, protocols, data streams, and IP addresses which can be analyzed offline. [49]

Wireshark Tool features

- The tool stores all the network packets and can be used for offline analysis.
- Determine network and browser packets.
- Provide the user with a functional GUI interface.
- Provides exclusive VoIP analysis
- It can also Inspect and decompress zip and gzip files

- It captures file formats like Tcpdump, Cisco Secure IDS IP log, and Microsoft network monitor.
- It has the ability to export the results in multiple formats like XML, CSV, PostScript, or plain text.

Wireshark usages by network security specialists

- **Acts as a sniffing tool:** Sniffing helps network security specialists/analysts verify network security devices' implementation and functionality and validates that data is traveling through secure channels. Security specialists use packet sniffing to identify the source of the attack, the time and duration of the attack, protocols and port numbers involved, and data transmitted for the attack. It detects the use of insecure protocol(s) for transmitting sensitive pieces of information.
- **For packet analysis:** The network analysts must understand network protocol suites and network communications.

  For example, a user hits ping to some website without knowing the source and destination IP address. The network packets flow from the source packet to the next available router. It traverses through network routers to reach the destination IP address. In such cases, the network analysts can use Wireshark to determine the routers involved in the network connection. Apart from router information, other information like the type of protocols used for transmissions of data packets, packet size, and the number of hops can be determined. [50] So, In a nutshell, Wireshark is a complete tool that can store all the network packet transmissions and could be used later for a detailed offline analysis.

Wireshark's harmful usage by hackers

Wireshark provides some exciting features used by hackers to exploit network communications. Following are some features of Wireshark used by hackers:

- Baselining network traffic:

  The hackers can identify all the host's normal network traffic, including regular patterns, login processes, and applications.

- Perform passive network discovery:

  A hacker can list visible hosts, port numbers, and addresses and listen to conversations via network analyzers and ARP traffic.

- Can detect unsecured applications that are not encrypted. Wireshark detects unusual applications and protocols. [51]

Wireshark is still one of the most used tools among hacker groups because of its detailed analysis features. We must use it wisely and abide by ethical hacking principles.

The web application attack and audit framework

W3AF is a pen-testing suite created by the software developers at Metasploit. Its ideal aim is to exploit security weaknesses in web-based applications. [44]

- **John the Ripper**

  http://www.openwall.com/john/

  JTR is an open-source platform that is a fast, efficient password cracker currently available for operating systems such as Unix, macOS, Windows, DOS, BeOS, and OpenVMS. Pen testers use it to detect weak passwords and address the inherent weaknesses in typical password use. [44]

5.3 Manual or Automated pen testing

Pen testing can be performed manually or automatically. Automated penetration testing and its advanced features play a critical role in protecting a system or a network. The U.S. Department of the interior defines penetration testing as "a controlled attack simulation that helps identify the susceptibility of an application, network, and operating system breaches." It implements defensive strategies to defend critical systems and intelligence. [52]

Pen testing can be performed using automated testing tools, which optimizes resources by automating elements of the testing process to identify the vulnerability continuously and without human intervention. Insights critical to the organization's ability are to refine its security policies and patch detected vulnerabilities. Penetration testing must expose security flaws that allow attackers user, system, network, or application access. This process involves collecting information about possible targets, identifying potential entry points, attempting to break in virtually or actually, and reporting the information discovered to the security team. [52]

The automated testing tool saves time and typically produces better penetration test results when compared to manual efforts. Costly breaches, data loss, compromised systems, users, and applications bring high risk to the enterprise. Automated pen testing is an effective tool to prevent real-time attacks and mitigate these vulnerabilities. They also deliver fewer false alarms and lower AppSec costs. [52]

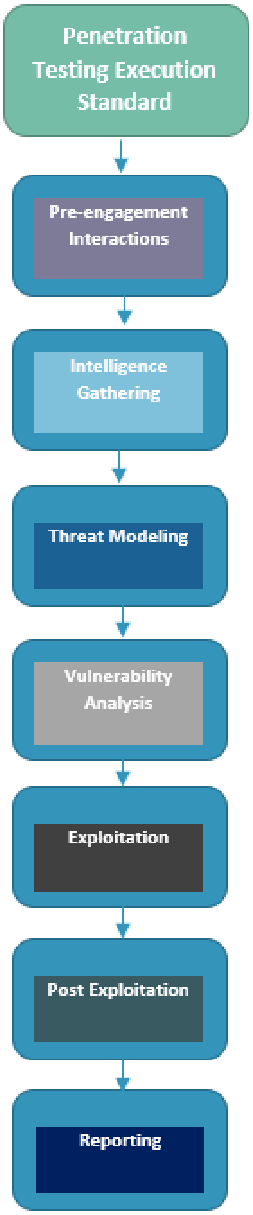


Figure 39: Operational Phases of PTES Standard *[44]*

5.3.1 Penetration testing industry standards

Some of the penetration standards include the Information Systems Security Assessment Framework (ISAF), the Open-Source Security Testing Methodology Manual (OSSTMM), the NIST SP 800-115, and the Open Information Systems Security Group (OISSG) [53]. The Open Web Application Security Project (OWASP) offers pen testing methodologies, guides, frameworks, and a Penetration Testing Execution Standard (PTES). PTES is the current standard implemented in 2010 and separates penetration testing into seven phases, as shown in fig.12 [44].

Organizations regularly perform penetration testing to identify their exposures and block holes. These tests help the organization to take a proactive stance as its search for vulnerabilities in its infrastructure (hardware), applications (software), and people to develop continuous adequate controls to keep up with evolving cyber threats [44]. The standard NIST (SP800-115) provides guidelines for preparing, performing checks, and reviewing information security. It provides the core safety testing and evaluation elements to ensure those methods. The penetration testing process is divided into four based on this standard, as shown in the figure.13 [46]

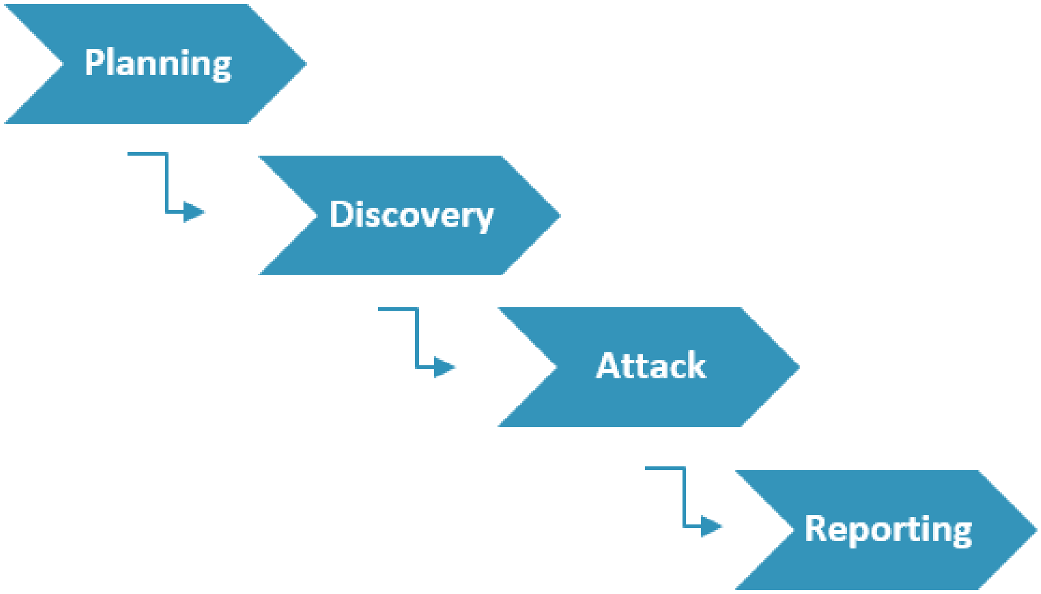


Figure 40: NIST standard phase *[46]*

Chapter 6: Analysis to prevent future attacks

6.1 Adaptable cybersecurity

Cybersecurity transforms the concepts of usernames, passwords, and IP addresses, turning them into practical elements that support their underlying functions. The evolving digital ecosystem became attainable as the traditional elements of computing mainframes, operating systems, applications, and networking became atomized, abstracted, and virtualized. Adaptable cybersecurity is considered to achieve a broader ecosystem, apply, protect and become more resilient in the face of cybercrime.

Security practices are based on the concept of trust. Cybersecurity should focus on authenticating identities in requests for any protected resource. The data, network, workload, data flows, and the underlying infrastructure supports them. Legacy security is not enough to secure a modern IT ecosystem that contains remote workers, workplaces, partners, and customer interactions to protect the data that need to access. Security functions focus on external threats, internal errors, and leaks that are not severe.

Context should hold the key to access rather than depending on users' email addresses and passwords to grant access to a secure LAN. Context means considering a variety of factors. What identity needs access? What is its status? What device is being used? How secure is it? Is it managed? And when was the access initiated? [54].

6.2 Zero Trust

A Zero Trust Security Model focuses on its purpose to deploy security. Zero Trust protects against unauthorized access to digital resources by enforcing controls. They are granular, risk-based, and adaptive for every access request. Zero Trust is based on six core principles and associated technologies [54].

1. **Never trust. Always verify**

   In today's world, most work occurs in a potentially compromised networked environment. Given the evolution of threats, the secure approach requires validation for every identity before establishing particular access.

2. **Purpose-driven access**

   Earlier, a "one-time password" was sent to an email address via internet protocols, which is no longer sufficient now as they can be compromised easily. Access must be contextual and time-bound to require desired outcomes. Password-less multifactor authentication (MFA) is more secure and faster than multiple password resets and insecure email delivery.

3. **Continuous risk discovery, real-time treatment**

A “find to fix” approach automated, with greater agility and operational rigor, should replace most IT organizations' long audit, testing, and remediation cycles. Cybersecurity is to manage risk, but most solutions focus on compliance rather than risk. Compliance is always necessary for regulatory adherence. Accurate technology risk is contextual, and measures that dynamically calibrate controls can address the threats.

4. **Information-centric security**

   Due to the current proliferation of cloud-based models and computing needs, businesses can center their proprietary information, expertise data, and uses at the business core, across which IT services can secure the perimeter.

5. **Security as a culture**

   “a chain is only as strong as its weakest link.” Data is only as secure as its most vulnerable vector. Cybersecurity is a culture that makes everyone responsible empowers all users to sense and act on cybersecurity matters. [54]

6.3 Implementing adaptable cybersecurity

Agile methods and DevOps provide another essential layer of protection covered by embedding and automating identity, threat, and vulnerability management. DevOps teams often need to use cloud platforms, untested or open-source software, and sandboxes. DevOps teams navigate more freely while minimizing security risks by securely managing the “secrets” to authenticate across the different platforms, applications, containers, and tools involved in development and testing.

State-of-the-art approaches and industry standards should be fundamental for any technical solution. Based on cybersecurity professionals worldwide, an approach to both resilient and adaptable cybersecurity is highly recommended. Cybersecurity must be comprehensive, adaptive, and collaborative according to the NIST standard of McKinsey analysis [54]. Companies must assess threats and develop controls for the most critical, as shown in figure 14.

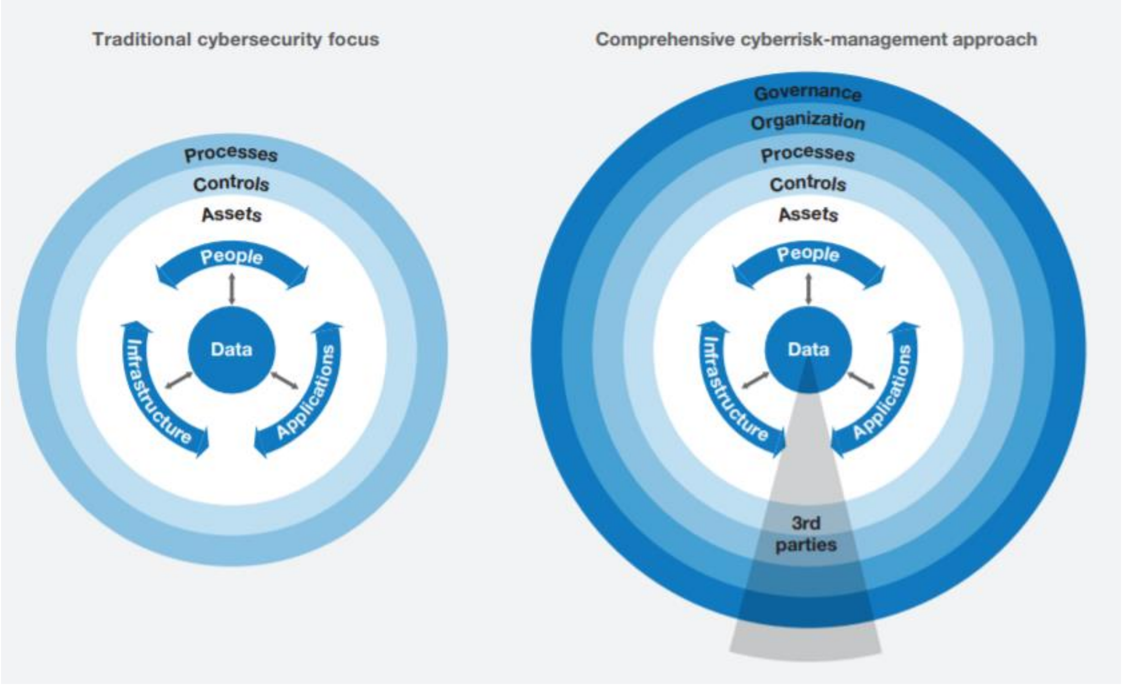


Figure 41: NIST - McKinsey Analysis *[54]*

The following steps can implement adaptable cybersecurity.

1. **Building resilience, step by step**

   Successful cyber strategies are built one step at a time, drawing on a comprehensive understanding of relevant business processes and the mindset of prospective attackers. Three key steps are prioritizing assets and risks, improving controls and processes, and establishing effective governance.

2. **Prioritize assets and risks by criticality**

   Companies should take stock of their cyber risk capabilities and compare them with industry benchmarks. With that knowledge, realistic aspirations for their resilience level can be set. IT, OT, and IoT products should be consolidated into one operating model. The entire system should be covered, including third parties.

3. **Consolidate and establish universal governance**

   Most security organizations are driven by analog instability; hence the resulting structures, decision rights, and processes are insufficient to deal with cyber risk. A state-of-the-art cybersecurity function should link the responsibility among physical security, information security, business continuity, and crisis management to reduce the conflicts of interest and replication of processes. It should align its cybersecurity tasks with relevant industry standards to work efficiently and manage incidents. The organizational structure must clarify responsibilities and relationships among corporate headquarters, regional teams, subsidiaries and set guidelines. Robust architecture must be established for data, systems, and security to ensure "security by design" and improve long-term digital resilience. [55]

The organization must set up a company-wide ideal governance structure designed on strong cyber risk culture. The cybersecurity unit should be held responsible for cybersecurity, and the following governance is recommended

- Report directly
- Possess the overall cyber risk budget
- Accountable for implementing any portfolio of initiatives
- A regular report on the progress of risk remediation on all cyber risk-related decisions such as outsourcing and exceptions from security controls
- Establish and ensure coverage of cyber risk-related activities
  Build awareness campaigns and training programs to cover the latest threats
- Enforce effective communication and incentive structures for cybersecurity controls
- Stage frequent, realistic attack and crisis simulations within the organization

Assets	Threats	Controls
Data	Data breach, Misuse, or Manipulation of information	Data protection (encryption), data-recovery capability, boundary defense
People	Identity theft, Man in the middle, social engineering, Authorization abuse	Controlled access, account monitoring, security skills and training, background screening, awareness, and social control
Infrastructure	Denial of service, Manipulation of hardware, Botnets, Network intrusion, Malware	privileged access control, monitoring audit logs, malware defenses, network controls (configuration, ports), inventory, secure configuration, continuous vulnerability assessment
Applications	Software manipulation, Unauthorized software installation, information system misuse, denial of service	Email, web browser protections, application software security, inventory, secure configuration, continuous vulnerability assessment

Table 5: Assess threats and develop controls *[56]*

6.4 Protecting critical digital assets

All data and systems are not created equally, while some data, systems, and applications are more critical than others. Some are exposed to risk are presumably to be targeted. Critical assets and sensitivity levels vary widely according to the workplace. For hospital systems, the most sensitive asset is patient information risks of personal data theft, ransom, and breach. The attacks can be sophisticated or straightforward based on the objectives varying from financial reward or geopolitical advantage. Cybersecurity teams are built on three pillars [57].

- **Microsegmentation**

  It helps to identify the "hot spots" of risk to trace a target based on threat monitoring and mitigation

- **Change of culture**

  It makes spiteful or inattentive probability events less presumably to place a reactive posture. To combat negligence and co-opting, often conduct rudimentary cybersecurity training and phishing testing. Targeted interventions help to see and feel the importance of cyber-hygiene

- **Prediction**

  Prediction allows us to identify and disrupt insider activities much earlier in the threat life cycle using predictive insider-persona analytics. The primary personas that present a risk are well established and studied at length. This analysis can identify a group or individual more likely to represent a threat even before the event occurs. Companies can mitigate the threat by exhibiting the below steps in the predictive analysis. [58]

1. **Insight-led intelligence for human and automated decisions**

   A deep and comprehensive view of threats contextualized to ensure priorities followed by identification, risk measures, complete insight into high-value assets, process to prioritize cybersecurity efforts, a clear plan of action for deploying both proactive and reactive proficiency to achieve cybersecurity objectives [58]

2. **Keeping bad actors at the bay**

   Access to protected assets should require contexts and how to define a particular request or connection. Once access for an asset is defined, it can be micro-segmented and governed in a Zero Trust context.

   - Identity: human, nonhuman, privileged (e.g., sysadmins)?
   - Device: laptop, desktop, mobile, IoT sensor (managed or unmanaged)?
   - Location: on-premises, via the internet, from the local plant, or where the company has no presence?
   - Time of day and duration: why is a database accessed for hours beginning at 3 a.m.?

   Double Trouble: Cyber risk created by two types of workers

   Malicious Insiders

They purposefully seek to benefit themselves at the organization’s expense or harm the organization directly by stealing valuable data, committing fraud for financial gain, publicly exposing sensitive information to attract attention, or sabotaging IT systems.

Negligent Insiders

They do not harm an organization intentionally but expose the organization to risk through their mistakes or carelessness. For example, an employee may create a vulnerability, which attackers can exploit directly by sharing personal information online. A developer may misconfigure a company’s simple storage service, or someone may lose a hard drive with sensitive data. Employees force themselves personally undemanding to spear-phishing attacks [58]

3. **Scalable and adaptive access**

   It is an understatement to mention that despite the COVID-19 pandemic, ‘work’ has undergone too many changes. Timeworn concepts around workplaces, location dependency, fixed local working hours, presenteeism, and high-touch governance pave the way to a distributed model to enlarge business opportunities. The pandemic accelerated a trend where working from home was announced as a permanent strategy or employee option by workforce management policies and practices. This forward-looking approach requires an equally forward-focused security strategy and includes adaptive and scalable access independent of device location or identity. VPNs are legacy security solutions and cannot adequately accommodate remote workforces, whereas remote-access-as-a-service solutions and multi-factor authentication methods address modified security requirements. [57]

6.5 Real-time detection and immediate action

In addition to prevention and protection, real-time action in an agile fashion is the last line of defense for plugging vulnerabilities and responding to attacks. It means exploiting the existing security features or adopting service-based security options from trusted sources. To promptly address threats equal to the risk, whether instantly or over time.

In addition to human access, companies must secure automation and analytics for a strategic cybersecurity architecture. AI and ML enable instant investigation, immediate response, and real-time detection of cyber threats.

Resiliency for business continuity

Finally, organizations must design and execute continuity, backup, and recovery programs for critical infrastructure using zero-latency required for risk management and compliance. The government has been issuing mandatory updates for critical IT infrastructure.

Risk focused, context-aware cybersecurity

The following obstacles are often perceived as barriers to risk-focused, context-aware security.

1. Budget
2. Pressures of innovation

Cyber security is critical and with the growing variety of threat and vulnerability vectors, investing in specific solutions to counter cyber threats is not a sustainable strategy. Instead, today's complex cybersecurity is based on consumption. Using cybersecurity as a service, a company can improve financially by accessing the latest defense systems, usually with associated features that can be upgraded at no additional cost. In addition, this approach allows us to add more services at package prices as needed. Offerings can include remote access, identity management, vulnerability management, threat intelligence, digital forensics, encryption, and multi-factor authentication.

While compliance-focused cybersecurity can be considered essential, the inside is often a brake on collaboration, innovation, and transformation. In contrast, the adaptive nature of risk-based cybersecurity reflects this perception and reality. Leading cyber security practices and services are flexible enough to meet the needs of changing new opportunities. By removing many of the same limitations of legacy solutions, cybersecurity can instead enable development, collaboration, replication, and modernization in ways not previously possible.

Call to action

Given the evolving nature of potential organizations, cybersecurity is an ongoing process. However, the cycle of discovery, identification, development, and deployment may become automated and intelligent as the security culture and resources grow. Adaption to changing threats and opportunities must be considered as well.

Discover

A comprehensive assessment is the first step to determining cybersecurity updates and understanding the current security landscape, real threat profiles, and risks.

Define

Create a directional roadmap to determine security measures. A roadmap can identify initial goals that can be implemented in the short term and whether existing tools can be reused to achieve goals or require new tools. Finally, the roadmap outlines a tactical and strategic implementation plan based on the current threat landscape and the organization's view of ultimate maturity.

Develop

Assuming there is no immediate threat that requires extensive and immediate attention, tactical work can be done in support of the defense, often with a scrum team with few dedicated resources. Other areas identified in the roadmap may pose less immediate risks to sensitive data but are relatively easy to fix, so solutions should be implemented at this point, not when encountering a problem. Many, if not more, are paving the way for a more strategic phase of cybersecurity maturity.

Deliver to scale

Designing and implementing security to support its original purpose and growth and even expanding into new markets and segments as part of its business strategy must be implemented to allow the organization to scale horizontally and cross-functionally. The strategic security delivery program can be augmented with multiple streams of scrum teams working consistently and faster to achieve a higher level of cybersecurity maturity to achieve this goal.

There are many initiatives that an organization can take in the strategic phase. The first is micro-segmentation, which separates sensitive sources regardless of their physical location, which can be done at the network or host level. The second is access to the Zero Trust Network (ZTNA), which creates a defined software environment that helps selectively and securely expose sensitive applications to authorized and verified identities. Context-based authentication and risk-based micro-licensing can also be developed at this stage. The organization can also take timely precautions to grant access only when and for as long as necessary so that privileged identities are not permanently denied access. Finally, security tools can extend their event and security incident management capabilities with advanced tools that analyze user behavior and entities such as devices, applications, services, data warehouses, and anything else with an IP address.

6.6 Shifting to an active-defense model

Active defense enables to attack and distract attackers in real-time by combining threat intelligence sources and analyzing IT performance. This approach is based on the military community's self-defense experience in turbulent attack environments like Afghanistan and Iraq. Commanders are beginning to house operators, planners, and intelligence analysts in a tent where they can provide special operations teams with real-time, continuous intelligence to detect and respond to threats faster. Integrated and more accurate information makes it easier for units to monitor conversations, set goals, and increase the number of tasks to complete overnight. The main elements of the active defense model are

- **Anticipate attacks before they happen**

  We need to look at the threat perspective to see if someone is talking about it or someone down the chain, identify software and network vulnerabilities and identify potential intruders before they occur. It is an informative and intensive service which is very important. Adding cybersecurity experts in the process help organizations gain their insights. Third-party threat intelligence experts monitor a variety of sources. It includes tracking conversations and conversations in places like dark websites that require specific software to access and providing user anonymity to assess evolving threats to the company or its suppliers.

- **Detect and respond to attacks in real-time**

  Early detection relies on tracking network patterns and user behavior that deviate from the norm. The challenge is understanding what is natural because everything constantly changes, and human behavior is unpredictable. Intrusion detection and anomaly detection are two standard methods. Intrusion detection systems (IDS) look for exploits based on known attack patterns. However, because these systems are trained to recognize specific threats, they can ignore these emerging systems. It can also identify inappropriate activity from legitimate activity, such as non-malicious internal connections with tagged language or web addresses, continuous scans for network vulnerabilities, or system attack issues that have already been modified. Anomaly detection models work in the opposite direction. Instead of known attack signatures, look for behaviors that deviate from standard network patterns, such as unusually high volume. Active defense companies use threat detection and anomalous defense systems to provide broader threat detection.

- **Alarms to contain attacks**

  Decoy servers and deceptions are other tools deployed as active defense. Deceptions lure attackers into a dummy environment to gain additional intelligence of attackers. Foothold into the trap triggers an alarm, alerting the deterrents, operations center and triggering software agents to be contained in the network to terminate access. Some companies salt the environments with false information to confound attackers. Intruders, when performing a breach, usually return by the same gateway. Deceptions and traps need to be convincing to keep intruders inside long enough to gather valuable insights. Repetitive visits can be used to record the act attackers use to gain file, server access, and update the defenses.

- **Ring architecture to protect critical assets**

  Over the longer term, businesses need to construct layers of defense to keep the company's most critical assets deeply buried. Ring architectures, for instance, allow organizations to store data in different layers depending on the value and sensitivity of those assets. Each layer requires a specific key and authorization protocol to manage access. The active defense also requires an IT plan that organizes and prioritizes security-related technology spending. Otherwise, it can be tempting to protect everything and create vulnerabilities when spending and systems prove too challenging to maintain. [59]

6.7 Making a secure transition to the public cloud

Companies need a systematic, proactive approach to adapt their cybersecurity capabilities for the public cloud. The following four practices can help develop a consistent, practical approach to public cloud cybersecurity.

1. **Develop a cloud-centric cybersecurity model**
   It is essential to choose how to manage the cloud perimeter and re-architect applications to align the cloud strategy.

2. **Redesign the complete set of cybersecurity controls for the public cloud**
   Determine who provides it and how rigorous they need to be for each control.

3. **Clarify internal responsibilities for cybersecurity**
   The public cloud requires a shared security model, with customers responsible for specific functions. It may be different from a traditional outsourcing arrangement and redesign internal processes accordingly.

4. **Applying DevOps to cybersecurity**
   Enhance highly-automated security services available to developers via APIs, just as they do for infrastructure services. [60]

6.7.1 Developing a cloud-centric cybersecurity model

On-premises controls it on public cloud platforms without reconfiguration. Moreover, these controls do not provide visibility and protection across workloads and cloud platforms even after reconfiguration. New strategies were required to achieve these limitations.

The most effective approach to re-evaluate the cybersecurity paradigm involves the following:

1. How the network perimeter is defined
2. Should application architecture change for the public cloud?

The perimeter definition determines the topology and the boundary for the cloud-cybersecurity model. Moreover, choices regarding application architecture can guide the incorporation of security controls within the applications. These two critical choices also inform one another.

Perimeter security

- **Backhauling**

  Routing traffic over local area networks is a way for half of the cloud enthusiasts to manage perimeter security. This model is attractive to organizations that need on-premises access to most cloud workloads and want to customize their workload migration options to fit their architecture. Even organizations with limited cloud security experience benefit from connectivity, as it allows them to continue using internal security tools they already know well. However, the deployment may not be widespread for long.

- **Adopting CSP-provided controls by default**

  Using CSP security controls can be less expensive than other environment models, but it complicates the security of a multi-cloud environment.

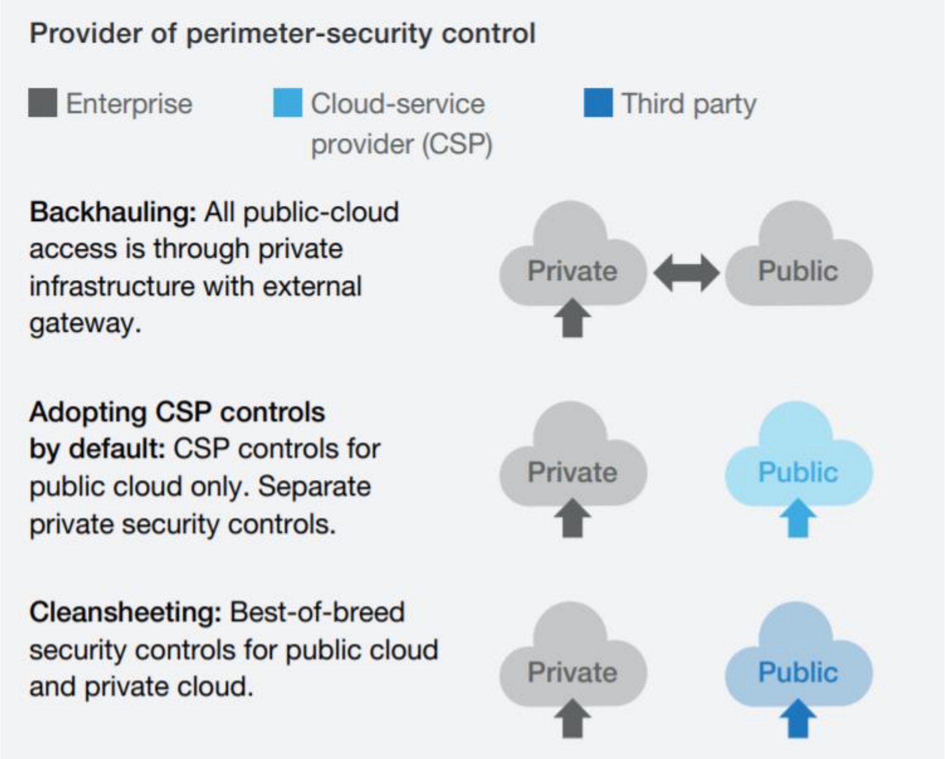

Provider of perimeter-security control
Enterprise
Cloud-service provider (CSP)
Third party
Backhauling: All public-cloud access is through private infrastructure with external gateway.
Private
Public
Adopting CSP controls by default: CSP controls for public cloud only. Separate private security controls.
Private
Public
Cleansheeting: Best-of-breed security controls for public cloud and private cloud.
Private
Public

Figure 42: Architecture model of perimeter security control *[60]*

- **Cleansheeting**

  It includes the design of a "virtual perimeter" and the development of cloud-specific controls from solutions that provide various external providers. This approach enables better perimeter security solutions and can be switched in and out as required. Since changes make technical demands, clean sheeting practice allows having enough cyber security solutions. Despite the high cost and complexity of clean sheeting, this approach can support multi-purpose environments to replace the needs efficiently.

Holistic Strategy

A holistic approach to cybersecurity can close the security gaps and impact governance, organizational structure, and processes. It is based on a precise overview of the risk situation that requires specific risk reporting. The goal is to enable organizations to focus their defenses on the most likely and most threatening cyber risk scenarios and balance effective resilience with effective operations. Strict controls only apply to crucial assets. The holistic approach explains the path to the root cause in four phases

Identify risks and risk appetite

Identify the risk appetite from the most threatening risk scenarios. Once the risk has been identified, an assessment is made in this phase of existing controls and vulnerabilities. The risk appetite may vary according to the value of the target of the threatened asset. The chief measure of cyber resilience is the security of the target's most valuable assets. The prioritization of identified risks is of utmost importance to top management.

Analysis and evaluation

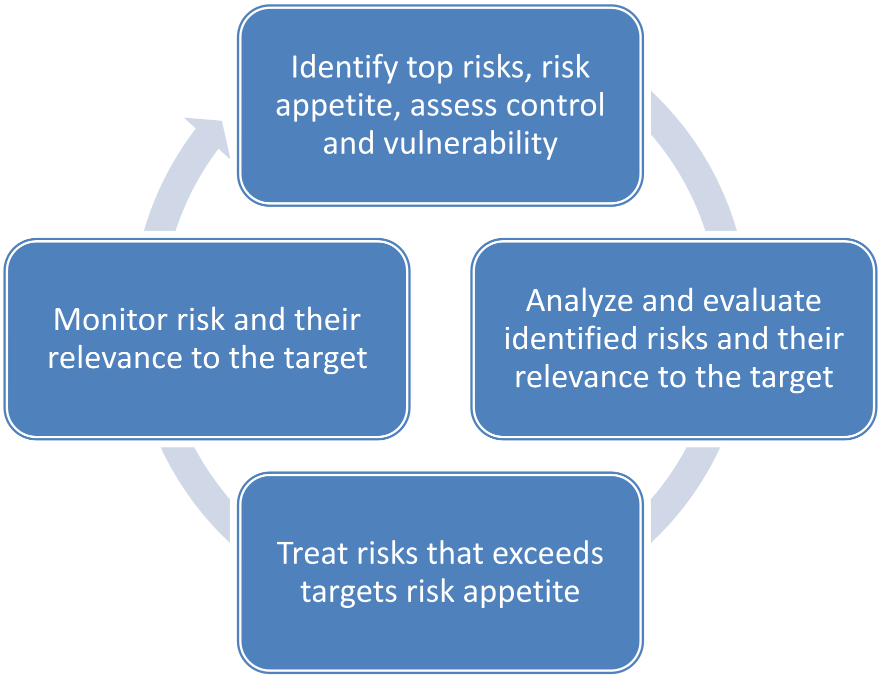
Internal/external experts evaluate each risk in terms of likelihood of occurrence and potential impact, including, as applicable, regulatory, reputational, operational, and financial impact.

Treatment

Once risks are identified and prioritized according to their likelihood and impact, risk owners and management performance need to work together to view risk mitigation actions comprehensively. Initiatives should be evaluated based on the likelihood of a risk event and its effectiveness in reducing the event's impact. By considering the impact of mitigation initiatives, risk professionals determine whether residual risk for each initial risk is now included in the organization's risk appetite parameters. If residual risk levels exceed these limits, additional mitigation plans can be developed and implemented.

Monitoring

Critical tools for promoting discipline across the organization include updating planned schedule status on critical cyber threats, remedial strategies, and remedial action. Over time, the metrics and metrics used in these updates become the leading language in the organization's risk discussions. [61]



Identify top risks, risk appetite, assess control and vulnerability

Analyze and evaluate identified risks and their relevance to the target

Treat risks that exceeds targets risk appetite

Monitor risk and their relevance to the target

Figure 43: Root-cause mitigation path *[61]*

Chapter 7: Conclusion

The Internet of Things (IoT) refers to the global infrastructure of the information society and realizes progressive benefits by connecting and developing physical and virtual objects based on existing information and communication technology. The Internet of Things has become one of the most advanced and exciting innovations in research, making it economically attractive to the business world. It includes connecting different devices and pairing devices with people. We live in the period of big data, and there is always a need to apply AI/ML to analyze big data collected in the cloud quickly and accurately. IoT cybersecurity issues include significant cyber-attacks, data fraud, remote access, hacking, and security vulnerabilities. However, AI is now a serious threat to cloud vulnerabilities and IoT device networks.

Artificial Intelligence and Machine Learning have made noticeable changes in implementing and processing various computer engineering and networking devices. Due to its highly adaptable cyber-physical system nature, AI comprises many interconnected devices, so data movement and analytics occur in a complex-wide area network. Advances in new technologies and the digitization of everything worldwide lead to increased cyber-attacks. It also provides a single platform for creating new attacks, from simple DDoS attacks to advanced WannaCry ransomware attacks. Phase 1 of this research provides a detailed description of IoT and its applications, cloud computing, artificial intelligence, and machine learning.

Identity and access management supports multiple public cloud environments and integrates identity and access management controls on-premises and public cloud resources. Cloud data encryption in hibernation must be standardized. Managing key encryption for cloud workloads must be established. Defining a perimeter is highly recommended in this research. Nowadays, many organizations are into the virtual perimeter model. The transition to these perimeter-control models may typically involve developing clean sheet designs that draw services like application firewall, web gateway, and network monitoring from third parties that support multiple clouds. Monitoring cloud-based operations in all aspects is a necessity. Cloud-based applications must use security tools or templates to rely on current security information and event management (SIEM) tools. It helps them maintain a single view of their on-premises and cloud workloads. Server-side endpoint control and user endpoint controls are clearly distinguished in this research. Phase 2 of this research paper uses a software implementation that explains the regulatory governance to tackle emerging cybersecurity issues.

Regulations on data protection govern most cybersecurity programs, such as the European Union’s General Data Protection Regulation. Personally identifiable information, data location and sovereignty, and resilience are essential for an effective cyber defense strategy. Resilience assumes that attacks are a constant feature of the digital business environment, and some of these attacks inevitably lead to breaches. Therefore, creating enough Resilience to continue the business when dealing with hacking and recovery after hacking is the most critical component of today's cyber defense strategy.

A forward-looking strategy takes a never trust, always verify approach to combat emerging cybersecurity issues. Advanced technology such as AI/ML, automation, cloud computing, and Agile development address them to access their data and processes, acknowledging that threats evolve and require existing capabilities. Through leveraging such services, a more resilient and adaptable cybersecurity model can be embraced by positioning itself to survive new disputes and take advantage of emerging digital ecosystems.

Phase 3 provides a predictive analysis using AI/ML to prevent future attacks. This research paper exhibits the future development of 5G networks, 3GPP, and MIMO as they play a specific role in cloud computing. AI/ML is at entry-level in NGN that has access to complete network infrastructure, distributed architecture and its users, automation, and simulation. The virtual network infrastructure gives AI/ML convergence by increasing the complexity of the network due to the improvement of customer use and satisfaction. AI/ML and IoT create a dynamic network environment. This change creates a dynamic and agile network where it learns autonomously to improve performance and uses its resources effectively, simultaneously improving the full power of 5G benefits. It may improve IoT support robust security in the end with increased revenue. This research also approaches resolving security issues by learning, analyzing, and identifying cyber attacks. Our AI/ML solutions outfit the history of attacks, recognize future attacks, and walled protection from zero-day attacks.

References

[1] "Overview of Internet of Things - Recommendations," *International Telecommunication Union - Telecommunication,* 2012.

[2] The Power of IoT, Examples, and Applications - Leverage, eBook, 2018.

[3] U. Priyadarshiny, "IoT real-time use cases, IoT applications in different domains," 2020.

[4] R. J. Sternberg, "Intelligence - Dialogues in Clinical Neuroscience," *MIT Press University Online,* 2008.

[5] W. Duch, "Studies in Computational Intelligence," 2007.

[6] J. X. Neapolitan, "Artificial intelligence - with an introduction to machine learning," no. Second edition.

[7] M. Flasiński, "Introduction to Artificial Intelligence," *Springer International Publishing,* 2016.

[8] "Artificial Intelligence - Overview," *Tutorialspoint,* p, 2021.

[9] Z. Saleh, "Artificial Intelligence Definition, Ethics, and Standards," 2019.

[10] B. Goertzel, "Concept, State of the Art, and Future Prospects," *Artificial General Intelligence,* 2014.

[11] N. Bostrom, "Superintelligence: Paths, Dangers, Strategies (Illustrated ed.)," *Oxford University Press,* 2014.

[12] A. Dönmez, "Introduction to Machine Learning," *The MIT Press,* no. 2nd ed., ISBN: 978-0-262-01243-0, p. 584 pages, 2013, 2nd ed., Cambridge, MA: 2010 Natural Language Engineering, 19(2), 285–288.

[13] N. J. Nilsson, INTRODUCTION TO MACHINE LEARNING AN EARLY DRAFT OF A PROPOSED TEXTBOOK, Department of Computer Science. Machine Learning, 56(2), 387–399, 2005.

[14] S. Ray, "A Quick Review of Machine Learning Algorithms," in *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon),* Faridabad, India., 2019.

[15] "How To Build Automation Scripts without Code," *HelpSystems.*

[16] M. Mohammed, "Machine Learning," *Amsterdam University Press,* 2016.

[17] O. Omer, "Introduction to Machine Learning," *The Wikipedia Guide.*

[18] N. Weiss, "Rule-based machine learning methods for functional prediction," *Journal of Artificial Intelligence research*, Journal of Artificial Intelligence Research, 3, 383-403. 1995.

[19] G. Bonaccorso, "Mastering Machine Learning Algorithms," in *Expert techniques to implement popular machine learning algorithms and fine-tune your models,* Packt Publishing, 2018.

[20] "Introduction To Genetic Algorithms In Machine Learning," *Software Testing Help,* 18 January 2021.

[21] H. Taylor, "Cyber security - Introduction to Cybersecurity," *Prey project,* 2018.

[22] M. Chapple, "Network security - TechTarget," *Outsourcing security services in the enterprise,* 2013.

[23] A. Harper, "Biggest security challenges for IoT," *Peerbits*, 2018.

[24] B. J. Steven, "What is threat modeling, Application and software security," *Synopsys*, 2011.

[25] A. Kothari, "AI Frameworks," *Geekflare,* 2020 .

[26] E. Pot, "Introducing Tensor Flow datasets," *Tensor Flow Blogs*, 26 February 2019.

[27] H. Bouma, "Object recognition using deep convolutional neural networks with complete transfer and partially frozen layers," *ResearchGate,* 30 March 2022.

[28] K. Hong, "Artificial Neutral Networks (ANN) - Deep Learning: Theano," *BogoToBogo,* 2020.

[29] S. Sanyal, "AI, AWS DeepLens, IoT," *AWS Machine Learning,* 2020.

[30] N. Tsaku, "Torch dataset for grid search," Github.

[31] S. Deb, "Top 10 Machine learning frameworks," *Edureka*, 2021.

[32] Lale, "Example of Scikit learn error checking," ResearcgGate.net, 2022.

[33] "Driverless AI," *doc.h20.ai*, 2021.

[34] G. Polat, "Security Issues in IoT challenges and Countermeasures," *Isaca Journal,* 2019.

[35] M. Chowdhury, "Cloud computing trends," *Analytics Insight,* 2021.

[36] "What is cloud computing," *Fastmetrics business,* 2021.

[37] S. Vennam, "Cloud computing," *IBM - Blogs,* 2020.

[38] "Cloud Deployment Models and Hybrid Cloud Computing," *CRM Trilogix,* 2001.

[39] "Cloud computing," *Salesforce,* 2021.

[40] A. Ohri, "Cloud computing and cloud monitoring tools," *Jigsaw Academy,* 2020.

[41] O. Dharmadhikari, "Leveraging Machine Learning and Artificial Intelligence for 5G," *CableLabs,* 2019.

[42] A. Khullar, "Cybersecurity risk management," *Infosys Knowledge Institute,* 2020.

[43] "Cybersecurity tools," *Software testing help,* January 2022.

[44] "Penetration Testing".*Contrast security* .

[45] M. Funk, "Web Application Penetration Testing Checklist," *Cybers Guards,* 2019.

[46] A. E. Abu-Dabaseh, "Automated Penetration Testing: An Overview," 2018.

[47] "Top 10 Ethical Hacking Tools in 2021- Leaders in Ethical Hacking," *Edureka,* 2021.

[48] "What is Wireshark? What this essential tool does and how to use it," *CSO Online,* 2021.

[49] "The Ethical Hacking Tools You Should Use and Why," *CyberVista,* 2021.

[50] "Wireshark Network Security," *Packt.,* 2021.

[51] "Top 10 Uses of Wireshark for Hackers Part I," *The Ethical Hacker Network,* 2021.

[52] Interior, "Penetration Testing," Penetration Testing. (2018, July 31). U.S. Department of the Interior 2018.

[53] C. Chebbi, "Advanced Infrastructure Penetration Testing," *Defend your systems from methodized and proficient attackers*, Packt Publishing, 2018.

[54] McKinsey, Perspectives on transforming cybersecurity - McKinsey and global risk practice, McKinsey Digital, 2019.

[55] R. Riemenschnitter, "A new posture for cybersecurity in a networked world," 2018.

[56] "European Union Agency For Network and Information Security," The Sans Institute, 2018.

[57] P. Kaminski, "Protecting your critical digital assets," 2017.

[58] W. Tucker Bailey, "Insider Threat - The Human element of Cyber Risk," 2018.

[59] R. Brad Brown, "Use active defense: To survive the advanced cyber threats," 2017.

[60] R. Arul Elumalai, "Making a secure transition to the public cloud," 2018.

[61] S. Jim Boehm, "Cyber risk measurement and the holistic cybersecurity approach," 2018.

[62] J. Cano, "Cyberattacks the instability of security and control knowledge," *ISACA JOURNAL*, 2016.

[63] A. Girma, Analysis of Security Vulnerability and Analytics of Internet of Things (IoT) Platform, vol. 738, 2018.

[64] R. Riemenschnitter, "A new posture for cybersecurity in the networked world," 2018.