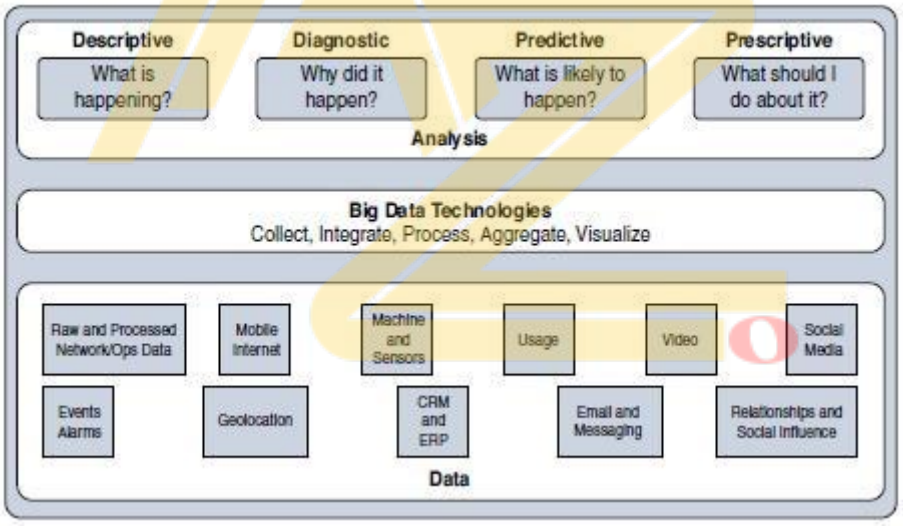
1. Differentiate between structure and unstructured data

Answer:

Structured Data	Unstructured Data
Structured data means that the data follows a model or schema that defines how the data is represented or organized	Unstructured data lacks a logical schema for understanding and decoding the data through traditional programming means.
Traditional relational database management system (RDBMS).	Data type includes text, speech, images, and video.
Structured data is easily formatted, stored, queried, and processed	Unstructured Data cannot be easily formatted stored, queried and processed
Structured data is more easily managed and processed due to its well-defined organization.	Unstructured data can be harder to deal with and typically requires very different analytics tools for processing the data.
It has been the core type of data used for making business decisions.	80% of a business's data is unstructured

2. Explain with a block diagram the types of Data analysis

Answer:



Types of Data Analysis Results

- **Descriptive:** Descriptive data analysis tells you what is happening, either now or in the past. For example, a thermometer in a truck engine reports temperature values every second. From a descriptive analysis perspective, you can pull this data at any moment to gain insight into the current operating condition of the truck engine. If the temperature value is too high, then there may be a cooling problem or the engine may be experiencing too much load.
- **Diagnostic:** When you are interested in the "why," diagnostic data analysis can provide the answer. Continuing with the example of the temperature sensor in the truck engine, you might wonder why the truck engine failed. Diagnostic analysis

might show that the temperature of the engine was too high, and the engine overheated. Applying diagnostic analysis across the data generated by a wide range of smart objects can provide a clear picture of why a problem or an event occurred.

- **Predictive:** Predictive analysis aims to foretell problems or issues before they occur. For example, with historical values of temperatures for the truck engine, predictive analysis could provide an estimate on the remaining life of certain components in the engine. These components could then be proactively replaced before failure occurs. Or perhaps if temperature values of the truck engine start to rise slowly over time, this could indicate the need for an oil change or some other sort of engine cooling maintenance.
- **Prescriptive:** Prescriptive analysis goes a step beyond predictive and recommends solutions for upcoming problems. A prescriptive analysis of the temperature data from a truck engine might calculate various alternatives to cost-effectively maintain our truck. These calculations could range from the cost necessary for more frequent oil changes and cooling maintenance to installing new cooling equipment on the engine or upgrading to a lease on a model with a more powerful engine. Prescriptive analysis looks at a variety of factors and makes the appropriate recommendation.

3. List and Explain Iot data analytics challenges.

Answer:

- **Scaling problems:** Due to the large number of smart objects in most IoT networks that continually send data, relational databases can grow incredibly large very quickly. This can result in performance issues that can be costly to resolve, often requiring more hardware and architecture changes.
- **Volatility of data:** With relational databases, it is critical that the schema be designed correctly from the beginning. Changing it later can slow or stop the database from operating. Due to the lack of flexibility, revisions to the schema must be kept at a minimum. IoT data, however, is volatile in the sense that the data model is likely to change and evolve over time. A dynamic schema is often required so that data model changes can be made daily or even hourly.
- **Live streaming** nature of its data and with managing data at the network level. It is valuable only if it is possible to analyze and respond to it in real-time. Real-time analysis of streaming data allows you to detect patterns or anomalies that could indicate a problem or a situation that needs some kind of immediate response.
- **Network Data or Network Analytics**: With the large numbers of smart objects in IoT networks that are communicating and streaming data, it can be challenging to ensure that these data flows are effectively managed, monitored, and secure. Network analytics tools provide the capability to detect irregular patterns or other problems in the flow of IoT data through a network.

4. Explain the domains which revolve around the common applications of ML for IOT.

Answer:

- **Monitoring:** Smart objects monitor the environment where they operate. Data is processed to better understand the conditions of operations. These conditions can refer to external factors, such as air temperature, humidity, or presence of carbon dioxide in a mine, or to operational internal factors, such as the pressure of a pump, the viscosity of oil flowing in a pipe. ML can be used with monitoring to detect early failure conditions
- **Behaviour control:** Monitoring commonly works in conjunction with behaviour control. When a given set of parameters reach a target threshold—defined in advance or learned dynamically through deviation from mean values monitoring functions generate an alarm. This alarm can be relayed to a human, but a more efficient and more advanced system would trigger a corrective action, such as increasing the flow of fresh air in the mine tunnel, turning the robot arm, or reducing the oil pressure in the pipe.
- **Operations optimization:** Behaviour control typically aims at taking corrective actions based on thresholds. However, analyzing data can also lead to changes that improve the overall process. For example, a water purification plant in a smart city can implement a system to monitor the efficiency of the purification process based on which chemical is used, at what temperature, and associated to what stirring. Neural networks can combine multiples of such units, in one or several layers, to estimate the best chemical and stirring mix for a target air temperature. This intelligence can help the plant reduce its consumption of chemicals while still operating at the same purification efficiency level. As a result of the learning, behaviour control results in different machine actions. The objective is not merely to pilot the operations but to improve the efficiency and the result of these operations.
- **Self-healing, self-optimizing:** A fast-developing aspect of deep learning is the closed loop. ML-based monitoring triggers changes in machine behaviour and operations optimizations. The ML engine can be programmed to dynamically monitor and combine new parameters and automatically deduce and implement new optimizations when the results demonstrate a possible gain. The system becomes self-learning and self-optimizing. It also detects new K-means deviations that result in pre-detection of new potential defects, allowing the system to self-heal. The healing is not literal, as external factors have to intervene, but the diagnosis is automated.

5. List and describe the "Three Vs" to categorize big Data.

Answer:

- **Velocity:** Velocity refers to how quickly data is being collected and analyzed. Hadoop Distributed File System is designed to ingest and process data very quickly. Smart objects can generate machine and sensor data at a very fast rate and require database or file systems capable of equally fast ingest functions.

- **Variety:** Variety refers to different types of data. Often you see data categorized as structured, semi-structured, or unstructured. Different database technologies may only be capable of accepting one of these types. Hadoop is able to collect and store all three types. This can be beneficial when combining machine data from IoT devices

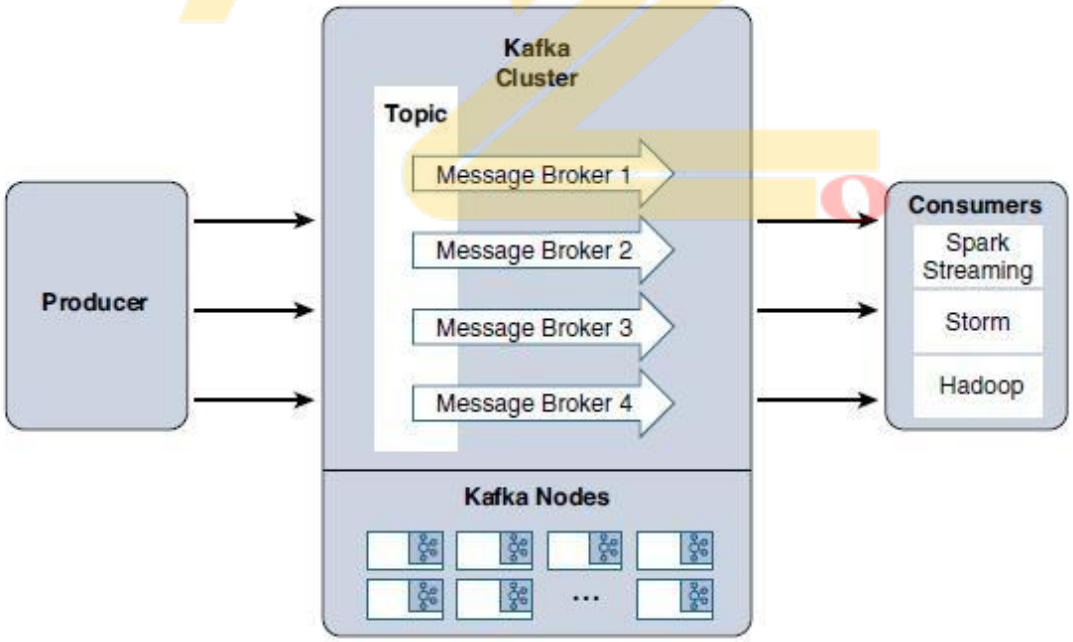
that is very structured in nature with data from other sources, such as social media or multimedia that is unstructured.

- **Volume:** Volume refers to the scale of the data. Typically, this is measured from gigabytes on the very low end to petabytes or even exabytes of data on the other extreme. Generally, big data implementations scale beyond what is available on locally attached storage disks on a single node. It is common to see clusters of servers that consist of dozens, hundreds, or even thousands of nodes for some large deployments.

6. Write a short note on Apache kafka.

Answer:

- Apache Kafka is a distributed publisher-subscriber messaging system that is built to be scalable and fast.
- It is composed of topics, or message brokers, where producers write data and consumers read data from these topics.
- Due to the distributed nature of Kafka, it can run in a clustered configuration that can handle many producers and consumers simultaneously and exchanges information between nodes, allowing topics to be distributed over multiple nodes.
- The goal of Kafka is to provide a simple way to connect to data sources and allow consumers to connect to that data in the way they would like.



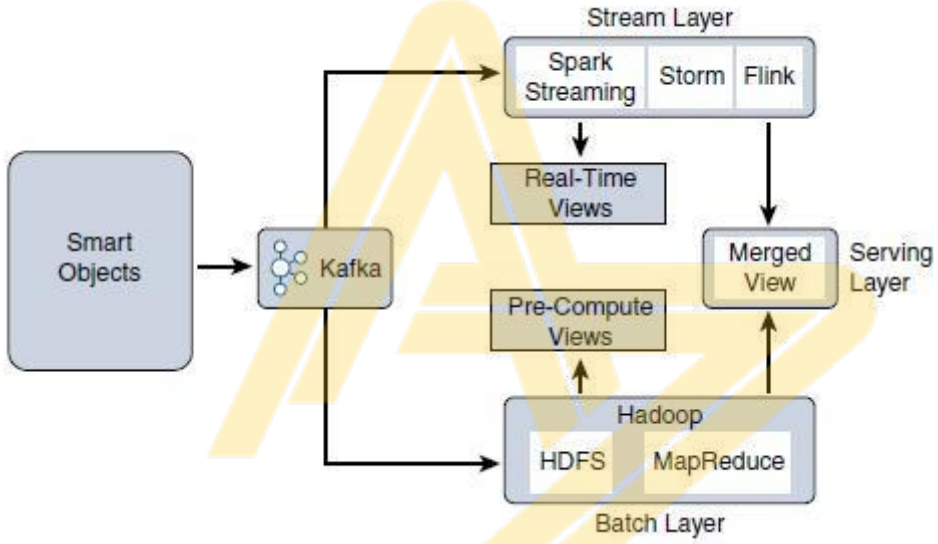
Apache Kafka Data Flow

- **Apache Spark** is an in-memory distributed data analytics platform designed to accelerate processes in the Hadoop ecosystem.

- The "in-memory" characteristic of Spark is what enables it to run jobs very quickly. At each stage of a Map Reduce operation, the data is read and written back to the disk, which means latency is introduced through each disk operation.
- Real-time processing is done by a component of the Apache Spark project called Spark Streaming. Spark Streaming is an extension of Spark Core that is responsible for taking live streamed data from a messaging system, like Kafka, and dividing it into smaller micro batches. These micro batches are called discretized streams
- **Apache Storm and Apache Flink** are other Hadoop ecosystem projects designed for distributed stream processing and are commonly deployed for IoT use cases. Storm can pull data from Kafka and process it in a near-real-time fashion, and so can Apache Flink.

7. Write a short note on Lambda architecture.

Answer:



Lambda Architecture

1. Lambda is a data management system that consists of two layers for ingesting data (Batch and Stream) and one layer for providing the combined data (Serving).
2. These layers allow for the packages like Spark and Map Reduce, to operate on the data independently, focusing on the key attributes for which they are designed and optimized.
3. Data is taken from a message broker, commonly Kafka, and processed by each layer in parallel, and the resulting data is delivered to a data store where additional processing or queries can be run.
4. Layers of Lambda Architecture are:

   - **Stream layer:** This layer is responsible for near-real-time processing of events. Technologies such as Spark Streaming, Storm, or Flink are used to quickly ingest, process, and analyze data on this layer.

   - **Batch layer:** The Batch layer consists of a batch-processing engine and data store. If an organization is using other parts of the Hadoop ecosystem for the other layers, Map Reduce and HDFS can easily fit.

- **Serving layer:** The Serving layer is a data store and mediator that decides which of the ingest layers to query based on the expected result or view into the data. The Serving layer is often used by the data consumers to access both stream and batch layers simultaneously.

5. The Lambda Architecture can provide a robust system for collecting and processing massive amounts of data and the flexibility of being able to analyze that data at different rates.

6. One **limitation** of this type of architecture is its place in the network. Due to the processing and storage requirements of many of these pieces, the vast majority of these deployments are either in data centres or in the cloud. This could limit the effectiveness of the analytics to respond rapidly enough if the processing systems are milliseconds or seconds away from the device generating the data.

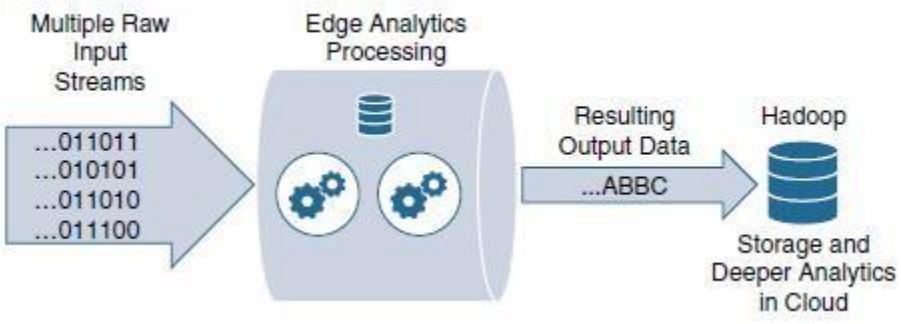
8. List and explain edge analytics core functions and illustrate edge analytics processing unit.

Answer:

- **Raw input data:** This is the raw data coming from the sensors into the analytics processing unit.

- **Analytics processing unit (APU):** The APU filters and combines data streams, organizes them by time windows, and performs various analytical functions. It is at this point that the results may be acted on by micro services running in the APU.

- **Output streams:** The data that is output is organized into insightful streams and is used to influence the behaviour of smart objects, and passed on for storage and further processing in the cloud. Communication with the cloud often happens through a standard publisher/subscriber messaging protocol, such as MQTT.



Edge Analytics Processing Unit

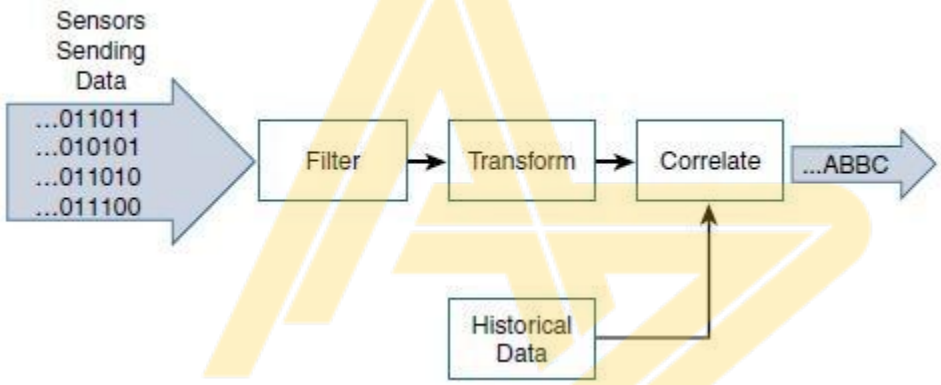
APU needs to perform the following functions:

a. Filter: The streaming data generated by IoT endpoints is likely to be very large, and most of it is irrelevant. The filtering function identifies the information that is considered important.

b. Transform: In the data warehousing world, Extract, Transform, and Load (ETL) operations are used to manipulate the data structure into a form that can be used for other purposes. Analogous to data warehouse ETL operations, in streaming analytics, once the data is filtered; it needs to be formatted for processing.

c. Time: As the real-time streaming data flows, a timing context needs to be established. This could be to correlated average temperature readings from sensors on a minute-by-minute basis. The APU is programmed to report the average temperature every minute from the sensors, based on an average of the past two minutes.

d. Correlate: Streaming data analytics becomes most useful when multiple data streams are combined from different types of sensors. Different types of data come from different instruments, but when this data is combined and analyzed, it provides an invaluable picture of the situation. Another key aspect is combining and correlating real-time measurements with pre-existing, or historical, data.

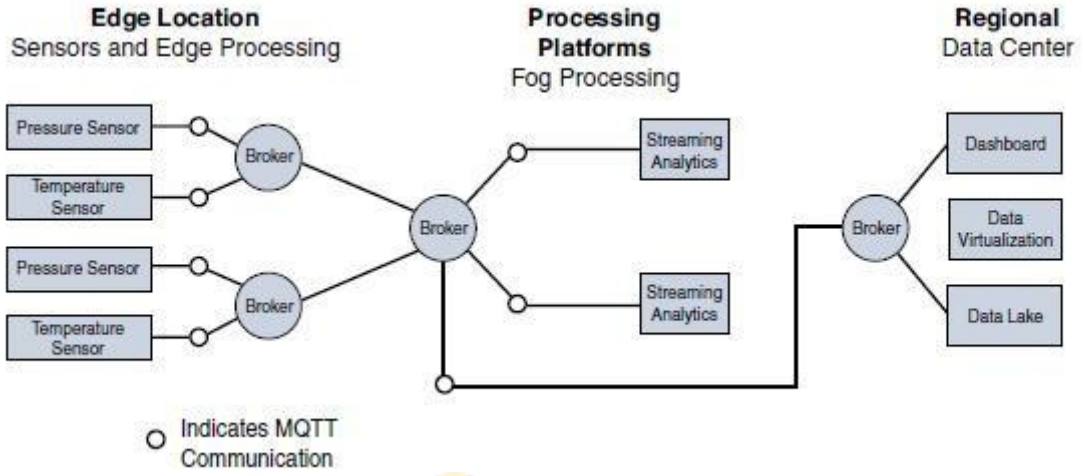


Correlating Data Streams with Historical Data

e. Match patterns: Once the data streams are properly cleaned, transformed, and correlated with other live streams as well as historical data sets, pattern matching operations are used to gain deeper insights to the data. The patterns can be simple relationships, or they may be complex, based on the criteria defined by the application. Machine learning may be leveraged to identify these patterns.

f. Improve business intelligence: Ultimately, the value of edge analytics is in the improvements to business intelligence that were not previously available.

9. Distributed analytics throughout the IoT system.

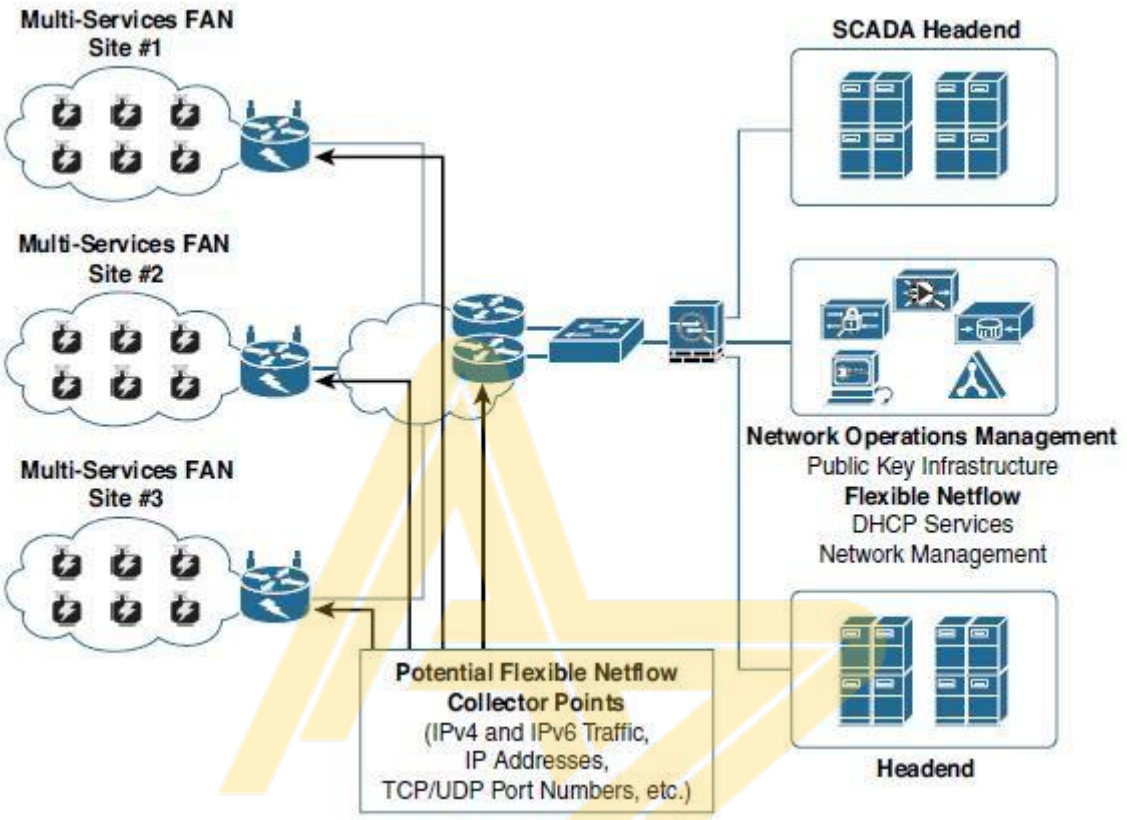


Distributed Analytics throughout the IoT System

- Streaming analytics may be performed directly at the edge, in the fog, or in the cloud data centre. There are no hard-and-fast rules dictating where analytics should be done, but there are a few guiding principles.
- Fog analytics allows you to see beyond one device, giving you visibility into an aggregation of edge nodes and allowing correlating data from a wider set.
- While there may be some value in doing analytics directly on the edge, the sensors communicate via MQTT through a message broker to the fog analytics node, allowing a broader data set.
- An example of an oil drilling company that is measuring both pressure and temperature on an oil rig.The fog node is located on the same oil rig and performs streaming analytics from several edge devices, giving it better insights due to the expanded data set. It may not be able to respond to an event as quickly as analytics performed directly on the edge device, but it is still close to responding in real-time as events occur. Once the fog node is finished with the data, it communicates the results to the cloud through a message broker via MQTT for deeper historical analysis through big data analytics tools.

10. Demonstrate smart grid FAN analytics with Net-Flow example.

Answer:



Smart Grid FAN Analytics with Net Flow

- Figure shows field area network (FAN) traffic analytics performed on the aggregation router in a smart grid.
- IoT endpoints, contrary to generic computing platforms, are designed to directly communicate with a very small number of specific application servers, such as an IoT message or data broker, or specific application servers and network management systems. Therefore, it could be said that IoT solutions and use cases tightly couple devices and applications.
- Network analytics has the power to analyze details of communications patterns made by protocols and correlate this across the network. It allows you to understand what should be considered normal behavior in a network and to quickly identify anomalies that suggest network problems due to suboptimal paths, intrusive malware, or excessive congestion.
- Network analytics offer capabilities to cope with capacity planning for scalable IoT deployment as well as security monitoring in order to detect abnormal traffic volume and patterns such as an unusual traffic spike for a normally quiet protocol for both centralized or distributed architectures, such as fog computing.

11. Explain the benefits of flow analytics or network analytics.

Answer:

The benefits of flow analytics, in addition to other network management services, are as follows:

Network traffic monitoring and profiling: Flow collection from the network layer provides global and distributed near-real-time monitoring capabilities. IPv4 and IPv6 network wide traffic volume and pattern analysis helps administrators proactively detect problems and quickly troubleshoot and resolve problems when they occur.

Application traffic monitoring and profiling: Monitoring and profiling can be used to gain a detailed time-based view of IoT access services, such as the application layer protocols, including MQTT, CoAP, and DNP3, as well as the associated applications that are being used over the network.

Capacity planning: Flow analytics can be used to track and anticipate IoT traffic growth and help in the planning of upgrades when deploying new locations or services by analyzing captured data over a long period of time. This analysis affords the opportunity to track and anticipate IoT network growth on a continual basis.

Security analysis: Because most IoT devices typically generate a low volume of traffic and always send their data to the same server(s), any change in network traffic behaviour may indicate a cyber security event, such as a denial of service (DoS) attack. Security can be enforced by ensuring that no traffic is sent outside the scope of the IoT domain.

Accounting: In field area networks, routers or gateways are often physically isolated and leverage public cellular services and VPNs for backhaul. Deployments may have thousands of gateways connecting the last-mile IoT infrastructure over a cellular network. Flow monitoring can thus be leveraged to analyze and optimize the billing, in complement with other dedicated applications, such as Cisco Jasper, with a broader scope than just monitoring data flow.

Data warehousing and data mining: Flow data can be warehouse for later retrieval and analysis in support of proactive analysis of multiservice IoT infrastructures and applications.

12. Write a shot note on flexible Net-Flow Architecture.

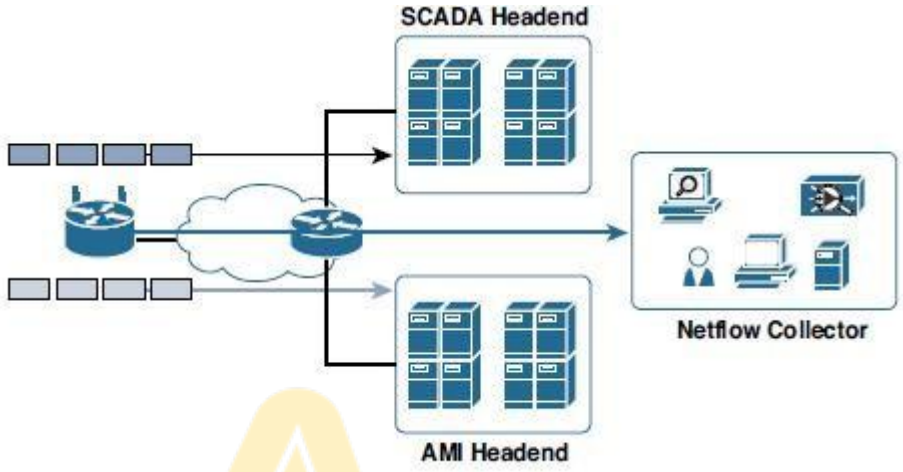
Answer:

FNF is a flow technology developed by Cisco Systems that is widely deployed all over the world. Key advantages of FNF are as follows:

- Flexibility, scalability, and aggregation of flow data
- Ability to monitor a wide range of packet information and produce new information about network behaviour:
- Enhanced network anomaly and security detection

- User-configurable flow information for performing customized traffic identification and ability to focus and monitor specific network behaviour
- Convergence of multiple accounting technologies into one accounting mechanism

FNF Components:



Flexible Net Flow

- **FNF Flow Monitor (Net Flow cache):** The FNF Flow Monitor describes the Net Flow cache or information stored in the cache. The Flow Monitor contains the flow record definitions with key fields (used to create a flow, unique per flow record: match statement) and non-key fields (collected with the flow as attributes or characteristics of a flow) within the cache.

- **FNF flow record:** A flow record is a set of key and non-key Net Flow field values used to characterize flows in the Net Flow cache. Flow records may be predefined for ease of use or customized and user defined. A typical predefined record aggregates flow data and allows users to target common applications for Net Flow. User-defined records allow selections of specific key or non-key fields in the flow record.

- **FNF Exporter:** There are two primary methods for accessing Net Flow data: Using the show commands at the command-line interface (CLI), and using an application reporting tool. Net Flow Export to the Net Flow reporting collector. The Flexible Net Flow Exporter allows the user to define where the export can be sent, the type of transport for the export, and properties for the export. Multiple exporters can be configured per Flow Monitor.

- **Flow export timers:** Timers indicate how often flows should be exported to the collection and reporting server.

- **Net Flow export format:** This simply indicates the type of flow reporting format.

- **Net Flow server for collection and reporting:** This is the destination of the flow export. It is often done with an analytics tool that looks for anomalies in the traffic patterns.

- **Flexible Net Flow in Multiservice IoT Networks:**

It is recommended that FNF be configured on the routers that aggregate connections from the last mile's routers. This gives a global view of all services flowing between the core network in the cloud and the IoT last-mile network.

Challenges with deploying flow analytics tools in an IoT network include the following:

a. The distributed nature of fog and edge computing may mean that traffic flows are processed in places that might not support flow analytics, and visibility is thus lost.

b. IPv4 and IPv6 native interfaces sometimes need to inspect inside VPN tunnels, which may impact the router's performance.

c. Additional network management traffic is generated by FNF reporting devices. The added cost of increasing bandwidth thus needs to be reviewed, especially if the backhaul network uses cellular or satellite communications.

13. Discuss the common challenges faced in OT security

OR

Discuss the common challenges faced IoT with respect to security

Answer:

- **Erosion of Network Architecture:** Two of the major challenges in securing industrial environments have been initial design and ongoing maintenance. The initial design challenges arose from the concept that networks were safe due to physical separation from the enterprise with minimal or no connectivity to the outside world, and the assumption that attackers lacked sufficient knowledge to carry out security attacks. Over time, what may have been a solid design to begin with is eroded through ad hoc updates and individual changes to hardware and machinery without consideration for the broader network impact. This led to miscalculations of expanding networks and the introduction of wireless communication in a standalone fashion, without consideration of the impact to the original security design. These uncontrolled or poorly controlled OT network evolutions have, in many cases, over time led to weak or inadequate network and systems security.

- **Pervasive Legacy Systems:** Due to the static nature and long lifecycles of equipment in industrial environments, many operational systems may be deemed legacy systems. Legacy components are not restricted to isolated network segments but have now been consolidated into the IT operational environment. Legacy components are not restricted to isolated network segments but have now been consolidated into the IT operational environment. From a security perspective, this is potentially dangerous as many devices may have historical vulnerabilities or weaknesses that have not been patched and updated.

- **Insecure Operational Protocols:** Many industrial control protocols, particularly those that are serial based, were designed without inherent strong security

requirements. Their operation was often within an assumed secure network. **Common industrial protocols and their respective security concerns are as follows:**

- Modbus: Authentication of communicating endpoints was not a default operation because it would allow an inappropriate source to send improper commands to the recipient. Some older and serial-based versions of Modbus communicate via broadcast. The ability to curb the broadcast function does not exist in some versions. There is potential for a recipient to act on a command that was not specifically targeting it.
- DNP3: participants allow for unsolicited responses, which could trigger an undesired response. The missing security element here is the ability to establish trust in the system's state and thus the ability to trust the veracity of the information being presented.
- ICCP (Inter-Control Center Communications Protocol): System did not require authentication for communication. Second, encryption across the protocol was not enabled as a default condition, thus exposing connections to man-in-the-middle (MITM) and replay attacks.
- OPC (OLE for Process Control): Dependence on the Remote Procedure Call (RPC) protocol, which creates two classes of exposure. The first requires you to clearly understand the much vulnerability associated with RPC, and the second requires you to identify the level of risk these vulnerabilities bring to a specific network.
- International Electro technical Commission (IEC) Protocols: Three message types were initially defined: MMS (Manufacturing Message Specification), GOOSE (Generic Object Oriented Substation Event), and SV (Sampled Values). Authentication is embedded in MMS, but it is based on clear-text passwords, and authentication is not available in GOOSE or SV. Firmware is typically not signed, which means there is no way to verify its authenticity or integrity. GOOSE and SV have limited message integrity, which makes it relatively easy to impersonate a publisher.

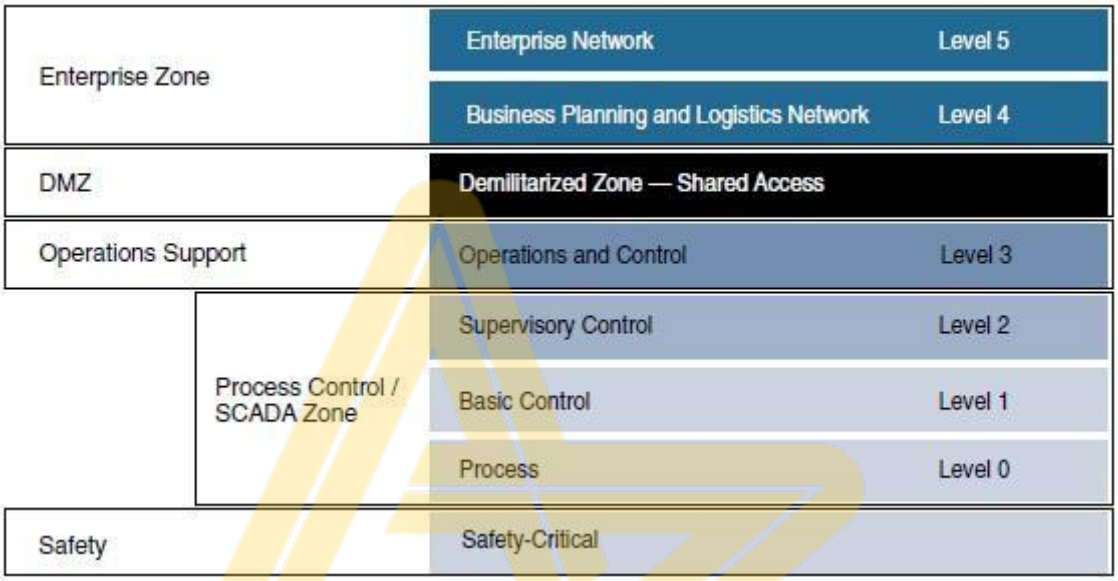
- **Device Insecurity:** Installation base of legacy systems, control and communication elements themselves have a history of vulnerabilities. Many of the systems utilize software packages that can be easily downloaded and worked against. They operate on common hardware and standard operating systems. Components used within those applications are well known to traditionally IT-focused security researchers. There is little need to develop new tools or techniques when those that have long been in place are sufficiently adequate to breach the target's defences

- **Dependence on External Vendors:** Direct and on-demand access to critical systems on the plant floor or in the field is sometimes written directly into contracts or is required for valid product warranties. This has clear benefits in many industries as it allows vendors to remotely manage and monitor equipment and to proactively alert the customer if problems are beginning to creep in. While contracts may be written to describe equipment monitoring and management requirements with explicit statements of what type of access is required and under what conditions, they generally fail to address questions of shared liability for security breaches or processes to ensure communication security.

- **Security Knowledge:** In the industrial operations space, the technical investment is primarily in connectivity and compute. It has seen far less investment in security relative to its IT counterpart. Due to the importance of security in the industrial space, all likely attack surfaces are treated as unsafe.

14. Write a short note on Purdue Model for control hierarchy with diagram.

Answer:



The Logical Framework Based on the Purdue Model for Control Hierarchy

This model identifies levels of operations and defines each level:

- Enterprise zone:
  - **Level 5: Enterprise network:** Corporate-level applications such as Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), document management, and services such as Internet access and VPN entry from the outside world exist at this level.
  - **Level 4: Business planning and logistics network:** The IT services exist at this level and may include scheduling systems, material flow applications, optimization and planning systems, and local IT services such as phone, email, printing, and security monitoring.

- Industrial demilitarized zone
  - **DMZ:** The DMZ provides a buffer zone where services and data can be shared between the operational and enterprise zones. It also allows for easy segmentation of organizational control. By default, no traffic should traverse the DMZ; everything should originate from or terminate on this area.

- Operational zone:
  - **Level 3: Operations and control:** This level includes the functions involved in managing the workflows to produce the desired end products and for monitoring and controlling the entire operational system.

- **Level 2: Supervisory control:** This level includes zone control rooms, controller status, control system network/application administration, and other control related applications.
- **Level 1: Basic control:** At this level, controllers and IEDs, dedicated HMIs (human-machine interface), and other applications may talk to each other to run part or all of the control function.
- **Level 0: Process:** This is where devices such as sensors and actuators and machines such as drives, motors, and robots communicate with controllers or IEDs.

- Safety zone:
  - **Safety-critical:** This level includes devices, sensors, and other equipment used to manage the safety functions of the control system.

15. Compare the nature of traffic flows across IT and OT Networks.

Answer:

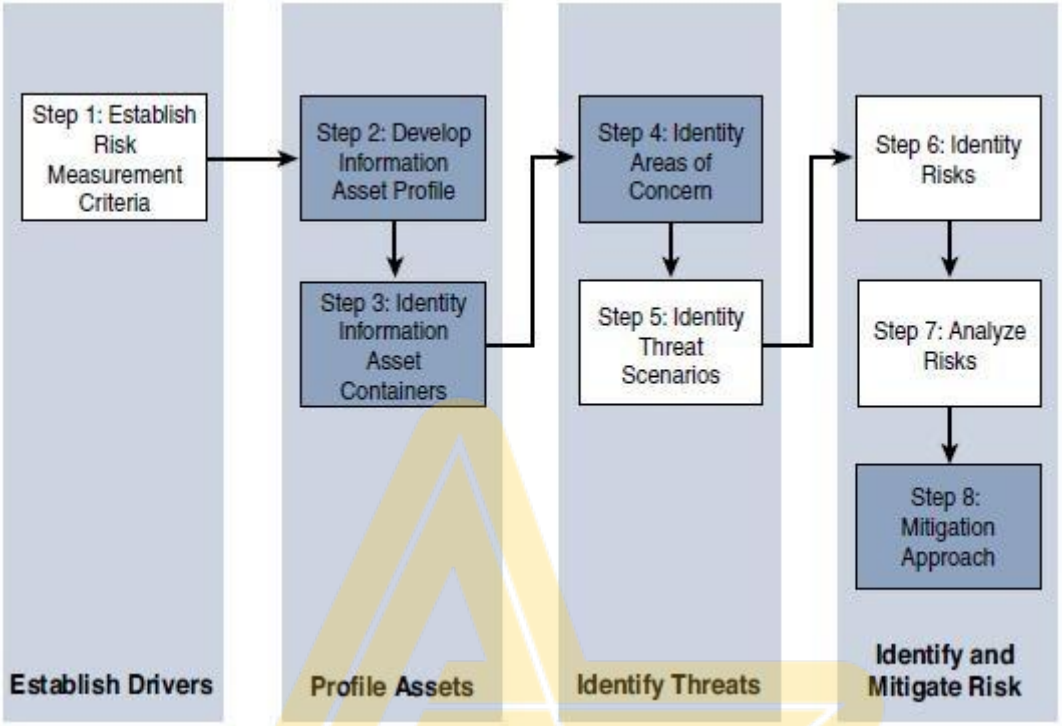
IT networks: In an IT environment, there are many diverse data flows. The communication data flows that emanate from a typical IT endpoint travel relatively far. They frequently traverse the network through layers of switches and eventually make their way to a set of local or remote servers, which they may connect to directly. Data in the form of email, file transfers, or print services will likely all make its way to the central data center, where it is responded to, or triggers actions in more local services, such as a printer. In the case of email or web browsing, the endpoint initiates actions that leave the confines of the enterprise network and potentially travel around the earth.

OT networks: By comparison, in an OT environment (Levels 0–3), there are typically two types of operational traffic. The first is local traffic that may be contained within a specific package or area to provide local monitoring and closed-loop control. This is the traffic that is used for real-time (or near-real-time) processes and does not need to leave the process control levels. The second type of traffic is used for monitoring and control of areas or zones or the overall system. SCADA traffic is a good example of this, where information about remote devices or summary information from a function is shared at a system level so that operators can understand how the overall system, or parts of it, is operating. They can then implement appropriate control commands based on this information.

16. Illustrate Formal Risk Analysis structures Octave and Fair.

Answer:

OCTAVE:



OCTAVE

- The first step of the OCTAVE Allegro methodology is to establish a risk measurement criterion. The point of having a risk measurement criterion is that at any point in the later stages, prioritization can take place against the reference model.

- The second step is to develop an information asset profile. This profile is populated with assets, a prioritization of assets, attributes associated with each asset, including owners, custodians, people, explicit security requirements, and technology assets.

- The third step is to identify information asset containers. Roughly speaking, this is the range of transports and possible locations where the information might reside. The emphasis is on the container level rather than the asset level. The value is to reduce potential inhibitors within the container for information operation.

- The fourth step is to identify areas of concern. Judgments are made through a mapping of security-related attributes to more business-focused use cases. The analyst looks to risk profiles and delves into the previously mentioned risk analysis.

- The Fifth step is where threat scenarios are identified. Threats are broadly (and properly) identified as potential undesirable events. This definition means that results from both malevolent and accidental causes are viable threats.

- At the sixth step risks are identified. Within OCTAVE, risk is the possibility of an undesired outcome. This is extended to focus on how the organization is impacted.

- The seventh step is risk analysis, with the effort placed on qualitative evaluation of the impacts of the risk. Here the risk measurement criteria defined in the first step are explicitly brought into the process.

- Mitigation is applied at the eighth step. There are three outputs or decisions to be taken at this stage. One may be to accept a risk and do nothing, other than document the situation, potential outcomes, and reasons for accepting the risk. The second is to mitigate the risk with whatever control effort is required. The final possible action is to defer a decision, meaning risk is neither accepted nor mitigated.

FAIR:

- FAIR (Factor Analysis of Information Risk) is a technical standard for risk definition from The Open Group. FAIR has clear applications within operational technology.
- FAIR places emphasis on both unambiguous definitions and the idea that risk and associated attributes are measurable. Measurable, quantifiable metrics are a key area of emphasis, which should lend itself well to an operational world with a richness of operational data.
- FAIR has a definition of risk as the probable frequency and probable magnitude of loss. A clear hierarchy of sub-elements emerges, with one side of the taxonomy focused on frequency and the other on magnitude.
- Loss even frequency is the result of a threat agent acting on an asset with a resulting loss to the organization. This happens with a given frequency called the threat event frequency (TEF), in which a specified time window becomes a probability. Vulnerability here is not necessarily some compute asset weakness, but is more broadly defined as the probability that the targeted asset will fail as a result of the actions applied.
- The other side of the risk taxonomy is the probable loss magnitude (PLM), which begins to quantify the impacts, with the emphasis again being on measurable metrics.
- FAIR defines six forms of loss, four of them externally focused and two internally focused. Of particular value for operational teams are productivity and replacement loss. Response loss is also reasonably measured, with fines and judgments easy to measure but difficult to predict.

VTU Previous year Questions:

1. Discuss Big Data Analytics Tools and Technology (July 2019)
2. With a case study relate the concept of securing IOT. (July 2019)
3. Explain in detail how IT and OT security practices and systems vary in real time. (July 2019)
4. Discuss OCTAVE and FAIR formal risk analysis. (July 2019)

Exercise Questions

1. Explain i) NOSQL databases, ii) Hadoop iii) YARN
2. Explain i)Supervised Learning, ii)Unsupervised Learning iii) Neural Network
3. The Phased Application of Security in an Operational Environment.