Internal Assessment Test II – JUNE 2022

Sub:	Internet of Things (SET 1)					Sub Code:	18CS81	Branch:	CSE	
Date:	4/6/22	Duration:	90 mins	Max Marks:	50	Sem / Sec:	VIII Sem A/B/C		OBE	

	Answer any FIVE FULL Questions	MARKS	CO	RBT
1	Explain the IoT Access Technologies IEEE 802.15.4 IEEE 802.15.4g and 802.15.4e	[10]	CO2	L2
2	Explain the IoT Access Technologies LoRaWAN	[10]	CO2	L2
3	What do you mean by data analytics for IoT? Explain their Challenges.	[10]	CO4	L2
4	Explain in detail how IT and OT security practices and system vary in real time	[10]	CO4	L3
5	Explain how Machine Learning aids in IoT data analysis	[10]	CO4	L2
6	Summarize the following i. Massively Parallel Processing Databases ii. Hadoop iii. No SQL	[10]	CO4	L2

1. **Explain the IoT Access Technologies IEEE 802.15.4 IEEE 802.15.4g and 802.15.4e**

IEEE 802.15.4:

IEEE 802.15.4 is a wireless access technology for low-cost and low-data-rate devices that are powered or run on batteries.

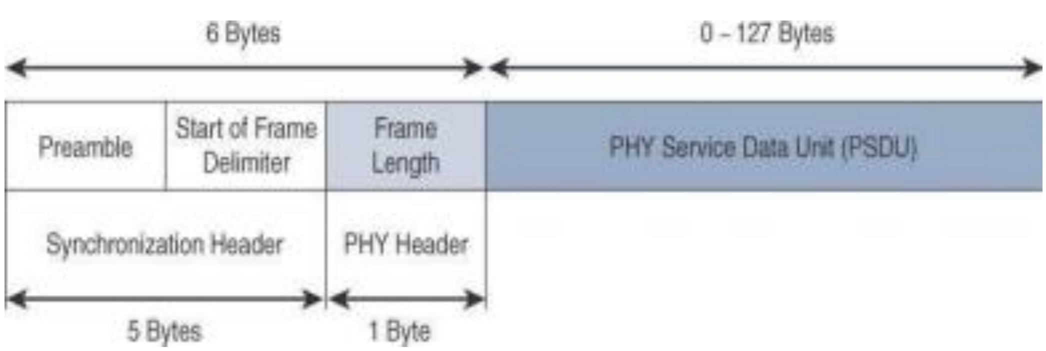
This access technology enables easy installation using a compact protocol stack while remaining both simple and flexible.

IEEE 802.15.4 is commonly found in the following types of deployments:

- Home and building automation
- Automotive networks
- Industrial wireless sensor networks
- Interactive toys and remote controls

Criticisms of IEEE 802.15.4 often focus on its MAC reliability, unbounded Latency, and susceptibility to interference and multipath fading.

Interference and multipath fading occur with IEEE 802.15.4 because it lacks a Frequency-hopping technique.



IEEE 802.15.4 PHY Format

- · The PHY Header portion of the PHY frame is shown in above figure is simply a frame length value.
- · It lets the receiver know how much total data to expect in the PHY service data unit PSDU) portion of the 802.4.15 PHY. The PSDU is the data field or payload.
- · The IEEE 802.15.4 MAC layer manages access to the PHY channel by defining how devices in the same area will share the frequencies allocated.

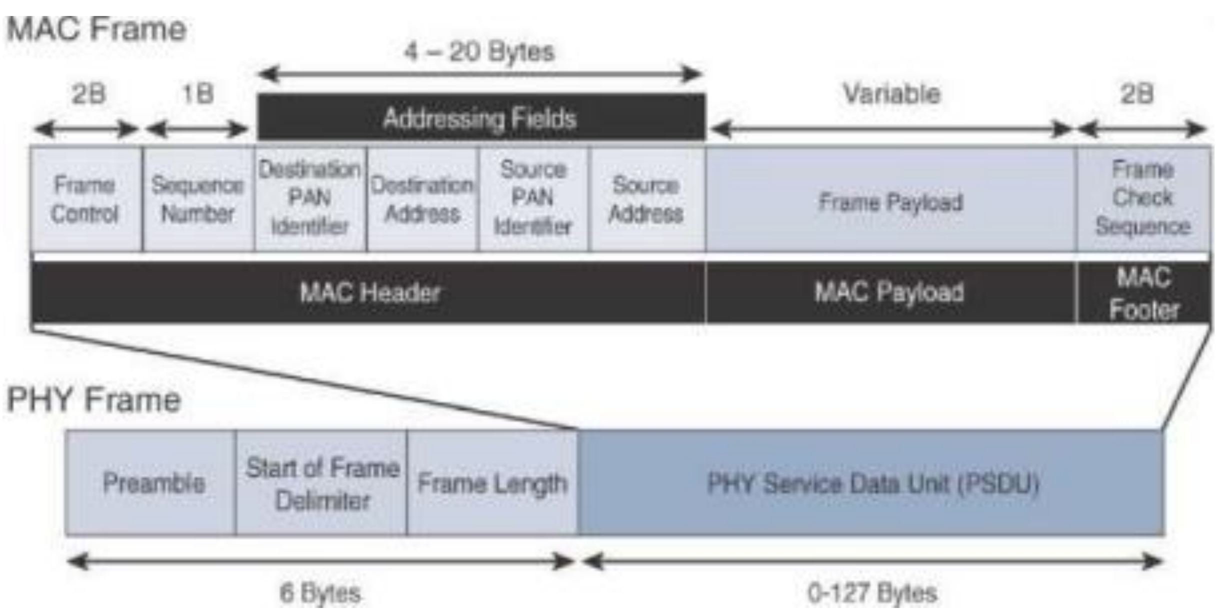
802.15.4 MAC layer

· at this layer, the scheduling and routing of data frames are also coordinated.

The 802.15.4 MAC layer performs the following tasks:

- Network beaconing for devices acting as coordinators (New devices use beacons to join an 802.15.4 network)
- PAN association and disassociation by a device
- Device security
- Reliable link communications between two peer MAC entities
- The MAC layer achieves these tasks by using various predefined frame types. In fact, four types of MAC frames are specified in 802.15.4:
- Data frame: Handles all transfers of data
- Beacon frame: Used in the transmission of beacons from a PAN coordinator ○ Acknowledgement frame: Confirms the successful reception of a frame ○ MAC command frame: Responsible for control communication between devices

· Each of these four 802.15.4 MAC frame types follows the frame format shown in below figure, notice that the MAC frame is carried as the PHY payload.



IEEE 802.15.4 MAC Format

The MAC Header field is composed of the Frame Control, Sequence Number and the Addressing fields. The Frame Control field defines attributes such as frame type, addressing modes, and other control flags. The Sequence Number field indicates the sequence identifier for the frame. The Addressing field specifies the Source and Destination PAN Identifier fields as well as the Source and Destination Address fields. The MAC Payload field varies by individual frame type. The MAC Footer field is nothing more than a frame check sequence (FCS). An FCS is a calculation based on the data in the frame that is used by the receiving side to confirm the integrity of the data in the frame.

IEEE 802.15.4g and 802.15.4e

IEEE 802.15.4g-2012 is also an amendment to the IEEE 802.15.4-2011 standard, and just like 802.15.4e-2012, it has been fully integrated into the core IEEE 802.15.4-2015 specification. 802.15.4g seeks to optimize large outdoor wireless mesh networks for field area Networks (FANs) · This technology applies to IoT use cases such as the following:

- Distribution automation and industrial supervisory control and data acquisition (SCADA) environments for remote monitoring and control
- Public lighting
- Environmental wireless sensors in smart cities
- Electrical vehicle charging stations
- Smart parking meters
- Microgrids
- Renewable energy

2. Explain the IoT Access Technologies LoRaWAN.

LoRaWAN:

It is an unlicensed-band LPWA(Low-Power Wide-Area) technology.

Standardization and Alliances

- Optimized for long-range, two-way communications and low power consumption, the technology evolved from Layer 1 to a broader scope through the creation of the LoRa Alliance.
- The LoRa Alliance quickly achieved industry support and currently has hundreds of members.
- LoRa Alliance uses the term LoRaWAN to refer to its architecture and its specifications that describe end-to-end LoRaWAN communications and protocols.

MAC Layer

The LoRaWAN specification documents three classes of LoRaWAN devices:

Class A:

- This class is the default implementation.
- Optimized for battery-powered nodes, it allows bidirectional communications, where a given node is able to receive downstream traffic after transmitting.
- Two receive windows are available after each transmission.

Class B:

- This class was designated "experimental" in LoRaWAN 1.0.1 until it can be better defined.
- A Class B node or endpoint should get additional receive windows compared to Class A, but gateways must be synchronized through a beaconing process.

Class C:

- This class is particularly adapted for powered nodes.
- This classification enables a node to be continuously listening by keeping its receive window open when not transmitting.
- LoRaWAN messages, either uplink or downlink, have a PHY payload composed of a 1-byte MAC header, a variable-byte MAC payload, and a MIC that is 4 bytes in length.

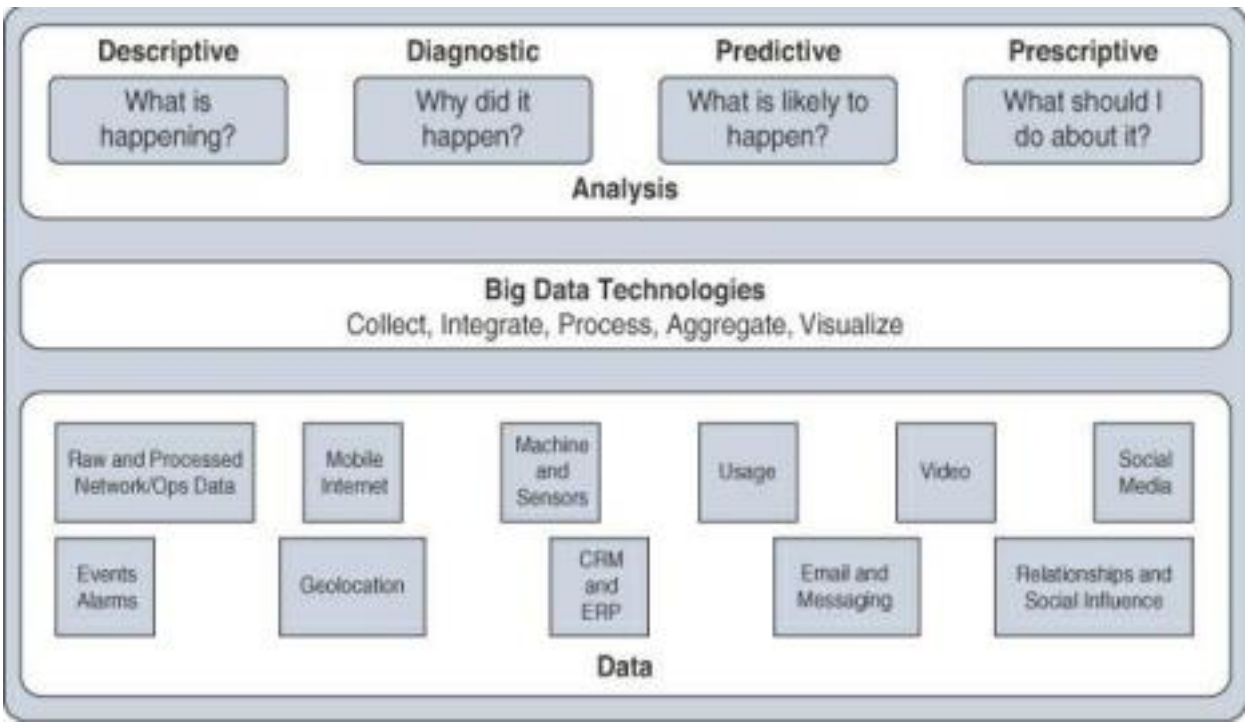
The MAC payload size depends on the frequency band and the data rate, ranging from 59 to 230 bytes for the 863–870 MHz band and 19 to 250 bytes for the 902–928 MHz band.

- In version 1.0.x, LoRaWAN utilizes six MAC message types
- **Join request:** over-the-air (OTA) activation and joining the network.
- **Join accept messages**: over-the-air (OTA) activation and joining the network.
- **Unconfirmed data up/down message** : End device does not need to acknowledge
- **Confirmed data up/down message** : A message that must be acknowledged
- **Uplink messages:** These messages are sent from endpoints to the network server and are relayed by one or more LoRaWAN gateways
- **Downlink messages:** These messages flow from the network server to a single

3. What do you mean by data analytics for IoT? Explain their Challenges.

In the world of IoT, the creation of massive amounts of data from sensors is common and one of the biggest challenges—not only from a transport perspective but also from a data management standpoint. A great example of the deluge of data that can be generated by IoT is found in the commercial aviation industry and the sensors that are deployed throughout an aircraft. Modern jet engines are fitted with thousands of sensors that generate a whopping 10GB of data per second. The potential for a petabyte (PB) of data per day per commercial airplane is not farfetched—and this is just for one airplane. Across the world, there are approximately 100,000 commercial flights per day. The amount of IoT data coming just from the commercial airline business is overwhelming. This example is but one of many that highlight the big data problem that is being exacerbated by IoT. Analyzing this amount of data in the most efficient manner possible falls under the umbrella of data analytics. Data analytics must be able to offer actionable insights and knowledge from data, no matter the amount or style, in a timely manner, or the full benefits of IoT cannot be realized. The true importance of IoT data from smart objects is realized only when the analysis of the data leads to actionable business intelligence and insights. Data analysis is typically broken down by the types of results that are produced. there are four 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.



Types of Data Analysis Results

Diagnostic:

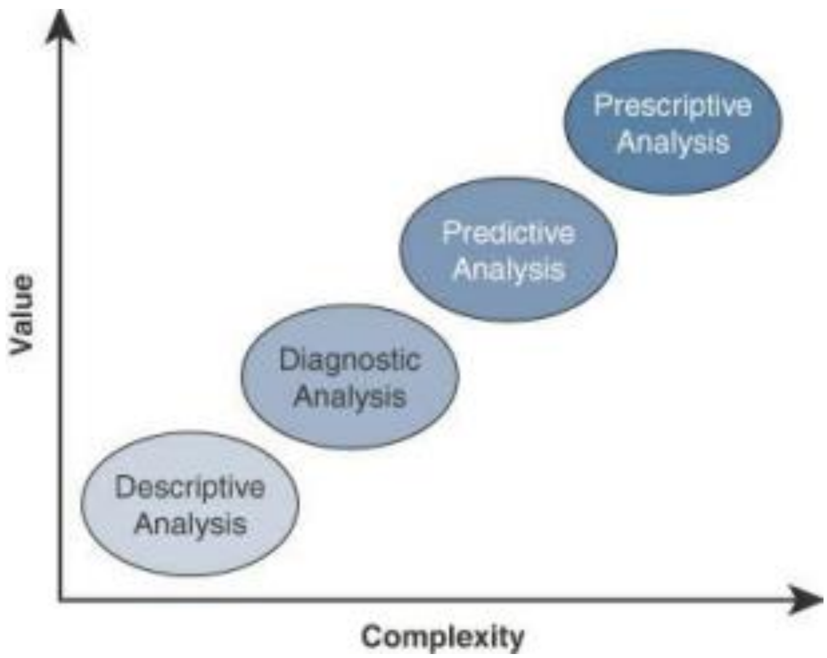
When you are interested in the "why," diagnostic data analysis canprovide the answer. Continuing with the example of the temperature sensor in the truck engine, you might wonder why the truck engine failed.· Diagnostic analysis might show that the temperature of the engine was too high, and the engine 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.



Both predictive and prescriptive analyses are more resource intensive and increase complexity, but the value they provide is much greater than the value from descriptive and diagnostic analysis.

IoT data places two specific challenges on a relational database:

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.
● A dynamic schema is often required so that data model changes can be made daily or even hourly.

4. Explain how Machine Learning aids in IoT data analysis.

Machine learning, deep learning, neural networks, and convolutional Networks are words you have probably heard in relation to big data and IoT. ML is indeed central to IoT. Data collected by smart objects needs to be analyzed, and intelligent actions need to be taken based on these analyses. Performing this kind of operation manually is almost impossible.
Machine learning is, in fact, part of a larger set of technologies commonly grouped under the term artificial intelligence (AI)
ML is concerned with any process where the computer needs to receive a set of data that is processed to help perform a task with more efficiency.
ML is a vast field but can be simply divided in two main categories: supervised and unsupervised learning.

Supervised Learning
In supervised learning, the machine is trained with input for which there is a known correct answer.
For example, suppose that you are training a system to recognize when there is a human in a mine tunnel. A sensor equipped with a basic camera can capture shapes and return them to a computing system that is responsible for determining whether the shape is a human or something else (such as a vehicle, a pile of ore, a rock, a piece of wood, and so on).

• With supervised learning techniques, hundreds or thousands of images are fed into the machine, and each image is labeled (human or nonhuman in this case). **This is called the training set.** • An algorithm is used to determine common parameters and common differences between the images.
• The comparison is usually done at the scale of the entire image, or pixel by pixel. Images are resized to have the same characteristics (resolution, color depth,position of the central figure, and so on), and each point is analyzed.

• Human images have certain types of shapes and pixels in certain locations (which correspond to the position of the face, legs, mouth, and so on).
• Each new image is compared to the set of known “good images,” and a deviation is calculated to determine how different the new image is from the average human image and, therefore, the **probability that what is shown is a human figure. This process is called classification.**
After training, the machine should be able to recognize human shapes. Before real field deployments, the machine is usually tested with unlabeled pictures— **this is called the validation or the test set, depending on the ML system used**— to verify that the recognition level is at acceptable thresholds. If the machine does not reach the level of success expected, more training is needed.

Unsupervised:
For example, you may decide to group the engines by the sound they make at a given temperature. A standard function to operate this grouping, K-means clustering, finds the mean values for a group of engines (for example, mean value for temperature, mean frequency for sound). Grouping the engines this way can quickly reveal several types of engines that all belong to the same category (for example, small engine of chainsaw type, medium engine of lawnmower type).
All engines of the same type produce sounds and temperatures in the same range as the other members of the same group.
There will occasionally be an engine in the group that displays unusual characteristics (slightly out of expected temperature or sound range). This is the engine that you send for manual evaluation. The computing process associated with this determination is called unsupervised learning. This type of learning is unsupervised because there is not a “good” or “bad” answer known in advance. It is the variation from a group behavior that allows the computer to learn that something is different.

5. **Explain in detail how IT and OT security practices and system vary in real time**

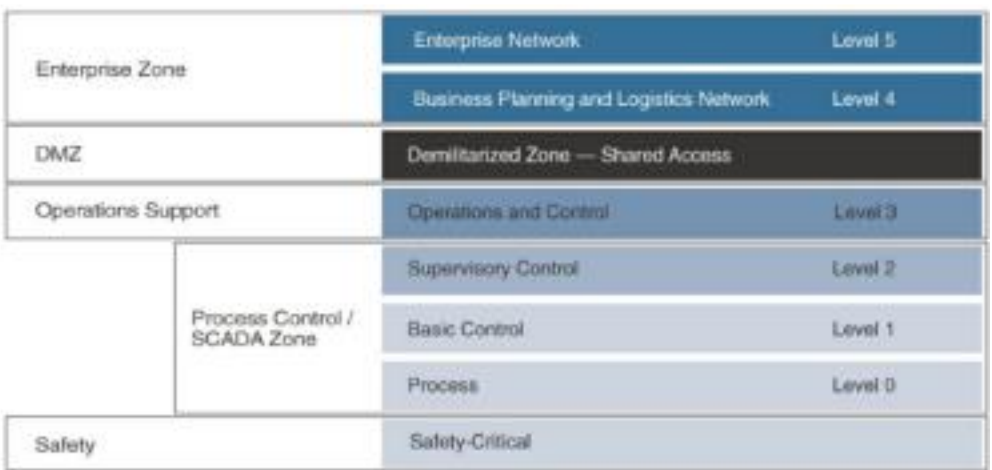
   The differences between an enterprise IT environment and an industrial focused OT deployment are important to understand because they have a direct impact on the security practice applied to them.

Purdue Model for Control Hierarchy This model identifies levels of operations and defines each level. The enterprise and operational domains are separated into different zones and kept in strict isolation via an industrial demilitarized zone (DMZ):

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.

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.

• This could include production scheduling, reliability assurance, system wide control optimization, security management, network management, and potentially other required IT services.

• **Level 2**: Supervisory control: This level includes zone control rooms, controller status, control system network/application administration, and other control-related applications, such as human machine interface (HMI)

• **Level 1**: Basic control: At this level, controllers and dedicated HMIs, 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 HMIs

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

6. **Summarize the following i. Massively Parallel Processing Databases ii. Hadoop iii. No SQL**

Big data analytics can consist of many different software pieces that together collect, store, manipulate, and analyze all different data types. It helps to better understand the landscape by defining what big data is and what it is not. Generally, the industry looks to the "three Vs" to categorize big data:

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.

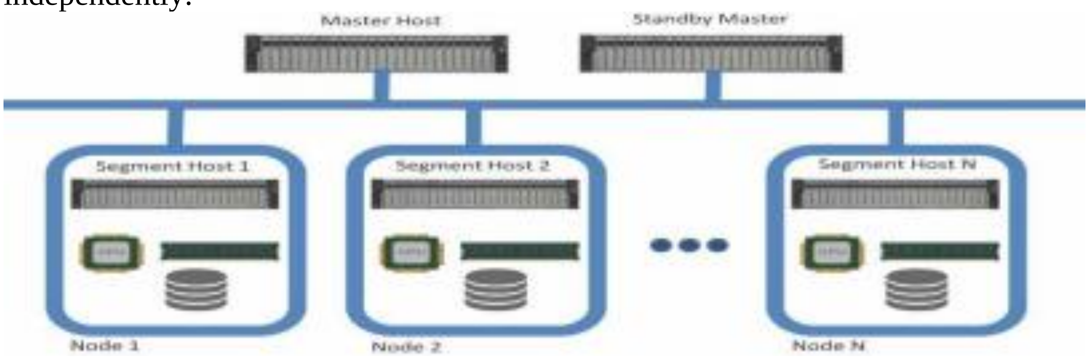
Volume: Volume refers to the scale of the data. Typically, this is measured from gigabytes on the very low end to petabytes or even Exabyte's of data on the other extreme. Generally, big data implementations scale beyond what is available on locally attached storage disks on a single node. It is common to see clusters of servers that consist of dozens, hundreds, or even thousands of nodes for some large deployments. The characteristics of big data can be defined by the sources and types of data. Machine data, which is generated by IoT devices and is typically unstructured data. Transactional data, which is from sources that produce data from transactions on these systems, and, have high volume and structured.

social data sources, which are typically high volume and structured.
Fourth is enterprise data, which is data that is lower in volume and very structured
● The three most popular technologies and techniques in the data management massively parallel processing systems, NoSQL, and Hadoop.

Massively Parallel Processing Databases

Massively parallel processing (MPP) databases were built on the concept of the relational data warehouses but are designed to be much faster, to be efficient, and to support reduced query times.

To accomplish this, MPP databases take advantage of multiple nodes (computers) designed in a scale- out architecture such that both data and processing are distributed across multiple systems. An MPP architecture typically contains a single master node that is responsible for the coordination of all the data storage and processing across the cluster. It operates in a "shared-nothing" fashion, with each node containing local processing, memory, and storage and operating independently.



NOSQL

Data storage is optimized across the nodes in a structured SQL-like format that allows data analysts to work with the data using common SQL tools and applications. Because data stored on MPPs must still conform to this relational structure, it may not be the only database type used in an IoT Implementation. NoSQL Databases NoSQL ("not only SQL") is a class of databases that support Semi-structured and unstructured data, in addition to the structured data handled by data warehouses and MPPs. NoSQL is not a specific database technology; rather, it is an umbrella term that encompasses several different types of databases, including the following:

Document stores: This type of database stores semi-structured data,such as XML or JSON. Document stores generally have query engines and indexing features that allow for many optimized queries.

Key-value stores: This type of database stores associative arrays where a key is paired with an associated value. These databases are easy to build and easy to scale.Wide-column stores: This type of database stores similar to a key-value store, but the formatting of the values can vary from row to row,even in the same table.

Graph stores: This type of database is organized based on the relationships between elements. Graph stores are commonly used for social media or natural language processing, where the connections between data are very relevant.

Expanding NoSQL databases to other nodes is similar to expansion in other distributed data systems, where additional hosts are managed by a master node or process.

 This expansion can be automated by some NoSQL implementations or can be provisioned manually.

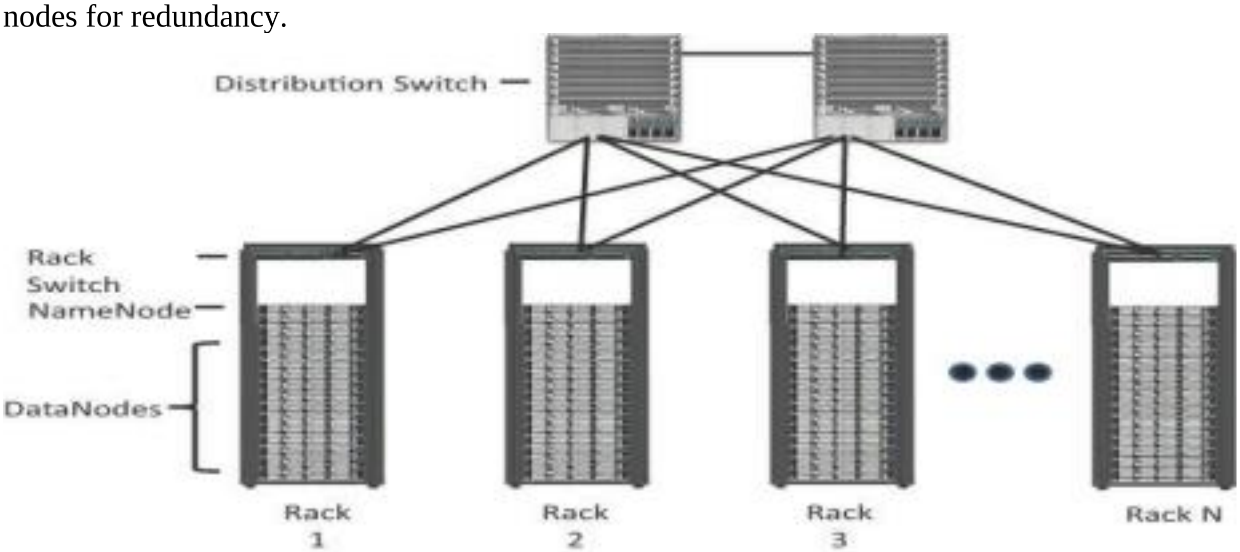
 This level of flexibility makes NoSQL a good candidate for holding machine and sensor data associated with smart objects.

Hadoop

Hadoop is the most recent entrant into the data management market, but it is arguably the most popular choice as a data repository and processing engine. Hadoop was originally developed as a result of projects at Google and Yahoo!, and the original intent for Hadoop was to index millions of websites and quickly return search results for open source search engines. The project had two key elements:

Hadoop Distributed File System (HDFS): A system for storing data across multiple nodes
MapReduce: A distributed processing engine that splits a large task into smaller ones that can be run in parallel Both MapReduce and HDFS take advantage of this distributed architecture to store and process massive amounts of data and are thus able to leverage resources from all nodes in the cluster. For HDFS, this capability is handled by specialized nodes in the cluster, including NameNodes and DataNodes

 NameNodes: These are a critical piece in data adds, moves, deletes,and reads on HDFS. They coordinate where the data is stored, and maintain a map of where each block of data is stored and where it is replicated.
 DataNodes: These are the servers where the data is stored at the direction of the NameNode. It is common to have many DataNodes in a Hadoop cluster to store the data. Data blocks are distributed across several nodes and often are replicated three, four, or more times across nodes for redundancy.



MapReduce leverages a similar model to batch process the data stored on the cluster nodes.
YARN
 Introduced with version 2.0 of Hadoop, YARN (Yet Another Resource Negotiator) was designed to enhance the functionality of MapReduce.
 YARN was developed to take over the resource negotiation and job/task tracking, allowing MapReduce to be responsible only for data processing. o With the development of a dedicated cluster resource scheduler, Hadoop was able to add additional data processing modules to its core feature set, including interactive SQL and real-time processing, in addition to batch processing using MapReduce.