**Big Data-I**

1. **How mahout plays an important role in data mining (Hadoop)?**

**Ans:** Mahout is a machine learning library that plays a crucial role in data mining within the Hadoop ecosystem. It provides scalable implementations of various machine learning algorithms, making it easier to analyze and extract valuable insights from large datasets.

1. **Differentiate b/w Hive and Pig.**

**Ans:** Hive and Pig are both high-level data processing languages in the Hadoop ecosystem, but they serve different purposes:

Hive: Hive is a data warehousing and SQL-like query language that allows users to write SQL-like queries (HQL) for querying and analyzing data stored in Hadoop's HDFS. It's best suited for users familiar with SQL.

Pig: Pig is a scripting language designed for data ETL (Extract, Transform, Load) tasks. It offers a more flexible approach to data processing, making it suitable for complex data transformations.

1. **Whate are the Characteristics of Big Data.**

**Ans:** Volume: Refers to the massive amount of data generated, collected, and stored, often beyond the capacity of traditional databases.

Velocity: Describes the high speed at which data is generated and the need to process it in real-time or near-real-time.

Variety: Indicates the diverse types of data, including structured, semi-structured, and unstructured data, such as text, images, videos, and more.

1. **Define YARN.**

**Ans:** YARN is a resource management and job scheduling component in the Hadoop ecosystem. It serves as the resource manager for Hadoop clusters, enabling efficient and centralized resource allocation and management. YARN allows multiple applications to share and utilize cluster resources more effectively, improving overall cluster utilization and performance.

1. **Differentiate b/w HDFS and RDBS.**

**Ans**: HDFS (Hadoop Distributed File System):

* HDFS is a distributed file system designed for storing and processing large volumes of data across clusters of commodity hardware.
* It is optimized for handling large files and is fault-tolerant, meaning it can handle hardware failures without data loss.
* HDFS is primarily used in the context of big data processing frameworks like Hadoop for distributed storage.

**RDBMS (Relational Database Management System):**

* RDBMS is a database management system that stores and manages data in structured tables with rows and columns.
* It is designed for transactional data and provides features like data integrity, ACID (Atomicity, Consistency, Isolation, Durability) properties, and SQL support.
* RDBMS is suitable for structured data and is commonly used in traditional enterprise applications.

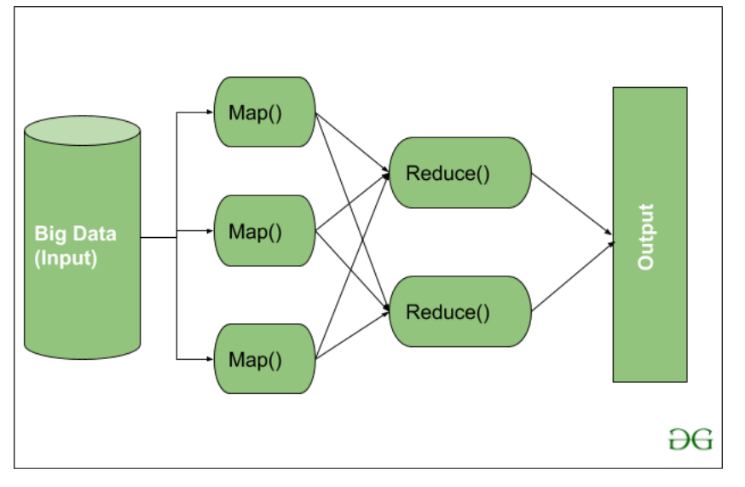
1. **Hive Meta Store.**

**Ans:**

* The Hive Meta Store, often referred to as the Hive Meta store, is a centralized metadata repository used by Apache Hive, a data warehousing and SQL-like query language for Hadoop.
* It stores metadata about tables, partitions, columns, and other objects created and managed by Hive.
* The Hive Meta store helps in maintaining a schema for data stored in Hadoop and allows users to query and analyze data using SQL-like syntax through Hive.

1. **Explain the ecosystem of Hadoop.**

**Ans:** The Hadoop ecosystem is a collection of open-source software tools and frameworks designed to store, process, and analyze large volumes of data. Some key components of the Hadoop ecosystem include:

* HDFS (Hadoop Distributed File System): A distributed file system that stores data across multiple nodes in a Hadoop cluster.
* MapReduce: A programming model and processing framework for distributed data processing.
* YARN (Yet Another Resource Negotiator): A resource management and job scheduling framework.
* Hive: A data warehousing and SQL-like query language.
* Pig: A scripting language for data ETL (Extract, Transform, Load) tasks.
* Spark: A fast and general-purpose cluster computing framework.
* Sqoop: A tool for transferring data between Hadoop and relational databases.
* Zookeeper: A distributed coordination service.
* Mahout: A machine learning library for scalable data mining.

These components work together to enable the processing of large-scale data, making Hadoop a powerful platform for big data analytics.

1. **Discuss the use cases of Big Data Analytics.**

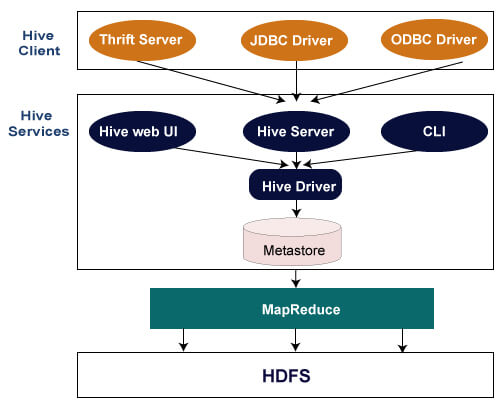
**Ans** Big Data analytics has a wide range of applications across various industries. Some common use cases include:

* Customer Analytics: Analyzing customer data to improve marketing, customer service, and retention.
* Fraud Detection: Identifying fraudulent activities in financial transactions.
* Predictive Maintenance: Predicting when equipment or machinery needs maintenance to avoid downtime.
* Healthcare Analytics: Analyzing patient data for better diagnosis and treatment.
* Recommendation Systems: Providing personalized recommendations in e-commerce and content streaming platforms.
* Supply Chain Optimization: Optimizing logistics and supply chain operations.
* Social Media Analysis: Analyzing social media data for sentiment analysis and market research.
* Cybersecurity: Detecting and preventing cyber threats.
* Environmental Monitoring: Analyzing sensor data for environmental research and conservation efforts.

1. **Discuss Hive Architecture.**

**Ans:** Hive is a data warehousing and SQL-like query language for Hadoop. Its architecture consists of the following components:

* Metastore: Stores metadata about tables, partitions, and columns.
* Driver: Manages query compilation, optimization, and execution.
* Compiler: Translates HQL (Hive Query Language) into a series of MapReduce jobs.
* Execution Engine: Executes the MapReduce jobs generated by the compiler.
* Hive CLI (Command Line Interface): Provides an interactive interface for users to submit Hive queries.
* Thrift Server: Allows remote clients to submit Hive queries using various programming languages.



1. **Compare the following HDFS and GPFS.**

HDFS (Hadoop Distributed File System):

* Open-source distributed file system.
* Designed for big data storage and processing in Hadoop clusters.
* Provides high fault tolerance through data replication.
* Optimized for sequential data access patterns.
* Scales horizontally by adding more commodity hardware.

GPFS (General Parallel File System):

* Proprietary parallel file system developed by IBM.
* Used in high-performance computing (HPC) and enterprise environments.
* Supports both parallel and distributed data access.
* Offers features like advanced data management, data compression, and snapshots.
* Scales vertically by adding more resources to a single server.

**DWDM-I**

1. **Write any to advantages of data mart or Dmart.**

**Ans:** Data Marts play a vital role in data mining by providing a focused and efficient way to access and analyze data that is tailored to the specific needs of business analysts, data scientists, and other users within an organization.

Advantages of Data Mart (Dmart):

* Focused Data: Data marts contain specific, tailored data subsets, making it easier for users to access relevant information for their particular needs.
* Improved Performance: They typically offer faster query performance compared to larger data warehouses due to their smaller size and specialized focus.

1. **Name and briefly explain three type of data preprocessing methods or Data Preprocessing.**

**Ans: Data Cleaning:**

Data cleaning, also known as data cleansing or data scrubbing, is the process of identifying and rectifying errors or inconsistencies in a dataset. It involves:

* Handling missing values: Removing or imputing missing data points to avoid gaps in the dataset.
* Noise reduction: Smoothing noisy data by applying techniques like filtering or aggregation.
* Data deduplication: Removing duplicate records to ensure data integrity.

**Data Transformation:**

Data transformation involves converting the format or structure of data into a more suitable form for analysis. Key aspects of data transformation include:

* Normalization: Scaling numerical attributes to a standard range (e.g., between 0 and 1) to eliminate variations in scale.
* Encoding categorical data: Converting categorical variables into numerical representations for analysis, often through techniques like one-hot encoding.

**Data Reduction:**

Data reduction methods aim to reduce the volume of data while preserving its quality and meaningful information. These methods include:

* Sampling: Selecting a representative subset of data points from a larger dataset, which can significantly reduce computational requirements for analysis.
* Histogram analysis: Analyzing data distributions to identify patterns and reduce data size by capturing the most relevant aspects of the data.

These data preprocessing methods are crucial for preparing data for analysis in data mining, machine learning, and other data-driven tasks, as they ensure that the data is clean, well-structured, and suitable for accurate modeling and insights extraction.

1. **List any four challenges in Data Mining. Or Data Mining & Challenges.**

Ans: **Data Quality:**

Data in real-world applications often contain errors, missing values, inconsistencies, and noise. Low-quality data can lead to inaccurate or unreliable results in data mining. Ensuring data quality is a fundamental challenge in the field.

**Scalability:**

Data mining algorithms must handle increasingly large datasets, often with millions or billions of records.

**Privacy and Security:**

Data mining involves the analysis of sensitive and personal information. Ensuring the privacy and security of data while extracting valuable insights is a significant challenge.

**Complexity of Algorithms:**

Many data mining algorithms, such as neural networks, decision trees, and association rule mining, can be complex and computationally intensive. Implementing and fine-tuning these algorithms can be challenging, especially for non-experts in the field.

1. **ROLAP, MOLAP.**

ROLAP (Relational Online Analytical Processing) is a data storage and querying approach often used in data mining and business intelligence applications. In a nutshell:

**Data Storage:** ROLAP stores data for data mining in relational databases. It utilizes the tabular structure of relational databases to organize and manage data. This allows data miners to work with structured and well-organized data that can be easily queried using SQL and other relational database tools.

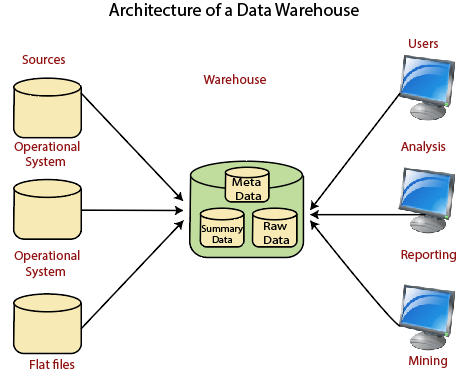
**Querying:** Data mining tasks often require complex querying and analysis of large datasets.

**MOLAP (Multidimensional Online Analytical Processing)**: is a data storage and querying approach frequently employed in data mining and business intelligence applications. Here are its key aspects in data mining:

**Data Storage (1 Mark):** MOLAP stores data for data mining in specialized multidimensional cubes or structures. These cubes are designed to efficiently represent and store multidimensional data, making them highly suitable for complex data mining tasks. The data in MOLAP systems is pre-aggregated and optimized for rapid retrieval and analysis.

**Querying (1 Mark):** Data mining often involves querying and analyzing large volumes of data from various dimensions and measures.

1. **Apriori Algo.**

**Ans:** Apriori algorithm refers to the algorithm which is used to calculate the association rules between objects. It means how two or more objects are related to one another. In other words, we can say that the apriori algorithm is an association rule leaning that analyzes that people who bought product A also bought product B.

1. **Architecture of data warehousing.**

Ans: A data warehouse architecture is a method of defining the overall architecture of data communication processing and presentation that exist for end-clients computing within the enterprise. Each data warehouse is different, but all are characterized by standard vital components.

1. **Data Cube / OLAP operation**

**|Ans: Data Cube (1 Mark):** A data cube is a multidimensional structure that stores data in a way that allows for efficient and flexible analysis. It typically includes dimensions (attributes), measures (data values), and hierarchies. Data cubes are essential for performing multidimensional analysis in OLAP systems.

**OLAP Operations (1 Mark):** OLAP operations are actions applied to data cubes for analysis. Common OLAP operations include roll-up (aggregating data to higher levels), drill-down (exploring more detailed data), pivot (changing the viewpoint), slice (selecting a specific subset), dice (selecting a subcube), drill-through (accessing detailed data), ranking (assigning ranks to data elements), and filtering (applying conditions). These operations provide users with the flexibility to explore and derive insights from multidimensional data efficiently.

1. **Differentiate among ROLAP, MOLAP, and HOLAP.**

Ans: **ROLAP (Relational Online Analytical Processing):** ROLAP is an OLAP approach that stores data in relational databases, using tables and SQL queries to represent and analyze multidimensional data. It provides flexibility for complex data models and diverse data sources.

**MOLAP (Multidimensional Online Analytical Processing):** MOLAP is an OLAP approach that stores data in specialized multidimensional cubes, pre-aggregated and optimized for efficient multidimensional data analysis. It excels in query performance and interactive analysis.

**HOLAP (Hybrid Online Analytical Processing):** HOLAP combines elements of ROLAP and MOLAP by allowing users to store summary data in multidimensional cubes for performance while retaining detailed data in relational databases for flexibility. It offers a balance between flexibility and query speed.

1. **Explain any four types of data cube operations with suitable example.**

**Ans:** Cube operations refer to various actions and manipulations performed on data cubes in the context of Online Analytical Processing (OLAP) to analyze and extract meaningful insights from multidimensional data. Here are some common cube operations:

**Roll-Up (Drill-Up):** Roll-up means looking at data from a higher, summary level. It's like going from detailed monthly sales to quarterly or yearly sales.

**Drill-Down (Roll-Down):** Drill-down is the opposite; it's diving into more detailed data. It's like going from yearly sales to monthly or daily sales.

**Slice:** Slicing is like cutting out a specific part of data. It's selecting data for a particular category or time period, such as "sales for a specific product in a given month."

**Dice:** Dicing is like taking a smaller piece of a pizza. It's selecting data for a combination of categories or conditions, such as "sales for a specific product in a given month and region."

1. **Naive Algorithm & example**
2. **Apriori Algorithm.**

**NSC-I**

1. **Define the cryptography with the help of an example. 4-2**

**Ans:** Cryptography is the practice and study of techniques used to secure communication and information from unauthorized access or modification. It involves the use of mathematical algorithms and principles to transform plaintext (unencrypted data) into ciphertext (encrypted data) in order to protect its confidentiality, integrity, and authenticity.

Exp: Original Message (Plaintext): "HELLO"

Shift Value: 3

Encrypted Message (Ciphertext): "KHOOR"

Encrypted Message (Ciphertext): "KHOOR"

Shift Value: 3 (opposite of the encryption shift)

Decrypted Message (Plaintext): "HELLO".

1. **Define DEC in cryptography. -2**

**Ans: DEC** stands for "Data Encryption Standard." It is a symmetric-key block cipher that was developed by the United States National Bureau of Standards (now known as the American National Standards Institute) in the 1970s. The DEC algorithm is widely used for encrypting data in various applications, including secure communication protocols like SSL/TLS and PGP.

1. **What are the block chipper design principles. -2**

Block Cipher is an encryption algorithm that works with a symmetric key in a deterministic way. The plain text is divided into several blocks of equal size. If the length of the plain text does not allow block division of equal size, padding is done over the plain text. His type of encryption method can encrypt on blocks of 128 bits, the key can be 128, 192, or 256 bits.

**Principals:**

**The number of encryption rounds** − the number of encryption rounds that the plain text will go through explains the decoding difficulty and hence establishes security.

**Key scheduling algorithm** − the generation of the keys for each of the rounds is defined by this algorithm.

**Function Design**.

1. **How do you define vulnerability? -2**

**Ans:** A vulnerability is a weakness that can be exploited by cybercriminals to gain unauthorized access to a computer system. After exploiting a vulnerability, a[cyberattack](https://www.upguard.com/blog/cyber-attack) can run malicious code, install[malware](https://www.upguard.com/blog/malware), and even steal[sensitive data](https://www.upguard.com/blog/sensitive-data).

1. **Difference between AEC and DES. 4-2**

|  |  |  |
| --- | --- | --- |
| 1. | AES stands for [Advanced Encryption Standard](https://www.geeksforgeeks.org/advanced-encryption-standard-aes/) | DES stands for [Data Encryption Standard](https://www.geeksforgeeks.org/data-encryption-standard-des-set-1/) |
| 2. | The date of creation is 2001. | The date of creation is 1977. |
| 3. | Byte-Oriented. | Bit-Oriented. |
| 4. | Key length can be 128-bits, 192-bits, and 256-bits. | The key length is 56 bits in DES. |
| 6. | It is faster than DES. | It is slower than AES. |
| 8. | It is flexible. | It is not flexible. |

Ans:

1. **Differentiate b/w security service and security mechanism?4-2**

**Ans:** Security services are high-level functions that provide protection to computer systems and networks. These services are typically provided by a security system or a security application, and they include:

1. Authentication: The process of verifying the identity of a user, device, or system.
2. Authorization: The process of granting or denying access to resources based on a user's identity and privileges.
3. Confidentiality: The protection of sensitive information from unauthorized access, use, or disclosure.

Security mechanisms are the specific techniques and technologies used to provide security services. These mechanisms can be hardware-based, software-based, or a combination of both. Some common security mechanisms include:  
  
1. Firewalls: Network devices that control incoming and outgoing network traffic based on a set of security rules.  
2. Encryption: The process of converting plaintext data into unreadable ciphertext to protect it from unauthorized access.

1. **Write the 3 components in the CIA models?4-2**

**Ans:** The CIA model, also known as the CIA triad, is a widely accepted framework used in information security to ensure the confidentiality, integrity, and availability of data. These three components are essential for maintaining the security and protection of sensitive information.  
  
1. **Confidentiality**: Confidentiality refers to the assurance that information is only accessible to authorized individuals or entities.   
2. **Integrity**: Integrity focuses on the accuracy, consistency, and trustworthiness of data throughout its lifecycle.   
**3. Availability:** Availability ensures that information and resources are accessible and usable when needed by authorized users.

1. **Differentiate b/w Euclidean and extended Euclidean algorithm? -4**

**Ans: The Euclidean algorithm** is a method for finding the greatest common divisor (GCD) of two integers, and it is based on the principle of iterative division. The algorithm starts with two integers, a and b, and repeatedly divides the larger number (b) by the smaller number (a) until the quotient is 0. The remaining number (b) is the GCD of a and b.

**The extended Euclidean algorithm** is an extension of the Euclidean algorithm that allows us to find the solution of a system of linear equations ax + by = c, where a, b, and c are integers. The algorithm involves a series of transformations that convert the system of equations into a set of simpler equations, which can be solved using the Euclidean algorithm.

**1 GCD Computation**:

The Euclidean algorithm is limited to finding the GCD of two integers.

while the extended Euclidean algorithm can be used to find the GCD of two integers as well as to solve systems of linear equations.

2. **Number of Operations**:

The Euclidean algorithm typically requires fewer operations than the extended Euclidean algorithm, especially for larger inputs.

However, the extended Euclidean algorithm can be more efficient for certain types of inputs, such as when the coefficients of the system of linear equations are relatively prime.

3. **Applications**:

The Euclidean algorithm has a wider range of applications, including cryptography, coding theory, and computer science,

while the extended Euclidean algorithm is primarily used in number theory and algebraic geometry.

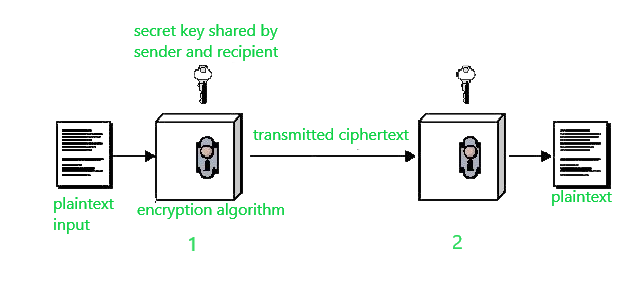
1. **Discuss the Fermat theorem in cryptography?-4**

**Ans:** In cryptography, the Fermat theorem is used in the design of cryptographic protocols, particularly in the area of public-key cryptography. Public-key cryptography relies on the difficulty of certain mathematical problems, such as factoring large numbers and computing discrete logarithms, to ensure the security of the encrypted data. The Fermat theorem is used in the design of these cryptographic protocols because it provides a way to prove the security of the system without having to explicitly compute the discrete logarithm of a number.

a^(p-1) ≡ 1 (mod p)

1. **Explain conventional encryption model? Discuss each component with diagram.8**

**Ans:** The conventional encryption model, also known as symmetric encryption or secret key encryption, is a widely used method of encrypting and decrypting data. In this model, the same key is used for both encryption and decryption processes. The key is kept secret and shared only between the sender and the receiver.

****

Conventional encryption has mainly 5 ingredients:

Plain text –   
It is the original data that is given to the algorithm as an input.

Encryption algorithm –   
This encryption algorithm performs various transformations on plain text to convert it into ciphertext. 

Secret key –   
The secret key is also an input to the algorithm. The encryption algorithm will produce different outputs based on the keys used at that time. 

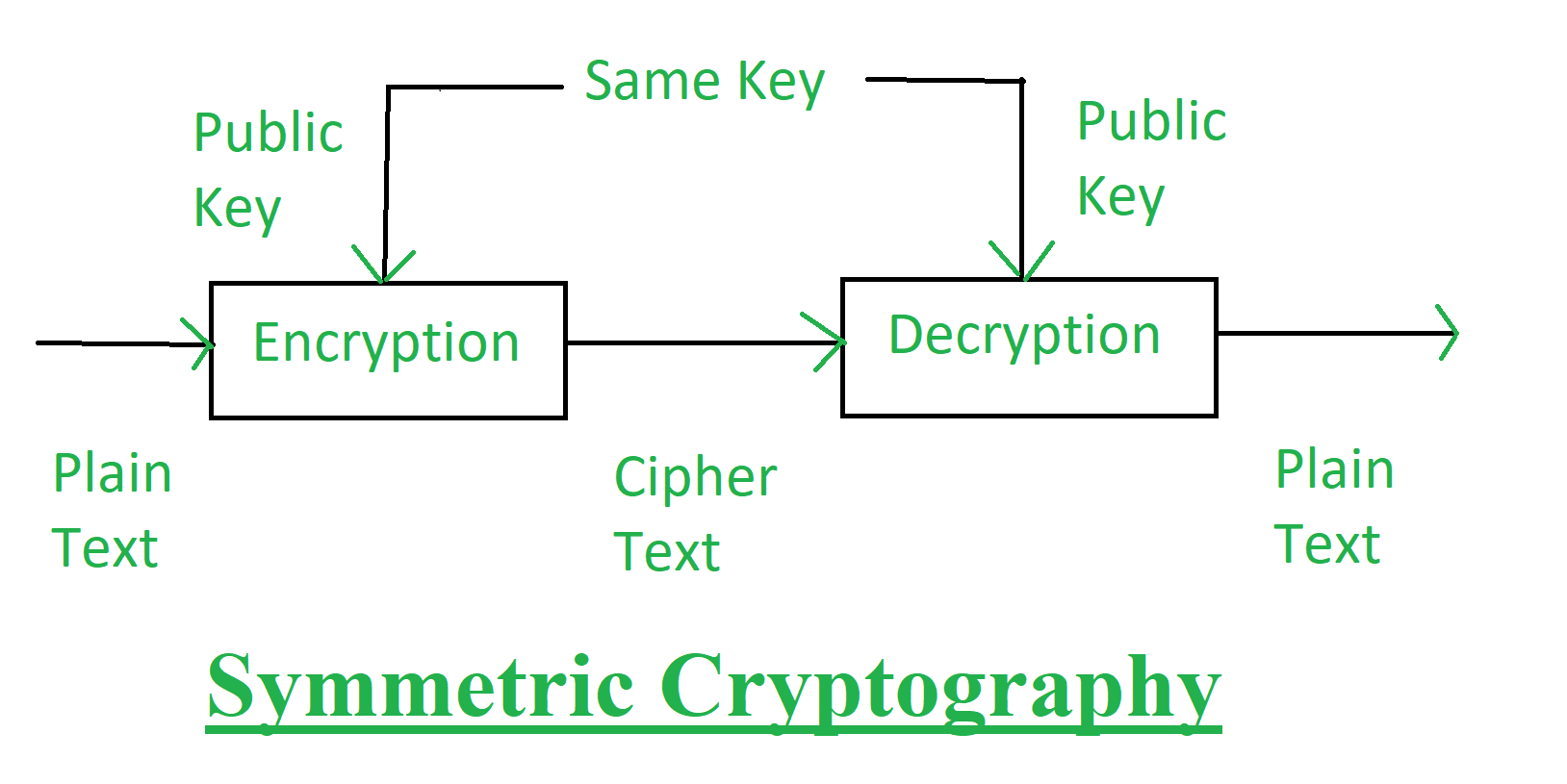
Ciphertext –   
It contains encrypted information because it contains a form of original plaintext that is unreadable by a human or computer without proper cipher to decrypt it. It is output from the algorithm. 

Decryption algorithm –   
This is used to run encryption algorithms in reverse. Ciphertext and Secret key is input here and it produces plain text as output.

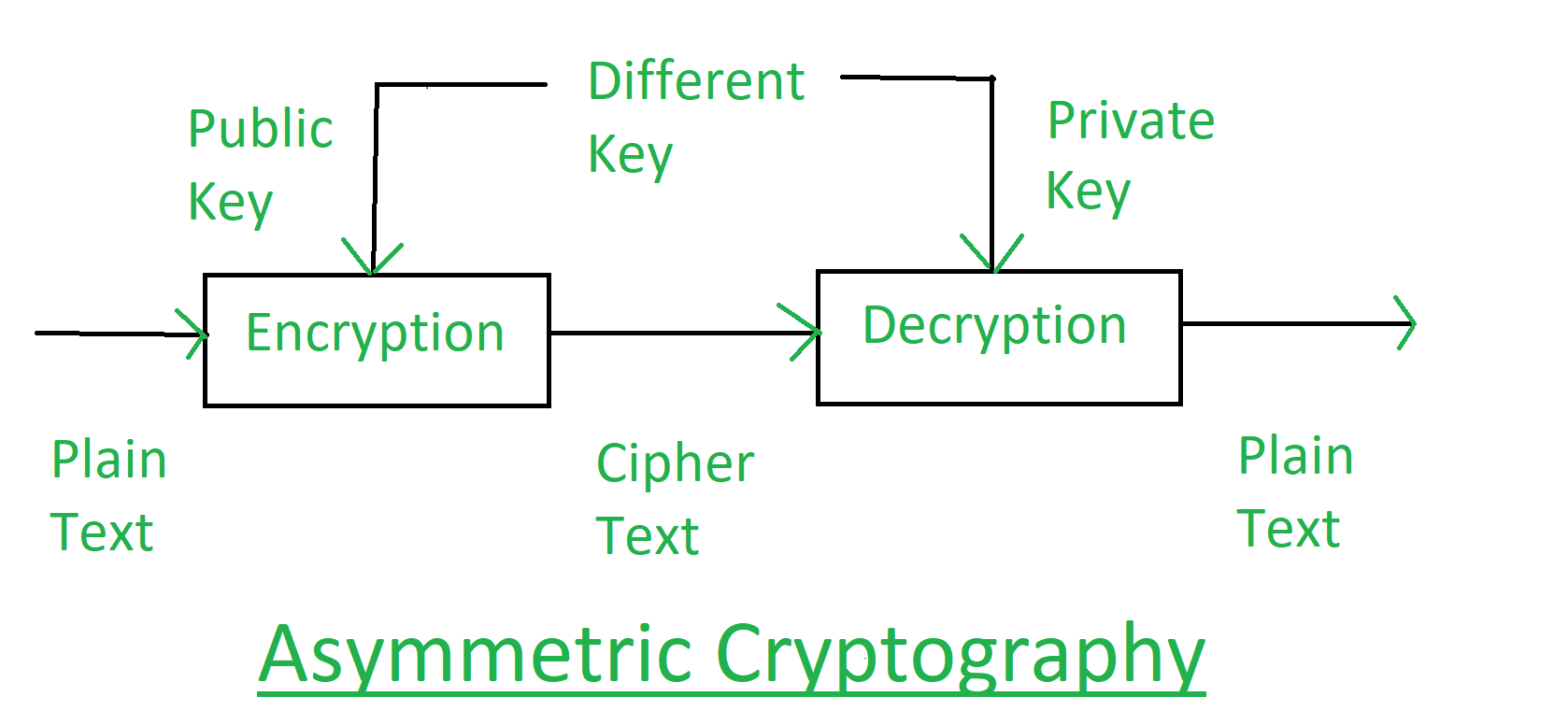
1. **Explain the different classical cryptography technique in details.8**

**Ans: Classical cryptography** refers to the traditional methods of encrypting and decrypting information that were used before the advent of computers and modern cryptographic techniques. It primarily relies on symmetric key algorithms, where the same key is used for both encryption and decryption.

**Symmetric cryptography,** also known as secret-key cryptography, is a cryptographic method where the same key is used for both encryption and decryption of data. In this approach, the sender and receiver must share the same secret key in advance.



**Asymmetric cryptography**, also known as public-key cryptography, is a cryptographic method that uses two different but mathematically related keys: a public key and a private key. These keys are generated as a pair, where what one key encrypts, only the other corresponding key can decrypt.  
In asymmetric cryptography, each user has a unique pair of keys. The public key is freely distributed to anyone who wants to communicate with that user, while the private key remains confidential and known only to its owner. The public key can be used to encrypt messages, while the private key is used for decryption.



**DL-I-R**

1. **Difference between overfitting & underfitting. - 2**

**Ans:**

|  |  |  |
| --- | --- | --- |
|  | Over Fitting | Under Fitting |
| Complexity of the Model | More complex model | More simple model. |
| Regularization | Less Regularization | More regularization |
| Quantity of features | A large quantity of features | A smaller quantity of features. |
| Data | Data cleaning, cross validation. | Data cleaning, cross validation. |

1. **Difference between forward & Back Propagation.**

**Ans:**

|  |  |  |
| --- | --- | --- |
| Aspect | Forward Propagation | Back Propagation |
| Process Definition | Computing the output of a neural network given an input. | Updating the network's parameters based on prediction error. |
| Usage Phase | Training and inference. | Primarily during training. |
| Computational Complexity | Relatively straightforward, single pass. | More complex, multiple passes and calculations. |
| Parameter Adjustment | No parameter adjustments. | Parameters adjusted to learn from errors. |
| Information Flow Type | Feed-forward, no feedback. | Introduces feedback through error propagation. |

1. **Gradient Descent.**

**Ans:** Gradient descent is an optimization algorithm used to minimize the loss function in machine learning. It is a first-order optimization algorithm that iteratively adjusts the parameters of a model to find the values that minimize the loss function.

**Steps:** 1- Initialize, 2- Compute for loss function. 3-Compute for Gradient, 4-Update.

1. **Define Bias & variance.**

**Ans: Bias** is the difference between the actual value and the predicted value.

**Variance:** The amount by which the model prediction would change if we estimate it using a different dataset.

1. **Computational graph.**

**Ans:** A computational graph, also known as a computation graph or a directed acyclic graph (DAG), is a graphical representation of mathematical operations and their dependencies in deep learning models. It is a fundamental concept used in deep learning frameworks to represent and optimize the computations involved in training and inference processes.  
  
1**. Structure:** A computational graph consists of nodes and edges. Nodes represent mathematical operations or variables, while edges represent the flow of data between these operations. Each node performs a specific operation on its inputs and produces an output.

**2. Forward and backward propagation:** Computational graphs are particularly useful for implementing forward and backward propagation algorithms, which are essential for training deep learning models.

3**. Efficiency and parallelism:** Computational graphs enable efficient execution of deep learning models by allowing parallelization of computations.

4. **Automatic differentiation:** Another key advantage of computational graphs is their ability to automatically compute gradients using automatic differentiation techniques. By representing mathematical operations as nodes in the graph.

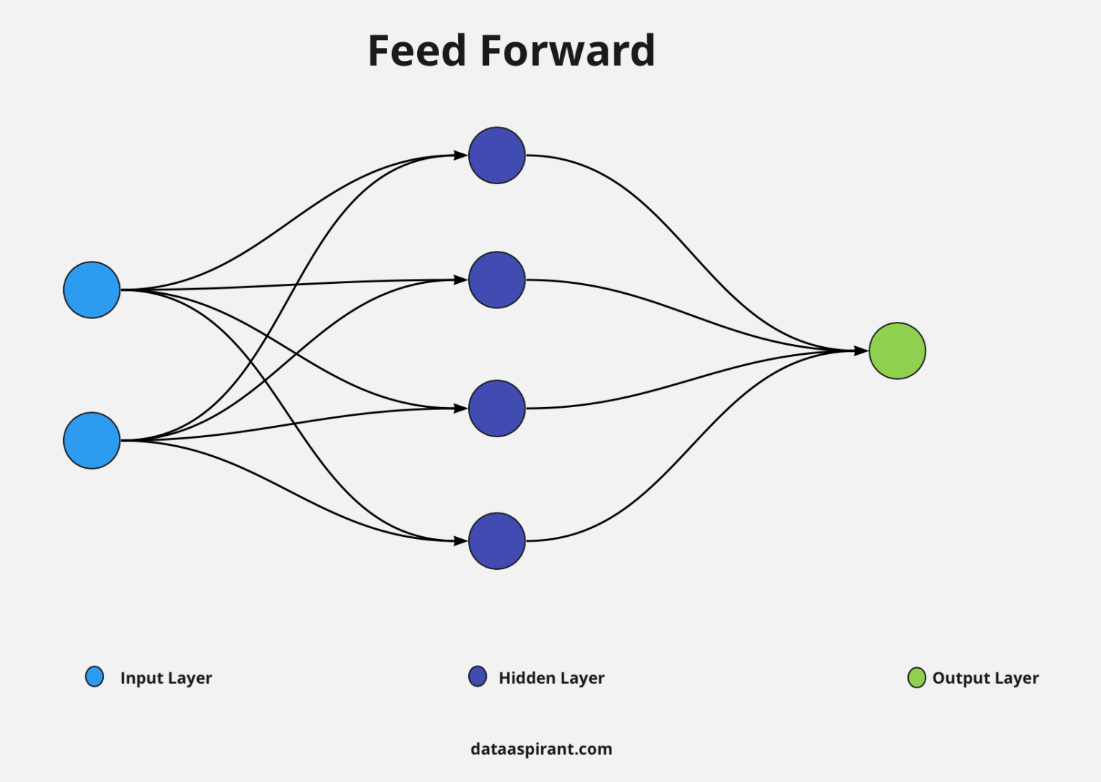
1. **Difference between supervised & unsupervised learning.**

**Ans:**

|  |  |  |
| --- | --- | --- |
|  | Supervised Learning | Unsupervised Learning. |
| Data Labeling | 1. Required labeled data | Work with unlabeled data. |
| Object | 1. Predicting outputs based on inputs. | Discovering patterns and structures. |
| Training Process | 1. Minimizing prediction errors with labels. | Find Patterns or grouping data. |
| Aspect | 1. Supervised Learning | Unsupervised. |
| Data Preparation | 1. Requires Labeling | Working with raw, unlabeled data. |
| Guided Learning | 1. Guided by explicit labels | Operates without explicit guidance |

1. **Feed forward neural networks.**

**Ans:** Feed forward neural networks are a type of artificial neural network where the information flows in one direction, from input layer to output layer, without any feedback loops. This type of network is widely used in various applications such as image recognition, speech recognition, and natural language processing.



**Input Layer:** It is starting layer of the network that has a weight associated with the signals.

**Hidden Layer:** This layer lies after the input layer and contains multiple neurons that perform all computations and pass the result to the output unit.

**Output Layer:** It is a layer that contains output units or neurons and receives processed data from the hidden layer, if there are further hidden layers connected to it then it passes the weighted unit to the connected hidden layer for further processing to get the desired result.

|  |  |
| --- | --- |
| **Advantages** | **Dis-advantages** |
| Simple to Implement | Limited Memory |
| Fast Training | Difficult to Interpret |
| Flexible | Overfitting |

**Application:**

* Natural Language Processing
* Speech Recognition
* Image Recognition

**Limitations:** It has some limitations like sometimes information about the neighborhood is lost and, in that case, it becomes difficult to process further all steps are needed to be performed again and it does not support back propagation so the network cannot learn or correct the fault of the previous stage.

**AWSN-I**

1. **Differentiate b/w Wireless Adhoc and Wireless Sensor Network?**
2. **Define me use of Routing Protocol in Wireless Networks?**
3. **List any three application of wireless sensor networks?**
4. **Differentiate b/w proactive and reactive Routing Protocols.**
5. **Discuss the concepts, Architecture and Application of Wireless Sensor Networks in details?**
6. **List issues and goals of designing Routing Protocols for Adhoc Wireless Networks in details?**
7. **Discuss the Adhoc On-Demand Distance Vector Routing Protocol in details with example?**

**Data Whare House and Data Mining-II**

1. **Define the terms Minimum Support and Confidence in context of Association rule mining.**
2. **List two advantages of Apriorism algorithm for association rule mining.**
3. **Identify and four application areas of association rule mining.**
4. **List two advantages of FP-Growth algorithm for association rule mining.**
5. **Recall the steps for Navie algorithm for association rule mining and give one example.**
6. **Draw the itemset for dynamic item Set Counting algorithm with Minium support 25% and m = 2 for the following set transactions:**
   * 1. **Transaction ID Itema Itemb Itemc**

**T1 1 1 0**

**T2 1 0 0**

**T3 0 1 1**

**T4 0 0 0.**

1. **Find Association Rules using Apriorism Algorithm with minimum support 50% and confidence 75% for the following set of transaction and also explain the algorithm.**

**Transaction ID Items**

**T1 Bread, Cornflakes, Eggs, Jam.**

**T2 Bread, Cornflakes, Jam.**

**T3 Bread, Milk, Tea**

**T4 Bread, Jam, Milk**

**T5 Cornflakes, Jam, Milk**

**NSC-II**

1. **Define DES in cryptography?**
2. **What are the block cipher design principles.?**
3. **What is digital signature and how it created.**
4. **Define Hash functions.**
5. **Explain Diffie Hellman key exchange problem with the help of an example?**
6. **Discuss RSA algorithm, Explain with example.**
7. **Explain the different Classical Cryptographic techniques in details.**