UNIT I: INTRODUCTION TO BIG DATA AND ANALYTICS
Topic 4: Data Warehouse and Hadoop Environment

1. <u>Data Warehouse</u> **:** It is a technique for gathering and managing information from different sources to supply significant commercial enterprise insights. A Data warehouse is commonly used to join and analyze commercial enterprise information from heterogeneous sources. It acts as the heart of the BI system which is constructed for data evaluation and reporting.

2. <u>Hadoop</u> **:** It is an open-source software program framework for storing information and strolling applications on clusters of commodity hardware. It offers large storage for any sort of data, extensive processing strength, and the potential to deal with actually limitless concurrent duties or jobs.

Difference between Data Warehouse and Hadoop:

S.No.	Data Warehouse	Hadoop
1.	In this, we first analyze the data and then further do the processing.	It can process various types of data such as Structured data, unstructured data, or raw data.
2.	It is convenient for storing a small volume of data.	It deals with a large volume of data.
3.	It uses schema-for-write logic to process the data.	It deals with schema-for-read logic to process the data.
4.	It is very less agile as compared to Hadoop.	It is more agile as compared to Data Warehouse.
5.	It is of fixed configuration.	It can be configured or reconfigured, accordingly.
6.	It has high security for storing different data.	Security is a great concern and It is improving and working on it.
7.	It is mainly used by business professionals.	It mainly deals with Data Engineering and Data Science.