**Assignment 1: Explain Data Engineering and Data Warehousing Concepts**

**Data Engineering**

It is basically the building of systems that can organize and manage data and mold it according to the needs of the user.

* A Data engineer creates a system that can store the data for it’s use by data scientists or business analytical use.
* A Data Scientist on the other hand performs some linear algebra and some calculations on the data to gain insights from the existing data.

In data Engineering there is the ETL which stands for Extract, Load And Transform the data.

* Extract-extracting data from different sources available
* Transform- to take raw data and process it into useful information.
* Load-loading the transformed data in the system to make it accessible to the user.

4 properties of big data are:

* Volume-Amount of data
* Velocity-How fast data is coming to us
* Variety-How vast the data is
* Veracity-How reliable the data is.

**Data Warehousing**

Data Warehouse is a database used for data reporting and analysis. The data stored in the warehouse is uploaded from the operational systems (such as marketing, sales etc.)

Features of Data Warehousing:

* Subject Oriented-Data is subjected to modelling and analysis.
* Integrity-Ensuring consistency in naming conventions, attribute measures, etc. among different data sources.
* Non-Volatile-Once the data entered the data warehouse, it can never be removed.
* Time Variant-Data can be stored for a very large amount of time.

DSS-Decision Support Systems: Required for decision making needs based on the diverse information available.

DSS can be implemented through:

* OLTP-Online Transaction Processing
* OLAP -Online Analytics Processing

OLTP: It is a methodology designed to give end users efficient access to extensive datasets. It operates swiftly and intuitively. It's great tasks transaction-oriented applications, mainly focusing on tasks such as data entry and retrieval for transaction processing. An ATM is a great example of OLTP.

Advantages of OLTP:

* Simple and Efficient
* Data integrity
* Faster Query Processing

Disadvantages of OLTP:

* Requires instant update.
* Not suitable for data analytics