**ASSESSMENT OF MARGINAL WORKERS IN**

**TAMIL NADU.**

**TEAM MEMBER :**VELMURUGAN P

**Phase 3: Development Part 1**

**INTRODUCTION:**

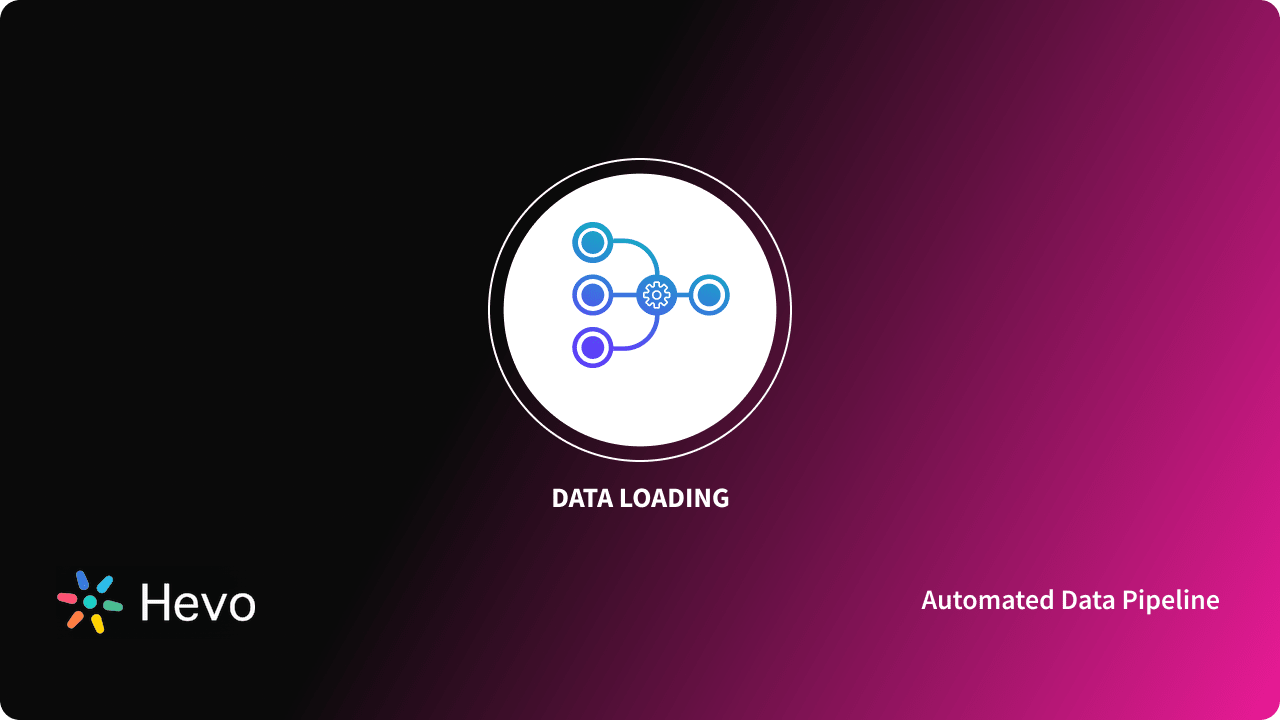
Apart from that, there is a group named marginal workers who work less than 183 days in a year for economical productivity. The workers comprise 312 million main workers and **88 million** marginal workers (i.e., those who did not work for at least 183 days in the preceding 12 months to the census taking).



Marginal refers to the focus on the cost or benefit of the next unit or individual, for example, **the cost to produce one more widget or the profit earned by adding one more worker**. Companies use marginal analysis as a decision-making tool to help them maximize their potential profits

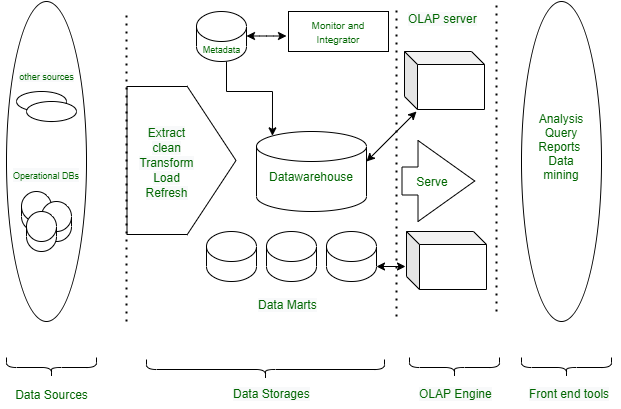
**DATA LOADING IN DATA SET:**

Data Loading is defined as **copying data from one electronic file or database into another**. Data loading implies converting from one format into another; for example, from one type of production database into a decision support database from a different vendor.



**TYPES OF LOADING IN DATA WAREHOUSE:**

* Initial Load: For the very first time loading all the data warehouse tables.
* Incremental Load: Periodically applying ongoing changes as per the requirement. ...
* Full Refresh: Deleting the contents of a table and reloading it with fresh data.
* Data are observations or measurements (unprocessed or processed) represented as text, numbers, or multimedia. A dataset is a structured collection of data generally associated with a unique body of work.
* The entire dataset is dumped, or loaded, and is then completely replaced (i.e. deleted and replaced) with the new, updated dataset.
* Truncating the target feature class table, thus removing all features but leaving the schema intact, followed by loading (or reloading, as the case may be) the data back into the table.



**DATA LOADING EXAMPLE PROGRAM:**

with open(‘sample.csv', 'r') as f:

# Read the column names

column\_names = f.readline().strip().split(',')

# Initialize a list to store the data

# It will hold row values

data = []

# Loop over the remaining lines in the file

for j in f:

{

j = j.strip()

values = j.split(',')

# Append the values to the data list

data.append(values)

}

# Create a dictionary with a list comprehension to hold the data

df = {column\_names[i]: [row[i] for row in data] for i in range(len(column\_names))}

# Create a dataframe from the dictionary

df = pd.DataFrame(df)

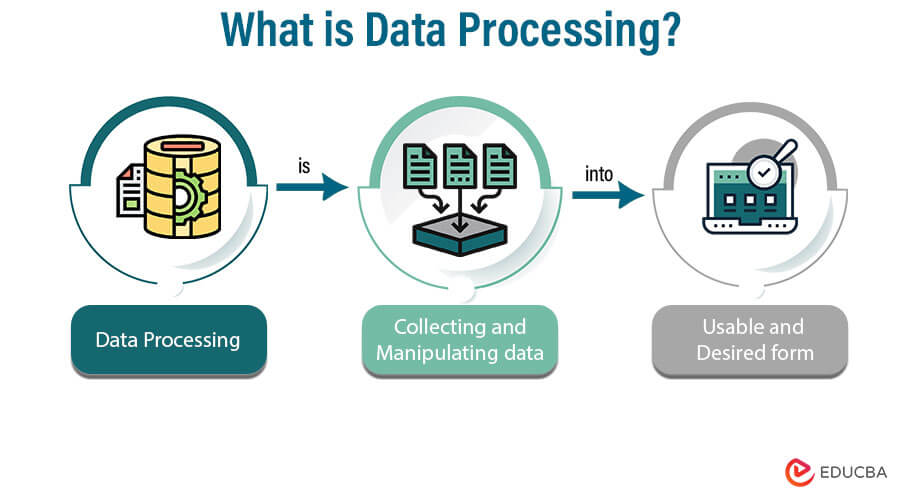
**OUTPUT:**

**DATA PROCESSING:**

* Data processing **occurs when data is collected and translated into usable information**.
* Usually performed by a data scientist or team of data scientists, it is important for data processing to be done correctly as not to negatively affect the end product, or data output.

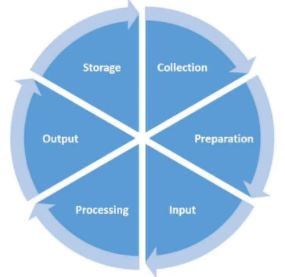
**STAGES OF DATA PROCESSING:**

* Data collection.
* Data input.
* Data processing.
* Data output.

****

**DATA PROCESSING CYCLE:**

The data processing cycle consists of **a series of steps where raw data (input) is fed into a system to produce actionable insights (output)**. Each step is taken in a specific order, but the entire process is repeated in a cyclic manner**.**



**TYPES OF DATA PROCESSING:**

* Manual Data Processing.
* Mechanical Data Processing.
* Electronic Data Processing.
* Batch Data Processing.
* Real-time Data Processing.
* Online Data Processing.
* Automatic Data Processing.

**DATA MANIPULATION:**

There are some python libraries that are useful for data scientists to do Data Manipulation, Machine Learning, Data Visualization, and Statistical Analysis. Libraries like **NumPy and pandas** offer powerful tools for manipulating data in CSV or Excel. Matplotlib offers charts and plots for visualization.

**DATA MANIPULATION LIBRARIES PYTHON IN PANDAS:**

 Installing Pandas.

 Creating DataFrame.

 Adding data in DataFrame using Append Function.

 Getting Shape and information of the data.

 Getting Statistical Analysis of Data.

 Dropping Columns from Data.

 Dropping Rows from Data

**PROGRAM:**

# Importing the pandas library

import pandas as pd

# creating a dataframe object

student\_register = pd.DataFrame()

# assigning values to the

# rows and columns of the dataframe

student\_register['Name'] = ['Abhijit','Smriti','Akash', 'Roshni']

student\_register['Age'] = [20, 19, 20, 14]

student\_register['Student'] = [False, True,True, False]

print(student\_register)

**OUTPUT:**

|  |  |  |  |
| --- | --- | --- | --- |
| 0  1  2  3 | Name  Abhijit  Smriti  Akash  Roshni | Age  20  19  20  14 | STUDENT  False  True  True  False |