**Experiment 8**

**Aim: Implementation of any one hierarchical clustering method**

**Theory:**

A Hierarchical clustering method works via grouping data into a tree of clusters. Hierarchical clustering begins by treating every data point as a separate cluster. Then, it repeatedly executes the subsequent steps:

Identify the 2 clusters which can be closest together, and

Merge the 2 maximum comparable clusters. We need to continue these steps until all the clusters are merged together.

In Hierarchical Clustering, the aim is to produce a hierarchical series of nested clusters. A diagram called Dendrogram (A Dendrogram is a tree-like diagram that statistics the sequences of merges or splits) graphically represents this hierarchy and is an inverted tree that describes the order in which factors are merged (bottom-up view) or clusters are broken up (top-down view).

Hierarchical clustering is a method of cluster analysis in data mining that creates a hierarchical representation of the clusters in a dataset. The method starts by treating each data point as a separate cluster and then iteratively combines the closest clusters until a stopping criterion is reached. The result of hierarchical clustering is a tree-like structure, called a dendrogram, which illustrates the hierarchical relationships among the clusters.

Agglomerative: Initially consider every data point as an individual Cluster and at every step, merge the nearest pairs of the cluster. (It is a bottom-up method). At first, every dataset is considered an individual entity or cluster. At every iteration, the clusters merge with different clusters until one cluster is formed.

The algorithm for Agglomerative Hierarchical Clustering is:

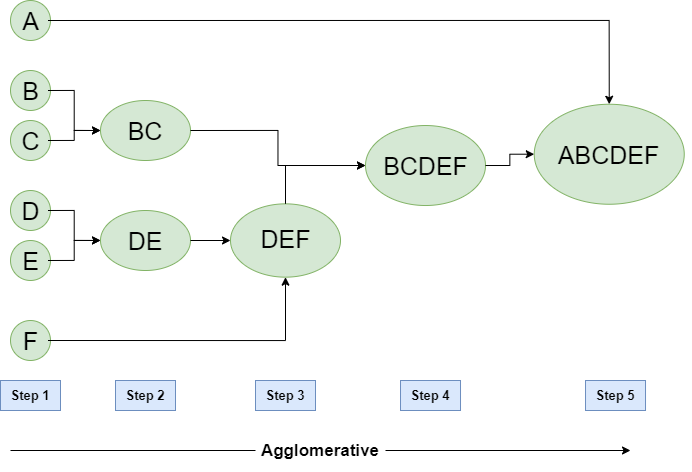
Calculate the similarity of one cluster with all the other clusters (calculate proximity matrix)

Consider every data point as an individual cluster

Merge the clusters which are highly similar or close to each other.

Recalculate the proximity matrix for each cluster

Repeat Steps 3 and 4 until only a single cluster remains.



**Fig 9.1 Agglomerative Hierarchical Clustering**

**Program:**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import warnings

warnings.filterwarnings('ignore')

from sklearn import preprocessing

import scipy.cluster.hierarchy as sch

from sklearn.cluster import AgglomerativeClustering

df = pd.read\_csv('Mall\_Customers.csv')

label\_encoder = preprocessing.LabelEncoder()

df['Gender'] = label\_encoder.fit\_transform(df['Gender'])

linkage\_matrix = sch.linkage(df, method="ward")

agglomerative = AgglomerativeClustering(n\_clusters=None, affinity='euclidean', linkage='ward', distance\_threshold=0)

agglomerative.fit(df)

cluster\_labels = agglomerative.labels\_

plt.figure(1, figsize=(16, 8))

dendrogram = sch.dendrogram(linkage\_matrix)

plt.title('Dendrogram')

plt.xlabel('Customers')

plt.ylabel('Euclidean distances')

plt.show()

distance\_matrix = sch.distance.pdist(df)

distance\_matrix = sch.distance.squareform(distance\_matrix)

distance\_matrix

