**What is Hierarchical clustering?**

**Hierarchical clustering** is characterized by the development of a hierarchy or tree-like structure.

* **Agglomerative clustering** starts with each object in a separate cluster. Clusters are formed by grouping objects into bigger and bigger clusters.
* **Divisive clustering** starts with all the objects grouped in a single cluster. Clusters are divided or split until each object is in a separate cluster.

It is a clustering algorithm, which clusters the data points in group. This algorithm follows agglomerative approach i.e. it starts with each data point as cluster and goes on merging the clusters based on similarity.

**Hierarchical algorithm:**

1. Start by assigning each item to its own cluster, so that if you have N items, you now have N clusters, each containing just one item.
2. Find the closest pair of clusters and merge them into a single cluster, so that now you have one less cluster.
3. Compute distances between the new cluster and each of the old clusters.
4. Repeat steps 2 and 3 until all items are clustered into a single cluster of size N.

**key questions:**

1. **How do you represent the cluster with more than one data Point?**
2. **How do you determine the nearness of the Cluster ?**
3. **When to stop combining the clusters**

**Linkages:**

Linkages are the distances (distance is directly proportional to similarity) between clusters. As stated in above algorithm, we go on merging the 2 nearest clusters. But the problem here is how to find the distance between 2 clusters?

There are multiple ways to find the clusters:

