Discovering Correlated Frequent Patterns in Big Data Using CPGrowthPlus Algorithm

In this tutorial, we will discuss two approaches to find correlated patterns in big data using CPGrowthPlus algorithm.

1. **Basic approach:** Here, we present the steps to discover correlated patterns using a single minimum support value
2. **Advanced approach:** Here, we generalize the basic approach by presenting the steps to discover correlated patterns using multiple minimum support values.

Basic approach: Executing CPGrowthPlus on a single dataset at a particular minimum support value

Step 1: Import the CPGrowthPlus algorithm

In [1]:
```
from PAMI.correlatedPattern.basic import CPGrowthPlus  as alg
```

Step 2: Specify the following input parameters

In [2]:
```
inputFile = 'transactional_T10I4D100K.csv'
minAllConfCount=0.1
minimumSupportCount=100  #Users can also specify this constraint between 0 to 1.

seperator='¥t'
```

Step 3: Execute the CPGrowthPlus algorithm

In [3]:
```
obj = alg.CPGrowthPlus(iFile=inputFile, minSup=minimumSupportCount,  minAllConf=minA
obj.startMine()            #Start the mining process
```

```
Correlated Frequent patterns were generated successfully using CorrelatedPatternGrow
th algorithm
```

Step 4: Storing the generated patterns

Step 4.1: Storing the generated patterns in a file

In [4]:
```
obj.savePatterns(outFile='frequentPatternsMinSupCount100.txt')
```

Step 4.2. Storing the generated patterns in a data frame

In [5]:
```
frequentPatternsDF= obj.getPatternsAsDataFrame()
```

Step 5: Getting the statistics

Step 5.1: Total number of discovered patterns

In [6]:
```
print('Total No of patterns: ' + str(len(frequentPatternsDF)))
```

```
Total No of patterns: 5758
```

Step 5.2: Runtime consumed by the mining algorithm

In [7]:
```
print('Runtime: ' + str(obj.getRuntime()))
```

```
Runtime: 13.133056163787842
```

In [8]:
```
##### Step 5.3: Total Memory consumed by the mining algorithm
```

In [9]:
```
print('Memory (RSS): ' + str(obj.getMemoryRSS()))
print('Memory (USS): ' + str(obj.getMemoryUSS()))
```

```
Memory (RSS): 401604608
Memory (USS): 363081728
```