**Compact Representation of Frequent Itemset**

**Introduction**

What happens when you have a large market basket data with over a hundred items?   
The number of frequent itemsets grows exponentially and this in turn creates an issue with storage and it is for this purpose that alternative representations have been derived which reduce the initial set but can be used to generate all other frequent itemsets.  The Maximal and Closed Frequent Itemsets are two such representations that are subsets of the larger frequent itemset that will be discussed in this section.

**Maximal Frequent Itemset**

***Definition***

It is a frequent itemset for which none of its immediate supersets are frequent.

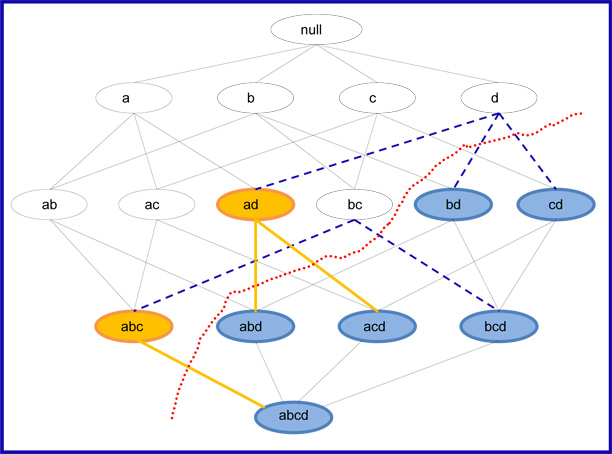
***Identification***

1. Examine the frequent itemsets that appear at the border between the infrequent and frequent itemsets.
2. Identify all of its immediate supersets.
3. If none of the immediate supersets are frequent, the itemset is maximal frequent.

***Illustration***

For instance consider the diagram shown below, the lattice is divided into two groups, red dashed line serves as the demarcation, the itemsets above the line that are blank are frequent itemsets and the blue ones below the red dashed line are infrequent.

* In order to find the maximal frequent itemset, you first identify the frequent itemsets at the border namely *d, bc, ad* and *abc*.
* Then identify their immediate supersets,  the supersets for *d, bc*  are characterized by the blue dashed line and if you trace the lattice you notice that for d, there are three supersets and one of them, ad is frequent and this can’t be maximal frequent,  for *bc* there are two supersets namely*abc* and*bcd abc* is frequent and so *bc* is NOT maximal frequent.
* The supersets for ad and abc are characterized by a solid orange line, the superset for *abc* is *abcd* and being that it is infrequent, *abcd* is maximal frequent. For ad, there are two supersets *abd* and *acd*, both of them are infrequent and so ***ad*** is also maximal frequent.



### *Summary*

This representation is valuable because they provide the most compact representation of frequent itemsets, so when space is an issue or when we are presented with a very large data set using these representations proves helpful. But this representation is not without its limitations. One of the major advantages of using the Apriori Algorithm to find frequent itemsets is that the support of all frequent itemsets are available so during the rule generation stage we don’t have to collect this information again, this advantage disappear when we use maximal frequent itemset and so another representation is presented on the next page that resolves the space issue and also keeps this important advantage involving the support.