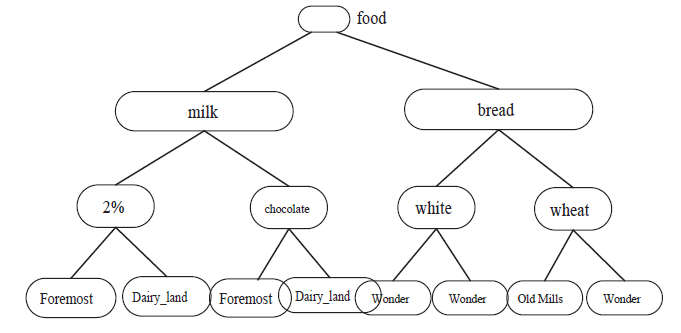
**7.1 Ans:**



* Scan the above database and encode each item by its level. In the above example, consider the item “chocolate milk foremost”, the encoding will be done as follows “122” here the first digit represents the item milk the second digit represents the item chocolate and third digit represents the item foremost. Similarly if we consider other example such as “bread wheat wonder” it can be encoded as “212”. By doing this encoding, with the help of encoding digits we can discover which are items are frequent items and which are not. Here the frequent items are discovered with respect to each level.
* Once the scanning through the database is done, next step is to find out the frequent item sets appeared on the encoded data and join these item sets with the candidate item sets. Then join this item sets with the level 3 item sets and do this till no level items are left and no level is going to generate any a new candidate.
* Once all the item sets were found in all the levels, generate a corresponding strong multilevel association rules.

As per the above observation it is understood that in single level association rules generation will most cost effective. Where as in case of multilevel association rules generation involves more candidate item sets generation and it happens at a degree of levels. So usually it takes more time and effort compare to the single level association rules. Hence the cost generating multilevel association rules is always more when it compare with the candidate generation of single level generation.

**7.3 Ans:**

As per the quantitative rule it is said that the average wage for females is only $7.9 per hour which is less than average mean wage.

*sex* = *female 🡺mean wage* D $7.90/*hr* .(*overall mean wage* D $9.02=*hr*),

When we seriously examine the above statement or constraint the left hand side represents the female population which is nothing but a subset of overall population. On the other side the behaviour of this subset i.e. the mean wage for females. It is basically deeply correlated or associated with the confidence level. The high confidence level can be take it for granted if it is being pass through the various statistical tests for instance z-score.

There are two forms of rules. One through the Categorical implies quantitative rules and other one is Quantitative implies quantitative rules. When this will become hard to identify is that the left hand side has two or more quantitative attributes.

**7.7 Ans:**

**a.** The constraint is succinct and monotonic. Initially mine all the data using a pattern called Blue ray dvd movie. Then using the data mining query language extracts only Blue Ray DVDs which are having at least one dvd movie. This can be mined efficiently with FP-growth.

**b.** This constraint is anti-monotonic. This can be mined efficiently using Apriori. Here the data will be selected for which the prices of the items are less than 150$.

**c.** This constraint is monotonic. The best way to get the best results would be mining with FP-Growth. The approach to do this is as follows.

* First collect all the frequent items in the list.
* Now pay attention to only those constraints and include only them from the FP-trees and exclude all others.
* Now it does a filter kind of operation i.e. extract only those patterns for which the price is at least 200$.
* Get rid of all the data for which the items are frequently occurring and it overall cost is less than 200$.

**d.** The constraint is nonconvertible. The best way to mine this is using FP-Growth. And the process mentioned below.

* Get the frequent items list and arrange them in ascending or descending order.
* Now get rid of all the data for which the conditional pattern is not matching. Here in this case mine only the data for which the cost of prices is more than 100$. And purge all the data otherwise.
* Similarly do it for anther constraint, mine all the data for which the cost is less than 500$ and prune or exclude all the data otherwise.

**References:**

<http://www.crisismanagement.com.cn/templates/blue/down_list/llzt_dsj/Data%20Mining%20Concepts%20and%20Techniques%203rd%20Edition.pdf>