**FP Tree construction by compressing the DB representing frequent items**

Compressing the transactional database to mine association rules by finding frequent itemsets into a frequent pattern tree or FP-tree. This also retains the itemset association information.

So let’s start with a small transaction data to understand the construction of the FP tree. The transaction which we consider here suppose consists of 5 items such as-

Asparagus (A), Corn (C), Beans (B), Tomatoes (T) & Squash (S)

**Table 1**

|  |  |
| --- | --- |
| **Transaction ID** | **List of items in the transaction** |
| T1 | B , A , T |
| T2 | A , C |
| T3 | A , S |
| T4 | B , A , C |
| T5 | B , S |
| T6 | A , S |
| T7 | B , S |
| T8 | B , A , S , T |
| T9 | B , A , S |

So for example, for the first transaction T1 consists of three items such as Beans (B), Asparagus (A), and Tomatoes (T). Similarly, the transaction T6 contains the items Asparagus (A) and Squash (S). Let us also consider the minimum support for this small transaction data to be 2. Hence, min\_support = 2.

First of all, we need to create a table of item counts in the whole transactional database as below:

**Table 2**

|  |  |
| --- | --- |
| **Item** | **Support Count** |
| Beans (B) | 6 |
| Asparagus (A) | 7 |
| Squash (S) | 6 |
| Corn (C) | 2 |
| Tomatoes (T) | 2 |

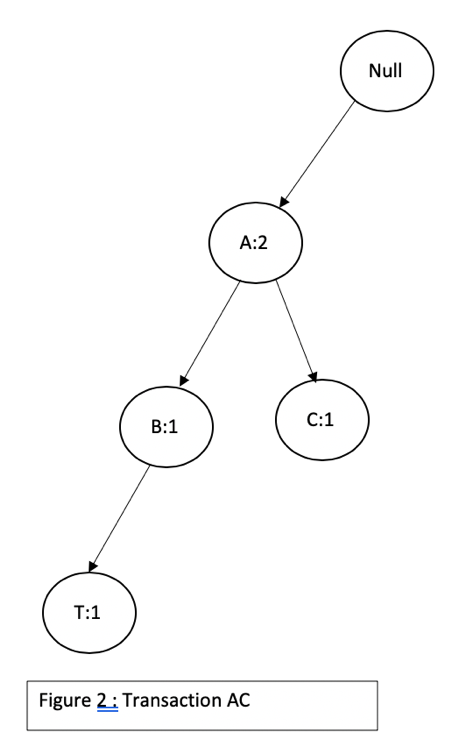
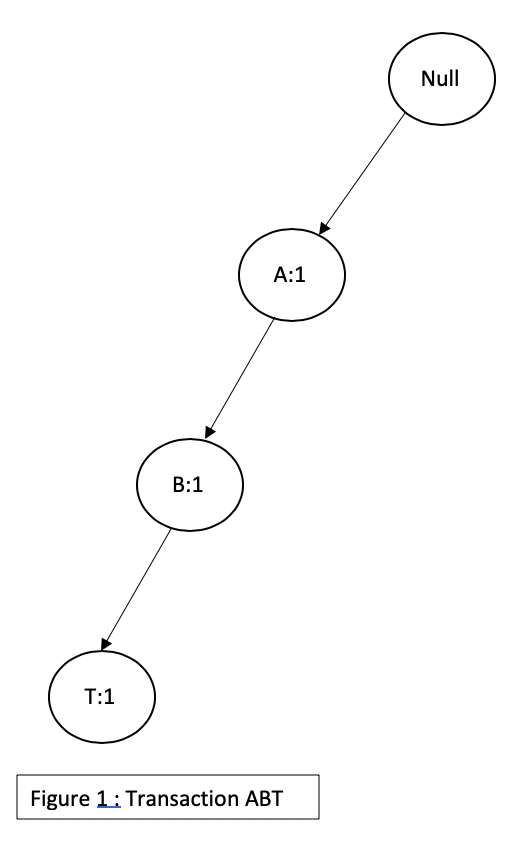
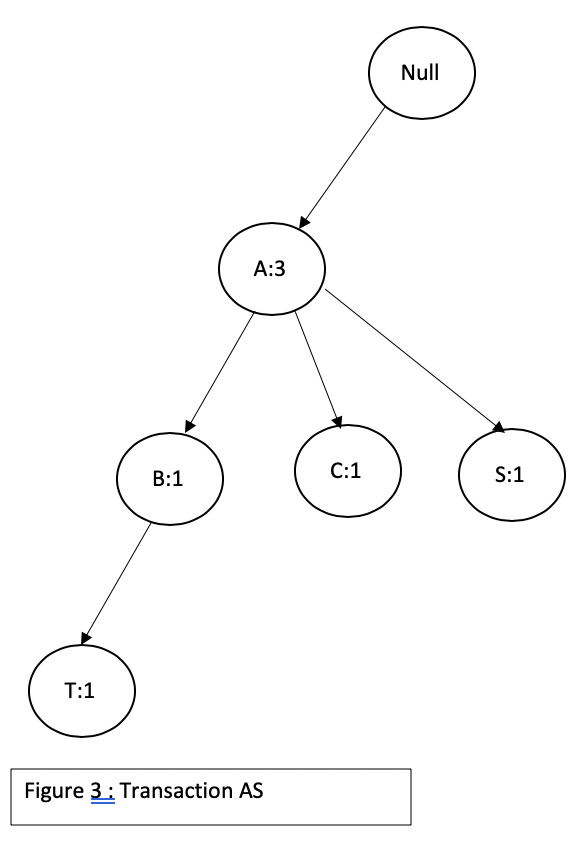
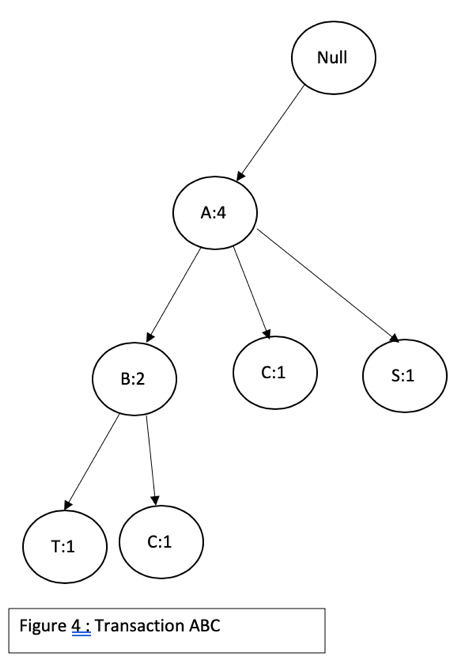
This is simply the count of each item, such as if we see Squash (S) has been bought in 6 transactions viz, T3, T5, T6, T7, T8 & T9, so the support count is 6 for Squash.

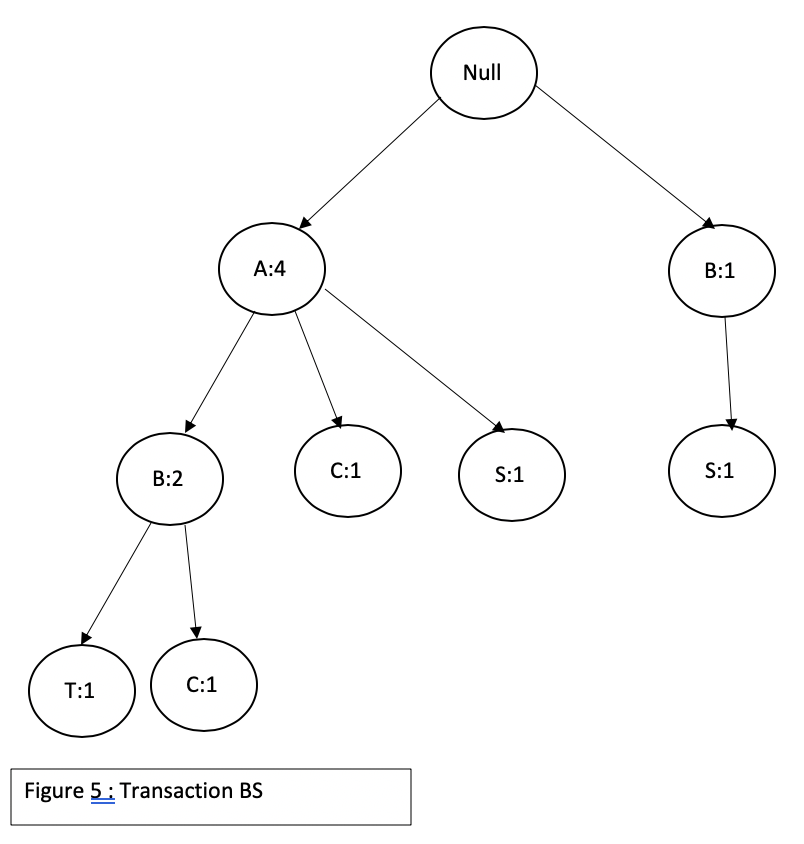
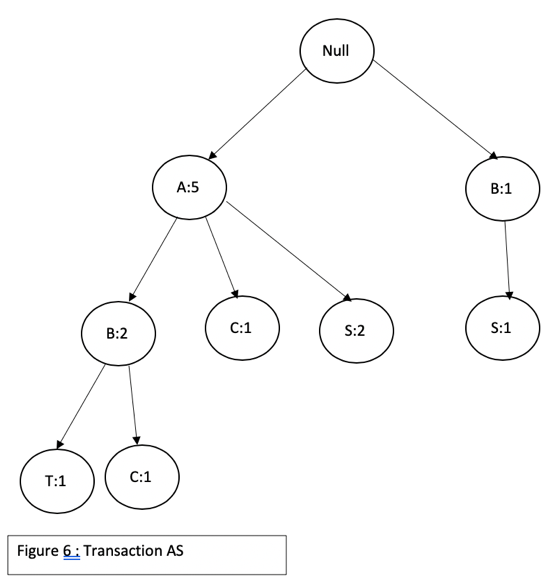
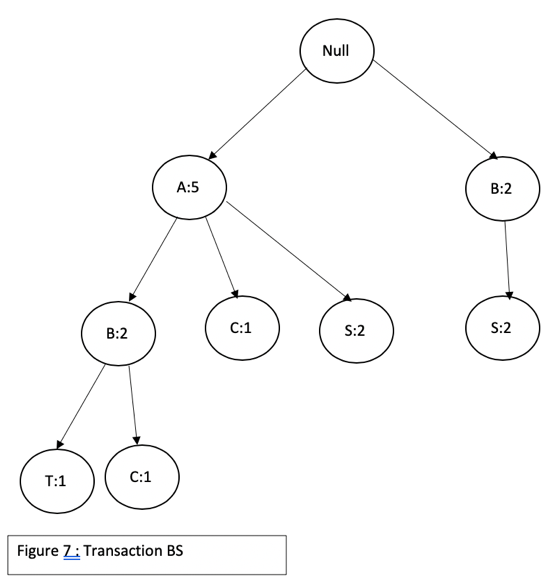
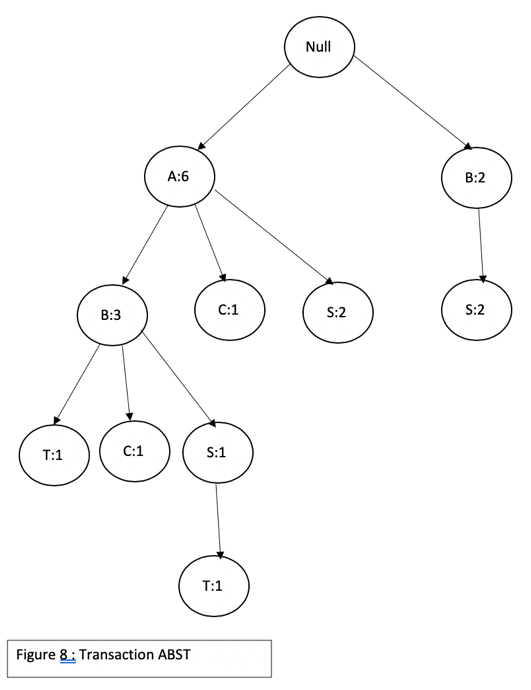
Next, we just need to sort the item list in descending order for their support count. Hence the table of support count may now be as represented in table 3 below:

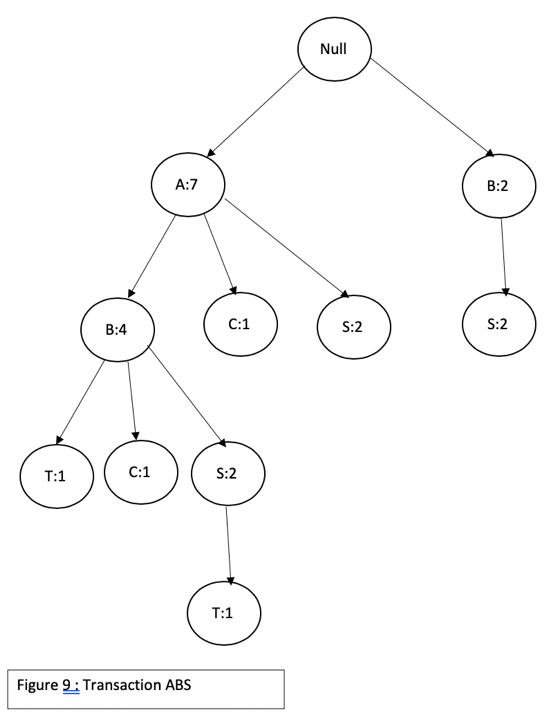
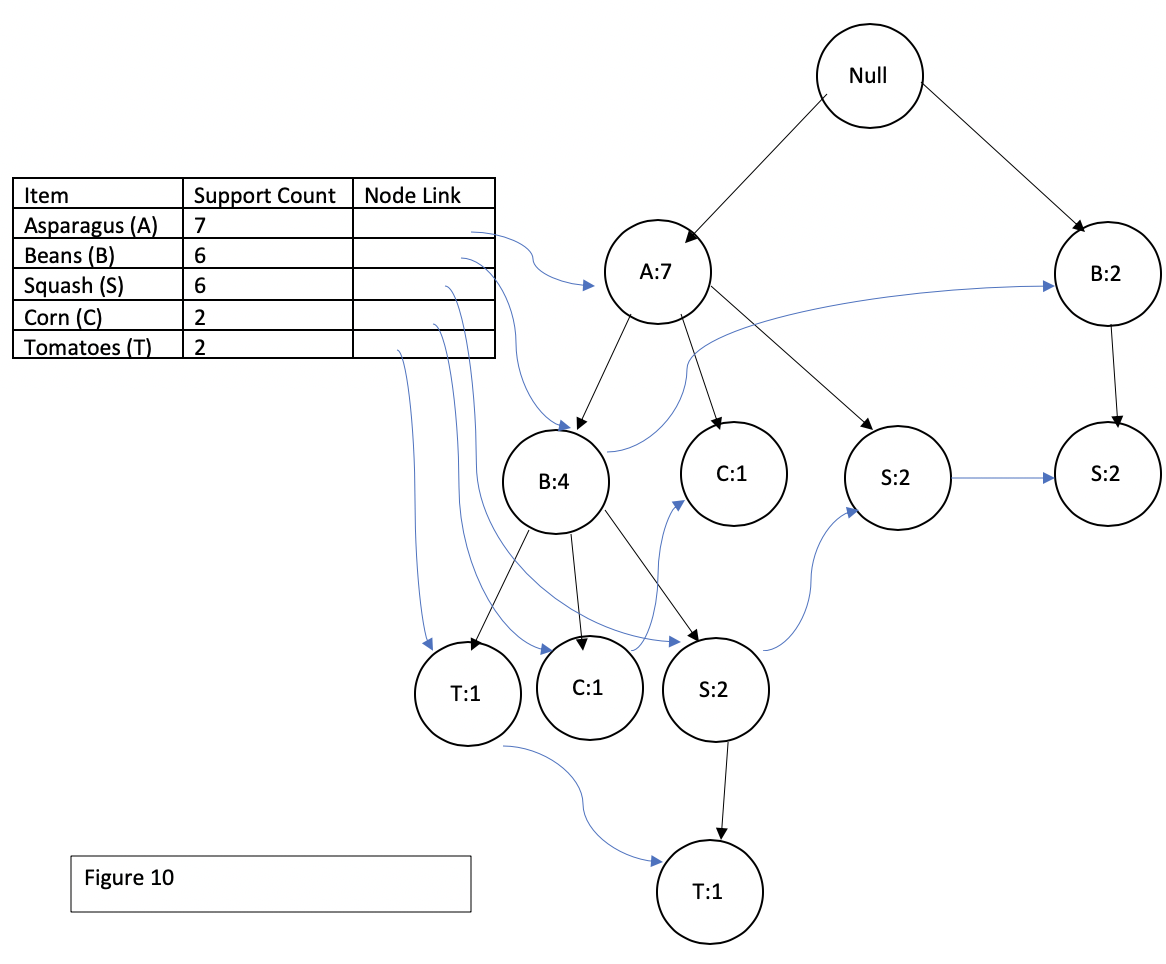
**Table 3**

|  |  |
| --- | --- |
| **Item** | **Support Count** |
| Asparagus (A) | 7 |
| Beans (B) | 6 |
| Squash (S) | 6 |
| Corn (C) | 2 |
| Tomatoes (T) | 2 |

|  |  |
| --- | --- |
| **Transaction ID** | **List of items in the transaction** |
| T1 | A , B , T |
| T2 | A , C |
| T3 | A , S |
| T4 | A , B , C |
| T5 | B , S |
| T6 | A , S |
| T7 | B , S |
| T8 | A , B , S , T |
| T9 | A , B , S |







**Table 4**

|  |  |  |  |
| --- | --- | --- | --- |
| **Item** | **Conditional Pattern base** | **Conditional FP tree** | **Frequent Pattern Generation** |
| Tomatoes (T) | {{A,B:1},{A,B,S:1}} | <A:2,B:2> | {A,T:2},{B,T:2},{A,B,T:2} |
| Corn (C) | {{A,B:1},{A:1}} | <A:2> | {A,C:2} |
| Squash (S) | {{A,B:2},{A:2},{B:2}} | <A:4,B:2>,<B:2> | {A,S:4},{B,S:4},{A,B,S:2} |
| Bean (B) | {{A:4}} | <A:4> | {A,B:4} |