How to convert CFG to Chomsky normal form (CNF)?

Step 1. Add new start symbol S_0 and the rule S_0->S, where S was the original start variable.

Step 2. Eliminate ε - productions. Remove all productions of the form $A \rightarrow \varepsilon$ if A is not a start variable, and for each occurrence of A on the RHS of a rule, we add a new rule with this occurrence deleted. For example, for production $R \rightarrow uAvAw$ add the following three productions: $R \rightarrow uvAw$, $R \rightarrow uAvw$, $R \rightarrow uvw$.

Step 3. Unit rules. Remove unit productions of the form $A \rightarrow B$. Whenever $B \rightarrow u$ appears, add $A \rightarrow u$ unless this was a unit rule previously removed (here u is a string of variables and terminals).

Step 4. Eliminate RHS with more than two symbols. Replace productions of the form
$A \rightarrow u_1 u_2 \ldots u_k$ (where $k \geq 3$ and u_i is a variable or terminal symbol for all $1 \leq i \leq k$) with the following productions:

$A \rightarrow u_1 A_1$
$A_1 \rightarrow u_2 A_2$
$A_2 \rightarrow u_3 A_3$
…
$A_{k-2} \rightarrow u_{k-1} u_k$ where $A_1, \ldots, A_{k-2}$ are new variables.

Replace any terminal u_i in the obtained rules with the new variable U_i and add rule
$U_i \rightarrow u_i$.

Example – Let us take an example to convert CFG to CNF. Consider the given grammar G1:

S → ASB

A → aAS|a|ε

B → SbS|A|bb

Step 1. Add a new start variable S_0 and rule S_0->S. Therefore, the grammar will become:

S_0->S

S → ASB

A → aAS|a|ε

B → SbS|A|bb

Step 2. As grammar contains null production A-> ε, its removal from the grammar yields:

S_0->S

S → ASB|SB

A → aAS|aS|a

B → SbS | A | ε | bb

Now, it creates null production B→ ε, its removal from the grammar yields:

S_0->S

S → ASB|AS|SB|S

A → aAS|aS|a

B → SbS|A|bb

Step 3. Remove unit rules. Remove B->A from the grammar yields:

S_0→S

S → ASB|AS| SB| S

A → aAS|aS|a

B → SbS|bb|aAS|aS|a

Also, removal of unit production S->S from grammar yields:

S_0→S

S → ASB|AS| SB

A → aAS|aS|a

B → SbS|bb|aAS|aS|a

Also, removal of unit production S_0->S from grammar yields:

S_0→ ASB|AS| SB

S → ASB|AS| SB

A → aAS|aS|a

B → SbS|bb|aAS|aS|a

Step 4. Convert remaining rules. First, convert each rule with 3 or more symbols in the right hand side (RHS). Use new variables X_1, X_2, X_3:

S_0→ AX_1|AS|SB

X_1→ SB

S → AX_1|AS|SB

A → aX_2|aS|a

X_2→ AS

B → SX_3|bb|aX_2|aS|a

X_3→ bS

Next, replace terminals in the obtained rules with new variables. We get the following grammar:

S_0-> AX_1|AS|SB

X_1→ SB

S → AX_1|AS|SB

A → $U_a$$X_2$|$U_a$S|a

X_2→ AS

B → SX_3|$U_b$$U_b$|$U_a$$X_2$|$U_a$S|a

X_3→ U_bS

U_a → a

U_b → b

So this is the required CNF for given grammar.