

ABSTRACT : Listings are essential part of clinical reporting. The file size of some of the listings (e.g.: lab, vitals, questionnaires) can go up to 400-500 MB. The time it takes to run and open in validated system is huge. It also poses challenge in storage of huge listing in document management system. Manual split requires revisiting .rtf file again and again after each rerun whether the file size has crossed the limit or not. The method we follow to automatically check and re-split the file if it increases from the threshold size. This paper is discussing the challenges and the method of splitting the listings based on minimum size of file you want. It contains set of macros which automatically splits the file into smaller file size which will easier to open and store.

Background

- Difficult to know the file size of all the final documents being produced in study.
- Currently being manually checked
- Requirement to generate listings having size < 30MB

Challenges

- What is the split criteria?
- How to split file (Ex: 90MB file to three 30 MB files or something else)?
- How many number files to be created (3 or more)?.
- What if file size is more than 30MB after splitting also.
- How to increment the number of files count in this case?
- How to delete files which will have size more than 30MB?
- Before deleting how to check whether the file is created or not?
- How to avoid hitting ADS dataset multiple times for multiple iterations?
- How to avoid creating multiple files having smaller size after all?

Usage

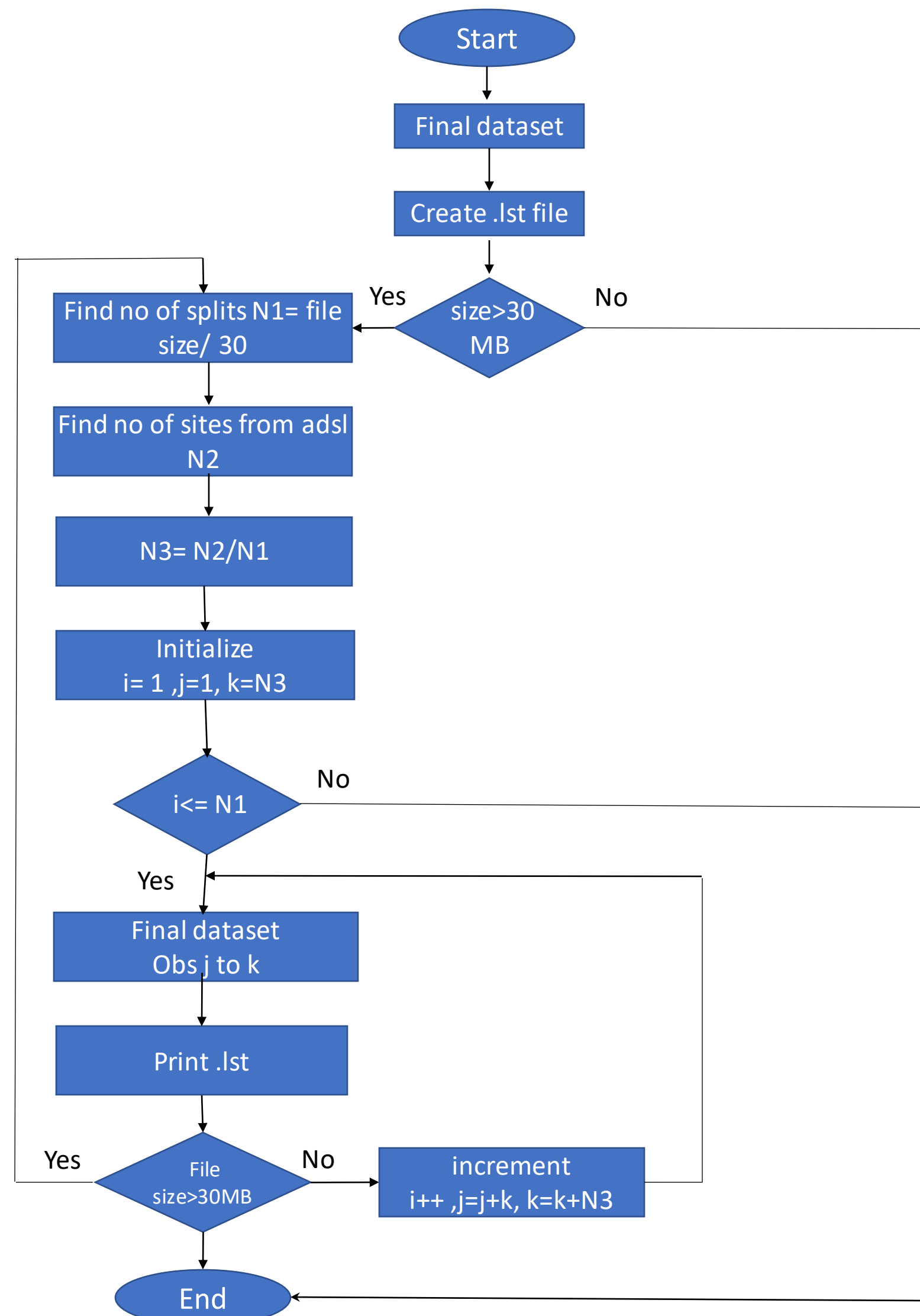
We can use this logic to split laboratory ,Questionnaire and Vital sign listings where we can see huge data.

Technology

SAS

We have used SAS MACRO's concept to code where the macro will take care of all the above challenges.

Flow chart



Steps to split files

- Create .lst file with all site id's if the split criteria is SITE ID
- Find the created .lst file size and estimate how many files can be created.
- Based on the number of files (n) create a loop for n number of listings.
- On the first run itself save the final reporting dataset to a temporary location .
- Use this dataset in loop and split sites for each iteration and generate .lst file.
- There might be a chance of having created file size >30MB , so at the end of each iteration need to check the file size .
- If split file size is >30MB then repeat the loop with same starting site id.
- Also, number of files that need to be created need to be changed as per re-split.
- Used proc print to create multiple files and log file.
- Delete temporary datasets as good practice to handle memory using PROC DELETE.
- Make sure that each time file name should be changed according to the iteration.

Outcome

- .lst files will be created with <30MB size approximately.
- One log file will be generated

Conclusion

Because of this automated macro we saved lot of time in calculating size of the file based on number of patients in a particular set of patient sites when we are splitting listing files based on site id.