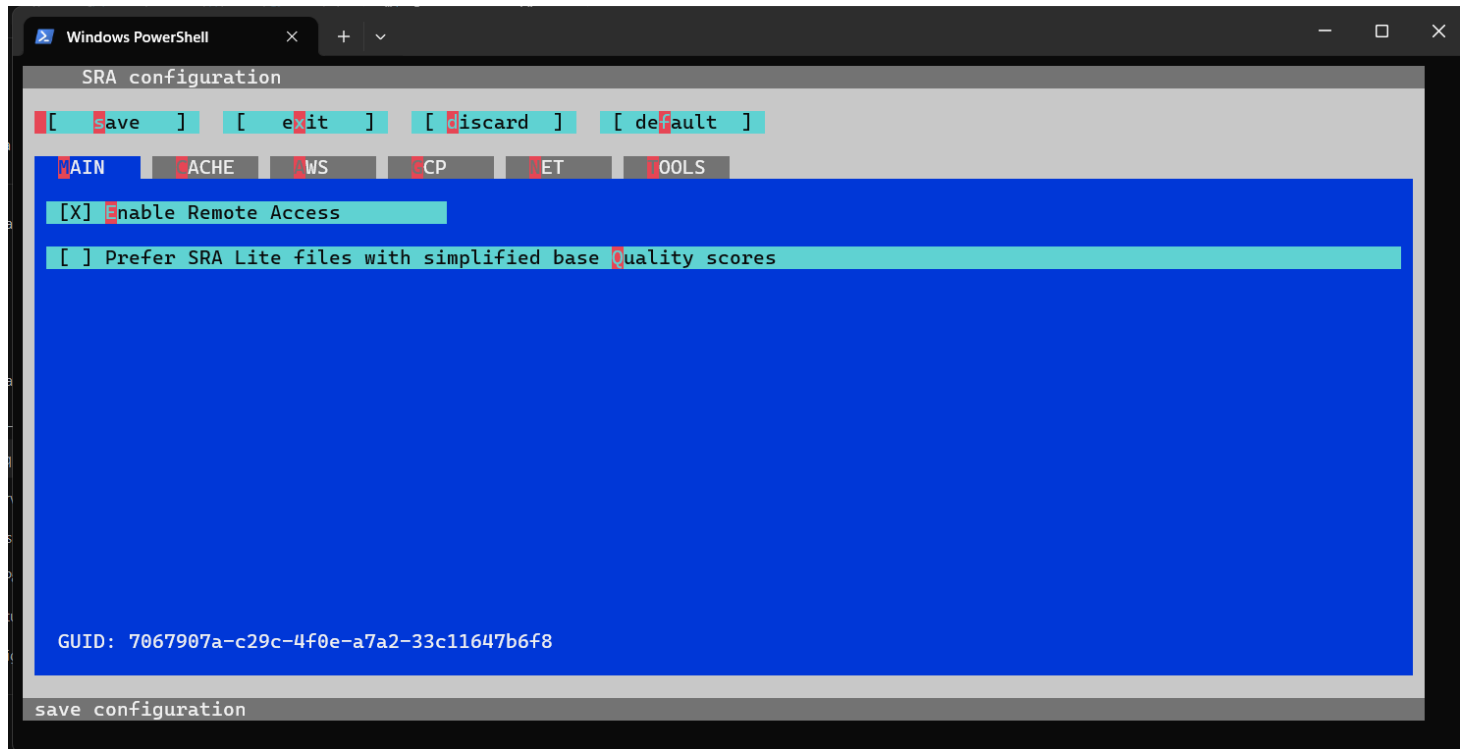Steps followed:

1. Download the SRA Toolkit for Windows
2. Extract the zip file and go to bin
3. Open from bin in the terminal
4. Run this command: `.\vdb-config.exe --interactive`
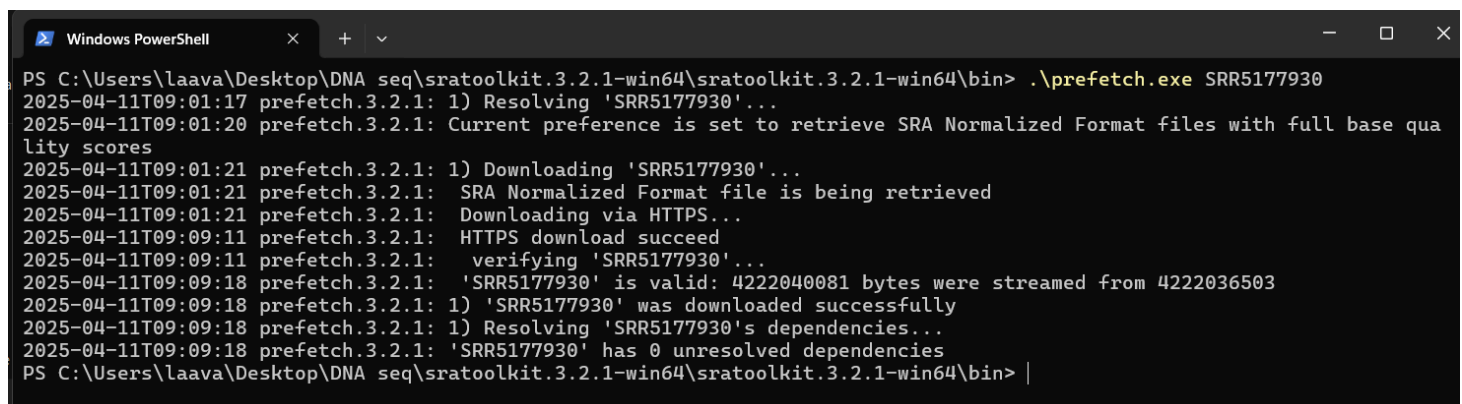5. Then you will be redirected to the configuration.

The configuration looks something like this:

Enable Remote Acces by pressing X

Then go to Cache (press E) - enable local caching

Prefetch

```
Windows PowerShell                    ×   +  ∨                                    —  □  ×

Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS D:\sratoolkit.3.2.1-win64\sratoolkit.3.2.1-win64\bin> .\fastq-dump.exe --split-files SRR5177930 --outdir "D:\fastqout
put"
Read 60300521 spots for SRR5177930
Written 60300521 spots for SRR5177930
PS D:\sratoolkit.3.2.1-win64\sratoolkit.3.2.1-win64\bin>
```

Then we convert the files to FastQ format so that we can read them in python.

Azure Setup

- Go to VSCode, bottom left for SSH setup
- Connect to Host
- Configure SSH host
- It will open a configuration file
- Enter details of Host

```
Host construction-project-1
   HostName 51.8.55.132
   User azureuser
   IdentityFile "/Users/aakashwalavalkar/Desktop/Azure VM details/construction-project-1_key.pem"
```

The identity file (.pem) needs to be in C:/Users/Name/.ssh/pem file
and paste the path for identity file in the code.

Then after you are in VSCode and you ave connected to host, go to File Explorer and Open
AzureUser. You will be able to see all the folders.

Then go to the terminal:

```
conda deactivate
```

to come out of an existing environment (can be seen in *()* before the file path).

Run the commands and you will be able to see these files:

```
(base) azureuser@construction-project-1:~$ conda deactivate
azureuser@construction-project-1:~$  conda activate dna_sequence
(dna_sequence) azureuser@construction-project-1:~$ cd dna_sequencing/
(dna_sequence) azureuser@construction-project-1:~/dna_sequencing$ ls
SRR5177930  SRR5177930_1.fastq.gz  SRR5177930_2.fastq.gz  sratoolkit.3.2.1-ubuntu64  sratoolkit
(dna_sequence) azureuser@construction-project-1:~/dna_sequencing$
```

All working and package installation is to be done in this environment with respect to this project.