

Step 1: Install Apache Spark

1. **Download Apache Spark:** Go to the Apache Spark website (<https://spark.apache.org/downloads.html>) and download the latest version of Apache Spark.
2. **Extract Spark:** After downloading, extract the contents of the downloaded file to a directory of your choice. (Create DSBDAL folder in home directory and extract the contents)
3. **Set up Environment Variables:** Add Spark's `bin` directory to your `PATH` environment variable. You can do this by modifying your shell profile file (e.g., `.bashrc`, `.bash_profile`, `.zshrc`, etc.) (Our is bash)
 - a. **Open your Terminal:** Launch your terminal application. This process may differ slightly depending on your operating system (e.g., Terminal on macOS, Command Prompt or PowerShell on Windows, Terminal on Linux).
 - b. **Determine your Shell:** Before proceeding, determine which shell you're using. Common shells include Bash, Zsh, and Fish. You can typically find out your current shell by running the following command:

```
echo $SHELL
```

- c. **Edit the Shell Profile File:** Based on your shell, you'll edit the corresponding profile file:
 - i. For Bash (`~/.bashrc`):
- d. **Add Spark's bin Directory to PATH:** Inside the opened file, add the following line at the end:

```
export  
PATH="/home/student/DSBDAL/spark-3.5.1-bin-hadoop3/bin:$P  
ATH"
```

```
GNU nano 6.2 /home/student/.bashrc
# ~/.bashrc: executed by bash(1) for non-login shells.
# see /usr/share/doc/bash/examples/startup-files (in the package bash-doc)
# for examples

# If not running interactively, don't do anything
case $- in
  *(*) ;;
  *) return;;
esac

# don't put duplicate lines or lines starting with space in the history.
# See bash(1) for more options
HISTCONTROL=ignoreboth

# append to the history file, don't overwrite it
shopt -s histappend
export PATH="/home/student/DSBDAL/spark-3.5.1-bin-hadoop3/bin:$PATH"
# for setting history length see HISTSIZE and HISTFILESIZE in bash(1)
HISTSIZE=1000
HISTFILESIZE=2000

Read 133 lines
^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute  ^C Location
^X Exit      ^R Read File ^\ Replace   ^U Paste     ^J Justify  ^_ Go To Line
```

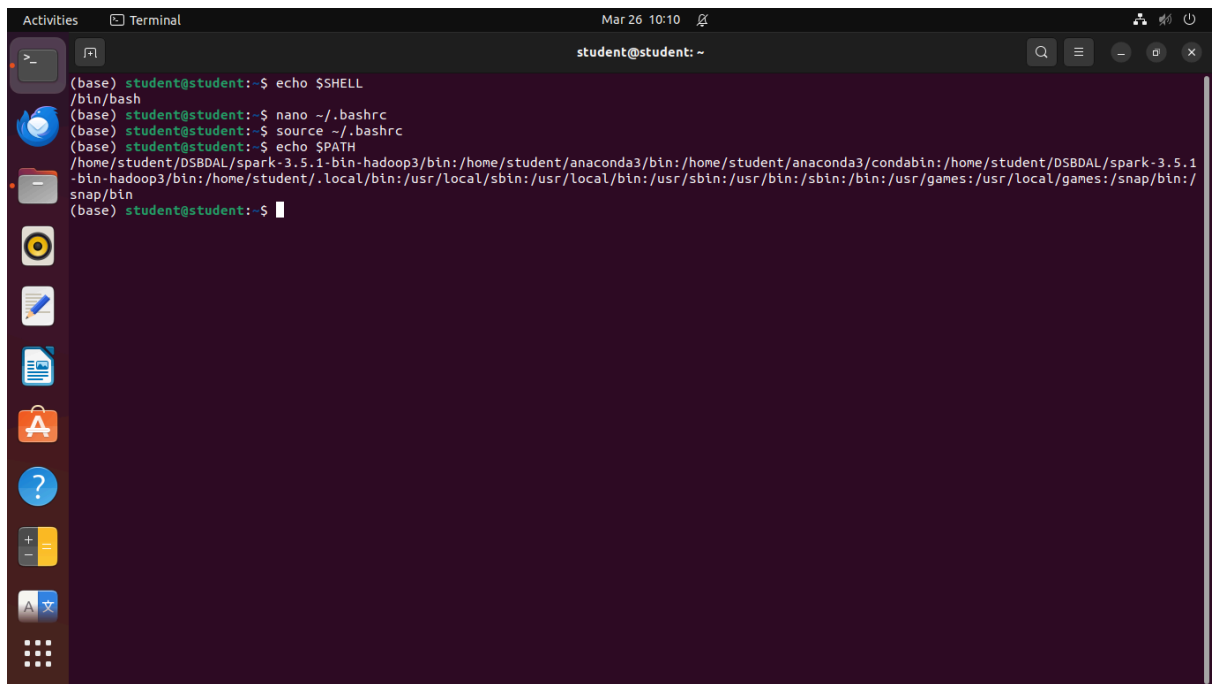
- e. **Save and Exit:** After adding the line, save the file and exit the editor. In Nano, you can do this by pressing **Ctrl + O** to write the file and **Ctrl + X** to exit.
- f. **Apply Changes:** To apply the changes to your current terminal session, either close and reopen the terminal or run:

```
source ~/.bashrc
```

- g. **Verify:** You can verify that Spark's **bin** directory has been added to your **PATH** by running:

```
echo $PATH
```

You should see the path to Spark's **bin** directory listed in the output.

A terminal window titled 'Terminal' with a dark background and light text. The window shows a series of commands and their outputs. The prompt is '(base) student@student: ~'. The commands and outputs are: 'echo \$SHELL' returns '/bin/bash'; 'nano ~/.bashrc' opens the nano editor; 'source ~/.bashrc' reloads the configuration; 'echo \$PATH' returns a long path including '/home/student/DSBDAL/spark-3.5.1-bin-hadoop3/bin:/home/student/.local/bin:/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/snap/bin'. The prompt returns to '(base) student@student: ~'.

```
(base) student@student: ~  
$ echo $SHELL  
/bin/bash  
(base) student@student: ~$ nano ~/.bashrc  
(base) student@student: ~$ source ~/.bashrc  
(base) student@student: ~$ echo $PATH  
/home/student/DSBDAL/spark-3.5.1-bin-hadoop3/bin:/home/student/anaconda3/bin:/home/student/anaconda3/condabin:/home/student/DSBDAL/spark-3.5.1  
-bin-hadoop3/bin:/home/student/.local/bin:/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/snap/bin:/  
snap/bin  
(base) student@student: ~$
```

4. To execute Code:

```
cd DSBDAL
```

```
spark-shell<WebLog_Processing.scala
```