Installation of nextflow on new Ubuntu with conda package manager

31 July 2024   12:58

1. Make sure you install the latest version of Java.
   ```
   sudo apt install default-jre
   ```
2. Use this command to confirm the Java Version – it should be above 11.
   ```
   java -version
   ```
3. Follow nextflow.io to install NextFlow on Linux machine.
   ```
   curl -s https://get.nextflow.io | bash
   chmod +x nextflow
   sudo mv nextflow /usr/local/bin
   ```
4. Use nextflow info to confirm the installation of NextFlow
   ```
   nextflow info
   ```
5. Installing Conda - Ref and then add channels Bioconda and conda-forge
   a. https://www.anaconda.com/download/success
   ```
   Or
   wget https://repo.anaconda.com/archive/Anaconda3-2024.06-1-Linux-x86_64.sh
   Chmod +x /home/talktoamitdhir/Downloads/Anaconda3-2024.06-1-Linux-x86_64.sh
   /home/talktoamitdhir/Downloads/Anaconda3-2024.06-1-Linux-x86_64.sh
   Yes
   Yes [when init is asked]
   /home/talktoamitdhir/anaconda3//bin/conda init
   conda config —set auto_activate_base false
   ```
6. `confirm if Conda is installed.`

   conda —h
7. Set up BiConda after installation of conda Ref
   ```
   conda config --add channels bioconda
   ```
   conda config --add channels conda-forge
   https://bioconda.github.io/
8. Confirm Conda channels Ref
   Conda config —show channels
9. Set up conda environment and create new conda Env.
   conda env list
   conda create -n bioinformatics
   conda activate bioinformatics
10. Install samtools in new conda environment.
    conda install –c bioconda samtools
    conda install –c bioconda fastqc
    conda install –c bioconda nf-core
11. Update Nextflow.
    nextflow self-update [updates and fixed the java issue]
12. Use following command to instantiate the git repo
    ```
    eval $(ssh-agent -s)
    ssh-add ubuntu
    ```
13. `Create a new FastQSheet.csv file.`
14. `Go to working directory and run following command`
    ```
    sudo nextflow run nf-core/rnaseq --input FastQSheet.csv --outdir "/"  --genome
    GRCh37
    ```
15. Follow these instructions for installation of the RNASeq.
    https://nf-co.re/docs/usage/installation/
    https://nf-co.re/rnaseq/3.14.0/
    https://nf-co.re/docs/usage/introduction#how-to-run-a-pipeline
    https://nf-co.re/docs/usage/configuration#custom-configuration-files

Helpful Links

1. Sample Data fastQ files - Link.
2. Wget command sample to download the files on Linux.
3. Use realpath command to get the path to the folder structure

4. IMPORTANT - Nfcore video link
5.

| EC2 Instance Connect | Session Manager | SSH client | EC2 serial console |

Instance ID

🗗 i-04603ea46e2540408 (preludeNF02)

1. Open an SSH client.

2. Locate your private key file. The key used to launch this instance is preludeNF02.pem

3. Run this command, if necessary, to ensure your key is not publicly viewable.
   chmod 400 "preludeNF02.pem"

4. Connect to your instance using its Public DNS:
   ec2-54-210-50-85.compute-1.amazonaws.com

Example:

ssh -i "preludeNF02.pem" ubuntu@ec2-54-210-50-85.compute-1.amazonaws.com

6.

> (i) **Note:** In most cases, the guessed username is correct. However, read your AMI usage instructions to check if the AMI owner has changed the default AMI username.

```
https://d-90678d2a22.awsapps.com/start/#/console?account_id=
896191365710&role_name=Sridhar_Raidi


896191365710

https://us-east-1.signin.aws/platform/login?workflowStateHandle=
87b75226-fabd-4b90-8b64-7163e0c7da14


Connection Type SSH using putty and ppk file
EC2 Public IP Address : 54.210.50.85
Port 22
use ppk as authentication file.

Connection Type terminal using pem file
Go to /c/Users folder
ssh -i hopde/OneDrive/Desktop/aws/preludeNF02.pem
ubuntu@ec2-54-210-50-85.compute-1.amazonaws.com
```

Code to login to EC2
ssh -I hopde/OneDrive/Desktop/aws/preludeNF02.pem ubuntu@ec2-54-210-50-85.compute-1.amazon.aws.com

Code to copy from local to EC2
scp -i ~/preludeNF02.pem local_image_file ubuntu@ec2-54-210-50-85.compute-1.amazon.aws.com:/home/user/images/

App.pluto.bio
MultiQC

Requirement

https://github.com/nf-core/rnaseq/tree/3.14.0

Workflow: Fastq Files to Final Algorithm Processing

1. **Start with Fastq Files:**
   - Input: Raw Fastq files containing the RNA sequencing data.

2. **NF-Core RNA-Seq Pipeline:**
   (Public Pipeline) https://nf-co.re/docs/nf-core-tools/installation
   - **QC and Alignment:**
     - Use the NF-Core RNA-Seq pipeline to process the Fastq files. This pipeline will handle alignment of reads and will QC data.
     - **This Output will be generated:** Aligned

reads (BAM files) and QC reports.

https://nf-co.re/events/2022/bytesize-32-nf-core-rnaseq
nf-core/rnaseq (nf-core/bytesize #32)

nf-core/rnaseq (nf-core/bytesize #32)

Sample Name	duplnt	% Dups	5'-3' bias	M Aligned	% Alignable	% Proper Pair
sample_15ng_75_23_18107_G04_2						
sample_15ng_75_23_R00724_3_B04	0.14%	88.8%	20.65	15.0	95.2%	7.2%
sample_15ng_75_23_R00724_3_B04_1						
sample_15ng_75_23_R00724_3_B04_2						
sample_15ng_75_23_R00728_5_C04	0.12%	89.7%	10.61	19.8	94.4%	5.9%
sample_15ng_75_23_R00728_5_C04_1						
sample_15ng_75_23_R00728_5_C04_2						
sample_15ng_75_23_R00746_7_D04	0.13%	88.9%	11.80	16.2	95.7%	7.1%

Showing 238/238 rows and 22/29 columns.

Copy table Configure Columns Plot

Open in