

Experiment – 2

(i) To setup and install Hadoop in all three modes.

```
aarushi@aarushiLT: ~$ apt-get update -y
Reading package lists... Done
E: Could not open lock file /var/lib/apt/lists/lock - open (13: Permission denied)
E: Unable to lock directory /var/lib/apt/lists/
W: Problem unlinking the file /var/cache/apt/pkgcache.bin - RemoveCaches (13: Permission denied)
W: Problem unlinking the file /var/cache/apt/srcpkgcache.bin - RemoveCaches (13: Permission denied)
aarushi@aarushiLT ~$ sudo apt-get update -y
[sudo] password for aarushi:
Hit:1 http://repo.mysql.com/apt/debian buster InRelease
Hit:2 http://deb.debian.org/debian bullseye InRelease
Hit:3 https://cli.github.com/packages stable InRelease
Hit:4 http://deb.debian.org/debian bullseye-updates InRelease
Hit:5 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_64 InRelease
Ign:6 http://developer.download.nvidia.com/compute/machine-learning/repos/ubuntu1804/x86_64 InRelease
Hit:7 http://developer.download.nvidia.com/compute/machine-learning/repos/ubuntu1804/x86_64 Release
Reading package lists... Done
aarushi@aarushiLT ~$ sudo apt-get install default-jdk default-jre -y
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
default-jdk is already the newest version (2:1.11-72).
default-jre is already the newest version (2:1.11-72).
default-jre set to manually installed.
0 upgraded, 0 newly installed, 0 to remove and 2 not upgraded.
aarushi@aarushiLT ~$ java -version
openjdk version "11.0.16" 2022-07-19
OpenJDK Runtime Environment (build 11.0.16+8-post-Debian-1deb11u1)
OpenJDK 64-Bit Server VM (build 11.0.16+8-post-Debian-1deb11u1, mixed mode, sharing)
aarushi@aarushiLT ~$
```

```
aarushi@aarushiLT: ~$ java -version
openjdk version "11.0.16" 2022-07-19
OpenJDK Runtime Environment (build 11.0.16+8-post-Debian-1deb11u1)
OpenJDK 64-Bit Server VM (build 11.0.16+8-post-Debian-1deb11u1, mixed mode, sharing)
aarushi@aarushiLT ~$ adduser hadoop
zsh: command not found: adduser
aarushi@aarushiLT ~$ sudo adduser hadoop
Adding user `hadoop' ...
Adding new group `hadoop' (1002) ...
Adding new user `hadoop' (1002) with group `hadoop' ...
Creating home directory `/home/hadoop' ...
Copying files from `/etc/skel' ...
New password:
Retype new password:
passwd: password updated successfully
Changing the user information for hadoop
Enter the new value, or press ENTER for the default
Full Name []: Aarushi Vohra
Room Number []:
Work Phone []:
Home Phone []:
Other []:
Is the information correct? [Y/n]
aarushi@aarushiLT ~$
```

```
aarushi@aarushiLT: ~ 80x24
Other []:
Is the information correct? [Y/n]
aarushi@aarushiLT ~$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/aarushi/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/aarushi/.ssh/id_rsa
Your public key has been saved in /home/aarushi/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:Te9iPxfQahUnV97/A+vu1wKf7nIfUgcYPhGZpQ++4dM aarushi@aarushiLT
The key's randomart image is:
+---[RSA 3072]-----+
|           +=. o|
|           .o=o.+|
|           . *. .=o|
|           o o.+o..|
|           S . +=..o|
|           o+=. o|
|           o.==E=o|
|           . o+.B.+|
|           +@+o.|
+---[SHA256]-----+
aarushi@aarushiLT ~$
```

```
hadoop@aarushiLT: ~ 80x24
aarushi@aarushiLT ~$ sudo - hadoop 130 ↵
sudo: -: command not found
aarushi@aarushiLT ~$ su - hadoop 1 ↵
Password:
su: Authentication failure
aarushi@aarushiLT ~$ su - hadoop 1 ↵
Password:
hadoop@aarushiLT:~$ https://downloads.apache.org/hadoop/common/hadoop-3.3.4/hado
op-3.3.4.tar.gz
-bash: https://downloads.apache.org/hadoop/common/hadoop-3.3.4/hadoop-3.3.4.tar.
gz: No such file or directory
hadoop@aarushiLT:~$ wget https://downloads.apache.org/hadoop/common/hadoop-3.3.4
/hadoop-3.3.4.tar.gz
--2023-01-29 21:31:26-- https://downloads.apache.org/hadoop/common/hadoop-3.3.4
/hadoop-3.3.4.tar.gz
Resolving downloads.apache.org (downloads.apache.org)... 2a01:4f8:10a:201a::2, 2
a01:4f9:3a:2c57::2, 88.99.95.219, ...
Connecting to downloads.apache.org (downloads.apache.org)|2a01:4f8:10a:201a::2|:
443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 695457782 (663M) [application/x-gzip]
Saving to: 'hadoop-3.3.4.tar.gz'

hadoop-3.3.4.tar.gz 1%[ ] 10.66M 949KB/s eta 14m 37s
```

(ii) Explore various dataset repositories like Kaggle, UCI ML, Google Dataset, Reddit, MNIST etc.

The screenshot shows the Kaggle Datasets page. The left sidebar contains navigation links: Home, Competitions, Datasets (selected), Code, Discussions, Learn, and More. The main content area is titled 'Datasets' and shows a search bar with 'Search 3,315 datasets'. A filter for 'Classification' is applied. Below the search bar, there are three dataset cards: 'Hotel Reservations Dataset' (Ahsan Raza, Updated 24 days ago, Usability 10.0, 1 File (CSV) - 491 kB), 'Home Loan Approval' (Rushikesh Konapure, Updated 16 days ago, Usability 10.0, 2 Files (CSV) - 13 kB), and 'Superstore Marketing Campaign Dataset' (Ahsan Raza, Updated a month ago, Usability 10.0, 1 File (CSV) - 57 kB). Each card shows a 'Hotness' score and a 'Silver' or 'Bronze' medal.

The screenshot shows the UCI Machine Learning Repository homepage. The header features the UCI logo and the text 'Machine Learning Repository' and 'Center for Machine Learning and Intelligent Systems'. A search bar is present. Below the header, there is a banner for the 'beta version' of the repository. The main content area is titled 'Welcome to the UC Irvine Machine Learning Repository!' and contains a paragraph about the repository's mission and a link to the 'About' page. Below this, there are three columns: 'Latest News' (listing updates from 2017 to 2018), 'Newest Data Sets' (listing 'Average Localization Error (ALE) in sensor node localization process in WSNs', '9mers from cullpdb', and 'TamilSentMix'), and 'Most Popular Data Sets (hits since 2007)' (listing 'Iris', 'Adult', and 'Dry Bean Dataset').

Experiment-2 x Dataset Search x +

← → ↺ https://datasetsearch.research.google.com/search?src=0&query=iris flower dataset&docid=L2cvMTFrNHZA2MWZyNA

Google iris flower dataset

▼ Last updated ▼ Download format ▼ Usage rights ▼ Topic Free Saved data sets

98 data sets found

kaggle IRIS-FLOWER-DATASETS

kaggle.com

zip

Updated Dec 26, 2022

iris Dataset

paperswithcode.com

Data from: iris-data

data.world

csv, zip

Updated Nov 26, 2022

kaggle

IRIS-FLOWER-DATASETS

A multivariate dataset

Explore at: [Kaggle | kaggle.com](#)

zip(2025 bytes)

Data set updated

Dec 26, 2022

Authors

Simran Sharma

Licence

[CC0 1.0 Universal Public Domain Dedication](#)

Experiment-2 x Datasets x +

← → × https://www.reddit.com/r/datasets/

reddit Search Reddit Get App Log In

FEEDS

- Home
- Popular

TOPICS

- Gaming
- Sports
- Business, Economics, ...
- Crypto
- Television
- Celebrity

Create an account to follow your favorite communities and start taking part in conversations.

[Join Reddit](#)

Datasets

r/datasets

Posts mod

[Hot](#) [New](#) [Top](#) ...

Posted by 48 minutes ago

request UK National/International Days/Weeks

Does anyone have a clean dataset on National / International Days / Weeks centred around the UK before I start scrapping?

Currently volunteering at a children's charity and they would like to start themed weeks so the dataset would be very handy.

0 Comments → Share Save ...

About Community

A place to share, find, and discuss Datasets.

Created Oct 8, 2009

170k Members **60** Online

Related Communities

[r/uberdrivers](#) 339,108 members [Join](#)