

## Maximizing Your Time with a CBHDS Biostatistician/Bioinformatician

### Data Transfer and Management

Please consider the following guidelines when using Virginia Tech's [Advanced Research Computing \(ARC\)](#) for data storage. This document is created based on [VT ARC documentation](#). The purpose of this document is to provide general guidelines for the use of ARC for data transfer and management at the Center for Biostatistics and Health Data Sciences (CBHDS). Note that You do not need to have any prior experience with high-performance computing (HPC) to follow this document. **Before you start following the instructions below, please make sure you are connected to VT's network.** If you have trouble with any section of the instructions provided in this document, please email Dr. Missi Zhang at [missizxm@vt.edu](mailto:missizxm@vt.edu).

#### **ARC Services**

ARC is a unit within VT's Division of Information Technology, providing centralized research computing infrastructure and support the university's research community. VT ARC provides high-performance computing systems, large-scale data storage, big data processing, software, and consulting services. It is designed for researchers from different disciplines or in different locations to collaborate on a project. ARC is very helpful when you have large jobs to run (e.g., high resolution or large-scale datasets), many jobs to run (parameter sweeps, many datasets), or needs for specialty hardware (GPU, memory, high bandwidth storage, fast network, scale). ARC also provides introductory training sessions via the TLOS [Professional Development Network](#) and there are computational scientists available in the ARC team to provide the necessary support.

#### **For VT Users**

ARC is available to all Virginia Tech faculty at no cost. Researchers and groups have the option to add computing costs to grants or contracts through ARC's Cost Center for requesting additional compute or storage resources. Departments and faculty can also purchase priority access to an ARC system for up to five years through ARC's Investment Computing Program.

## Links to ARC References and Additional Resources

- [Office Hours](#)
- [Request a Consultation](#)
- [Create an ARC User Account](#)
- [New ARC Users Info](#)
- [Video tutorials](#)
- [Computing Resources](#)
- [Storage Resources](#)
- [Visualization Services](#)
- [ARC Faculty & Staff Directory](#)

## Getting Started with ARC

This section contains instructions for creating user accounts, requesting project storage space and computation allocations, managing projects and user lists, as well as transferring and managing data on ARC. **Before you follow these instructions, please make sure you are connected to VT's campus network.** If you are not using VT's campus network, please use VT's Ivanti Remote Access VPN. Instructions about how to connect to VT's Remote Access VPN can be found [here](#).

### 1. Create an ARC user account.

ARC accounts are based on VT PID accounts and authentication.

- a valid VT PID is required and logins to ARC systems are authenticated with VT PID **username and password** plus **DUO second factor authentication**.
- All current faculty, staff, and students at VT have a VT PID.
- Collaborators external to VT can get a "[sponsored VT PID](#)" when that is requested by a faculty member at VT.

**Anyone with an active VT PID can request access to ARC system.** The [account creation form](#) will require you to confirm your acceptance of VT usage policies, and then create your account. Once your account is created, you can login in ARC.

Next, you will need to create a project and request allocations for this project. An allocation is a system time account requested and managed by a single person (e.g., a project Principal

Investigator [PI]). Many users (e.g., Co-Investigators [Co-I] or graduate students) can then be granted access to a single allocation. There are two types of project allocations:

**Research** Allocations: for research projects & managed by the project's PI

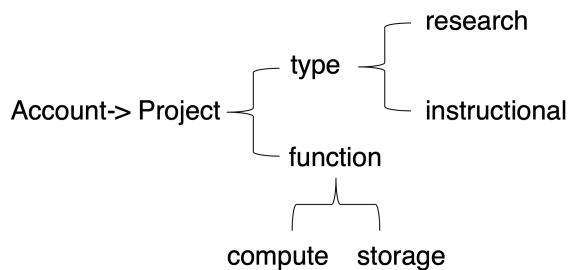
- Typically granted for a single year and can be renewed annually for the length of the project.
- Multi-year research allocations may be granted through negotiation with ARC.
- Can be used for CBHDS data science projects.

Who is **eligible** for research allocations?

- a current faculty member or post-doctoral researcher at VT.
- an employee of VT and the PI for a research computing-related project.
- an employee of VT and the Co-I for a research computing-related project led by a non-VT PI  
Adjunct professors must provide a letter from their department chair, indicating that they are qualified to lead an internal research project, before their project and allocation requests can be approved.

**Instructional** Allocations: for academic classes & managed by the faculty member/instructor

- typically smaller, available for shorter time periods
- limited to a select set of systems.
- Can be used for training workshops.



*Figure 1.1. The relationship between project allocation type and function.*

## 2. Create a new project.

Once your ARC account is approved, you can go to [ColdFront](#) to create a project, add users, request allocations, and add grant/publication information.

- 2.1. Log into [ColdFront](#) using your ARC account username and password.
- 2.2. On the home page, your username should appear on the top right corner (as depicted in Figure 2.1 below). To create a new project, click “Projects” and then click “Add a project”.

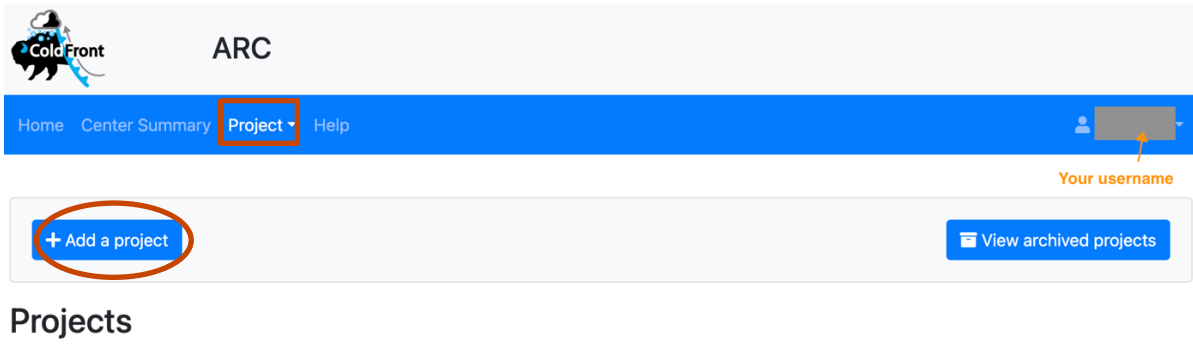


Figure 2.1. The home page of ColdFront.

- 2.3. In order to create a project, you will need to fill and submit the following form below with information on the project title, description, and your department.

Title\*

Please fill out this field.

Description\*

We do not have information about your research. Please provide a detailed description of your work and update your field of science. Thank you!

Department\*

Center for Biostatistics & HDS

Save Cancel

Figure 2.2. A project request form.

3. Set up new allocations.
  - 3.1. Once you have created a new project, log into [ColdFront](#) using your ARC account username and password to set up new allocations.
  - 3.2. On the home page, you will see the recently created project on the project list on the left. Click the project title/name to visit the project management page.
  - 3.3. In order to authorize a student or a lab member to access your project, please click “Add

Users” as shown in Figure 3.1. Then use a VT PID username or email address to find and add this person to your user list.

The authorized users include regular users and project managers. Managers are allowed to grant user access and edit the project information.

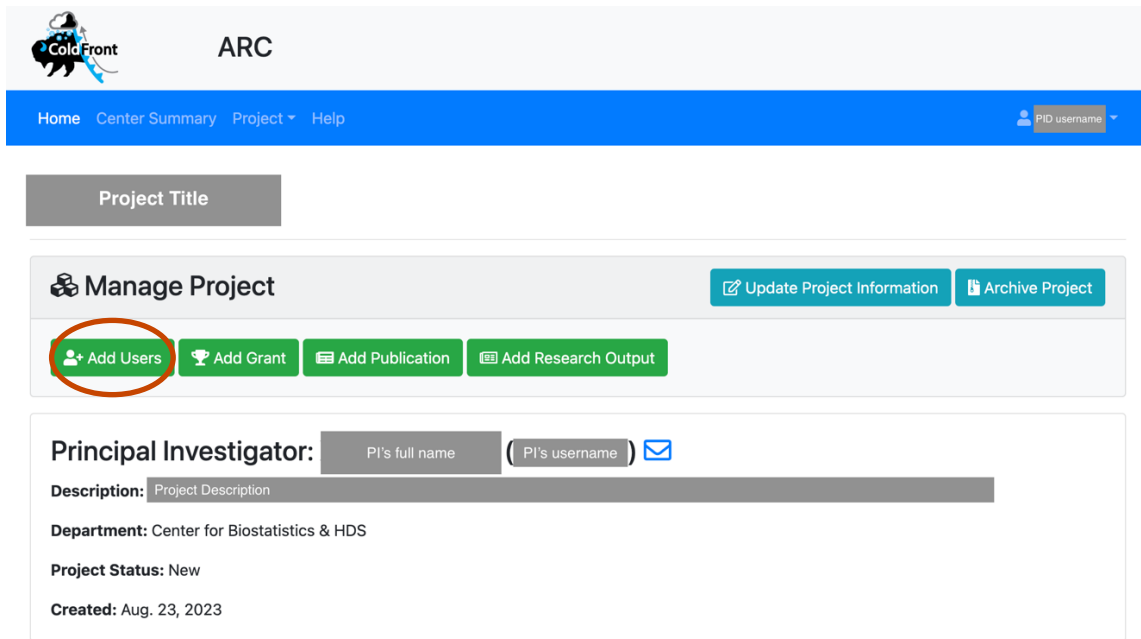


Figure 3.1. A project management page.

**3.4. To request storage and computational resources, please click the “Request Resource Allocation” as shown below in Figure 3.2 and select the allocation type.**

Then you’ll need to fill out two forms. “Storage allocation” is for requesting project storage space, and “compute allocation” is for requesting computation resources. ARC provides 0.6 million core-hour/month and 25TB storage for free to all VT PIs. Extra storage space and computation time can be granted upon request at a [cost](#).

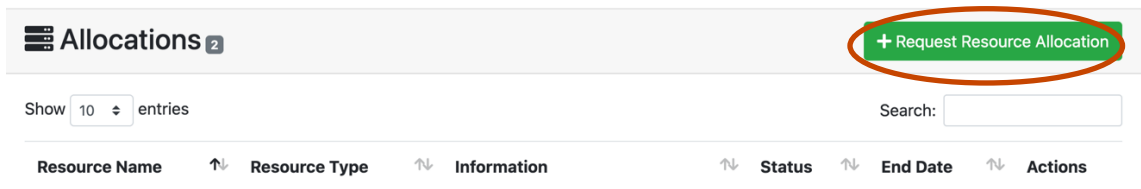


Figure 3.2. Requesting resource allocations for a project.

You will need to select the allocation type. Please revisit [Section 1](#) of this document for details on allocation types.

## Request New Allocation

Project: Bioinformatics-CBHDS

The following ARC resources are available to request for this project. If you need access to more than one of these, please submit a separate allocation request for each resource. For each request you must provide the justification for how you intend to use the resource to further the research goals of your project.

Resource\*

Compute (Free) (Cluster)

Justification\*

Tell us what your resource requirements are.

Justification for requesting this allocation.

**Core hours (up to 800k/month)**  
**(TinkerCliffs only, no monthly usage limits on other clusters)\***

800000

Users

Select All Users

Select users in your project to add to this allocation.

Select account name to associate with resource. [Click here to create an account name!](#)

Allocation type\*

research

instructional

I have uploaded the relevant publication and grant information\*

Submit

Back to Project

*Figure 3.3. Requesting computing hours for a project.*

Please click the circled line “Click here to create an account name!” to create an account name. This is your slurm account name. You and the authorized users on the project user list will need this account name for running data analysis jobs on ARC.

## Request New Allocation

Project: Bioinformatics-CBHDS

The following ARC resources are available to request for this project. If you need access to more than one of these, please submit a separate allocation request for each resource. For each request you must provide the justification for how you intend to use the resource to further the research goals of your project.

Resource\*

Justification\*

Tell us what your resource requirements are.

Justification for requesting this allocation.

Terabytes (first 25 TB free. [Click here for more details](#))\*

Users

Select All Users

[Redacted]

Select users in your project to add to this allocation.

Select account name to associate with resource. [Click here to create an account name!](#)

Allocation type\*

research

instructional

I have uploaded the relevant publication and grant information\*

Figure 3.4. Requesting storage for a project.

## 4. Manage ARC projects.

4.1. To manage ARC accounts and project directories (folders), please go to [ColdFront](#).

4.2. On the home page, you'll see a project list on the left. All projects provided to you in this list are the ARC projects that you currently have access to.

4.3. To find and review the specific information of a project, simply click the name of that project (Figure 4.1.):

### Projects »

- █
- █ CBHDS\_capstone
- █ CBHDS\_Projects <= Click to enter
- █
- █

### Allocations »

Project	Name	Resource	Status
█		Compute (Free) (Cluster)	Active
CBHDS_capstone	cbhds_capstone	Compute (Free) (Cluster)	Active
CBHDS_Projects	cbhds	Compute (Free) (Cluster)	Active
█		Compute (Free) (Cluster)	Active
█		Compute (Free) (Cluster)	Active

Figure 4.1. Reviewing your projects on the ColdFront home page.

#### 4.4. You add or remove users and request additional storage and computational resources on the project management page.

Once you enter the project management page, scroll down to the “users” and “allocations” sections to manage the user list and find the account information. The project directory path is also shown in the “allocations” section. The authorized users include regular users and project managers. Managers are allowed to grant user access and edit the project information.

### Users 3

[Email Project Users](#)
[Add Users](#)
[Remove Users](#)

Username	Name	Email	Role <span style="font-size: 0.8em;">1</span>	Status	<input type="checkbox"/> Enable Notifications <span style="font-size: 0.8em;">1</span>	Actions
█	█	█	Manager	Active	<input checked="" type="checkbox"/>	<a href="#">User</a>
█	█	█	Manager	Active	<input checked="" type="checkbox"/>	<a href="#">User</a>
█	█	█	Manager	Active	<input checked="" type="checkbox"/>	<a href="#">User</a>

### Allocations 2

[+ Request Resource Allocation](#)

Show  entries
Search:

Resource Name	Resource Type	Information	Status	End Date	Actions
Compute (Free)	Cluster	slurm_account_name: █	Active	2025-06-07	<a href="#">Info</a>
Project (Free)	Storage	/projects/cbhds-bioinfo/ (25TB) <span style="color: orange; font-size: 0.8em;">The project directory path</span>	Active		<a href="#">Info</a>

Showing 1 to 2 of 2 entries

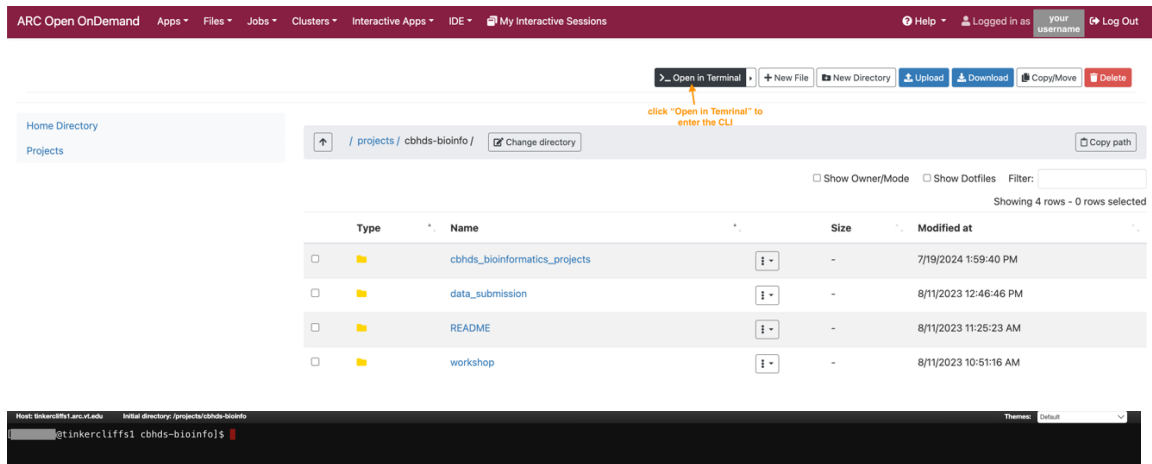
Previous
1
Next

Figure 4.2. User and allocation management on the ColdFront project management page.



**4.5. To access your project on ARC, please go to [the project web page on Open OnDemand](#).**

All the authorized users on the user list of a project have access to the corresponding project directory. Note that newly added users will acquire immediate access to a project through the terminal (Command Line Interface, CLI), but it takes longer to activate the GUI (Graphic User Interface) access.



*Figure 4.3. Entering the Command Line Interface (bottom) from the Graphic User Interface (top) of an ARC hpc cluster.*

## 5. Data Transfer and Management

If you cannot have an ARC account, please contact a CBHDS bioinformatician to submit your data in a different way.

### 5.1. Connect to VT's campus network and grant a CBHDS bioinformatician access to your project directory (following [Steps 4.1 through 4.4](#)).

### 5.2. Go to [ARC Open OnDemand](#).

Once you log in using your VT credentials (PID and password), you will see your username at the top right corner of the dashboard (home page) of [ARC Open OnDemand](#). OnDemand provides multiple features including file management and transfer, job management, shell access, and interactive apps. For data transfer, please go to "Files" and then click "Projects" (Figure 5.1.).

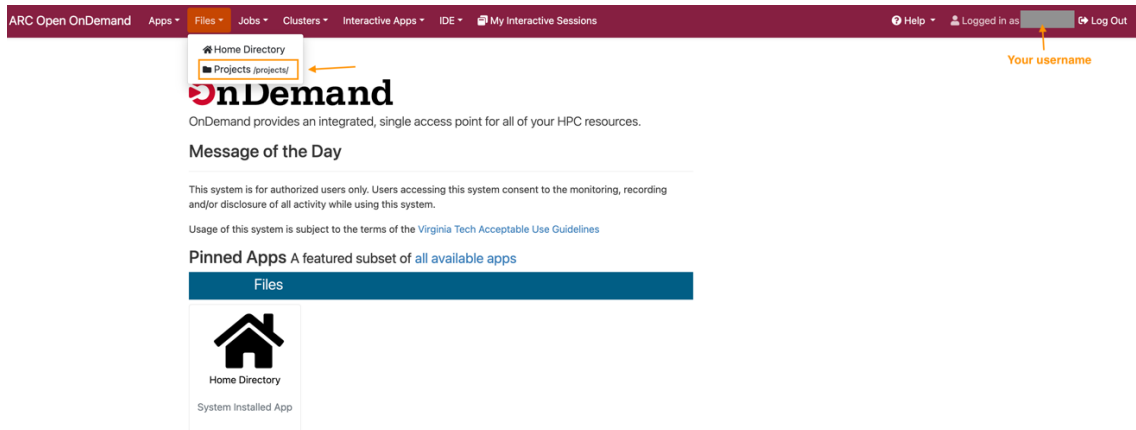


Figure 5.1. Accessing ARC projects from ARC Open OnDemand.

5.3. Once you enter [the project web page](#), you will see a list of project directories or folders (Figure 5.2.). To submit data to CBHDS bioinformatics, please go to your project directory. In this case, we will use “cbhds-bioinfo” as an example. Click the project name to enter the project directory.

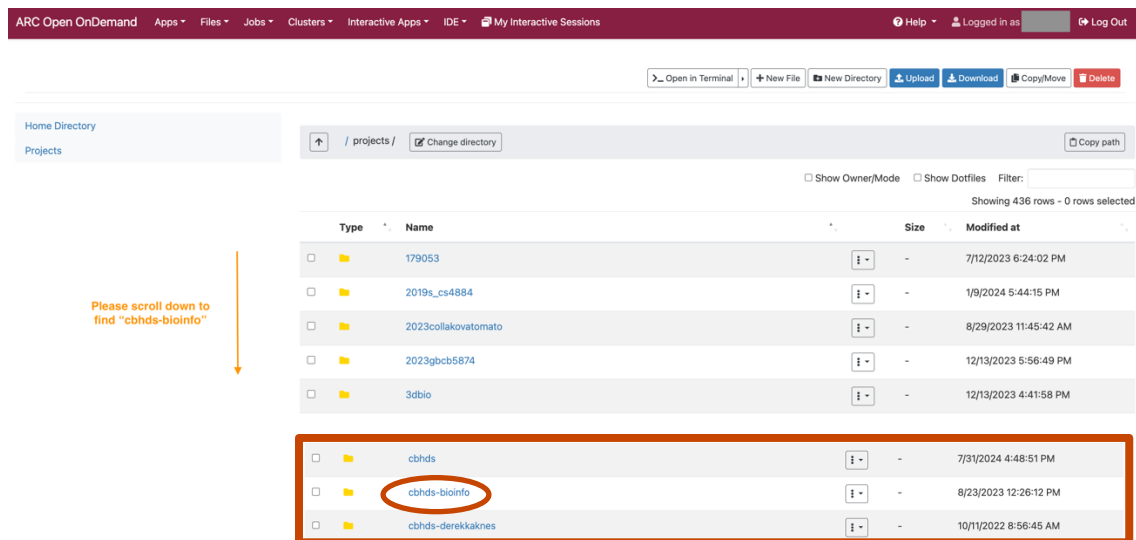


Figure 5.2. The project web page of ARC Open OnDemand.

After you enter your project directory, please click “new directory” to create a new directory and name it “data\_submission” (Figure 5.3.).

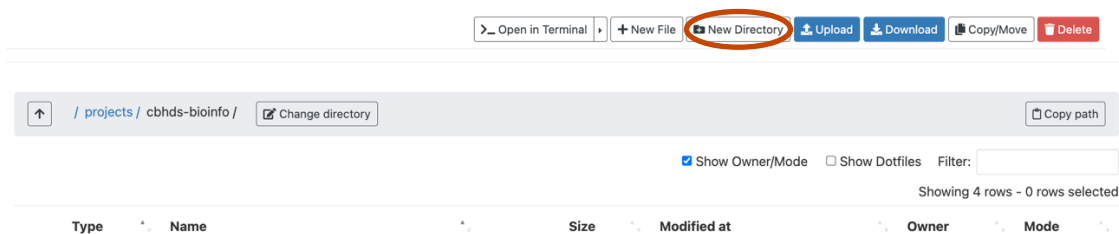


Figure 5.3. creating a new directory.

Click the name “data\_submission” to enter this directory (Figure 5.4.).

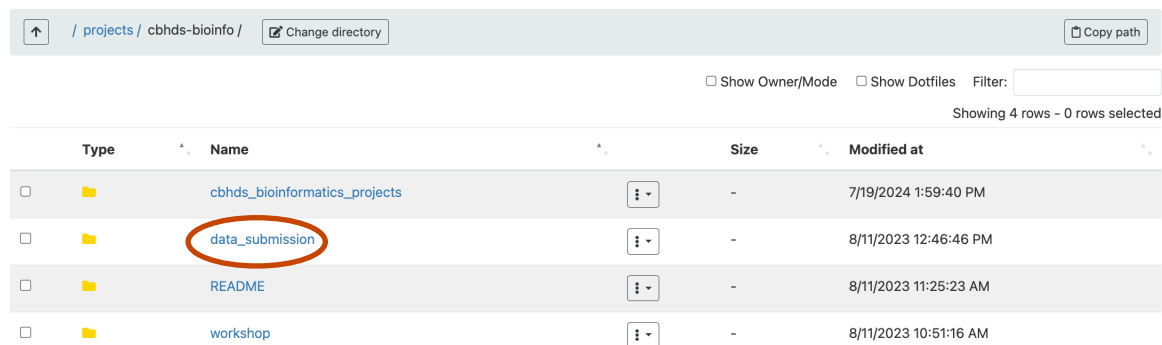


Figure 5.4. The “data\_submission” subdirectory located in our project directory, “cbhds-bioinfo”.

**5.4. Once you enter the “data\_submission” directory, you can click “new directory” to create another new directory. Please name this new directory use the format of last name and first & middle name initials, and the data type. For example: “Zhang-XM\_RNASEQ”. Next, please click “>\_ Open in Terminal” (Figure 5.5.).**

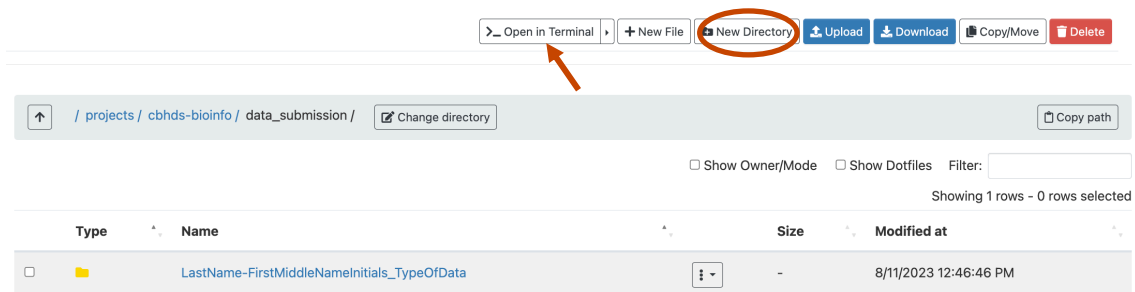
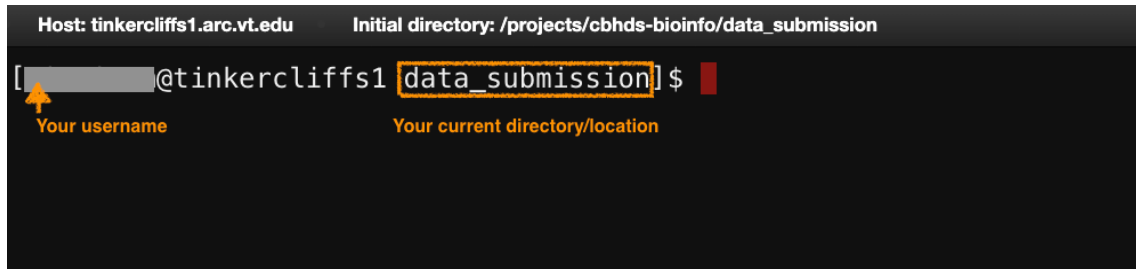


Figure 5.5. creating a raw data directory and entering the terminal interface.

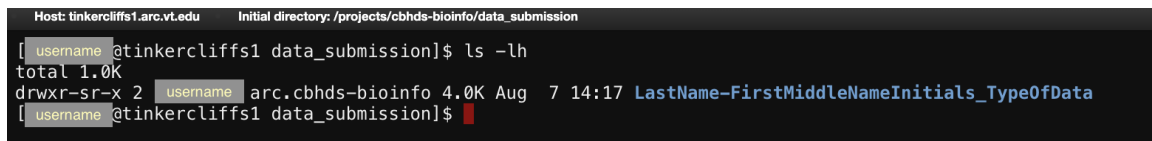
5.5. Then you will see a terminal (Command Line Interface) window similar to Figure 5.6.



```
Host: tinkerc cliffs1.arc.vt.edu Initial directory: /projects/cbhds-bioinfo/data_submission
[username@tinkerc cliffs1 data_submission]$
```

Figure 5.6. Shell access to the tinkerc cliffs1 cluster.

5.6. Type “ls (space)-lh” and hit the “enter/return” key on your keyboard, and you should see your new directory (Figure 5.7.).



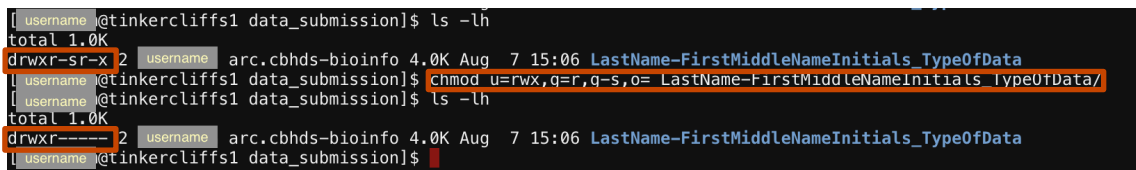
```
Host: tinkerc cliffs1.arc.vt.edu Initial directory: /projects/cbhds-bioinfo/data_submission
[username@tinkerc cliffs1 data_submission]$ ls -lh
total 1.0K
drwxr-sr-x 2 username arc.cbhds-bioinfo 4.0K Aug 7 14:17 LastName-FirstMiddleNameInitials_TypeOfData
[username@tinkerc cliffs1 data_submission]$
```

Figure 5.7. Listing the files and folders available in your current directory “data\_submission”.

5.7. Now type “chmod (space)u=rwx,g=r,g-s,o= (space)LastName-FirstMiddleNameInitials\_TypeOfData/”. Please replace FirstMiddleNameInitials\_TypeOfData with your directory name. Then hit “enter/return”.

5.8. Type “ls (space)-lh” and hit the “enter/return”.

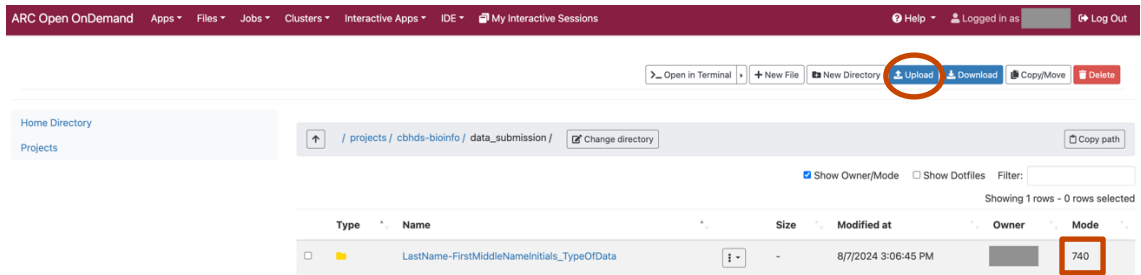
Notice the output in the first column changed to “drwxr-----” from “drwxr-sr-x”? Now, everything in this new directory is fully accessible to you and it’s readable and downloadable to the CBHDS bioinformatics team.



```
[username@tinkerc cliffs1 data_submission]$ ls -lh
total 1.0K
drwxr-sr-x 2 username arc.cbhds-bioinfo 4.0K Aug 7 15:06 LastName-FirstMiddleNameInitials_TypeOfData
[username@tinkerc cliffs1 data_submission]$ chmod u=rwx,g=r,g-s,o= LastName-FirstMiddleNameInitials_TypeOfData/
[username@tinkerc cliffs1 data_submission]$ ls -lh
total 1.0K
drwxr----- 2 username arc.cbhds-bioinfo 4.0K Aug 7 15:06 LastName-FirstMiddleNameInitials_TypeOfData
[username@tinkerc cliffs1 data_submission]$
```

Figure 5.8. Changing the accessibility of a directory.

5.9. Go back to the webpage to confirm that the mode of your directory is “740”. Now, you can click “Upload” to upload your data to the directory you just created (Figure 5.9.).



*Figure 5.9. Ready to upload the data files.*

Note that it is highly recommended to change the directory permission ([Step 5.4-5.8](#)) before uploading your data. Changing the permission to “drwxr----” (“740” permission mode) will ensure that you have the full access to your raw data while our CBHDS bioinformatics team can only read or download those files.