

# AGAVE: Accessing the power of UA-HPC through CyVerse Discovery Environment

Muhammad Noon

Bio Computing Facility  
Arizona Research Labs

# UA – HIGH PERFORMANCE COMPUTING

- ▶ UA offers research computing that includes High Performance/Throughput computing
- ▶ Resources exclusively available to the entire campus community free of charge
- ▶ Recently added new nodes and storage

# PBS SCRIPT ASSISTANT

**Selection**

Nodes  1

Memory Per Node  23G

Cores (ncpus)  12

Per Core Memory  2 GB

Job Type  Serial

Queue  Standard

```
### Your job will use 1 node, 12 cores, and 23gb of memory total.  
#PBS -q standard  
#PBS -l jobtype=serial  
#PBS -l select=1:ncpus=12:mem=23gb
```

### Script Builder

Do you want to build a script for the PBS directives? If so, fill in the next fields.  
Required fields start with a \*

Specify a name for the job \*  Alphanumeric, no spaces or special characters

Specify the group to be charged for time \*  Alphanumeric, no spaces or special characters

Does your job use less than a whole node?  yes Y = [pack:shared] N = [free]

Specify cputime in hh:mm:ss or hhh:mm:ss \*  If your job runs longer than your specified time, it will be terminated

Walltime calculates to [cputime x nodes x cores]  Do you want something more? [ hh:mm:ss]

Error details are in a separate file. To blend them in one file choose oe.  Blended

<https://hpc.toolshed.uits.arizona.edu/pbscalc/calc.php>

# CYVERSE

- ▶ NSF funded project that provides life scientists with powerful computational infrastructure to handle huge datasets and complex analyses
- ▶ Provides data storage, bioinformatics tools, image analysis, cloud services and APIs
  - ✓ Discovery Environment: offers hundreds of bioinformatics apps and manage data in a simple web interface

# CYVERSE DISCOVERY ENVIRONMENT



## Discovery Environment

The Discovery Environment integrates powerful, community-recommended software tools into a system that:

- Makes big data management easy. Upload, organize, edit, view and search with ease!
- Has 500+ scientific apps that utilize compute clusters and HPC resources as needed.
- Hides the complexity needed to do these tasks.

[Log in with your CyVerse ID](#)

[Forgot Password?](#) [Register Now](#)

Minimum screen resolution supported: 1024 x 768

©2016 CyVerse. CyVerse is funded by a grant from the National Science Foundation (#DBI-0735191, #DBI-1265383).

[The Discovery Environment Manual](#)

[Getting Started with CyVerse](#)

[The Main CyVerse Homepage](#)

[Ask CyVerse](#)

[Need Help?](#)

# CYVERSE DISCOVERY ENVIRONMENT



## CyVerse Authentication Service

Enter your Username and Password

Username:

mnoon

Password:

\*\*\*\*\*

Warn me before logging me into other sites.

[clear](#) [Register](#)

### Additional Help?

[Need to reset your password?](#)

[Contact Support](#)

[Other questions?](#)

For security reasons, please Log Out and Exit your web browser when you are done accessing services that require authentication!

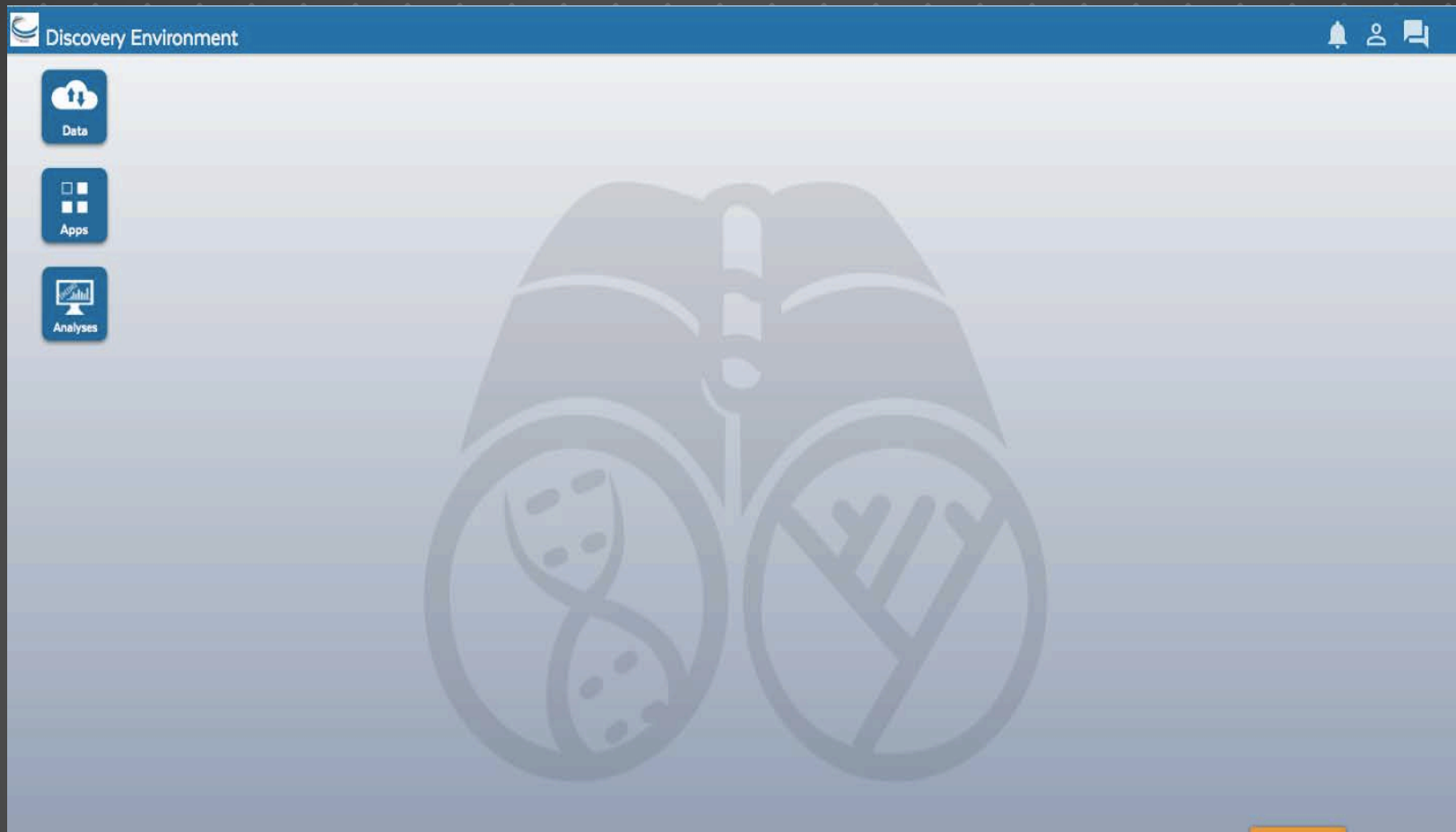
Languages:

[English](#) | [Spanish](#) | [French](#) | [Russian](#) | [Nederlands](#) | [Svenskt](#) | [Italiano](#) | [Urdu](#) | [Chinese \(Simplified\)](#) | [Deutsch](#) | [Japanese](#) | [Croatian](#) | [Czech](#) | [Slovenian](#) | [Catalan](#) | [Macedonian](#) | [Polish](#)

Copyright © 2005 - 2010 JASig, Inc. All rights reserved.  
Powered by JASig Central Authentication Service 4.0.1



# CYVERSE DISCOVERY ENVIRONMENT



# CYVERSE DISCOVERY ENVIRONMENT

The screenshot displays the Cyverse Discovery Environment interface, which is a web-based platform for managing and analyzing data. The interface is divided into several main sections:

- Navigation:** A sidebar on the left shows a tree view of the user's data space, including folders like 'Community Data', 'Shared With Me', 'Trash', and 'Favorites'. The main content area shows a file listing for the user 'mnoon' at the path '/iplant/home/mnoon'. The listing includes columns for Name, Last Modified, and Size, with various file types such as folders, scripts (e.g., 'batch\_1\_all.phylip'), and FASTA files (e.g., 'yeastgenes.fa').
- Apps:** A panel on the right titled 'Apps' shows search results for 'blast'. It lists 16 applications, including '16sblaster', 'Add GO to Blastp-u...', 'Blast2Seqs', 'Blastn-2.2.26', 'Blastp a subset of u...', 'Blastp-2.2.29+', 'Bulk Blast 2 Sequen...', 'Create BLAST datab...', and 'DeltaBLAST-2.2.29+'. Each entry includes the name, the person who integrated it, and a star rating.
- Analyses:** A panel at the bottom right shows a list of analyses. The selected analysis is 'SRA-Import-0.1.0\_analys...'. The table below shows details for various analyses, including their names, owners, the apps used, start and end dates, and their status (Completed or Failed).

Name	Owner	App	Start Date	End Date	Status
NCBI_SRA_Toolkit_fastq...	mnoon@iplantcollab...	NCBI SRA Tool...	2016 Mar 21 14:36:20	2016 Mar 21 18:24:18	Completed
NCBI_SRA_Toolkit_fastq...	mnoon@iplantcollab...	NCBI SRA Tool...	2016 Mar 21 10:43:19	2016 Mar 21 14:04:37	Completed
<input checked="" type="checkbox"/> SRA-Import-0.1.0_analys...	mnoon@iplantcollab...	SRA-Import-0.1.0	2016 Mar 18 10:03:16	2016 Mar 18 19:14:34	Completed
NCBI_SRA_Import_1.2_a...	mnoon@iplantcollab...	NCBI SRA Imp...	2016 Mar 17 16:36:30	2016 Mar 17 22:48:37	Failed
DNAPARS_pyrad	mnoon@iplantcollab...	DNAPARS	2015 Aug 3 10:05:34	2015 Aug 3 11:34:45	Completed
DNAPARS_all_pop_phy	mnoon@iplantcollab...	DNAPARS	2015 Jul 30 13:41:37	2015 Aug 4 12:18:56	Failed
RAxML-7.3.0_pyrad	mnoon@iplantcollab...	RAxML-7.3.0	2015 Jul 29 16:28:56	2015 Jul 29 17:25:45	Completed
DNAPARS_pop_kwn_copy	mnoon@iplantcollab...	DNAPARS	2015 Jul 29 15:30:44	2015 Jul 29 15:32:19	Completed
DNAPARS_pop_kwn	mnoon@iplantcollab...	DNAPARS	2015 Jul 29 15:14:15	2015 Jul 29 15:17:00	Completed
DNAPARS_pyrad_phy	mnoon@iplantcollab...	DNAPARS	2015 Jul 29 15:01:53	2015 Jul 29 16:25:50	Failed



# CYVERSE

- **Graphical User Interface**
- **Max memory = 1TB**
- **Max no. of jobs = 1-5**
- **Max wall time = 48 hours**

# UA-HPC

- **Command-line job submission**
- **Ocelote has one node with 2TB of memory**
- **Max no. of jobs = 500**
- **Max wall time = 240 hours**
- **Standard queue allocation = 24k hours compute time/month**

- **Jobs such as BLAST alignments of whole genome can't be run on CyVerse Discovery Environment**
- **People are uncomfortable with command-line to utilize many tools**

# AGAVE-API (Application Programming Interface)

- ▶ AGAVE-API bridges the gap between the HPC and web worlds by supporting interfaces
- ▶ It offers to combine graphical user interface of CyVerse Discovery Environment with UA-HPC resources
- ▶ We can develop and add an app to Discovery Environment that utilizes UA-HPC for computation on back-end
- ▶ Benefits:
  - ▶ More control of memory/wall time that UA-HPC offers
  - ▶ Attract more users who are reluctant to use command-line tools

# SETTING UP AGAVE-API

1. Install CyVerse Software Development Kit (SDK)
2. Create a client and set of Api keys (local system communicates with CyVerse API)
3. Obtain authentication token for the access (expires after 4 hours)
4. Set up CyVerse/UA-HPC system
  - ▶ System.json file that has credentials for login, ssh-keys, information about available physical and storage resources, computing environment, where to move data and run computing tasks)

# SETTING UP AGAVE-API ( CONTD.)

## 5. Create CyVerse application for UA-HPC

- Create a test wrapper script to test-run your tool on UA-HPC system with required memory and wall time
- Copy test wrapper script, with binary of the tool, libraries and input files and put them in the CyVerse Data store
- Write Agave-api wrapper script
- Write App.json file that defines app inputs, output and parameters options it would offer to the user

# SUMMARY

▶ We create four files:

1. System.json
2. App.json
3. Wrapper\_test.sh
4. Wrapper.sh

➤ We can use Principal Investigator's credentials for setting up system.json and whole group can use the application

# SYSTEM.JSON

```
1 {
2   "id": "mnoon-uahpc",
3   "name": "MuhammadNoonUAHPC-Ocelote",
4   "status": "UP",
5   "type": "EXECUTION",
6   "description": "2016 UA HPC",
7   "site": "hpc.arizona.edu",
8   "executionType": "HPC",
9   "default": true,
10  "scratchDir": "/gsfs1/xdisk/mnoon",
11  "queues": [
12    {
13      "name": "standard",
14      "maxJobs": 15,
15      "maxNodes": 12,
16      "maxProcessorsPerNode": 28,
17      "maxRequestedTime": "60:00:00",
18      "defaultMaxRunTime": "59:59:00",
19      "maxMemoryPerNode": "168GB",
20      "customDirectives": "#PBS -l select=1:ncpus=8:mem=48gb\n#PBS -l place=free:shared\n#PBS -l pvmem=100gb\n#PBS",
21      "default": true
22    },
23    {
24      "name": "smp",
25      "maxJobs": 30,
26      "maxNodes": 1,
27      "maxProcessorsPerNode": 16,
28      "maxRequestedTime": "48:00:00",
29      "maxMemoryPerNode": "256GB",
30      "customDirectives": "-A iPlant-Collabs",
31      "default": true
32    }
33  ],
34  "login": {
35    "host": "login.ocelote.hpc.arizona.edu",
36    "port": 22,
37    "protocol": "SSH",
38    "rootDir": "/",
39    "homeDir": "/gsfs2/home/u8/mnoon",
40    "auth": {
41      "username": "mnoon",
42      "publicKey": "ssh-rsa AAAAB3NzaC1yc2EAAAABIwAAAQEA2k\AeD6mV\jM3JgInKe\k5eQ\jXwMG9NEp9y0Tg+\6PUhgMqXkLrk",
43      "privateKey": "-----BEGIN RSA PRIVATE KEY-----\n*****",
44      "type": "SSHKEYS",
45      "default": true
46    },
47    "proxy": {
48      "name": "Bastion host",
49      "host": "hpc.arizona.edu",
```

# APP.JSON

```
1 {
2   "name": "UAHPC-NCBIblast",
3   "parallelism": "PTHREAD",
4   "version": "2.4",
5   "revision": 1,
6   "helpURI": "http://blast.ncbi.nlm.nih.gov/Blast.cgi",
7   "label": "NCBIblast",
8   "shortDescription": "NCBI blast sequence similarity search run on UA HPC",
9   "longDescription": "Run a similarty search program form the NCBI blast suite on UA HPC",
10  "datePublished": "08/18/16",
11  "tags": [
12    "blast",
13    "NCBI"
14  ],
15  "ontology": [
16    "http://sswapmeet.sswap.info/sequenceServices/SequenceServices"
17  ],
18  "defaultMaxRunTime": "01:00:00",
19  "defaultQueue": "standard",
20  "defaultNodeCount": 1,
21  "defaultMemoryPerNode": 28,
22  "defaultProcessorsPerNode": 14,
23  "executionSystem": "sjmiller-uahpco",
24  "executionType": "HPC",
25  "deploymentPath": "sjmiller/apps/blast/2.4/UAHPC",
26  "templatePath": "wrapper.sh",
27  "testPath": "test/test.sh",
28  "checkpointable": false,
29  "modules": [
30    "purge",
31    "load blast"
32  ],

```

# APP.JSON

```
33 ▼ "inputs": [  
34 ▼   {  
35     "id": "queryseq_file",  
36     "value": {  
37       "default": "",  
38       "validator": "",  
39       "visible": true,  
40       "required": false,  
41       "order": 1  
42     },  
43     "details": {  
44       "label": "fasta query file",  
45       "description": "Sequence fasta query file"  
46     },  
47     "semantics": {  
48       "ontology": [  
49         "http://sswapmeet.sswap.info/sequence/FASTA"  
50       ],  
51       "minCardinality": 1,  
52       "fileTypes": [  
53         "fasta-0"  
54       ]  
55     }  
56   }  
57 ],
```



# APP.JSON

```
58  ▼ "parameters": [  
59  │   {  
60  │     "id": "evaluate",  
61  │     "value": {  
62  │       "default": "1e-10",  
63  │       "required": false,  
64  │       "visible": true,  
65  │       "type": "string"  
66  │     },  
67  │     "details": {  
68  │       "label": "evaluate",  
69  │       "description": "Expectation threshold for retaining hits",  
70  │       "argument": "-evaluate ",  
71  │       "showArgument": true,  
72  │       "visible": true,  
73  │       "order": 2  
74  │     },  
75  │     "semantics": {  
76  │       "ontology": [  
77  │         "xs:string"  
78  │       ]  
79  │     },  
80  │   },  
81  │   {  
82  │     "id": "num_descrip",  
83  │     "value": {  
84  │       "default": 20,  
85  │       "type": "number",  
86  │       "required": false,  
87  │       "visible": true,  
88  │       "order": 3  
89  │     },  
90  │     "details": {  
91  │       "label": "num_descrip",  
92  │       "description": "Number of hit descriptions to display",  
93  │       "argument": "-num_descriptions ",  
94  │       "showArgument": true,  
95  │       "visible": true  
96  │     },  
97  │     "semantics": {  
98  │       "ontology": [  
99  │         "xs:number"  
100  │       ]  
101  │     },  
102  │   },  
]
```

# TEST.SH

```
1  #!/bin/bash
2  #PBS -N oGMblastp
3  #PBS -m bea
4  #PBS -M mnoon@email.arizona.edu
5  #PBS -W group_list=nirav
6  #PBS -q standard
7  #PBS -l select=1:ncpus=28:mem=188gb:pcmem=6gb
8  #PBS -l cput=28:0:0
9  #PBS -l walltime=1:0:0
10 #PBS -l pvmem=23Gb
11
12 module load blast
13
14 cd $PBS_O_WORKDIR
15 blast_db="/genome/nr"
16 queryseq_file="GallusMitoProt100.fasta"
17 num_descrip=20
18 num_align=20
19 time blastp -db ${blast_db} -query ${queryseq_file} -out ${queryseq_file}.blp -evalue 1e-30 -num_descriptions ${num_descrip} -num_al
```

# SUPPORT

- ▶ Active Agave-api support via Slack channel
- ▶ UA-HPC support
- ▶ CyVerse support via Slack channel
- ▶ UA Genome Analytics Services group provides consulting services and can help develop an application for your group for a minimal charge

<http://genomeanalytics.arizona.edu/>