Solr Notes

NOTE: To access the servers, you will need a public and private key. Follow the steps under the section "Creating a Key Pair Using Amazon EC2" in this guide created by Amazon: https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-key-pairs.html#having-ec2-create-your-key-pair

- For Windows users who use PuTTY, you will need to create a private key in order to connect to the server. To create the private key using PuTTYgen, follow the steps in this guide: https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/putty.html
- To access the servers on other operating systems or through other means, you can view the guides Amazon provides here: https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/AccessingInstances.html
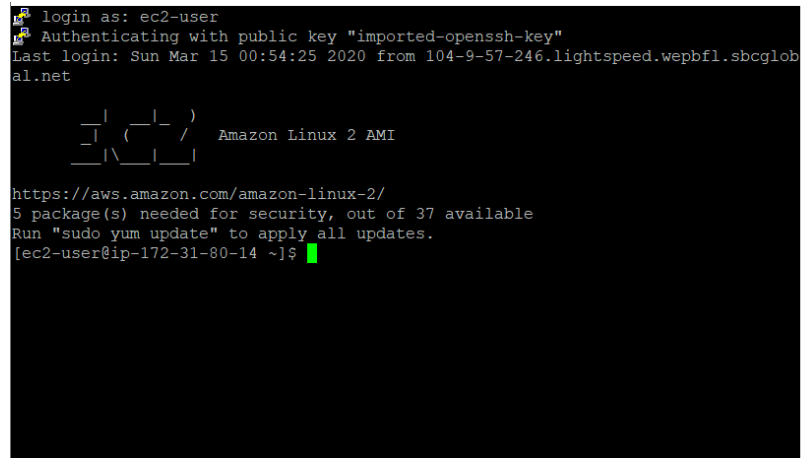
Currently installed:

- Solr v8.4.0
- Java Database Connector v.8.0.19
- Java OpenJDK v1.8.0_222
- Python2 v2.7.16
- Python3 v3.7.6

1) Solr Notes

- How to start Solr
  - First log into the server using the login name provided to you. If no name login name was provided for you, use "ec2-user".
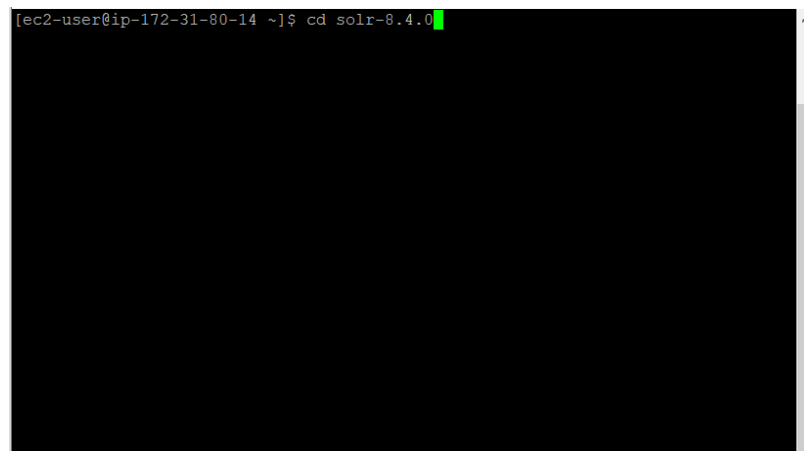
  ```
  login as: ec2-user
  Authenticating with public key "imported-openssh-key"
  Last login: Sun Mar 15 00:54:25 2020 from 104-9-57-246.lightspeed.wepbfl.sbcglob
  al.net

         __|  __|_  )
         _|  (     /   Amazon Linux 2 AMI
        ___|\___|___|

  https://aws.amazon.com/amazon-linux-2/
  5 package(s) needed for security, out of 37 available
  Run "sudo yum update" to apply all updates.
  [ec2-user@ip-172-31-80-14 ~]$
  ```

  - Enter the command "cd solr-8.4.0" to enter the solr directory.

  ```
  [ec2-user@ip-172-31-80-14 ~]$ cd solr-8.4.0
  ```

  - While logged into the "solr1" server enter the command:
    bin/solr -c -p 8983 -h <AWS Public DNS 1>
    where the AWS Public DNS is provided in the EC2 Management Console. The Public DNS used in the screenshot

below (ec2-35-171-20-243.compute-1.amazonaws.com) will be different from the current one as the server instance will have been stopped at one point since this document has been written.

- -c command means that Solr is entering Cloud mode which is needed since we are using Zookeeper to connect our two servers together. Zookeeper ensures that if any configuration files of one server is changed, those same files will be updated/transferred/deleted on the other server.
- -p command is the port that the Solr server will be running on. The first Solr instance is usually set to run on port 8983.
- -h is the host name of the Solr instance.


```
[ec2-user@ip-172-31-80-14 ~]$ cd solr-8.4.0
[ec2-user@ip-172-31-80-14 solr-8.4.0]$ bin/solr -c -p 8983 -h ec2-35-171-20-243.
compute-1.amazonaws.com
```

- o The command prompt should look like this with a message stating that the Solr server has been started.

```
[ec2-user@ip-172-31-80-14 ~]$ cd solr-8.4.0
[ec2-user@ip-172-31-80-14 solr-8.4.0]$ bin/solr -c -p 8983 -h ec2-35-171-20-243.
compute-1.amazonaws.com
*** [WARN] *** Your open file limit is currently 1024.
 It should be set to 65000 to avoid operational disruption.
 If you no longer wish to see this warning, set SOLR_ULIMIT_CHECKS to false in y
our profile or solr.in.sh
*** [WARN] ***  Your Max Processes Limit is currently 3792.
 It should be set to 65000 to avoid operational disruption.
 If you no longer wish to see this warning, set SOLR_ULIMIT_CHECKS to false in y
our profile or solr.in.sh
Waiting up to 180 seconds to see Solr running on port 8983 [-]
Started Solr server on port 8983 (pid=3557). Happy searching!

[ec2-user@ip-172-31-80-14 solr-8.4.0]$ █
```

o   To access the Solr Admin UI enter the following link into your
    search bar:

    <AWS Public DNS 1>:8983/solr

    The link that will be used based off the screenshots will be:

    http://ec2-35-171-20-243.compute-1.amazonaws.com:8983/solr

o Now we will start up the second Solr instance. Repeat steps 1 and 2 that we did for the first Solr instance. You should be in the "solr-8.4.0" directory.

o Enter the command:

bin/solr start -c -p 8983 -h <AWS Public DNS 2> -z <AWS Public DNS 1>:9983

  ▪ <AWS Public DNS 2> will be the Public DNS for the "solr2" server. In the screenshot below, this is

    ec2-3-95-237-89.compute-1.amazonaws.com

  ▪ <AWS Public DNS 1> will be the Public DNS for the "solr1" server, the one we used in steps 3-5 above. In the screenshot below, this is

    ec2-35-171-20-243.compute-1.amazonaws.com

  ▪ -z command creates a Zookeeper instance on the "solr1" server which connects the two Solr instances together. It often opens on port 9983.

```
[ec2-user@ip-172-31-90-67 solr-8.4.0]$ bin/solr start -c -p 8983 -h ec2-3-95-237-89.
compute-1.amazonaws.com -z ec2-35-171-20-243.compute-1.amazonaws.com:9983
```
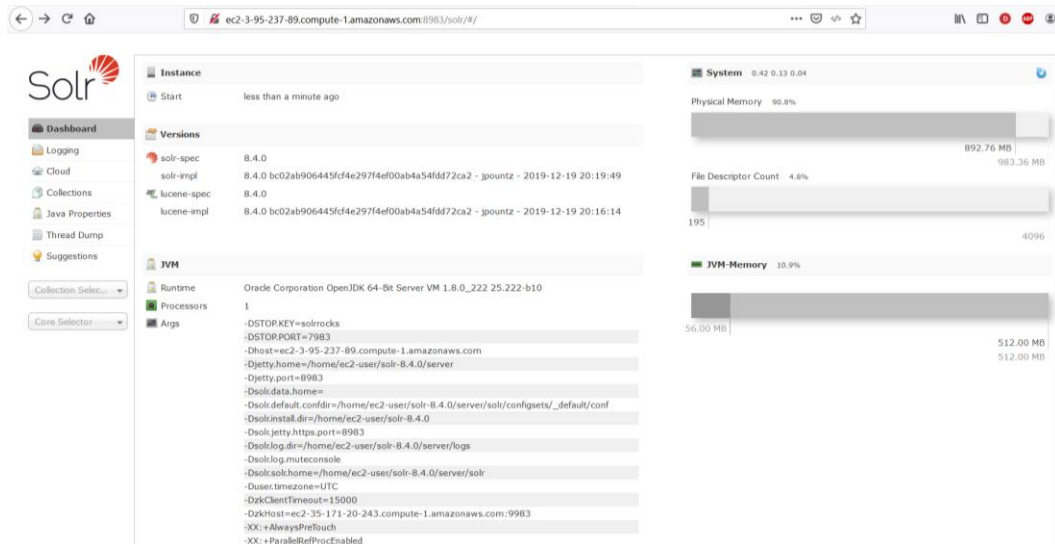
o You should now see the same success message that is seen in step 4.

```
[ec2-user@ip-172-31-90-67 solr-8.4.0]$ bin/solr start -c -p 8983 -h ec2-3-95-237-89.
compute-1.amazonaws.com -z ec2-35-171-20-243.compute-1.amazonaws.com:9983
*** [WARN] *** Your open file limit is currently 1024.
 It should be set to 65000 to avoid operational disruption.
 If you no longer wish to see this warning, set SOLR_ULIMIT_CHECKS to false in your
profile or solr.in.sh
*** [WARN] ***  Your Max Processes Limit is currently 3792.
 It should be set to 65000 to avoid operational disruption.
 If you no longer wish to see this warning, set SOLR_ULIMIT_CHECKS to false in your
profile or solr.in.sh
Waiting up to 180 seconds to see Solr running on port 8983 [\]
Started Solr server on port 8983 (pid=5235). Happy searching!
                                                                      [ec2-u
ser@ip-172-31-90-67 solr-8.4.0]$
```

o To access the Solr Admin UI for the "solr2" server, enter the following link in your search bar:

<AWS Public DNS 2>:8983/solr

The link that will be used based off the screenshots will be:
http://ec2-3-95-237-89.compute-1.amazonaws.com:8983/solr

- o To stop the running Solr instances, enter the command:

  bin/solr stop -all

- Retrieving Zookeeper config files
  - o Zookeeper configuration files are not accessible if you were to look at the server files in Filezilla or another application. The problem with these files being hidden is that Solr must know what columns are in your database so that when it processes all the documents, it stores all the data properly in Solr's index. If the configuration files are not setup accordingly, Solr will not be able to process the data from your database or perform queries.
  - o To access the configuration files run the following commands on the Solr server that is running Zookeeper. It has been standard for the "solr1" server to be the Zookeeper host throughout the course of this project. If you follow the steps
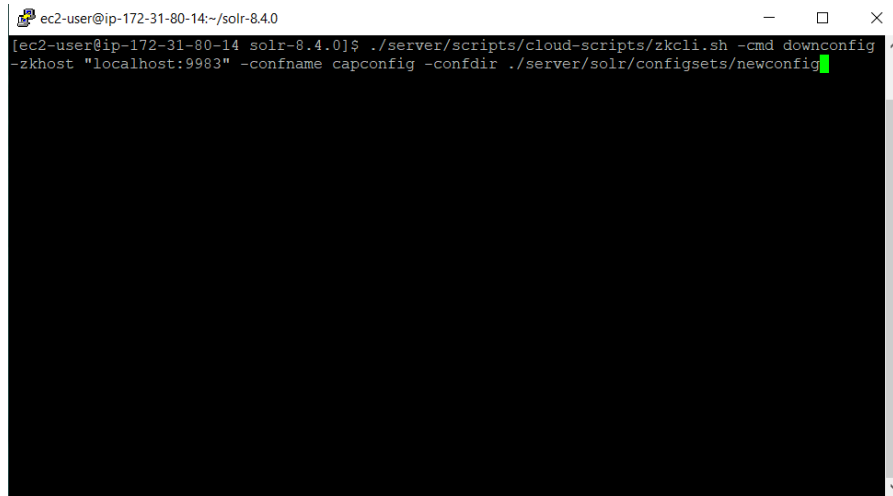
above on "How to start Solr", the "solr1" server will be the
Zookeeper host.

1) Login into the "solr1" server.
2) Enter the command "cd solr-8.4.0".
3) Enter the command:

./server/scripts/cloud-scripts/zkcli.sh -cmd downconfig
-zkhost "localhost:9983" -confname capconfig -confdir
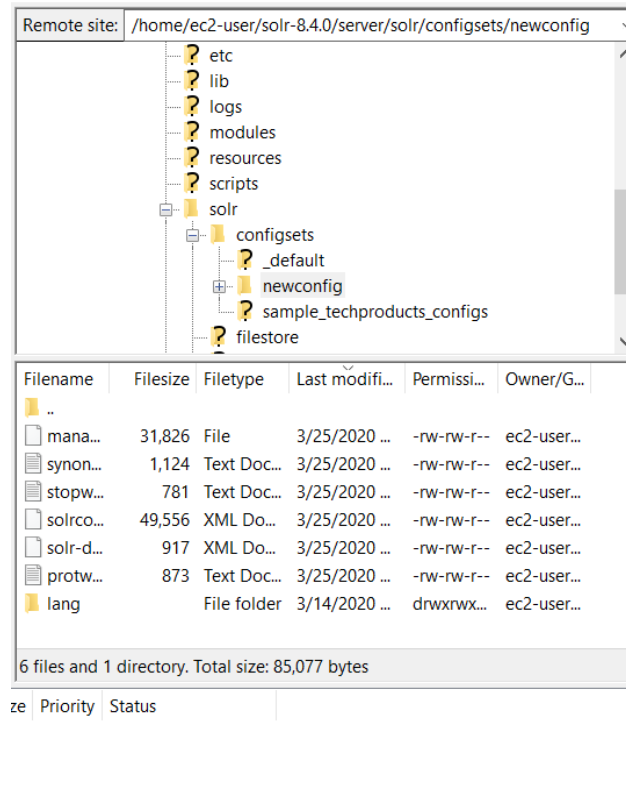./server/solr/configsets/newconfig

- The zkcli.sh is the script that allows access to the
  configuration files of Zookeeper.
- -cmd downconfig is the Zookeeper command we
  want to run. In this case we are downloading Solr
  configuration files.
- -zkhost "localhost:9983" is where Zookeeper is
  running. Since we are logged in to "solr1" which is
  the Zookeeper host, we use "localhost" instead of
  the DNS. 9983 is the port that Zookeeper is set to
  run on.
- -confname capconfig specifies which set of
  configuration files we are to download. The
  configuration name for this project is capconfig.
- -confdir ./server/solr/configsets/newconfig is the
  location where the Zookeeper configuration files
  will be downloaded to. The "newconfig" folder will

dually serve as the destination where the config files will be downloaded to and the location where we will upload the new config files if they need to be updated.

4) Once the config files are in the newconfig folder, you can download them off Filezilla and modify them as need be. The only files that you may want to modify are managed-schema.xml, solr-data-config.xml, and solrconfig.xml.

5) Once you have updated the config files, upload them back into the same directory that you downloaded them from (pictured above).

6) Now that the new files have been uploaded on to the server, enter the command bin/solr zk upconfig -n capconfig -d ./server/solr/configsets/newconfig

   ▪ zk upconfig is the Zookeeper command we want to run. We are uploading config files to Zookeeper.

   ▪ -n capconfig is the name of the configuration that we are uploading to. Capconfig is the configuration name for the collection we have created on Solr.

- -d ./server/solr/configsets/newconfig is the directory that we are uploading from.

```
[ec2-user@ip-172-31-80-14 ~]$ bin/solr zk upconfig -n capconfig -d ./server/solr
/configsets/newconfig
```

7) Once you entered the command, the terminal should display all the files that have now been uploaded to Zookeeper. Both Solr instances will now have the same configuration files.