BlueData EPIC for AWS Quick Start Guide

bluedata.zendesk.com/hc/en-us/articles/228218068-BlueData-EPIC-for-AWS-Quick-Start-Guide

BlueData Support
Updated 1 year ago

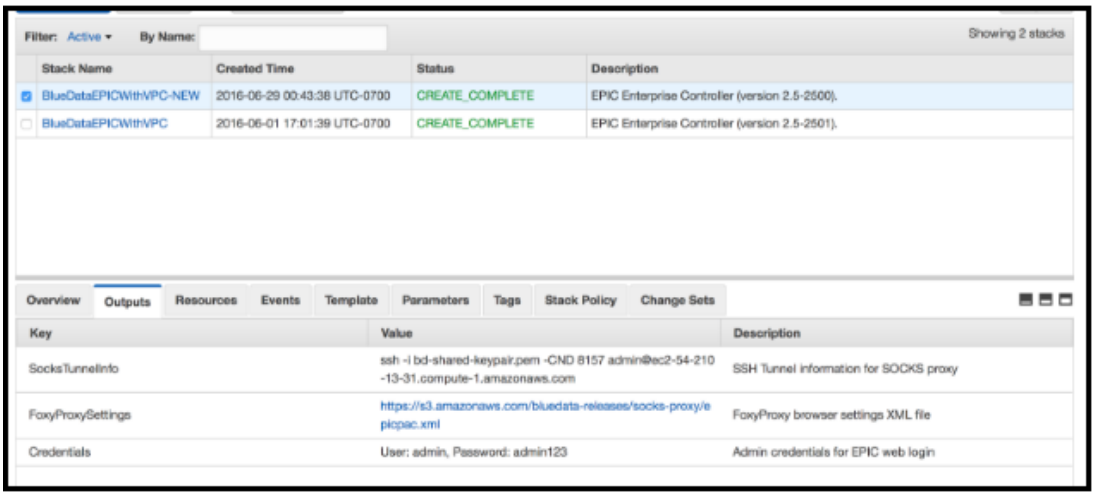
Follow

SITE ADMINISTRATOR

1. Login to AWS. Best practice is to not use the root user account.
2. Click on Security Credentials and create a few IAM users. These could be mapped to the individual tenants in BlueData.
    - These users must have full EC2 access policy.
    - Note down the security access key for these users.
    - These access keys will be required post-launch of BlueData UI to configure tenants.
3. Under EC2, set up a key pair. This will be used to launch the CloudFormation stack outlined in the next step.
4. SETUP: Please view the detailed video walkthrough here.
    - Click on the following BlueData-provided URL.
        - **EXAMPLE URL. DO NOT USE.** https://console.aws.amazon.com/cloudformation/home?region=us-east-1#/stacks/new?stackName=EPIC-2.5-2501&templateURL=https:%2F%2Fs3.amazonaws.com%2Fbluedata-releases%2Fcfn-templates%2Fepic-cfn-byol-2.5-2501.json
        - NOTE: New URLs will be provided for new versions during DA.
    - Follow the wizard. You will be prompted to choose the right key pair. Select a key pair that you already have, otherwise, you will not be able to access the instance and/or SSH into the instance.
    - Stack will be setup

      NOTE: AWS sometimes takes you to the "Create New Stack" screen. Make sure to go back on EC2 and click on "EC2 Management Console" (you should see an EC2 instance getting started).
    - Go to EC2 Management Console and confirm that there is an EC2 instance running with your stack name.
    - Note the public IP address because we will need to run an SSH proxy to this instance from end user's local machine.
5. Click on the "Outputs" tab of the created stack (click on CloudFormation in the AWS home screen that lists all services).

Download the FoxyProxy Setting (https://s3.amazonaws.com/bluedata-releases/socks-proxy/epicpac.xml).



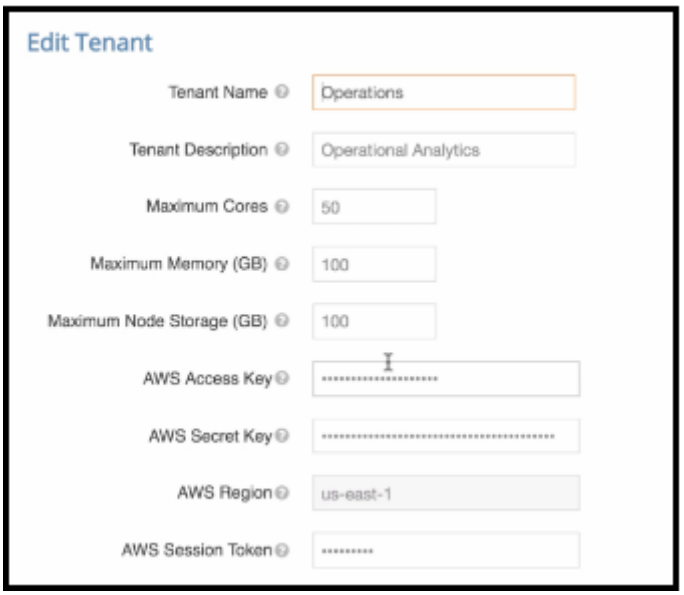
6. Configure FireFox or Chrome with FoxyProxy.
    - Click on "Add-Ons."
    - Search for FoxyProxy.
    - Install Add-On.
    - FireFox
        - Click on "Add New Proxy."
        - On the "General" tab - input a name (e.g. epic-socks-proxy)
        - Click the radio button titled "Manual Proxy Configuration."
            - Input host = localhost
            - Port = 8157
        - On the "URL Patterns" tab, import the epicpac.xml file.
        - Enable the proxy (you can disable it from FireFox when not required).

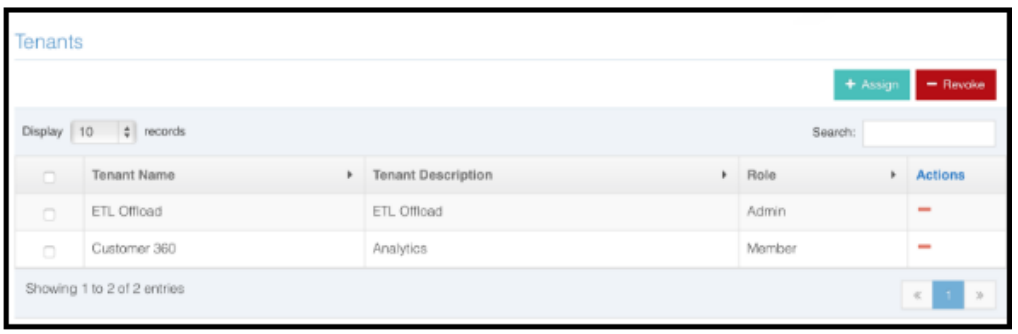
7. Run SSH proxy to the BlueData EC2 Controller instance.
    - **Administrators**

      Open a terminal window and run the command:
      ssh -i <AWS Key Pair> -CND 8157 admin@e<EC2 instance address of EPIC Controller>

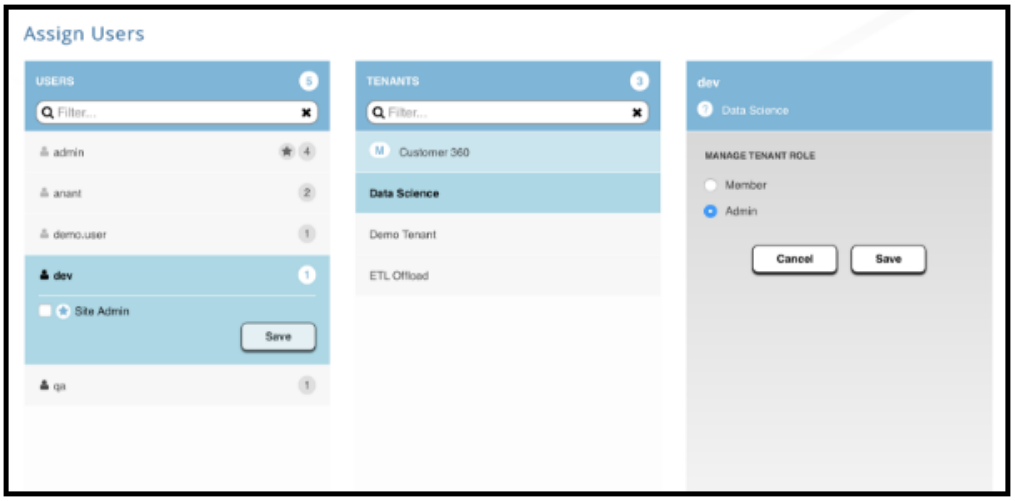
    - [IMPORTANT] First Time Login, App Store Image Install and Tenant Setup
        1. After setting up the SSH tunnel from the local machine and the FoxyProxy, access the BlueData Web UI.
        2. Open the browser (Chrome/FireFox) where FoxyProxy is setup/enabled and access the URL: http://localhost.
        3. Login with default credentials: **username:** admin and **password:** admin123

4. You will arrive at the App Store page by default and may install the images of your choice (e.g. Spark and CDH 5.4.3 are supported during DA).
5. Click on Tenants link of the left navigation bar.
    - Create a new Tenant and specify quotas for CPU, memory, and storage.
    - BlueData recommends 100 cores, 100GB of memory, and 600GB of storage in order to create a 5 worker cluster of a specific platform (e.g. Spark or Hadoop).
    - BlueData will limit the number of EC2 instances that can be launched to the quota set above.
    - You must also provide the Access Key and Secret Key (as shown in the screenshot below). These keys can be obtained from the "Security and Credentials" section (a dropdown under your login name in the AWS Web UI). **The AWS Session Token is optional.**
6. Click on "Users" link on the left navigation bar and create a few users.
7. Post user creation, click on the "Assign" link to attach users to one or more Tenants (as shown in the screenshot below).





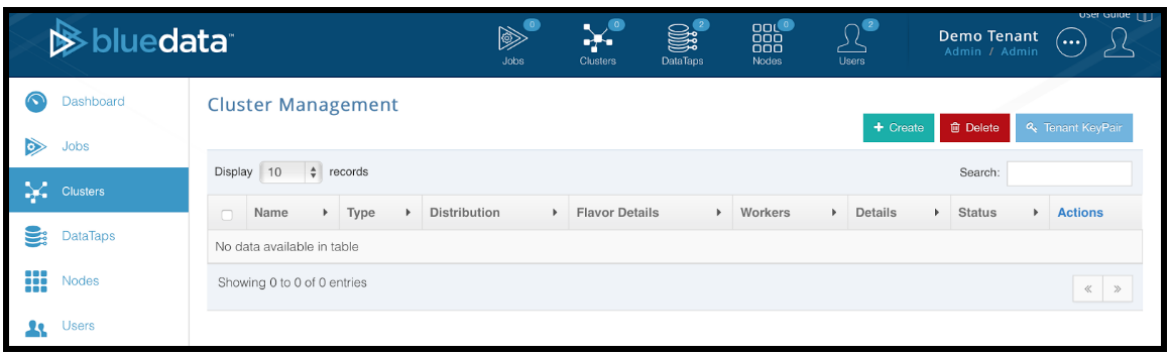
Select User Name on the left panel, select Tenant Name in the middle panel, and select type of Role in the right panel (e.g. Tenant Member or Tenant Admin).



TENANT USERS (i.e. end users who create clusters)

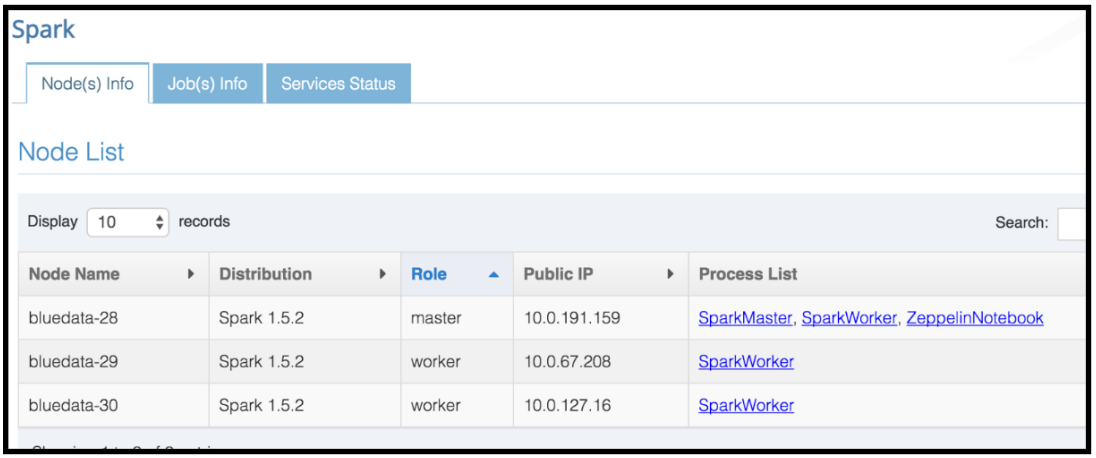
1. Open a terminal window and run the command:
   ssh-i <TenantKeyPair> -N _C -D 8157 tenant@$CONTROLLER_PUBLIC_DNS

2. Open the browser (Chrome/FireFox) where FoxyProxy is setup/enabled and access the URL: http://localhost.
3. NOTE: TenantKeyPair needs to be downloaded from the BlueData UI on the Cluster Summary page. See screenshot below (top right hand corner).

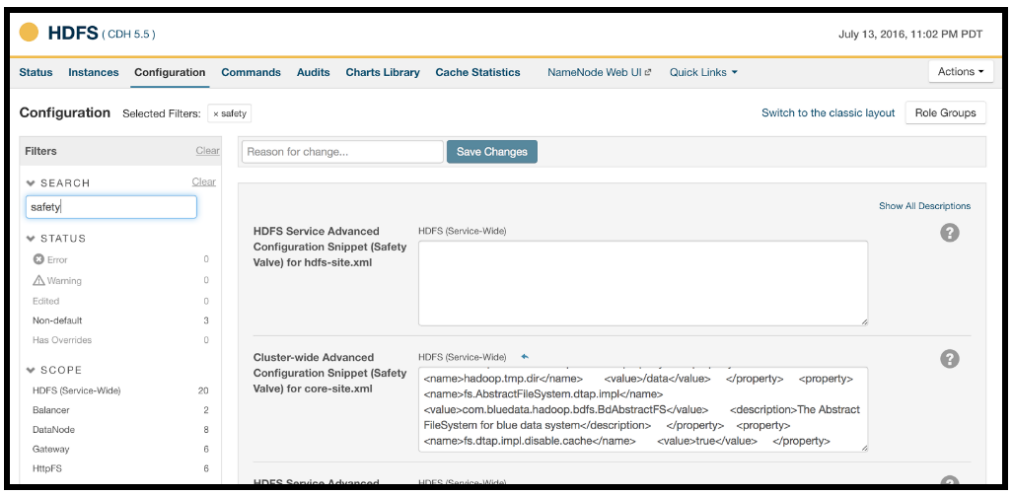


   - Download the key pair.
   - Change permissions by running chmod 400 <TenantKeyPairName>.pem.
4. Log into virtual nodes of clusters (e.g. Spark, CDH, etc.).
   - Download the Tenant key pair from the BlueData Tenant User Interface (as described above).
   - Run chmod 400 <TenantKeyPair>.pem
   - ssh -o ProxyCommand='nc -x localhost:8157 %h %p' -i <TenantKeyPair> bluedata@<ip address of the virtual node from the BlueData UI>
     - EXAMPLE SCREENSHOT AND COMMAND EXECUTED. Note the IP address of the virtual node = 10.0.191.159.
     - Run command:
       ssh -o ProxyCommand='nc -x localhost:8157 %h %p' -i BD_Ops.pem bluedata@10.0.191.159

   - Now you can run spark-shell and other command line utilities in the virtual node/container.



5. Setting up connectivity to S3 will be automatically done in DA v2.
   - **Cloudera/CDH Cluster**
     1. After cluster is created, click "Cluster Details" to view NameNodes, IP addresses, etc.
     2. Click on the "Cloudera Manager" link on the Master Node.
     3. Login (default login/password = admin/admin).
     4. Click on "HDFS" in the list of services.
     5. Click on "Configuration" tab and search for text 'safety' as shown in the screenshot below.



     6. For the safety value "Cluster-wide Advanced Configuration Snippet (Safety valve for core-site.xml)," add the following snippet at the end:
        <property>

        <name>fs.s3n.awsAccessKeyId</name>

        <value>YOUR ACCESS KEY ID THAT YOU SET UP FOR THE TENANT</value>

        </property>

        <property>

        <name>fs.s3n.awsSecretAccessKey</name>

        <value>YOUR SECRET ACCESS KEY THAT YOU SET UP FOR THE TENANT</value>

        </property>

- Spark Cluster (Tested for Spark 1.5.2 image only)
    1. After the cluster is created, ssh into the master virtual node.
    2. Go to /etc/hadoop/conf.
    3. Open core-site.xml.
    4. Append the following properties to this file:
       <property>

       <name>fs.s3n.awsAccessKeyId</name>

       <value>YOUR ACCESS KEY ID THAT YOU SET UP FOR THE TENANT</value>

       </property>

       <property>

       <name>fs.s3n.awsSecretAccessKey</name>

       <value>YOUR SECRET ACCESS KEY THAT YOU SET UP FOR THE TENANT</value>

       </property>

OTHER DOCUMENTATION RESOURCES

USER & ADMIN GUIDE (General Concepts of BlueData EPIC) can be found here.