Cluster Deployment

1. Login to cloudbreak
2. Add the “hdpsecure” blueprint
3. Add the “database-setup” recipe as a post-ambari-install
4. Add the “create-tuto-users” recipe as a post-ambari-install
5. Add the “create-hdfs-homes” recipe as a post-**cluster**-install
6. Create a cluster and pick the “hdpsecure” blueprint. Make sure you turn on “Advanced Mode”
    a. Standard_D12_v2 are the recommended machines for the 3 masters.
    b. Standard_D3_v2 are the recommended machines for workers.
    c. Ambari should be installed on “Master1”
7. On the recipe page, select “create tuto users” on all machines
8. On the recipe page, select “database-setup” & “create-hdfs-homes”on the master1 machine
9. Add port #6080 to the Master 1 security group for Ranger
10. Add port #21000 to the Master 2 security group for Atlas
11. Enter the Ambari password and select or create a new SSH key
12. Hit create and go fetch a coffee ;)
13. When cluster creation is complete, Cloubreak will have a link to Ambari UI. You can use “admin” as a username for Ambari + the password selected during cluster creation in Cloudbreak.
14. Optional : If you have less than 3 workers, you should reduce the replication factor on HDFS to 1 to avoid false error message.
    a. Open the HDFS config’s tab
    b. Select “Advanced”
    c. In the “General” section change the “Block replication” factor from 3 to 1.

Tag based policy with ranger and atlas tutorial :
Ref : https://hortonworks.com/tutorial/tag-based-policies-with-apache-ranger-and-apache-atlas/

1. Enable Ranger audit to Solr. SKIP, already configured properly
2. Restart all services affected SKIP, already configured properly
3. Explore general information - you might have to change the user’s password since they were imported from local users - this would use LDAP in a production environment.
    3.1. Click on the “Admin” button on top and select “Manage Ambari”
    3.2. Click on groups and add a group called “hadoop”
    3.3. Add a user called “maria_dev” in that group. Make sure she’s an admin to simplify the process
    3.4. Repeat the step above for a user “raj_ops”
4. Explore Sandbox user personas policy.

4.1. Acces the ranger UI : make sure you use the External IP if you're not connected to the internal network of you VM. When using the external IP, make sure your security rules allow port 6080 to be accessible on the "master1" machine.
4.2. You can connect using "admin/admin"
4.3. Click on "hive" resource board.
4.4. Edit policy called ""all - hiveservice".
4.5. Add the "admin" user and save. ← by default, even the admin doesn't have access
4.6. Create a new policy named "admin_default"
4.6.1. Enter "default" in the database
4.6.2. Enter * in the table and hive column fields
4.6.3. Select the "admin" user
4.6.4. Click on "select all" in the permission and save.

5. Access without tag bases policies
5.1. Go to Hive view 2.0
5.2. Create employee table
5.3. Verify table creation
5.4. Add data to table :
5.4.1. Create a text file with the following content :

```
111-111-111,James,San Jose
222-222-222,Mike,Santa Clara
333-333-333,Robert,Fremont
```

5.4.2. Using the file view, upload this file to "/apps/hive/warehouse/employee"
5.5. Skip
5.6. Skip
5.7. Select new table in Hive view
6. Create ranger policy to limit access of Hive data
6.1. Create ranger policy to restrict employee table access
6.1.1. Remember to use the externally accessible IP
6.1.2. Use the "hive" service
6.1.3. Enter the data as prescribed
6.1.4. Idem
6.1.5. Verify your new policy is there
6.1.6. Skip the delete
6.2. Verify the ranger policy is in effect
6.2.1. A select * will fail

    6.2.2. A select name should work
    6.2.3. Visit ranger to check the audit
7. Create Atlas tag to classify data. Login using “admin/admin”. The steps can be followed verbatim
8. Use as-is.