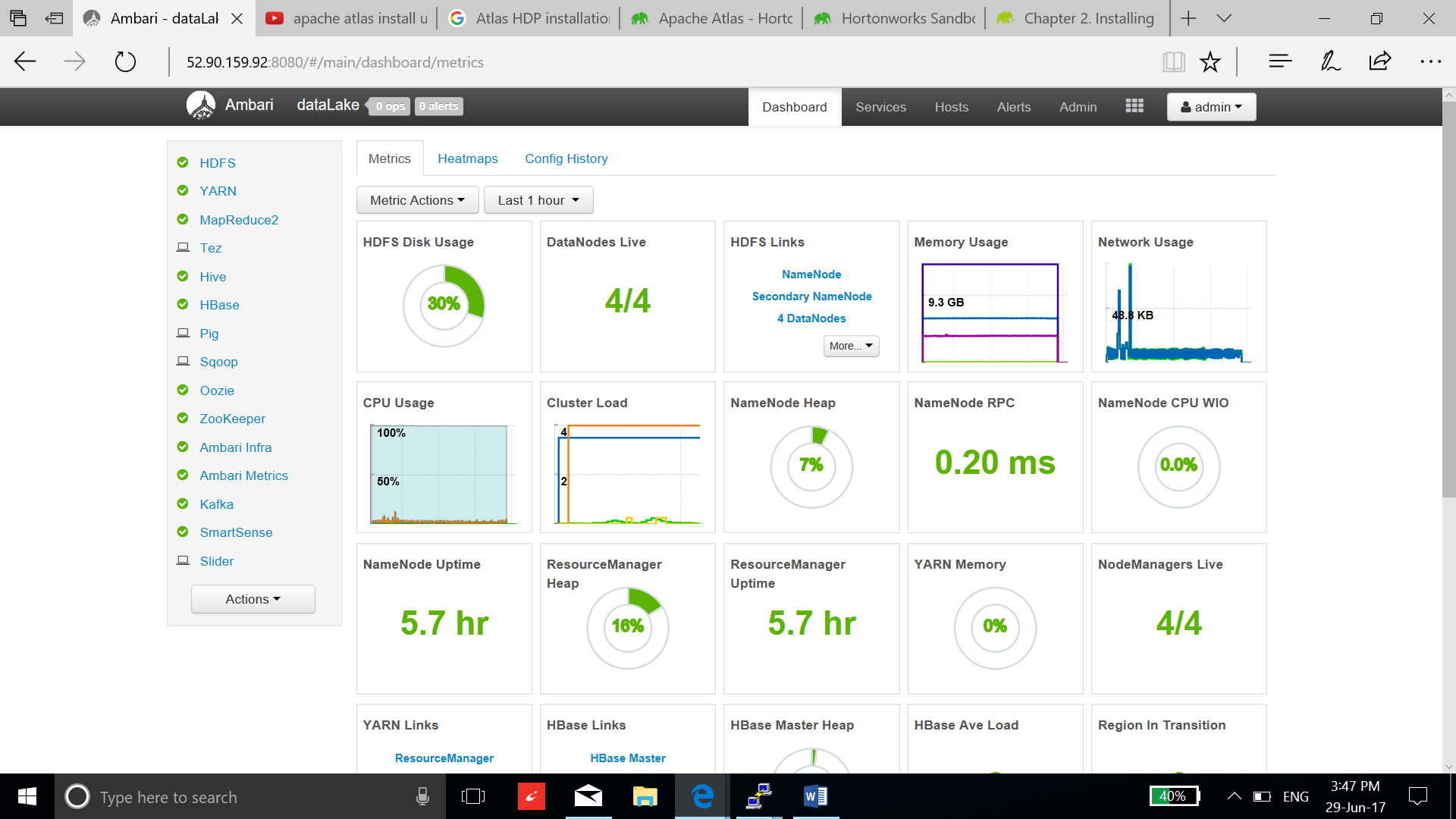
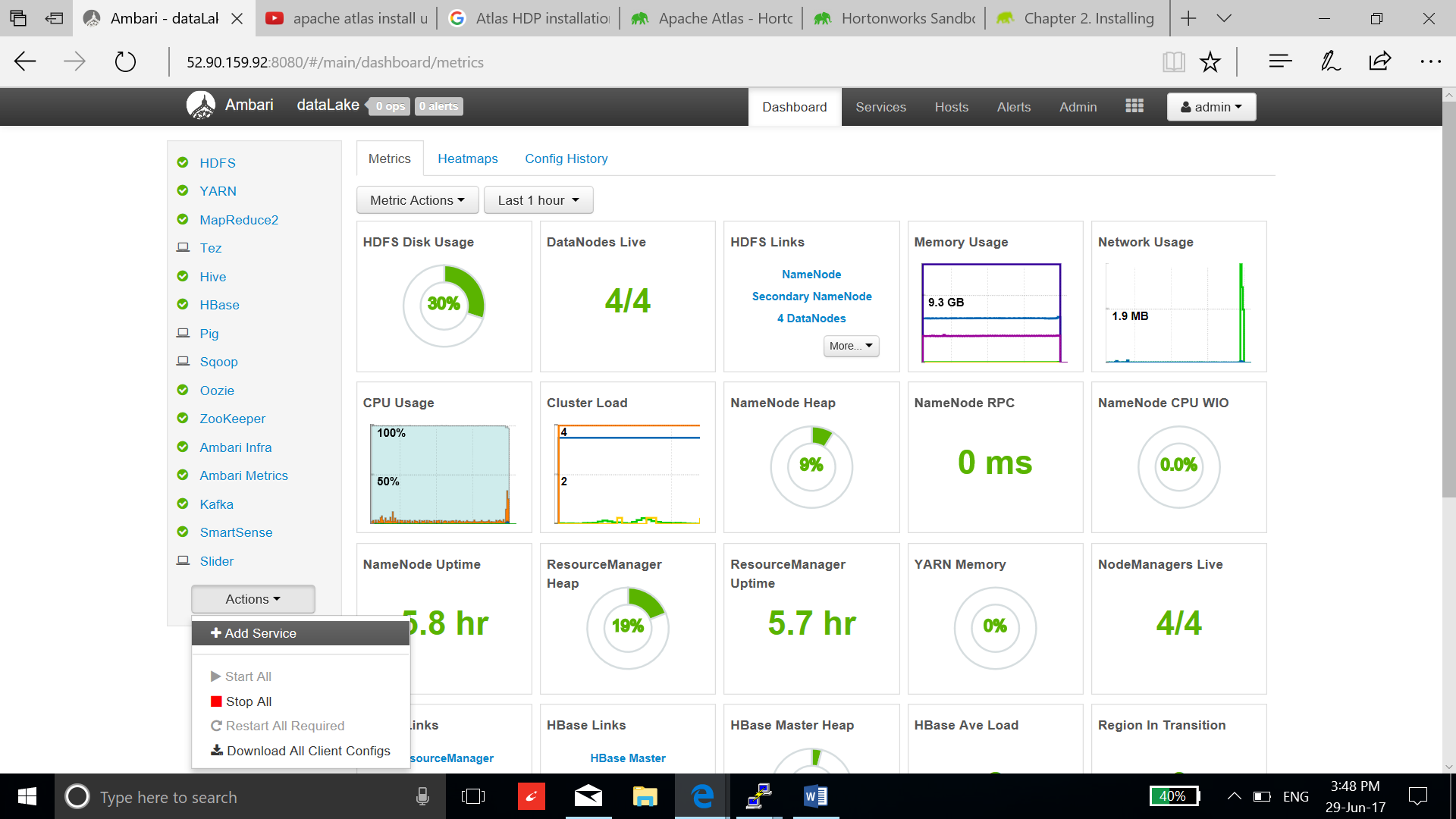
Steps to install Atlas:

1. Prerequisite:

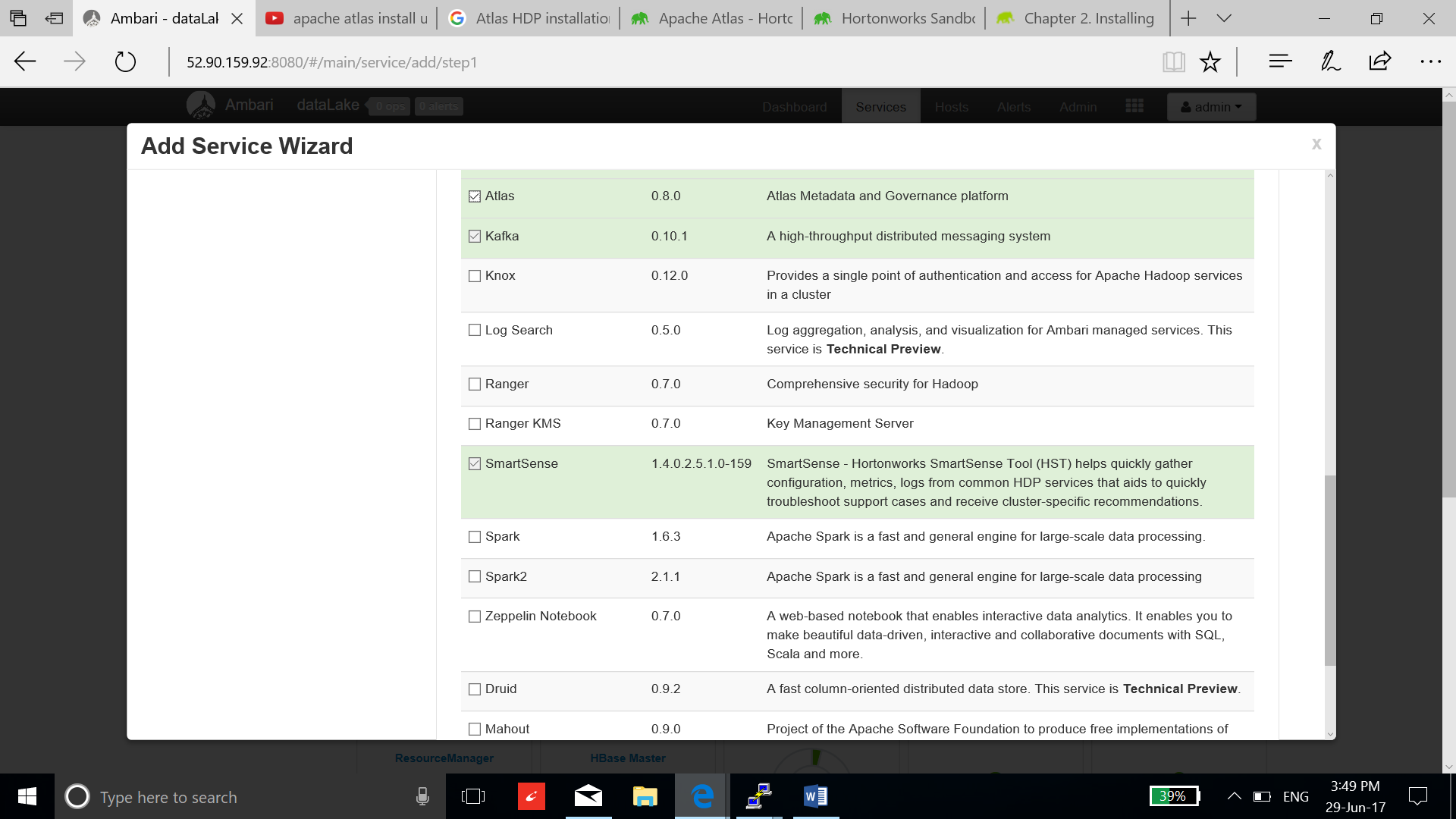
* Ambari Infra
* HBase
* Kafka
* Use the default configuration and add all services first.



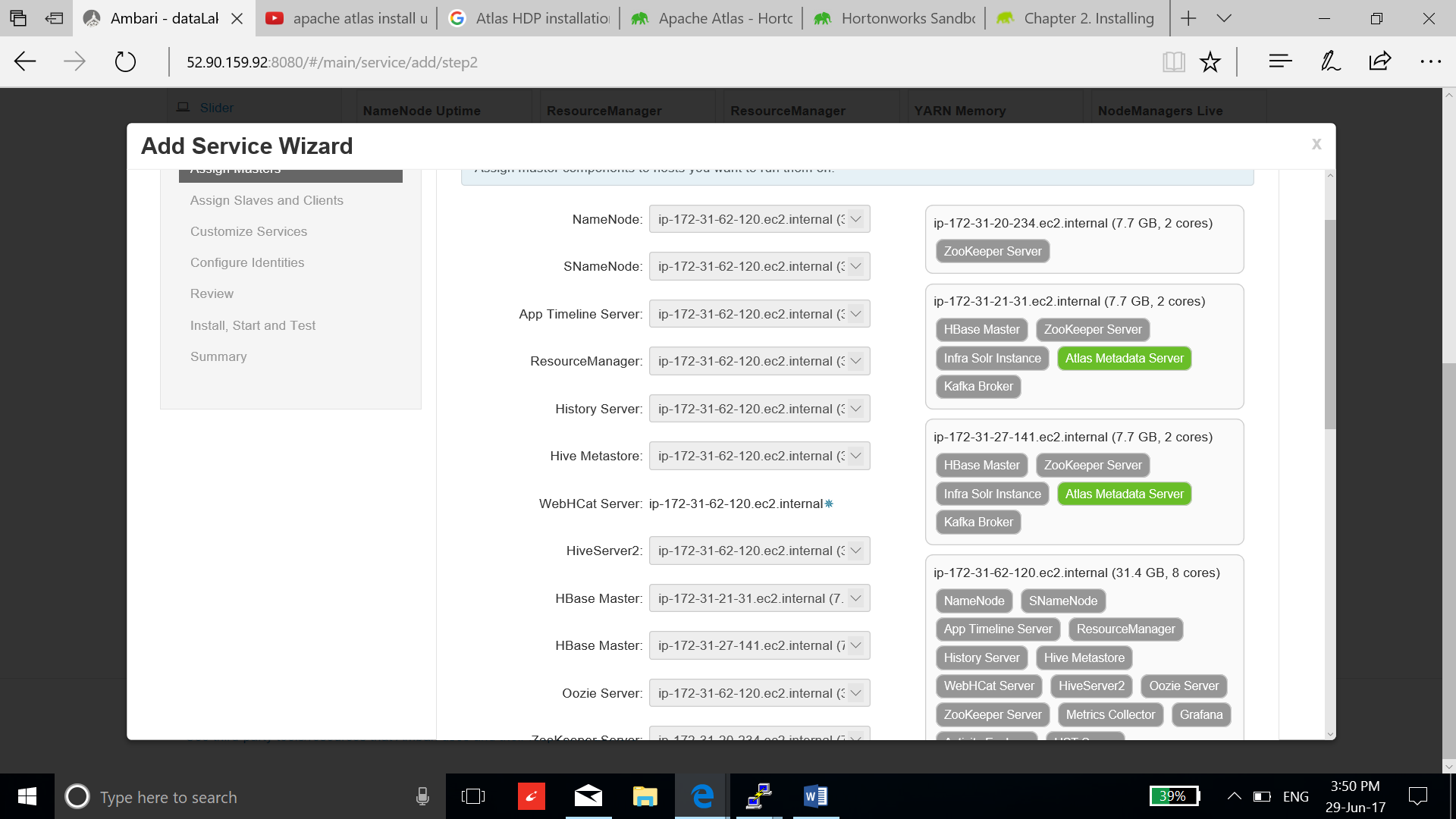
1. Select Add services:



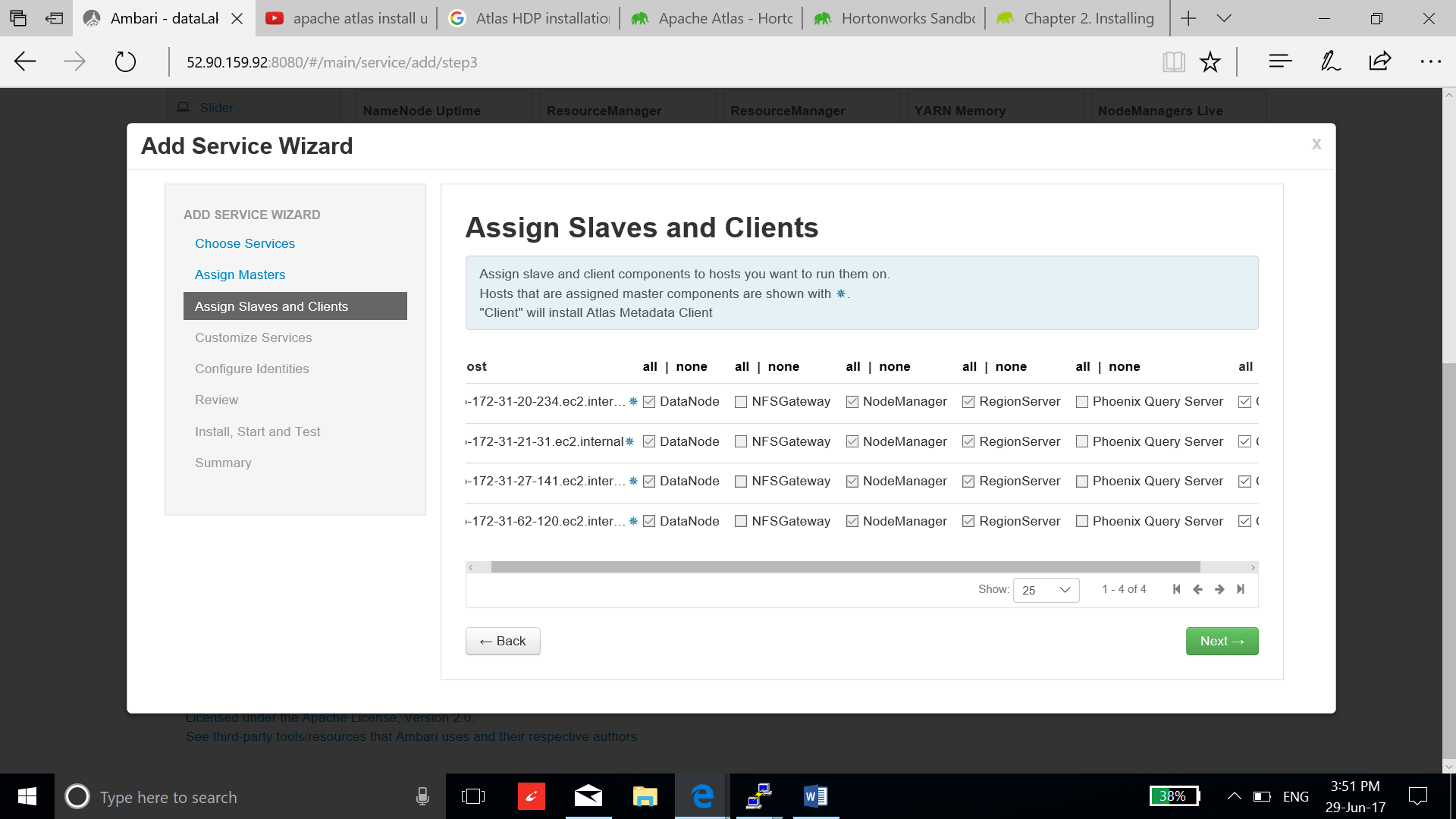
1. Add Atlas service:



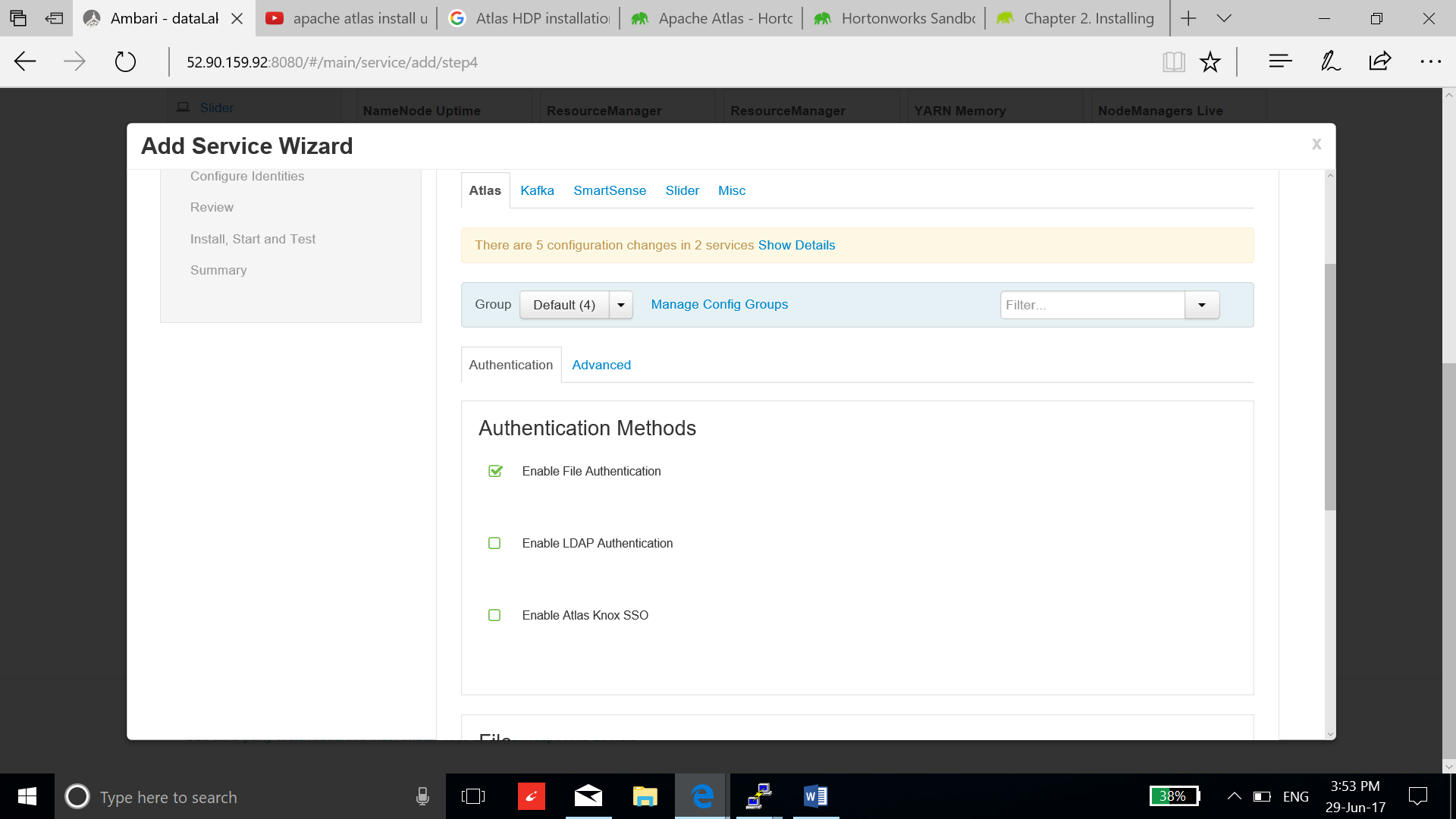
1. Add the master nodes of the Atlas server:



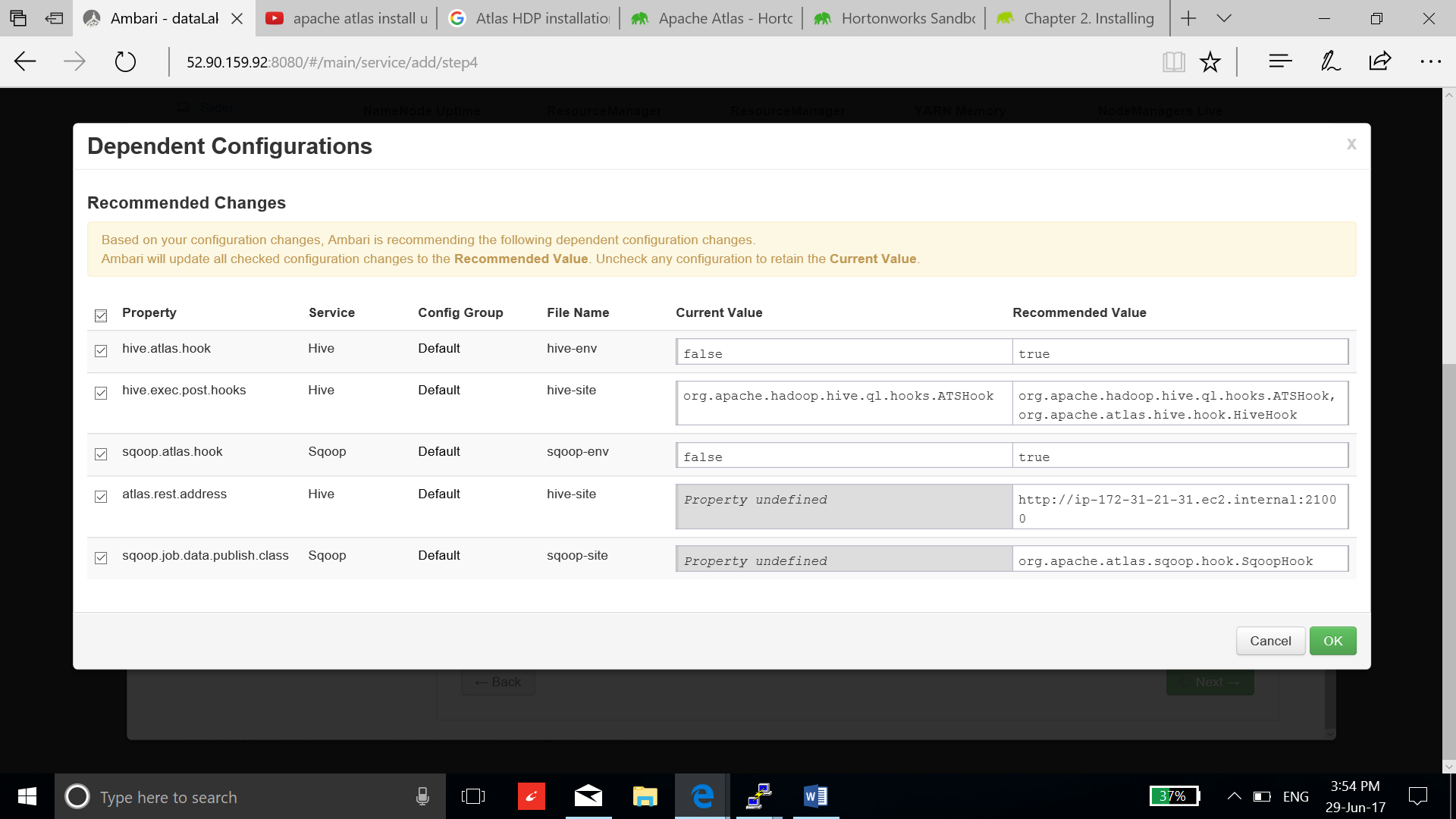
1. Add the client node:

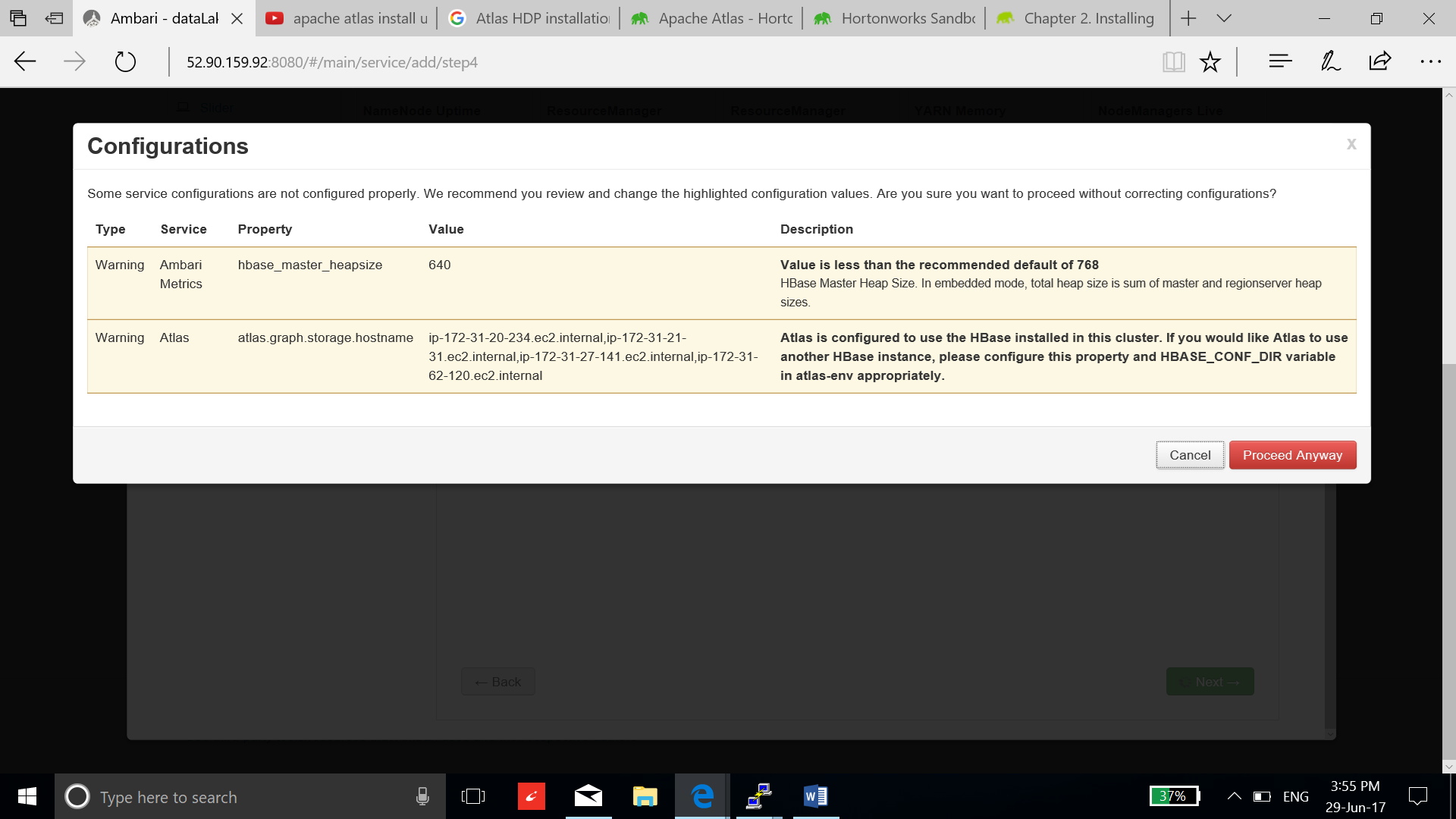


1. Keep the default setting:

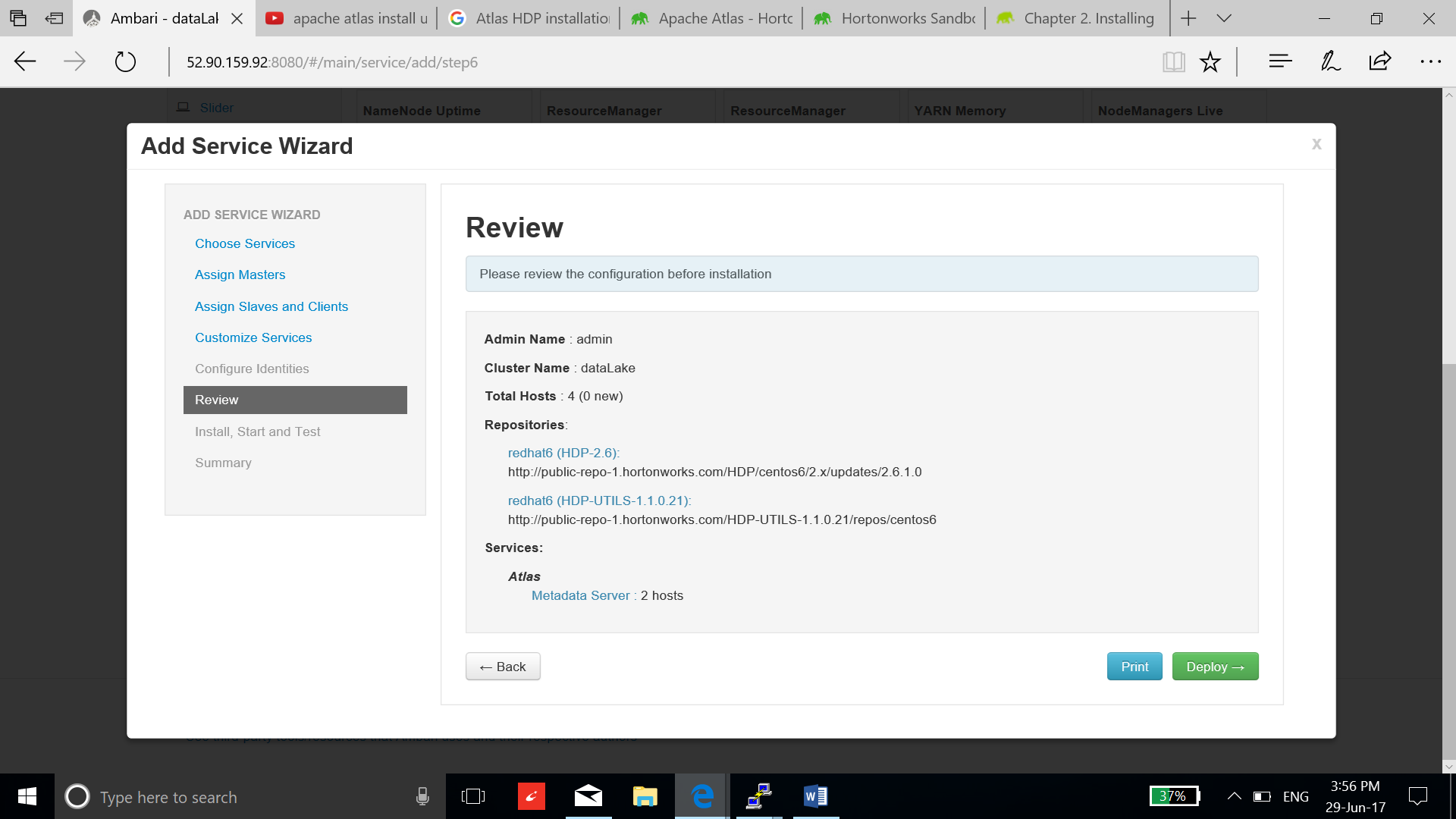


1. Accept the suggested settings and select proceed anyways.

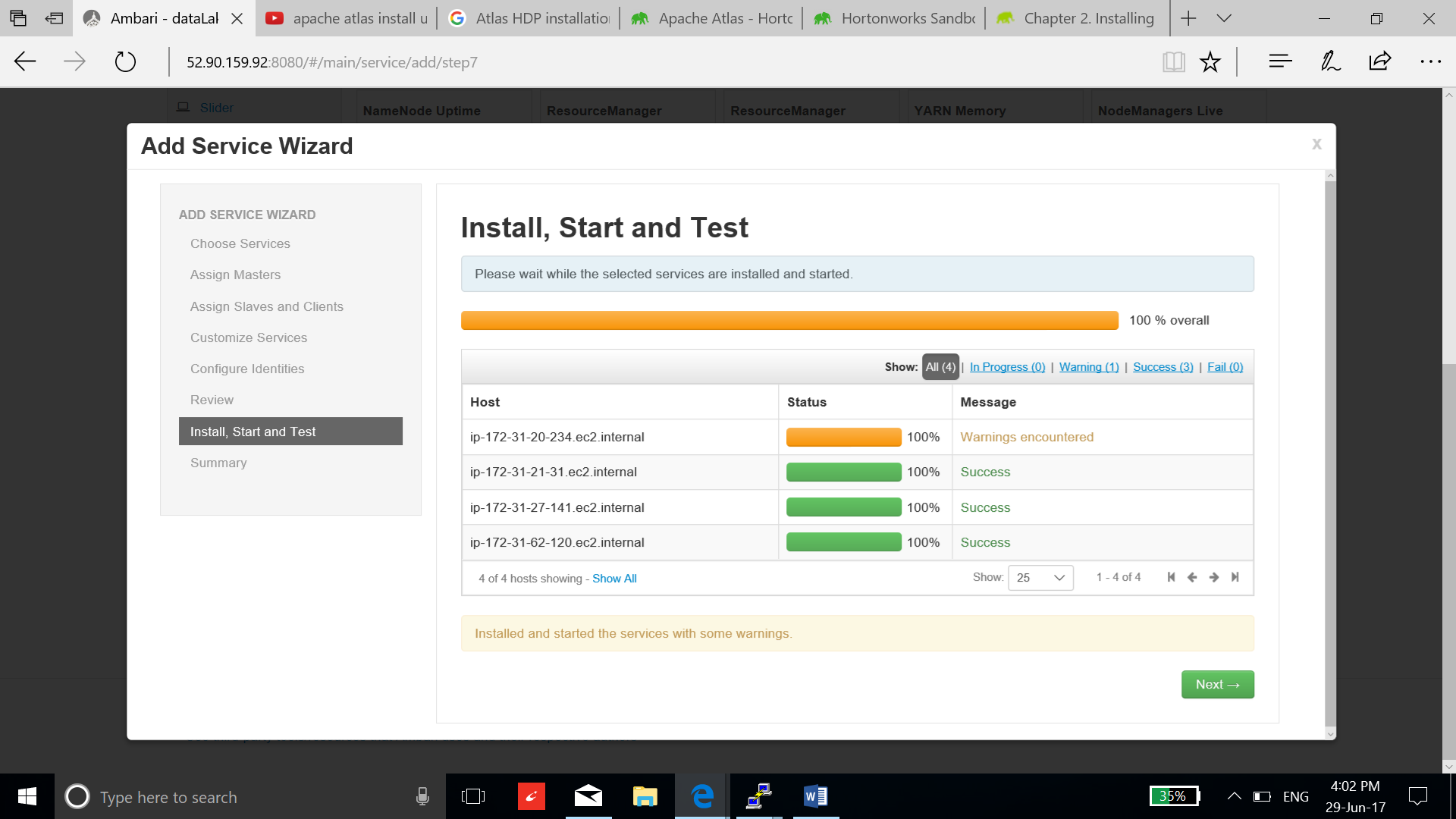




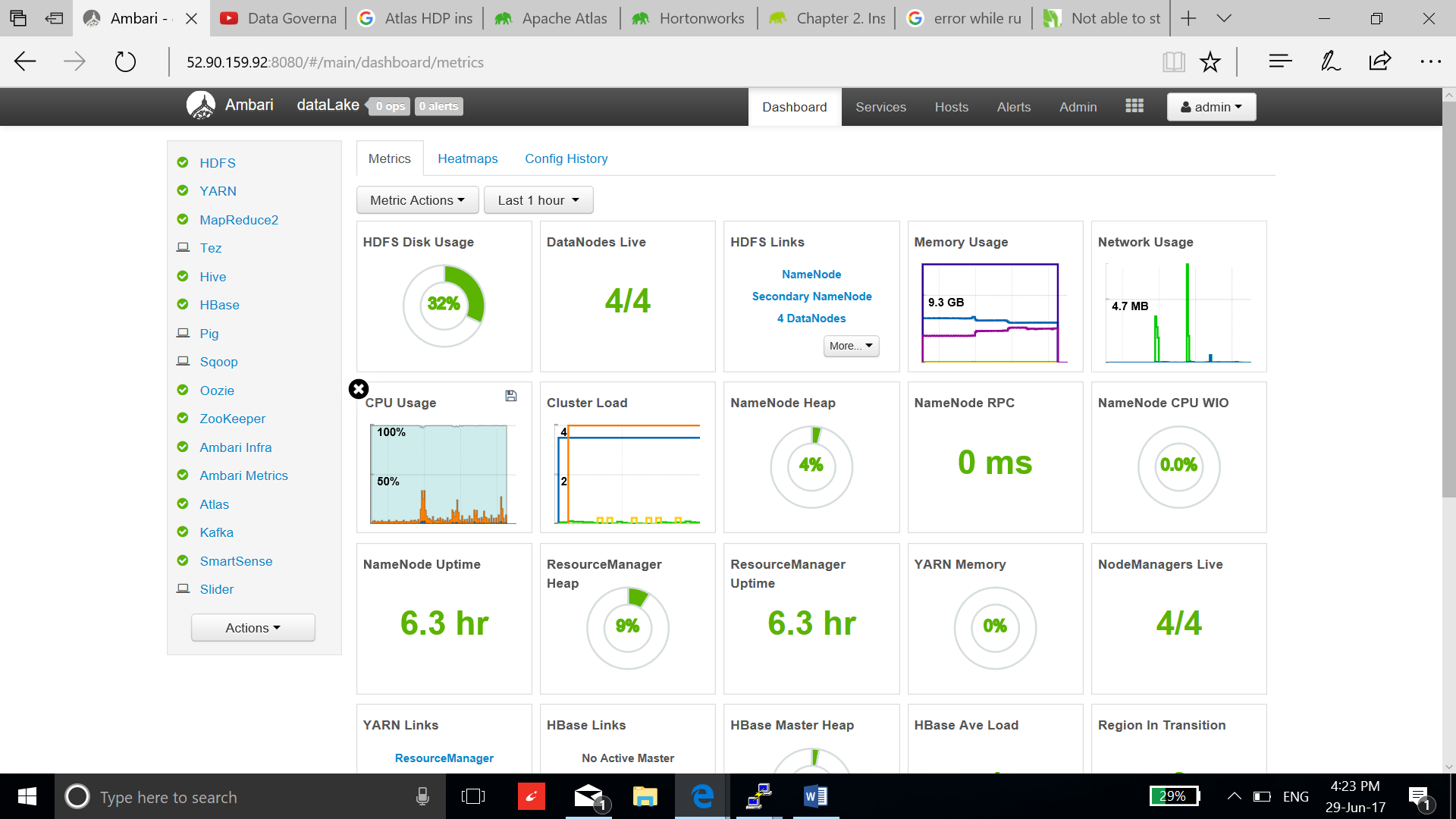
1. Select Deploy and wait for the installation:



1. Installation has completed, restart all the required application and troubleshoot the warning message.



1. After the restart, the warning will go away, warning was due to connection failure.

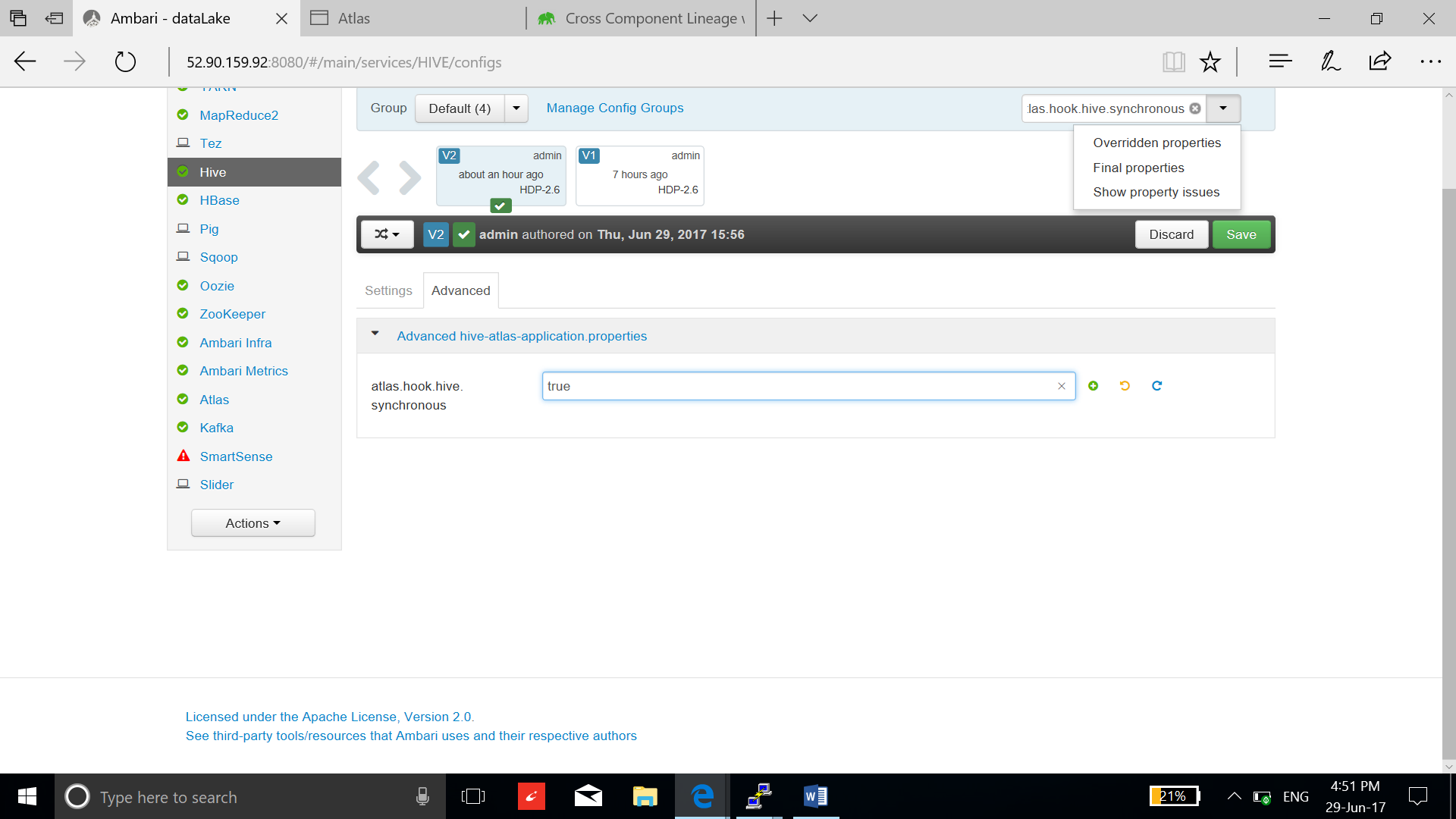


1. To run Taxonomy feature, go to Atlas/config/advanced/custom application properties and add below property:

**atlas.feature.taxonomy.enable=true**

**Validate the installation by checking the data lineage:**

1. To activate hive lineage in Atlas, synchronize hive with Atlas. Go to hive/config/advanced and search for **atlas.hook.hive.synchronous.** Set the parameter **true, save and restart hive.**



1. Sqoop-hive lineage:

* Download the postgres driver and copy it in the ambari-server sqoop library:

**curl -L 'http://jdbc.postgresql.org/download/postgresql-9.2-1002.jdbc4.jar' -o postgresql-9.2-1002.jdbc4.jar**

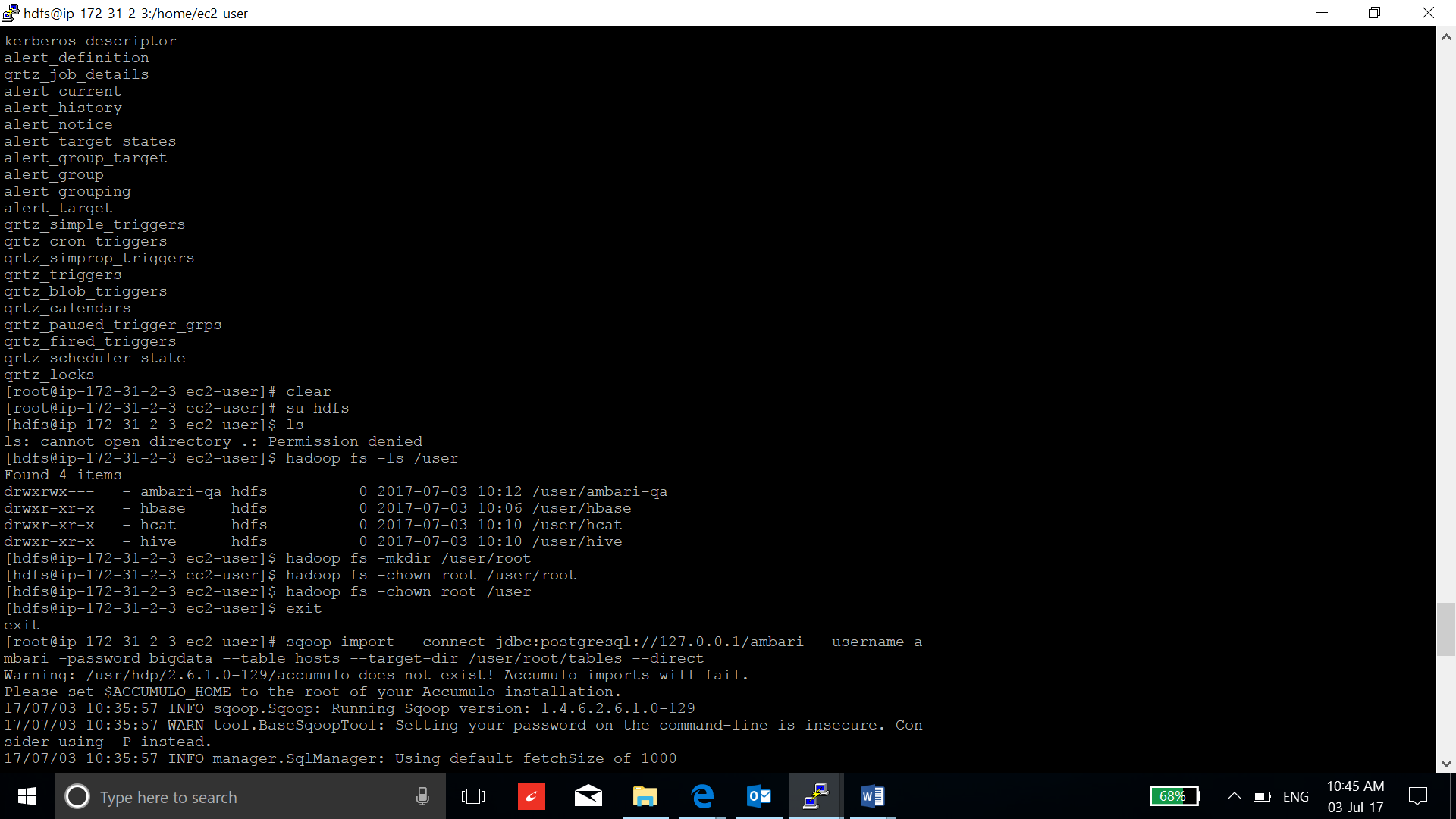
**sudo cp postgresql-9.2-1002.jdbc4.jar /usr/hdp/current/sqoop-client/lib/**

* Open security connection to the postgresql database by editing pg\_hba.conf file. Add the host configuration line in the starting of the file and restart the postgresql:

**host all all 0.0.0.0/0 md5**

**/etc/init.d/postgresql restart**

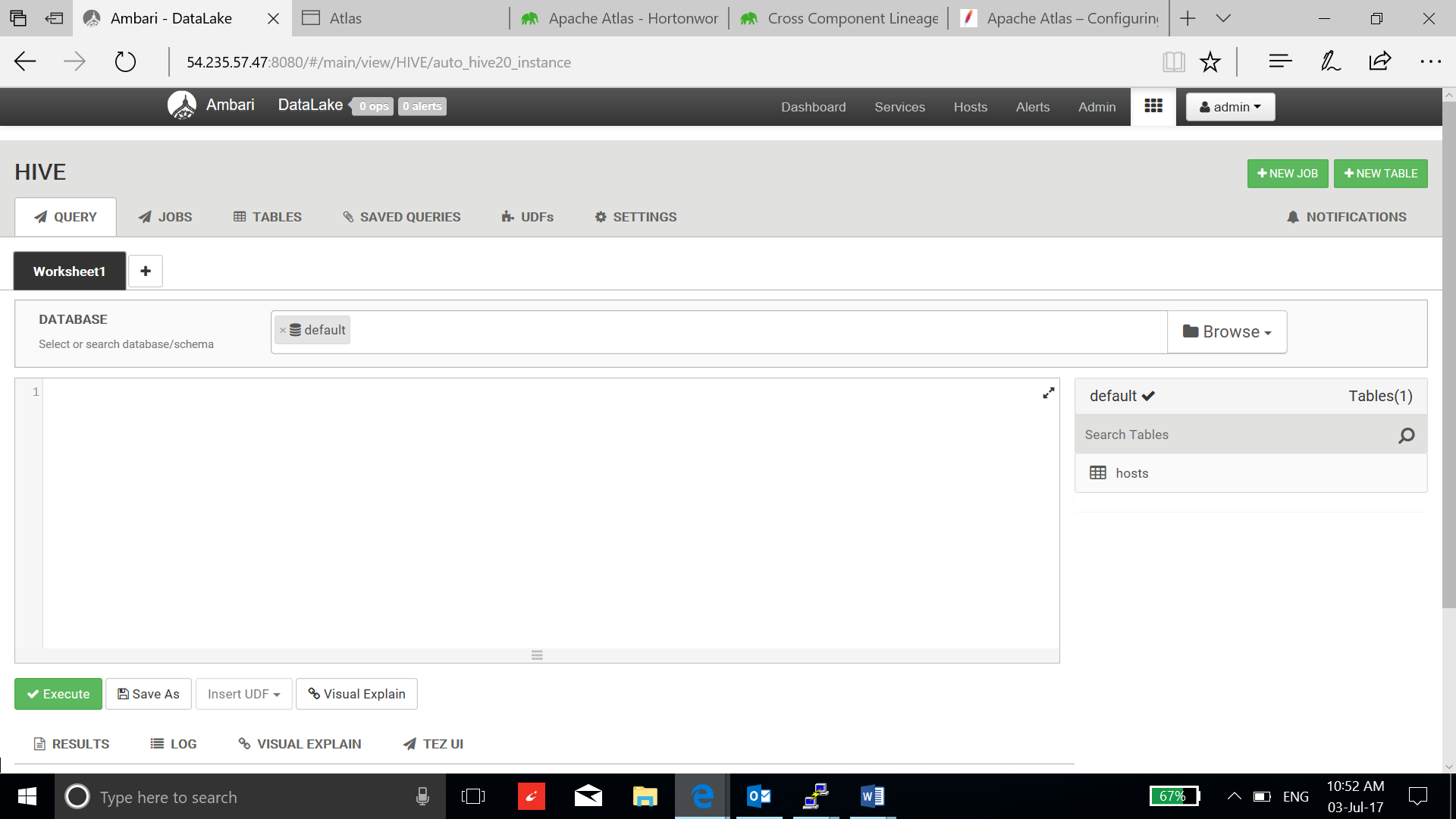
* Create a /user/root directory in HDFS:



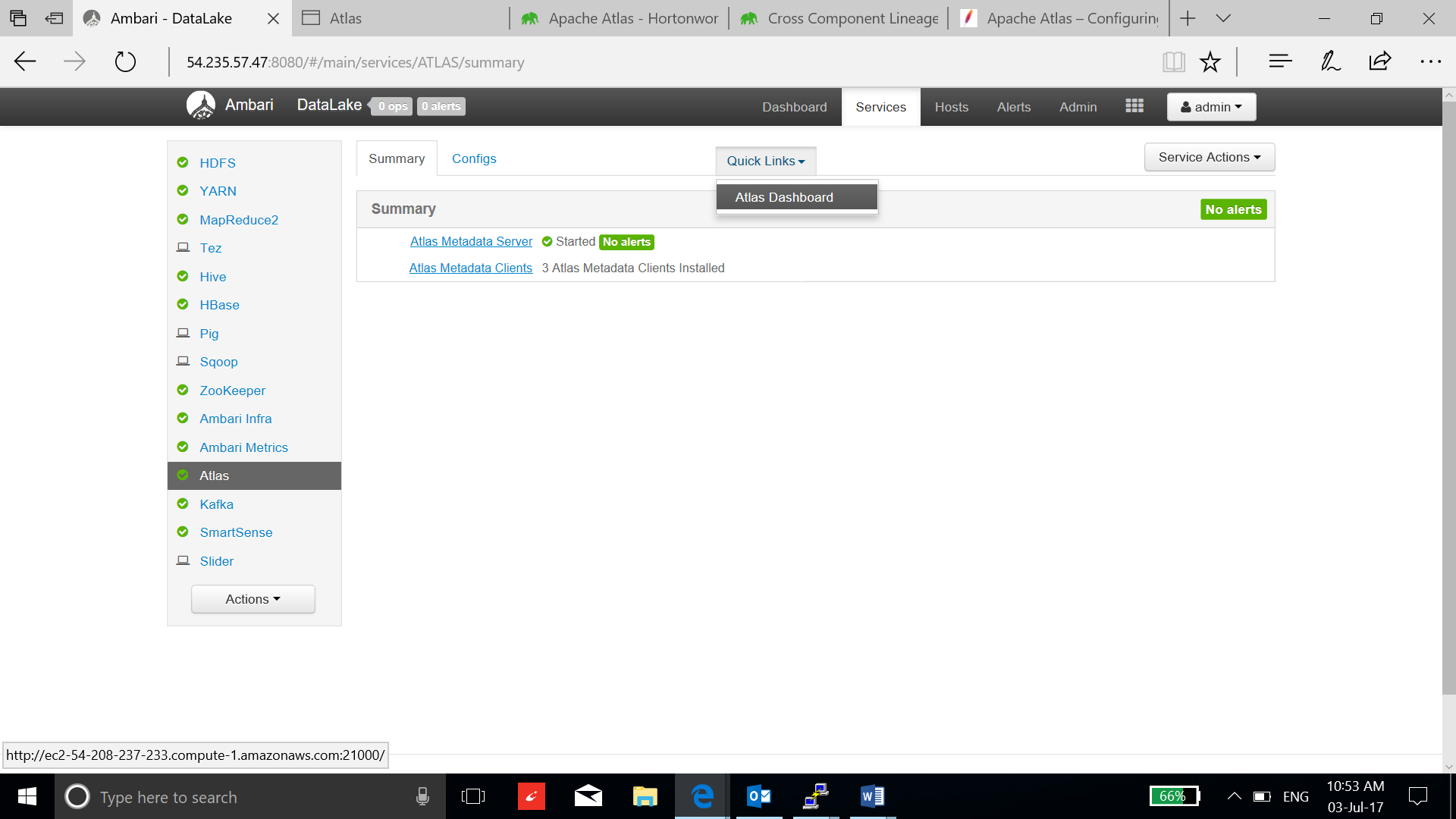
* Transfer the table from ambari database in postgresql on server instance to the hive:

**sqoop import --connect jdbc:postgresql://127.0.0.1/ambari --username a mbari -password bigdata --table hosts --hive-import --create-hive-table –direct**

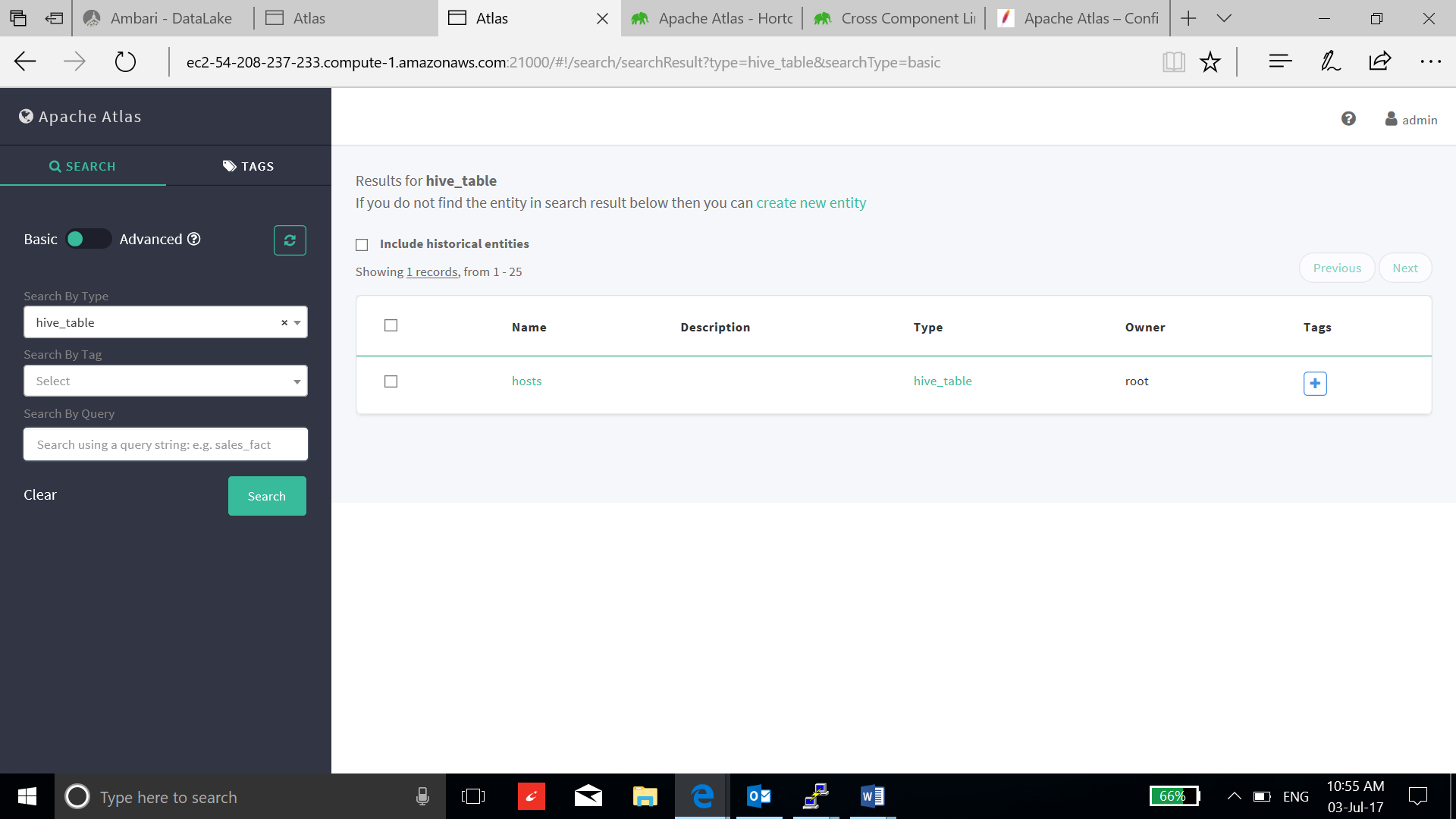
* Go the hive2 view on ambari webpage and verify the transfer of the table:



* Now go to the Atlas dashboard, logon using admin as a user with “admin” password.



* Search hive\_table as a type and hosts as table\_name:

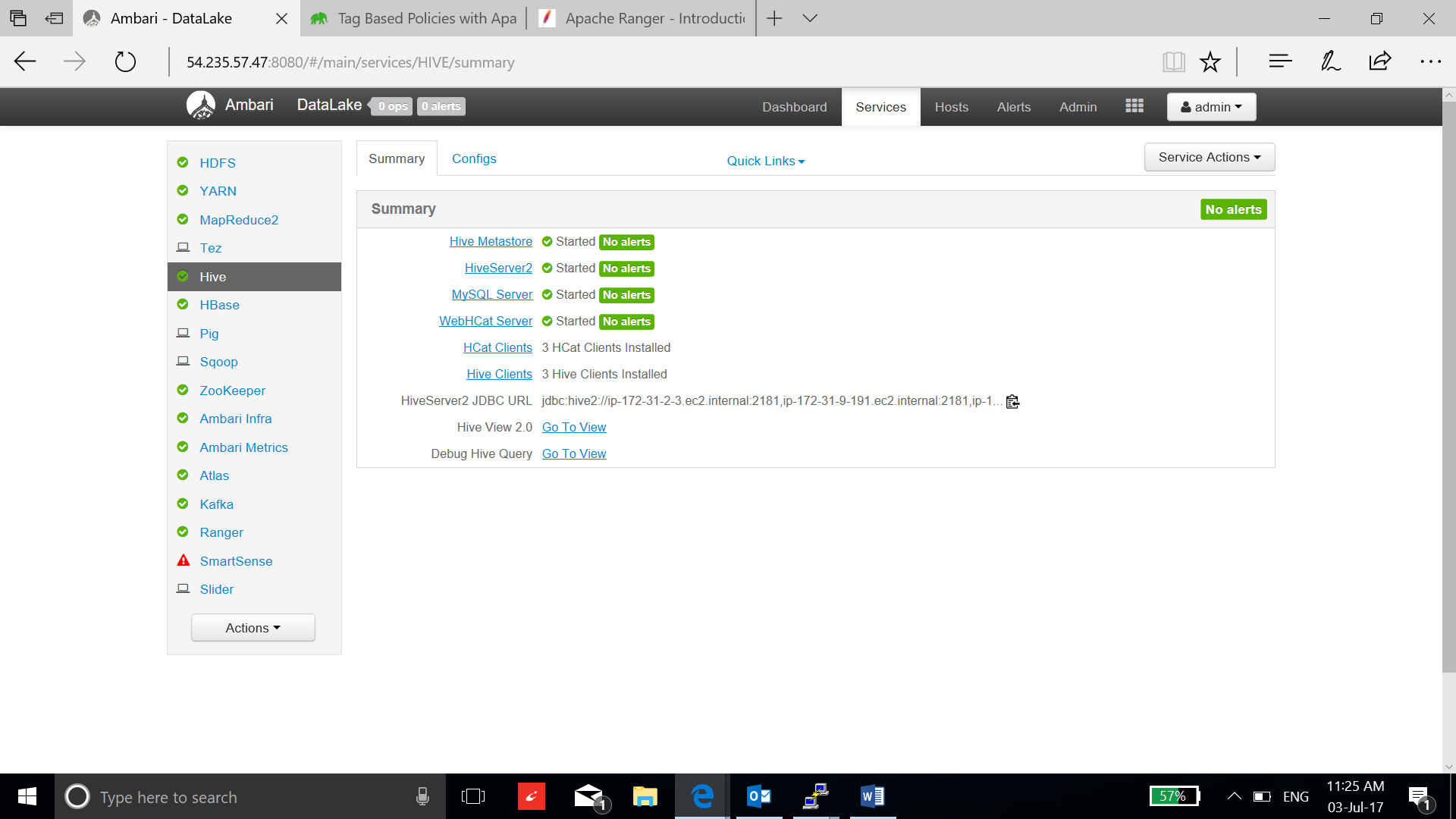


* Click the hosts table and you can see the lineage and many more parameters:

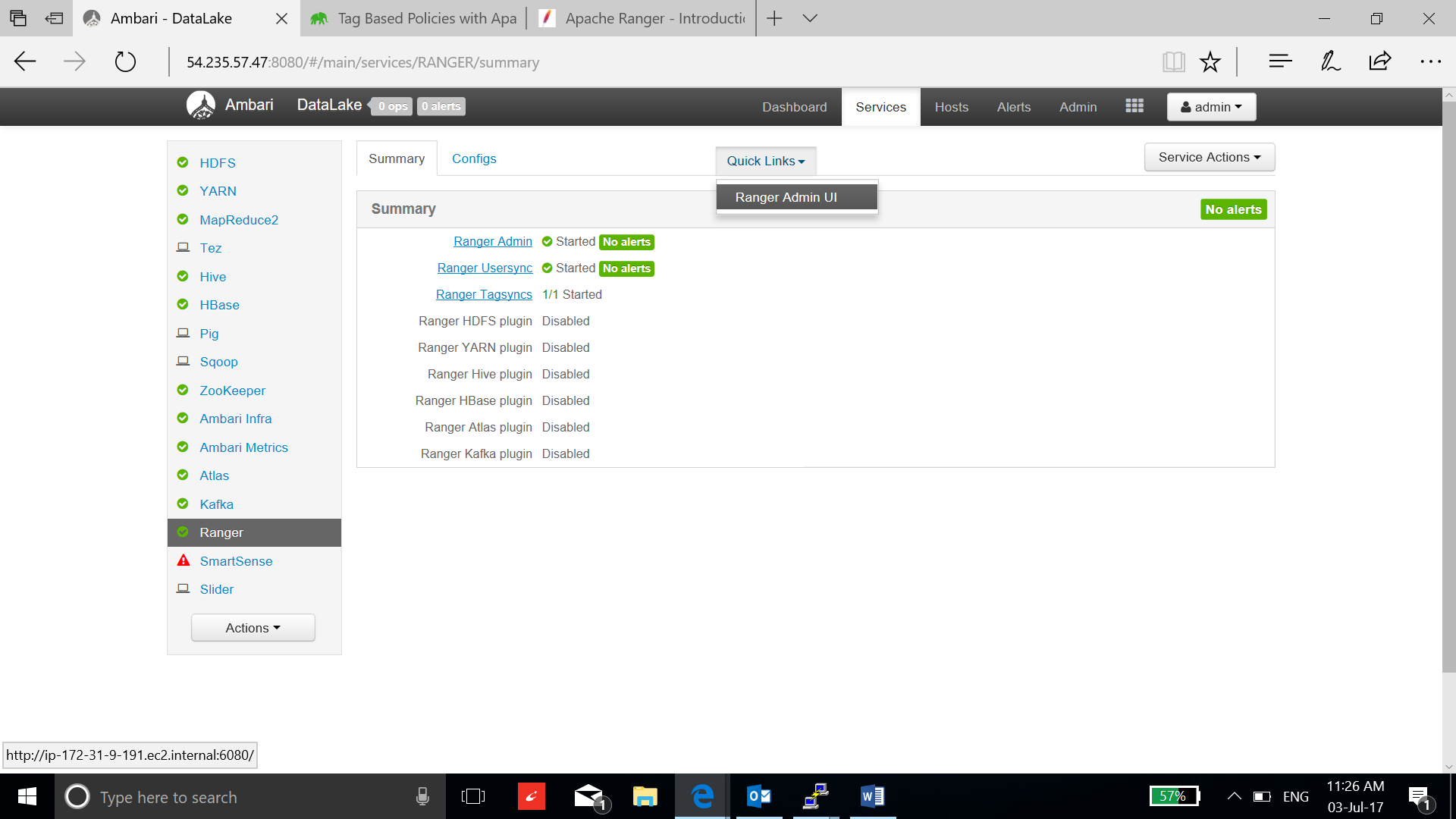


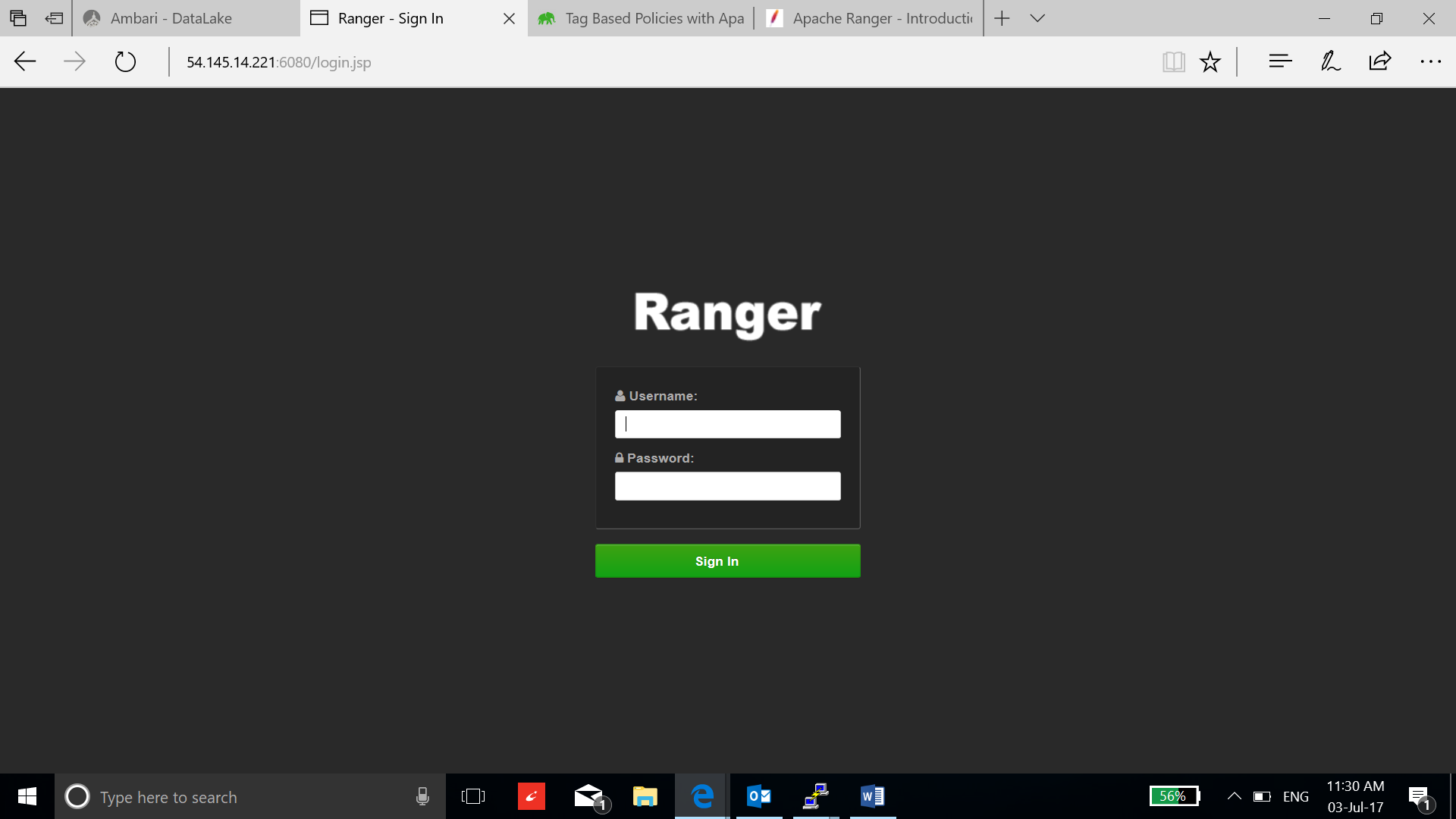
1. Tag based policies using Apache Ranger:

* Install Ranger using the installation guide of ranger:



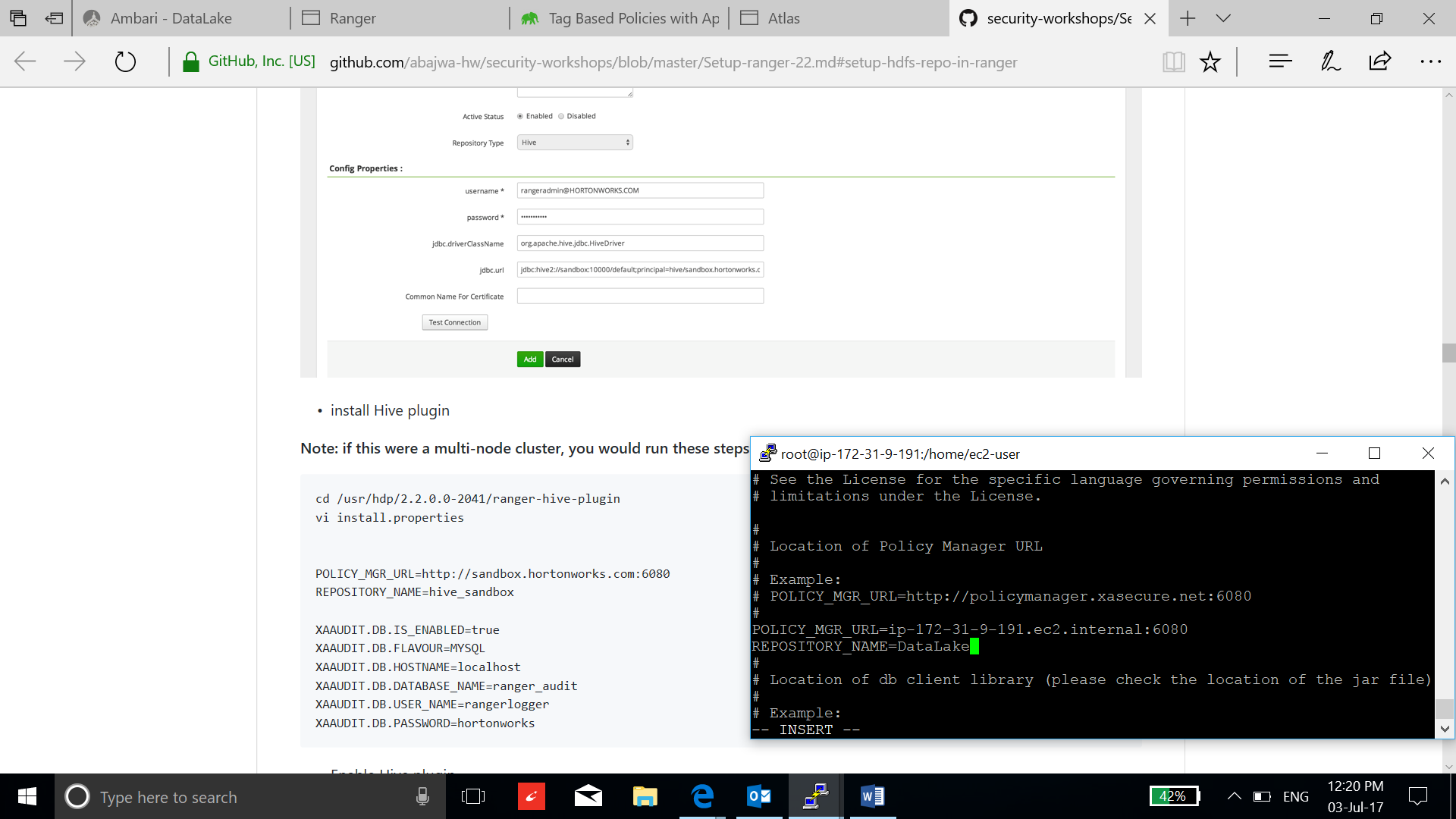
* Go to ranger admin UI, sign on using username: admin and password: admin:



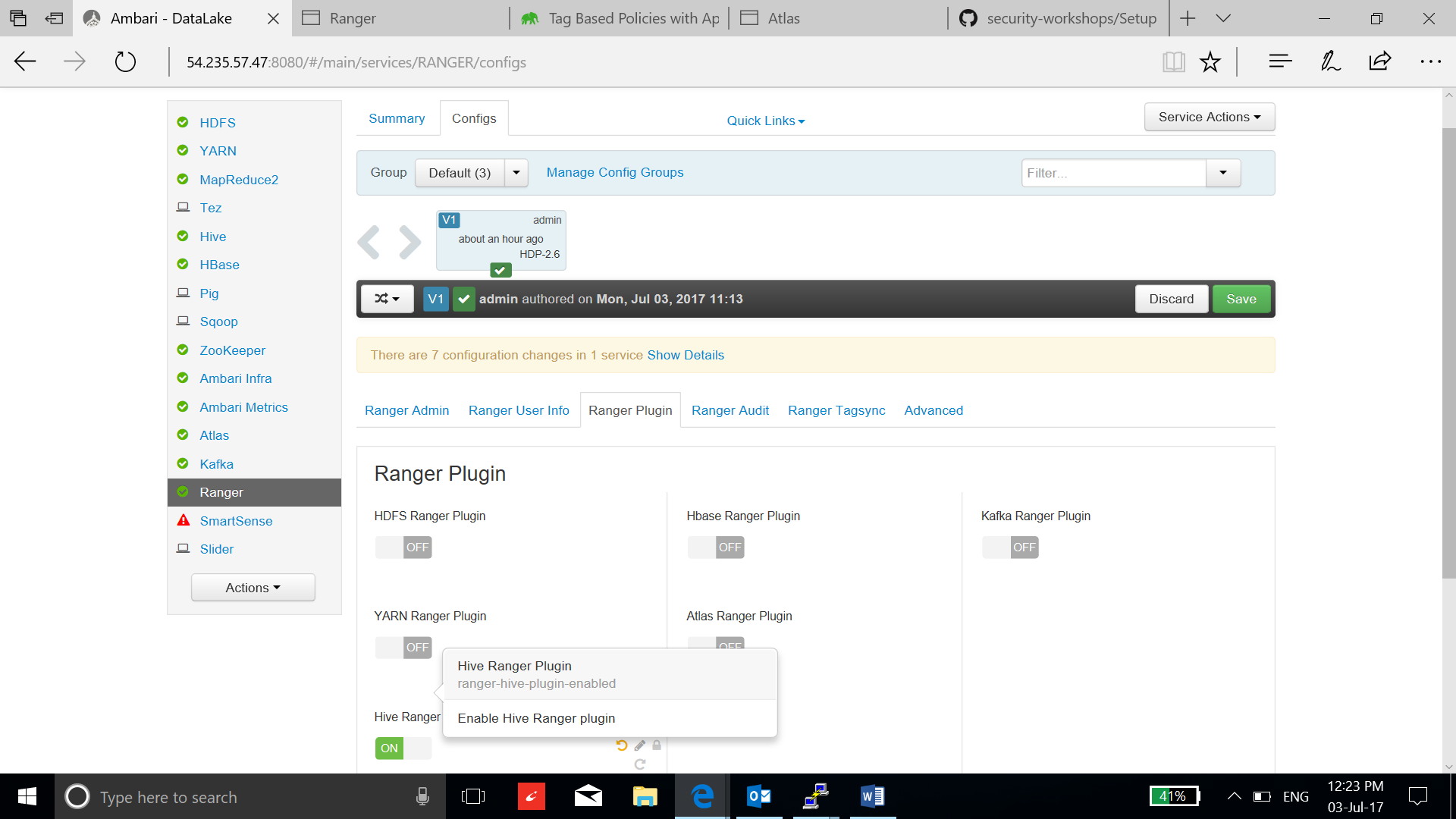


* Add the following parameter to the instance running hive:

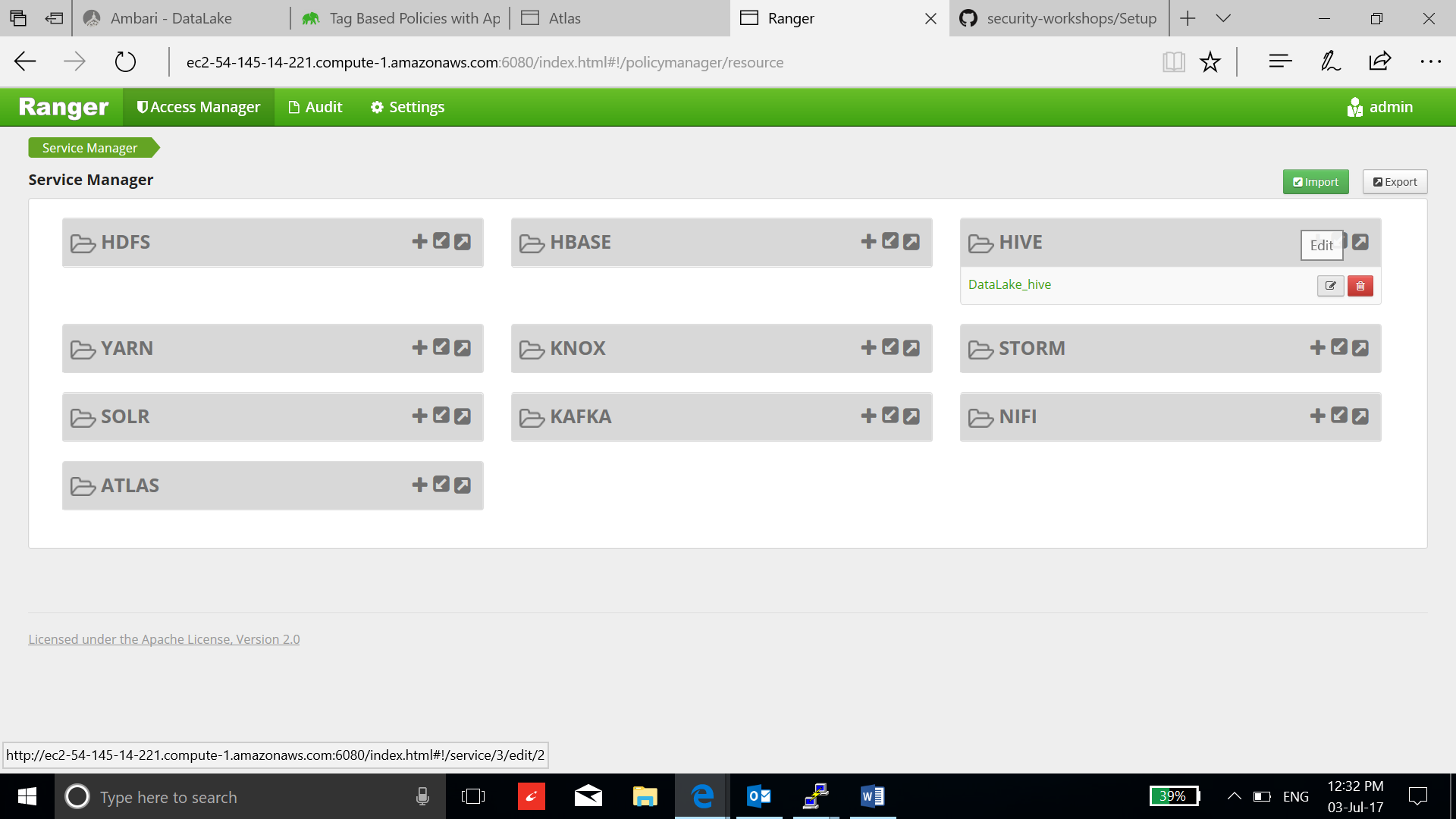
**vi /usr/hdp/2.6.1.0-129/ranger-hive-plugin/install.properties**



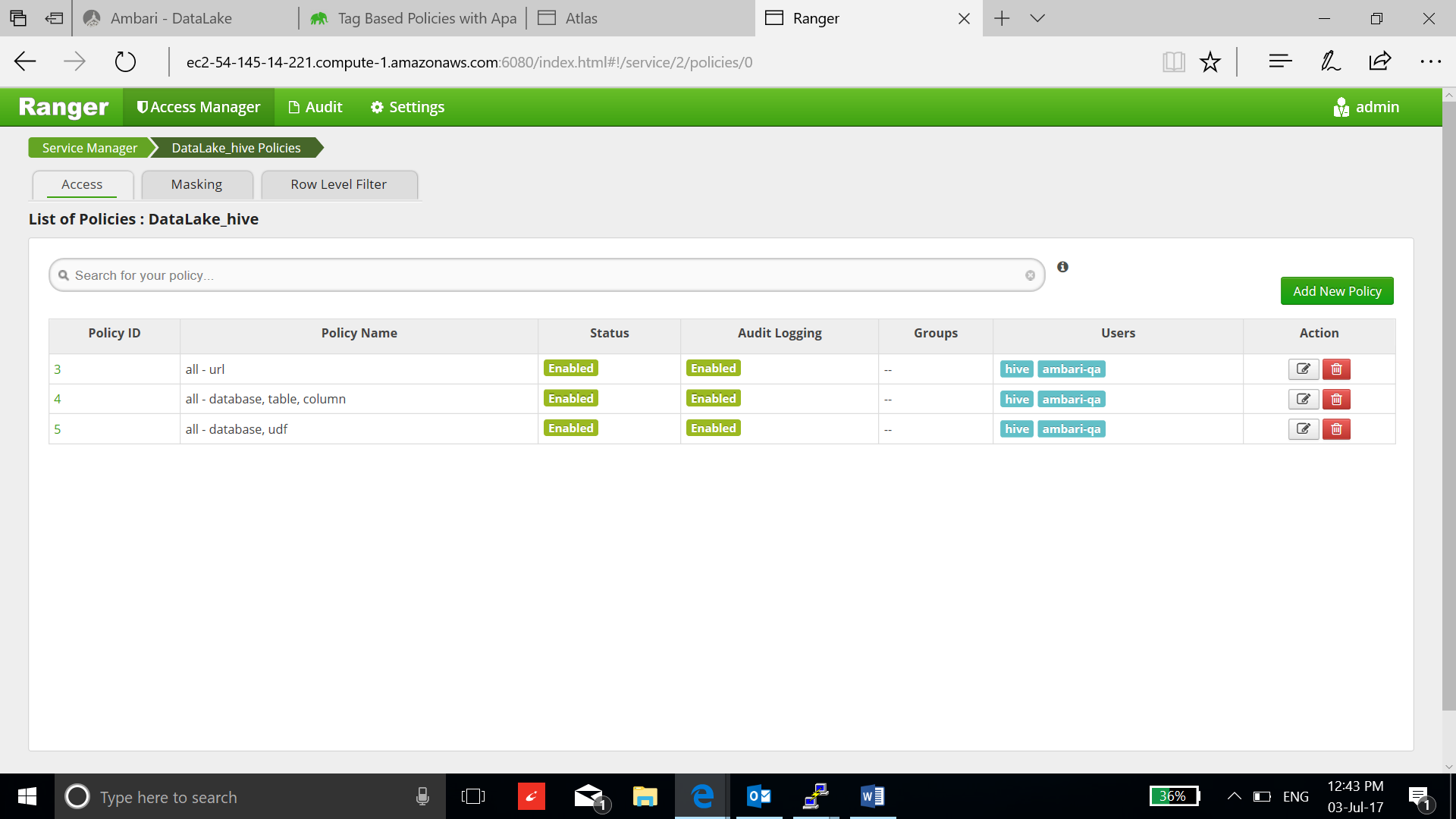
* Enable hive plugin by going to the ranger configuration, save configuration and restart hive and ranger:

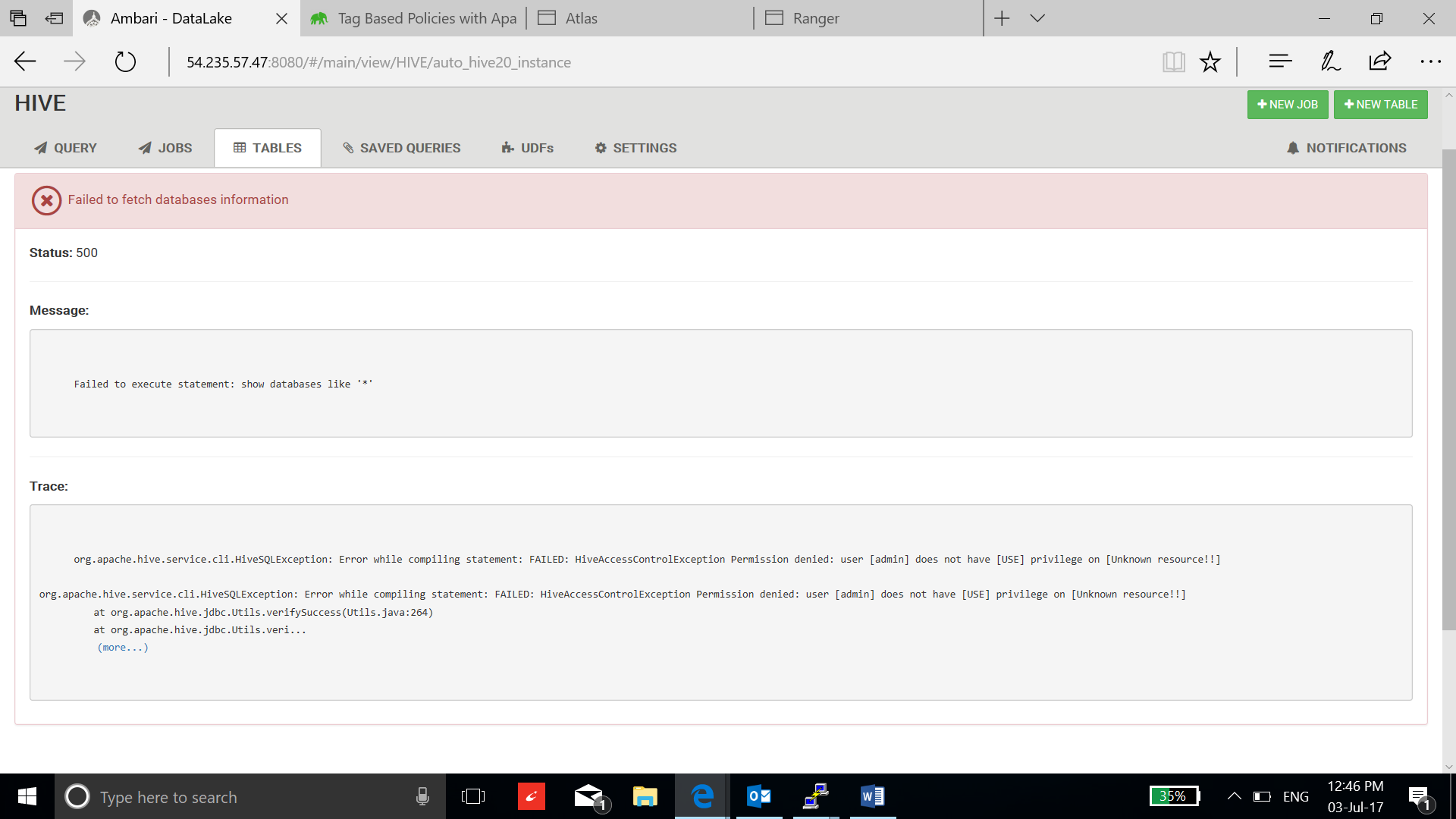


* Select a Data\_hive policy by clicking:

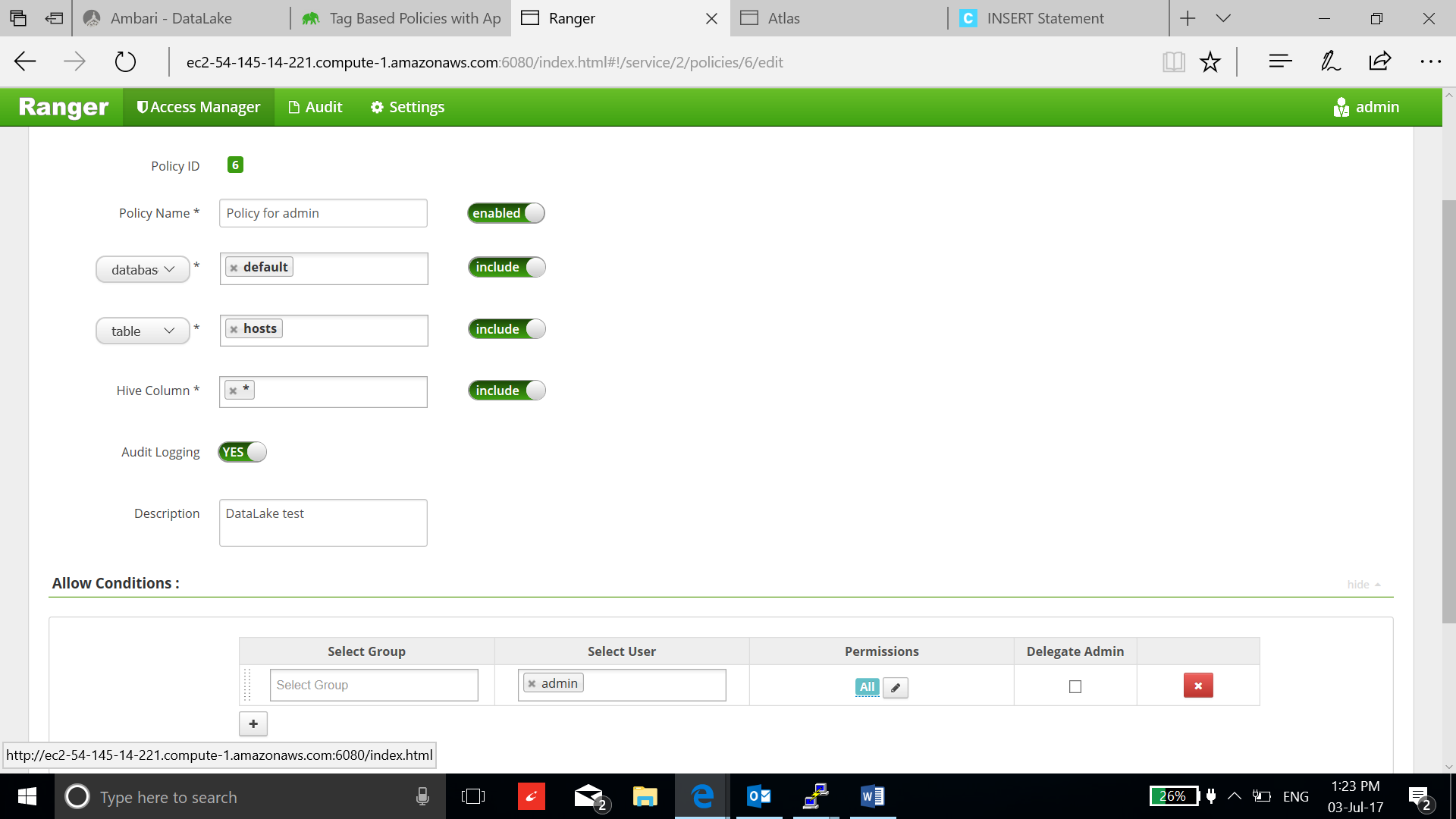


* Use can see user admin has no access to the hive:





* Add new policy to give access to the user admin:





* Create a table using hive2 interface:

**create table employee (ssn string, name string, location string)**

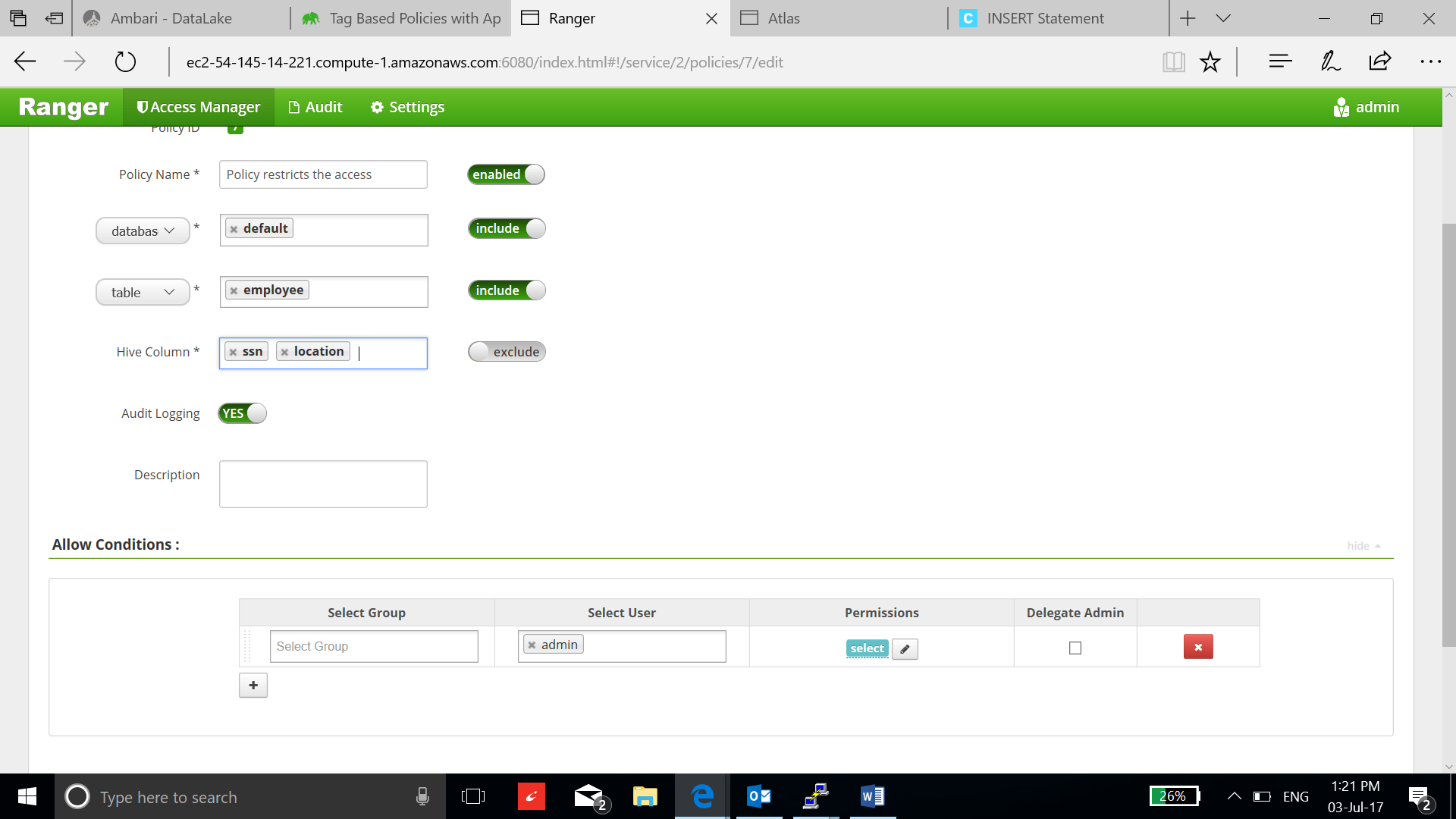
**row format delimited**

**fields terminated by ','**

**stored as textfile;**

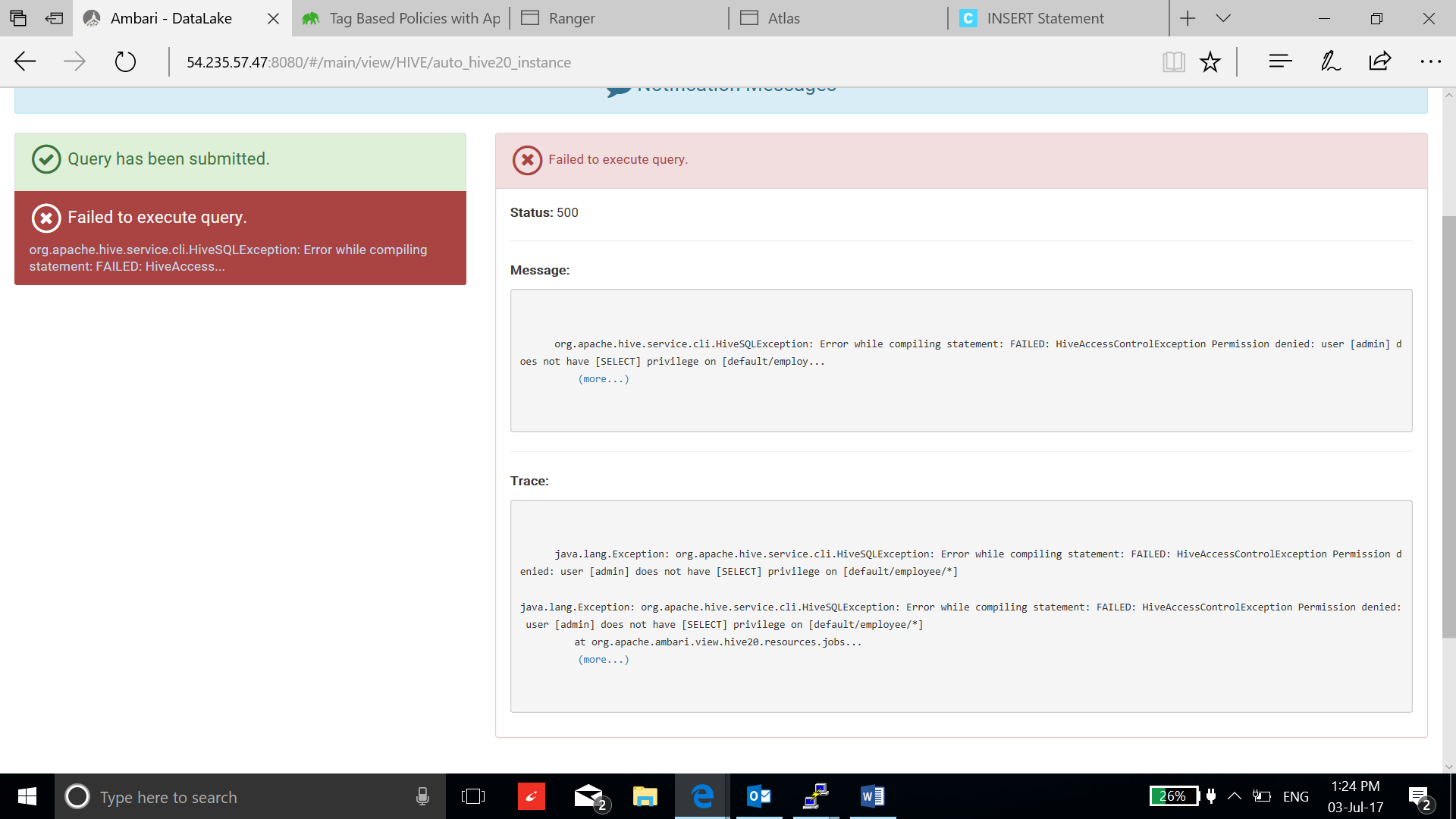
**insert into employee values ("111-111-111","James","San Josen"),("222-222-222","Mike","Santa Claran"),("333-333-333","Robert","Fremont");**

* Add new policy to restrict the column of the table employee:

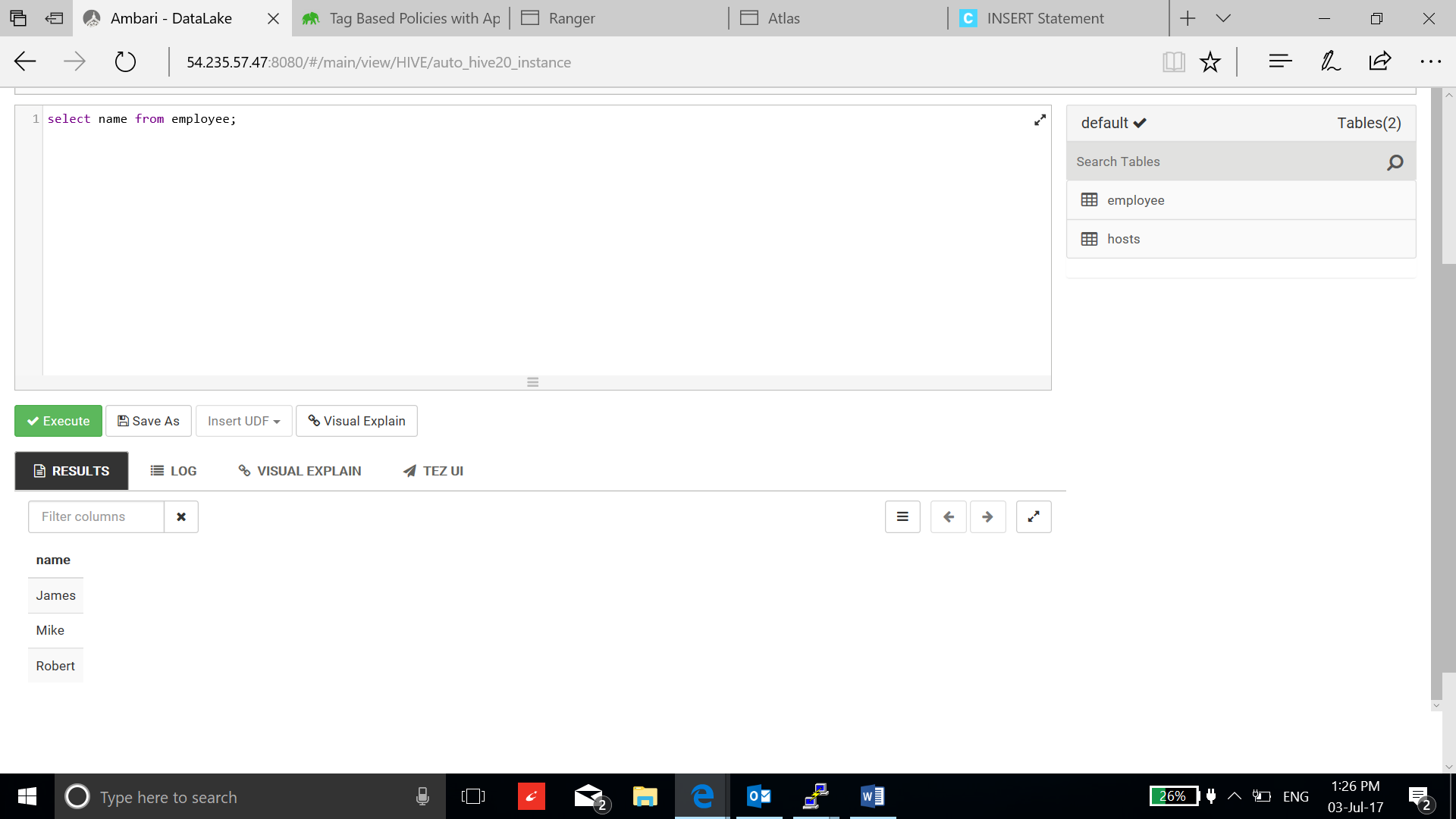


* Go to hive and execute select statement on employee table:

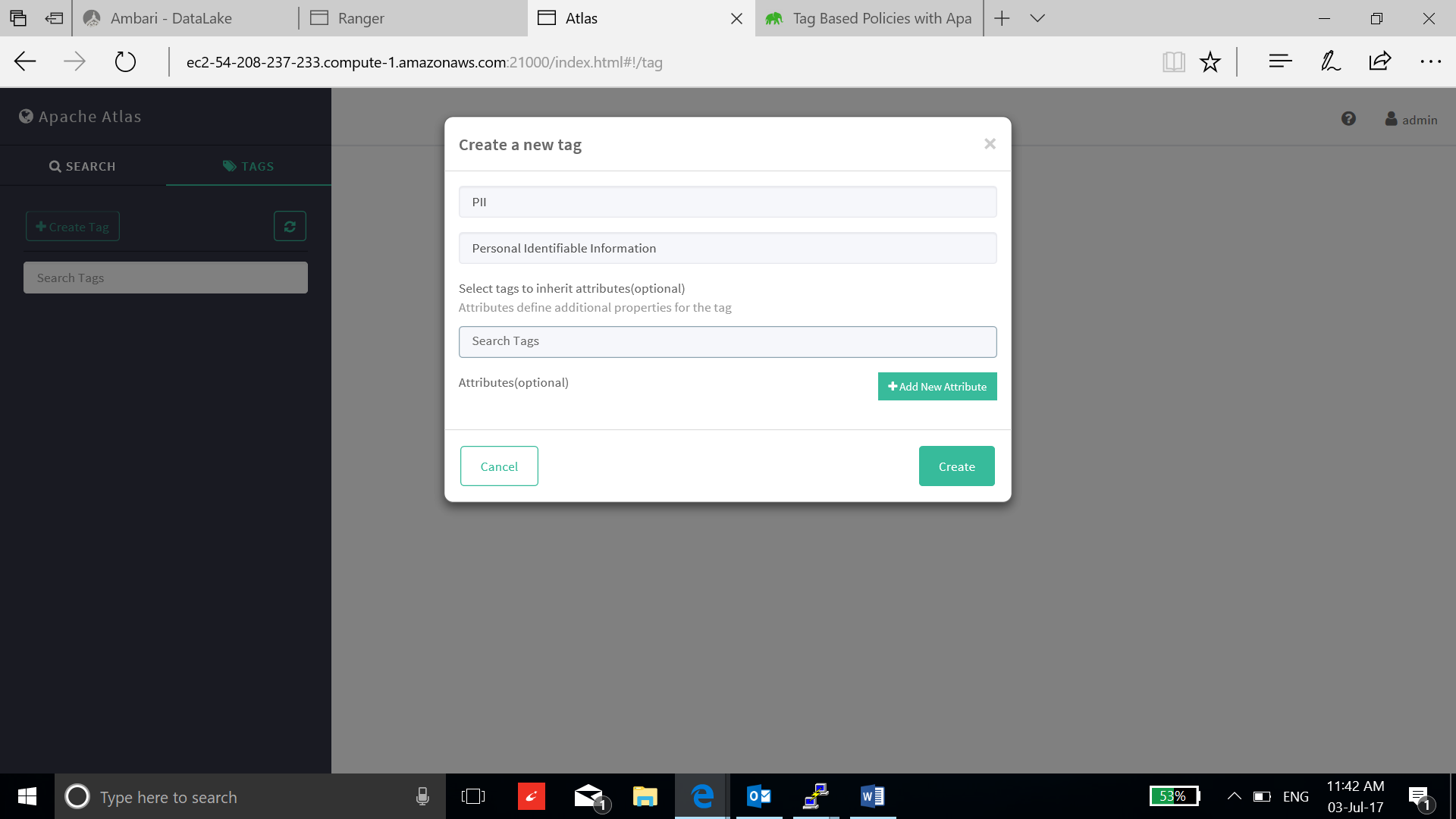
**select \* from employee;**



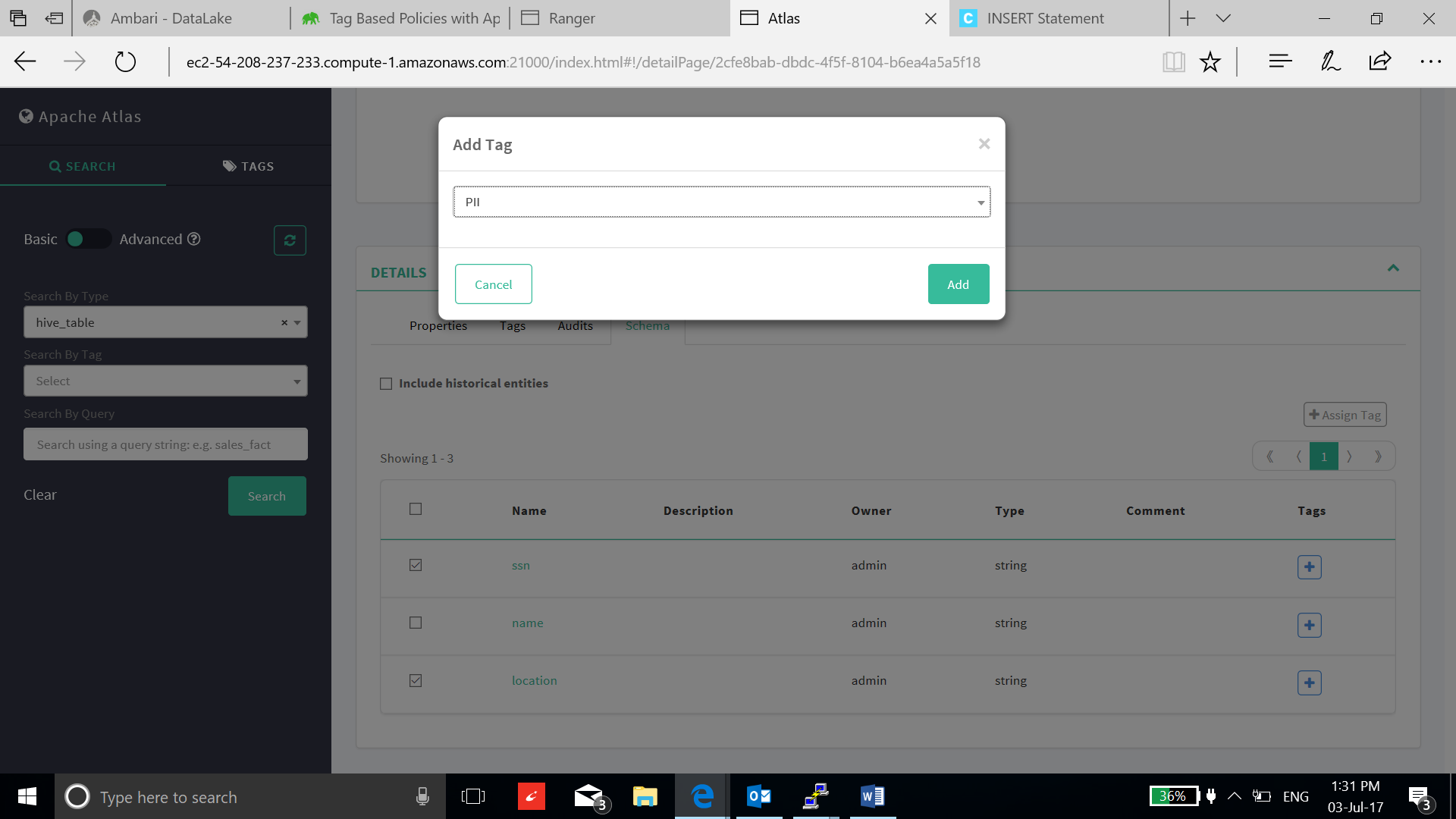
* Execute select statement on name:



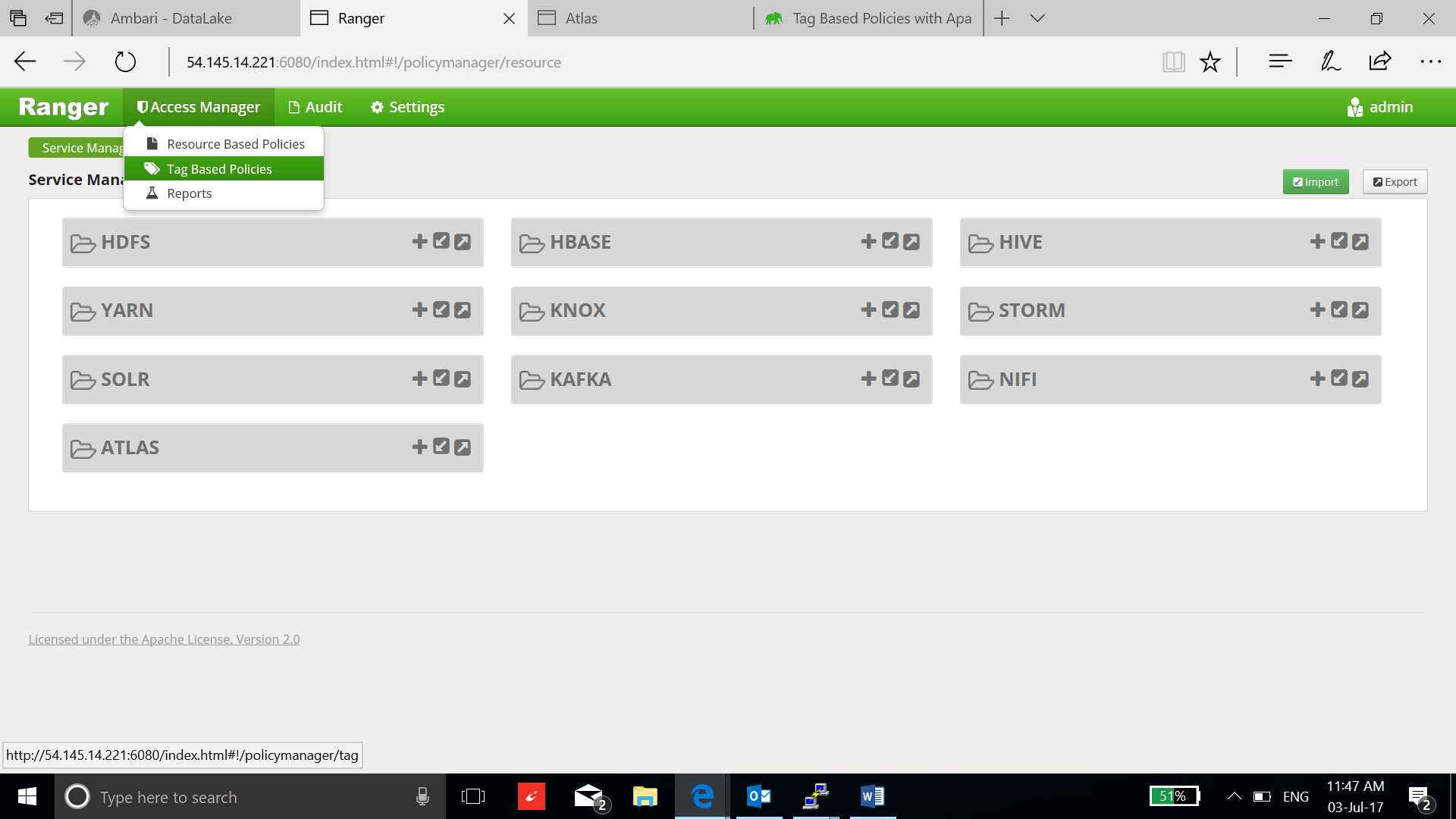
* Go to Atlas dashboard, go to tag tab and create a tag:



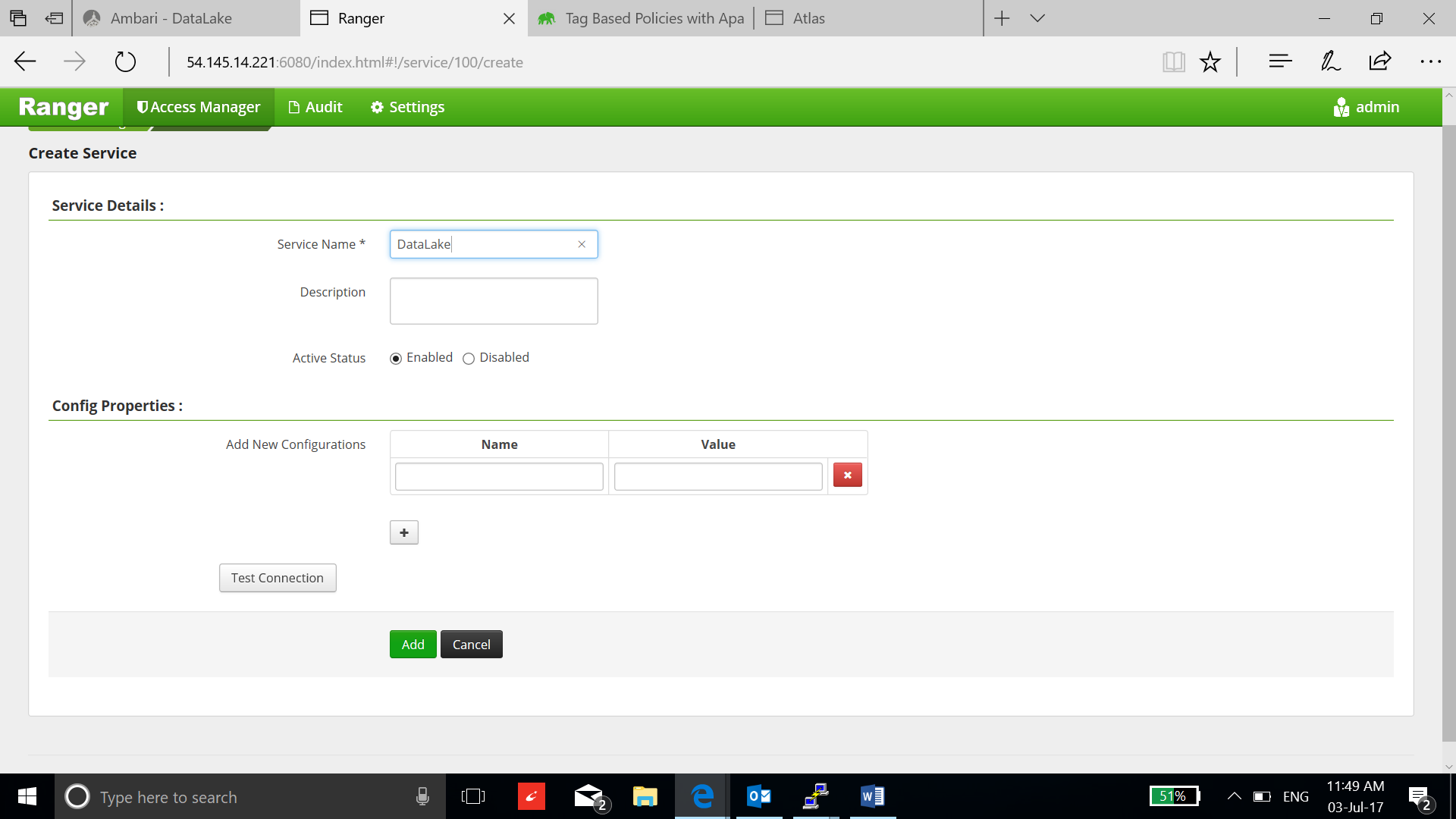
* Now search for hive\_tables and select the employee table. Go to schema and add tags to the ssn and location column using + tab:



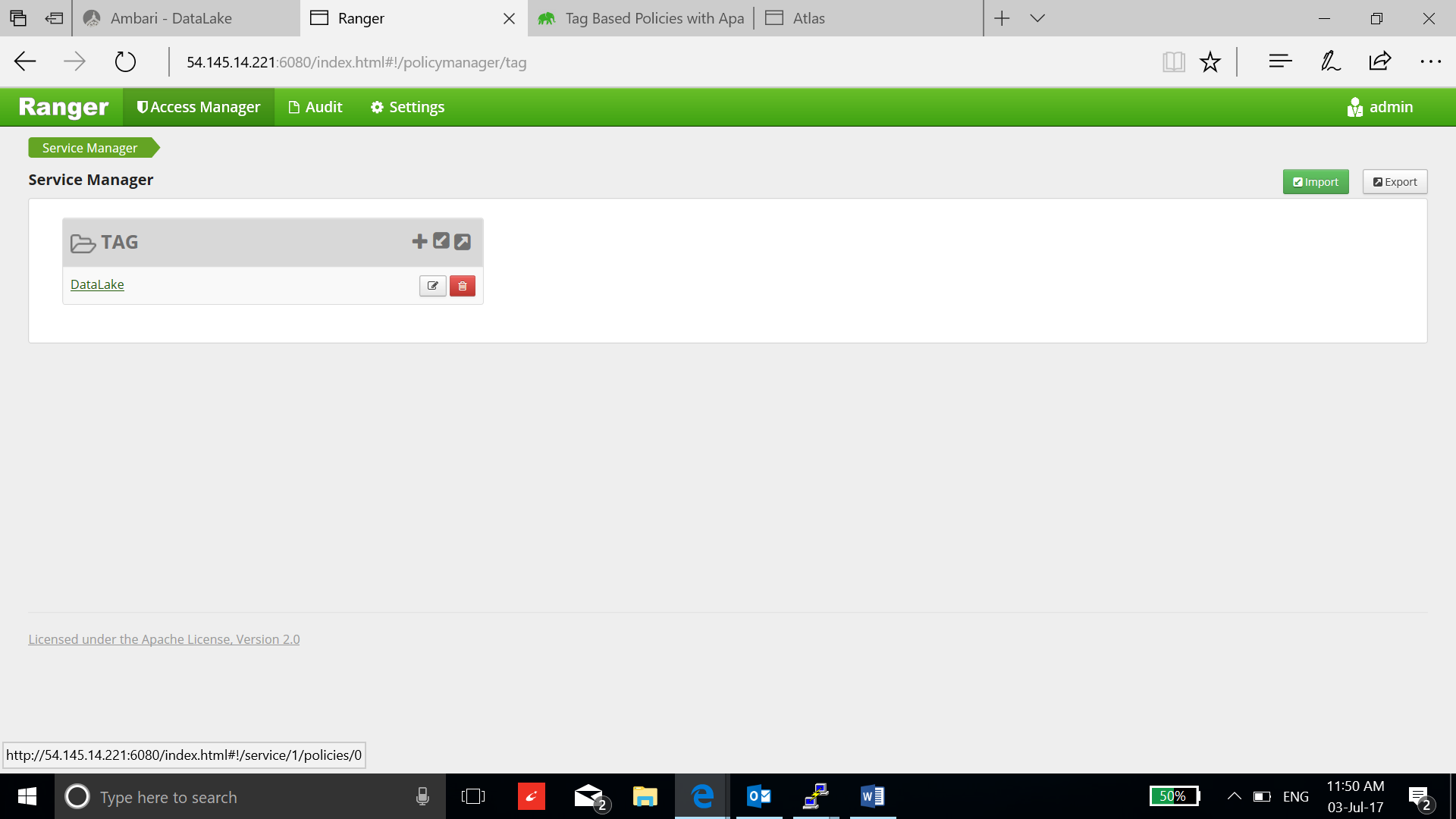
* Go to Ranger, click access manager and select the Tag Based Policies:



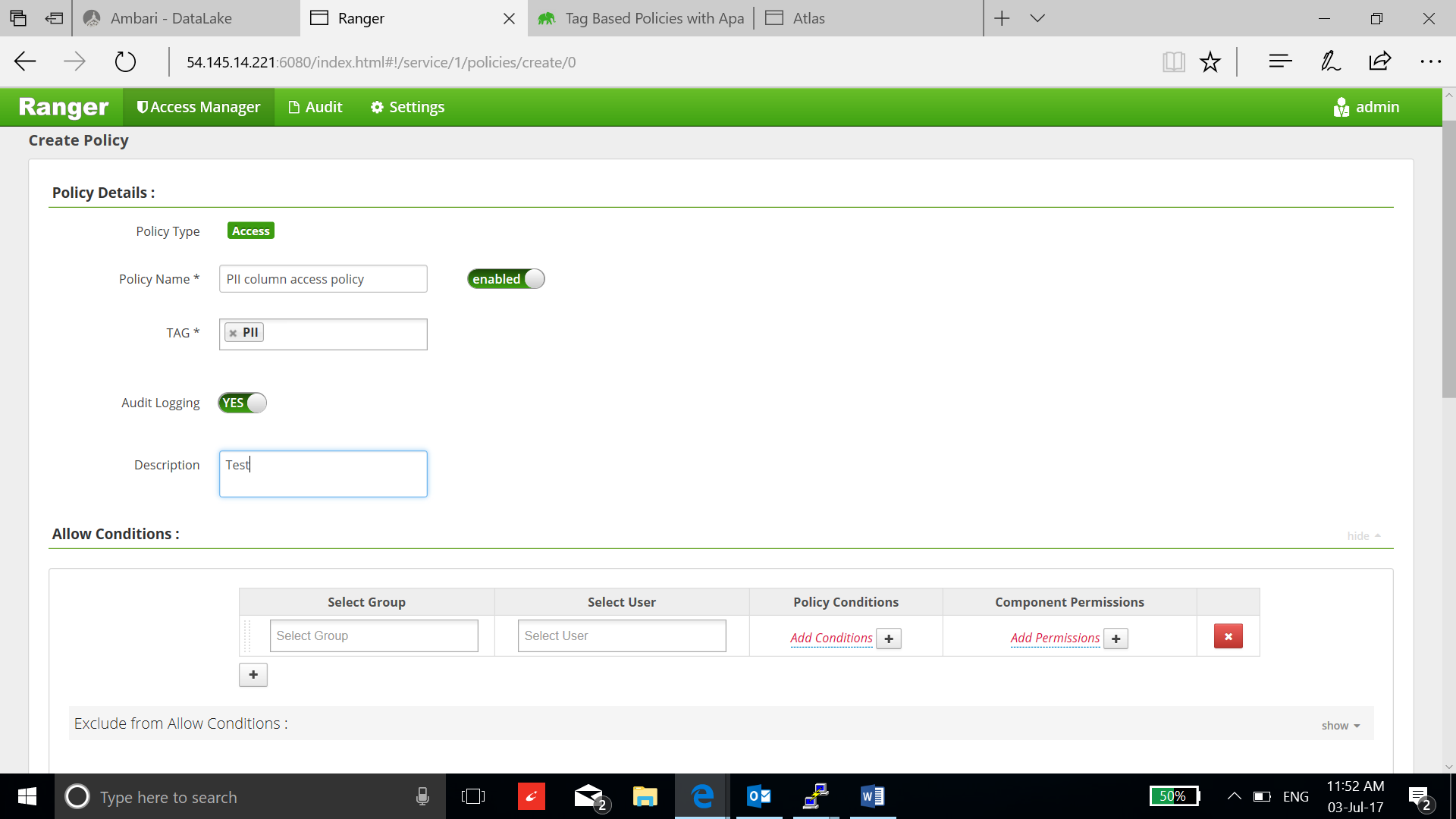
* Now click on + tab to add a tag:

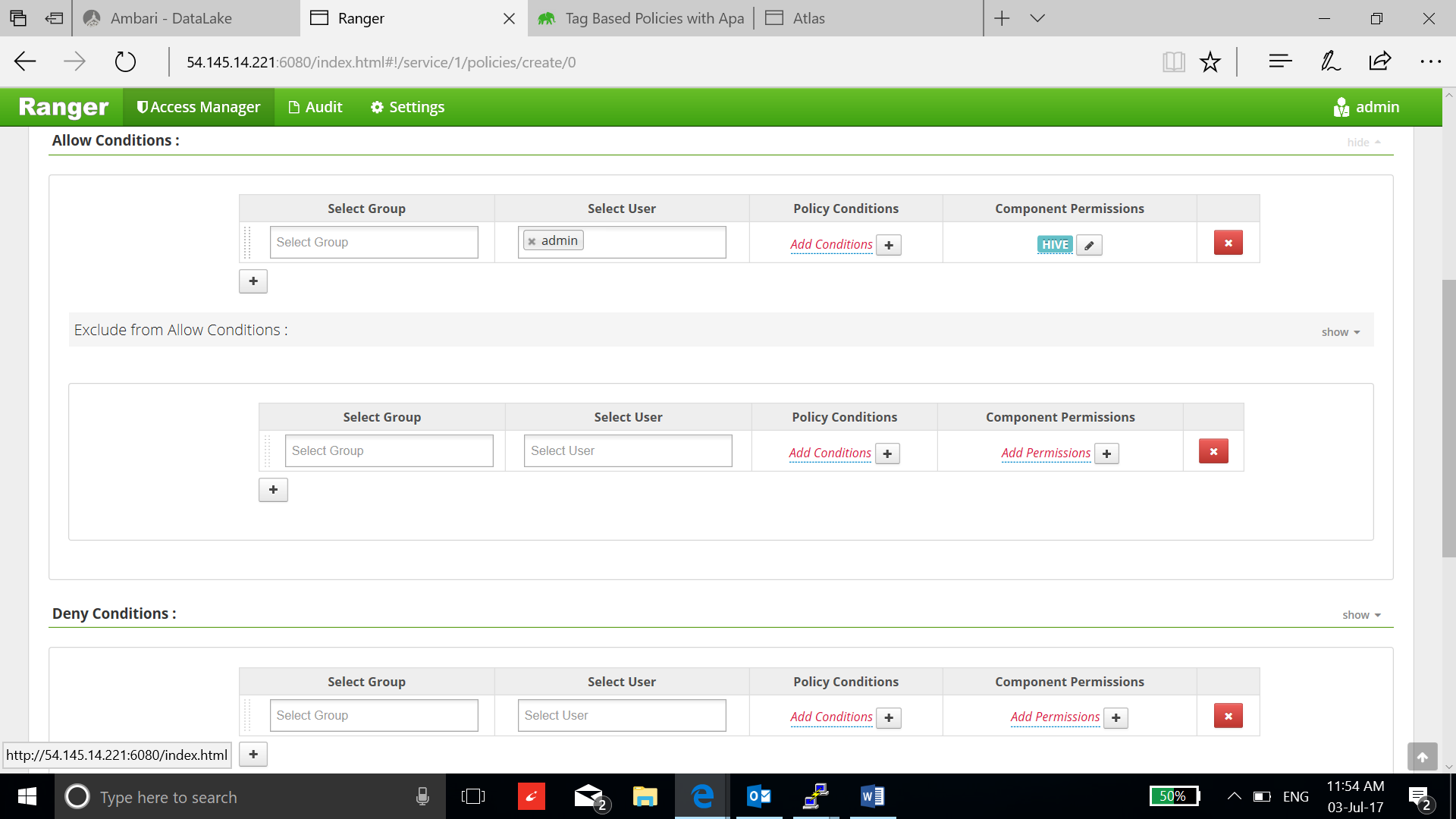


* Select the dataLake and add new policies:

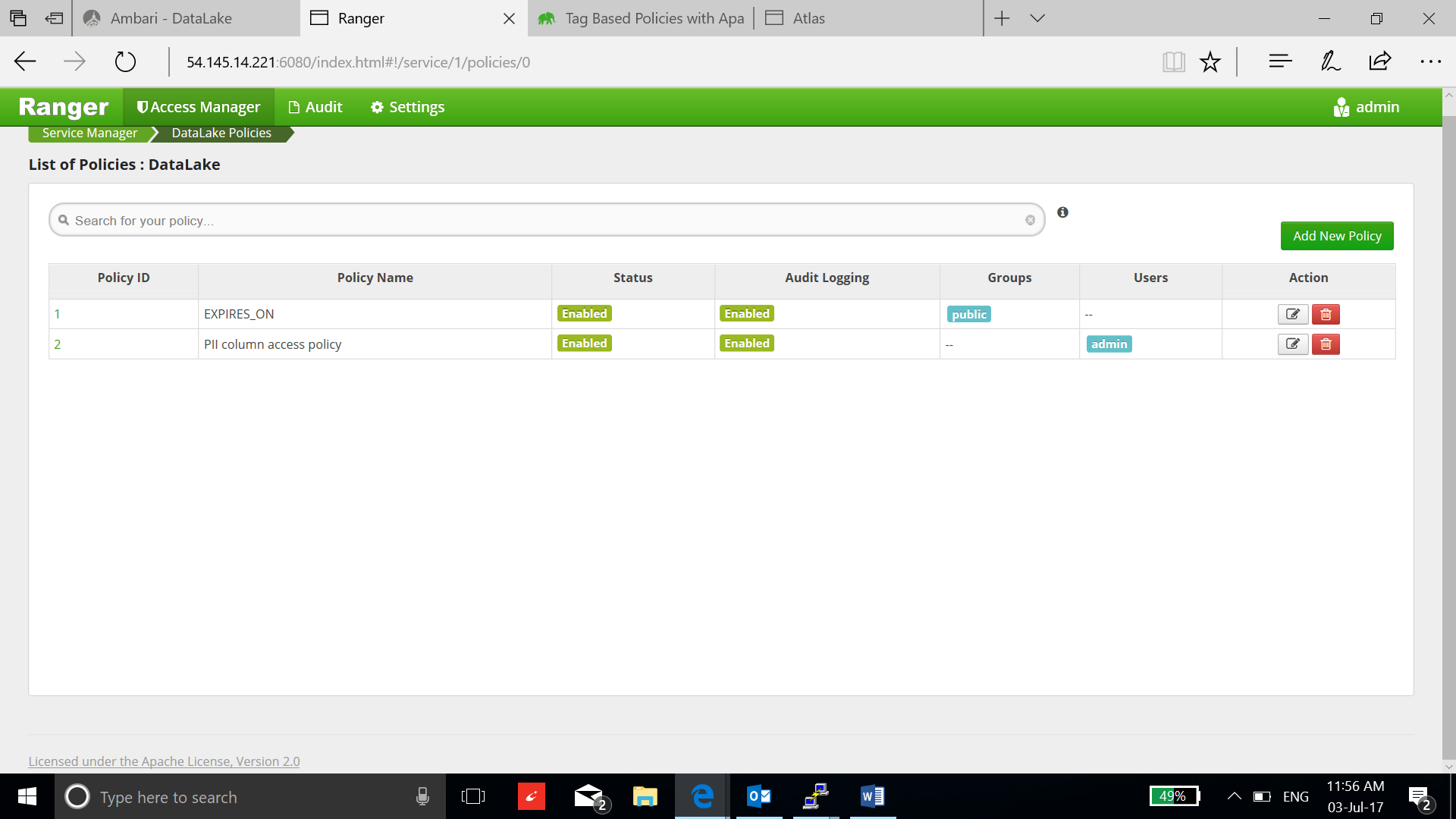


* Enter the details in the policy:

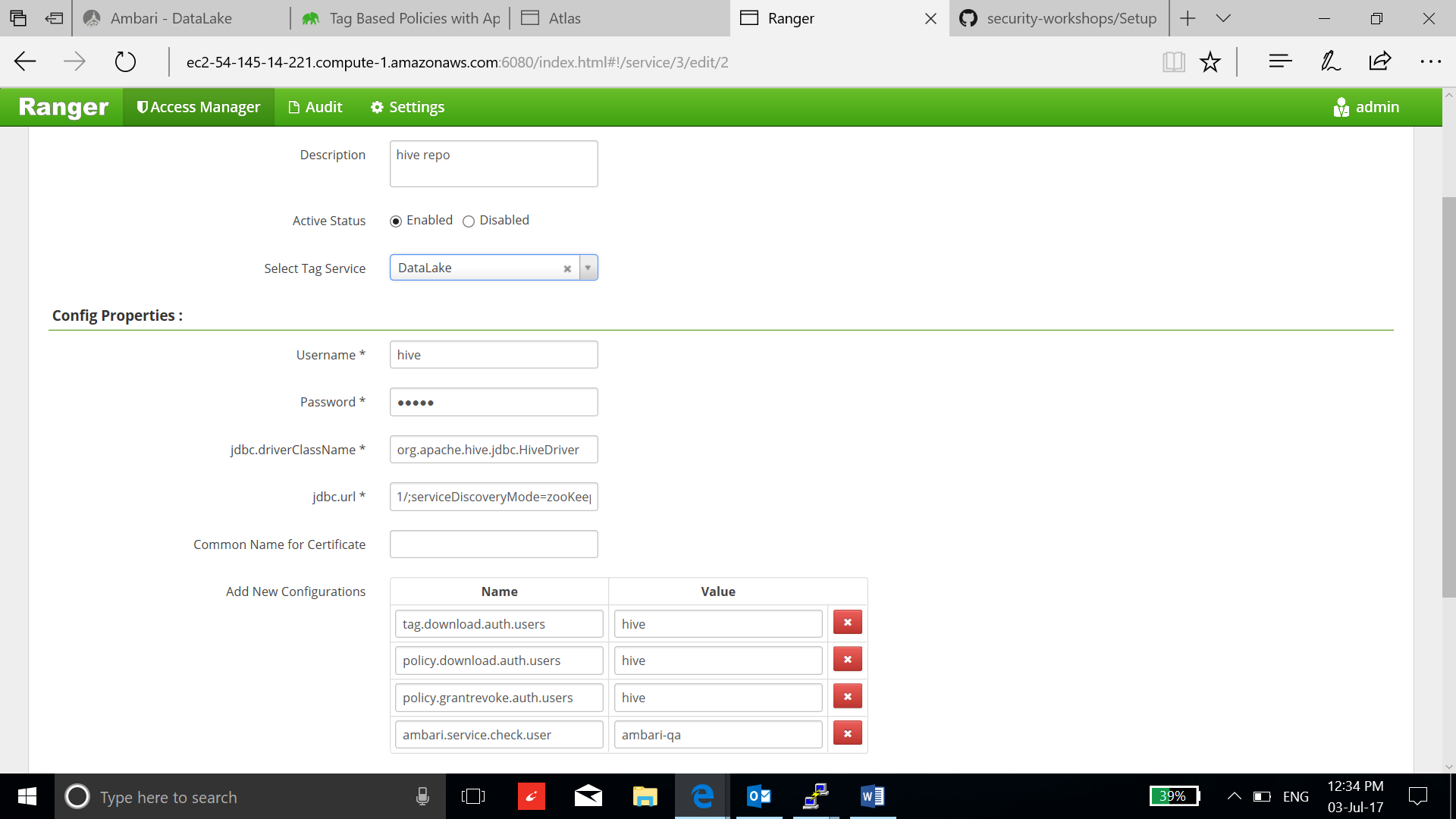




* This signifies that only admin user is allowed to do any operation on the columns that are specified by PII tag. Click Add



* Select Tag service as the once we appointed to FII tag and select save:



* The Ranger tag based policy is now enabled for **admin** user. You can test it by running the query on all columns in employee table.

