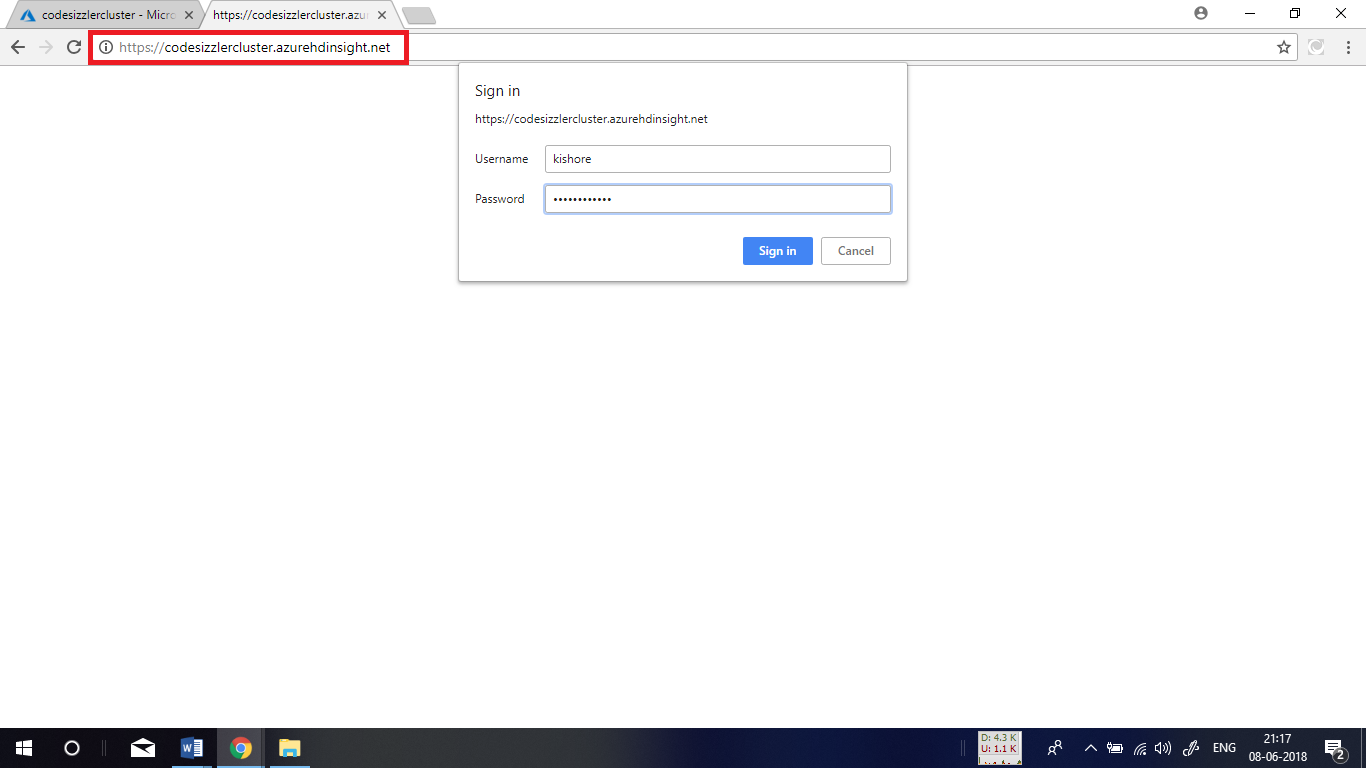
# **Managing HDInsight Cluster using Ambari Web UI**

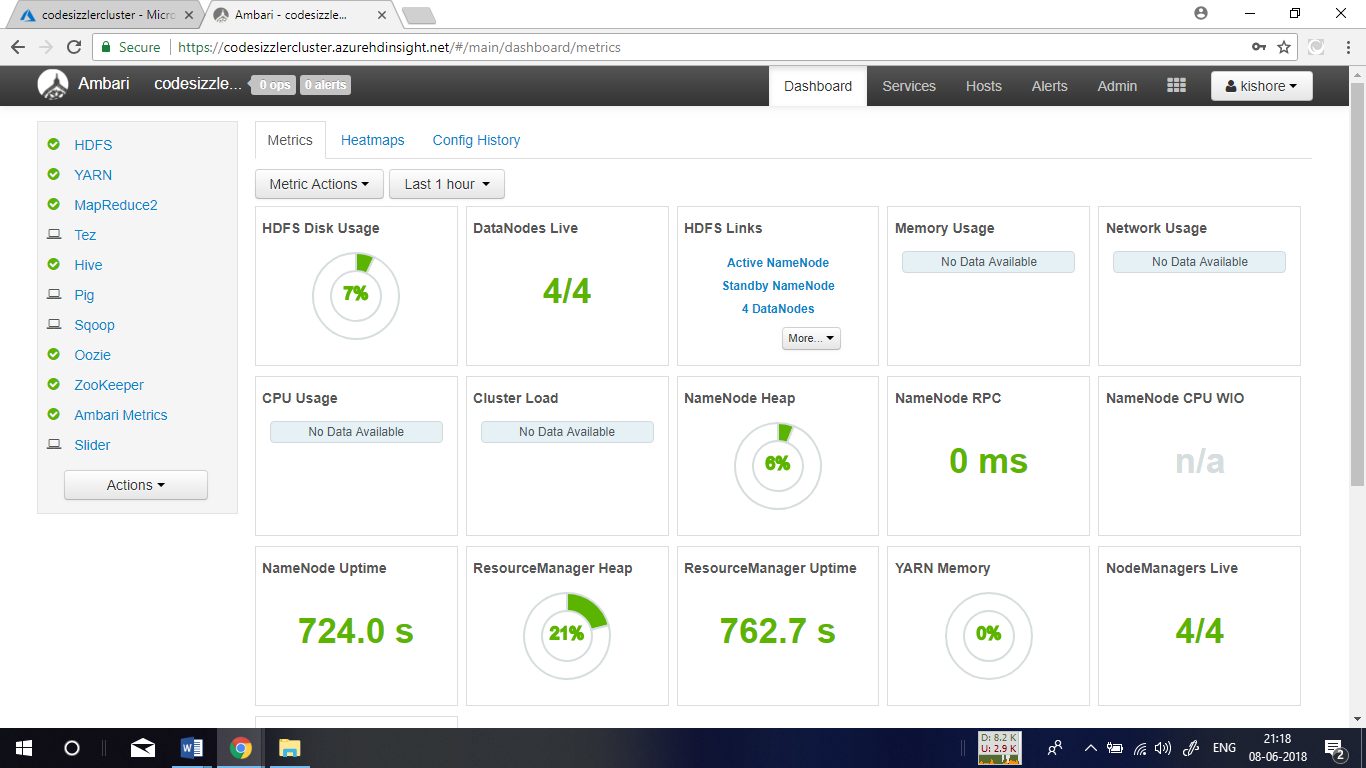
This demo explains about managing and monitoring HDInsight using Ambari Web Interface. For this, make sure that you have a HDInsight cluster deployed in Azure.

**Logging into Cluster:**

Open the link of the cluster that you created in a browser. It will open the login pane. Enter the username and password. This will take you the **Ambari** web UI.

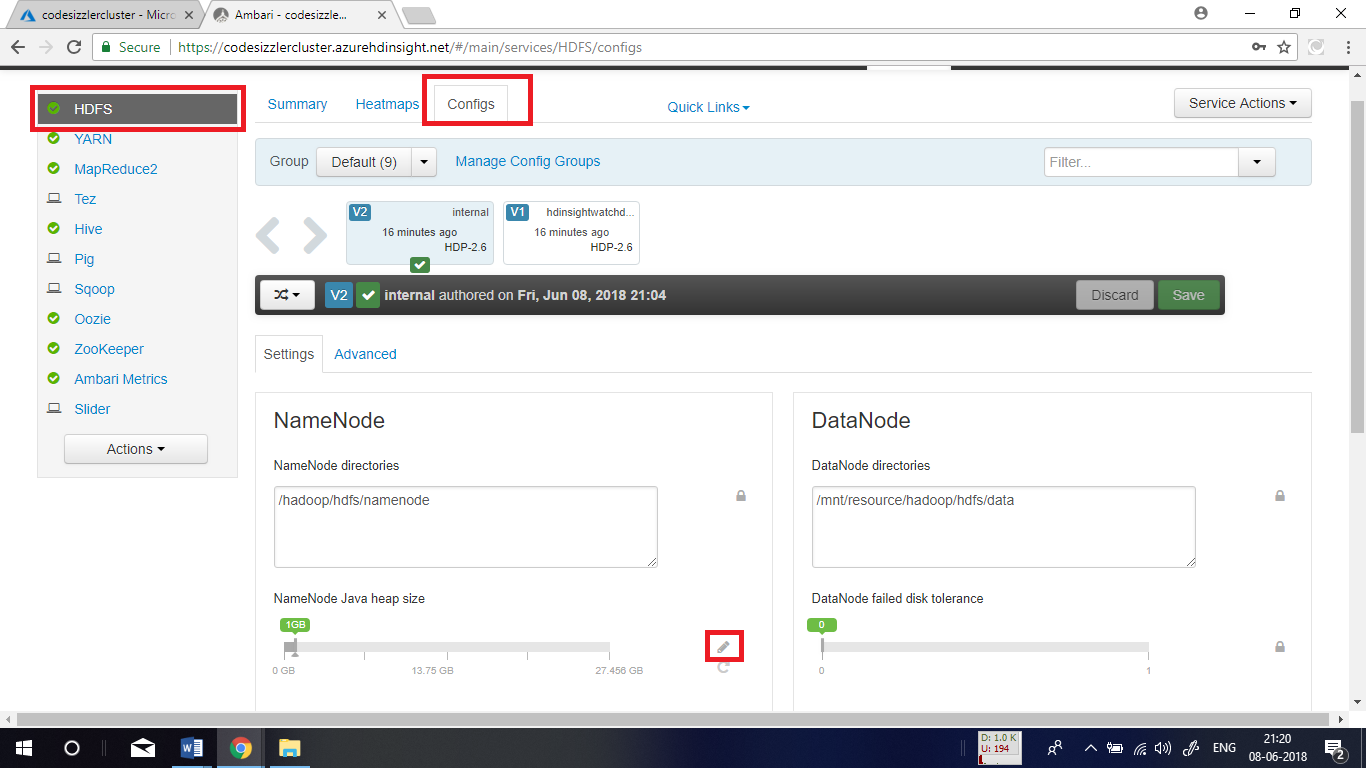


This is how the **Ambari UI** dashboard look like. In here, you can manage and monitor all your resources.

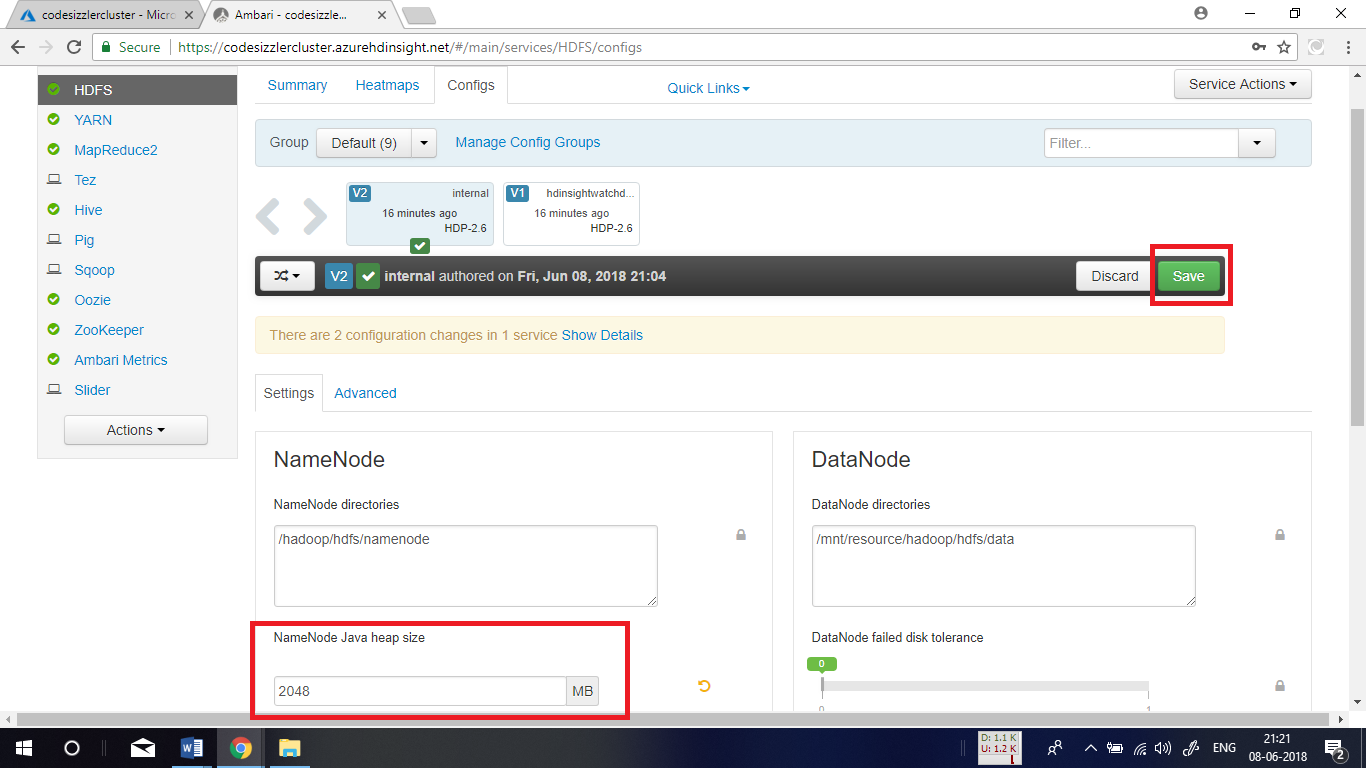


**Modifying NameNode Java Heap Size:**

The NameNode Java heap size depends on many factors such as the load on the cluster, the numbers of files, and the numbers of blocks. The default size of 1 GB works well with most clusters, although some workloads can require more or less memory. To modify the NameNode Java heap size Select **HDFS** from the Services sidebar and navigate to the **Configs** tab. Find the setting **NameNode Java heap size**. Select the **pen** icon beside the setting name.

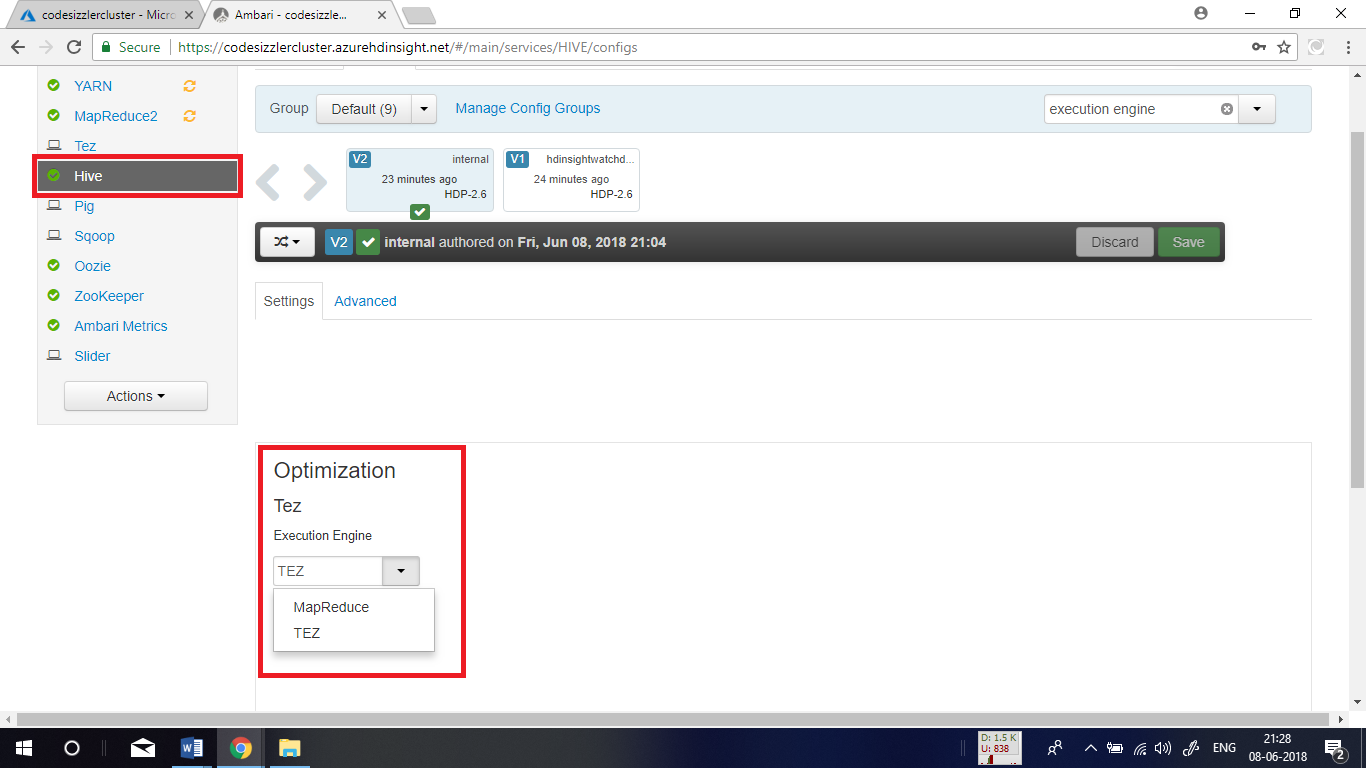


This will allow you to change the size of heap. Change the value in there and click on save button. You might have to restart the cluster after saving the changes if it prompts you to.



**Changing Hive Execution Engine:**

By default, HDInsight cluster is equipped with **Tez** as its default execution engine which works faster than **MapReduce**. To optimize Hive changing its execution engine. To change the configuration, in the **Hive execution engine**, go to **Hive** menu in the left side and change the configuration to desired one and click **Save.**



**Tuning Mappers:**

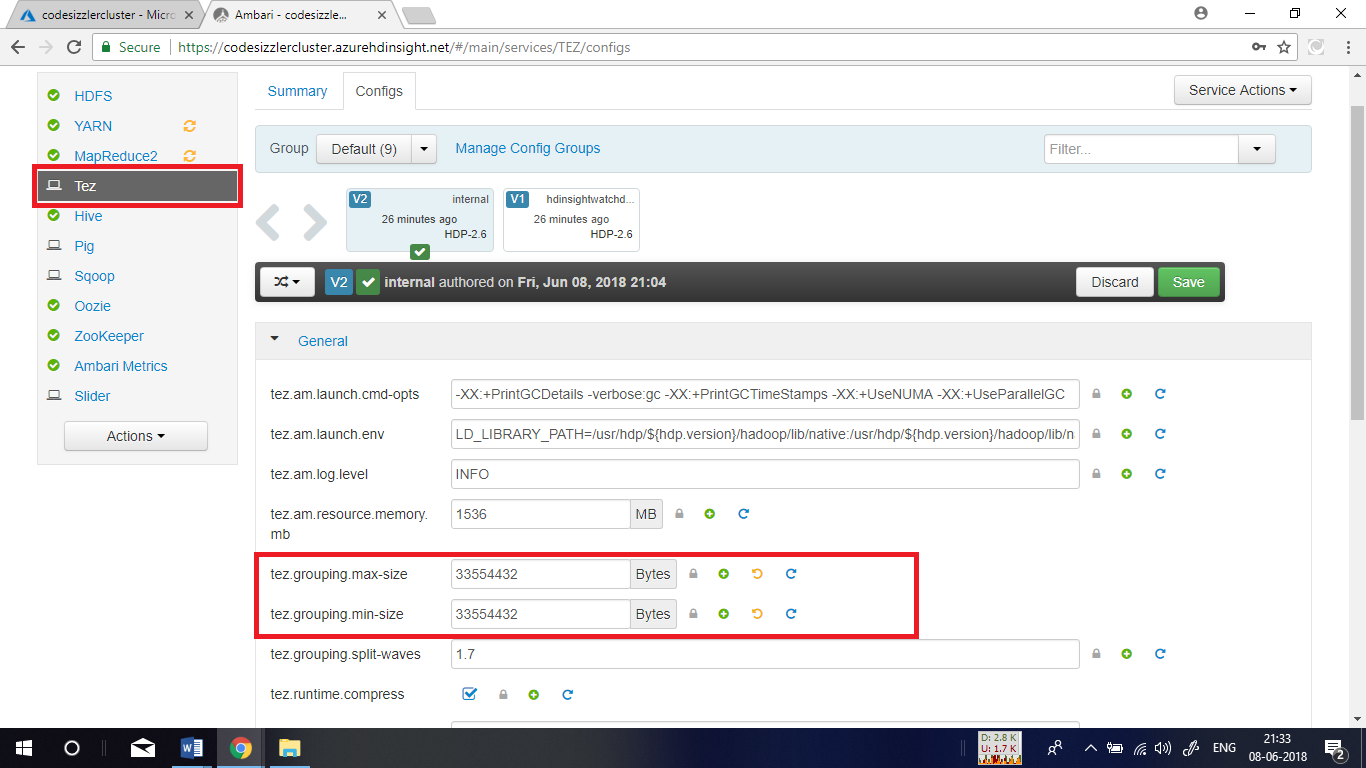
Hadoop tries to split (*map*) a single file into multiple files and process the resulting files in parallel. The number of mappers depends on the number of splits. The following two configuration parameters drive the number of splits for the Tez execution engine:

* tez.grouping.min-size: Lower limit on the size of a grouped split, with a default value of 16 MB (16,777,216 bytes).
* tez.grouping.max-size: Upper limit on the size of a grouped split, with a default value of 1 GB (1,073,741,824 bytes).

As a performance rule of thumb, decrease both of these parameters to improve latency, increase for more throughput.

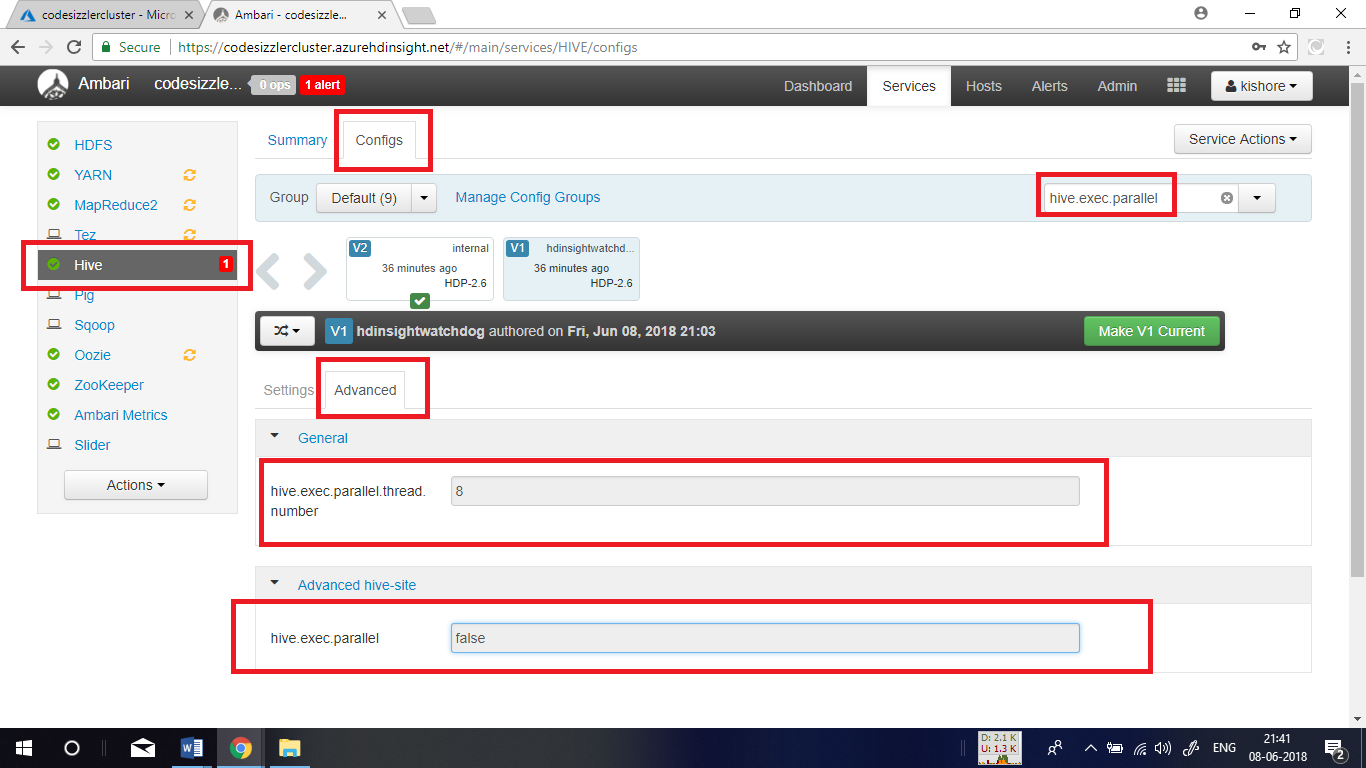
For example, to set four mapper tasks for a data size of 128 MB, you would set both parameters to 32 MB each (33,554,432 bytes).

To modify these limit parameters, go to **Tez** in the left side menu and under configs option, find the **tez.grouping.max**-size and **tez.grouping.min-size** parameters. There, set both the parameters as **33554432** bytes which is equivalent to 32 MB. These changes affect all Tez jobs across the server. To get an optimal result, choose appropriate parameter values.



**Enabling Parallel Execution:**

A Hive query is executed in one or more stages. If the independent stages can be run in parallel, that will increase query performance. To enable parallel query execution, navigate to the Hive **Config** tab and search for the **hive.exec.parallel** property. The default value is false. Change the value to true, and then press **Enter** to save the value. To limit the number of jobs to be run in parallel, modify the **hive.exec.parallel.thread.number** property. The default value is 8.



**Pig Optimization:**

Pig properties can be modified from the Ambari web UI to tune Pig queries. Modifying Pig properties from Ambari directly modifies the Pig properties in the **/etc/pig/2.4.2.0-258.0/pig.properties** file.To modify Pig properties, navigate to the Pig **Configs** tab, and then expand the **Advanced pig-properties** pane.Find, uncomment, and change the value of the property you wish to modify.Select **Save** on the top right side of the window to save the new value. Some properties may require a service restart.

