**KERBEROS vs NON-KERBEROS:**

What is Kerberos:

* Kerberos secures Hadoop cluster.
* Kerberos Authentication is a cryptographic network authentication system developed at (and by) MIT in the 1980s. It's one answer to the age-old network manager's utopian dream of "single network sign-on" (login).
* Let’s first understand what Kerberos is!. It is an authentication protocol which uses “tickets” to allow nodes to identify themselves.

Talking about how it secures the cluster, it is pretty much similar to your building’s security guard checking the sticker on your car before you enter the premises and make sure that you are who you are and authorize your entry.

Hadoop can use the Kerberos protocol to ensure that when someone makes a request, they really are who they say they are. This mechanism is used throughout the cluster.

In a secure Hadoop configuration, all of the Hadoop daemons use Kerberos to perform mutual authentication, which means that when two daemons talk to each other, they each make sure that the other daemon is who it says it is.

Additionally, this allows the NameNode and ResourceManager/JobTracker to ensure that any HDFS or MR requests are being executed with the appropriate authorization level.

* First Look at this term Authentication and Authorization:

Authentication is the process of determining whether someone is who they claim to be.

Authorization is the function of specifying access rights to resources.

Hadoop has the ability to require authentication, in the form of Kerberos principals. Kerberos is an authentication protocol which uses “tickets” to allow nodes to identify themselves.

Kerberos protocol ensures that when someone makes a request, they really are who they say they are. This mechanism is used throughout the cluster. In a secure Hadoop configuration, all of the Hadoop daemons use Kerberos to perform mutual authentication, which means that when two daemons talk to each other, they each make sure that the other daemon is who it says it is. Additionally, this allows the NameNode and JobTracker to ensure that any HDFS or MR requests are being executed with the appropriate authorization level.

**How do I configure Hadoop to work with Kerberos authentication?**

* You can use Cloudera Manager to setup Hadoop with Kerberos: It's a fairly complicated process and it's not just about enabling passwordless ssh between nodes.