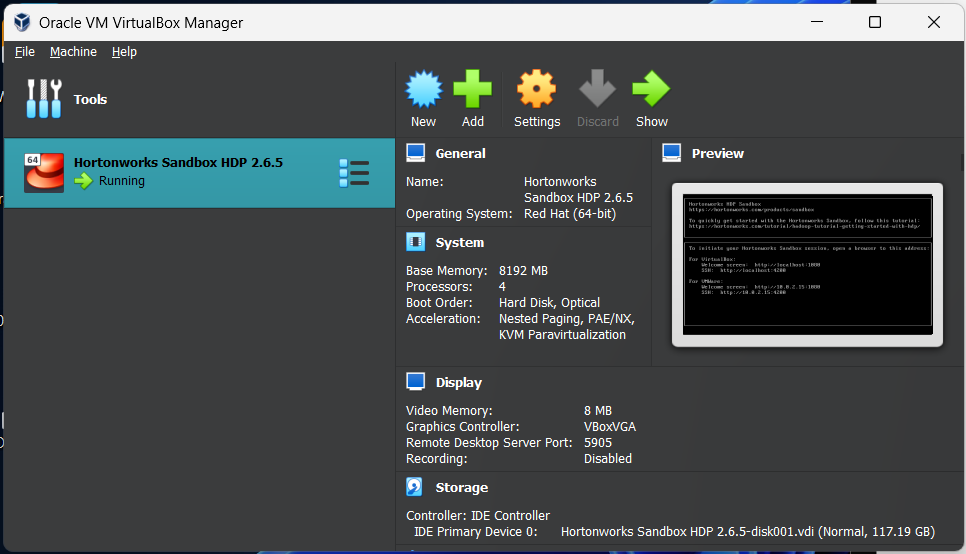
1. Set up hadoop cluster

* Download Hortonworks-sandbox 2.6.5 from <https://www.cloudera.com/downloads/hortonworks-sandbox/hdp.html>
* Dowload Virtualbox and install
* Install Hortonworks-sandbox 2.6.5 to Virtualbox



* 1. Set up Spark Stream to take raw transaction data , run Predict Model and save to HDFS



1. Read transaction from Kafka Topic
2. Transaction go to Predict Model to classify fraud or legit – if this transaction is likelihood fraudulent, use Kafka to notify user
3. Save transaction data to HDFS
   * 1. Set up Kafka

* Create topic:

Go to folder bin of kafka  
cd /usr/hdp/current/kafka-broker/bin   
Now we can create topic

./kafka-topics.sh --create --zookeeper sandbox-hdp.hortonworks.com:2181 --replication-factor 1 --partitions 1 --topic transactions

* + 1. Run Spark Stream
* Download spark-streaming-kafka-assembly 2.10 jar in link:

<https://search.maven.org/artifact/org.apache.spark/spark-streaming-kafka-assembly_2.10/1.6.0/jar>

linus command:

wget <https://search.maven.org/remotecontent?filepath=org/apache/spark/spark-streaming-kafka-assembly_2.10/1.6.0/spark-streaming-kafka-assembly_2.10-1.6.0.jar>

* Download streamingTransactionAnalysis.py in our team github
* Run spark streaming to analysis

spark-submit –jars spark-streaming-kafka-assembly\_2.10-1.5.2.jar /streamingTransactionAnalysis.py

* New transactionChecked.csv is created and ready to visualize.