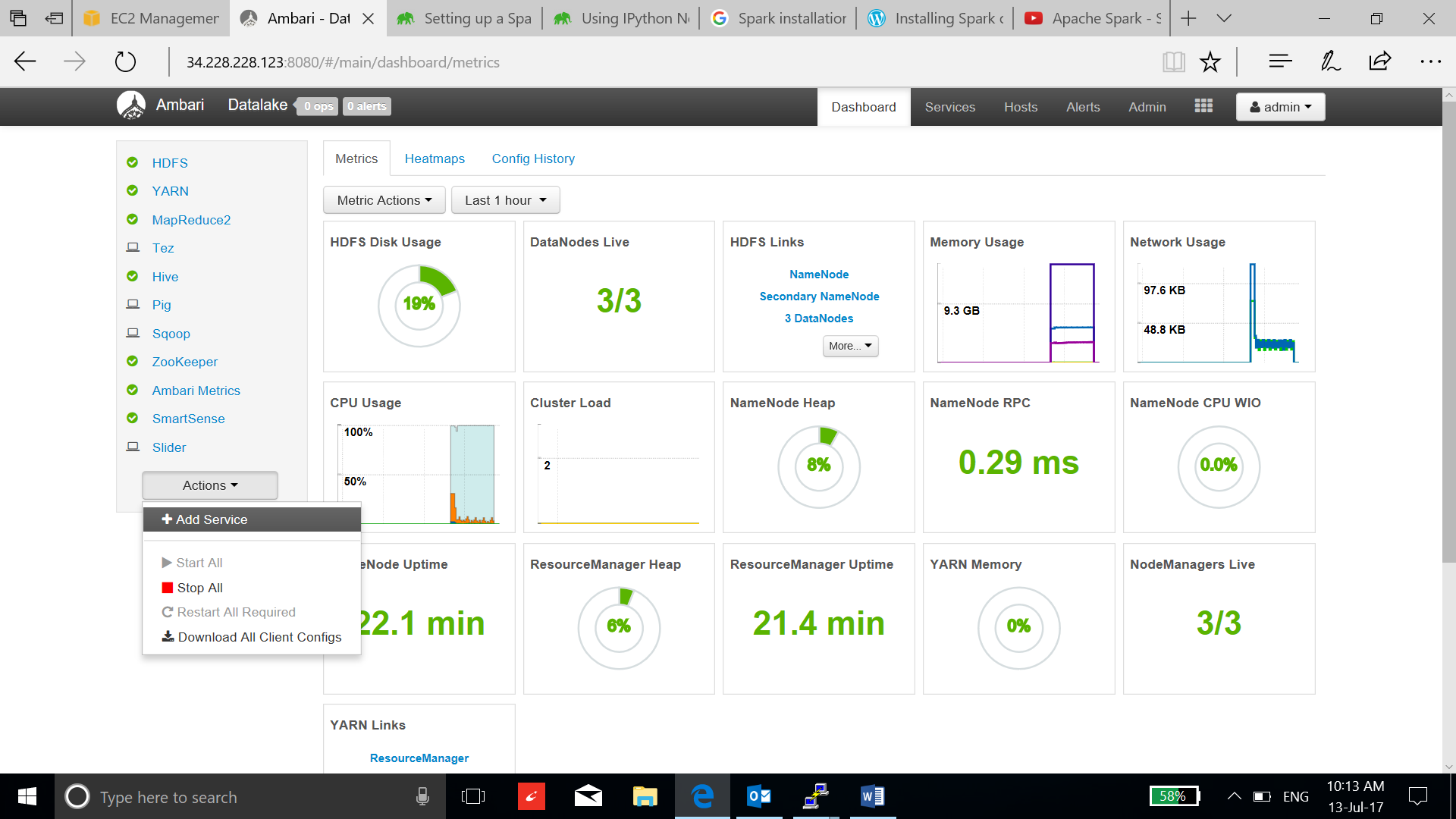
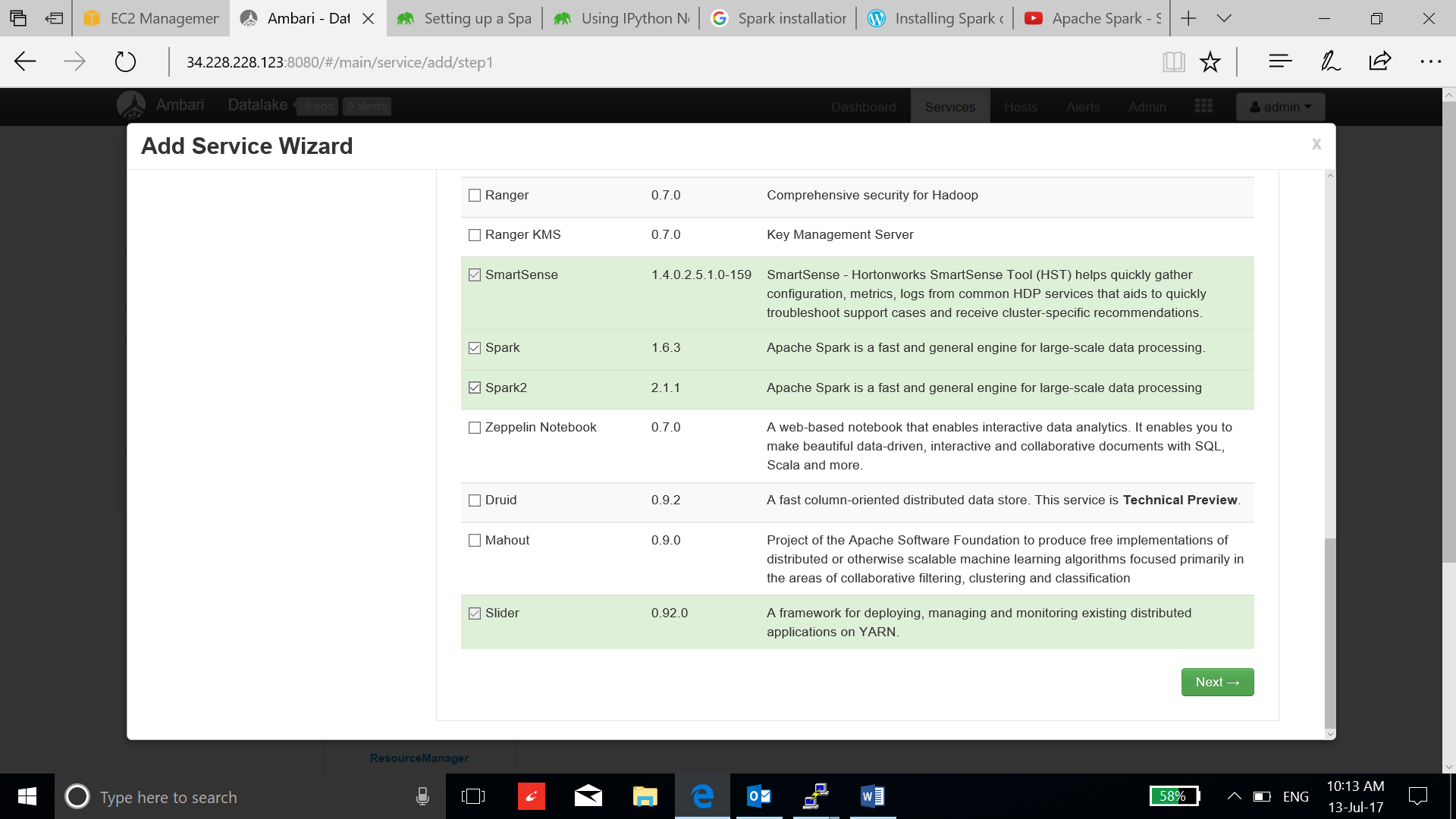
Steps to setup Spark:

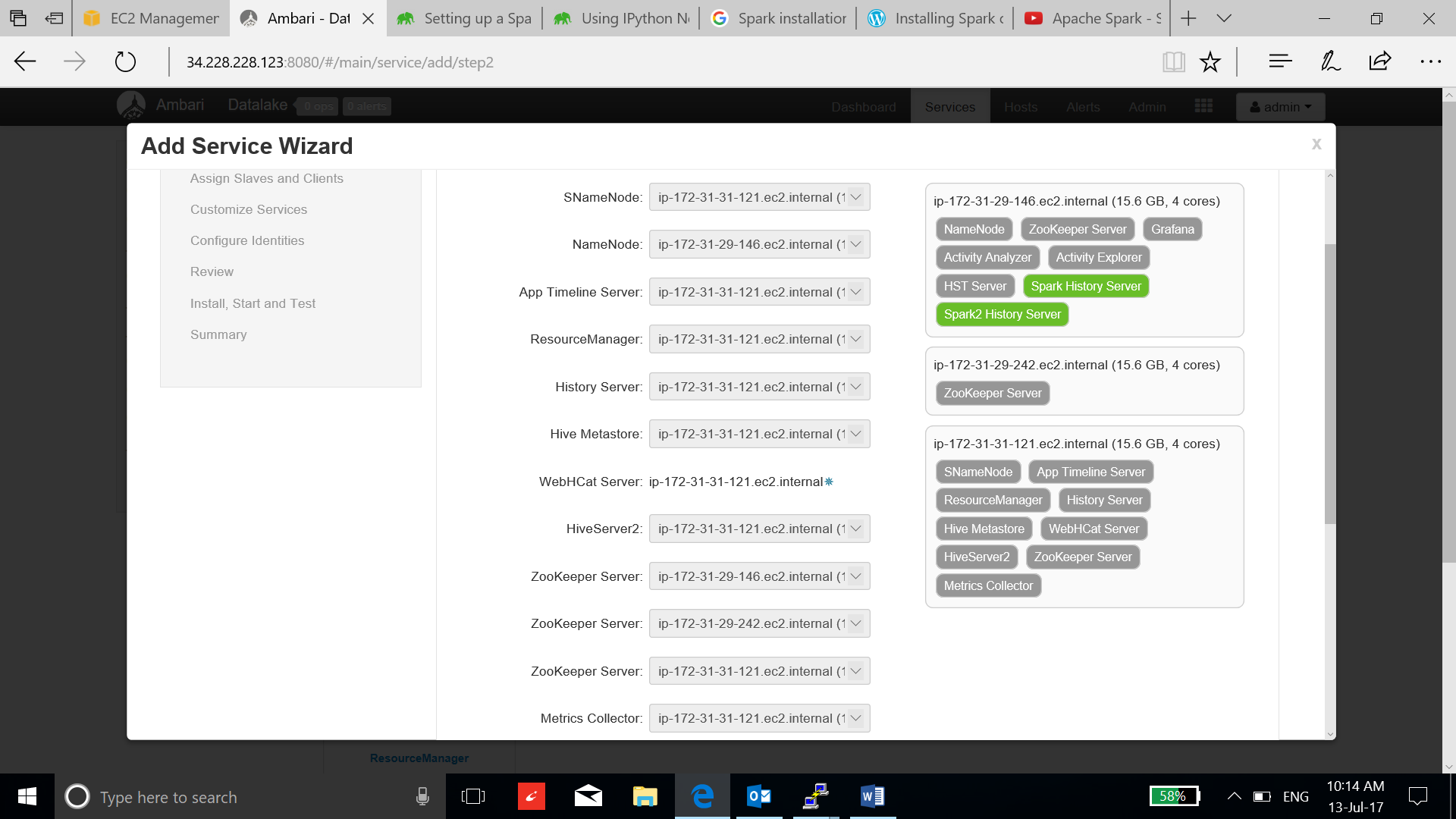
1. Go to Ambari Dashboard and select Add services:



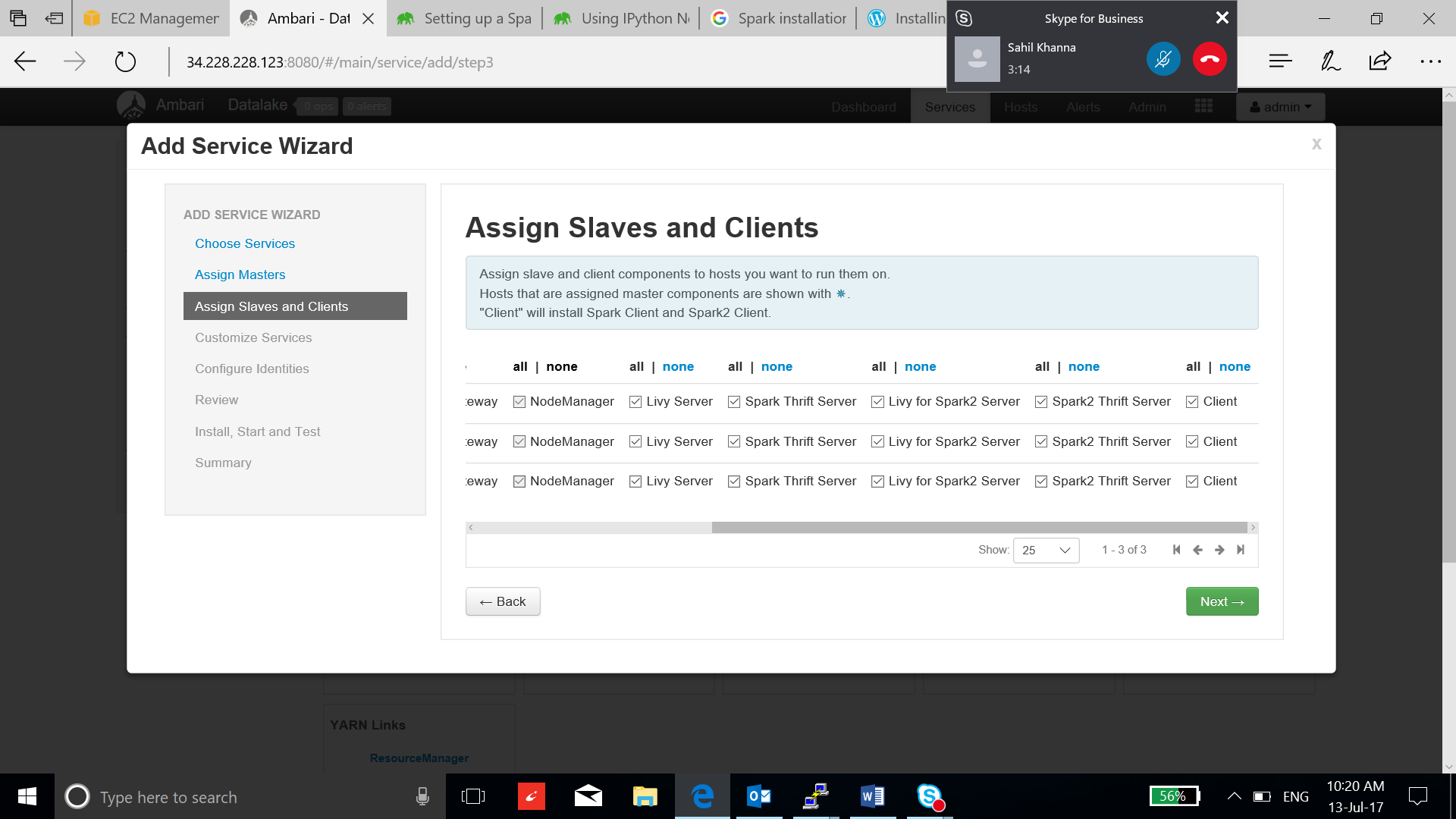
1. Select the spark services and click next:



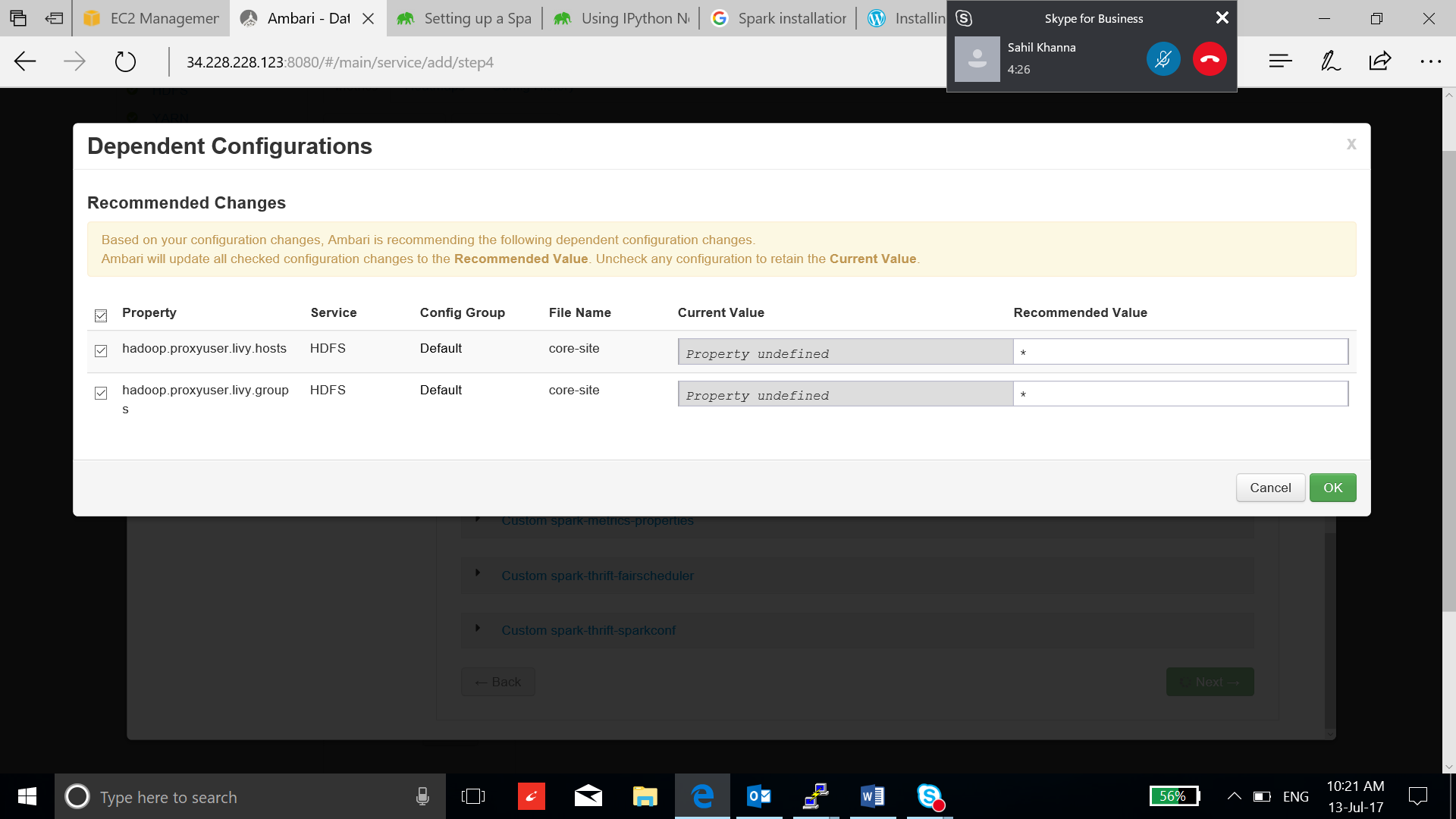
1. Select the Master node:



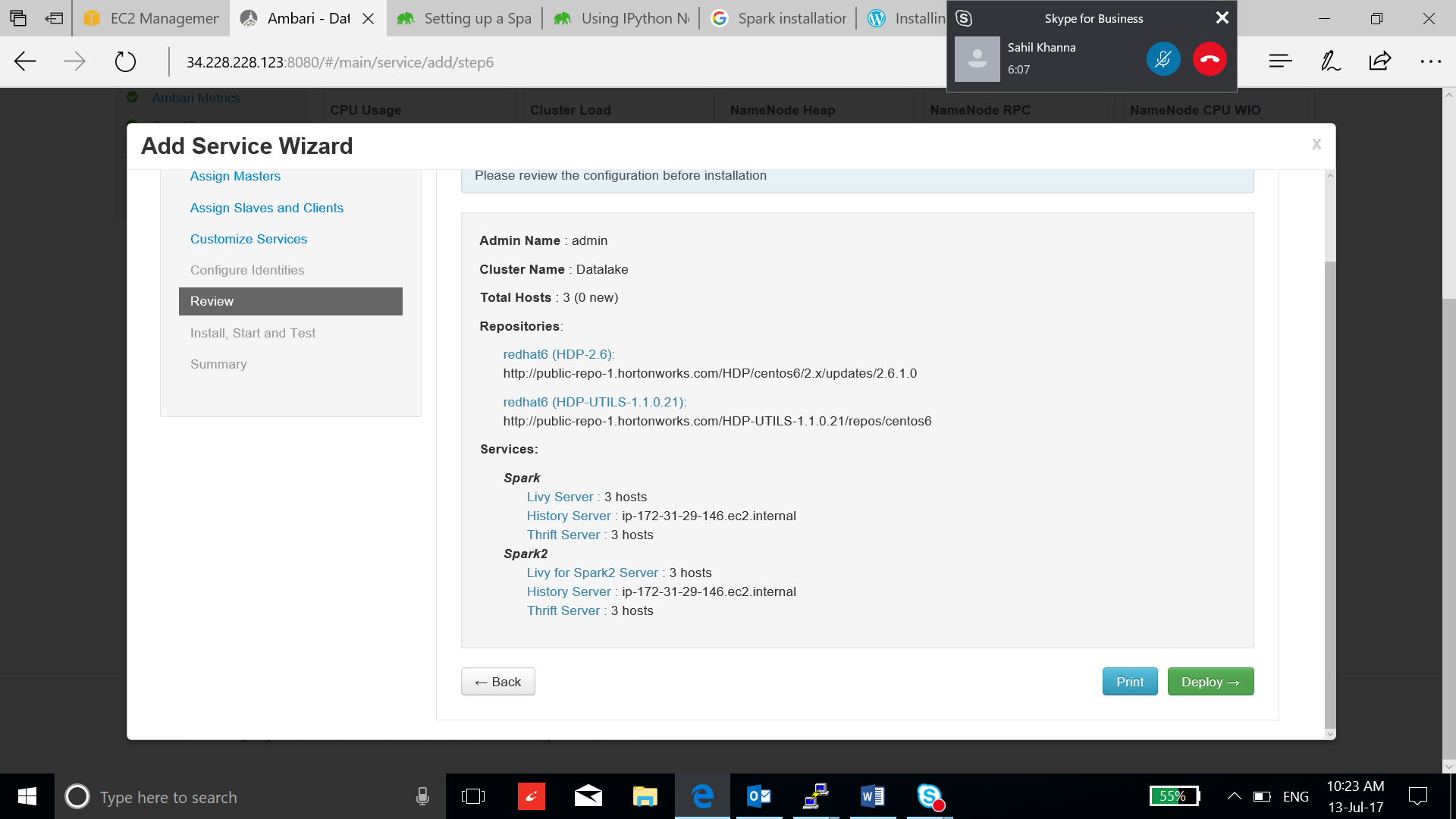
1. Add clients:



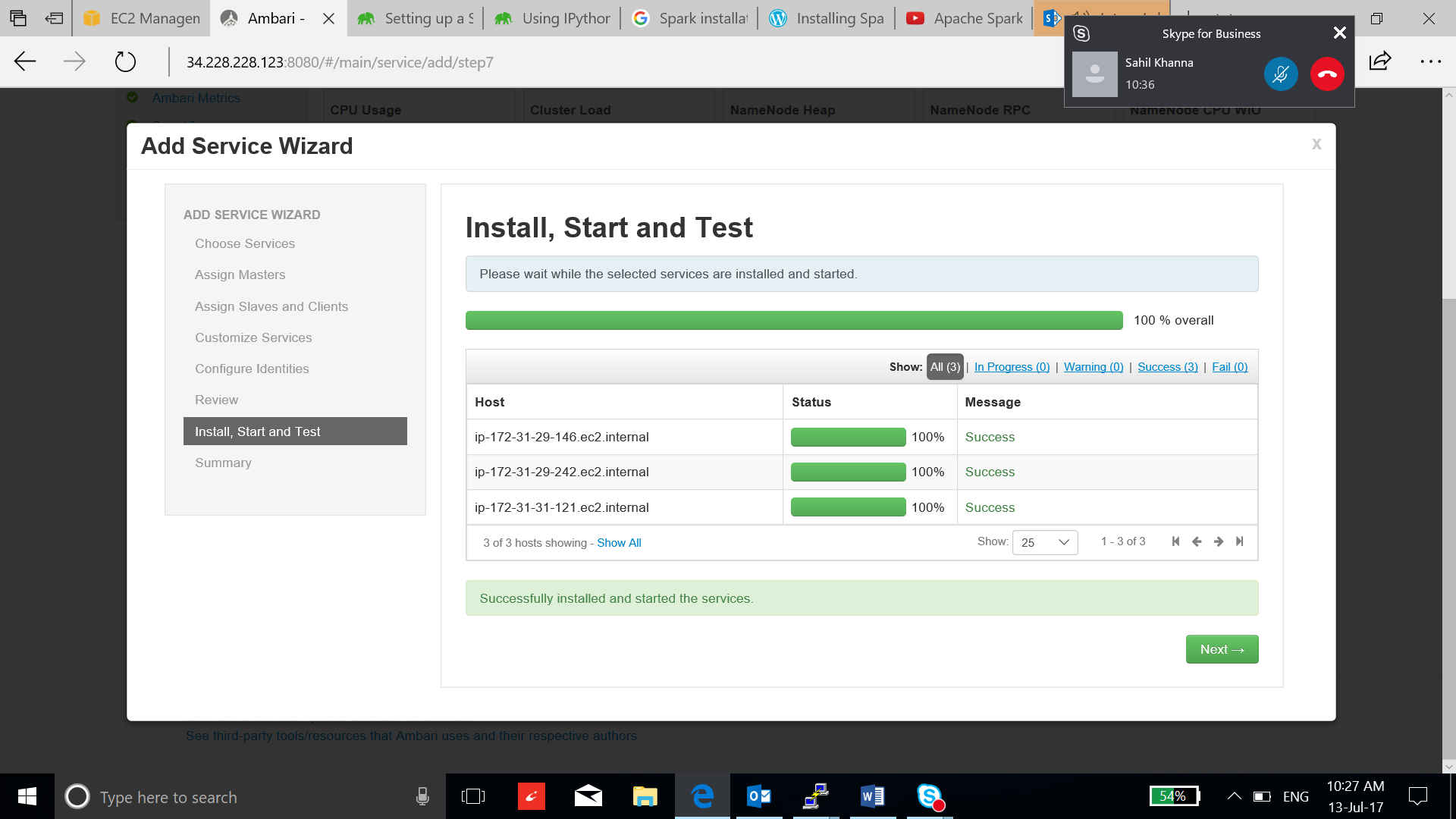
1. Keep the default configuration and accept the suggested recommendations:



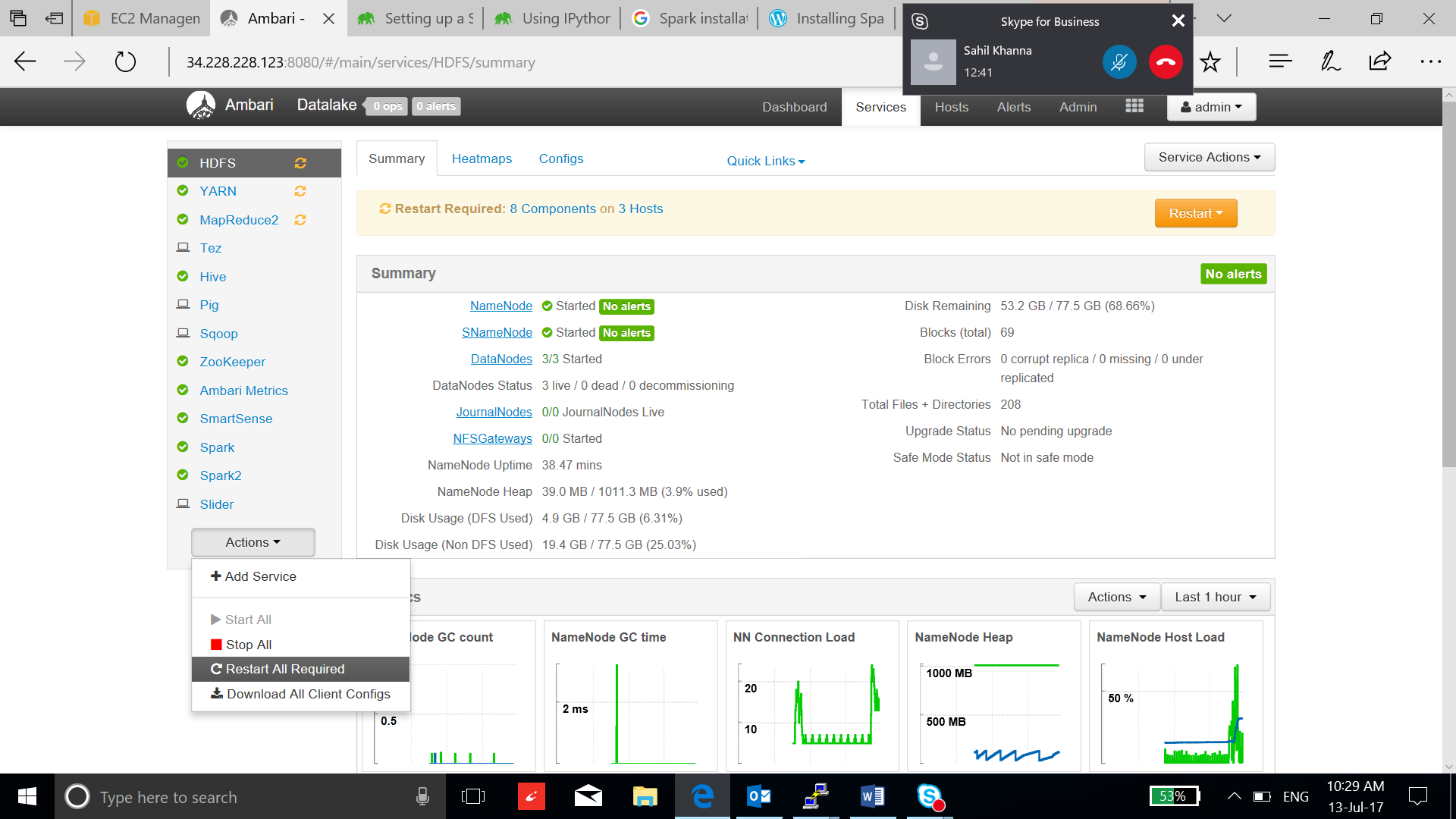
1. Review the repositories and select deploy:



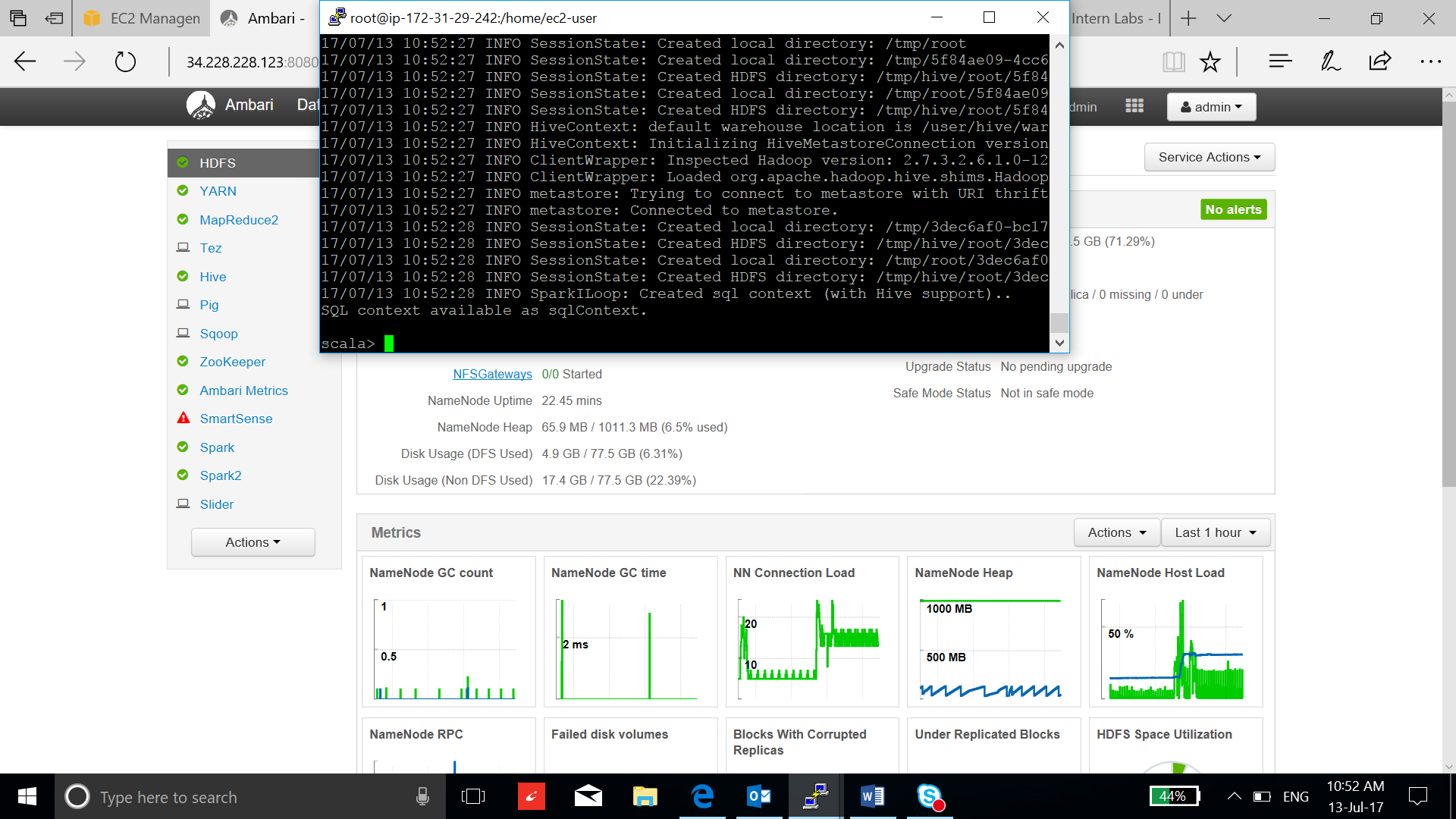
1. After the installation is complete and go to Ambari Dashboard, restart the required services:



1. After the restart completes, ssh to the instance and verify the installation:



1. Run “spark-shell” or use “pyspark” to verify:



1. Install the jupyter notebook:

* **wget** <https://repo.continuum.io/archive/Anaconda2-4.1.1-Linux-x86_64.sh>

**bash Anaconda2-4.1.1-Linux-x86\_64.sh (select yes to add variable to environment)**

**source .bashrc**

* **ipython profile create pyspark**

**vi ~/start\_ipython\_notebook.sh**

**copy:**

**#!/bin/bash**

**IPYTHON\_OPTS="notebook --port 8889 \**

**--notebook-dir='/usr/hdp/current/spark-client/' \**

**--ip='\*' --no-browser" pyspark**

**chmod +x start\_ipython\_notebook.sh**

**./start\_ipython\_notebook.sh**

1. Go to the browser, type <public IP address:8889> to open Jupyter notebook.
2. Change the spark version:

**export SPARK\_MAJOR\_VERSION=2**