BIG DATA CYBERSECURITY

LAB 1 INSTANTIATE A METRON VIRTUAL MACHINE

INSTANTIATE A SINGLE-NODE HADOOP & METRON CLUSTER IN THE AMAZON WEB SERVICES CLOUD

Lab Description: Apache Metron is an open-source cybersecurity software running in conjunction with several applications from the Apache Hadoop ecosystem. Metron is a tool for security monitoring and analysis integrating a number of big data technologies and capable of log aggregation, network packet capture indexing, storage, analytics and enrichment, and for cybersecurity network telemetry monitoring using threat information.

Metron has four major features: (1) A mechanism for capturing, storing and normalizing cybersecurity network telemetry data streaming at high rates, (2) Enriching telemetry data in real-time by adding threat intelligence, geolocation and DNS information, (3) Efficient storage of collected data allowing data mining and analytics, extraction and reconstruction of full network packets, and long term storage for visibility over time and machine learning application, and (4) a convenient graphical user interface for managing Metron settings, analyzing alerts and monitoring summaries [1].

Metron analytics consists of several components. Model-as-a-Service (MaaS) allows to deploy models, e.g. for determining whether a destination host in a Squid proxy log belongs to a domain generation algorithm threat [2]. The Metron Profiler generates a profile of a server, user, subnet or an application using a feature extraction mechanism. Profiles can be used by Metron MaaS for identifying anomalous behavior [3]. Metron Profiler Client provides with Java API and API for Stellar—Metron domain specific language, which allows simple computations and transformation of the cybersecurity data. The intended use of the Profiler Client is applying profile data for model scoring [4, 5]. Additionally, Metron Stellar has a number of statistics and math functions, which may be used for advanced analytics.

Metron interface consists of four major modules: (1) Alerts for browsing, filtering and acting on alerts; (2) Config for configuration and maintenance; (3) REST exposing Metron features through a REST/JSON interface, and (4) REST Client for serializing and de-serialilzing requests and responses passed to and from MaaS models [6].

Metron platform contains a number of Apache Storm topologies and topology attributes for streaming, enriching, indexing, and storing network telemetry in Hadoop. Metron is also capable of consuming data from sensors and sensor stubs [7].

The paragraphs above introduced a number of new concepts, which will be explained in the following lab assignments. In this assignment, we will instantiate a single-node Hadoop and Metron cluster in the Amazon Web Services (AWS) cloud. Apache Hadoop and related big data applications are typically used on a number of machines, which may range from one to several thousand. This is needed for fast processing of big datasets reaching terabytes of data. For this series of lab assignments, Metron, Hadoop and several related applications from the big data stack were installed on a single virtual machine (VM) in the AWS cloud for the purpose of saving budget. A custom Amazon Machine Image (AMI) was created from the VM. This AMI serves as a template for created a VM with a preinstalled operating system and all the applications needed for completing the lab exercises.

It is recommended for the lab instructor to apply for the AWS Educate program providing students and instructors with free credits, which can be applied towards multiple AWS resources [19].

Note: when creating a virtual machine, an error message similar to the one below may appear.

An error occurred (InstanceLimitExceeded) when calling the RunInstances operation: You have requested more instances than your current instance limit of 1 allows for the specified instance type. Please visit http://aws.amazon.com/contact-us/ec2-request to request an adjustment to this limit. [8]

Navigating to the URL provided in the message, completing and submitting a form usually resolves the issue.

Lab Files that are Needed: metron-cluster.ppk, metron-cluster-pem-public, aws_metron_script.sh, ssd-volume.json.

LAB EXERCISE/STEP 1

Obtain a public and private key pair from the lab instructor. Note: Normally, private keys are not shared. Instead, a user would generate a key pair comprised of public and private keys. This assignment is simplified to reduce student workload.

LAB EXERCISE/STEP 2

Create and activate an AWS account and login to the AWS Console as described in [9].

LAB EXERCISE/STEP 3

In the top right corner of the AWS Console, click your name, then click My Billing Dashboard. Next, click Credits in the left-hand-side menu. Enter a promo code given by your lab instructor and security check characters displayed as an image. Then click the Redeem button.

LAB EXERCISE/STEP 4

In the top right corner, click the region name to display a menu of available AWS data center locations. It is recommended to select a data center close to your physical location for better data transfer and response time.

In the top left corner, click Services then click EC2. In the menu on the left, locate and click Key Pairs under Network & Security. Click Import Key Pair, give your key pair a name and upload the public key file metron-cluster-pem-public or paste its contents, and then click the Import button.

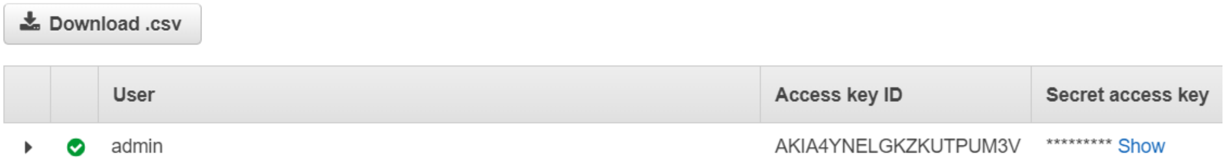
LAB EXERCISE/STEP 5

Create an Administrator Identity and Access Management (IAM) user [10].

Access to billing data should be enabled for the IAM admin user that we are about to create. In the top navigation bar, click your account name, and then click My Account. Next to IAM User and Role Access to Billing Information, choose Edit. Check the check box to Activate IAM Access and click Update.

Navigate to AWS IAM by following this link https://console.aws.amazon.com/iam/ or by clicking the corresponding item in the AWS Console menu. In the left-hand-side menu, click Users and then click the Add User button. In the user name box, type administrator. Below, for access type check the box next to Programmatic access and click the Next: Permissions button. On the new screen, click Attach existing policies directly. A table showing policy names will appear. Check the box next to AdministratorAccess. If this policy name does not show up, use the search bar to locate it. Click Next: Tags in the bottom right corner. Without making changes, click Next: Review. On the Review screen, click Create user.

When the user is successfully created, a new screen with a Success message will show up. This screen will also contain Access key ID and Secret access key. These keys should either be copied from the screen or downloaded as a CSV file.



		User	Access key ID	Secret access key
▸	✓	admin	AKIA4YNELGKZKUTPUM3V	********* Show

LAB EXERCISE/STEP 6

AWS Command Line Interface (CLI) is a proprietary tool for controlling and automating multiple AWS services from the command line and via scripts [11].

Install the AWS CLI on your workstation computer. Installation instructions and downloadable bundled installers are available at [20]

LAB EXERCISE/STEP 7

AWS CLI needs to be configured for your account [12]. To do so, open a terminal on your computer and type:

```
aws configure
```

The configuration tool will prompt for four values.

```
AWS Access Key ID [None]: YourKeyID
AWS Secret Access Key [None]: YourSecretAccessKey
Default region name [None]: us-east-2
Default output format [None]: json
```

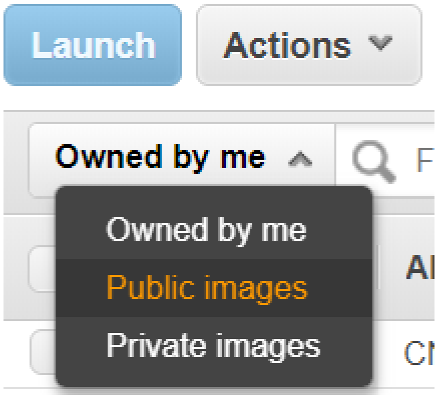
YourKeyID and YourSecretAccessKey must be changed to your AWS credentials obtained earlier in this assignment. It is recommended to change the default region name to the one, which is geographically closer to your physical location.

LAB EXERCISE/STEP 8

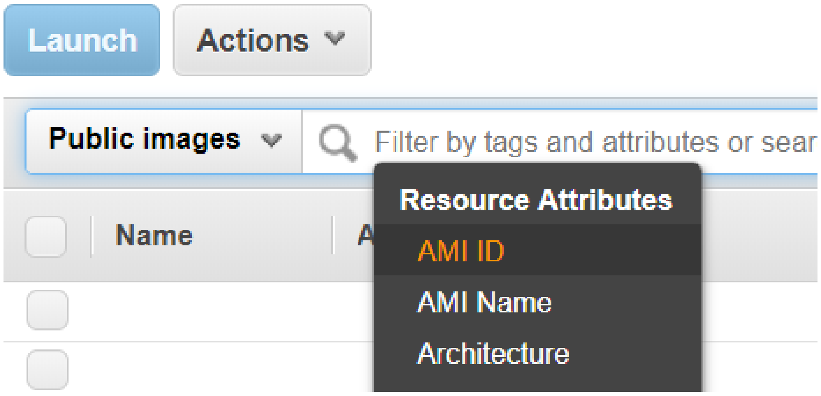
Start a single-node Metron cluster from an AWS Amazon Machine Image (AMI). An easy way to understand AMI is to think about it as a template for a virtual machine commonly containing preinstalled operating system, drivers, software and utilities for a particular task as well as launch permissions and other necessary information [13].

While preparing these laboratory assignments, we set up an AWS AMI with installed Hadoop Ecosystem, Apache Metron and a few other applications. It is possible to manually locate this AMI by its ID, ami-049c5743a562ad3ef.

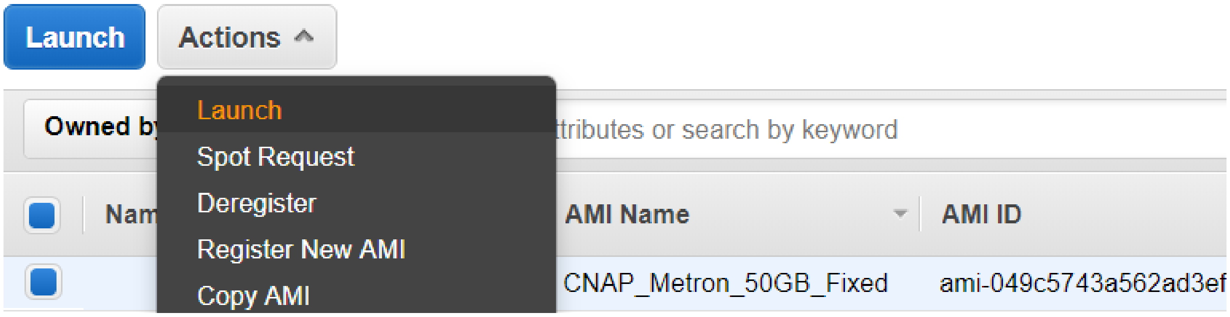
Task: manually start a new Elastic Cloud Compute (EC2) virtual machine from the shared Metron AMI. To do so, navigate to the AWS Console, click Owned by me and in the drop-down menu click Public images.



Then, click somewhere in the search bar. A Resource Attributes menu will show up. Click the AMI ID entry.

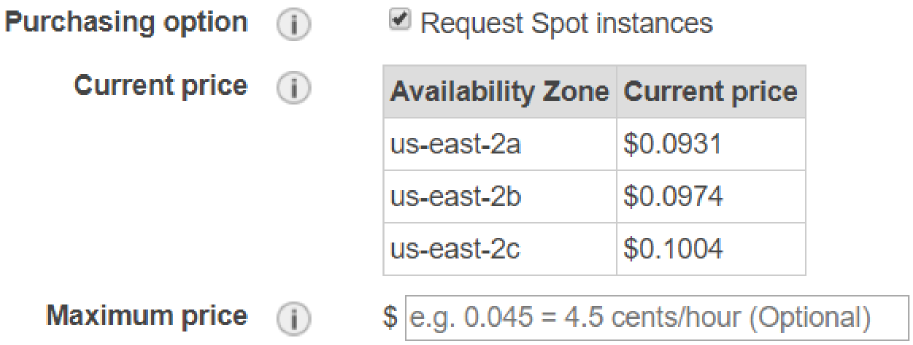


Then type or copy and paste the ID of the shared Metron AMI given above. Then press Enter. Next, click the row displaying details for the CNAP_Metron_50GB_fixed AMI. Click the Actions button and then click Launch.



Nam	AMI Name	AMI ID
	CNAP_Metron_50GB_Fixed	ami-049c5743a562ad3ef

The system will display a new screen, Step 2 Choose an instance type screen. We recommend to pick the memory-optimized r4.2xlarge instance type based on experiment results. Clicking Next: Configure Instance Details will open Step 3. Optionally, it is possible to check the Request Spot instances box. The Spot pricing may reduce costs by up to 90%. AWS EC2 may terminate, stop, or hibernate a Spot Instance when the Spot price exceeds the maximum price for your request or capacity is no longer available [14]. It is also possible to set the maximum price.



Purchasing option ☑ Request Spot instances

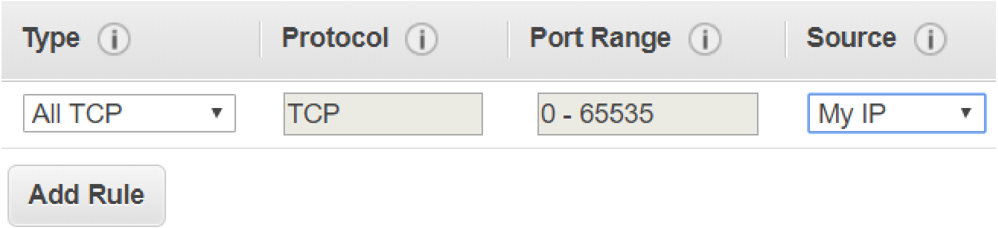
Current price

Availability Zone	Current price
us-east-2a	$0.0931
us-east-2b	$0.0974
us-east-2c	$0.1004

Maximum price $ e.g. 0.045 = 4.5 cents/hour (Optional)

Click Next: Add Storage. The default values on this screen are acceptable. It is possible to change the Volume Type to Provisioned IOPS SSD for better performance. Click Next: Add Tags. Add a tag on this screen with key Name and value Metron 50GB. The Name tag will later be displayed in the EC2 instances table and will help distinguish them. Otherwise, it will be possible to set or change its Name later manually.

Click Next: Configure Security Group. Apache Metron and its related applications use a number of ports, which may be accessed from a client computer. To avoid a hassle specifying all the different ports, it is possible to open all TCP ports with a restriction that they will be only available from IP address of your client computer:



Type	Protocol	Port Range	Source
All TCP	TCP	0 - 65535	My IP

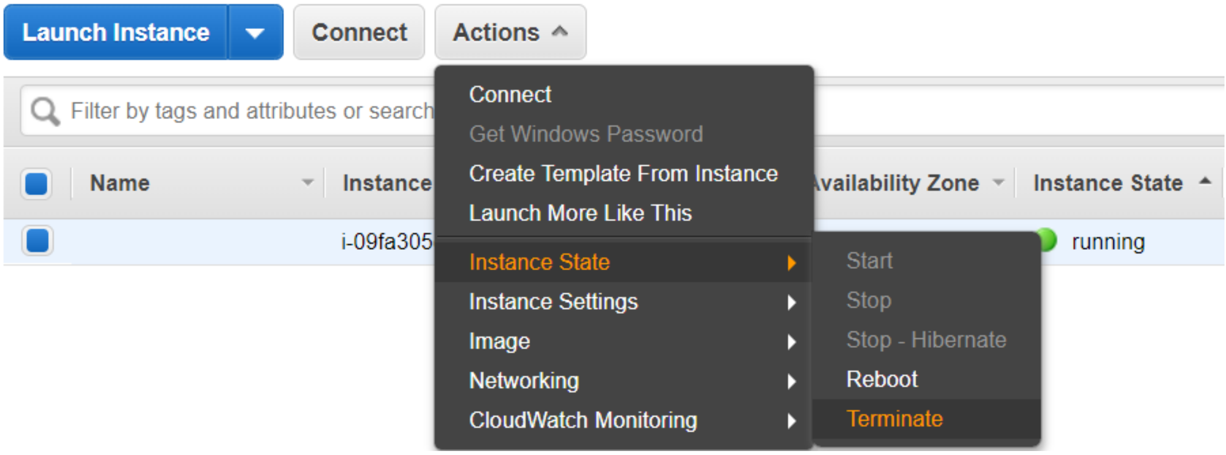
Clicking Preview and Launch will open the next screen displaying the summary of virtual machine settings. Expand all the items on this screen.

Make a screenshot and store it in the submission document.

Click Launch. A window asking to select either existing or new key pair will pop up. Select the metron-cluster key uploaded earlier, check the "I acknowledge …" box, and click the blue button, which would either display Request Spot Instances or Launch Instances depending on the choice you made earlier.

LAB EXERCISE/STEP 9

Terminate the created instance to avoid budget depletion. Click Services in the top bar of the AWS Console and click EC2. Click Instances in the menu on the left. The new screen will show a table of your EC2 instances. Ensure that the new instance is selected. Click Actions → Instance State → Terminate.



LAB EXERCISE/STEP 10

Start Metron virtual machine using AWS CLI. This way of starting a single-node cluster has its advantages since several important commands can be passed via a shell script to the virtual machine.

Pick or create a new working directory on your local computer for this step. Copy the aws_metron_script.sh file containing important commands for a virtual machine and the ssd-volume.json file with parameters for the virtual machine storage volume into the working directory.

Open a terminal in your local operating system and change directory to the working directory, where aws_metron_script.sh and ssd-volume.json have been copied. This is needed since these files are used by the AWS CLI command below. Run this command to start a new AWS EC2 virtual machine while changing several parameters:

```
aws ec2 run-instances --image-id ami-049c5743a562ad3ef --count 1 --
instance-type r4.2xlarge --key-name metron-cluster-pem-public  --
security-group-ids sg-68d29f09 --block-device-mappings file://ssd-
volume.json --user-data file://aws_metron_script.sh  --tag-specifications
ResourceType=instance,Tags=[{Key=Name,Value=MetronServer}]
ResourceType=volume,Tags=[{Key=Name,Value=MetronServer}]
```

Be sure two change the two values in bold:

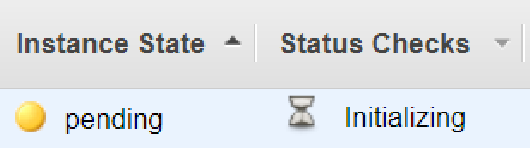
Metron-cluster-pem-public should be replaced the exact name you gave to the public key uploaded to AWS.

Sg-68d29f09 should be replaced with the exact ID of the security group you created in AWS. To find it in AWS Console, navigate to Services → EC2 → Security Groups.

After executing this command, a JSON-formatted response showing various values related to the created EC2 instance is displayed in the terminal.

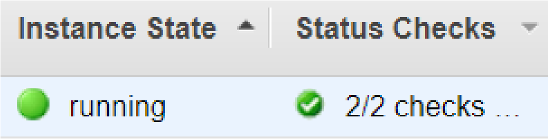
LAB EXERCISE/STEP 11

Using a web browser, navigate to AWS Console → Services → EC2 → Instances. Brief details of the new virtual machine will be displayed in a table. While the instance is created, its state goes from pending to running and its status checks change from Initializing to 2/2 checks passed.



Instance State	Status Checks
● pending	⌛ Initializing

→



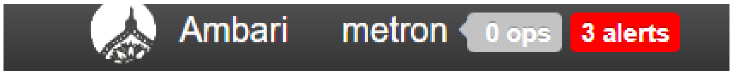
Instance State	Status Checks
● running	✓ 2/2 checks ...

Now, it is possible to connect to the instance via SSH or to Ambari interface via HTTP. Locate your instance public IPv4 address, and use it to modify the URL below to navigate to Ambari login screen:

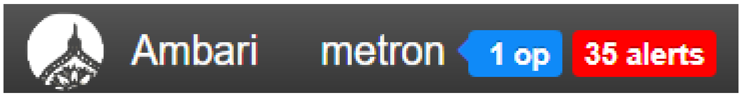
http://ipv4address:8080

where ipv4address must be substituted with the actual one.

In the login screen, enter admin twice, once as a username and once as a password. Initially, Ambari will show that zero operations are pending, three alerts are active and services are in unknown state. The number of alerts may grow to 35 or a close number.



Meanwhile, the virtual machine is executing important commands provided in the aws_metron_script.sh file. After approximately a minute, you may notice that one operation is pending.

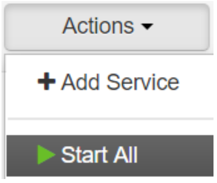


If you click the blue icon, a new window will display the overall progress of past and current operations and their titles. On the top, there will be Start All Services in its initial phase slowly growing to 100%. Click anywhere in the pending operation row. It will display that the operation is being done on a single node named metronserver.localdomain. Click that row once more to see a scrollable list of completed and scheduled tasks.

Make a screenshot and store it in the submission document.

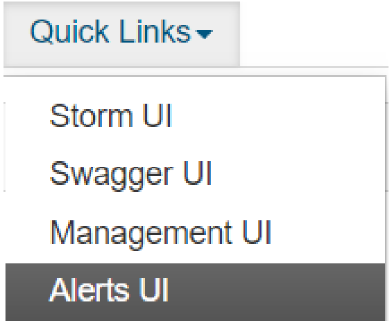
Click the green OK button in the bottom right corner to exit the progress window. Overall, it may take 10-15 minutes to start all services of this distribution of Apache Hadoop with Metron. After services start, they will have round green check box icons next to them. There might be a few alerts as well.

Note: if a virtual machine running Metron is restarted, all services must be started using Ambari. To do so, Ambari dashboard must be scrolled down until the Actions drop-down menu becomes visible. The Start All menu item will start the services.



LAB EXERCISE/STEP 12

Locate and click Metron in the left-hand-side menu. In the Metron screen, locate the Quick Links drop down menu.



Using the menu, open in new tabs Storm UI, Swagger UI, Management UI and Alerts UI. Storm UI does not require a password. Swagger UI, Management UI and Alerts UI use metron as a username and Smoothmetron2 as a password.

Make a screenshot of each component: Storm UI, Swagger UI, Management UI and Alerts UI. Store screenshots.

Navigate to Storm UI. This view consists of four sections: Cluster Summary showing brief details of Apache Storm, Nimbus Summary displays a few metrics for Nimbus, a Storm service responsible for distributing Storm code, assigning Storm tasks to Storm Supervisors, and monitoring for Storm failures [15]. Topology Summary shows details on installed Storm topologies, which are real-time applications consisting of spouts and bolts for processing streaming data [16]. Supervisor Summary shows several data points on Storm Supervisor, which responsibility is to receive a task from Nimbus and passing it to worker processes. The Nimbus Configuration table consists of more than two hundred lines of Nimbus settings.

Navigate to Swagger UI, which is a web interface generated by a third-party tool for Metron API. It allows to view auto-generated documentation and interact with Metron API [17].

Task: expand the line for Alerts UI Controller, type threat in the Value box and click the Try it out! button.

Make a screenshot of the response and store it in the submission document.

Navigate to the Management UI, which displays the status and metrics for available sensors. Click the line corresponding to the Bro sensor. An information panel will open on the right side showing configuration and status details for the sensor parses topology.

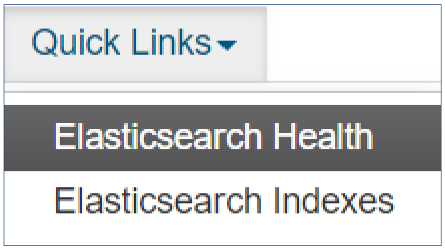
Make a screenshot and store it in the submission document.

This view also allows to delete, enable, disable, start or stop available parses topologies.

Navigate to the Alerts UI, which is designed to display alerts generated by Metron and to take additional actions, such as open, dismiss, escalate, resolve or add to an alert. This view also allows querying and filtering the data. This view will be used in a subsequent laboratory assignment in bigger detail.

LAB EXERCISE/STEP 13

Click Elasticsearch in the left-hand-side menu. Locate the Quick Links menu in the middle of the upper side of the screen.



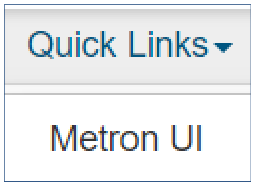
The Health screen displays summary data on the Elasticsearch service and the Indexes screen shows a brief overview on each existing index. These views do not require a password.

Task: open both views. Verify that Metron cluster and .kibana index have a green status.

Make a screenshot and store it in the submission document.

LAB EXERCISE/STEP 14

Locate and click Kibana in the menu on the left side. Locate and click the Quick Links menu in the view.



Open the Metron UI screen. Kibana Metron UI does not require a password. This component allows to produce analytics dashboards, to analyze and aggregate data from Elasticsearch indices. It consists of several components: Discover, Visualize, Dashboard, Timelion, Dev Tools and Management.

LAB EXERCISE/STEP 15

Be sure to terminate your AWS Metron VM to avoid budget depletion. Resources created in a cloud environment under your account have associated costs. AWS offers more than 60 products at a free tier with associated free tier usage limits [18].

PUZZLER

If a Metron virtual machine is started manually using AWS Console, Ambari shows that all services are not started. On the other hand, if a Metron virtual machine is created using AWS CLI, after some time all services are up and running. Explain the difference.

What to submit

Submit a Word (or other text editor) document with embedded screenshots made as requested in the assignment and a brief description for each screenshot.

References

[1] The Apache Software Foundation, "Apache Metron," [Online]. Available: https://metron.apache.org/current-book/index.html. [Accessed: Aug. 12, 2019].

[2] The Apache Software Foundation, "Model Management Infrastructure," [Online]. Available: https://metron.apache.org/current-book/metron-analytics/metron-maas-service/index.html. [Accessed: Aug. 12, 2019].

[3] The Apache Software Foundation, “Metron Profiler,” [Online]. Available: https://metron.apache.org/current-book/metron-analytics/metron-profiler/index.html. [Accessed: Aug. 12, 2019].

[4] The Apache Software Foundation, “Metron Profiler Client,” [Online]. Available: https://metron.apache.org/current-book/metron-analytics/metron-profiler-client/index.html. [Accessed: Aug. 12, 2019].

[5] The Apache Software Foundation, “Stellar Language,” [Online]. Available: https://metron.apache.org/current-book/metron-stellar/stellar-common/index.html. [Accessed: Aug. 12, 2019].

[6] The Apache Software Foundation, “Metron Interface,” [Online]. Available: https://metron.apache.org/current-book/metron-interface/index.html. [Accessed: Aug. 12, 2019].

[7] The Apache Software Foundation, “Sensor Stubs,” [Online]. Available: https://metron.apache.org/current-book/metron-deployment/roles/sensor-stubs/index.html. [Accessed: Aug. 12, 2019].

[8] AWS “Create case. Service limit increase,” [Online]. Available: http://aws.amazon.com/contact-us/ec2-request. [Accessed: Aug. 12, 2019].

[9] AWS, “How do I create and activate a new Amazon Web Services account?,” [Online]. Available: https://aws.amazon.com/premiumsupport/knowledge-center/create-and-activate-aws-account/. [Accessed: Aug. 12, 2019].

[10] AWS, “Creating Your First IAM Admin User and Group,” [Online]. Available: https://docs.aws.amazon.com/IAM/latest/UserGuide/getting-started_create-admin-group.html. [Accessed: Aug. 12, 2019].

[11] AWS, “AWS Command Line Interface,” [Online]. Available: https://aws.amazon.com/cli/. [Accessed: Aug. 12, 2019].

[12] AWS, “Quickly Configuring the AWS CLI,” [Online]. Available: https://docs.aws.amazon.com/cli/latest/userguide/cli-chap-configure.html#cli-quick-configuration. [Accessed: Aug. 12, 2019].

[13] AWS, “Amazon Machine Images (AMI),” [Online]. Available: https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/AMIs.html. [Accessed: Aug. 12, 2019].

[14] AWS, “Spot Instances,” [Online]. Available: https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-spot-instances.html?icmpid=docs_ec2_console. [Accessed: Aug. 12, 2019].

[15] Apache Storm, “Tutorial,” Available: https://storm.apache.org/releases/current/Tutorial.html. [Accessed: Aug. 12, 2019].

[16] Apache Storm, “Concepts,” Available: https://storm.apache.org/releases/1.2.3/Concepts.html

[17] Swagger Supported by Smartbear, "Swagger UI," Available: https://swagger.io/tools/swagger-ui/. [Accessed: Aug. 12, 2019].

[18] AWS, "AWS Free Tier," [Online]. Available: https://aws.amazon.com/free/. [Accessed Aug 8, 2019].

[19] AWS, "AWS Educate. Teach Tomorrow's Cloud Workforce Today," [Online]. Available: https://aws.amazon.com/education/awseducate/. [Accessed Aug 8, 2019].

[20] AWS, "Installing the AWS CLI," [Online]. Available: https://docs.aws.amazon.com/cli/latest/userguide/cli-chap-install.html. [Accessed Aug 8, 2019].