Try it out objective

Use this hands-on to get started with data streams and one of the process by which data can be injected from IaaS to a streaming service. Install and configure the agent on an instance and send log data to the stream, then to firehose and finally to S3.

The goal

The following are the goals of this hands-on:

1. Create the data stream & a delivery stream

2. Agent installation & understand the need of IAM

3. Log (or any data) collection in S3 (forming a data lake)

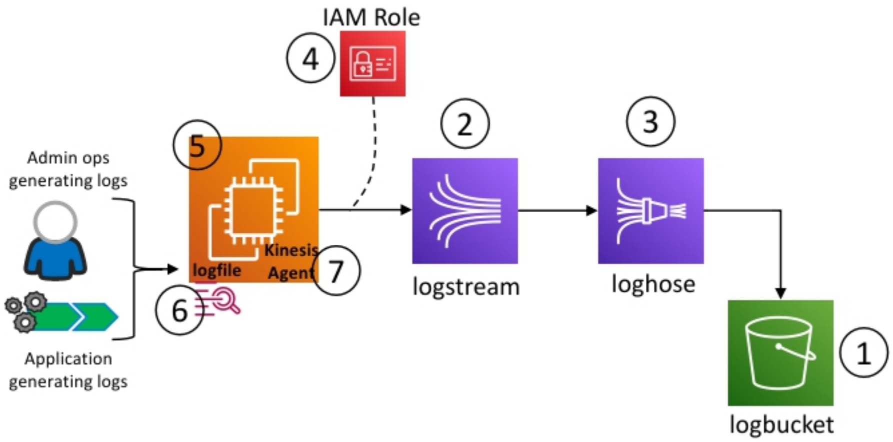
Note: in the diagram to the right -

“logstream” is the name of the Kinesis data stream

“loghose” is the name of the Firehose delivery stream

“logbucket” is the representation of the bucket

The numbers represent the order of setting up the flow.



Important - this exercise **will not work in AWS Academy**, a personal AWS account is mandatory. This exercise is for technical learners only.

Please note if a field (short for text field/text area/checkbox/radio/dropdown/list or any other UI element) is not specified in the following steps, it means the default value of the field set by AWS needs to be used. No change is needed for those fields as part of this hands-on.

This exercise will work with multiple services, please use a dedicated browser instance with only the tabs that are needed for this exercise, otherwise it may/will lead to confusion.

A. Hands-On: Create S3 Bucket

1. Visit S3 management console at https://s3.console.aws.amazon.com/s3/home?region=us-east-1&region=us-east-1 (you will be required to sign in)
2. On the right hand side, there will be a button called **Create Bucket**. Click on it.
3. Under the **General Configuration** card make the following changes –
   a) For **Bucket name** field use the following naming -

   ```
   pgpccmmddhhmm
   ```

   NOTE: mmddhhmm represents Month, day, hour and minute, a valid value could be pgpcc12251538 (25th December 15hrs 38mins), do not use this example value from here, use the actual date and time at the time of doing this exercise.

   b) Keep the default **US East(N. Virginia) us-east-1 region.**
4. Under **Object Ownership** card, ensure the radio button for **ACLs disabled** is selected.
5. Under **Block Public Access settings for this bucket** card, checkbox for **Block all public access** should be checked.
6. Under **Bucket Versioning** card ensure **Disable** is selected.
7. **Tag** is optional for the user.
8. Under **Default encryption** card ensure **Disable** is selected.
9. Under **Advanced Settings** card ensure **Disable** is selected.
10. Click on **Create Table** button at the bottom of the page to create your bucket.

B. Hands-On: Create Amazon Kinesis Data Stream

1. Visit Amazon Kinesis management console at https://console.aws.amazon.com/kinesis/home?region=us-east-1#/home
2. Select **Kinesis Data Stream** radio from the **Get Started** section and click on **Create Data Stream** button.
3. Under the **Date stream** name field paste the following value –

```
logstream
```

4. Under **Data stream capacity** card select **provisioned** radio button and type **1** in **provisioned shard** field.
5. At the bottom of the page, click **Create Data Stream** button.
6. In the left navigation, click on **Delivery Streams**
7. Click on **Create Delivery Stream** button.
8. Under **Choose source and destination** card select the following –
   a) In the **Source** field, select **Amazon Kinesis Data Streams** from the dropdown.
   b) In the **Destination** field, select **Amazon S3** from the dropdown.
9. Under **Source** settings card, browse and select **logstream** as **Kinesis Data Stream.**
10. Under **Delivery Stream Name** card paste the following value –

```
loghose
```

11. Under **Transform and convert records** card make the following changes –
    a) In **Data Transformation** field ensure **Disabled** radio button is selected
    b) In **Record format conversion** field ensure **Disabled** radio button is selected.
12. Under **Destination settings** card make the following changes –
    a) In the **S3 bucket name** field browse and select the S3 bucket created above (pgpccmmddhhmm)
    b) In the **Dynamic partitioning** field ensure **Disabled** radio button is selected.
    c) In the **S3 buffer hints** field, keep the buffer size as **5 MiB** and change the **buffer interval time** to **60 seconds** by typing it manually.

   d) In the **Compression for data records** field ensure **Disabled** radio button is selected

   e) In the **Encryption for data records** field ensure **Disabled** radio button is selected

13. At the bottom of the page click on **Create Delivery Stream** button.

C. Hands-On: Launch an EC2 Instance

1. Visit EC2 management console at https://console.aws.amazon.com/ec2/v2/home?region=us-east-1#Instances:sort=instanceId
2. Click on **Launch Instance** button and create new instance with following specifications (Refer CloudWithAWS-TIO-1-v1.1 under 'Cloud Computing with AWS' module)
   a) Amazon Linux 2 AMI

   b) t2 micro

   c) Security group - port 22 should be open

   Leave all other fields, dropdowns, checkboxes, radio buttons etc to their default
3. After creating the instance successfully, click on the checkbox to the left of the instance name displayed in the **Instances** table. (no action is necessary if the checkbox is already selected)
4. Click on the **Connect** button towards the top of the screen.
5. **Connect to your Instance pop** up will appear, select the **EC2 instance connect (browser-based SSH connection)** radio button.
6. Click on **Connect** button at the bottom of that page.
7. Terminal window is now open to use.

D. Hands-On: Install, Configure & Test the Kinesis agent from the EC2 instance

1. To setup and configure the agent, copy and paste the following two commands one after the other –

```
sudo yum update

sudo yum install -y aws-kinesis-agent
```

2. We are going to use a stream and simulate it by picking up the logs from an application log file, let's call it myapp.log located at /opt. Copy and paste the following three commands one after the other –

```
sudo mkdir /opt/myapp

sudo chown ec2-user:ec2-user -R /opt/myapp

touch /opt/myapp/myapp.log
```

3. To configure the agent to push the data to the stream, run the following two commands one after the other –

```
wget https://d6opu47qoi4ee.cloudfront.net/agent.json

sudo mv -f agent.json /etc/aws-kinesis/agent.json
```

4. To start the Kinesis Agent and see the agent logs, copy and paste the following two commands one after the other –

```
sudo service aws-kinesis-agent start

tail -25 /var/log/aws-kinesis-agent/aws-kinesis-agent.log
```

5. To throw some logs in the application log file and check the agent log, copy and paste the following two commands one after the other –

```
echo "This is the first log entry" > /opt/myapp/myapp.log

tail -25 /var/log/aws-kinesis-agent/aws-kinesis-agent.log
```

Notice the message “**1 record parsed but 0 records sent to destination**”, it is because the **IAM is missing**

6. After attaching the IAM role to the instance (Refer to section E). Copy paste the following command in the terminal (*if the terminal window tab becomes non responsive then refresh the tab and the terminal will become active again*) –

```
tail -25 /var/log/aws-kinesis-agent/aws-kinesis-agent.log
```

Now you will see 1 record sent successfully.

7. Now to **append multiple lines** to the log file, copy and paste the following lines with about **5 seconds gap** between each –

```
echo "Line 2 in the log" >> /opt/myapp/myapp.log

echo "Line 3 in the log" >> /opt/myapp/myapp.log

echo "Line 4 in the log" >> /opt/myapp/myapp.log

echo "Line 5 in the log" >> /opt/myapp/myapp.log
```

8. Go to **S3 Management Console** and see the **folder structure**, **download the log file** and open to **see the logs**.

9. Go to **Step F** for cleanup

E. Hands-On: Attach IAM Role to the EC2 Instance

1. Visit IAM console at https://console.aws.amazon.com/iamv2/home?#/home

2. In the left navigation, under **Access management** click on **Roles**

3. Click on **Create role** button.

4. Select **AWS service** in the field of **trusted entity.**

5. Select **EC2** in the field of **common use case.**

6. Click on **Next: Permission** button at the bottom of the page.

7. In this step you need to select the policies.
   a) Paste the following policy in Search field (just above the table that lists the policies) –

```
AmazonKinesisFullAccess
```

   b) Click on the checkbox on the left of the policy name.
   c) Repeat the above process for the following policy as well.

```
CloudWatchFullAccess
```

8. Click on **Next: Tags** button at the bottom of the page.

9. **Tag** is optional, no change is needed

10. Click on **Next: Review** button at the bottom of the page.

11. In the field of **Role name** paste the following –

```
IAMRoleKinesis
```

Make sure both policies are there in the Policies field.

12. Click on **Create Role** button at the bottom of the page.
13. Go to your EC2 management console and select the instance (checkbox to the left of the listing)
14. Click on **Actions dropdown** then click on **Security** and finally click on **Modify IAM role**
15. Attach the IAM role by **selecting it from the dropdown**.
16. Click the **Save** button.

Go back to **Step D point 6** to continue with the kinesis agent exercise

F. Hands-On: Cleaning up!

1. In the respective service management console cleanup (delete/terminate) the following resources -
   a) EC2
   b) IAM Role
   c) Kinesis Firehose
   d) Kinesis Stream
   e) S3 bucket (will require the “Empty” bucket step before bucket deletion)