

# Project execution documentation

## Step1 : Create EMR cluster in aws

1. Sign in to the AWS Management Console and open the Amazon EMR console
2. Choose Create cluster.
3. On the Create Cluster - Quick Options page, accept the default values except for the following fields:
  - a. Enter a Cluster name that helps you identify the cluster, for example, dezyre-cluster
  - b. Under Security and access, choose the EC2 key pair that you created in [Create an Amazon EC2 Key Pair](#).
  - c. In the applications, choose your required services , for our project we choose the first option
  - d. In hardware configuration, we choose 1 master and 1 core node of the instance type m5.xlarge
  - e. Enable auto scaling option
  - f. Choose your keypair and leave permissions to default.
4. Choose Create cluster.

How would you rate your experience with this service console? ☆ ☆ ☆ ☆ ☆

aws Services ▾

projectpro ▾ Ohio ▾ Support ▾

Create Cluster - Quick Options [Go to advanced options](#)

General Configuration

Cluster name

☒ Logging ⓘ

S3 folder

Launch mode ☒ Cluster ⓘ ☐ Step execution ⓘ

Software configuration

Release

Applications

☒ Core Hadoop: Hadoop 2.10.0, Hive 2.3.7, Hue 4.7.1, Mahout 0.13.0, Pig 0.17.0, and Tez 0.9.2

☐ HBase: HBase 1.4.13, Hadoop 2.10.0, Hive 2.3.7, Hue 4.7.1, Phoenix 4.14.3, and ZooKeeper 3.4.14

☐ Presto: Presto 0.238.3 with Hadoop 2.10.0 HDFS and Hive 2.3.7 Metastore

☐ Spark: Spark 2.4.6 on Hadoop 2.10.0 YARN and Zeppelin 0.8.2

☐ Use AWS Glue Data Catalog for table metadata ⓘ

Hardware configuration

Instance type

The selected instance type adds 64 GiB of GP2 EBS storage per instance by default. [Learn more](#)

Number of instances  (1 master and 2 core nodes)

Cluster scaling ☐ scale cluster nodes based on workload

## Hardware configuration

**Instance type** m5.xlarge The selected instance type adds 64 GiB of GP2 EBS storage per instance by default. [Learn more](#)

**Number of instances** 2 (1 master and 1 core nodes)

**Cluster scaling** ☒ scale cluster nodes based on workload

**EMR-managed scaling**  
EMR will automatically increase and decrease the number of instances in core and task nodes based on workload. Set a minimum and maximum limit of the number of instances for the cluster nodes. Master nodes do not scale. [Learn more](#)

**Core and task units**

Minimum: 1

Maximum: 2

## Security and access

**EC2 key pair** dezyre-keypair [Learn how to create an EC2 key pair.](#)

**Permissions** ☒ Default ☐ Custom  
Use default IAM roles. If roles are not present, they will be automatically created for you with managed policies for automatic policy updates.

**EMR role** EMR\_DefaultRole

**EC2 instance profile** EMR\_EC2\_DefaultRole

[Cancel](#) [Create cluster](#)

- Cluster creation tasks takes 5-10 minutes . You may need to choose the refresh icon on the right or refresh your browser to receive updates.
- Under Network and hardware, find the Master and Core instance status. The status goes from Provisioning to Bootstrapping to Waiting during the cluster creation process. For more information, see [Understanding the Cluster Lifecycle](#).
- As soon as you see the links for Security groups for Master and Security Groups for Core & Task, you can move on to the next step, but you may want to wait until the cluster starts successfully and is in the Waiting state.

Clone

Terminate

AWS CLI export

Cluster: dezyre-clusterStarting

Summary

Application user interfaces

Monitoring

Hardware

Configurations

Events

Steps

Bootstrap actions

Summary

ID: j-36W2U13LMG7WU

Creation date: 2020-10-31 20:27 (UTC+5:30)

Elapsed time: 1 second

After last step completes: Cluster waits

Termination protection: Off [Change](#)

Tags: -- [View All / Edit](#)

Master public DNS: --

Configuration details

Release label: emr-5.31.0

Hadoop distribution: Amazon 2.10.0

Applications: Hive 2.3.7, Hue 4.7.1, Mahout 0.13.0, Pig 0.17.0, Tez 0.9.2

Log URI: s3://dezyre-bucket/

EMRFS consistent view: Disabled

Custom AMI ID: --

Application user interfaces

Persistent user interfaces --

On-cluster user -- interfaces

Security and access

Key name: dezyre-keypair

EC2 instance profile: EMR\_EC2\_DefaultRole

EMR role: EMR\_DefaultRole

Visible to all users: All [Change](#)

Security groups for Master:

Security groups for Core & Task:

Network and hardware

Availability zone: --

Subnet ID: [subnet-f957cab5](#)

Master: Provisioning 1 m5.xlarge

Core: Provisioning 1 m5.xlarge

Task: --

Cluster scaling: EMR-managed scaling

aws

Services

projectpro

Ohio

Support

Amazon EMR

Clusters

Notebooks

Git repositories

Security configurations

Block public access

VPC subnets

Events

Help

What's new

Clone

Terminate

AWS CLI export

Cluster: dezyre-clusterStarting

Summary

Application user interfaces

Monitoring

Hardware

Configurations

Events

Steps

Bootstrap actions

Summary

ID: j-36W2U13LMG7WU

Creation date: 2020-10-31 20:27 (UTC+5:30)

Elapsed time: 31 seconds

After last step completes: Cluster waits

Termination protection: Off [Change](#)

Tags: -- [View All / Edit](#)

Master public DNS: --

Configuration details

Release label: emr-5.31.0

Hadoop distribution: Amazon 2.10.0

Applications: Hive 2.3.7, Hue 4.7.1, Mahout 0.13.0, Pig 0.17.0, Tez 0.9.2

Log URI: s3://dezyre-bucket/

EMRFS consistent view: Disabled

Custom AMI ID: --

Application user interfaces

Persistent user interfaces --

On-cluster user -- interfaces

Security and access

Key name: dezyre-keypair

EC2 instance profile: EMR\_EC2\_DefaultRole

EMR role: EMR\_DefaultRole

Visible to all users: All [Change](#)

Security groups for Master: [sg-0195194765414ec26](#) (ElasticMapReduce-master)

Security groups for Core & Task: [sg-0a584ce26d5cdab681](#) (ElasticMapReduce-slave)

Network and hardware

Availability zone: us-east-2c

Subnet ID: [subnet-f957cab5](#)

Master: Provisioning 1 m5.xlarge

Core: Provisioning 1 m5.xlarge

Task: --

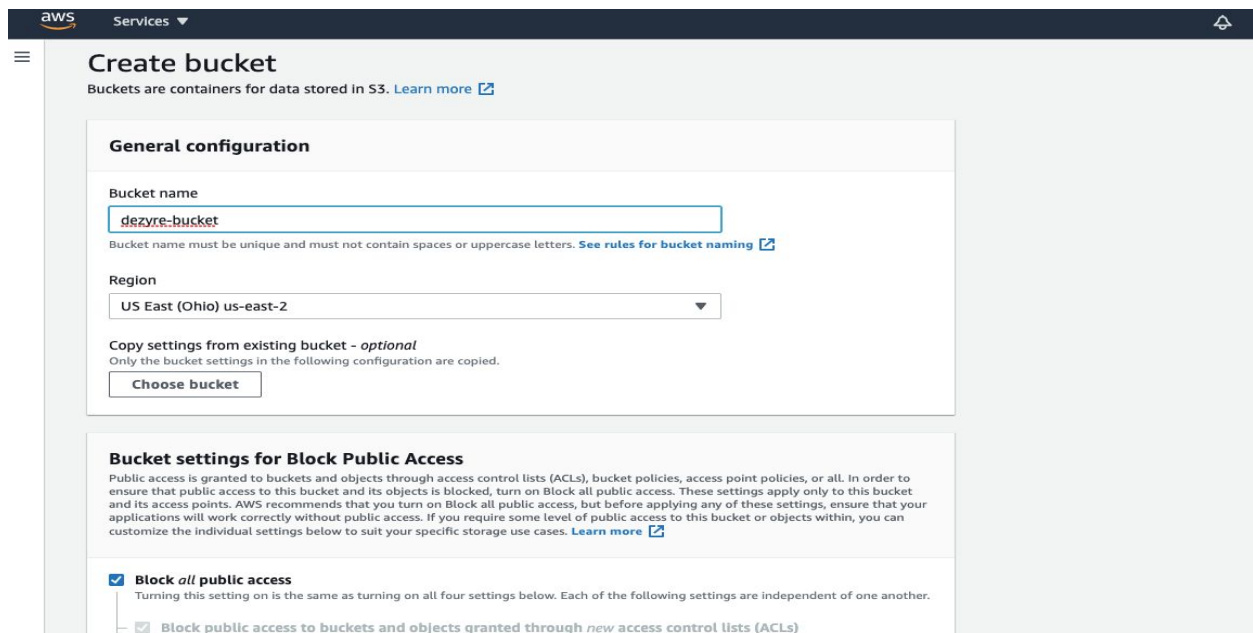
Cluster scaling: EMR-managed scaling

## Step 2: Create S3 bucket

Give a unique name to the S3 (please add steps) to create and upload file

1. Sign in to the AWS Management Console and open the Amazon S3 console
2. Choose Create Bucket.  
The Create a Bucket dialog box opens.
3. Enter a bucket name, such as **dezyre-bucket**.  
This name should be globally unique, and cannot be the same name used by another bucket.
4. Select the Region for your bucket. To avoid paying cross-region bandwidth charges, create the Amazon S3 bucket in the same region as your cluster.
5. Choose Create.

You created a bucket with the URI `s3n://dezyre-bucket/` as shown in the below.



The screenshot shows the 'Create bucket' dialog box in the AWS Management Console. The 'General configuration' section is active, showing the 'Bucket name' field with the value 'dezyre-bucket' and the 'Region' dropdown set to 'US East (Ohio) us-east-2'. Below this, there is a section for 'Bucket settings for Block Public Access' with a checkbox for 'Block all public access' checked. The dialog box also includes a 'Choose bucket' button and a 'Learn more' link.

**Create bucket**  
Buckets are containers for data stored in S3. [Learn more](#)

**General configuration**

Bucket name  
  
Bucket name must be unique and must not contain spaces or uppercase letters. [See rules for bucket naming](#)

Region

Copy settings from existing bucket - *optional*  
Only the bucket settings in the following configuration are copied.

**Bucket settings for Block Public Access**  
Public access is granted to buckets and objects through access control lists (ACLs), bucket policies, access point policies, or all. In order to ensure that public access to this bucket and its objects is blocked, turn on Block all public access. These settings apply only to this bucket and its access points. AWS recommends that you turn on Block all public access, but before applying any of these settings, ensure that your applications will work correctly without public access. If you require some level of public access to this bucket or objects within, you can customize the individual settings below to suit your specific storage use cases. [Learn more](#)

☒ **Block all public access**  
Turning this setting on is the same as turning on all four settings below. Each of the following settings are independent of one another.

☒ Block public access to buckets and objects granted through *new* access control lists (ACLs)

**Bucket Versioning**  
Versioning is a means of keeping multiple variants of an object in the same bucket. You can use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket. With versioning, you can easily recover from both unintended user actions and application failures. [Learn more](#)

Bucket Versioning  
☒ Disable  
☐ Enable

**Tags (0) - optional**  
Track storage cost or other criteria by tagging your bucket. [Learn more](#)

No tags associated with this bucket.

[Add tag](#)

**Default encryption**  
Automatically encrypt new objects stored in this bucket. [Learn more](#)

Server-side encryption  
☒ Disable  
☐ Enable

► **Advanced settings**

ⓘ After creating the bucket you can upload files and folders to the bucket, and configure additional bucket settings.

[Cancel](#) [Create bucket](#)

When Amazon S3 successfully creates your bucket, the console displays your empty bucket in the Buckets panel.

1. Create a folder.
  - a. Click the name of the new bucket.
  - b. Click the Actions button, and click **Create Folder** in the drop-down list.
  - c. Name the new folder **load**.

Upload the data files to the new Amazon S3 bucket.

- a. Click the name of the data folder.
- b. In the Upload - Select Files wizard, click **Add Files**.  
A file selection dialog box opens.
- c. Select all of the files you downloaded and extracted, and then click **Open**.
- d. Click **Start Upload**.

Here after creating the s3 will store data in a bucket.

Step 3: Connect to the emr master node using SSH

## Connect to the Master Node Using SSH

You can connect to the Amazon EMR master node using SSH to run interactive queries, examine log files, submit Linux commands, and so on.

[Learn more](#)

Windows Mac / Linux

1. Open a terminal window. On Mac OS X, choose Applications > Utilities > Terminal. On other Linux distributions, terminal is typically found at Applications > Accessories > Terminal.
2. To establish a connection to the master node, type the following command. Replace ~/dezyre-keypair.pem with the location and filename of the private key file (.pem) used to launch the cluster.

```
ssh -i ~/dezyre-keypair.pem hadoop@ec2-13-58-124-139.us-east-2.compute.amazonaws.com
```

3. Type yes to dismiss the security warning.

Close

aws Services

Amazon EMR

Clusters

Notebooks

Git repositories

Security configurations

Block public access

VPC subnets

Events

Help

What's new

Clone Terminate AWS CLI export

Cluster: dezyre-cluster **Running** Running step

Summary Application user interfaces Monitoring Hardware Configurations Events Steps Bootstrap actions

**Summary**

ID: j-36W2U13LMG7WU  
 Creation date: 2020-10-31 20:27 (UTC+5:30)  
 Elapsed time: 7 minutes  
 After last step completes: Cluster waits  
 Termination protection: Off [Change](#)  
 Tags: -- [View All / Edit](#)  
 Master public DNS: ec2-13-58-124-139.us-east-2.compute.amazonaws.com  
[Connect to the Master Node Using SSH](#)

**Configuration details**

Release label: emr-5.31.0  
 Hadoop distribution: Amazon 2.10.0  
 Applications: Hive 2.3.7, Hue 4.7.1, Mahout 0.13.0, Pig 0.17.0, Tez 0.9.2  
 Log URI: s3://dezyre-bucket/  
 EMRFS consistent view: Disabled  
 Custom AMI ID: --

**Application user interfaces**

Persistent user interfaces: [YARN timeline server](#), [Tez UI](#)  
 On-cluster user interfaces: Not Enabled [Enable an SSH Connection interfaces](#)

**Network and hardware**

Availability zone: us-east-2c  
 Subnet ID: [subnet-f957cab5](#)  
 Master: **Running** 1 m5.xlarge  
 Core: **Running** 1 m5.xlarge  
 Task: --  
 Cluster scaling: EMR-managed scaling

**Security and access**

Key name: dezyre-keypair  
 EC2 instance profile: EMR\_EC2\_DefaultRole  
 EMR role: EMR\_DefaultRole  
 Visible to all users: All [Change](#)  
 Security groups for Master: [sg-0195134765414ec25](#) (ElasticMapReduce-master)

Once we open EC2 instance view, we would see the instances created and running

Instances (6) Info

Filter instances

	Name	Instance ID	Instance state	Instance type	Status check	Alarm Status	Availability zone	Public IPv4 DNS
<input type="checkbox"/>	Project4	i-0abe4f78736b1768e	<b>Running</b>	t2.small	2/2 checks ...	No alarms	us-east-2a	ec2-18-220-218-183.u...
<input type="checkbox"/>	-	i-0a5342f8c96637450	<b>Running</b>	m5.xlarge	2/2 checks ...	No alarms	us-east-2c	ec2-18-224-94-38.us...
<input type="checkbox"/>	-	i-0c17deace143e14ff	<b>Running</b>	m5.xlarge	2/2 checks ...	No alarms	us-east-2c	ec2-13-58-124-139.us...

Now, ssh to emr master node as follows :

```
srikantha@Srikanths-MacBook-Air Downloads % ssh -i dezyre-keypair.pem hadoop@ec2-13-58-124-139.us-east-2.compute.amazonaws.com
```

The authenticity of host 'ec2-13-58-124-139.us-east-2.compute.amazonaws.com (13.58.124.139)' can't be established.  
ECDSA key fingerprint is SHA256:hvWp0w4KiS8zheALmZ1FqMS2YBkVnK4o7E3QTSZGfLc.

Are you sure you want to continue connecting (yes/no/[fingerprint])? yes

Warning: Permanently added 'ec2-13-58-124-139.us-east-2.compute.amazonaws.com,13.58.124.139' (ECDSA) to the list of known hosts.

Last login: Sat Oct 31 15:08:23 2020

```
as
states _|_(_|_) / Amazon Linux 2 AMI Security group ID Security group name VPC ID Description
is _|\_|_|_| sg-009ca1af03bd63dc quicksightaccess vpc-5d74a036 allow quicksight access
https://aws.amazon.com/amazon-linux-2/ sg-0195134765414ec25 ElasticMapReduce-master vpc-5d74a036 Master group for ElasticMapReduce
Run "sudo yum update" to apply all updates. sg-0e584ce26d5dab681 NVIDIA Deep Learning Stack vpc-5d74a036 This security group was created for the NVIDIA Deep Learning Stack
-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory vpc-5d74a036 This security group was created for the NVIDIA Deep Learning Stack
```

```
evasions
EEEEEEEEEEEEEEEEEE MMMMMM MMMMM RRRRRRRRRRRRRR
E:::E M:::M M:::M R:::R
EE:::EEEEEEEEEE E M:::M M:::M R:::R
E:::E EEEEE M:::M M:::M RR:::R R:::R
E:::E M:::M M:::M M:::M R:::R R:::R
E:::EEEEEEEEEE M:::M M:::M M:::M R:::RRRRRR:::R
E:::E M:::M M:::M M:::M R:::RR:::R
E:::EEEEEEEEEE M:::M M:::M M:::M R:::RRRRRR:::R
E:::E M:::M M:::M M:::M R:::R R:::R
E:::E EEEEE M:::M M:::M M:::M R:::R R:::R
EE:::EEEEEEEEEE E M:::M M:::M R:::R
E:::E M:::M M:::M RR:::R R:::R
EEEEEEEEEEEEEEEEEE MMMMMM MMMMM RRRRRRR 0-65533 RRRRRR sg-0195134765414ec25 (ElasticMapReduce-master) -
[hadoop@ip-172-31-45-7 ~]$ TCP 0-65535 sg-0e584ce26d5dab681 (ElasticMapReduce-slave) -
```

## Step 4 : Creating the External Table from data in Amazon S3

Outside data sources are referenced in Amazon EMR by creating an EXTERNAL TABLE. This simply creates a reference to the data; no data is moved yet.

1. Once logged into the master node, start the Hive shell:

```
$ hive
```

2. Define the source using a CREATE TABLE statement. For this example,

```
[hadoop@ip-172-31-45-7 ~]$ hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive> CREATE TABLE statement. For this example, we will only use English 1-grams dataset.
```

```
[hadoop@ip-172-31-45-7 ~]$ hive
Creating a Replica Table in HDFS
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive> show databases;
OK
We create a replica table to store the results on HDFS required for Impala. In the replica table we'll use Parquet instead
of Sequence File format. Parquet is a column-oriented binary file format intended to be highly efficient for running large-
default
Time taken: 0.523 seconds, Fetched: 1 row(s)
hive> use default;
OK
1. Create the replica table in Hive:
Time taken: 0.06 seconds
hive> show tables;
OK
CREATE TABLE eng_1M_1gram_parquet(token STRING, year INT, frequency INT, pages
INT, books INT) ROW FORMAT SERDE 'parquet.hive.serde.ParquetHiveSerDe' STORED AS
Time taken: 0.065 seconds
hive> 'parquet.hive.DeprecatedParquetInputFormat' outputformat
hive> 'parquet.hive.DeprecatedParquetOutputFormat';
```

```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS sales_table(
> region string, sales_channel string,
> country string, order_priority string,
> item_type string, order_date string,
> sales_channel string, order_id STRING,
> order_priority string, ship_date string,
> order_date string, units_sold STRING,
> order_id STRING, unit_price STRING,
> ship_date string, unit_cost STRING,
> units_sold STRING, total_revenue STRING,
> unit_price STRING, total_cost STRING,
> unit_cost STRING, total_profit STRING
> )
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ','
> LOCATION 's3://dezyre-bucket/';
OK
Time taken: 2.69 seconds
hive> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';
```



← → ↻ 🏠 p-36w2u13img7wu.emrappui-prod.us-east-2.amazonaws.com/applicationhistory

People - Home Srikanth Ayalasom... allsec MyTE Gaudium Dashboa... Catalog | Qwiklabs EPF Member Home Support Home - A... Outlook Web App » Other Bookmarks

hadoop

All Applications

Logged in as: dr.who

Application History

- About
- Applications
- FINISHED
- FAILED
- KILLED

Tools

Show 20 entries

ID

User

Name

Application Type

Queue

Application Priority

StartTime

FinishTime

State

FinalStatus

Progress

Tracking UI

application\_1604156601999\_0001

hadoop

HIVE-80644572-b75-49a6-b88d-dad90187fa37

TEZ

default

0

Sat Oct 31 20:53:52 +0550 2020

Sat Oct 31 20:56:36 +0550 2020

FINISHED

ENDED

History

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

```
Time taken: 0.089 seconds, Fetched: 4 row(s)
hive> CREATE EXTERNAL TABLE sales_tbl(
  > region STRING,
  > country STRING,
  > item_type STRING,
  > sales_channel STRING,
  > order_priority STRING,
  > order_date STRING,
  > order_id INT,
  > ship_date STRING,
  > units_sold INT,
  > unit_price INT,
  > unit_cost INT,
  > total_revenue INT,
  > total_cost INT,
  > total_profit INT,
  > ) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION 's3://dezyre-bucket/dezyre'
  > TBLPROPERTIES ("skip.header.line.count"="1");
OK
Time taken: 0.271 seconds
hive> select * from sales_tbl limit 5;
OK
Middle East and North Africa Libya Cosmetics Offline M 10/18/2014 686800706 10-31-2014 8446 437 263 3692591 2224085 1468506
North America Canada Vegetables Online M 11-07-2011 185941302 12-08-2011 3018 154 90 464953 274426 190526
Middle East and North Africa Libya Baby Food Offline C 10/31/2016 246222341 12-09-2016 1517 255 159 387259 241840 145419
Asia Japan Cereal Offline C 04-10-2010 161442649 05-12-2010 3322 205 117 683335 389039 294295
Sub-Saharan Africa Chad Fruits Offline H 8/16/2011 645713555 8/31/2011 9845 9 6 91853 68127 23726
Time taken: 0.089 seconds, Fetched: 5 row(s)
hive> set hive.cli.print.header=true;
hive> select * from sales_tbl limit 5;
OK
sales_tbl.region sales_tbl.country sales_tbl.item_type sales_tbl.sales_channel sales_tbl.order_priority sales_tbl.order_date sales_tbl.order_id
sales_tbl.ship_date sales_tbl.units_sold sales_tbl.unit_price sales_tbl.unit_cost sales_tbl.total_revenue sales_tbl.total_cost sales_tbl.total_profit
```

Here we do using the external table processing the data parsing the columns and creating the new final table and storing data into it.

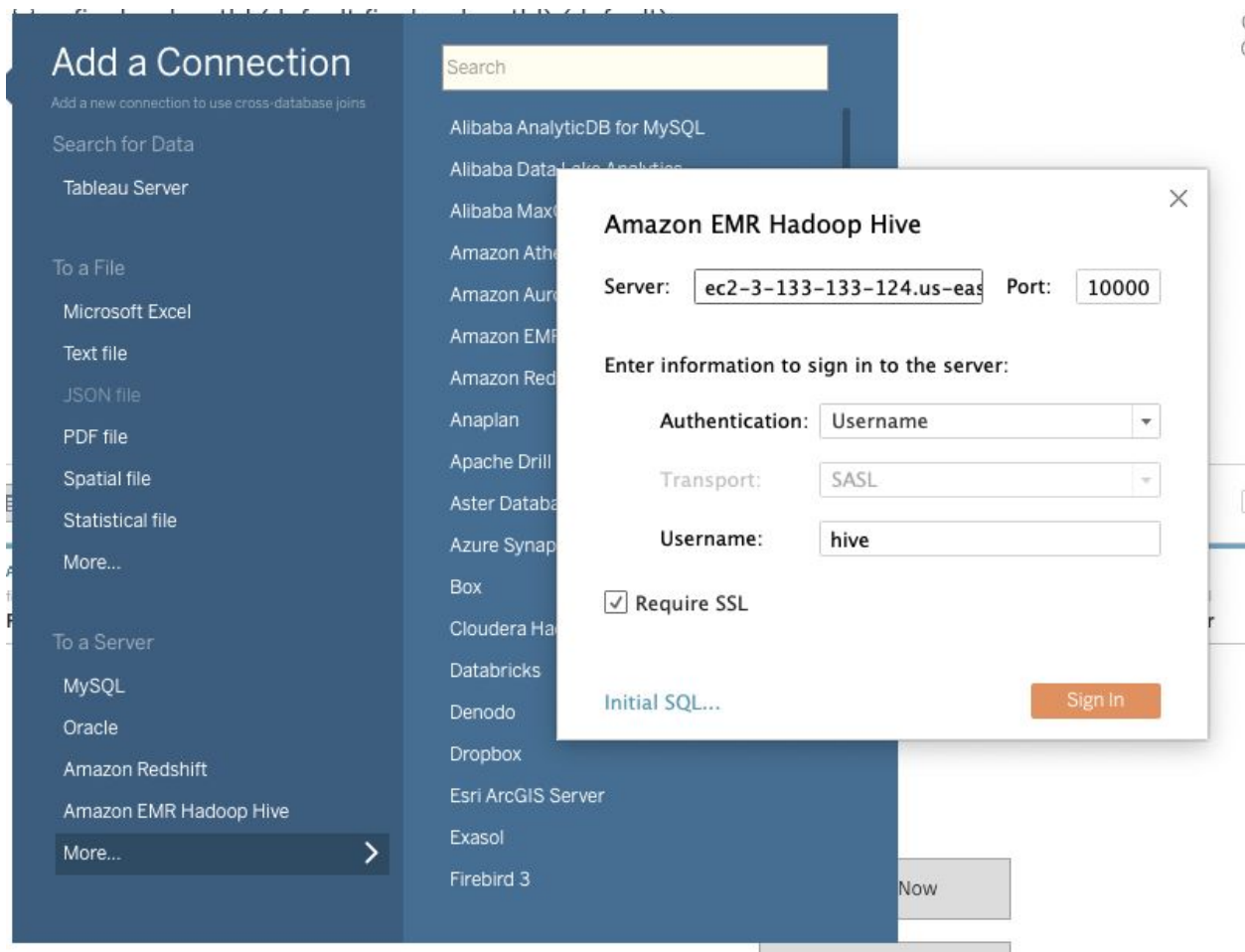
--we add scripts insert statement and final table here.

## Step 5 : Using Tableau to Visualize the Data from EMR HIVE

For the next steps, you'll need the Tableau Desktop trial version installed on a Windows or MacOSX machine.

To connect the Tableau desktop to the EMR Hadoop Hive on AWS, please follow the steps below:

1. Install the ODBC driver on your machine with Tableau Desktop, required for connecting Tableau Desktop to HIVE on Amazon EMR.
  - a. Download the drivers.
  - b. Unzip the downloaded file. This should create a folder named "AmazonEMRHadoopHiveODBC."
  - c. Navigate to the required package for installing the driver.  
Windows: AmazonEMRHadoopHive.msi  
MacOSX: AmazonEMRHadoopHive.dmg
  - d. Run the package above and follow the prompts to install the ODBC driver.
2. Modify the Amazon EMR cluster's Master Security Group so Tableau can connect with the AmazonEMRHadoopHive server running on the master node of the Amazon EMR cluster.
  - a. Click the Amazon EC2 tab in the AWS Management Console to open the Amazon EC2 console.
  - b. In the navigation pane, select Security Groups under the Network and Security group.
  - c. In the Security Groups list, select Elastic MapReduce-master.
  - d. In the lower pane, click the Inbound tab.
  - e. In the Port Range field type 10000. Leave the default value in the Source field.
  - f. Click Add Rule, and then click Apply Rule Changes.
3. Follow the steps as directed by Tableau to enable Amazon EMR Hadoop Hive as a data connection option in Tableau.
4. Got to Add a connexion -> Click on more-> Select Amazon EMR Hadoop Hive and a pop up window appears to enter details .
5. Give your server DNS , port number as 10000
6. Choose Authentication as Username from the dropdown
7. Specify Username as hive
8. Check the RequireSSL box and click on Sign in



- Search for your Database and press enter
- Then search for your schema and press enter
- Drag the table in to work area in workbook in tableau as shown in the below images
- Now, navigate to your sheets and make different visualisations

Tableau - Book1 - Tableau license expires in 5 days

← → ↺ ↻

final\_sales\_tbl (default.final\_sales\_tbl) (default)

Connection

Live

Extract

Filters

0

Add

final\_sales\_tbl

Need more data?

Drag tables here to relate them. [Learn more](#)

Sort fields

Data source order

Show aliases

Show hidden fields

rows

Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order Year	Ship Date	Shipped Year	Units Sold
--------	---------	-----------	---------------	----------------	------------	------------	-----------	--------------	------------

Update Now

Automatically Update

Connections

Add

ec2-3-133-13...amazonaws.com

Amazon EMR Hadoop Hive

Schema

default

Table

final\_sales\_tbl

Exact Contains Starts with

final\_sales\_tbl...inal\_sales\_tbl)

New Custom SQL

Tableau - Book1 - Tableau license expires in 5 days

← → ↺ ↻

final\_sales\_tbl (default.final\_sales\_tbl) (default)

Connection

Live

Extract

Filters

0

Add

final\_sales\_tbl

Need more data?

Drag tables here to relate them. [Learn more](#)

Sort fields

Data source order

Show aliases

Show hidden fields

100 rows

Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order Year	Ship Date	Shipped Year
Sub-Saharan Africa	Mauritius	Personal Care	Online	C	2016-01-01	2016	2016-07-01	2016
Australia and Oceania	Fiji	Vegetables	Offline	H	2013-12-24	2013	2014-01-19	2014
Sub-Saharan Africa	Niger	Baby Food	Online	M	2016-03-01	2016	2016-01-02	2016
Middle East and Nort...	Algeria	Personal Care	Online	H	2011-02-20	2011	2011-09-03	2011
Sub-Saharan Africa	Guinea-Bissau	Office Supplies	Offline	C	2014-05-21	2014	2014-12-06	2014
Sub-Saharan Africa	South Sudan	Meat	Online	C	2013-09-30	2013	2013-01-10	2013
Asia	Philippines	Baby Food	Online	L	2014-02-23	2014	2014-03-23	2014
Asia	Japan	Cereal	Offline	C	2010-10-04	2010	2010-12-05	2010
Sub-Saharan Africa	Gabon	Household	Offline	C	2013-11-11	2013	2013-12-24	2013
Asia	Malaysia	Snacks	Offline	M	2012-06-10	2012	2012-05-11	2012
North America	Canada	Vegetables	Online	M	2011-07-11	2011	2011-08-12	2011
Sub-Saharan Africa	Sudan	Fruits	Online	L	2012-12-26	2012	2013-01-31	2013

Connections

Add

ec2-3-133-13...amazonaws.com

Amazon EMR Hadoop Hive

Schema

default

Table

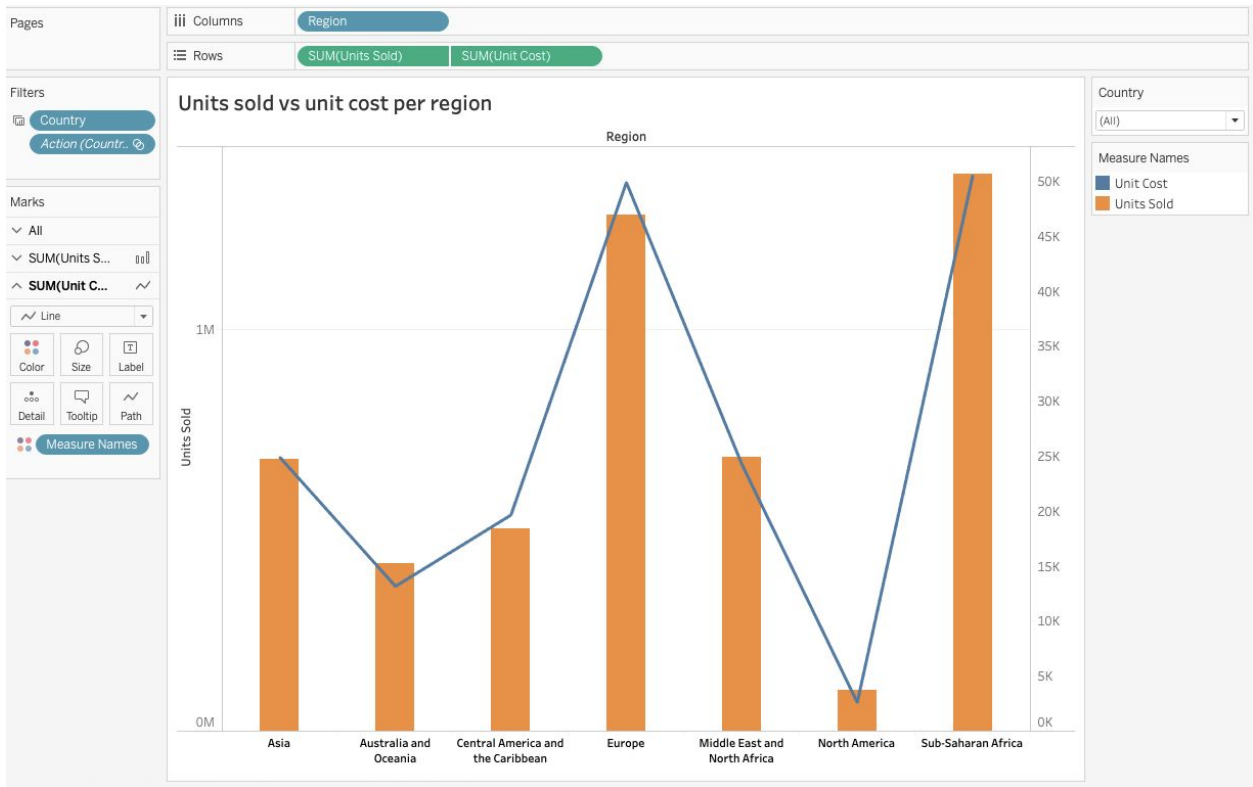
final\_sales\_tbl

Exact Contains Starts with

final\_sales\_tbl...inal\_sales\_tbl)

New Custom SQL

Creating different types of visualizations on sales data using tableau



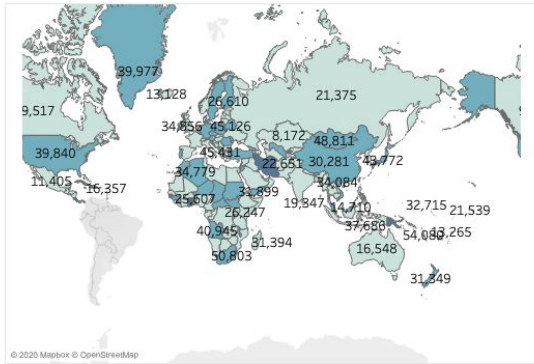




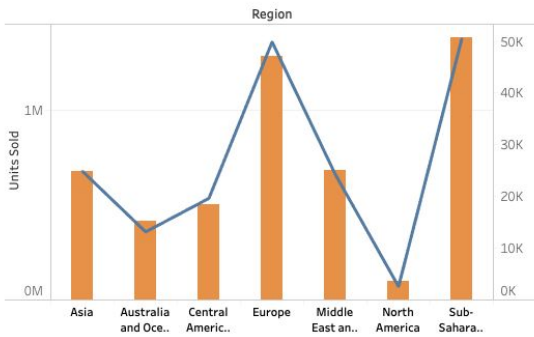


## Sales dashboard

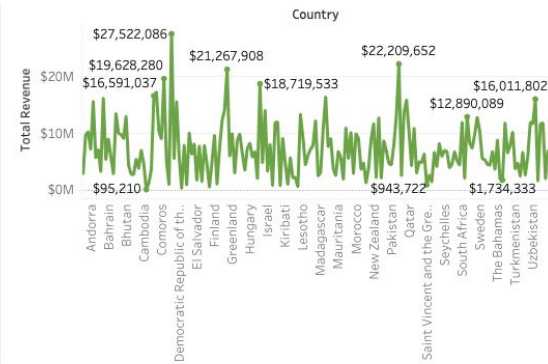
### Units sold by country



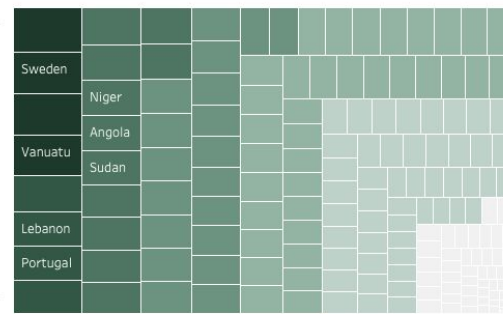
### Units sold vs unit cost per region



### Total revenue and cost per country



### Count of order by country



## Filters

Sales Channel

☐ Offline

Country

(All)



## Sales dashboard

Units sold by country



Total revenue and cost per country

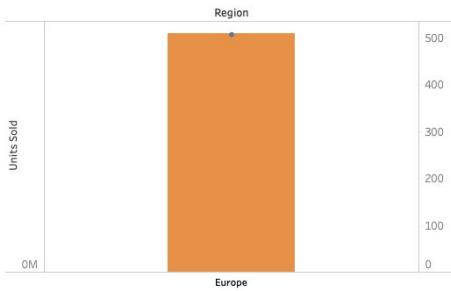


### Filters

Sales Channel  
☐ Offline  
☒ Online

Country  
(All) ▼

Units sold vs unit cost per region



Count of order by country

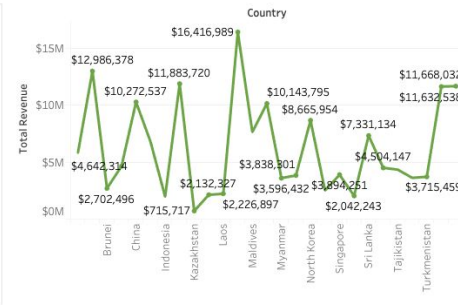


## Sales dashboard

Units sold by country



Total revenue and cost per country



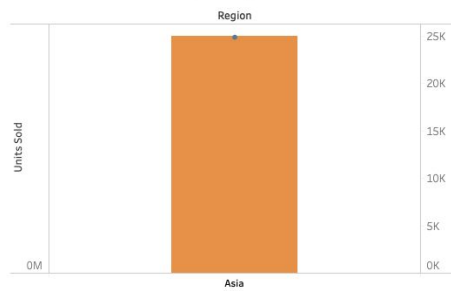
### Filters

Sales Channel  
☐ Offline  
☒ Online

Country  
(All) ▼

Region  
Asia ▼

Units sold vs unit cost per region



Count of order by country

