

|  |
| --- |
| , RD Dep.  **AWS Cloud for Data Engineering** |
| IAM S3 |

# Tasks

**NB:**

* **FOR ALL YOUR TASKS AND ACTIVITIES PLEASE USE Frankfurt (****eu-central-1) REGION AND ENABLE EPAM VPN EU.**
* **FOR ALL RESOURCES YOU CREATE ADD MANDATORY TAG “owner=student”**
* **ALL YOUR REPORTS SHOULD BE COMMITED TO GIT, NOT S3**
* **DO NOT COMMIT OR SHARE ANY CREDENTIALS**

1. Establish connection to AWS account:
   1. Login **AWS Console** with SSO using link provided in the email from Auto EPM-CSUP Notification
   2. Choose “User” to login AWS Console
   3. After you logged in the console in the right upper corner click drop down button

A screenshot of a computer

Description automatically generated with medium confidence

* 1. Click “Switch role” button
  2. Account name – 260586643565

Role – students-role

Display Name (optional) – Fill in the name you like.

Color (optional) – Chose the one you like.

* 1. To use **AWS CLI** please firstly install the client (NB: PLEASE INSTALL VERSION 2 ONLY) https://docs.aws.amazon.com/cli/latest/userguide/getting-started-install.html and after that open SSO link from point b and choose “Access keys”
  2. Run *aws configure sso* command <https://docs.aws.amazon.com/cli/latest/userguide/sso-configure-profile-token.html>:

*SSO session name = user-role*

*SSO start URL = https://epam.awsapps.com/start#*

*SSO region = us-east-1*

*SSO registration scopes –* *keep blank and press enter*

A window in your browser will be opened, enter or verify the code and press continue:

*CLI default client Region = eu-central-1*

*CLI default output format - keep blank and press enter*

*CLI profile name –* *user-role-260586643565*

* 1. Check your aws config file. The config file is located at ~/.aws/config on Linux or macOS, or at C:\Users\USERNAME\.aws\config on Windows. <https://docs.aws.amazon.com/cli/latest/userguide/cli-chap-configure.html>

You should have something like:

*[profile* *user-role-260586643565]*

*sso\_session = user-role*

*sso\_account\_id = 260586643565*

*sso\_role\_name = User*

*region = eu-central-1*

*[sso-session user-role]*

*sso\_start\_url = https://epam.awsapps.com/start#/*

*sso\_region = us-east-1*

*sso\_registration\_scopes = sso:account:access*

* 1. Now assume *students-role* from *user-role-260586643565*, as **only** *students-role* has necessary permissions. To do this add new profile for *students-role* into config file. Here is documentation how you could do this <https://docs.aws.amazon.com/cli/latest/userguide/cli-configure-files.html> [IAM Role section].  
     Profile name should be *students-role,* region = *eu-central-1*
  2. Now run *aws sts get-caller-identity --profile students-role* in case you did all right you should see something like  
     A close-up of words

     Description automatically generated

1. Working with AWS services:
   1. Create your own bucket with recognizable naming (to see who the owner from the name).
   2. Use students-profile to run any aws cli command for s3: at least to put files using api call from local PC to the Cloud.
   3. Upload you tables from you previously created oracle BL\_DM schema to S3 bucket with appropriate with next folders hierarchy: <your\_**bucket**>/<folder\_di\_dwh\_**database**>/<folder\_**schema**\_bl\_dm>/table1/….files…  
      ( <names> should be substituted by or own )

**NOTE:** Please, read AWS documentation before the next step: <https://docs.aws.amazon.com/athena/latest/ug/glue-best-practices.html>

Pay attention at the crawlers naming convention (especially dashes and underscores usage).

* 1. Add path to your tables ( <your\_**bucket**>/<folder\_di\_dwh\_**database**>/<folder\_**schema**\_bl\_dm>/… ) into existing Glue Crawler or create your own Crawler to find your tables in buckets and put them into the Glue DataCatalog Database. Run not very simple select on the tables from your s3 path, i.e. execute query in Athena. Don't forget about limits of 10000 ... 50000 rows. Show us properties of the tables, query results.

All screens, queries with results, intermediate steps result, and mentioned in the task activities please consolidate into the Word-based report and put it into your own bucket.