AWS. Task_01.

Access key in aws console is created.

Access key ID	Created	Last used	Status	
AKIATZLBMUBWYNBXM3TB	2022-03-29 16:18 UTC+0300	2022-03-29 19:38 UTC+0300 with s3 in eu-central-1	Active | Make inactive	✖

SSH keys for AWS CodeCommit

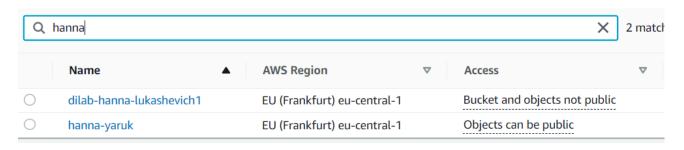Use SSH public keys to authenticate access to AWS CodeCommit repositories. Learn more

Additional profile with mfa token was created with aws cli.

> aws sts get-session-token --serial-number arn:aws:iam::260586643565:mfa/hanna_yaruk@epam.com --token-code <passcode from duo>

```
PS D:\> aws s3 ls

An error occurred (AccessDenied) when calling the ListBuckets operation: Access Denied
PS D:\> aws s3 ls --profile mfa_token
2022-03-29 03:28:42 aksana-shchukina-bucket
2022-03-28 21:05:19 alexandra-makhnach-aws-bucket
2022-03-29 14:37:08 andrei-bialetski-bucket-cli
2022-03-28 21:57:21 andrei-bialetski-bucket-console
2022-03-28 21:52:37 andrei-matveyeu
2022-03-29 13:25:18 anton-sandyuk-bucket
2022-03-29 12:46:35 artiom-dolzhenko-aws-bucket
2021-11-19 15:48:32 athena-bucket-260586643565
2021-11-25 10:34:03 cf-templates-1cjanlre8stsp-eu-central-1
2022-02-27 17:32:04 config-bucket-260586643565
2021-11-17 11:35:11 config-bucket-260586643565-eu-central-1
2022-01-21 13:46:03 dalab-project-stage
2022-01-25 11:56:03 dalab-project-stage-gz
2022-03-29 13:27:20 dilab-hanna-lukashevich1
2022-03-25 13:17:10 dilabbucket
2022-03-29 11:30:21 epam-di-lab-anastasiya-viktarovich
```

S3 bucket was created (aws cli):

```
PS D:\> aws s3 mb s3://hanna-yaruk --region eu-central-1 --profile mfa_token
make_bucket: hanna-yaruk
PS D:\>
```

🔍 hanna   ✕   2 match

	Name ▲	AWS Region ▽	Access ▽
○	dilab-hanna-lukashevich1	EU (Frankfurt) eu-central-1	Bucket and objects not public
○	hanna-yaruk	EU (Frankfurt) eu-central-1	Objects can be public

Limit public access (aws cli):

```
PS D:\> aws s3 mb s3://hanna-yaruk --region eu-central-1 --profile mfa_token
make_bucket: hanna-yaruk
PS D:\> aws s3api put-public-access-block --bucket hanna-yaruk --profile mfa_token --public-access-block-configuration "BlockPublicAcls=true,IgnorePublicAcl
s=true,BlockPublicPolicy=true,RestrictPublicBuckets=true"
PS D:\> |
```

Buckets (41) Info

Buckets are containers for data stored in S3. Learn more ↗

🔍 hanna	✕

2 matc

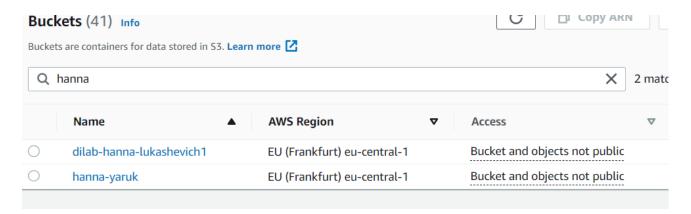|	Name ▲	AWS Region ▽	Access ▽
○	dilab-hanna-lukashevich1	EU (Frankfurt) eu-central-1	Bucket and objects not public
○	hanna-yaruk	EU (Frankfurt) eu-central-1	Bucket and objects not public

Bucket creation logs in CloudTrail:

Delete bucket with aws cli (testing activity):
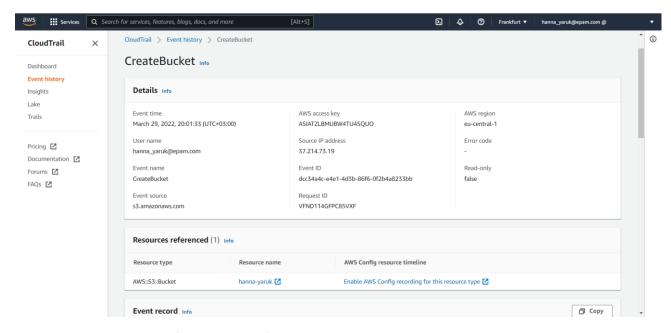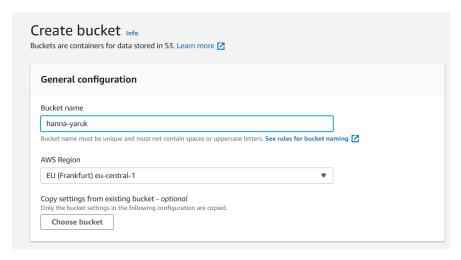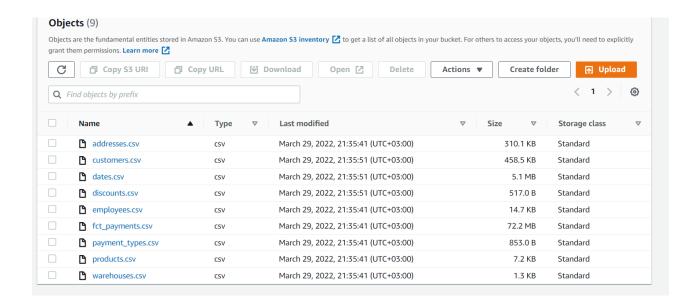
```
PS D:\> aws s3 mb s3://hanna-yaruk --region eu-central-1 --profile mfa_token
make_bucket: hanna-yaruk
PS D:\> aws s3api put-public-access-block --bucket hanna-yaruk --profile mfa
s=true,BlockPublicPolicy=true,RestrictPublicBuckets=true"
PS D:\> aws s3api delete-bucket --bucket hanna-yaruk --profile mfa_token
PS D:\> |
```

Creating bucket in web console

Uploading local files to s3 bucket

```
PS D:\Lab_2022\AWS\online_shop> aws s3 cp . s3://hanna-yaruk --recursive --profile mfa_token
upload: .\employees.csv to s3://hanna-yaruk/employees.csv
upload: .\payment_types.csv to s3://hanna-yaruk/payment_types.csv
upload: .\products.csv to s3://hanna-yaruk/products.csv
upload: .\addresses.csv to s3://hanna-yaruk/addresses.csv
upload: .\warehouses.csv to s3://hanna-yaruk/warehouses.csv
upload: .\discounts.csv to s3://hanna-yaruk/discounts.csv
upload: .\customers.csv to s3://hanna-yaruk/customers.csv
upload: .\fct_payments.csv to s3://hanna-yaruk/fct_payments.csv
upload: .\dates.csv to s3://hanna-yaruk/dates.csv
PS D:\Lab_2022\AWS\online_shop>
```

Objects (9)

Objects are the fundamental entities stored in Amazon S3. You can use **Amazon S3 inventory** ☑ to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. **Learn more** ☑

	Name ▲	Type ▽	Last modified ▽	Size ▽	Storage class ▽
☐	addresses.csv	csv	March 29, 2022, 21:35:41 (UTC+03:00)	310.1 KB	Standard
☐	customers.csv	csv	March 29, 2022, 21:35:51 (UTC+03:00)	458.5 KB	Standard
☐	dates.csv	csv	March 29, 2022, 21:35:51 (UTC+03:00)	5.1 MB	Standard
☐	discounts.csv	csv	March 29, 2022, 21:35:51 (UTC+03:00)	517.0 B	Standard
☐	employees.csv	csv	March 29, 2022, 21:35:41 (UTC+03:00)	14.7 KB	Standard
☐	fct_payments.csv	csv	March 29, 2022, 21:35:41 (UTC+03:00)	72.2 MB	Standard
☐	payment_types.csv	csv	March 29, 2022, 21:35:41 (UTC+03:00)	853.0 B	Standard
☐	products.csv	csv	March 29, 2022, 21:35:41 (UTC+03:00)	7.2 KB	Standard
☐	warehouses.csv	csv	March 29, 2022, 21:35:41 (UTC+03:00)	1.3 KB	Standard

Move files to folders:

```
PS D:\Lab_2022\AWS> aws s3 sync . s3://hanna-yaruk --profile mfa_token
upload: online_shop\bl_dm\payment_types\payment_types.csv to s3://hanna-yaruk/online_shop/bl_dm/payment_types/payment_types.csv
upload: online_shop\bl_dm\products\products.csv to s3://hanna-yaruk/online_shop/bl_dm/products/products.csv
upload: online_shop\bl_dm\warehouses\warehouses.csv to s3://hanna-yaruk/online_shop/bl_dm/warehouses/warehouses.csv
upload: online_shop\bl_dm\employees\employees.csv to s3://hanna-yaruk/online_shop/bl_dm/employees/employees.csv
upload: online_shop\bl_dm\customers\customers.csv to s3://hanna-yaruk/online_shop/bl_dm/customers/customers.csv
upload: online_shop\bl_dm\addresses\addresses.csv to s3://hanna-yaruk/online_shop/bl_dm/addresses/addresses.csv
upload: online_shop\bl_dm\dates\dates.csv to s3://hanna-yaruk/online_shop/bl_dm/dates/dates.csv
upload: online_shop\bl_dm\discounts\discounts.csv to s3://hanna-yaruk/online_shop/bl_dm/discounts/discounts.csv
upload: online_shop\bl_dm\fct_payments\fct_payments.csv to s3://hanna-yaruk/online_shop/bl_dm/fct_payments/fct_payments.csv
PS D:\Lab_2022\AWS> aws s3 sync . s3://hanna-yaruk --delete --profile mfa_token
delete: s3://hanna-yaruk/addresses.csv
delete: s3://hanna-yaruk/customers.csv
delete: s3://hanna-yaruk/discounts.csv
delete: s3://hanna-yaruk/dates.csv
delete: s3://hanna-yaruk/warehouses.csv
delete: s3://hanna-yaruk/employees.csv
delete: s3://hanna-yaruk/products.csv
delete: s3://hanna-yaruk/fct_payments.csv
delete: s3://hanna-yaruk/payment_types.csv
PS D:\Lab_2022\AWS>
```

hanna-yaruk Info

Objects	Properties	Permissions	Metrics	Management	Access Points

Objects (1)

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory ↗ to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more ↗

Copy S3 URI | Copy URL | Download | Open | Delete | Actions ▼ | Create folder | Upload

Find objects by prefix

	Name ▲	Type ▽	Last modified ▽	Size ▽	Storage class ▽
☐	📁 online_shop/	Folder	-	-	-

Crawler creation:

Crawlers

A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog.

Crawler **hanna_yaruk_crawler** was created to run on demand. Run it now?   ✖

User preferences

Add crawler | Run crawler | Action ▼ | Filter by attributes   Showing: 1 - 31 ‹ › ↻ ❓

	Name	Schedule	Status	Logs	Last runtime	Median runtime	Tables updated	Tables added
☐	hanna_lukashevich_crawler		Ready	Logs	55 secs	55 secs	0	8
☑	hanna_yaruk_crawler		Ready		0 secs	0 secs	0	0

All tables are added to crawler

	Name	Schedule	Status	Logs	Last runtime	Median runtime	Tables updated	Tables added
☐	hanna_yaruk_crawler		Ready	Logs	59 secs	59 secs	0	9

Not all tables were recognized correctly (all files have to be .csv)

	Name	Database	Location	Classification	Last updated	Deprecated
☐	addresses	hanna_yaruk	s3://hanna-yaruk/online_shop…	Unknown	29 March 2022 10:07 AM UT…	
☐	dates	hanna_yaruk	s3://hanna-yaruk/online_shop…	csv	29 March 2022 10:07 AM UT…	
☐	warehouses	hanna_yaruk	s3://hanna-yaruk/online_shop…	Unknown	29 March 2022 10:07 AM UT…	
☐	products	hanna_yaruk	s3://hanna-yaruk/online_shop…	csv	29 March 2022 10:07 AM UT…	
☐	customers	hanna_yaruk	s3://hanna-yaruk/online_shop…	Unknown	29 March 2022 10:07 AM UT…	
☐	discounts	hanna_yaruk	s3://hanna-yaruk/online_shop…	csv	29 March 2022 10:07 AM UT…	
☐	payment_types	hanna_yaruk	s3://hanna-yaruk/online_shop…	csv	29 March 2022 10:07 AM UT…	
☐	employees	hanna_yaruk	s3://hanna-yaruk/online_shop…	Unknown	29 March 2022 10:07 AM UT…	
☐	fct_payments	hanna_yaruk	s3://hanna-yaruk/online_shop…	csv	29 March 2022 10:07 AM UT…	

While analyzing the problem was found the difference in data format and – more critical – data in csv was written with mistakes (1 long line was divided into 2). Fixing of all these problems is in process – try to change dates format and to write files in json format. Unfortunately, results will not be represented in time – testing and fixing will take unknown amount of time. Files will be updated as the results of the work.

After debugging all tables were recognized as .csv

Tables A table is the metadata definition that represents your data, including its schema. A table can be used as a source or target in a job definition.

	Name	Database	Location	Classification	Last updated	Deprecated
☐	addresses	hanna_yaruk	s3://hanna-yaruk/online_shop…	csv	1 April 2022 12:01 PM UTC+3	
☐	dates	hanna_yaruk	s3://hanna-yaruk/online_shop…	csv	1 April 2022 12:01 PM UTC+3	
☐	warehouses	hanna_yaruk	s3://hanna-yaruk/online_shop…	csv	1 April 2022 12:01 PM UTC+3	
☐	products	hanna_yaruk	s3://hanna-yaruk/online_shop…	csv	1 April 2022 12:01 PM UTC+3	
☐	discounts	hanna_yaruk	s3://hanna-yaruk/online_shop…	csv	1 April 2022 12:01 PM UTC+3	
☐	customers	hanna_yaruk	s3://hanna-yaruk/online_shop…	csv	1 April 2022 12:01 PM UTC+3	
☐	employees	hanna_yaruk	s3://hanna-yaruk/online_shop…	csv	1 April 2022 12:01 PM UTC+3	
☐	payment_types	hanna_yaruk	s3://hanna-yaruk/online_shop…	csv	1 April 2022 12:01 PM UTC+3	
☐	fct_payments	hanna_yaruk	s3://hanna-yaruk/online_shop…	csv	1 April 2022 12:01 PM UTC+3	

The data were data types are aligned with Oracle data.

The query in Athena works correctly:

```sql
1  select * from
2 ▼ (select
3  c.customer_dm_surr_id,
4      c.first_name,
5      c.last_name,
6      c.country_name,
7      payment_type,
8      sum(quantity) as money
9  from customers c
10 left join fct_payments dd on c.customer_dm_surr_id = dd.customer_dm_surr_id
11 left join payment_types pt on pt.payment_type_surr_id = dd.payment_type_dm_surr_id
12 group by c.country_name, c.customer_dm_surr_id, c.first_name, c.last_name, payment_type, quantity
13 order by money, c.customer_dm_surr_id DESC)
14 limit 10;
```

SQL   Ln 13, Col 38

Run again   Cancel   Save ▼   Clear   Create ▼

Results (10)   🗐 Copy   **Download results**

🔍 Search rows   ‹ 1 › ⚙

#	customer_dm_surr_id	first_name	last_name	country_name	payment_type	money
1	26029	h?l?ne	szachniewicz	france	visaelectron	1
2	25630	gar?on	lanchbery	france	cash	1
3	25557	g?rel	luca	france	cash	1
4	25440	ma?line	lowthorpe	france	visaelectron	1
5	25438	ma?lys	spreag	france	visaelectron	1
6	25336	c?lia	gregore	france	visaelectron	1
7	25336	c?lia	gregore	france	cash	1
8	25335	abraham	cake	belgium	visaelectron	1
9	25199	el	duprey	belgium	visaelectron	1
10	24985	lesya	layfield	belgium	visaelectron	1

To compare here are the results from Oracle

	CUSTOMER_DM_SURR_ID	FIRST_NAME	LAST_NAME	COUNTRY_NAME	PAYMENT_TYPE	MONEY
1	26029	Hélène	Szachniewicz	France	VisaElectron	1
2	25630	Garçon	Lanchbery	France	cash	1
3	25557	Görel	Luca	France	cash	1
4	25440	Maëline	Lowthorpe	France	VisaElectron	1
5	25438	Maëlys	Spreag	France	VisaElectron	1
6	25336	Célia	Gregore	France	cash	1
7	25336	Célia	Gregore	France	VisaElectron	1
8	25335	Abraham	Cake	Belgium	VisaElectron	1
9	25199	El	Duprey	Belgium	VisaElectron	1
10	24985	Lesya	Layfield	Belgium	VisaElectron	1