

School of Computer Science, Engineering and Applications (SCSEA)

B. Tech TY (CCSA)

Subject: Cloud Architecture And Protocol

Name of the Student: Sahil S. Mandawgade

PRN: 20220802265

Title of Practical: 7. AWS Macie: Recognizing and Assessing Sensitive Data with PAN Card Details in Stored Documents.

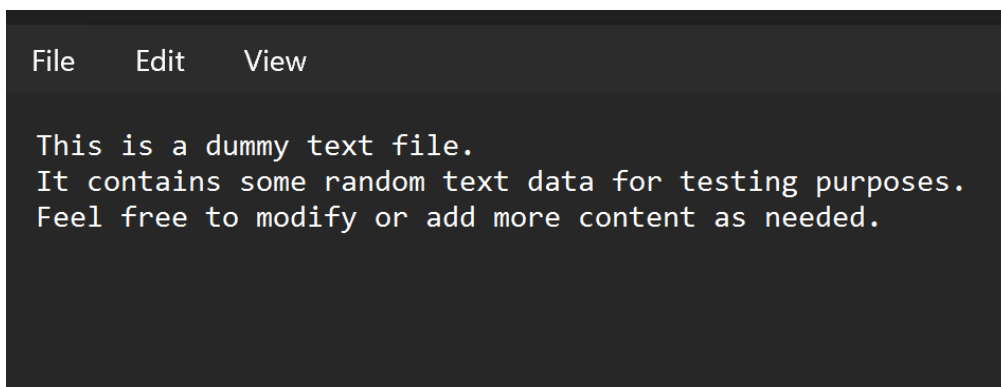
Step 1: Create a S3 Bucket and Upload some sample csv files in it.

- Sign in to the AWS Management Console.
- Navigate to Amazon S3
- Create a bucket (if you don't have one already)
- In the S3 Bucket, upload four files:
 - One .txt file - dummy.txt – contains random text
 - Three .csv files - Sample1.csv (Contains Employee IDs).

Sample2.csv (Contains PAN Card numbers).

Sample3.csv (Contains non-sensitive information).

- Click on Upload.



EmpId	Name	State
MH-01	Amit Sharma	Maharashtra
MH-02	Priya Singh	Maharashtra
MH-03	Rohit Verma	Maharashtra
MH-04	Sneha Joshi	Maharashtra

EmpId	Name	State	Pancard
MH-01-JP	Amit Sharma	Maharashtra	ABCDE1234F
MH-02-KR	Priya Singh	Maharashtra	WXYZA5678G
MH-03-LM	Rohit Verma	Maharashtra	LMNOS9012H
MH-04-NP	Sneha Joshi	Maharashtra	PQRSW3456J

School of Computer Science, Engineering and Applications (SCSEA)

B. Tech TY (CCSA)

Subject: Cloud Architecture And Protocol

Name of the Student: Sahil S. Mandawgade

PRN: 20220802265

Title of Practical: 7. AWS Macie: Recognizing and Assessing Sensitive Data with PAN Card Details in Stored Documents.

City	State
Mumbai	Maharashtra
Pune	Maharashtra
Nagpur	Maharashtra
Nashik	Maharashtra

bucket-2265 [Info](#)

Objects | Properties | Permissions | Metrics | Management | Access Points

Objects (4) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#) [Actions](#) [Create folder](#) [Upload](#)

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](#)

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	dummy.txt	txt	February 27, 2025, 14:11:02 (UTC+05:30)	142.0 B	Standard
<input type="checkbox"/>	sample-1.csv	csv	February 27, 2025, 14:11:02 (UTC+05:30)	137.0 B	Standard
<input type="checkbox"/>	sample-2.csv	csv	February 27, 2025, 14:11:02 (UTC+05:30)	198.0 B	Standard
<input type="checkbox"/>	sample-3.csv	csv	February 27, 2025, 14:11:03 (UTC+05:30)	85.0 B	Standard

Step 2: Create Two Macie Jobs.

- Open Amazon Macie in the AWS Console.
- Enable Macie.
- Go to "Jobs" and click "Create Job".
- In Step 1, Choose your S3 Bucket - (bucket-2265) and click on next.
- In Step 2, review and click next.
- In Step 3, Select 'One Time Job'.
- Under 'Additional settings', in 'Object criteria', type 'csv' in the box and click on 'Include' and the click on next.
- In Step 4, keep everything as is (recommended) and click on next.

School of Computer Science, Engineering and Applications (SCSEA)

B. Tech TY (CCSA)

Subject: Cloud Architecture And Protocol

Name of the Student: Sahil S. Mandawgade

PRN: 20220802265

Title of Practical: 7. AWS Macie: Recognizing and Assessing Sensitive Data with PAN Card Details in Stored Documents.

Step 1
Choose S3 buckets

Choose S3 buckets [Info](#)

A job can analyze objects in one or more S3 buckets. Specify how you want to choose buckets that contain objects for the job to analyze.

Step 2
Review S3 buckets

☒ **Select specific buckets**

Manually select each bucket that contains objects for the job to analyze. If the job runs more than once, it analyzes objects in the same buckets each time it runs.

☐ **Specify bucket criteria**

Enter criteria that determine which buckets contain objects for the job to analyze. If the job runs more than once, it can analyze objects in different buckets each time it runs, as your bucket inventory changes over time.

Step 3
Refine the scope

Select S3 buckets (1/1)

This table lists S3 buckets for your account. Select the check box for each bucket to include in the job's analysis.

 Add filter criteria

<input checked="" type="checkbox"/>	Sensitivity	Bucket	Account	Classifiable objects
<input checked="" type="checkbox"/>	50	bucket-2265	908027405956	5

Step 7
Enter general settings

[Cancel](#)

[Next](#)

Step 1
Choose S3 buckets

Refine the scope [Info](#)

Use these settings to specify how often you want the job to run. You can also specify the depth and scope of the job's analysis.

Step 2
Review S3 buckets

Sensitive data discovery options

☐ **Scheduled job**

Update frequency

Daily

☒ **Include existing objects**

Select this option to analyze new and existing objects. To analyze only new objects, clear this option.

☒ **One-time job**

Analyze existing objects one time only

Sampling depth %

Sample a subset of objects based on depth percentage

Step 6
Select allow lists

School of Computer Science, Engineering and Applications (SCSEA)

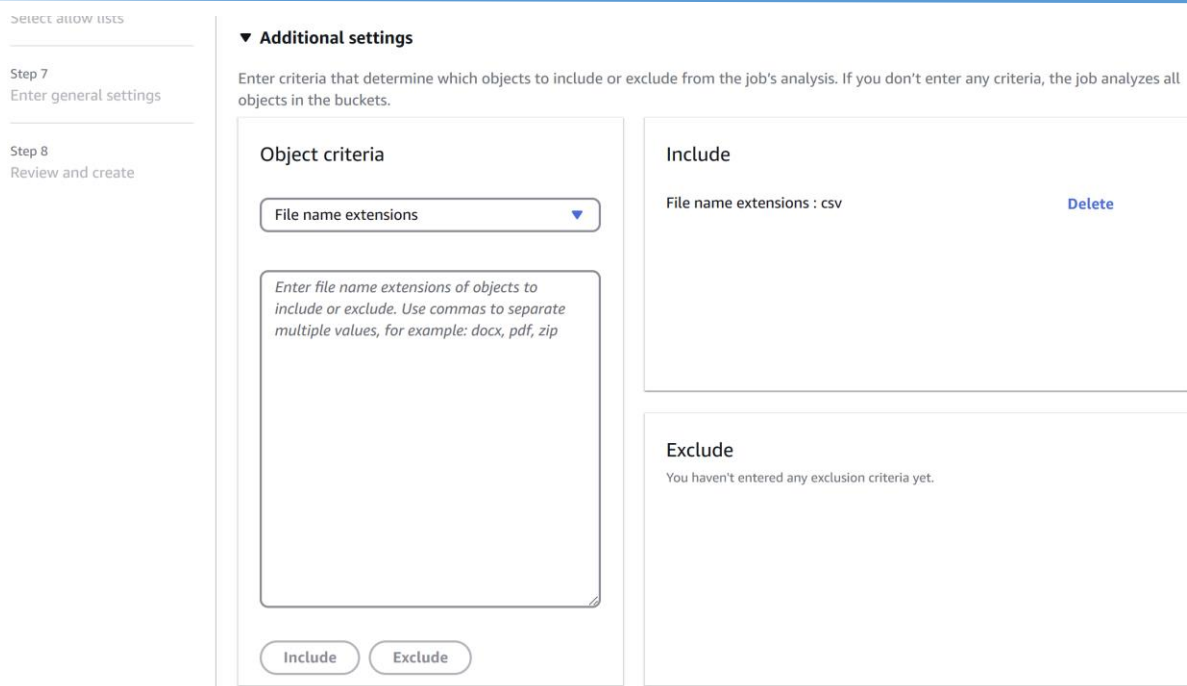
B. Tech TY (CCSA)

Subject: Cloud Architecture And Protocol

Name of the Student: Sahil S. Mandawgade

PRN: 20220802265

Title of Practical: 7. AWS Macie: Recognizing and Assessing Sensitive Data with PAN Card Details in Stored Documents.



The screenshot shows the 'Additional settings' page in AWS Macie. On the left, a sidebar lists steps: 'Step 7: Enter general settings' and 'Step 8: Review and create'. The main content area is titled 'Additional settings' and includes a description: 'Enter criteria that determine which objects to include or exclude from the job's analysis. If you don't enter any criteria, the job analyzes all objects in the buckets.' Below this, there are two main sections: 'Object criteria' and 'Include/Exclude' settings. The 'Object criteria' section has a dropdown menu for 'File name extensions' and a text area for 'Enter file name extensions of objects to include or exclude. Use commas to separate multiple values, for example: docx, pdf, zip'. At the bottom of this section are 'Include' and 'Exclude' buttons. The 'Include' section on the right shows 'File name extensions : csv' with a 'Delete' button. The 'Exclude' section on the right shows a message: 'You haven't entered any exclusion criteria yet.'

- In Step 5, click on 'Manage custom identifiers' to create a custom identifier.
 - Set name – 'PancardID'.
 - Add Regular Expression – '[A-Z]{5}[0-9]{4}[A-Z]{1}'
 - Keep Severity settings as 'Use Medium severity'.
 - Click on submit.
 - Note: You can check whether the regex is correct or not by testing sample data from csv files in 'Evaluate' box.
- Go back to Step 5 and refresh to find newly created identifier.
- Click on next.

School of Computer Science, Engineering and Applications (SCSEA)

B. Tech TY (CCSA)

Subject: Cloud Architecture And Protocol

Name of the Student: Sahil S. Mandawgade

PRN: 20220802265

Title of Practical: 7. AWS Macie: Recognizing and Assessing Sensitive Data with PAN Card Details in Stored Documents.

New custom data identifier [Info](#)

We recommend that you never enter sensitive data in the name or description of an identifier. Other users might be able to see the data in these fields.

Name

Description - optional

Regular expression

Enter the regular expression (regex) that defines the pattern to match.

Keywords - optional

Enter up to 50 keywords, separated by commas (,) that define specific text to match. Macie includes a result for text that contains any of these keywords, if the result matches the regex pattern and is within distance of 100 characters.

Evaluate

Test and refine the identifier before you submit it. You can't edit an identifier after you submit it.

Sample data

[Test](#)

Result

1 match

Step 1
[Choose S3 buckets](#)

Step 2
[Review S3 buckets](#)

Step 3
[Refine the scope](#)

Step 4
[Select managed data identifiers](#)

Step 5
Select custom data identifiers

Select custom data identifiers [Info](#)

A custom data identifier is a set of criteria that you define to detect sensitive data. Select each custom data identifier that you want the job to use.

Custom data identifiers

[Manage custom identifiers](#)

<input checked="" type="checkbox"/>	Identifier name	Description
<input checked="" type="checkbox"/>	PancardID	

[Cancel](#) [Previous](#) [Next](#)

- In Step 6, click on next.
- In Step 7, set job name – 'PanJob-2265' and click on next.
- In Step 8, click on submit.
- The job will take up to a few minutes to complete.

School of Computer Science, Engineering and Applications (SCSEA)

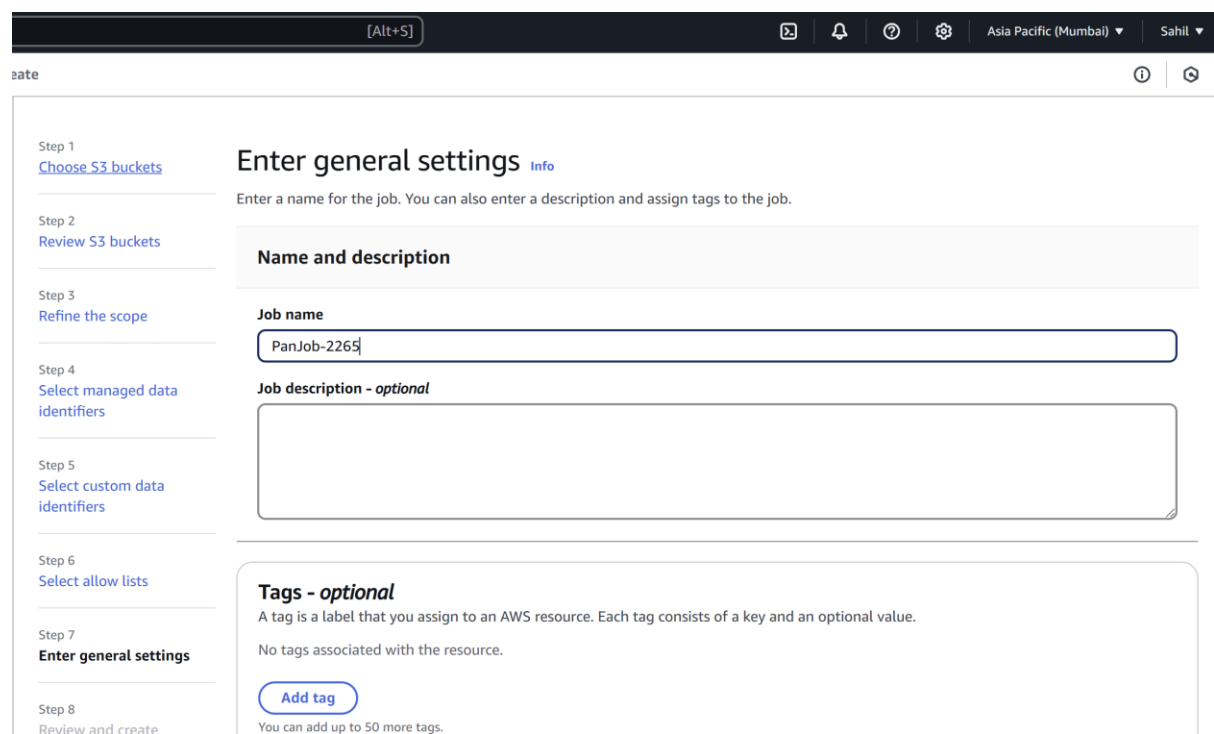
B. Tech TY (CCSA)

Subject: Cloud Architecture And Protocol

Name of the Student: Sahil S. Mandawgade

PRN: 20220802265

Title of Practical: 7. AWS Macie: Recognizing and Assessing Sensitive Data with PAN Card Details in Stored Documents.



[Alt+S]

Asia Pacific (Mumbai) Sahil

Step 1
[Choose S3 buckets](#)

Step 2
[Review S3 buckets](#)

Step 3
[Refine the scope](#)

Step 4
[Select managed data identifiers](#)

Step 5
[Select custom data identifiers](#)

Step 6
[Select allow lists](#)

Step 7
Enter general settings

Step 8
[Review and create](#)

Enter general settings [Info](#)

Enter a name for the job. You can also enter a description and assign tags to the job.

Name and description

Job name

PanJob-2265

Job description - optional

Tags - optional

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value.

No tags associated with the resource.

[Add tag](#)

You can add up to 50 more tags.

- Similarly, create another Job.
- Choose the same S3 Bucket.
- Select 'One-time job' under 'Sensitive data discovery options'.
- Create another data identifier named - 'EmpID'.
- Add Regular Expression - '[A-Z]{2}-[0-9]{2}'.
- Keep Severity as - Medium severity.
- Test sample data from csv file.
- Go back to creating job to select the Custom data identifier.
- Name the 2nd job as - 'EmpJob-2265'.

School of Computer Science, Engineering and Applications (SCSEA)

B. Tech TY (CCSA)

Subject: Cloud Architecture And Protocol

Name of the Student: Sahil S. Mandawgade

PRN: 20220802265

Title of Practical: 7. AWS Macie: Recognizing and Assessing Sensitive Data with PAN Card Details in Stored Documents.

New custom data identifier [Info](#)

We recommend that you never enter sensitive data in the name or description of an identifier. Other users might be able to see the data in these fields.

Name

Description - optional

Regular expression
Enter the regular expression (regex) that defines the pattern to match.

Keywords - optional
Enter up to 50 keywords, separated by commas (,) that define specific text to match. Macie includes a result for text that contains any of these keywords, if the result matches the regex pattern and is within distance of one of these words.

Evaluate

Test and refine the identifier before you submit it. You can't edit an identifier after you submit it.

Sample data

[Test](#)

Result
🕒 1 match

[Alt+S]

📧 🔔 ⚙️ ⚙️ Asia Pacific (Mumbai) ▼ Sahil ▼

[Create job](#)

Jobs (2) [Info](#)

Jobs analyze objects in S3 buckets to discover and report sensitive data. To help ensure accurate results for audits or investigations, you can't change the settings for an existing job.

[Add filter criteria](#)

<input type="checkbox"/>	Job name	Resources	
<input type="checkbox"/>	EmpJob-2265	1	C
<input type="checkbox"/>	PanJob-2265	1	C

PanJob-2265

Job ID: 2a002bbeb7f4ebd58a45c4ee9c8778c6

[Show results](#) 📄 ▼

[Refresh](#)

General information

Job ARN	arn:aws:macie2:ap-south-1:908027405956:classification-job/2a002bbeb7f4ebd58a45c4ee9c8778c6
Created	March 6, 2025, 11:58:38 (10 minutes ago)
Last run time	March 6, 2025, 11:58:44 (10 minutes ago)
Status	🕒 Active (Running)

Statistics

Approximate number of objects to process	0
Number of runs	1

Scope

Job type	One time
Sampling depth	100

Object inclusion criteria

File name extensions	csv
-----------------------------	-----

- As we can see, 2 Jobs have been created.

School of Computer Science, Engineering and Applications (SCSEA)

B. Tech TY (CCSA)

Subject: Cloud Architecture And Protocol

Name of the Student: Sahil S. Mandawgade

PRN: 20220802265

Title of Practical: 7. AWS Macie: Recognizing and Assessing Sensitive Data with PAN Card Details in Stored Documents.

Step 3: View Findings and Extract JSON Output.

- Click on the Job ID to open details.
- Click on "Show Findings".
- Select a finding (related to PAN or Employee ID).
- View and copy the JSON output.

Showing 3 of 3 Severity: Low: 0 Medium: 3 High: 0

Findings (3) Info Actions

This table lists findings for your organization. Select a finding to show its details. You can also filter, group, and sort findings based on specific fields and field values.

Suppress findings Saved rules: No saved rules

Finding status	Filter criteria			
Current	Add filter criteria			< 1 >
<input type="checkbox"/>	Severity	Finding type	Resources affected	Updated at
<input type="checkbox"/>	Medium	SensitiveData:S3Object/CustomIdentifier	bucket-2265/sample-1.csv	4 minutes ago
<input type="checkbox"/>	Medium	SensitiveData:S3Object/CustomIdentifier	bucket-2265/sample-2.csv	4 minutes ago
<input type="checkbox"/>	Medium	SensitiveData:S3Object/CustomIdentifier	bucket-2265/sample-2.csv	13 minutes ago

- We can see Sensitive Data count as 4.

School of Computer Science, Engineering and Applications (SCSEA)

B. Tech TY (CCSA)

Subject: Cloud Architecture And Protocol

Name of the Student: Sahil S. Mandawgade

PRN: 20220802265

Title of Practical: 7. AWS Macie: Recognizing and Assessing Sensitive Data with PAN Card Details in Stored Documents.

Showing 3 of 3 Severity: Low: 0 Medium: 3 High: 0

Findings (3) Info Actions

This table lists findings for your organization. Select a finding to show its details. You can also filter, group, and sort findings based on specific fields and field values.

[Suppress findings](#)

Saved rules [No saved rules](#)

Finding status Filter criteria

Current Add filter criteria

	Severity	Finding type
<input type="checkbox"/>	Medium	SensitiveData:S3Object/CustomIdentifier
<input type="checkbox"/>	Medium	SensitiveData:S3Object/CustomIdentifier
<input type="checkbox"/>	Medium	SensitiveData:S3Object/CustomIdentifier

Showing 1 of 1 Severity: Low: 0 Medium: 1 High: 0

Findings (1) Info Actions

This table lists findings for your organization. Select a finding to show its details. You can also filter, group, and sort findings based on specific fields and field values.

[Suppress findings](#)

Saved rules [No saved rules](#)

Finding status Filter criteria

Current

Job ID: d721c32a69
3a50eecdaf67b022d2c692 Save rule X

Add filter

	Severity	Finding type
<input type="checkbox"/>	Medium	SensitiveData:S3Object/CustomIdentifier

Resource [bucket-2265/sample-2.csv](#)

Created at March 6, 2025, 12:13:15 (4 minutes ago)

Updated at March 6, 2025, 12:13:15 (4 minutes ago)

Origin type SENSITIVE_DATA_DISCOVERY_JOB

Result

Job ID [61644cc04bb53b71caf63906ff3ee55f](#)

Details

Status COMPLETE

Size classified 198 B

MIME type text/csv

Detailed result location [s3://\[export-config-not-set\]/AWSLogs/908027405956/Macie/ap-south-1/61644cc04bb53b71caf63906ff3ee55f/908027405956/f5bfc5df-7aa5-3525-ba78-54ad8f4ea959.json.gz](#)

Sensitive data

Total count 4

Reveal samples [Enter settings](#)

Custom data identifiers

Emp id 4

Details

Status COMPLETE

Size classified 206 B

MIME type text/csv

Detailed result location [s3://\[export-config-not-set\]/AWSLogs/908027405956/Macie/ap-south-1/d721c32a693a50eecdaf67b022d2c692/908027405956/0f7b4adc-4526-393b-a38c-7f6276452654.json.gz](#)

Sensitive data

Total count 4

Reveal samples [Enter settings](#)

Custom data identifiers

Pancard id 4

Resources affected (S3 bucket)

Bucket name	bucket-2265
Public access	NOT_PUBLIC
Encryption required by bucket policy	No
Default encryption	AES256
Created at	February 6, 2025, 14:59:15 (1 month ago)
Owner	null



**D Y PATIL
INTERNATIONAL
UNIVERSITY**
AKURDI PUNE

School of Computer Science, Engineering and Applications (SCSEA)

B. Tech TY (CCSA)

Subject: Cloud Architecture And Protocol

Name of the Student: Sahil S. Mandawgade

PRN: 20220802265

Title of Practical: 7. AWS Macie: Recognizing and Assessing Sensitive Data with PAN Card Details in Stored Documents.

Occurrences of emp id



Read-only ⓘ

```
1 - {}  
2   "arn": "a0852f6b-8225-4b75-a6e6-7db6c7d65f55",  
3   "count": 4,  
4   "name": "EmpID",  
5   "occurrences": {  
6     "cells": [  
7       {  
8         "cellReference": null,  
9         "column": 1,  
10        "columnName": "EmpId",  
11        "row": 2  
12      },  
13     ],  
14     {  
15       "cellReference": null,  
16       "column": 1,  
17       "columnName": "EmpId",  
18       "row": 3  
19     },  
20     {  
21       "cellReference": null,  
22       "column": 1,  
23       "columnName": "EmpId",  
24       "row": 4  
25     },  
26     {  
27       "cellReference": null,  
28       "column": 1,  
29       "columnName": "EmpId",  
30       "row": 5  
31     }  
32   ]  
33 }
```

Cancel

Download



**D Y PATIL
INTERNATIONAL
UNIVERSITY**
AKURDI PUNE

School of Computer Science, Engineering and Applications (SCSEA)

B. Tech TY (CCSA)

Subject: Cloud Architecture And Protocol

Name of the Student: Sahil S. Mandawgade

PRN: 20220802265

Title of Practical: 7. AWS Macie: Recognizing and Assessing Sensitive Data with PAN Card Details in Stored Documents.

Occurrences of pancard id



Read-only ⓘ

```
1 {  
2   "arn": "837377bf-43a0-4858-b26f-6ca2cffe4040",  
3   "count": 4,  
4   "name": "PancardID",  
5   "occurrences": {  
6     "cells": [  
7       {  
8         "cellReference": null,  
9         "column": 4,  
10        "columnName": "Pancard",  
11        "row": 2  
12      },  
13      {  
14        "cellReference": null,  
15        "column": 4,  
16        "columnName": "Pancard",  
17        "row": 3  
18      },  
19      {  
20        "cellReference": null,  
21        "column": 4,  
22        "columnName": "Pancard",  
23        "row": 4  
24      },  
25      {  
26        "cellReference": null,  
27        "column": 4,  
28        "columnName": "Pancard",  
29        "row": 5  
30      }  
31    ]  
32  }  
33 }
```

Cancel

Download