

Using the Cloud Data Loss Prevention (DLP) API for Data Protection



Janani Ravi

CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

Classify and redact sensitive data before writing out to permanent storage

Flexible classification with 90+ pre-defined detectors for sensitive data

Custom detectors for specific use cases

Redaction and de-identification

Introducing the DLP API

Data Loss Prevention

Data loss prevention (DLP) is a strategy for making sure that end users do not send sensitive or critical information outside the corporate network

Data Loss Prevention (DLP) API

Google-provided API for identifying, protecting and redacting sensitive data, before it is exposed to users and other programs

DLP Building Blocks

**InfoType and
infoType Detectors**

Match Likelihood

Job Triggers

Actions

**Redaction and De-
identification**

Templates

DLP Building Blocks

**InfoType and
infoType Detectors**

Match Likelihood

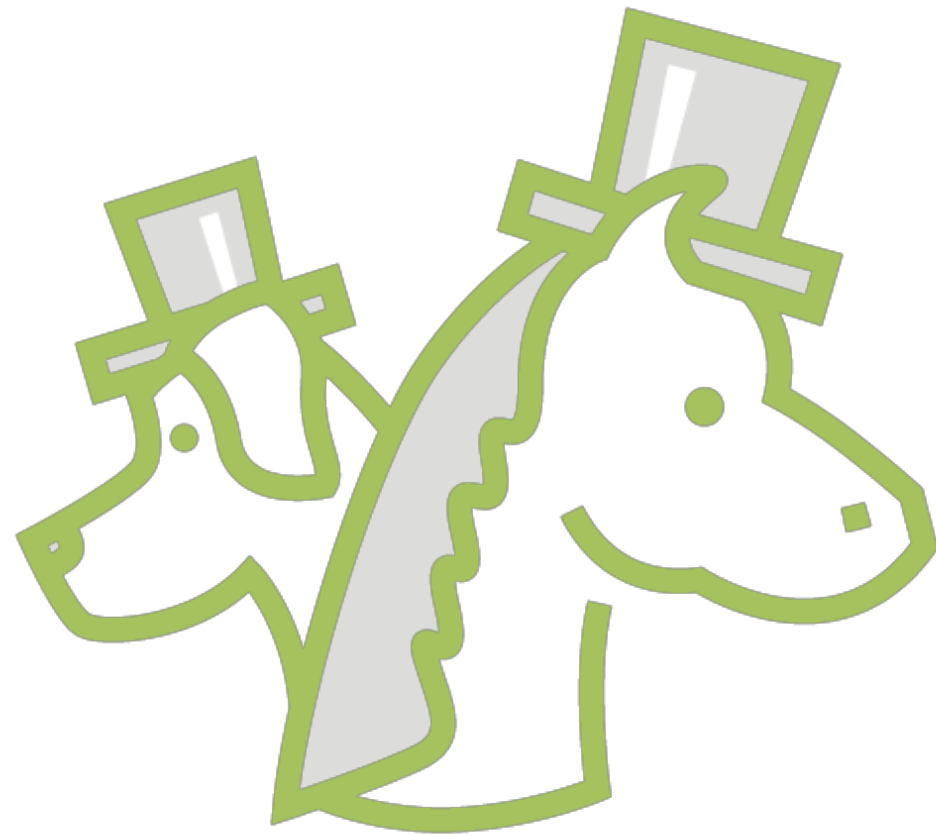
Job Triggers

Actions

Redaction and De-
identification

Templates

InfoTypes



Specific type of sensitive data

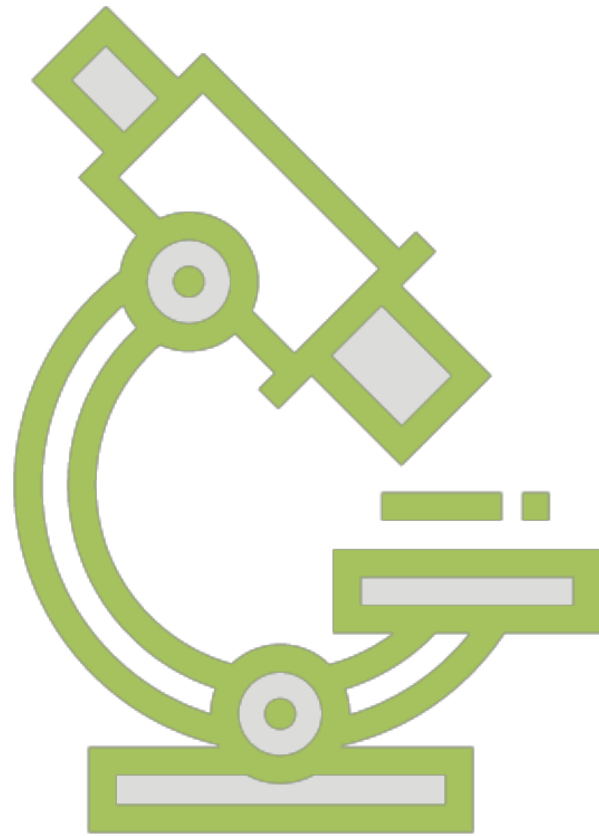
Global and country-specific lists

Long list of enumerated values

- CREDIT_CARD_NUMBER
- US_DRIVERS_LICENSE_NUMBER

<https://cloud.google.com/dlp/docs/infotypes-reference>

infoTypes Detector



Every InfoType has a corresponding detector

Built-in or custom

Built-in detectors are not 100% accurate

Do not rely on them for statutory compliance

DLP Building Blocks

InfoType and
infoType Detectors

Match Likelihood

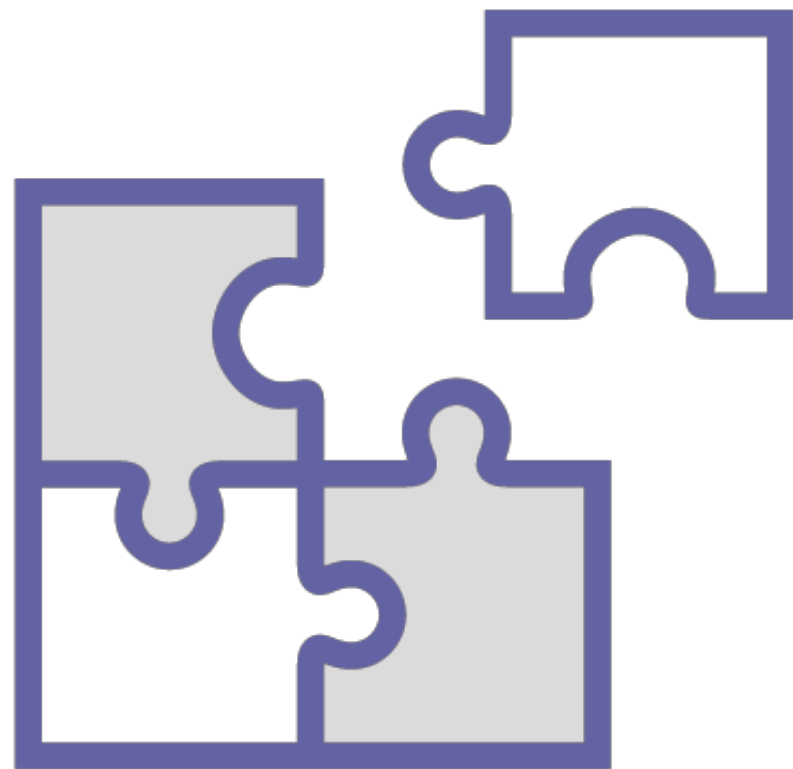
Job Triggers

Actions

Redaction and De-
identification

Templates

Match Likelihood



Results categorized based on how likely they are to represent a match

- LIKELIHOOD_UNSPECIFIED (same as POSSIBLE)
- VERY_UNLIKELY
- UNLIKELY
- POSSIBLE
- LIKELY
- VERY_LIKELY

DLP Building Blocks

InfoType and
infoType Detectors

Match Likelihood

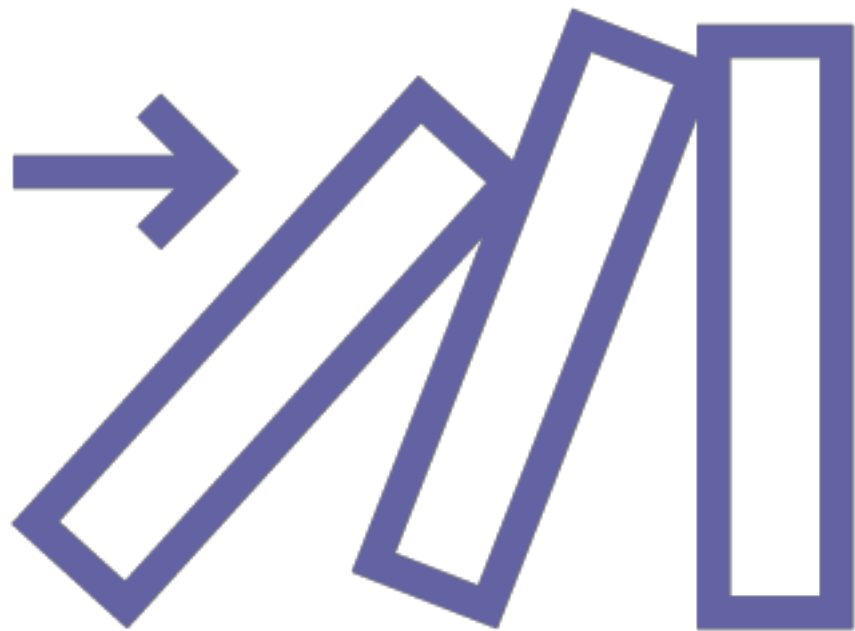
Job Triggers

Actions

Redaction and De-
identification

Templates

Job Trigger



Event that automates creation of DLP job

Scheduled triggers for checks at regular intervals

Can also configure to

- Limit scans to new content
- Trigger on file upload

DLP Building Blocks

InfoType and
infoType Detectors

Match Likelihood

Job Triggers

Actions

Redaction and De-
identification

Templates

Actions



Something that happens after DLP job successfully completes

Save DLP scan job results to BigQuery

Publish to Pub/Sub topic

DLP Building Blocks

InfoType and
infoType Detectors

Match Likelihood

Job Triggers

Actions

Redaction and De-
identification

Templates

Redaction and De-identification

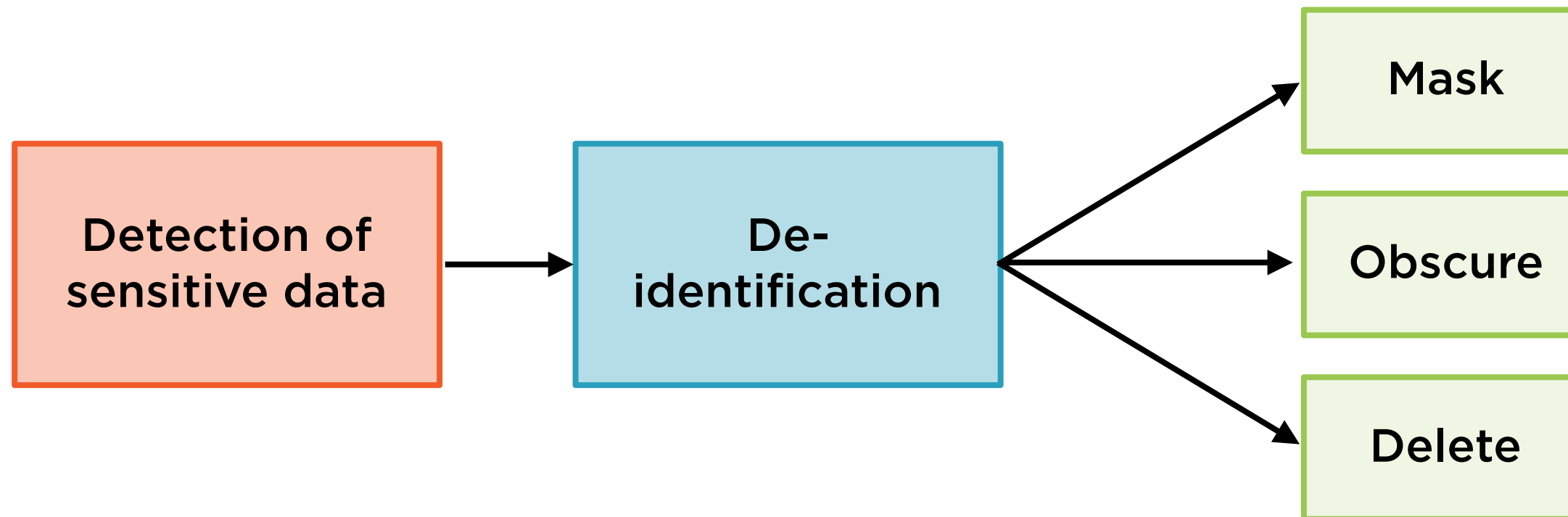


Classification of sensitive content whether text or images

Redaction removes sensitive matches

De-identification remove identifying features from data

De-identification of Sensitive Data



De-identification Techniques



Date shifting: Shift dates but preserve the sequence and duration of a period of time

Generalization: Making a value more generic, reducing how identifiable it is

- Bucketing to categorize data

Pseudonymization: Sensitive data replaced with surrogates or tokens

DLP Building Blocks

InfoType and
infoType Detectors

Match Likelihood

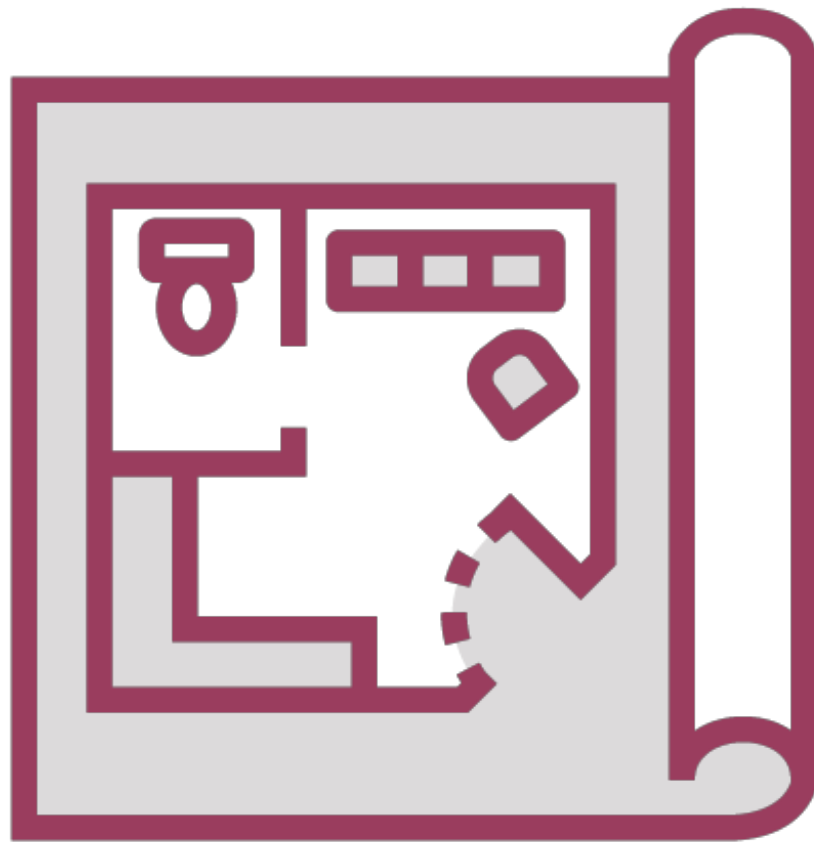
Job Triggers

Actions

Redaction and De-
identification

Templates

Templates



Create and persist configuration information to use with DLP APIs

- Inspection templates for inspection scan jobs
- De-identification templates

Risk Analysis

Risk analysis is the process of analyzing sensitive data to find properties that might increase the risk of subjects being identified.

Risk Metrics



k-anonymity metric

l-diversity metric

k-map metric

δ -presence metric

Not covered in this course

Custom InfoType Detectors

Custom InfoType Detectors

**Regular custom
dictionary
detectors**

**Stored custom
dictionary
detectors**

**Regular
expressions**

Regular Custom Dictionary Detectors

**Regular custom
dictionary
detectors**

Match a short list of words or phrases
Can act as its own unique detector

Stored Custom Dictionary Detectors

**Stored custom
dictionary
detectors**

More than a few words or phrases to scan

Can match on up to tens of millions of words or phrases


Components

- List of phrases
- Generated dictionary files

Regular Expressions



**Regular
expressions**



**Allows you to create your own infoType
detectors**

Inspection Rules



Exclusion rules

- Decrease the quantity or precision of findings returned

Hotword rules

- Increase the quantity or accuracy of findings returned

Demo

Authenticating to the DLP API using a service account

Performing a simple inspection of text content

Demo

Executing a DLP job to detect sensitive data in files stored in Cloud Storage buckets

Storing inspection results in BigQuery

Demo

**Registering and using custom
inspection templates**

Demo

**Redaction and de-identification of text
as well as image data**

Demo

**Creating and using a custom infoType
detector**

Demo

**Creating a Cloud Function to classify
data uploaded to a GCS bucket**

Summary

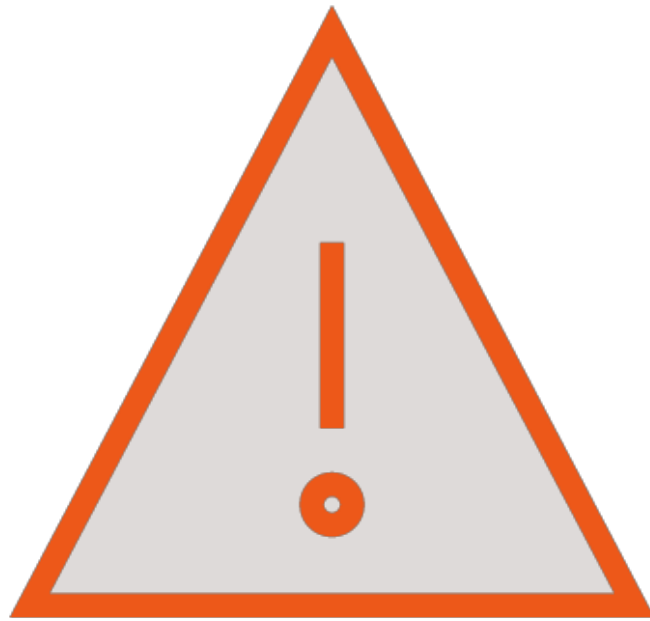
Classify and redact sensitive data before writing out to permanent storage

Flexible classification with 90+ pre-defined detectors for sensitive data

Custom detectors for specific use cases

Redaction and de-identification

Delete Resources



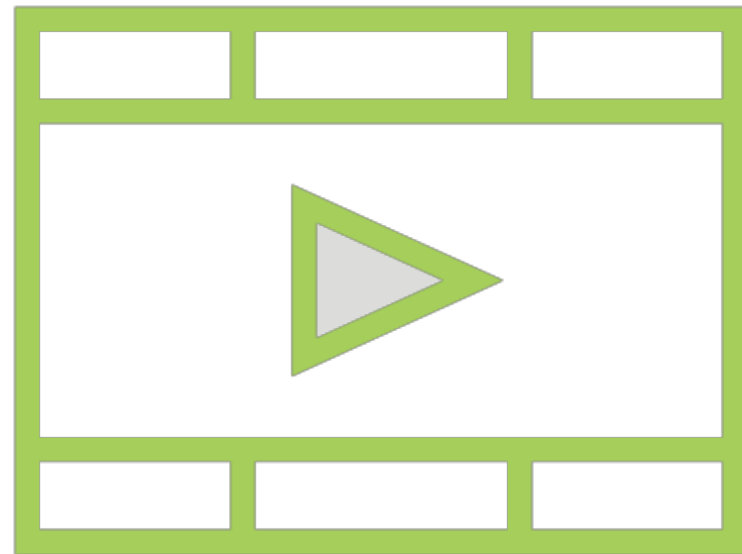
VM instances running the web application

Cloud Storage buckets

Clean up all resources linked to Forseti

- Cloud SQL instances
- VMs
- Storage buckets
- Deployments

Related Courses



Implementing Customer Managed Encryption Keys (CMEK) with Google Key Management Service

Ethical Hacking: Understanding Ethical Hacking

Implementing and Performing Vulnerability Management