Data Security Center

FAQs

Issue 02

Date 2024-09-25

HUAWEI

HUAWEI CLOUD COMPUTING TECHNOLOGIES CO., LTD.

Trademarks and Permissions


and other Huawei trademarks are the property of Huawei Technologies Co., Ltd.
All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei Cloud and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Cloud Computing Technologies Co., Ltd.

Address: Huawei Cloud Data Center Jiaoxinggong Road
Qianzhong Avenue
Gui'an New District
Gui Zhou 550029
People's Republic of China

Website: https://www.huaweicloud.com/intl/en-us/

Contents

1 Product Consulting

1.1 What is Data Security Center?

The Data Security Center (DSC) is a next-generation, cloud-native data security management platform. It offers fundamental data security features, including data classification and grading, data masking, data watermarking, and API data protection. The DSC visualizes the overall data security posture in the cloud via an asset map and facilitates comprehensive, one-stop data security operations.

1.2 Does DSC Store My Data Assets or Files?

DSC does not store your data or files. DSC only identifies, anonymizes, or watermarks the data from the data sources you authorize DSC to access.

The data identification results are displayed on the DSC console. For details, see **Identification Results**.

1.3 What Types of Unstructured Files Can DSC Parse?

Table 1-1, **Table 1-2**, and **Table 1-3** list the types of unstructured files that can be parsed by DSC.

Table 1-1 Text and code files

No.	File Type	No.	File Type
1	Access database file	74	PDF document
2	ARFF file	75	Perl source code
3	ASP file	76	PGP file
4	ATOM file	77	PHP source code
5	BAT file	78	PKCS7 digital certificate file

No.	File Type	No.	File Type
6	BCPL source code	79	Plist file
7	BIB file	80	PostgreSQL database file
8	C# source code	81	PostScript document
9	C/C+ source code	82	PowerPoint document
10	CAD SldWorks file	83	Properties file
11	CAD document	84	Publisher file
12	CBOR file	85	Python source code
13	CFG file	86	Quattro-Pro spreadsheet
14	CHM file	87	Redis database file
15	Com executable file	88	RSS file
16	CSS file	89	RTF document
17	DataX configuration file	90	Ruby source code
18	DBF file	91	R source code
19	DIF file	92	SAS7BDAT file
20	DITA file	93	SAS file
21	Djvu Document	94	Scala source code
22	DOS executable file	95	Shell script
23	D source code	96	SQLite 3 database file
24	ELF executable file	97	SQLServer database file
25	EPUB eBook file	98	SQL source code
26	Excel document	99	SSH public key
27	FDF document	100	SSH configuration file
28	Fictionbook XML file	101	SSH private key
29	FTP session file	102	Staroffice document
30	GnucCash financial XML file	103	Swift source code
31	Go source code	104	TAB file
32	Groovy source code	105	TCL source code
33	HDR file	106	TEXT file

No.	File Type	No.	File Type
34	HOCON file	107	TFF file
35	HTML file	108	TNEF file
36	HTM file	109	Tomcat Application configuration file
37	HWP file	110	Tomcat Users configuration file
38	Ibooks file	111	Tomcat configuration file
39	Iis configuration file	112	TOML file
40	Initialization file	113	TSD file
41	ISA-Tab file	114	TSV file
42	iWork document	115	VCS file
43	Java Jce Keystore file	116	Visio document
44	Java Keystore file	117	Visual Basic source code
45	JavaScript source code	118	Virtual Reality Modeling Language (VRML) code
46	Java source code	119	Web Archive file
47	JSON file	120	WebLogic configuration file
48	JSP source code	121	WebVTT file
49	LaTeX source code	122	Windowsinf file
50	Log file	123	Windows full-text search index
51	Lua source code	124	Windows precompilation file
52	MariaDB database file	125	WordPerfect document
53	Markdown document	126	DOC file
54	Matlab source code	127	WPD document
55	Mbox file	128	WPS document
56	MIME HTML file	129	XDP file
57	Microsoft Reader documentation	130	XFDF file

No.	File Type	No.	File Type
58	MongoDB database file	131	XHTML file
59	MRS configuration file	132	XLF file
60	Microsoft Works document	133	XLIFF file
61	MySQL database file	134	XLR file
62	NetCDF file	135	XLZ file
63	Objective-C source code	136	XML sitemap file
64	OBS configuration file	137	XML File
65	Office document	138	XMP file
66	OneNote file	139	XPS document
67	OpenDocument file	140	XPT file
68	OpenVPN configuration file	141	YAML file
69	Oracle database file	142	Common digital certificate files
70	Outlook file	143	Empty file
71	PASCAL source code	144	Configuration file **Windows Initialization**
72	PBM file	145	Other unencrypted text files
73	PCX file	146	Email document

Table 1-2 Compressed and binary files

No.	File Type	No.	File Type
1	7-Zip file	26	Lha compressed file
2	APK Android program	27	LZ4 compressed file
3	ARJ file	28	LZMA compressed file
4	AR file	29	MAT file
5	BGP file	30	NetCDF file
6	Brotli compressed file	31	Object file

No.	File Type	No.	File Type
7	Bzip2 compressed file	32	Pack200 compressed file
8	Bzip compressed file	33	RAR compressed file
9	Cabinet compressed file	34	ShareLib file
10	Core dump file	35	Snappy compressed file
11	CPIO compressed file	36	TAR compressed file
12	Deflate64 compressed file	37	TCP dump file
13	DMG file	38	Tika-Unix-Dump file
14	ELF executable file	39	UNIX compressed file
15	GDAL file	40	Xcompress compressed file
16	GRB file	41	XLZ compressed file
17	GRIB2 file	42	XPI Firefox plug-in installation package
18	GRIB file	43	XZ compressed file
19	GZIP file	44	ZIP compressed file
20	HDF file	45	Zlib compressed file
21	HE5 file	46	ZSTD compressed file
22	ISO-19139 geographic information file	47	ZSTD dictionary file
23	ISO compressed file	48	Z compressed file
24	JAR file	49	Executable file
25	Java Class file	50	Common compressed file

Table 1-3 Images

No.	File Type	No.	File Type
1	BMP file	4	JFIF file
2	PNM file	5	JPEG file
3	PNG file	6	TIFF file

2 Regions and AZs

2.1 What Are Regions and AZs?

Concepts

A region and availability zone (AZ) identify the location of a data center. You can create resources in a specific region and AZ.

- Regions are divided from the dimensions of geographical location and network latency. Public services, such as Elastic Cloud Server (ECS), Elastic Volume Service (EVS), Object Storage Service (OBS), Virtual Private Cloud (VPC), Elastic IP (EIP), and Image Management Service (IMS), are shared within the same region. Regions are classified as universal regions and dedicated regions. A universal region provides universal cloud services for common tenants. A dedicated region provides services of the same type only or for specific tenants.
- An AZ contains one or more physical data centers. Each AZ has independent cooling, fire extinguishing, moisture-proof, and electricity facilities. Within an AZ, computing, network, storage, and other resources are logically divided into multiple clusters. AZs within a region are interconnected using high-speed optical fibers to allow you to build cross-AZ high-availability systems.

Selecting a Region

If you or your users are in Europe, select the **EU-Dublin** region.

Selecting an AZ

When determining whether to deploy resources in the same AZ, consider your applications' requirements on disaster recovery (DR) and network latency.

- For high DR capability, deploy resources in different AZs in the same region.
- For low network latency, deploy resources in the same AZ.

2.2 Can DSC Be Used Across Regions?

DSC can only be used to identify and mask the sensitive data in resources of the region where DSC is located.

3 Asset Authorization

3.1 Agency Policies Obtained After the Access To Assets Is Allowed

After the access to cloud resources is allowed, DSC can access your OBS buckets, databases, big data assets, and asset map. **Table 3-1** describes the agency policies obtained after the access is allowed.

Table 3-1 Agency policies

Asset	Policy	Scope	Remarks
OBS	OBS Administrator	Global	Used to configure OBS logs, obtain the OBS bucket list, and download items form OBS.
	EVS ReadOnlyAccess	Regional	Used to obtain the EVS disk list.
	OBS Administrator	Global	Used to obtain the logs delivered by OBS.
Database	ECS ReadOnlyAccess	Regional	Used to obtain the list of ECSs where databases are built.
	RDS ReadOnlyAccess	Regional	Used to obtain the RDS database list and related information.
	DWS ReadOnlyAccess	Regional	Used to obtain the DWS instance list.
	VPC FullAccess	Regional	Used to establish network connection and create VPC ports and security group rules
	KMS CMKFullAccess	Regional	Used to perform encryption using KMS in data masking.

Asset	Policy	Scope	Remarks
	GaussDB ReadOnlyAccess	Regional	Used to obtain the GaussDB list.
Big Data	ECS ReadOnlyAccess	Regional	Used to obtain the list of ECSs where big data sources reside.
	CSS ReadOnlyAccess	Regional	Used to obtain the CSS data cluster list and data indexes.
	DLI Service User	Regional	Used to obtain the DLI queue and database.
	VPC FullAccess	Regional	Used to establish network connection and create VPC ports and security group rules.
	KMS CMKFullAccess	Regional	Used to perform encryption using KMS in data masking.
MRS	MRS CommonOperations	Regional	Used for cluster query and task creation.
Asset Map	Tenant Guest	Regional	Used to obtain the list of cloud services used for data storage and processing.
	OBS Administrator	Global	Used to configure OBS logs, obtain the OBS bucket list, and download items form OBS.
	EVS ReadOnlyAccess	Regional	Used to obtain the EVS disk list.
	OBS Administrator	Global	Used for OBS to deliver logs.
LTS	LTS ReadOnlyAccess	Region	Used to read LTS log groups or log streams.

3.2 How Do I Troubleshoot the Failure in Connecting to the Added Database?

DSC will check the connectivity of the added database. If the connection to the added database fails, perform the following operations to troubleshoot the problem:

Step 1 Check whether the IP address, account, password, and database name of the added database are correct.

- If no, correct it.

- If yes, go to **2**.

Step 2 Check whether all ports and protocols are bypassed in the outbound direction of the security group which the added database belongs.

- If no, add outbound rules for the security group. Add the database to DSC again after all ports and protocols are bypassed in the outbound direction of the security group. If the failure persists, go to **3**.
- If yes, go to **3**.

Step 3 Check whether the number of available IP addresses in the IP subnet corresponding to the database is 0.

At least one IP address is required for DSC to establish connection to the added database. If the number of available IP addresses in the IP subnet corresponding to the database is 0, add available IP addresses to the database.

----End

4 Sensitive Data Identification and Masking

4.1 What Services Can Use DSC to Scan for Sensitive Data?

DSC can scan data stored in OBS, RDS, CSS, DLI, or GaussDB(DWS) for sensitive information by using built-in and customized rules.

The following table lists the data sources supported by DSC and identification restrictions.

Table 4-1 Supported data sources

Data Source	Data Type	Restriction
RDS	MySQL, SQL Server, and PostgreSQL	The first 500 lines of data records are sampled and scanned. The QPS reaches 300 times per second.
CSS	Big data asset	N/A
OBS	More than 200 file types	Files larger than 200 MB or encrypted files in the OBS bucket cannot be scanned.
DWS	N/A	N/A
ECS	Data in MySQL, SQL Server, PostgreSQL, TDSQL, and Oracle databases, as well as Elasticsearch instances	N/A
Data Lake Insight (DLI)	Big data asset	N/A

4.2 How Long Does It Take for DSC to Identify and Mask Sensitive Data?

Identification Duration

The identification duration depends on the data volume, number of identification rules, and scan mode. The information provided in **Table 4-2** is for reference only.

Table 4-2 Identification duration

Data Source	Data Volume	Scan Mode	Duration (Minutes)
RDS	1,000 tables	Quick scan	5
CSS	10 million documents	Quick scan	15
OBS	100 MB	Quick scan	1
OBS	100 MB	Full scan	15

Data Masking Duration

DSC uses preset and customized masking algorithms to mask sensitive data stored in RDS, MRS, Hive,and Elasticsearch. **Table 4-3** describes the masking duration.

Table 4-3 Data masking duration

Data Source	Data Volume	Duration (Minutes)
RDS	10 million lines	40
Elasticsearch	10 million documents	40

4.3 What Are the Identification Templates Supported by DSC?

DSC data identification templates are industry-specific. These templates can make data comply with regulations. For details about the templates supported by DSC, see **Table 4-4**.

You can also customize templates. A maximum of 20 identification templates are supported.

Built-in Huawei Cloud Data Categorization and Leveling Templates

Table 4-4 Built-in categorization and leveling templates

Level-1 Category	Level-2 Category	Sensitivity Level	Built-in Rule
Personal information	Authoritative social ID	L3	ID card No. (Chinese mainland)
		L3	Passport No. (Chinese mainland)
		L3	Driver's license No. (Chinese mainland)
		L3	Exit-Entry Permit for Traveling to and from Hong Kong and Macau (EEP)
		L3	Taiwan compatriot permit
		L2	Car license plate number (Chinese mainland)
		L3	Military ID card number
		L3	American social security number (SSN)
		L3	ITIN
		L3	Social security information
		L2	Vehicle identification number
	General personal information	L1	Name (Simplified Chinese)
		L1	Name (English)
		L2	Nationality
		L2	Gender
		L2	Ethnicity

Level-1 Category	Level-2 Category	Sensitivity Level	Built-in Rule
		L2	Birthday
		L2	Birth place
		L2	Education level
		L2	Company
		L2	Industry
		L2	Telephone number (Chinese mainland)
		L3	Mobile number (Chinese mainland)
		L3	Email address
		L2	WeChat ID
		L2	QQ account
	Personal private information	L4	Marital status
		L4	Family member relationship
		L4	Religion
	Real-name authentication certificate	L4	Driving license image (Chinese mainland)
		L4	Bank card image (Chinese mainland)
		L4	ID card image (Chinese mainland)
		L4	Image of motor vehicle registration certificate (Chinese mainland)
		L4	Passport image (Chinese mainland)

Level-1 Category	Level-2 Category	Sensitivity Level	Built-in Rule
		L4	Auto insurance policy image (Chinese mainland)
		L4	Motor vehicle license image (Chinese mainland)
	Bank account information	L3	Bank account number
		L3	Credit card number
		L3	MasterCard credit card number
		L3	VISA credit card number
		L4	Credit card security code
Enterprise information	Enterprise ID information	L1	Business registration number
		L1	Unified social credit code
		L1	Taxpayer identification number (tax number)
		L1	Organization code
		L1	Business license image
	Publicly disclosed information	L1	Enterprise type
		L1	Operation status
	Enterprise internal information	L2	Enterprise delivery information
		L3	Enterprise planning information

Level-1 Category	Level-2 Category	Sensitivity Level	Built-in Rule
		L2	Enterprise requirement information
Device information	Terminal ID information	L2	International mobile equipment identity (IMEI)
		L2	Mobile equipment identity (MEID)
		L2	MAC address
		L2	SIM card IMSI information.
	IP address	L2	IPv4 address
		L2	IPv6 address
	Terminal configuration information	L3	Linux-Passwd file
		L3	Linux-Shadow file
General information	Time information	L1	Date
		L1	Time
	Location	L4	GPS data
		L4	Exact address (China)
		L2	Province (Chinese mainland)
		L2	Postal code (Chinese mainland)
		L2	City (Chinese mainland)
		L2	Municipality (China)
		L3	Address (Chinese mainland)
	Key credential information	L3	SSL Certificate
		L3	Access_Key_Id
		L4	Secret_Access_Key
		L3	AWS_ACCESS_KEY

Level-1 Category	Level-2 Category	Sensitivity Level	Built-in Rule
		L4	AWS_SECRET_KEY
		L4	Facebook_SECRET
		L4	IAM op_service account and password
		L4	GitHub_KEY
		L4	DSA private key
		L4	EC private key
		L4	Encryption private key
		L4	RSA private key
	System network information	L2	URL link
		L2	LDAP
		L1	OS

4.4 Does Data Masking Affect My Raw Data?

No. The sensitive data masking function only reads data, masks sensitive information, and saves the data in a specified path without changing your raw data.

CAUTION

- Do not fill in an existing service data table. Otherwise, services may be affected.
- Do not select an original data table as the target data table. Otherwise, the original data may be overwritten.

4.5 Does DSC Have Specific Requirements on the Character Set for Which Sensitive Data Is to Be Identified and Masked?

The DSC identification and masking functions have no requirements on the database encoding formats.

Only UTF-8 is supported for UDF based masking of MRS data sources.

5 Data Watermarking

5.1 Will the Source Data Be Modified During Data Watermarking?

The source data will not be modified during data watermarking.

DSC injects watermarks into the files stored in the OBS bucket or local directory and generates the watermarked files. The files will be automatically downloaded to the directory specified, and there is no any modification to the source data. For details, see section **Data Watermarking**.

5.2 Can the Watermark Be Extracted from a Damaged Document?

DSC data watermarking is highly robust. Watermarks are not easily removed during transmission or use. Even if the data carrier is tampered with or damaged, there is a high probability that watermarks are extracted.

- If several pages are deleted from a document, the watermarks can still be extracted.
- If an image is rotated, cropped, scaled, or retouched, the watermarks can still be extracted as long as the deformation is small.

5.3 What Are the Requirements on the Source Data To Be Watermarked?

Watermark injection is a process to embed atomic watermark information into data with different features. The more source data features, the more complete watermark information can be embedded, and the higher the extraction success rate is. In addition, even if some data is missing, watermark extraction is not affected. The data to be watermarked must meet the following requirements:

- The source data must contain 1000 lines or more.

If the source data contains less than 1000 lines, the watermark may fail to be extracted due to insufficient features.

- You are advised to select a column with various data values. If all the values of the column can be enumerated, the extraction may fail due to insufficient features.

  Common columns that can be embedded with watermarks include the address, name, UUID, amount, and total amount.

6 Data Usage Audit

6.1 Which Types of Abnormal Events Can Be Identified by DSC?

Currently, DSC can only identify abnormal events in OBS.

DSC identifies sensitive data based on its identification rules and monitors events related to the sensitive data. You can check results in the event list and handle the abnormal events as needed. **Table 6-1** lists the abnormal events that can be identified by DSC.

Table 6-1 Abnormal events that can be identified by DSC

Type	Event
Unauthorized data access	• Access sensitive files without granted permissions. • Download sensitive files.
Abnormal data operations	• Update sensitive files. • Append data to sensitive files. • Delete sensitive files. • Copy sensitive files.

Type	Event
Abnormal data management	• When a bucket is added, the system detects that the bucket is a public read or a public read/write bucket. • When a bucket is added, the system detects that the access/ACL access permissions of a private bucket are granted for anonymous users or registered user groups. • The policy of a bucket containing sensitive files is changed or deleted. • The ACL of a bucket containing sensitive files is changed or deleted. • The cross-region replication configuration of a bucket containing sensitive files is modified or deleted. • The ACL of a sensitive file is modified or deleted.

